

MERYEM MOLDUR

GUIDELINES FOR EFL WRITING INSTRUCTION IN THE AGE OF AI: AN
INTERPRETIVE META-SYNTHESIS

MERYEM MOLDUR

A THESIS SUBMITTED

FOR

THE DEGREE OF MASTER OF ARTS

IN

TEACHING ENGLISH AS A FOREIGN LANGUAGE

İHSAN DOĞRAMACI BILKENT UNIVERSITY

ANKARA

SEPTEMBER 2024

2024

Dedicated to my grandparents...



Guidelines for EFL Writing Instruction in the Age of AI: An Interpretive Meta-Synthesis

The Graduate School of Education

of

İhsan Doğramacı Bilkent University

by

Meryem Moldur

In Partial Fulfilment of the Requirements for the Degree of

Master of Arts

in

Teaching English as a Foreign Language

The Department of English Language Teaching

Ankara

September 2024

İHSAN DOĞRAMACI BILKENT UNIVERSITY

GRADUATE SCHOOL OF EDUCATION

Guidelines for EFL Writing Instruction in the Age of AI: An Interpretive Meta-Synthesis

Meryem Moldur

September 2024

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Teaching English as a Foreign Language.

Asst. Prof. Dr. Necmi Akşit (Advisor)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Teaching English as a Foreign Language.

Asst. Prof. Dr. Tijen Akşit (Examining Committee Member)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Teaching English as a Foreign Language.

Prof. Dr. A. Cendel Karaman, METU (Examining Committee Member)

Approval of the Graduate School of Education

Prof. Dr. Alipaşa AYAS (Director)

ABSTRACT

Guidelines for EFL Writing Instruction in the Age of AI: An Interpretive Meta-Synthesis

Meryem Moldur

MA in Teaching English as a Foreign Language

Advisor: Asst. Prof. Dr. Necmi Akşit

September 2024

This study aimed to develop guidelines for writing instruction for instructors, particularly those teaching English as a Foreign Language (EFL) in higher education contexts, to address the challenges and opportunities in an increasingly AI-integrated educational landscape. To achieve this, an interpretive meta-synthesis approach was employed, focusing on empirical quantitative and qualitative studies published in academic journals that explore the use of AI and the development of writing skills and processes involved. The researcher analyzed and synthesized the purposes, results, and implications of these studies. Through qualitative analysis using open coding and axial coding, core categories were established through selective coding. These core categories were then considered as instructional guidelines, guiding the development of guidelines for writing instruction for educators developing writing skills in English. The framework also made connections with existing international competency standards, discussing alignment with current standards, emerging gaps that are arising, policy impacts, and future directions.

Keywords: Artificial intelligence, writing instruction, AI writing tools, AI-enhanced classes, writing instruction guidelines

ÖZET

Yapay Zeka Çağında Yabancı Dil Olarak İngilizce Yazma Eğitimi İçin Yönergeler:

Yorumlayıcı Bir Meta-Sentez

Meryem Moldur

Yabancı Dil Olarak İngilizce Öğretimi Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Necmi Akşit

Eylül 2024

Bu çalışma, giderek artan bir şekilde yapay zeka entegre edilmiş eğitim ortamında karşılaşılan zorlukları ve fırsatları ele almak için, özellikle yükseköğretim bağlamında Yabancı Dil olarak İngilizce (EFL) öğreten öğretmenler için yazma öğretimi kılavuzları geliştirmeyi amaçlamıştır. Bunu gerçekleştirmek için, yapay zeka kullanımını ve yazma becerilerinin gelişimini ve ilgili süreçleri araştıran, akademik dergilerde yayınlanmış ampirik nicel ve nitel çalışmalara odaklanan yorumlayıcı bir meta-sentez yaklaşımı kullanılmıştır. Araştırmacı, bu çalışmaların amaçlarını, sonuçlarını ve çıkarımlarını analiz edip sentezlemiştir. Açık kodlama ve eksenli kodlama kullanılarak yapılan nitel analiz yoluyla, seçici kodlama ile temel kategoriler oluşturulmuştur. Bu temel kategoriler daha sonra öğretim kılavuzları olarak ele alınmış ve İngilizce yazma becerilerini geliştiren eğitimciler için yazma öğretimi kılavuzlarının geliştirilmesine yön vermiştir. Çerçeve ayrıca mevcut uluslararası yeterlilik standartlarıyla bağlantılar kurmuş, mevcut standartlarla uyumu, ortaya çıkan boşlukları, politika etkilerini ve gelecekteki yönelimleri tartışmıştır.

Anahtar kelimeler: Yapay zeka, yazma eğitimi, yapay zeka yazma araçları, yapay zeka destekli dersler, yazma öğretimi kılavuzları

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my advisor, Asst. Prof. Dr. Necmi Akşit, thank you for your irreplaceable guidance and continuous support throughout this journey. Your expertise, encouragement, and trust have been influential in shaping both my research and my growth as a scholar. I would also like to extend my sincere thanks to the members of my thesis jury: Asst. Prof. Dr. Tijen Akşit and Prof. Dr. A. Cendel Karaman. Your insightful comments and valuable feedback have greatly improved the quality of this work. I am grateful for your time and expertise throughout this process.

To my husband, Serkan, and daughters, Duru and Doğa, I owe a debt of gratitude for your endless patience, understanding, and love. Your sacrifices and support have made this achievement possible, and I am eternally grateful for your presence in my life.

I also wish to thank my friends, Özden and Funda, for their constant encouragement and for being a source of joy and relief during challenging times. Your friendship has been a pillar of strength throughout this process.

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZET.....	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES	xix
LIST OF FIGURES	xxvi
CHAPTER 1: INTRODUCTION	1
Introduction	1
Background	1
Problem	7
Purpose.....	9
Research Questions	10
Significance.....	10
Definition of Key Terms	111
CHAPTER 2: REVIEW OF RELATED LITERATURE	113
Introduction	113
New Literacies	14
AI in Education	15
Key Concepts of AI	19
Artificial Intelligence	19
Generative Artificial Intelligence	20
Machine Learning	22
Deep Learning.....	24
Natural Language Processing	25

AI in Writing Instruction.....	27
Technology-Enhanced Approaches to Writing Instruction	29
Flipped Learning.....	29
Digital Storytelling	30
Computer-Assisted Language Learning (CALL)	31
Blended Learning.....	33
AI-Enhanced Approaches to Writing Instruction	34
Automated Writing Evaluation (AWE) Systems.....	34
Intelligent Tutoring Systems (ITS)	37
Plagiarism Detection.....	38
AI-Powered Writing Assistant in Language Instruction.....	41
Adaptive Learning Systems	42
Collaborative Writing Platforms with AI Features.....	44
Common Competency Frameworks for EFL Teachers	46
What is a competency framework?.....	46
British Council Continuing Professional Development (CPD) Framework	47
The TESOL International Association's Standards for ESL/EFL Teachers of Adults.....	48
The European Profiling Grid (EPG)	49
The Cambridge English Teaching Framework	51
The Equals Framework for Language Teacher Training & Development.....	53
EdTech Masterplan 2030	54
The BALEAP Competency Framework for Teachers of English for Academic Purposes (TEAP)	56
The ISTE (International Society for Technology in Education) Standards	59

The UNESCO Draft AI Competency Frameworks for Teachers and School	
Students	60
Previous Studies that Focus on AI, Writing Instruction, and Teacher Competencies	
.....	63
CHAPTER 3: METHOD	66
Introduction	66
Research Design.....	66
Sampling	67
Method of Data Collection.....	68
Methods of Data Analysis.....	70
Trustworthiness	74
CHAPTER 4: RESULTS	78
Introduction	78
Good Academic Conduct in Writing.....	78
Academic Dishonesty	79
Purpose.....	79
Results.....	80
Implications	82
Academic Integrity.....	83
Purpose.....	83
Results.....	85
Implications	88
Plagiarism	90
Purpose.....	90
Results.....	91

Implications	92
AI-generated text Authorship and Ownership	93
Purpose.....	93
Results.....	95
Implications	97
Transparency and Disclosure	99
Purpose.....	99
Results.....	99
Implications	100
Ethical Use	101
Purpose.....	101
Results.....	102
Implications	103
Assessment and Writing.....	105
Feedback	105
Purpose.....	105
Results.....	109
Implications	115
Assessment of Writing.....	122
Purpose.....	122
Results.....	123
Implications	125
Automated Evaluation/ Scoring.....	128
Purpose.....	128
Results.....	129

Implications	131
Evaluating and Enhancing the Quality and Impact of Academic Content.....	133
Performance and Impact	134
Capabilities of AI.....	134
Purpose	134
Results	136
Implications	139
Accuracy of AI.....	142
Purpose	142
Results	143
Implications	146
Proficiency and Originality.....	148
Purpose	148
Results	148
Implications	149
Reliability and Quality.....	150
Purpose	150
Results	151
Implications	151
Usefulness for Scholars	152
Purpose	152
Results	153
Implications	154
Impact Efficacy.....	155
Purpose	155

Results	155
Implications	156
Detection	157
Purpose.....	157
Results.....	160
Implications	165
Identification	171
Purpose.....	171
Results.....	173
Implications	177
Machine Learning	181
Purpose.....	181
Results.....	182
Implications	184
Output Characteristics.....	186
Quality and Accuracy of Academic Content	186
Purpose	186
Results	188
Implications	192
Authenticity	196
Purpose	196
Results	197
Implications	198
Creativity	201
Purpose	201

Results	202
Implications	203
Originality	205
Purpose	205
Results	206
Implications	208
User Behavior and Pedagogical Implications of AI-Assisted Writing	210
Acceptance	210
Purpose.....	210
Results.....	212
Implications	213
Adoption	215
Purpose.....	215
Results.....	216
Implications	218
Attitudes and Challenges	220
Purpose.....	220
Results.....	221
Implications	223
Self-efficacy	225
Purpose.....	223
Results.....	226
Implications	229
Users and Uses	232
Purpose.....	232

Results.....	234
Implications	236
Trust	239
Purpose.....	239
Results.....	240
Implications	240
Motivation.....	241
Purpose.....	241
Results.....	241
Implications	242
Independence and Proficiency	243
Purpose.....	243
Results.....	243
Implications	244
AI-Integrated Writing Instruction and Outcomes	245
Writing Skills	246
Purpose.....	246
Results.....	248
Implications	253
Writing Performance.....	258
Purpose.....	258
Results.....	260
Implications	262
Writing Process	264
Purpose.....	264

Results.....	266
Implications	269
AI-Enhanced Classroom	272
Purpose.....	272
Results.....	274
Implications	277
Writing Quality	280
Purpose.....	280
Results.....	281
Implications	283
Scaffolding.....	284
Purpose.....	284
Results.....	284
Implications	285
AI-Assisted Writing	286
Purpose.....	286
Results.....	287
Implications	289
Diverse Writing Genres and Modalities.....	290
Creative Writing.....	291
Purpose.....	291
Results.....	292
Implications	294
Business Communication.....	296
Purpose.....	296

Results.....	297
Implications	298
Multimodal Writing	299
Purpose.....	299
Results.....	300
Implications	301
News Writing	302
Purpose.....	302
Results.....	303
Implications	303
Narrative English Writing.....	304
Purpose.....	304
Results.....	304
Implications	305
Graphic Storytelling.....	306
Purpose.....	306
Results.....	306
Implications	307
Digital Stories	308
Purpose.....	308
Results.....	308
Implications	309
Picture-cued Writing.....	311
Purpose.....	311
Results.....	311

Implications	312
Multiple Genres	312
Purpose.....	312
Results.....	313
Implications	314
Linguistic and Discourse Features of AI-Generated Texts.....	315
Textual and Linguistic Features	316
Purpose.....	316
Results.....	318
Implications	321
Discourse Components	325
Purpose.....	325
Results.....	326
Implications	326
Style	327
Purpose.....	327
Results.....	328
Implications	329
Register	330
Purpose.....	330
Results.....	332
Implications	332
CHAPTER 5: DISCUSSION.....	333
Introduction	333
Overview of the Study	333

Guidelines for Effective Writing Instruction for EFL Writing Instructors in	
AI-Enhanced Classes	335
Guideline Area 1: Good Academic Conduct in Writing	335
Alignment and Gaps	338
Policy Impact and Future Directions	341
Guideline Area 2: Assessment and Writing	343
Alignment and Gaps	346
Policy Impact and Future Directions	349
Guideline Area 3: Evaluating and Enhancing the Quality and Impact of	
Academic Content	350
Alignment and Gaps	355
Policy Impact and Future Directions	358
Guideline Area 4: User Behavior and Pedagogical Implications of	
AI-Assisted Writing	360
Alignment and Gaps	363
Policy Impact and Future Directions	366
Guideline Area 5: AI-Integrated Writing Instruction and Outcomes.....	367
Alignment and Gaps	370
Policy Impact and Future Directions	373
Guideline Area 6: Diverse Writing Genres and Modalities	375
Alignment and Gaps	378
Policy Impact and Future Directions	379
Guideline Area 7: Linguistic and Discourse Features of AI-Generated	
Texts.....	381

Alignment and Gaps	384
Policy Impact and Future Directions	386
Implications for Practice	387
Implications for Further Research.....	389
Future Study Implications from the Synthesized Studies	389
Future Study Implications for Interpretive Meta-Synthesis Studies.....	391
Limitations	393
REFERENCES.....	395
Appendix A	427

LIST OF TABLES

Table		Page
1	The Flow Diagram Structure by PRISMA.....	69
2	The Core/Overarching Categories.....	72
3	The Evaluative Criteria Used for Validity and Reliability in Qualitative Research.....	74
4	Academic Dishonesty: Purpose	79
5	Academic Dishonesty: Results	80
6	Academic Dishonesty: Implications.....	82
7	Academic Integrity: Purpose	83
8	Academic Integrity: Results	85
9	Academic Integrity: Implications.....	88
10	Plagiarism: Purpose	90
11	Plagiarism: Results.....	91
12	Plagiarism: Implications	92
13	AI-Generated Text Authorship and Ownership: Purpose	94
14	AI-Generated Text Authorship and Ownership: Results.....	95
15	AI-Generated Text Authorship and Ownership: Implications	97
16	Transparency and Disclosure: Purpose.....	99
17	Transparency and Disclosure: Results.....	99
18	Transparency and Disclosure: Implications.....	100
19	Ethical Use: Purpose	101
20	Ethical Use: Results.....	102
21	Ethical Use: Implications.....	103

22	Feedback: Purpose	106
23	Feedback: Results.....	109
24	Feedback: Implications.....	115
25	Assessment of Writing: Purpose.....	122
26	Assessment of Writing: Results.....	123
27	Assessment of Writing: Implications.....	125
28	Automated Evaluation or Scoring: Purpose	129
29	Automated Evaluation or Scoring: Results.....	130
30	Automated Evaluation or Scoring: Implications.....	131
31	Capabilities of AI: Purpose	134
32	Capabilities of AI: Results	136
33	Capabilities of AI: Implications	139
34	Accuracy of AI: Purpose	142
35	Accuracy of AI: Results	143
36	Accuracy of AI: Implications	146
37	Proficiency and Originality: Purpose	148
38	Proficiency and Originality: Results	149
39	Proficiency and Originality: Implications	149
40	Reliability and Quality: Purpose	150
41	Reliability and Quality: Results	151
42	Reliability and Quality: Implications	152
43	Usefulness for Scholars: Purpose	152
44	Usefulness for Scholars: Results	153
45	Usefulness for Scholars: Implications	154
46	Impact Efficacy: Purpose	155

47	Impact Efficacy: Results	156
48	Impact Efficacy: Implications	156
49	Detection: Purpose	157
50	Detection: Results	160
51	Detection: Implications	166
52	Identification: Purpose	171
53	Identification: Results	173
54	Identification: Implications	178
55	Machine Learning: Purpose	181
56	Machine Learning: Results	182
57	Machine Learning: Implications	184
58	The Quality and Accuracy of Academic Content: Purpose ..	187
59	The Quality and Accuracy of Academic Content: Results ...	189
60	The Quality and Accuracy of Academic Content: Implications	192
61	Authenticity: Purpose	196
62	Authenticity: Results	197
63	Authenticity: Implications	199
64	Creativity: Purpose.....	201
65	Creativity: Results.....	202
66	Creativity: Implications	203
67	Originality: Purpose	205
68	Originality: Results	206
69	Originality: Implications	208
70	Acceptance: Purpose	211

71	Acceptance: Results	212
72	Acceptance: Implications	213
73	Adoption: Purpose.....	215
74	Adoption: Results	216
75	Adoption: Implications	218
76	Attitudes and Challenges: Purpose	220
77	Attitudes and Challenges: Results	221
78	Attitudes and Challenges: Implications	223
79	Self-Efficacy: Purpose	225
80	Self-Efficacy: Results.....	226
81	Self-Efficacy: Implications	229
82	Users and Uses: Purpose.....	232
83	Users and Uses: Results.....	234
84	Users and Uses: Implications	236
85	Trust: Purpose	239
86	Trust: Results	239
87	Trust: Implications	240
88	Motivation: Purpose	241
89	Motivation: Results	241
90	Motivation: Implications.....	242
91	Independence and Proficiency: Purpose	243
92	Independence and Proficiency: Results	244
93	Independence and Proficiency: Implications	244
94	Writing Skills: Purpose.....	246
95	Writing Skills: Results.....	248

96	Writing Skills: Implications	253
97	Writing Performance: Purpose	258
98	Writing Performance: Results	260
99	Writing Performance: Implications	262
100	Writing Process: Purpose	264
101	Writing Process: Results.....	267
102	Writing Process: Implications	269
103	AI-Enhanced Classroom: Purpose	273
104	AI Enhanced Classroom: Results	274
105	AI Enhanced Classroom: Implications	278
106	Writing Quality: Purpose	281
107	Writing Quality: Results	282
108	Writing Quality: Implications	283
109	Scaffolding: Purpose	284
110	Scaffolding: Results	284
111	Scaffolding: Implications	285
112	AI Assisted Writing: Purpose	286
113	AI Assisted Writing: Results	287
114	AI Assisted Writing: Implications	289
115	Creative Writing: Purpose	291
116	Creative Writing: Results	292
117	Creative Writing: Implications	294
118	Business Communication: Purpose	296
119	Business Communication: Results	297
120	Business Communication: Implications	298

121	Multimodal Writing: Purpose	300
122	Multimodal Writing: Results	300
123	Multimodal Writing: Implications	301
124	News Writing: Purpose.....	302
125	News Writing: Results.....	303
126	News Writing: Implications	303
127	Narrative Writing: Purpose	304
128	Narrative Writing: Results	304
129	Narrative Writing: Implications	305
130	Graphic Storytelling: Purpose	306
131	Graphic Storytelling: Results	306
132	Graphic Storytelling: Implications	307
133	Digital Stories: Purpose	308
134	Digital Stories: Results	308
135	Digital Stories: Implications	309
136	Picture-Cued Writing: Purpose	311
137	Picture-Cued Writing: Results	311
138	Picture-Cued Writing: Implications	312
139	Multiple Genres: Purpose	313
140	Multiple Genres: Results	313
141	Multiple Genres: Implications	314
142	Textual and Linguistic Features: Purpose	316
143	Textual and Linguistic Features: Results	318
144	Textual and Linguistic Features: Implications	322
145	Discourse Components: Purpose	325

146	Discourse Components: Results	326
147	Discourse Components: Implications	326
148	Style: Purpose	327
149	Style: Results	328
150	Style: Implications	329
151	Register: Purpose	331
152	Register: Results	331
153	Register: Implications	332



LIST OF FIGURES

Figure		Page
1	Relationship between AI, ML, DL, and NLP	22
2	The Flow Diagram Followed in the Data Collection of the Study	69
3	Layered Open Coding: Parent & First Level Sample Categories and Codes	71
4	Relational Coding: Sample Categories and Codes Based on Recurring Patterns	72
5	Good Academic Conduct in Writing.....	78
6	Assessment and Writing	105
7	Evaluating and Enhancing the Quality and Impact of Academic Content	134
8	User Behavior and Pedagogical Implications of AI-Assisted Writing	210
9	AI-Integrated Writing Instruction and Outcomes	245
10	Diverse Writing Genres and Modalities.....	291
11	Linguistic and Discourse Features of AI-Generated Texts ...	315

CHAPTER 1: INTRODUCTION

"The technology itself is not transformative. It is the school, the pedagogy, that is transformative." Tanya Byron

Introduction

The introduction chapter of this study provides an overview of the research. It begins with background information to contextualize the study. The chapter then presents the specific problem being addressed, outlines the purpose of the research, and states the research questions guiding the study. Next, the significance of the study is discussed. Finally, the chapter concludes with definitions of key terms used throughout the study.

Background

The rapid pace of technological advancements in the 21st century, particularly the rise of Artificial Intelligence, has profoundly impacted education, affecting how we teach, learn, and interact with technology. AI has opened up new possibilities for integrating AI-driven tools and resources into teaching and learning practices. AI's influence on education is multifaceted and far-reaching, continuing to shape various aspects of teaching and learning across disciplines. According to UNESCO (2021), AI has the potential to address some of the biggest challenges in education today and could support the development of more inclusive and effective learning systems. The technology offers opportunities for personalized learning, intelligent tutoring systems, and automated administrative tasks, which can significantly enhance the educational experience for both students and educators.

Furthermore, the U.S. Department of Education (2023) emphasizes AI's role in promoting educational equity and accessibility. AI-powered tools can provide real-time language translation, assist students with disabilities, and offer adaptive learning experiences that cater to individual student needs. However, the report also highlights the importance of ensuring that AI implementation in education is guided by ethical considerations and a commitment to protecting student privacy and data security. Similarly, the European Commission (2022) underscores the transformative potential of AI in teacher education and professional development. AI can facilitate data-driven decision-making in educational policy and practice, enabling more targeted and effective interventions. The Commission also stresses the need for developing AI literacy among all educational stakeholders to maximize benefits and mitigate potential risks associated with AI integration in education.

Despite these promising applications, challenges remain. There is a recognized need for practical AI application integration across various educational contexts, from K-12 to higher education and professional training. Additionally, as AI technologies continue to evolve rapidly, there is an ongoing requirement for educators and institutions to stay informed about the latest developments and best practices in AI-enhanced education. Many studies underscore AI's positive impacts of AI on higher education, especially in enhancing student engagement and academic achievement, but concerns remain about educator readiness and ethical considerations (Alasadi & Baiz, 2023; Ateeq et al., 2024; Harry, 2023; Nguyen, 2023; Salloum, 2024; Sánchez-Ruiz et al., 2023).

As AI continues to reshape education, one area where its influence is particularly significant is writing instruction. AI offers numerous opportunities, particularly for enhancing feedback mechanisms, improving student engagement,

and elevating writing quality. In an investigation of the use of ChatGPT and other AI tools in English for Academic Purposes (EAP) classrooms, Wu (2024) highlights AI's potential to provide personalized feedback and improve discourse writing performance among English major students, though challenges remain in understanding academic language and specialized terminology. Similarly, Shen et al. (2023) found that an AI-based writing evaluation system, Pigai, significantly improved writing accuracy and complexity, particularly benefitting higher-level students. Furthermore, Roquet et al. (2024) examined adjunct AI instruction in EFL academic writing, and the results demonstrated that AI-enhanced courses can improve lexical diversity and accuracy in student writing, though this depends on students' initial language exposure. Further research by Hutson et al. (2024) on the integration of AI tools like ChatGPT in English composition demonstrated a shift from initial student skepticism to effective practical application, suggesting a hybrid pedagogical model that combines traditional methods with AI-driven enhancements.

However, writing instruction faces several significant hurdles across different educational contexts, including barriers to effective teaching methodologies, student engagement, and curriculum development. One of these challenges becomes evident in a study by Dickinson (2018), which highlights that English as an Additional Language (EAL) learners often encounter writing difficulties and advocates for Universal Design for Learning (UDL) to address these barriers through diversified instructional strategies. In another study, Kovanen et al. (2020) emphasize the difficulties STEM educators face in integrating writing instruction due to a lack of pedagogical training, large class sizes, and heavy content requirements. This study suggests adopting writing-as-process models to enhance learning. Moreover, the study by Dong et al. (2019) discuss the impact of high-stakes testing culture on

writing instruction in China, where teachers face the challenge of developing genuine writing skills while being pressured to perform well on standardized tests. In another study, Johnson (2018) explores the curriculum development challenges in academic writing programs, particularly in engaging students and enabling the transfer of writing knowledge across contexts. These diverse challenges across different educational contexts underscore the complexity of writing instruction and highlight the need for innovative, adaptive approaches to address the multifaceted nature of these issues while enhancing the effectiveness of writing pedagogy.

AI has the potential to significantly address and overcome many challenges in writing instruction by providing enhanced feedback mechanisms, supporting self-regulated learning, and improving the quality of student writing. A systematic review study by Alsaedi (2024) on the advantages and challenges of ChatGPT in writing instruction concludes that AI tools, like ChatGPT, can enhance writing instruction by improving efficiency, creativity, and personalized learning experiences while reducing teachers' workload through automated evaluation and feedback. This result suggests that AI can be effectively used to improve basic writing skills while requiring human oversight for more complex tasks (An et al., 2023; Lin et al., 2022; Wongras & Tanantong, 2023; Wu et al., 2022, as cited in Aljuaid, 2024).

Another study examines the impact of AI tools on English major students' discursive writing performance and finds that AI can provide personalized feedback and improve expression skills, though there are still limitations in understanding academic language and specialized terminology (Wu, 2024). The implications of AI-based natural language production for writing instruction are numerous, and strategies for integrating these tools must ensure students remain engaged in the

writing process (An et al., 2023; Anson & Straume, 2022; Lin et al., 2022; Wongras & Tanantong, 2023; Wu et al., 2022).

In light of these developments and challenges, this thesis aims to explore how recent studies in AI research and writing inform the development of guidelines for EFL writing instruction. By synthesizing relevant studies, this meta-synthesis seeks to bridge the gap between technological advancements and their practical applications in educational contexts. Through systematic analysis and interpretation of the literature, this study aims to provide a theoretically grounded and practically applicable outlook that can guide educators, curriculum developers, and policymakers in effectively integrating AI technologies in writing instruction.

As AI continues to influence educational practices, the field of writing instruction has undergone significant transformations, marked by changes in pedagogical approaches, the integration of technology, and shifting educational paradigms. This evolution is evident in both the methods of instruction and the tools used for writing.

Lang (2018) provides insights into the dynamic nature of writing instruction by analyzing five years of instructor feedback. The study reveals that while new instructors initially use a more consistent vocabulary, their feedback evolves over time, reflecting both local and broader instructional trends. This observation underscores writing pedagogy's adaptive nature and responsiveness to changing educational contexts. Building on this idea of evolution, Pennington (2013) discusses broader trends in writing pedagogy. The author highlights several significant shifts, including the movement from U.S.-centric approaches to more global perspectives, acknowledging the increasingly interconnected nature of education. Additionally, Pennington notes the disarticulation of writing from reading, a trend that has

implications for how we conceptualize literacy instruction. Perhaps most relevantly, the study addresses the challenges of maintaining author identity and text ownership in digital contexts. This concern has become increasingly relevant with the rise of AI-assisted writing tools.

The technological aspect of this evolution is further explored by Mugo et al. (2014), who offer a historical reflection on writing technologies. The study illustrates how these advancements have shaped instructional practices over time by tracing the development of writing systems, surfaces, and instruments. This historical perspective provides valuable context for understanding current technological integrations in writing instruction, including AI. The study by Jett (2013) brings this technological evolution into focus by exploring the transition from traditional pen-and-paper methods to digital writing in one-to-one laptop settings. This study reveals how technology not only changes the medium of writing but also influences pedagogical strategies and student engagement. The shift to digital writing environments sets the stage for understanding how AI might further transform these practices.

Furthermore, Smith (2017) advocates for a principled revolution in writing instruction, emphasizing the need for adaptable and context-specific practices. This call for revolution is particularly relevant in the context of AI integration, as it highlights the importance of bridging the gap between research and classroom implementation. Smith argues that the integration of new technologies, including AI, should be guided by comprehensive pedagogical principles rather than simply following technological trends. Finally, according to the results of a systemic review by Aljuaid (2024), a study aiming to investigate the impact of Artificial Intelligence (AI) tools on academic writing instruction in higher education, there is a need for

deeper and more comprehensive studies on how AI tools can be integrated into writing curricula. Such studies could help to enhance student engagement, improve writing proficiency, and promote academic integrity while addressing concerns over privacy, bias, and originality.

Collectively, these studies highlight the dynamic nature of writing instruction, constantly adapting to emerging technologies, pedagogical advances, and global contexts. This ongoing evolution underscores the need for writing instructors to develop competencies in integrating AI tools into their instruction. The study by Lee et al. (2024) states that in addition to staying up to date on AI advancements, the higher education industry must remain committed to continuous study and dialogue on optimal approaches to address AI's impact. By embracing AI, educators can address persistent challenges in writing instruction while enhancing student engagement and leveraging new opportunities for personalized learning.

Problem

AI has the potential to enhance language learning outcomes and contribute to effective teaching practices, but further research is needed to understand this innovation and its features, benefits, and potential impact. It is also essential to understand the process of AI adoption and its acceptance for the optimal integration of AI technologies into writing instruction (Malik et al., 2023; Nazari et al., 2021; Ouyang et al., 2022). Despite the rapid advancement and growing integration of AI technologies in educational settings, particularly in writing instruction, there remains a lack of comprehensive studies examining AI-focused research conducted within the context of writing and writing instruction. This gap in research could impact learning outcomes, as some aspects may remain unresolved, such as uncertainty about AI's impact on students' writing skills development (Aljuaid, 2024; Selim, 2024; Storey,

2023), questions about AI's effect on teacher's role in writing instruction (Hutson et al., 2024; Luo & Qiu, 2024; Nazim, 2024; Trần, 2024), and concerns regarding potential overdependence on AI (Gayed et al., 2022; Ouyang et al., 2022). There are also debates over how AI affects creativity and critical thinking (Liang, 2023; Livberber & Ayvaz, 2023; Malik et al., 2023), as well as a pressing need for guidelines on the ethical and effective integration of AI in writing curricula (Aljuaid, 2024; Afshan, 2024; Balta, 2023; Eva et al., 2023; Farina & Stevenson, 2024; Mahona et al., 2023).

Moreover, there is a risk of misusing or underutilizing AI in writing instruction (Cardon et al., 2023; Hutson, 2024; Rahayu et al., 2024), potentially leading to suboptimal learning outcomes or missed opportunities for pedagogical innovation in pedagogy. Thus, it is essential to understand AI's role in writing instruction to better inform educational policies, teacher training, and curriculum development.

By systematically identifying and categorizing key concepts in the existing body of research on AI in writing instruction, this study aims to provide a comprehensive overview of the current state of AI integration in this field. Through exploring the relationships between these identified key concepts, the research seeks to offer a deeper and broader understanding of AI's role in writing pedagogy. This analysis will shed light on the potential benefits, drawbacks, challenges, and opportunities associated with AI integration in writing instruction.

Additionally, these findings serve as a needs assessment, identifying areas where instructors require support and development. By utilizing the insights gained from examining the interconnections between various key concepts, this research aims to develop guidelines tailored specifically for EFL writing instructors,

providing valuable guidance on the effective and appropriate integration of AI technologies into writing pedagogy. Ultimately, this study intends to contribute to the ongoing conversation about AI's role, offering critical insights into how AI technologies can be best integrated into writing instruction.

Purpose

The purpose of this study is to explore how recent studies conducted within the context of AI research inform the development of guidelines for effective EFL writing instruction. Specifically, the study aims to

1. identify key concepts that emerge in these studies,
2. examine how key concepts interrelate,
3. analyze the core categories that emerge from these relationships in AI research related to writing instruction
4. develop guidelines for writing instruction

The ultimate goal is to inform educational practice and policy and to ensure that AI integration in writing instruction is not only effective but also aligned with the needs of both educators and students. These guidelines will address the evolving needs of language learners in an increasingly technology-driven educational environment and are designed to guide instructors, curriculum developers, and educational institutions aiming to enhance their writing instruction practices in the age of AI. The framework aims to utilize existing teacher competency standards. It also identifies potential gaps in current standards, explores the policy impact and future directions of AI integration into writing instruction, and ultimately intends to equip EFL writing educators with the essential skills and knowledge to function effectively, efficiently, and ethically.

Research Questions

This study aims to address the following research question and sub-questions:

How do recent studies conducted within the context of AI research inform the development of guidelines for EFL writing instruction?

- a. What key concepts emerge in these studies?
- b. How do these key concepts in these studies interrelate?
- c. What core categories emerge from the analysis of these relationships?
- d. How do these core categories inform the development of guidelines for writing instruction?

Significance

A major contribution of this research is the development of practical, research-based guidelines specifically tailored for EFL writing instructors. These guidelines aim to bridge the gap between AI advancements and pedagogical practice, ensuring that educators have the tools needed to integrate AI into EFL writing instruction effectively. Moreover, the significance of this study also lies in its ability to address the specific challenges educators face when using AI tools in writing instruction, with an emphasis on AI's complementary role in enhancing rather than replacing traditional teaching methods.

The significance of this interpretive meta-synthesis also stems from its focus on gaps in research and synthesizing the outcomes of the studies conducted on AI integration in writing instruction. Although AI is increasingly being integrated into educational practices, there is a need to systemically analyze and synthesize the use and impact of AI on writing instruction. By providing a deeper understanding of AI's role in writing pedagogy, the study offers a comprehensive platform for synthesizing

the purposes, results, and implications of the studies analyzed, thus contributing to a more robust body of knowledge.

In addition, this study is also significant for its potential to influence teacher education and professional development. By identifying key areas of focus, the research will inform teacher training programs and professional development initiatives, ensuring that writing instructors are equipped with the knowledge, skills, and strategies to use AI effectively in their writing classrooms. This is especially important as educators are expected to function efficiently in the increasingly complex educational settings where technology and pedagogy intersect.

The insights from this study will also guide curriculum development and educational policy. Policymakers and curriculum developers will benefit from the study's comprehensive analysis and synthesis of AI's role in writing instruction, as it provides evidence-based outcomes for AI adoption and ethical use of AI in writing curricula. This will support the development of well-informed policies that promote the responsible and effective integration of AI in education.

Finally, while the study focuses on EFL writing instruction, its findings will have broader educational implications. The insights gained from this research can inform discussions about AI integration across other language skills and disciplines, maximizing the benefits of AI across various educational settings.

Definition of Key Terms

Artificial Intelligence (AI) systems: are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate

with varying levels of autonomy by means of knowledge modeling and representation and by exploiting data and calculating correlations (UNESCO, 2023).



CHAPTER 2: REVIEW OF RELATED LITERATURE

Introduction

The field of writing instruction has been undergoing a big transformation with the introduction of Artificial Intelligence (AI) technologies. This chapter aims to explore the complex relationship between AI and writing instruction, with a particular focus on English as a Foreign Language (EFL) contexts. The primary objective is to identify recurring concepts in AI research related to writing instruction and to understand how these overarching concepts can contribute to the development of guidelines for writing instruction for EFL teachers.

To this end, this chapter will explore several key areas of AI and writing instruction. The chapter will begin by examining the broader context of AI in educational settings, a brief history of AI in education, and key concepts related to AI. Then, the focus will be on AI in writing instruction, and this section will delve into both technology-enhanced and AI-enhanced approaches to writing instruction and cover topics such as automated writing evaluation systems, intelligent tutoring systems, and AI-powered writing assistants. Later, there will be a discussion of what competencies and competency frameworks are before delving into a review of the existing competency frameworks for language teachers/students, with a particular focus on those that address technology integration or digital skills and writing instruction. Special attention will be given to frameworks that incorporate AI-related competencies. Finally, the study will examine relevant research studies that have explored the intersection of AI, writing instruction, and teacher competencies in the EFL context.

New Literacies

The concept of *new literacies* refers to the evolving set of skills, strategies, and dispositions that individuals need to effectively navigate and engage with the rapidly changing landscape of information and communication technologies in our digital age (Leu et al., 2013). In educational and language instruction contexts, "new literacies" extend beyond the traditional definitions of reading and writing and encompass a broader spectrum of digital, technological, and information-based competencies (Lankshear & Knobel, 2011). These competencies are increasingly essential for students and professionals alike as they help them adapt to the demands of a digitally driven world.

Some examples of these *new literacies* include:

- *Digital Literacy* is the ability to use digital technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills (Bawden, 2008).
- *Multimodal Literacy* is the ability to interpret, create, and communicate meaning through various modes, such as visual, audio, gestural, and spatial representations, in addition to traditional text (Kress, 2003).
- *Critical Digital Literacy* is the capacity to critically analyze, evaluate, and create digital content while understanding the social, cultural, and political implications of digital technologies (Pangrazio, 2016).
- *AI Literacy* is the knowledge and skills required to critically understand AI technologies, their applications, limitations, and societal impacts, as well as the ability to use AI tools effectively and responsibly (Long & Magerko, 2020).
- *Data Literacy* is the ability to read, understand, create, and communicate data as information (Wolff et al., 2016).

The concept of new literacies offers a valuable framework for introducing AI in education, as it emphasizes the evolving skills and competencies required to function effectively in the digital age for AI introduction. By introducing these new literacies into the curriculum, educators can not only equip students and teachers with the essential skills within the AI landscape but also foster a critical, responsible, and ethical mindset towards AI technologies.

AI in Education

With the advent and expansion of AI tools and applications in education, a new field has gradually emerged during the past few decades: artificial intelligence in education (AIED). According to many studies, the application of AI in educational settings is a rapidly developing field (EDUCAUSE, 2022; Lin et al., 2023; Nasir et al., 2024; Pranav et al., 2024; Yalçınalp et al., 2024). AIED aims to design machines that mimic human systems to support learning and teaching (Conati et al., 2018; Holmes & Tuomi, 2022; Tahiru, 2021). The implementation of these AI systems in educational institutions has the potential to fundamentally impact classroom instruction and scientific investigation methods positively (Chen et al., 2020; Harry, 2023; Sytnyk & Podlinskyayeva, 2024), improve students' learning experience and the overall quality of education (Srinivasa et al., 2022), analyze student input and provide corrective feedback instantly, generate automatic scoring and formative assessments, and help students with revisions during the learning process (Zhang & Aslan, 2020), support students with special needs and at the system level, incorporate predictive analysis to reduce dropout or assess new skillsets (Vincent-Lancrin & van der Vlies, 2020).

Many aspects of education have undergone significant change due to AI technologies that many educators and students have been commonly using. Several

studies highlight AI's potential for transformation in education, especially in higher education and engineering (Cota-Rivera et al., 2024; Holmes & Tuomi, 2022; Saaïda, 2023). These studies emphasize the benefits of AI-driven adaptive systems, which can potentially tailor education to individual student needs to improve engagement and retention. According to UNESCO (2023) and the U.S. Department of Education (2023) reports, AI can enhance personalized learning, improve accessibility for students with disabilities and multilingual learners, and streamline teaching tasks. While promoting lifelong learning capabilities, these technologies can also boost students' creativity and critical thinking skills. AI applications in education include intelligent tutoring systems, automated grading, personalized learning experiences, and the introduction of Virtual Reality for immersive learning (Crompton & Burke, 2023; Ali et al., 2024). In the book *Transforming Education with the Power of Artificial Intelligence*, Cota-Rivera et al. (2024) highlight that AI in education personalizes learning, analyzes strengths/weaknesses, adapts content, and enhances engagement. It also transforms higher education through audiovisual materials, mentoring, and information dissemination for students and staff. Additionally, AI's impact extends to special education as well. Singh and Jain (2024) explore the theoretical foundations and practical applications of AI in special education, and the results highlight how AI can enhance learning outcomes and promote inclusion for students with special needs by providing personalized education and real-time feedback.

However, the use of AI in education also presents significant challenges. Multiple researchers have identified similar drawbacks and limitations (Alasadi & Baiz, 2023; Nguyen, 2023; Seo et al., 2021; UNESCO, 2023). However, Nguyen (2023) makes a comprehensive summary of these challenges as follows:

- Related to *privacy and ethical concerns*, data collection and potential misuse raise significant privacy issues exacerbated by the lack of comprehensive regulations.
- There are also *high costs and implementation barriers*. The expenses associated with AI implementation can restrict its feasibility in many educational settings.
- *Lack of human interaction* is another issue. AI's inability to replicate the nuanced relationship between teachers and students poses a challenge, potentially leading to a decline in essential soft skills better developed through human interaction.
- *Limitations in providing personalized feedback and fostering critical thinking* suggest that while AI can offer personalized learning experiences, it still faces limitations in nurturing higher-order thinking skills.
- Related to *data accuracy and bias*, AI's reliance on accurate data can lead to the spread of biases if the input data is flawed or biased.
- *Lack of explainability* is another concern that has been raised. The "black box" nature of some AI systems makes it difficult to understand and manage their decision-making processes.
- There is also *limited contextual understanding*. AI may struggle with nuanced interpretation of complex or context-dependent information.
- *Digital literacy requirements* mean that learners and teachers must develop new digital literacies to utilize AI tools in education effectively (Nguyen, 2023).

In order to mitigate the risks and challenges caused by the use of AI in educational settings, studies emphasize fairness, accountability, transparency, and

explainability (Kazim & Koshiyama, 2020), ethical design, equity, privacy, and an ethical framework to navigate the integration of AI in educational settings (The Institute for Ethical AI in Education [IEAIED] 2021), the redefinition of teachers' fundamental duties, the instruction of students in the sensible use of AI, and the promotion of effective regulation of AI's deployment in education (Bu, 2022), the ethical codes that include AI and four basic building blocks of the ethical attitude of AI in education (autonomy, privacy, trust, and responsibility) (Gartner & Krašna, 2023). These studies and reports on mitigating risks in AI-based education emphasize various ethical considerations and frameworks to guide the responsible integration and use of AI technologies in educational settings and offer solutions to ensure responsible AI use.

In the report, *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations*, the U.S. Department of Education, Office of Educational Technology (2023) recommends a comprehensive approach to mitigate risks while leveraging AI's potential. It emphasizes keeping humans involved in AI-driven processes, aligning AI models with educational visions, and designing AI systems based on modern learning principles. The report stresses the importance of building trust in AI technologies and actively involving educators in understanding and implementing AI. It calls for focused research on AI's adaptability to diverse educational contexts and enhancing its safety. Finally, it advocates for the development of education-specific guidelines and guardrails to ensure the safe and effective use of AI in educational settings. To address the challenges, UNESCO (2023) recommends developing clear policies and guidelines for AI use in educational settings, providing comprehensive training for faculty on AI integration, and educating students about appropriate AI use and academic integrity. The

European Commission (2022) also emphasizes the need for ethical guidelines on AI and data use in teaching and learning. These reports stress the importance of ongoing research and collaboration between institutions, policymakers, and technology companies to develop best practices and ensure equitable educational access to AI technologies. These recommendations collectively aim to create a balanced approach that can harness AI's benefits while also addressing its potential risks in education. Some of these requirements clearly necessitate teachers to have the competencies to guide the students in an ethical and responsible use of AI in their learning practice.

To wrap up, AI holds great potential for transforming and enriching education with its various uses in numerous aspects of educational practices. However, it is crucial to thoughtfully address its limitations and ethical concerns to ensure that its implementation in education is effective and responsible. In order to achieve this, teacher competencies are crucial because they can enable educators to leverage AI tools to enhance learning outcomes while maintaining pedagogical integrity, addressing ethical concerns, and fostering students' critical writing skills.

Key Concepts of AI

Artificial intelligence

“AI refers to machine-based systems that can, given a set of human-defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems interact with us and act on our environment directly or indirectly. Often, they appear to operate autonomously and can adapt their behavior by learning about the context” (UNICEF, 2021, p. 16).

The phrase “artificial intelligence” was initially used at a 1956 workshop at Dartmouth College, referring to the “science and engineering of constructing intelligent devices, notably intelligent computer programs.” " (McCarthy et al., 2006,

p. 2). This was the beginning of AI research. Later, with the help of significant contributions to the field, there was the development of Symbolic AI, Expert Systems, and Knowledge-Based Systems. During the 1990s, the research into the field stopped, causing what was called the AI Winter. With Machine Learning Resurgence in the 2000s and Deep Learning and Big Data in the 2010s, the field of AI showed remarkable results and has begun to receive significant interest and investment from both the private and public sectors due to its potential to revolutionize various industries and improve the quality of life for people around the world (Haenlein & Kaplan, 2019).

Generative Artificial Intelligence

Generative Artificial Intelligence refers to a subset of artificial intelligence that focuses on creating or generating new data, primarily in the form of text, images, or other media, using neural networks and machine learning techniques. These models are designed to understand patterns and structures within existing data and use that understanding to produce novel content that is often indistinguishable from human-created content (Brown et al., et al. 2020).

Some developments have collectively advanced the field of Generative AI, each contributing to improved text generation, understanding, and manipulation capabilities that are crucial for applications in education and writing instruction.

First of all, there is *Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)*. Early architectures used for sequence generation, including text. LSTMs, introduced by Hochreiter and Schmidhuber (1997), were particularly important for handling long-term dependencies in language.

There is also *Generative Adversarial Networks (GANs)*. Introduced by Goodfellow et al. in 2014, GANs use two competing neural networks—a generator

and a discriminator—to create increasingly realistic data. While primarily used for image generation, GANs have also been applied to text generation (Goodfellow et al., 2014).

Variational Autoencoders (VAEs) are another important development. Proposed by Kingma & Welling (2013), VAEs are generative models that learn to encode and decode data, which are helpful for text generation and style transfer in writing.

Sequence-to-Sequence (Seq2Seq) Models were introduced by Sutskever et al. (2014), these models revolutionized machine translation and formed the basis for many text generation tasks.

Attention Mechanism was proposed by Bahdanau et al. (2014), attention allows models to focus on relevant input parts when generating output, significantly improving text quality.

Transformer Architecture was introduced by Vaswani et al. (2017), this architecture uses self-attention mechanisms and has become the foundation for most modern language models.

GPT (Generative Pre-trained Transformer) series were developed by OpenAI, these models have evolved from GPT (2018) to GPT-4 (2023). They use the transformer architecture and are pre-trained on vast text datasets, demonstrating impressive language generation capabilities (Brown et al., 2020).

BERT (Bidirectional Encoder Representations from Transformers) were introduced by Devlin et al. (2018), BERT and its variants improved natural language understanding, indirectly benefiting text generation tasks.

T5 (Text-to-Text Transfer Transformer) was proposed by Raffel et al. (2019). T5 frames all NLP tasks as text-to-text problems, offering a unified approach to various language tasks, including generation.

There is also *Few-shot and Zero-shot Learning*. Recent large language models like GPT-3 have demonstrated the ability to perform tasks with few or no specific examples, opening new possibilities for adaptive learning and writing assistance (Brown et al., 2020).

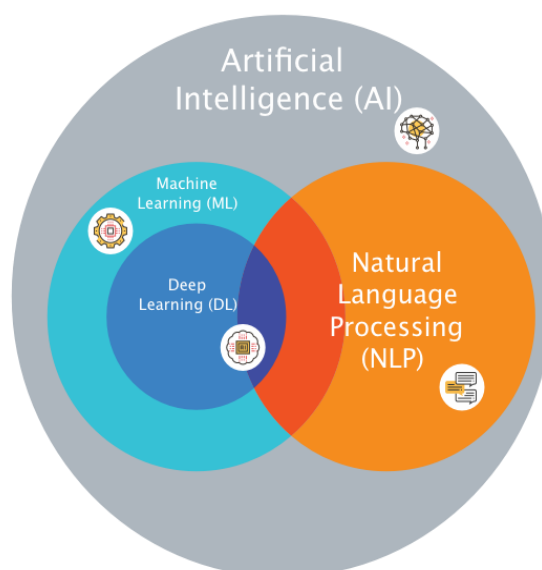
Multimodal Models are recent developments like DALL-E (OpenAI, 2021) and GPT-4 (OpenAI, 2023) integrate text and image understanding, potentially revolutionizing multimedia content creation in educational contexts.

Machine Learning

The following figure (Figure 1) shows the relationship between the key subsets or fields of AI. Understanding these relationships helps educators and researchers choose the appropriate techniques and tools for specific educational applications and understand the broader impacts these concepts create.

Figure 1

Relationship between AI, ML, DL, and NLP



Note. From *Real-World Natural Language Processing* (p.9), by M. Hagiwara, 2021, Manning Publications (<https://www.manning.com>). Copyright 2021 by Manning Publications.

AI is the broadest concept, and it encompasses any technique that enables machines to mimic human behavior. Machine Learning (ML), on the other hand, machine learning (ML) is a subset of artificial intelligence that focuses on developing algorithms and statistical models that enable computer systems to improve their performance on a specific task through experience (Mitchell, 1997).

Machine Learning has numerous applications in education. It has the potential to revolutionize teaching and learning processes. ML algorithms enable personalized learning by adapting educational content and pacing to individual student needs, thereby improving learning outcomes (Wang, 2022). Intelligent Tutoring Systems leverage ML to provide personalized feedback and guidance to students, which effectively enables one-on-one tutoring (VanLehn, 2011). In the realm of predictive analytics, ML models can forecast student performance, identify at-risk students and allow for timely interventions (Liz-Domínguez et al., 2019), facilitate automated grading, assist in the evaluation of both objective assessments and written responses, which saves valuable time for educators (Shermis & Burstein, 2013). According to Wan and Niu (2018), adaptive assessment, powered by ML, can adjust question difficulty based on student responses with more accurate evaluations. In the field of learning analytics, ML can process vast amounts of educational data, and this clearly helps provide insights into learning processes and institutional effectiveness (Siemens & Long, 2011). Lastly, the study by Winkler and Söllner (2018) concludes that ML-powered chatbots and virtual assistants can provide instant support to students by answering questions and guiding them through learning

materials. These diverse applications demonstrate the transformative potential of Machine Learning in enhancing educational experiences and outcomes.

Deep Learning

Deep Learning, a subset of machine learning based on artificial neural networks with multiple layers, has emerged as a powerful tool in various fields, including education. At its core, deep learning algorithms attempt to model high-level abstractions in data by using multiple processing layers, with each layer learning to transform its input data into a slightly more abstract and composite representation (LeCun et al., 2015). Deep learning has found numerous applications in the educational context. These applications have significantly enhanced the learning experience and educational outcomes. One of the primary applications is in Natural Language Processing (NLP), where deep learning models can understand and generate human-like text, which makes more sophisticated automated essay grading systems and personalized feedback on writing assignments possible (Alikaniotis et al., 2016). In adaptive learning, deep learning algorithms can create highly personalized learning paths by analyzing vast amounts of student data, such as performance history, learning style, and engagement patterns. This allows for the development of intelligent tutoring systems that can adapt in real-time to a student's needs. Thanks to this, students get tailored explanations, examples, and practice questions. According to Yang et al. (2019), deep learning has also revolutionized educational content creation, with models capable of generating diverse learning materials, such as practice questions, explanatory texts, and even educational videos.

Deep learning and related algorithms could also provide intelligent question-answering through natural language processing (Ansari et al., 2016). In the field of student performance prediction, deep learning models have shown superior accuracy

compared to traditional machine learning methods and allowed for earlier and more precise identification of at-risk students. Moreover, deep learning-based computer vision techniques have found applications in analyzing student engagement in physical and online classrooms. This provides valuable insights to educators about student attention and participation. Hernández-Blanco et al. (2019) highlight that in the domain of learning analytics, deep learning models can process complex, multimodal data from various learning platforms, and this provides detailed insights into learning behaviors and institutional effectiveness. Lastly, Cheng et al. (2020) propose that in language education, deep learning-powered speech recognition and synthesis have enabled more interactive and practical language learning applications, and this has allowed students to practice speaking and listening skills with AI-driven conversational partners. As deep learning continues to advance, its potential to transform education grows. This promises more personalized, effective, and engaging learning experiences for students across all educational levels.

Natural Language Processing (NLP)

Natural Language Processing (NLP), a branch of artificial intelligence that focuses on the interaction between computers and human language, has become increasingly influential in the field of education. NLP encompasses a range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis. It intends to achieve human-like language processing for various tasks or applications (Liddy, 2001). In educational contexts, NLP has numerous applications that are transforming teaching and learning processes. One of the most prominent applications is automated essay scoring. In these applications, NLP techniques are used to evaluate written responses. They help educators provide rapid feedback on aspects such as grammar, vocabulary,

coherence, and even content relevance (Shermis & Burstein, 2013). This not only saves time for educators but also allows students to receive immediate feedback on their writing. NLP is also crucial in the development of intelligent tutoring systems by enabling these systems to understand and respond to students' natural language inputs, thereby facilitating more interactive and personalized learning experiences (Graesser et al., 2012). In the realm of language learning, NLP gives power to applications that can analyze learners' pronunciation, grammar, and vocabulary usage and offer targeted feedback and personalized exercises (Vajjala & Banerjee, 2016). Text simplification is another area in which NLP is making significant contributions. The study by Siddharthan (2014) emphasizes that it automatically adapts complex texts to be appropriate for different reading levels. This makes educational materials more accessible to a broader range of learners. By helping to ensure that textbooks and other resources are appropriately matched to curriculum standards and student reading levels, NLP techniques are also employed in the content analysis of educational materials (McNamara et al., 2014). In the field of learning analytics, the research by Florence et al. (2019) identifies specific student learning outcomes that can be derived from the semantic analysis using the PSCE Speech Genre and NLP approach. This highlights the potential of NLP to uncover insights into student learning processes. Sentiment analysis, another application of NLP, is used to gauge student attitudes and emotions toward learning experiences by helping educators identify and address potential issues in real-time (Rani & Kumar, 2017). Moreover, NLP is fundamental to the development of chatbots and virtual assistants in educational settings. NLP enables these AI-powered tools to understand and respond to student questions, provide explanations, and offer learning support (Gonda et al., 2018).

As NLP technologies continue to advance, particularly with integrating deep learning techniques, their potential to enhance educational practices and outcomes grows. This means more personalized, interactive, and effective learning experiences in many educational domains.

Understanding AI, Generative AI, ML, DL, and NLP is crucial for developing an AI competency framework for EFL writing instructors because these technologies enable the creation of advanced tools that can enhance teaching effectiveness and student learning in a writing class. By leveraging AI and its subsets, both teachers and students can use automated feedback systems, personalized learning experiences, and innovative writing prompts following individual student needs. Overall, all these could ultimately improve writing instruction and outcomes in the EFL classroom.

AI in Writing Instruction

The integration of Artificial Intelligence (AI) into writing instruction has been a transformative shift in educational practices and has offered innovative solutions to enhance both teaching and learning practices in various educational settings. AI technologies, particularly those utilizing natural language processing and machine learning, can revolutionize writing instruction by providing real-time feedback, personalized learning experiences, and advanced tools for content creation for educators and students. These advancements are especially beneficial in the context of EFL writing instruction, where tailored feedback and individualized practice are crucial for the enhancement of writing skills and better learning outcomes.

AI's role in writing instruction can be better understood through two main lenses: technology-enhanced and AI-enhanced approaches. Technology-enhanced

approaches to writing instruction have laid the groundwork for incorporating AI using various digital tools and methodologies. These include *flipped learning*, *digital storytelling*, *computer-assisted language learning (CALL)*, and *blended learning*.

By building on these foundations, AI-enhanced writing instruction approaches have started offering even more sophisticated tools and systems. While *Automated Writing Evaluation (AWE)* systems provide instant feedback on writing and help students improve their skills through iterative practice, *Intelligent Tutoring Systems (ITS)* offer personalized instruction by adapting to the learner's needs in order to get effective learning outcomes. Furthermore, *AI-Powered Writing Assistants* support students in drafting and refining their writing for enhanced productivity and creativity. Additionally, *Adaptive learning systems* tailor educational content to the individual learner's pace and level, which could ensure that instruction is optimally challenging. Lastly, *Collaborative writing platforms with AI features* aim to facilitate joint writing projects with teamwork and collaborative learning opportunities.

This chapter will delve into these subcategories with the aim of exploring both technology-enhanced and AI-enhanced approaches to writing instruction. By examining these innovative methods, this study aims to uncover the specific skills, knowledge, and attitudes necessary for educators to successfully integrate AI tools into their teaching practices and the competencies they need to develop to make the best use of AI-enhanced writing support. The ultimate goal is to understand how these competencies can contribute to improved learning outcomes and support EFL students in developing their writing skills within an AI-enhanced writing instruction class.

Technology-Enhanced Approaches to Writing Instruction

Flipped Learning

Flipped learning is an instructional strategy that reverses the traditional learning environment by delivering instructional content, often online, outside the classroom (Bergmann & Sams, 2012). This approach allows for more interactive and engaging classroom activities, as students come prepared with prior knowledge gained from pre-class materials such as videos, readings, or lectures (Bishop & Verleger, 2013). According to Tucker (2012), flipped learning has enhanced student engagement, understanding, and material retention by shifting the focus from passive to active learning. In their study, Bergmann and Sams (2014) find that it facilitates differentiated instruction, as students can learn at their own pace outside of class and receive personalized support during class activities.

Writing instruction in flipped classrooms has demonstrated significant benefits by allowing students to engage with instructional content at their own pace before class, thus freeing up in-class time for interactive writing activities and personalized feedback (Engin, 2014). This approach helps students better understand the writing process through pre-class materials such as instructional videos, readings, and online exercises focused on writing techniques and grammar (Chen et al., 2014). In-class sessions are then dedicated to collaborative writing activities, peer review, and direct instructor interaction, enhancing students' writing skills and critical thinking abilities (Hung, 2015). This method not only improves writing proficiency but also fosters a more engaging and supportive learning environment (Akçayır & Akçayır, 2018). Additionally, using digital tools in flipped writing instruction allows for diverse and creative assignments, such as blogging, digital storytelling, and

multimedia projects, which can further motivate and engage students (Sun & Wu, 2016).

The integration of artificial intelligence in flipped classrooms has further enhanced the efficacy of this instructional model by providing personalized learning experiences and real-time feedback (Zawacki-Richter et al., 2019). AI-powered tools can analyze students' pre-class work and provide customized recommendations and adaptive learning paths. With the help of this, each student receives instruction tailored to their individual needs and learning pace (Huang et al., 2020). For instance, AI can be used to automatically grade assignments, highlight areas for improvement, and offer targeted exercises to address specific weaknesses (Luckin et al., 2016). In writing instruction, AI can assist in providing instant feedback on grammar, style, and coherence, helping students refine their writing before class (Perrotta, 2019). Moreover, AI-driven analytics can help instructors identify patterns in student performance and adjust their teaching strategies accordingly, making flipped classrooms more responsive and effective (Gong et al., 2020). Overall, integrating AI in flipped learning environments can significantly enhance student engagement, learning outcomes, and instructional efficiency.

Digital Storytelling

Digital storytelling is using digital tools to create and share narratives. It combines traditional storytelling elements with multimedia components such as images, audio, and video (Robin, 2008). This approach has evolved from simple text-based stories to complex, interactive multimedia presentations that can be easily shared online. Digital storytelling engages students by allowing them to create personal, meaningful stories that connect academic content with their own experiences (McLellan, 2006). According to Sadik (2008), this method not only

enhances students' creativity and technical skills but also provides a platform for them to express their voices and perspectives.

Digital storytelling significantly benefits writing skills by encouraging students to develop narratives, organize their thoughts coherently, and convey messages effectively (Sylvester & Greenidge, 2009). The process of crafting a digital story requires critical thinking, as students must plan, draft, revise, and edit their work, which improves their overall writing proficiency (Dreon et al., 2011). Additionally, integrating AI in digital storytelling further enhances these benefits by providing personalized feedback and support. AI-driven tools can assist in grammar and style correction, offer suggestions for narrative improvement, and even generate prompts to inspire creativity (Fang et al., 2023). These technologies enable students to refine their writing through iterative feedback loops, ultimately resulting in higher-quality narratives. Moreover, AI in digital storytelling utilizes ML, NLP, and tools like DALL-E and OpenCV to automate story generation. This creates personalized, interactive narratives with relevant images and revolutionizes industries beyond entertainment (Golwalker et al., 2024). Overall, the combination of digital storytelling and AI not only enhances writing skills but also fosters a deeper engagement with the learning process, which ultimately promotes both linguistic competence and digital literacy.

Computer-Assisted Language Learning (CALL)

Computer-Assisted Language Learning (CALL) refers to the use of computers and digital technologies to facilitate and enhance language teaching and learning processes (Levy, 1997). Over the years, CALL has evolved from simple text-based drills and practice exercises to sophisticated multimedia and interactive environments that provide immersive language learning experiences (Warschauer &

Healey, 1998). CALL encompasses a wide range of applications, including language learning software, online courses, mobile apps, and virtual reality environments that offer contextualized language practice (Chapelle, 2001). The development of CALL has been driven by advancements in technology and pedagogy, integrating elements such as multimedia content, gamification, and adaptive learning systems to create engaging and effective language learning experiences (Garrett, 2009).

CALL has significantly impacted writing instruction by providing tools and platforms that support the development of writing skills through interactive and personalized learning experiences (Hyland, 2003). For example, digital tools such as word processors, online writing labs, and collaborative writing platforms enable learners to practice and refine their writing in meaningful contexts (Warschauer, 2010). Additionally, the integration of artificial intelligence (AI) in CALL has further enhanced writing instruction by offering automated feedback and personalized support. AI-driven tools such as grammar checkers, style analyzers, and essay scoring systems provide instant feedback on various aspects of writing, including grammar, coherence, and style, helping learners to improve their writing skills iteratively (Heift & Schulze, 2015). Furthermore, when AI analyzes learners' writing patterns and offers targeted exercises and suggestions to address individual weaknesses, this provides learners with a more tailored learning experience (Kim et al., 2022). The use of AI in CALL also facilitates the creation of adaptive learning environments that adjust the difficulty level and type of writing tasks based on learners' progress and performance, and the final outcome is better learning (Ranalli, 2018). Overall, the integration of AI in CALL has transformed writing instruction by making it more interactive, personalized, and effective and has supported learners' development of linguistic competence and writing proficiency in the digital age.

Blended Learning

Blended learning, also known as hybrid learning, combines traditional face-to-face instruction with online educational materials and interactive digital platforms (Graham, 2006). This approach leverages the strengths of both in-person and online learning environments and offers a more flexible and personalized educational experience. According to Garrison and Vaughan (2008), the development of blended learning has been driven by advancements in technology and the growing recognition of the need for adaptive, student-centered learning models. It incorporates various digital tools and resources, such as learning management systems (LMS), video lectures, discussion forums, and virtual labs, to complement and enhance classroom activities (Horn & Staker, 2011). Bonk and Graham (2012) suggest that blended learning allows for differentiated instruction, enabling educators to tailor content and teaching methods to individual student needs and learning preferences.

Blended learning has significantly enhanced writing instruction by integrating digital tools and online resources that support the development of writing skills through both in-class and online activities (Means et al., 2010). In a blended learning environment, students can access instructional videos, interactive tutorials, and writing exercises online, allowing them to learn at their own pace and revisit materials as needed (Graham et al., 2013). Bernard et al. (2014) suggest that in-class sessions can then be dedicated to collaborative writing projects, peer review, and direct interaction with instructors, which reinforces learning and provides immediate feedback. The integration of artificial intelligence (AI) in blended learning has further enhanced writing instruction by offering personalized feedback and adaptive learning pathways. While Allen and Seaman (2013) conclude that AI-driven tools

such as automated essay scoring systems, grammar checkers, and writing analytics provide instant, detailed feedback on various aspects of writing, helping students improve their skills iteratively, McCarthy (2017) highlights that these tools can analyze writing patterns, suggest improvements, and identify areas for further practice, thus providing a tailored learning experience that addresses individual needs. Furthermore, AI can facilitate formative assessment by tracking student progress and performance over time, enabling educators to adjust instruction and provide targeted support as needed (Baker & Smith, 2019). Overall, the integration of AI in blended learning environments has transformed writing instruction by making it more interactive, personalized, and effective, supporting students' development of writing proficiency and digital literacy.

AI-Enhanced Approaches to Writing Instruction

Automated Writing Evaluation (AWE) Systems

Automated Writing Evaluation (AWE) refers to the use of computer technology to assess and provide feedback on students' written work. AWE systems analyze various aspects of writing, including grammar, syntax, coherence, and organization, and offer both formative and summative feedback. The origins of AWE date back to the 1960s, when early research focused on developing algorithms to analyze text for grammatical correctness. One of the first significant AWE systems, Project Essay Grade (PEG), was developed by Ellis Page during this period, and statistical models were used to evaluate writing based on various linguistic features (Page, 1966). In the 1990s and 2000s, advancements in Natural Language Processing (NLP) and machine learning enabled more sophisticated analysis of written text, leading to the emergence of commercial systems such as Criterion by ETS, IntelliMetric by Vantage Learning, and MyAccess! by Vantage. Recent

developments have seen modern AWE systems incorporate AI and deep learning techniques to enhance accuracy and provide more comprehensive feedback.

Additionally, these tools are increasingly integrated into learning management systems (LMS) and other educational technologies, which makes them more accessible and user-friendly. Some commonly used examples of AWE systems include Criterion (ETS), which provides diagnostic feedback and scores on various writing traits; Grammarly, which offers real-time grammar, punctuation, and style suggestions; WriteToLearn (Pearson), which combines automated essay scoring with reading comprehension assessment; and Turnitin's Revision Assistant, which focuses on providing formative feedback on higher-order concerns such as argumentation and coherence.

Automated Writing Evaluation (AWE) systems are crucial in providing feedback and correction for writing tasks, helping the development of second language writing. Recent research by Allen et al. (2024) has highlighted the shift towards evaluating source-based writing, emphasizing the integration of comprehension and writing skills in essays. Studies have shown that AWE tools like Grammarly, Quillbot, and ChatGPT can significantly enhance students' feedback literacy, particularly in areas such as vocabulary use, language mechanics, and grammar (Gozali et al., 2024; Mat, 2024; Gavina & Ibay-Pamo, 2024). Automated Writing Evaluation has revolutionized writing instruction by providing scalable, consistent, and objective feedback on student writing. Its widespread use is driven by the efficiency and accessibility it offers, the support it provides to teachers, and its ability to enhance student engagement and learning. As technology continues to advance, AWE systems are likely to become even more sophisticated and integral to writing education.

While AWE tools are beneficial in improving specific aspects of writing, they may still require a combination of traditional teaching methods and educator intervention to address higher-order writing skills like content organization and substance (Mat, 2024). Integrating AWE tools with teacher facilitation can substantially improve students' academic writing, especially in areas like grammar usage and verb tense accuracy (Gavina & Ibay-Pamo, 2024). The effectiveness of AWE systems varies based on feedback types and cannot fully replace human feedback yet (Wu & Yu, 2023).

Automated Writing Evaluation (AWE) systems, such as VWT, focus on lower-level aspects and receive positive student perception but lack revision strategies compared to teacher feedback in L2 writing practice (Chen et al., 2024). Systems like Pigai provide detailed feedback on technical aspects, aiding students in revising writing, reducing errors, and enhancing autonomous learning through interaction with machines (Yuan, 2023). Moreover, AWE systems provide immediate feedback and suggestions to improve English writing. Consequently, this benefits students' test performance, revision, and self-efficacy in writing skills (Wang et al., 2023).

In English as a Foreign Language (EFL) contexts, AWE systems enhance student writing proficiency by providing quality feedback, fostering autonomy, and reducing teacher workload. However, human intervention remains crucial for maximizing outcomes and addressing limitations (Aldosemani et al., 2023). AWE systems assess written work, provide feedback, and focus on sentence-level scoring for detailed feedback, effectively enhancing students' writing skills (Hossain & Mustafa, 2023).

Research exploring English learners' perceptions of Grammarly, an AI-based AWE program, shows positive impacts on grammar achievement and learner motivation in writing (Son, 2023). AWE systems must consider variance in student writing styles, such as syntactic complexity, to provide accurate assessment and relevant feedback, as shown in the study by Goldshtein et al. (2023).

Automated Writing Evaluation has clearly revolutionized writing instruction by providing scalable, consistent, and objective feedback on student writing. Its widespread use is mainly driven by the efficiency and accessibility it offers, the support it provides to teachers, and its ability to enhance student engagement and learning. As technology continues to advance, AWE systems are likely to become even more sophisticated, integral and useful to writing education.

Intelligent Tutoring Systems (ITS)

Intelligent Tutoring Systems (ITS) are advanced computer systems that provide personalized instruction and feedback by leveraging artificial intelligence to adapt to individual learner needs. They are defined as “The approach known as ITS has been pursued by researchers in education, psychology, and artificial intelligence. The goal of ITS is to provide the benefits of one-to-one instruction. It enables learners to practice their skills by carrying out tasks within highly interactive learning environments.” (Phobun, & Vicheanpanya, p.2, 2010)

Originating in the 1960s with rule-based systems like SCHOLAR, ITS evolved through the 1990s with cognitive models such as the Cognitive Tutor (Carbonell, 1970) and has recently incorporated AI and machine learning for greater adaptability and sophistication (Lin et al., 2023). These systems are widely used due to their ability to provide personalized learning it offers (VanLehn, 2011), immediate feedback (Anderson et al., 1995), enhanced engagement (Graesser et al., 2012),

scalability (Luckin & Holmes, 2017) and Data-Driven Insights (Baker & Yacef, 2009). Examples include Carnegie Learning's MATHia, ALEKS, and AutoTutor, which have demonstrated significant improvements in student performance and motivation. Despite challenges in development and implementation, ITS are valuable tools in modern education since they combine the efficiency of technology with the support of human instructors for ideal learning (Yuan, 2023; Dašić et al., 2016).

Plagiarism Detection

Plagiarism detection systems are sophisticated tools designed to identify instances of academic dishonesty by comparing submitted work against extensive databases of published material, online content, and other student submissions. Emerging in the early 2000s, these systems have evolved rapidly with the rise of digital content and the need for academic integrity. Early tools focused on simple text-matching techniques, while modern systems incorporate advanced natural language processing and machine learning to detect paraphrasing, synonym substitution, and other sophisticated attempts to avoid detection.

As more and more people rely on artificial intelligence (AI) for academic and professional writing, there has been an increase in advanced AI tools designed to detect plagiarism. These tools do not aim to identify the usual copy-paste jobs anymore; they aim to detect texts that have been generated by AI. However, how well these tools work is still up for debate, with questions about their accuracy and the ethics of using them.

While AI detection tools have shown some success in detecting AI-generated text, they often have difficulty when it comes to recognizing human-written content. A study by Elkhatat et al. (2023) suggests that while these tools are useful, they still need more refining as AI-written content gets more sophisticated and harder to tell

apart from human writing. This points to a bigger issue: how to stay ahead in this fast-moving field of AI-assisted writing and plagiarism detection. The situation gets even more difficult with the advanced AI tools that can paraphrase text so well that detecting AI-generated plagiarism becomes a challenge. Gupta (2023) discusses the need to keep evolving how we evaluate and catch these new forms of plagiarism. This means constantly updating AI detection tools so they can do their job effectively.

Some AI tools seem to perform better than others in real-world applications. Anil et al. (2023) found that Grammarly, a popular AI tool, was better at spotting plagiarism in AI-generated articles compared to other platforms. This suggests that certain AI tools could be a good, cost-effective choice for institutions looking to protect academic integrity. However, the accuracy of these tools must be questioned, depending on the specific technology they use. Ibrahim (2023) points out that tools like RoBERTa are somewhat helpful in detecting AI-assisted plagiarism, especially in writing by non-native English speakers, but they are not always reliable at identifying machine-generated text. This inconsistency raises concerns about the reliability of these tools for educators, particularly regarding the potential for both false positives and undetected instances of machine-generated text.

There are also some ethical concerns. Májovský et al. (2024) warn that if we rely too heavily on AI for plagiarism detection, we could end up in a kind of tech arms race, where increasingly advanced AI is met with equally advanced detection methods. They suggest we should also develop ethical guidelines to go along with these technological solutions, to make sure the process remains fair.

Generative AI tools like GPTZero are already being incorporated into plagiarism detection systems, with platforms like iThenticate expected to add similar

features to enhance academic integrity and originality checks (Davison et al., 2023). This points to a future where AI plays a central role in keeping academic work authentic and original. There is also ongoing research into improving AI-based plagiarism detection. Xiong et al. (2023) propose a new method that uses AI tools like BERT embeddings, enhanced ABC optimization, and reinforcement learning. This approach seems to outperform existing models on standard datasets, which is a promising development for creating stronger, more accurate plagiarism detection systems.

However, the rise of AI-generated research paper fabrication and plagiarism is creating new challenges. Elali and Rachid (2023) emphasize the growing importance of AI tools in catching these practices within the scientific community. Similarly, Abd-Elaal et al. (2022) highlight the need for academics to be more aware of the challenges posed by AI-powered tools like Automatic Article Generators (AAG), which can make it tough to detect computer-generated writing.

Finally, some researchers are turning to deep learning as a potential solution for text plagiarism detection. El-Rashidy et al. (2022) propose using intelligent deep learning systems, like Long Short-Term Memory (LSTM) networks, for more effective plagiarism detection. This method leverages AI's strengths to tackle the complexities of modern plagiarism.

In summary, while AI tools offer a lot of potential in the fight against plagiarism, they are not perfect. As AI-generated content continues to evolve, there will always be a need to keep developing these tools and consider the ethical implications to ensure they truly support academic integrity in this digital age.

AI-Powered Writing Assistants in Language Instruction

AI-Powered Writing Assistants represent a significant advancement in language learning technology since they offer sophisticated support for writing tasks in both first and second-language contexts. These tools leverage natural language processing, machine learning, and deep learning algorithms to provide real-time feedback, suggestions, and corrections on various aspects of writing, including grammar, vocabulary, style, and coherence (Warschauer & Grimes, 2019). AI writing assistants, such as Grammarly and ProWritingAid, and language-specific tools like Write & Improve for English learners have transformed the writing process by offering instantaneous, personalized guidance that adapts to individual user patterns and proficiency levels (Ranalli, 2018). These assistants can identify complex linguistic issues, suggest more appropriate phrasing, and even explain their recommendations, thereby supporting language acquisition and metalinguistic awareness. In another study, Li et al. (2024) also highlight the value users place on AI assistance for content generation, noting improvements in productivity and confidence, although these tools may reduce accountability and diversity in writing. Moreover, advanced AI writing tools can analyze text for tone, clarity, and audience appropriateness, helping learners develop a more nuanced understanding of effective communication in different contexts (Strobl et al., 2019). The integration of these tools in language classrooms has facilitated more efficient peer review processes, enabled teachers to provide more targeted instruction, and empowered learners to take greater control of their writing development (Godwin-Jones, 2018). As these tools continue to evolve, they promise to play an increasingly significant role in supporting language learners' writing skills, fostering autonomy, and preparing students for the technological demands of academic and professional writing in the

21st century. Cardon et al. (2023) emphasize the need for AI literacy among students to harness benefits like efficiency and idea generation while overcoming challenges such as diminished critical thinking and authenticity. However, the use of AI writing assistants also raises important considerations regarding academic integrity, over-reliance on technology, and the need for critical evaluation of AI-generated suggestions (Roe et al., 2023; Yeo, 2023). Furthermore, Sok and Heng (2023) explore the benefits and risks related to the use of ChatGPT in education and research, and they conclude that Academic integrity problems, unfair learning assessments, incorrect information, and excessive dependence on AI are among the hazards.

The benefits and challenges of AI-powered writing assistants in writing instruction are multi-faceted and require careful consideration. While these tools offer unique support for writing development, enhancing efficiency, personalization, and linguistic awareness, they also necessitate careful consideration of academic integrity, critical thinking skills, and the balance between technological assistance and autonomous learning.

Adaptive Learning Systems (ALS)

Adaptive Learning Systems (ALS) in language instruction represent a significant advancement in educational technology by offering personalized learning experiences that are tailored to individual student needs. These intelligent, data-driven platforms can adjust learning content, pace, and difficulty based on learner performance, preferences, and behaviors. Since it can leverage algorithms and machine learning, ALS creates customized learning paths in order to optimize the acquisition of language skills (VanLehn, 2011).

In language learning, ALS is utilized across various fields. They promote vocabulary acquisition by tailoring word lists and employing spaced repetition techniques based on learner progress. By observing error patterns, these systems adjust the complexity and frequency of exercises for grammar practice. In reading comprehension, ALS selects texts of appropriate difficulty and topics in alignment with learner interests. Moreover, pronunciation training benefits from focused practice on specific phonemes that challenge individual learners. Furthermore, ALS provides writing support through targeted feedback and suggestions that are usually adjusted to the learner's proficiency level.

Several examples of ALS have gained popularity in language learning. DuoLingo utilizes adaptive algorithms to personalize lesson plans and difficulty levels and thus ensures an engaging and appropriately challenging experience for language learners. Rosetta Stone employs adaptive recall techniques to reinforce learned content at optimal intervals, enhancing long-term retention. Babbel adapts its content based on learner performance and self-reported goals, which creates a more relevant learning journey. Busuu supports learners in their language acquisition process by leveraging AI to create personalized study plans and deliver personalized feedback.

The benefits of incorporating ALS in language instruction are various. Personalization is a key advantage, with tailored learning experiences improving engagement and outcomes (Kerr, 2016). These systems also enhance efficiency by optimizing learning time, focusing on areas that require improvement (Wang & Liao, 2011). In terms of motivation, ALS provides learners with a sense of progress and achievement through adaptive challenges, which is helpful for students for continued engagement (Tseng et al., 2008). Furthermore, the data-driven nature of these

systems offers detailed analytics on learner performance and thus provides valuable insights for both students and teachers (Godwin-Jones, 2017).

However, the implementation of ALS is not without challenges. One important ethical question is that data privacy concerns arise from collecting and analyzing large amounts of learner data (Rodríguez-Triana et al., 2016). Also, the initial development costs for effective adaptive systems can be substantial, and this requires significant investment in both technology and content creation (Khosravi et al., 2020). Without careful planning and adaptation, integrating these systems with the current curricula and traditional teaching methods often proves challenging (Yang et al., 2013). If crucial human aspects of language learning, such as cultural understanding and authentic communication, are neglected, there is also a risk of over-reliance on technology (Reinders & White, 2016). Lastly, ensuring algorithmic fairness and avoiding bias across diverse learner populations is another challenge in the development and implementation of ALS (Holstein et al., 2019).

As Adaptive Learning Systems continue to evolve, they hold great potential for transforming language instruction. By providing highly personalized, efficient, and engaging learning experiences, these systems can significantly enhance language acquisition processes. However, for their effective implementation, there needs to be careful consideration of both their potential benefits and limitations within the broader context of language education.

Collaborative Writing Platforms with AI Features

Collaborative Writing Platforms with AI Features represent a dynamic intersection of technology and pedagogy in language education. These platforms combine the benefits of collaborative learning with artificial intelligence to enhance the writing process for a more effective language learning process with better

outcomes. By integrating AI-powered tools such as real-time grammar and style checkers, automated feedback systems, and intelligent content suggestions, these platforms create a supportive environment for group writing projects and peer review (Strobl et al., 2019).

Popular examples of such platforms include Google Docs with its AI-enhanced grammar suggestions, Grammarly's collaborative features, and specialized language learning tools like EssayJack. These systems leverage natural language processing and machine learning algorithms to provide instant feedback on various aspects of writing, from basic grammar and spelling to more complex issues of style, coherence, and argumentation (Li et al., 2021). The AI components can identify patterns in student writing, suggest improvements, and even adapt to different genres and writing styles.

The benefits of these platforms are significant. They foster collaborative learning by allowing students to work together in real time, sharing ideas, and providing peer feedback. The AI features support this process by offering immediate suggestions and reducing the cognitive load on both learners and instructors. This combination of human collaboration and AI assistance can improve writing quality, increase learner autonomy, and enhance metalinguistic awareness (Warschauer & Grimes, 2019).

However, challenges exist in implementing these platforms effectively. There are concerns about over-reliance on AI suggestions, potentially stifling creativity or leading to homogenized writing styles. Additionally, issues of data privacy and the need for critical evaluation of AI-generated feedback require careful consideration. Educators must also adapt their teaching strategies to integrate these tools into their

curriculum, balancing technology with traditional writing instruction methods (Godwin-Jones, 2018).

As these platforms continue to evolve, they hold the potential to transform collaborative writing in language learning contexts significantly. Combining human creativity and interaction with AI-powered assistance offers a promising avenue for enhancing writing skills, promoting peer learning, and preparing students for increasingly digital and collaborative writing in academic and professional settings.

Common Competency Frameworks for Language Teachers

What is a competency framework?

According to Hyland (1994):

A competency framework is a structured model that delineates the specific competencies, including skills, knowledge, and behaviors, required to perform a job effectively. It serves as a foundation for developing educational programs, assessing performance, and guiding career progression within an organization. (Hyland, 1994, p.21)

In an educational setting, a competency framework is defined as:

A competency framework in education is a system that specifies the essential competencies — including skills, knowledge, and behaviors — that students or teachers need to acquire to be successful in their educational and professional pursuits. It provides a comprehensive outline for curriculum development, instructional strategies, and assessment methods. (European Commission, 2019)

These educational frameworks provide clarity and consistency (Spencer & Spencer, 1993), enhanced learning outcomes (Klein-Collins, 2012), personalized

learning (McClarty & Gaertner, 2015), improved assessment and accountability (Guskey, 2007), and facilitation of lifelong learning (European Commission, 2019).

British Council Continuing Professional Development (CPD) Framework

The British Council Continuing Professional Development Framework is a comprehensive guide designed to support English language teachers' professional growth. It outlines a set of professional practices and development stages. Teachers can use them to reflect on their skills, identify areas for improvement, and plan their professional development (British Council, 2015).

The framework is structured around four stages of development: *Awareness, Understanding, Engagement, and Integration*.

These stages are applied across 12 professional practices: *Planning lessons and courses, Understanding learners, Managing the lesson, Knowing the subject, Managing resources, Assessing learning, Integrating ICT, Taking responsibility for professional development, Using inclusive practices, Using multilingual approaches, Promoting 21st-century skills, Understanding educational policies and practice* (British Council, 2015).

While the framework does not explicitly mention AI, it does include competencies related to technology integration and writing instruction. Regarding technology integration: Under the "Integrating ICT" practice, teachers at the "Integration" stage are expected to: "Exploit the potential of ICT to develop learners' language skills," "Develop innovative learning materials using ICT," and "Evaluate the impact of using ICT on learning outcomes" (British Council, p. 13, 2015).

Regarding writing instruction, the "Knowing the subject" practice includes competencies related to teaching writing. At the "Integration" stage, teachers should:

"Support learners to develop their writing skills for various genres and purposes."

"Help learners to develop their writing process (planning, drafting, revising)"
(British Council, p. 10, 2015).

Although the framework does not explicitly state competencies for the use of AI in writing instruction, it encompasses useful areas for language teachers to develop themselves professionally.

The TESOL International Association's Standards for ESL/EFL Teachers of Adults

The TESOL International Association's Standards for ESL/EFL Teachers of Adults framework is designed to ensure high-quality English language instruction for adult learners. It consists of eight domains, each with its own set of standards and performance indicators: *Planning, Instructing, Assessing, Identity and Context, Language Proficiency, Learning, Content and Commitment*, and *Professionalism* (TESOL International Association, 2008).

While this framework does not explicitly mention AI, several standards are relevant to technology integration and writing instruction. Related to technology integration: Under the "Instructing" domain, Standard 3.C states: "Teachers use appropriate technology to enhance instruction." Furthermore, performance indicators for this standard include: "Using technological resources to support instructional goals, integrating technology to enhance language learning, and evaluating and adapting technological resources to meet learner needs and interests."

Related to writing instruction, the "Instructing" domain also includes Standard 3. B:

"Teachers create supportive environments that engage all learners in purposeful learning and promote respectful classroom interactions." This standard encompasses writing instruction, with performance indicators such as Providing activities that integrate listening, speaking, reading, and writing and Incorporating activities that help learners develop writing skills for real-life purposes.

The "Assessing" domain (Standard 4) is also relevant, as it includes performance indicators related to assessing writing skills and providing feedback, which could potentially involve AI-assisted tools. Professional Development:

The "Commitment and Professionalism" domain (Standard 8) emphasizes ongoing professional development, which could include staying current with technological advancements in language teaching:

"Teachers continue to grow in their understanding of the relationship of second language teaching and learning to the community of English language teaching professionals, the broader teaching community, and communities at large, and use these understandings to inform and change themselves and these communities."

While not explicitly mentioning AI or advanced technologies, these standards provide a framework that can accommodate the integration of AI-enhanced teaching tools and methods. The emphasis on using appropriate technology, creating supportive learning environments, integrating skills, and ongoing professional development aligns well with the potential use of AI in language instruction.

The European Profiling Grid (EPG)

The European Profiling Grid (EPG) is a comprehensive framework designed to describe the fundamental competencies of language teachers. It was developed as part of a project co-funded by the European Commission and is aligned with the

principles of the Common European Framework of Reference for Languages (CEFR) (North et al., 2013).

The EPG is organized into four main categories:

Training and Qualifications

Key Teaching Competences

Enabling Competences

Professionalism

Each category is further divided into subcategories, and competencies are described across six developmental phases, from novice (1.1) to highly experienced and expert (3.2) (North et al., 2013).

Regarding AI-related competencies, the EPG was developed before the widespread integration of AI in language teaching, so it does not explicitly mention AI. However, it does include competencies related to digital technologies that could be extended to encompass AI:

In the "Enabling Competences" category, under "Digital Media," the EPG describes the following competencies at the highest level (3.2): "Can train colleagues in the use of digital media. Can advise colleagues and institutions on the choice and use of digital media" (North et al, 2017, p. 7).

While not explicitly mentioning AI, these competencies could be interpreted to include AI tools as part of digital media in language teaching. Related to writing instruction, the EPG addresses writing instruction primarily under the "Key Teaching Competences" category, specifically in the "Methodology: Knowledge and Skills" subcategory. At the highest level (3.2), it states that teachers: "Can apply appropriate strategies for developing different kinds of writing skills for a wide range of academic and professional purposes" (Rossner, 2017, p. 5). Additionally, under

"Lesson and Course Planning," at level 3.2, teachers: "Can design courses that are based on learners' specific needs (e.g., academic writing, business English)" (North et al, 2017, p. 6).

It is worth noting that while the EPG provides a comprehensive framework for language teacher competencies, it does not go into great detail about specific methodologies or approaches for writing instruction. Instead, it focuses on teachers' ability to adapt their teaching to various contexts and learner needs.

In conclusion, while the EPG does not explicitly address AI-related competencies, its structure allows for the integration of new technologies and methodologies as they emerge in the field of language teaching. As AI continues to impact education, frameworks like the EPG may need to be updated to include specific AI-related competencies for language teachers.

The Cambridge English Teaching Framework

The Cambridge English Teaching Framework is designed to describe the key competencies for effective teaching at different stages of a teacher's career. It was developed by Cambridge Assessment English, part of the University of Cambridge, in 2019. It provides a comprehensive guide for language teachers to assess their skills, identify areas for improvement, and plan their professional development. The framework is organized into five main categories:

Learning and the Learner

Teaching, Learning, and Assessment

Language Ability

Language Knowledge and Awareness

Professional Development and Values

Each category is further divided into subcategories and components. The framework describes teacher competencies across different stages of development, although these stages are not explicitly mentioned.

Related to writing Instruction, the framework addresses writing instruction under the "Teaching Language Skills" subcategory within the "Teaching, Learning and Assessment" category. Key points include

Understanding key concepts, principles, and techniques for teaching writing skills (e.g., genres, text types, pre-/while-/post-writing activities, sub-skills, strategies),

Knowledge of principles and techniques for supporting and giving feedback on learners' written language (e.g., process writing stages, peer-checking, drafting, use of error codes),

Understanding different genres, types of writing and their purposes, differences between spoken and written English, and levels of formality,

Ability to identify difficulties learners face in writing,

Skill in selecting appropriate models and tasks for writing instruction,

Demonstration of practical application through various teaching techniques in specific contexts to achieve desired learning outcomes

Although the framework does not explicitly mention AI-related competencies, it does include some digital and technology-related skills that could be relevant to AI integration.

Under "Using Digital Resources, there is:

- *Familiarity with key concepts related to digital teaching/learning (e.g., blended learning, flipped classroom),*

- *Core principles and techniques for selecting and using digital/online resources,*
- *Key practical skills (e.g., computer skills, typing, searching the internet, finding web-based materials, use of social media),*
- *Ability to use these techniques and skills effectively to achieve desired learning outcomes and promote learner autonomy.*

Under "Using Teaching Aids", there are core techniques for using digital teaching aids (e.g., interactive whiteboards, PowerPoint).

While not explicitly mentioning AI, these competencies provide a foundation that could be extended to include AI-related skills in language teaching. The framework's emphasis on continuous professional development also suggests an openness to incorporating new technologies and methodologies as they emerge in the field of language teaching.

The Equals Framework for Language Teacher Training & Development

The Equals Framework for Language Teacher Training & Development is a comprehensive tool designed to support the training and development of language teachers. Equals (Evaluation and Accreditation of Quality Language Services) is an international association of language education providers that aims to promote high-quality language teaching and learning. The framework is organized into three phases of a teacher's development: Initial Training, In-Service Training, and Continuous Professional Development. It covers five main areas of competence:

Planning Teaching and Learning

Teaching and Supporting Learning

Assessment of Learning

Language, Communication, and Culture

The Teacher as a Professional

For each competency area, the framework provides detailed descriptors of teacher competencies at different stages of development. It is designed to be adaptable to various contexts and can be used for different purposes, including self-assessment, peer observation, and formal evaluation. The framework aligns with international standards, including the Common European Framework of Reference for Languages (CEFR). It can be used by individual teachers for self-reflection, by teacher trainers for program design, and by institutions for staff development planning. The framework focuses on teaching skills and emphasizes the importance of cultural awareness, professional development, and the teacher's role in the broader educational community.

While not explicitly focused on AI, the framework does include competencies related to using digital technologies in language teaching, which can be relevant to AI integration.

EdTech Masterplan 2030

The "Transforming Education through Technology" Masterplan 2030, also known as "EdTech Masterplan 2030," is an initiative launched by Singapore's Ministry of Education (MOE) to leverage technology for enhancing teaching and learning. Building on previous plans, it addresses the post-COVID landscape where technology has become crucial for education. The vision of this master plan is "Technology-transformed learning, to prepare students for a technology-transformed world."

The master plan outlines several outcome goals for students, teachers, and learning environments. For students, it aims to develop digitally-empowered 21st-century learners who are self-directed, collaborative, digitally literate, and

innovative. Teachers are expected to become collaborative contributors, pedagogically proficient, data-literate learning designers, and continuous learners and innovators in educational technology. The plan also envisions flexible, multi-functional, and digitally connected learning spaces that can adapt to various teaching and learning needs.

The implementation of the master plan focuses on several strategic thrusts. These include providing an AI-enabled, resource-rich Singapore Student Learning Space (SLS) for greater customization of learning, enhancing feedback and assessment through digital tools, strengthening students' digital literacy and technological skills, emphasizing cyber wellness, and harnessing technology to enhance 21st Century Competencies (21CC) development. The plan also aims to establish a culture of sharing and adapting lesson resources and promote the use of e-Pedagogy in subject departments.

Key enablers support the implementation of these strategic thrusts at a system-wide level. These include enhancing schools' ability to use data for decision-making, redesigning learning spaces to support technology-enabled learning, and establishing a network of partnerships to facilitate resource sharing and solution crowdsourcing. The plan emphasizes professional development for teachers in areas such as e-Pedagogy, data literacy, and cyber wellness.

The EdTech Masterplan 2030 also highlights the importance of partnerships. It seeks to enhance public-private partnerships between MOE, other government agencies, Institutes of Higher Learning, and the private sector to develop and scale good EdTech practices and solutions more quickly. The plan also aims to keep MOE's EdTech efforts informed by the latest research and industry developments, benchmarking against international best practices. Additionally, it recognizes parents

as key partners, emphasizing their role in complementing schools' efforts in EdTech use, especially in maintaining consistent cyber wellness guidelines and practices at home.

In conclusion, the "Transforming Education through Technology" Masterplan 2030 presents a comprehensive approach to integrating technology into Singapore's education system. It aims to create a technology-enhanced learning environment that prepares students for a rapidly evolving digital world while supporting teachers in their professional development and leveraging partnerships to drive innovation in education.

The BALEAP Competency Framework for Teachers of English for Academic Purposes (EAP)

The BALEAP Competency Framework for Teachers of English for Academic Purposes (EAP) was developed by the British Association of Lecturers in English for Academic Purposes. Established to address the expanding need for specialized EAP teaching in higher education institutions worldwide, this framework was published in August 2008. It aims to provide a clear set of core competencies for professional EAP practitioners, aiding new teachers in the field and those responsible for their training. This framework ensures that teachers can effectively support students in acquiring the language, skills, and strategies necessary for academic success in higher education contexts.

The framework outlines several key competencies necessary for effective EAP instruction. These include understanding academic contexts, recognizing disciplinary differences, and possessing a high level of systemic language knowledge, particularly in discourse analysis. Teachers are expected to have a reasonable knowledge of the organizational, educational, and communicative

policies, practices, values, and conventions of universities. They should also be able to help students navigate these systems, understand university policies and procedures, and relate EAP course objectives to the requirements of university courses.

Understanding disciplinary differences is another crucial competency. EAP teachers must recognize how various disciplines influence knowledge communication and help students investigate and understand these disciplinary conventions. This includes working with subject specialists and raising students' awareness of discourse features specific to their fields. Teachers are also expected to thoroughly understand academic discourse, including grammar, syntax, cohesion, coherence, semantics, and pragmatics, and apply this knowledge in course organization, materials selection, and assessment.

The BALEAP framework explicitly addresses writing instruction within several competencies. In the area of text processing and text production, teachers are expected to understand approaches to text classification and discourse analysis. This includes identifying and analyzing academic genres and their functional and rhetorical features and training students to do the same. Teachers are tasked with incorporating explicit analysis of spoken and written texts into their teaching, staging and scaffolding the teaching of reading and listening for study purposes, and using a text-based approach to teach writing and speaking skills.

Moreover, teachers must be able to design programs that integrate text processing and text production, using cycles of analysis and synthesis of processes, skills, and tasks. This ensures that students develop their abilities to effectively produce and comprehend academic texts. Teachers are also encouraged to employ

various teaching methods and materials to cater to different learner needs, incorporating tasks and activities that develop individual and group work skills.

Assessment practices in the BALEAP framework are integral to writing instruction. Teachers are expected to be proficient in assessing academic language and skills tasks using formative and summative assessments. This involves selecting appropriate assessment modes and designing or evaluating assessment tools for EAP listening, speaking, reading, and writing, including integrated skills. Teachers must apply marking criteria consistently, provide constructive feedback on oral and written performance, and use assessment outcomes to inform their teaching strategies. This comprehensive approach ensures that writing instruction is continuously adapted to meet students' needs and promote their academic development.

While the BALEAP Competency Framework does not directly address AI, it implicitly supports the integration of AI-related technologies through its emphasis on digital literacy, autonomous learning, and data-informed assessment practices. These areas can be significantly enhanced by AI tools, thereby improving the overall effectiveness of EAP teaching and learning.

Overall, the BALEAP Competency Framework provides a structured and comprehensive guide for EAP teachers. It aims to support teachers to have the necessary skills and knowledge to support students' academic success. By emphasizing key competencies, writing instruction, assessment practices, and professional development, the framework helps maintain high standards in EAP teaching and learning.

The writing section of the BALEAP (British Association of Lecturers in English for Academic Purposes) Can Do Framework for EAP syllabus design and

assessment outlines key competencies for academic writing. It is structured into two main categories:

The Academic Context category includes three subcategories: W1.1 Approaching academic tasks, W1.2 Completing academic tasks, and W1.3 Dissertation or project writing. These focus on the practical aspects of engaging with and fulfilling various academic writing assignments. The Academic Discourse category is divided into three subcategories: W2.1 General, W2.2 Language specific, and W2.3 Genre specific. These subcategories address the linguistic and stylistic elements of academic writing, from general principles to language-specific features and genre conventions. The framework notes that some competencies, marked with an asterisk (*), have been identified as particularly important by multiple interviewees, highlighting their significance in the field. This structured approach provides a comprehensive guide for developing EAP writing syllabi and assessment criteria, ensuring that students acquire the necessary skills for effective academic writing across various contexts and disciplines.

The ISTE (International Society for Technology in Education) Standards

The ISTE (International Society for Technology in Education) Standards are a framework designed to guide educators, students, and leaders in effectively using technology to enhance learning and teaching. These standards emphasize the development of digital skills and competencies necessary for success in the modern educational environment (Crompton, 2023).

Basic Content of ISTE Standards for different members of educational world include:

The standards for *students* encourage students to take an active role in their learning by using technology to achieve their goals, build networks, and demonstrate

competency. Key areas include empowered learners, digital citizens, knowledge constructors, innovative designers, computational thinkers, creative communicators, and global collaborators.

Educators are guided to continuously improve their professional practices by learning from others, engaging in leadership roles, fostering digital citizenship, designing learner-driven activities, and using data to inform instruction. The focus is on professional growth, leadership, and effective use of technology to enhance learning outcomes.

Education Leaders are expected to use technology to promote equity and inclusion, establish strategic visions, and create a culture of innovation in schools. They should model continuous learning and encourage the use of technology to solve problems and improve educational practices.

For *Coaches*, these standards outline the role of coaches in supporting educators and leaders in the effective use of technology. Coaches are expected to inspire, collaborate, and provide professional learning to help educators integrate technology into their teaching practices.

Overall, the ISTE Standards aim to prepare all members of the educational community to develop in a connected, digital world by providing a comprehensive set of guidelines for integrating technology into education effectively.

The UNESCO Draft AI Competency Frameworks for Teachers and School Students

The UNESCO Draft AI Competency Frameworks for Teachers and School Students are currently being developed by UNESCO, with the release scheduled for Digital Learning Week 2024. These frameworks aim to provide comprehensive guidelines for integrating AI competencies into educational practices for both

teachers and students. The drafts outline various competencies across multiple dimensions with a structured approach to AI literacy and its applications in education (UNESCO, 2024)

The AI Competency Framework for Teachers (AI CFT) includes several key areas for educators to incorporate AI into their teaching practices effectively. These areas are categorized into different aspects, each with a progression from basic understanding to advanced application and creation:

Human-centered Mindset focuses on human agency, accountability, and social responsibility in the context of AI. It emphasizes the ethical principles surrounding AI, promoting safe and responsible use and encouraging the co-creation of AI ethics.

AI Foundations & Applications includes AI's basic techniques and applications, advancing to application skills and, ultimately, the ability to create with AI. Teachers are expected to develop a foundational understanding of AI, apply these concepts in educational settings, and innovate by creating new AI applications.

AI Pedagogy covers the integration of AI-assisted teaching methods, moving towards AI-pedagogy integration and AI-enhanced pedagogical transformation. It highlights the importance of leveraging AI to improve teaching methods and educational outcomes.

AI for Professional Development focuses on how AI can enable lifelong professional learning, enhance organizational learning, and support professional transformation. It underscores the role of AI in continuous professional development for teachers.

The draft AI competency frameworks highlight several areas that are relevant to writing instruction. Specifically, the integration of AI in teaching practices can significantly enhance the teaching and learning of writing skills:

Related to *AI Foundations & Applications*, teachers can leverage AI tools to assist in teaching writing. This includes using AI-driven applications to provide immediate feedback on grammar, style, and coherence, thereby helping students improve their writing skills in real time.

AI Pedagogy is about how the use of AI-assisted teaching methods can transform writing instruction. For instance, AI can be integrated into writing tasks to provide personalized prompts, suggest vocabulary enhancements, and even evaluate the structure of essays. AI tools can support teachers in designing more engaging and interactive writing exercises.

Human-centered Mindset highlights the importance of understanding the ethical implications of AI in writing instruction. Teachers must ensure that the use of AI respects student privacy and promotes fair and unbiased feedback. Encouraging students to reflect critically on the use of AI in writing can also foster a deeper understanding of the technology and its impact.

AI for Professional Development means teachers can use AI to enhance their own professional skills in writing instruction. This includes participating in AI-driven training modules, using AI to analyze and improve their teaching methods, and staying updated with the latest advancements in AI applications for education.

The UNESCO Draft AI Competency Frameworks provide a structured approach to incorporating AI into educational practices, with significant implications for writing instruction. By developing AI literacy and integrating AI tools, teachers

can enhance their teaching methods and support students in developing strong writing skills in an increasingly digital world.

Overall, while these competencies do not explicitly address AI, they try to provide a foundation for integrating AI tools in writing instruction. The competencies related to the use of digital skills, in particular, align with the potential use of AI in the classroom. However, it is essential to note that there is currently no comprehensive framework specifically addressing the intersection of writing instruction and AI. This highlights a gap in existing competency frameworks for language teachers.

Previous Studies that Focus on AI, Writing Instruction, and Teacher Competencies

Bruneault et al. (2022) present a competency-based framework for integrating AI ethics into higher education. Their approach, grounded in John Dewey's pragmatist philosophy, identifies core ethical competencies such as ethical sensitivity, reflexive capacities, and dialogical capacities. The study emphasizes the importance of combining ethical training with technical, social, and normative knowledge to address the multifaceted nature of AI ethics effectively. Their comprehensive framework aims to guide the development of educational programs that equip students with the necessary skills to navigate the ethical landscape of AI.

Mikeladze et al. (2023) critically examine AI competence frameworks for educators and identify five distinct types: (1) existing competence model-oriented, (2) competence areas/domains-oriented, (3) process-driven, (4) AI systems-driven, and (5) competence level-driven. The authors highlight the need for empirical research to validate these frameworks and stress the importance of balancing

theoretical models with practical applications. Their review serves as a foundation for further exploration of AI competence frameworks in education.

Kim and Kwon (2022) investigated the AI teaching competencies of elementary school teachers in South Korea. Their study, involving 67 teachers in AI-leading schools, reveals a significant gap in teachers' confidence regarding AI content knowledge compared to their technological and pedagogical knowledge. The research identifies five key themes influencing AI education and derives 22 AI competencies for teachers. The framework outlines key competencies in content knowledge (CK), pedagogical knowledge (PK), technological knowledge (TK), pedagogical content knowledge (PCK), technological content knowledge (TCK), and technological pedagogical knowledge (TPK). CK focuses on a foundational understanding of AI, ethics, and the AI workforce. PK emphasizes classroom management, project-based learning, and fostering a growth mindset. TK highlights the integration and openness of AI tools in education. PCK involves designing AI curricula tailored to student development, while TCK covers AI literacy and ethical considerations. TPK involves using AI tools for assessment and organizing student work digitally. The framework aims to contribute to the understanding of AI teaching practices in the context of South Korean education.

Muttaqin (2022) examines the competencies required by educators to adapt to the evolving landscape of AI in education. The study emphasizes the importance of professional, pedagogical, personal, and social competencies for the effective integration of AI technologies in teaching practices. The study argues for enhanced pedagogical skills and strategic use of digital tools to foster engaging learning environments and concludes that comprehensive training and support systems are crucial for improving educators' AI-related competencies.

Long and Magerko (2020) explore the concept of AI literacy, defining it as a set of competencies that enable individuals to critically assess AI technologies, communicate and collaborate with AI systems, and effectively use AI in various contexts. The authors synthesize interdisciplinary research to outline core competencies of AI literacy and provide design considerations for creating AI learning experiences. Their work serves as a foundational guide for researchers and educators seeking to enhance AI literacy and inform future empirical studies and curriculum development.

While these studies provide valuable insights into AI competencies and literacy in education, there is currently no research that specifically addresses an AI competency framework for EFL writing teachers through an interpretive meta-synthesis approach. The existing studies have focused on general education contexts, ethics training, or specific geographic locations, but none have targeted the unique needs of EFL writing instruction. This interpretive meta-synthesis study fills this gap by synthesizing existing knowledge to develop a specialized AI competency framework for EFL writing teachers. This approach allows for a comprehensive understanding of the intersections between AI technology, language teaching pedagogy, and writing instruction, and offers new perspectives and practical guidelines for integrating AI in EFL writing classrooms. When this study contributes a novel and much-needed framework to the field, it thus addresses a specific area that has not been explored in depth by previous research.

CHAPTER 3: METHOD

Introduction

This chapter details the methodology and research design used to address the research questions. It provides an overview of the sampling strategy, data collection and analysis procedures in this study. Additionally, the chapter addresses issues of trustworthiness and ethical considerations.

Research Design

This study aims to explore the current landscape of AI research in the context of writing instruction with the ultimate goal of developing comprehensive and well-informed guidelines for effective EFL writing instruction in the context of AI integration in writing instruction. These guidelines are designed to enhance teachers' ability to effectively and appropriately incorporate AI technologies into their writing pedagogy.

To this end, the study employed meta-synthesis to integrate both quantitative and qualitative research studies (Barnett-Page & Thomas, 2009; Edwards & Kaimal, 2016; Major & Savin-Baden, 2010; Thorne et al., 2004). In this study, interpretations from both qualitative and quantitative studies constituted second-order constructs. Through qualitative analysis using open coding and relational coding, these second-order constructs were extracted to create third-order constructs. The researcher then generated fourth-order constructs, first in the form of core categories, which served as the for formulating research-based guidelines for writing instruction (Drisko, 2020; Schütz, 1962; Sim & Mengshoel, 2022; Zimmer, 2004).

Although this study is fundamentally qualitative in nature, the Preferred

Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 was utilized to ensure a focused, objective, and transparent selection of studies despite PRISMA's typical association with quantitative research. The PRISMA 2020 guidelines are an updated set of standards designed to improve the transparency, completeness, and accuracy of systematic reviews and meta-analyses. These guidelines are a crucial tool for researchers because of their role in helping researchers plan, conduct, and report their reviews in a comprehensive and structured manner (Joshua et al., 2023; Sewell et al., 2023). The PRISMA 2020 statement consists of a 27-item checklist and a flow diagram, and these cover all aspects of the review process, from the formulation of research questions to the presentation of results. By following these guidelines, researchers can ensure that their reviews are methodologically sound, reproducible, and minimize bias (Page et al., 2021). The PRISMA 2020 guidelines are widely accepted in the academic community and are often required by journals for the publication of systematic reviews and meta-analyses. They enhance the quality of individual studies and contribute to the overall reliability and usefulness of synthesized evidence in informing policy, practice, and future research directions across various fields, including education and technology integration (Moher et al., 2009).

Sampling

This study employed a purposive sampling strategy to select relevant recent articles for inclusion. Specifically, the study focused on articles published in quartile 1 (Q1) and 2 (Q2) journals as ranked by the 2023 Scimago Journal Rank (n.d.). This criterion was chosen to ensure the inclusion of high-quality and impactful research in the field of AI and writing instruction. The SJR ranking is a widely recognized metric that measures the scientific influence and prestige of scholarly journals based

on citation data. By focusing on these high-ranking publications, the study aimed to incorporate research that has undergone rigorous peer review, demonstrated significant scholarly impact, and represented the forefront of knowledge in the field. A total of 103 empirical articles are included in this study.

Method of Data Collection

Data collection was composed of three phases guided by PRISMA 2020 (Page et al., 2021):

- I. Identification
- II. Screening
- III. Eligibility
- IV. Inclusion

The identification phase involved a comprehensive search across four major academic databases: ProQuest, EBSCOhost, Scopus, and Web of Science. The search strategy employed a combination of keywords related to AI and writing tailored to each database's specific requirements. The keywords included “artificial intelligence” OR “AI” OR “a.i”. OR “ChatGPT” (Title) and “writing” (Title) OR “essays” (Title). These keywords were combined using Boolean operators to create comprehensive search strings. The search was conducted between the dates 05.04.2024 and 10.05.2024. The search was limited to peer-reviewed empirical articles published in English within the last five years to ensure the currency of the information, given the rapid developments in AI technology. This systematic search strategy aimed to identify a diverse and representative sample of studies examining the intersection of AI and writing instruction while forming a foundation for in-depth analysis (Table 1).

Table 1

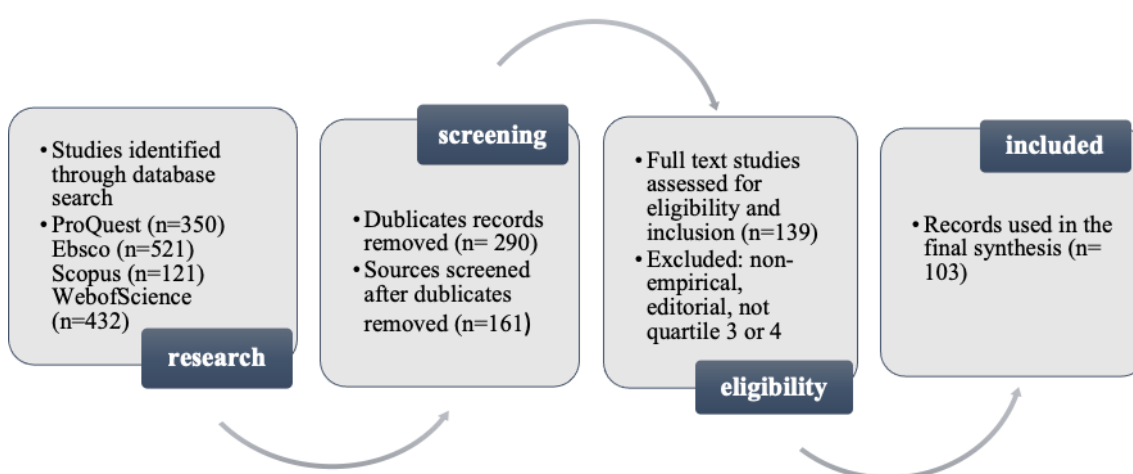
The Flow Diagram Structure by PRISMA

PRISMA Flow	
<i>Identification</i>	Number of records identified through database searching. Number of records identified through other sources.
<i>Screening</i>	Number of records after duplicates removed. Number of records screened. Number of records excluded.
<i>Eligibility</i>	Number of full-text articles assessed for eligibility. Number of full-text articles excluded, with reasons.
<i>Included</i>	Number of studies included in the qualitative synthesis. Number of studies included in quantitative synthesis (meta-analysis).

The screening phase involved removing 290 duplicates from the initial pool. The eligibility phase focused on selecting the full texts of the remaining records and excluding those published in Q3 and Q4 articles. Following these rigorous screening procedures, a final set of 103 studies was selected for in-depth- analysis (Figure 2).

Figure 2

The Flow Diagram Followed in the Data Collection of the Study



Note. Adapted from "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," by M. J. Page et al., 2021, BMJ, 372(71), p. 5 (<https://doi.org/10.1136/bmj.n71>).

In a nutshell, the PRISMA flow diagram illustrates the systematic review process for this meta-synthesis study. It begins with the research phase, where 1,424 studies were identified through searches in four databases: ProQuest (350), Ebsco (521), Scopus (121), and Web of Science (332). The screening phase follows, removing 290 duplicate records and further screening the remaining 1,134 studies, leaving 161 for full-text assessment. In the eligibility phase, these 161 studies underwent detailed evaluation, excluding 58 that did not meet specific criteria (e.g. not published in Q1/Q2 journals), resulting in 139 eligible studies. Finally, the included phase shows that 103 records were used in the final synthesis. This systematic approach aimed to ensure a thorough and unbiased selection of studies to enhance the reliability of the meta-synthesis results.

Methods of Data Analysis

Following the selection of 103 studies, a systematic inductive coding process was initiated, unfolding in four distinct phases. First, the researcher thoroughly read each article to gain an initial understanding of the scope and focus of each article. Next, the abstract of each study was analyzed to identify key elements, including purpose and research questions. Subsequently, the researcher concentrated on the results and discussion sections. Finally, the researcher examined the implications reported by the authors in terms of research and practice.

The data analysis followed a systematic and iterative approach. Initially, a layered open coding strategy was employed, involving both parent coding and first-level coding. In the parent coding phase, broader, overarching codes were assigned to capture the general concepts within the data. Subsequently, first-level coding was used to generate more specific sub-categories and codes, adding greater detail to the analysis (Major & Savin-Baden, 2012; Miles et al., 2014). This initial phase was

exploratory and descriptive in nature. To systematically organize the data, keywords related to writing instruction and AI were extracted from each study and recorded in a spreadsheet (Akşit, 1998; Yıldırım & Şimşek, 2021) (Figure 3). This process yielded a wide range of codes and categories, the *efficacy of ChatGPT-generated abstracts*, *AI-assisted plagiarism*, *essay writing performance*, and *writing process*.

Figure 3

Layered Open Coding: Parent & First Level Sample Categories and Codes

	A	B	C	D	E	F	G	H
1	article	title	abstract	purpose	research questions	Code	parent coding	first coding
20	P6	Not quite eye to A	Generative artificial inte	With the goal of identifying son	1. What are undergraduate student	INT	perceptions on the app	educators' (n = 68
21	P6	Not quite eye to A.I.: student and teacher	With the goal of identifying son	With the goal of identifying son	1. What are undergraduate student	WP	writing process	GenAI in the writin
22	P7	ChatGPT-3.5 as w	ChatGPT-3.5, an AI language model capable of text gene	However, it is still unknown how Ci	However, it is still unknown how Ci	WAT	ChatGPT-3.5 as an essay	how ChatGPT perf
23	P7	ChatGPT-3.5 as writing assistance in students' essays	However, it is still unknown how Ci	However, it is still unknown how Ci	However, it is still unknown how Ci	W-Pe	essay-writing performan	whether ChatGPT
24	P8	AI-generated feed	The question of how ge	to report on two longitudinal st	1. Does the application of AI-gener	FD	AI-generated feedback	the efficacy of the
25	P8	AI-generated feedback on writing: insights	to report on two longitudinal st	1. Does the application of AI-gener	1. Does the application of AI-gener	AWE	the efficacy of ChatGPT	the efficacy of Che
26	P8	AI-generated feedback on writing: insights	to report on two longitudinal st	1. Does the application of AI-gener	1. Does the application of AI-gener	FD	AI-generated feedback	the preference for
27	P8	AI-generated feedback on writing: insights	to report on two longitudinal st	1. Does the application of AI-gener	1. Does the application of AI-gener	LP	linguistic progress / lang	the efficacy of the
28	P9	Using AI-based de	The release of ChatGPT	to (a) assess the effectiveness c	(a) How effective is GPT-2 Output	DT	the effectiveness of ava	the comparison of
29	P9	Using AI-based detectors to control AI-as	to (a) assess the effectiveness c	(a) How effective is GPT-2 Output	(a) How effective is GPT-2 Output	PL	AI-assisted plagiarism	the challenge of A
30	P9	Using AI-based detectors to control AI-as	to (a) assess the effectiveness c	(a) How effective is GPT-2 Output	(a) How effective is GPT-2 Output	CL-E	classification effectiveness	the difference in
31	P9	Using AI-based detectors to control AI-as	to (a) assess the effectiveness c	(a) How effective is GPT-2 Output	(a) How effective is GPT-2 Output	G	to guide educators	an in-depth invest
32	P10	Strengths and We	Background: ChatGPT is	to evaluate the ability of GPT-4	In this study, an evaluation was per	AB	GPT-4's ability to genera	GPT-4's ability to j
33	P10	Strengths and Weaknesses of ChatGPT Models for Scientific Writing Abou	In this study, an evaluation was per	In this study, an evaluation was per	In this study, an evaluation was per	Q	quality and characteristi	focusing on their c
34	P11	Exploring the Role	This study examined the	This study examined the robust	1. How robust are LLMs (GPT-4, GP	WA	the role of four large lan	the robustness an
35	P11	Exploring the Role of Artificial Intelligence in Facilitating Assessment of Wr	1. How robust are LLMs (GPT-4, GP	1. How robust are LLMs (GPT-4, GP	1. How robust are LLMs (GPT-4, GP	AC	assessing writing accurat	the role of four lai
36	P12	Empowering Child	ChatGPT is a state-of-th	to explore the potential of Chat	Is there a functional relation bet	WAT	ChatGPT-assisted writin	the potential of C

Following this initial coding, where raw data is examined line by line and assigned descriptive labels, relational or axial coding was utilized, a more interpretive phase, where patterns and relationships within the coded data were identified. This process involved grouping codes that shared common concepts or characteristics into broader categories, as illustrated in Figure 4. This phase involved connecting codes to broader categories such as *textual features*, *perceptions*, and *detection* (Major & Savin-Baden, 2012; Miles et al., 2014).

Figure 4*Relational Coding: Sample Categories and Codes Based on Recurring Patterns*

1	Recurring Themes / Patterns from the Studies									
2	AA	attitude assessment	CL	classification	FLU	fluency	P	personalization	Ss	sustainability
3	AAWRW	AI-assisted writing in the work	CI	creative identity	H vs AI	Human vs AI	PA	practical applications	SW	scientific writing
4	AB	ability	CIBW	collaborative writing	H&A CL	human AI collaboration	PE	prompt engineering	SWS	scholarly writing stand
5	AC	accuracy	CO	correctness	H-AI-Int	human AI interaction	PL	plagiarism	T	trust
6	ACC	acceptance	COMP	comparison	HCA	human critical appraisal	PnG	policies and guidelines	TAT	teacher assistance too
7	AcMc	academic misconduct	Conf	confidence	HL	hallucination	PR	precision	TF	textual features
8	ACU	acceptable use	COT	contextualization	ID	identification of AI-generated texts by hu	PRF	proficiency	TnS	tone and style
9	AD	academic dishonesty	CP	cognitive processes	ImpOWP	impact on writing practice	PV	prediction variables	TP	text policies
10	ADP	AI adoption	CPB	capability	InAIin	inaccurate information	Q	quality	Tr	training set
11	AE	academic emotion	CR	creativity	IND	indicators of AI text	R	reliability	TRP	transparency
12	AEE	automated essay evaluation	CrA	critical analysis	INDP	independence	RD	readability	TV	task value
13	AI-INT	AI integration	CRD	credibility	INQ	inequality	RE	readiness	TW	thesis writing
14	AI-lit	AI literacy	Cri	criteria	INT	integrity	Rex	reviewer experience	V	voice
15	AI-S	AI-assisted writing scoring	CrPr	current practices	IoWE	identification of writing errors	RK	register knowledge	WAT	writing assistance tool
16	AI-T	AI training	CU	comment updating	IP	international publication	RL	relationship bw	WFL	writing feedback litera
17	AJ	academic judgement	CW	creative writing	IRV	independent research variables	RSP	responsibility	WI	writing instruction
18	AJ	AI journalism	DC	discourse components	LC	language characteristics	RT	reflective thinking	WM	writing motivation
19	AS	assessment	DC	declaring the use of generative	LF	linguistic features	Rtn	retention	WO	writing performance
20	A-S	automated scoring	DMC	digital multimodal composing	Lim	limitations	RW	research writing	WP	writing performance
21	AS-S	assessment strategies	DS	digital stories	LP	linguistic progress	Sed	self editing	WSc	writing score
22	ATT	attitudes	DT	detection	M	motivation	Self	self-efficacy	WSK	writing skills

In the final stage of analysis, selective or structural coding was used to integrate the multiple relational categories into core categories (Figure 4). This process aimed to identify central unifying concepts that capture the essence of the findings and provide a concise yet comprehensive understanding of the complex interplay between AI and writing instruction. These overarching categories served as a foundation for informing the development of guidelines to inform a competency framework (Major & Savin-Baden, 2012; Miles et al., 2014).

Table 2*Core/Overarching Categories*

Core/Overarching Categories	Relational Coding: Connecting Categories
Good Academic Conduct in Writing	✓ Academic dishonesty ✓ Academic integrity ✓ Plagiarism ✓ AI-generated text Authorship and Ownership ✓ Transparency and Disclosure ✓ Ethical use
Assessment and Writing	✓ Feedback ✓ Automated Evaluation/ Scoring ✓ Assessment of Writing

Evaluating and Enhancing
the Quality and Impact of
Academic Content

- ✓ Performance and Impact
 - Capabilities
 - Accuracy
 - Proficiency and Originality
 - Reliability and Quality
 - Usefulness for Scholars
 - Impact Efficacy
- ✓ Output Characteristics
 - Quality and Accuracy of Academic Content
 - Authenticity
 - Creativity
 - Originality

User Behavior and
Pedagogical Implications of
AI-Assisted Writing

- ✓ Detection
- ✓ Identification
- ✓ Machine Learning
- ✓ Acceptance
- ✓ Adoption
- ✓ Attitudes and Challenges
- ✓ Self-efficacy
- ✓ Users and Uses
- ✓ Trust
- ✓ Motivation
- ✓ Independence and Proficiency

AI-Integrated Writing
Instruction and Outcomes

- ✓ Writing Skills
- ✓ Writing Performances
- ✓ Writing Process
- ✓ AI-Enhanced Classroom
- ✓ Writing Quality
- ✓ Scaffolding
- ✓ AI-Assisted Writing

Diverse Writing Genres and
Modalities

- ✓ Creative Writing
 - ✓ Business Communication
 - ✓ Multimodal Writing
 - ✓ News Writing
 - ✓ Narrative English Writing
 - ✓ Graphic Storytelling
 - ✓ Digital Stories
 - ✓ Picture-cued Writing
 - ✓ Multiple Genres (Essays, Stories, Poetry, and Python code)
-

Linguistic and Discourse Features of AI-Generated Texts	✓ Textual and Linguistic Features ✓ Discourse Components ✓ Style Register

Throughout this process, a spreadsheet program (Akşit, 1998; Yıldırım & Şimşek, 2021) was utilized to organize, sift, sort, and analyze the data, facilitating a systematic and comprehensive examination of the compiled studies.

Trustworthiness

The researcher used Lincoln and Guba's (1985) evaluative criteria to establish validity and reliability (Table 3).

Table 3

The Evaluative Criteria Used for Validity and Reliability in Qualitative Research

	The Criteria (Lincoln and Guba, 1985)	The Techniques
Criterion 1	Credibility refers to the confidence in the truth of the findings.	• Using second-order interpretations derived from the studies • Transparency
Criterion 2	Transferability refers to the extent to which the finding can be applied or transferred to other contexts or settings.	• Thick description • Purposeful sampling
Criterion 3	Dependability refers to the consistency of the findings over time and across researchers.	• Constant double/triple checking with the thesis advisor during coding and categorizing, and making constant arrangements based on given feedback.
Criterion 4	Confirmability refers to the degree the which the findings are shaped by the data collected rather than the researcher's biases.	• Using direct quotes from the studies

This study adhered to four key criteria to ensure the validity and reliability of the findings. Credibility, the first criterion, was established through the use of

second-order interpretations derived from the included studies and maintaining transparency throughout the research process. This involved extracting and analyzing the original authors' interpretations rather than raw data and providing clear, detailed descriptions of selection and analysis methods. Transferability, the second criterion, was addressed through thick description and purposeful sampling. The study included rich, contextual details about the original studies and the findings while also selecting studies that offered diverse, relevant information to the research questions.

The third criterion, dependability, was ensured through regular consultation with the thesis advisor on a weekly basis. This involved regular double and triple-checking during the coding and categorizing processes, with ongoing adjustments based on feedback. This collaborative approach helped maintain consistency in the findings over time.

Finally, confirmability was established by using direct quotes from the original studies. By grounding the interpretations in the original text, the study demonstrated that the findings were shaped by the data rather than personal biases. These four criteria and their associated steps worked together to enhance the overall objectivity of this meta-synthesis and to create a solid foundation for the research findings.

The study also followed The ENTREQ (Enhancing Transparency in Reporting the Synthesis of Qualitative Research) statement to ensure transparency and comprehensiveness in the synthesis and reporting. The ENTREQ statement consists of 21 items grouped into five main domains. It aims to help researchers provide a comprehensive account of the synthesis process (Tong et al., 2012).

The ENTREQ statement, developed by Tong et al. in 2012, is a comprehensive guideline designed to enhance transparency, reproducibility, and

critical appraisal of qualitative syntheses by providing a structured framework for researchers to comprehensively report their methods and findings, including meta-syntheses. It comprises 21 items grouped into five main domains: Introduction, Methods, Literature Search and Selection, Analysis and Synthesis Process, and Results. These items cover crucial aspects of the synthesis process, from stating the research question and identifying the synthesis methodology to detailing the search strategy and study selection criteria to describe the data extraction, analysis, and synthesis procedures.

By addressing each applicable item, this study aimed to ensure a clear, detailed account of the synthesis process. First, the objectives and research questions were clearly stated, with a solid rationale for the synthesis. Inclusion and exclusion criteria for selecting studies were detailed, and a comprehensive search strategy was developed with specific databases, keywords, and the search timeframe explained. The study selection process was documented, with decisions made collaboratively by the thesis advisor to minimize bias. Data extraction was systematically recorded, specifying what data was extracted and how it was used in the synthesis. The quality of the included studies was assessed using established criteria to ensure the synthesis was built on robust and credible evidence. During the synthesis process, the chosen approach was clearly explained, along with how discrepancies and differences in findings were addressed. The final synthesis provided detailed findings from the included studies, which highlighted the patterns, concepts, and new insights that emerged from the analysis.

By implementing these two different trustworthiness measures, as a researcher, I wanted to establish a robust foundation for addressing each of the research sub-questions. For sub-question (a), I wanted to ensure that the key concepts

I identified were firmly grounded in the original studies and not misinterpreted, preserving the integrity of the source material. I aimed my analysis of the interrelationships between these concepts, addressing sub-question (b), to become consistent and verifiable, allowing other researchers to follow my analytical process. The core categories I developed in response to sub-question (c) ensured credibility and transferability, making them applicable across various EFL writing instruction contexts. Finally, the guidelines I proposed for writing instruction, as outlined in sub-question (d), were based on a rigorous and transparent process. This methodological rigor aims not only to enhance the validity of the findings but also to increase the potential for suggested guidelines to be adopted in practical settings and to serve as a springboard for further research in the field of AI-informed EFL writing instruction.

CHAPTER 4: RESULTS

Introduction

This chapter outlines the findings derived from the comprehensive analysis of the selected corpus of literature. The results are systematically categorized according to overarching core categories that emerged from the analysis. Within each thematic category, first, the research objectives and the research questions of the constituent studies are given. These objectives were derived from the clearly stated purposes and the research questions stated in the articles. Then, the researcher synthesizes the empirical results from these investigations. Finally, the practical implications that the articles suggest are presented.

Good Academic Conduct in Writing

Among the core concepts identified in the analysis, several pertain to *Good Academic Conduct in Writing*. Figure 5 shows the subcategories.

Figure 5

Good Academic Conduct in Writing



Within this broader category, distinct subcategories emerged, including *academic dishonesty*, *academic integrity*, *plagiarism*, *authorship* and *ownership*, *transparency* and *disclosure*, and *ethical use*.

Academic Dishonesty

Purpose

Two of the studies out of 103 articles aim to investigate academic dishonesty. Both of the studies look at teacher perceptions to gather their views related to academic dishonesty. One study also investigates students' perceptions related to academic dishonesty.

Table 4

Academic Dishonesty: Purpose

Title	Purpose	Research Question	Author
Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence	"To examine teachers' perceptions of academic dishonesty in the writings of EFL students in the context of AI"	"RQ1: How do teachers perceive the use of AI by their students for academic purposes? RQ2: How do teachers reflect upon the impact of AI on the prevalence of and commitment to academic dishonesty in students' writings? RQ3: What strategies and approaches do teachers employ to address and deter AI-driven academic dishonesty and how do they reflect on the effectiveness of these interventions?"	Mohammadkarimi (2023)
Comparing Student and Writing Instructor Perceptions of Academic Dishonesty When Collaborators Are Artificial Intelligence or Human	"To delve into the multifaceted role of artificial intelligence (AI) in writing instruction with a particular emphasis on AI-based writing tools like Grammarly and QuillBot"	"RQ1: For writing instructors is there a differential perception of academic dishonesty in collaborative writing scenarios when the identity of the collaborator is varied between a human and an AI agent? If so what are the differences? RQ2: For students is there a differential perception of academic dishonesty in collaborative writing scenarios when the identity of the collaborator is varied between a human and an AI agent? If so what are the differences? RQ3: How do the answers in RQ1 and RQ2 compare?"	Gallagher and Wagner (2024)

The first study (Mohammadkarimi, 2023) aims to investigate teachers' perceptions within the context of AI, specifically how teachers perceive the use of AI by EFL students for academic purposes, the impact of AI on academic dishonesty, and the strategies teachers employ to address AI-driven dishonesty.

The second study (Gallagher & Wagner, 2024) focuses on collaborative writing scenarios and aims to understand and compare the views of students and teachers in terms of their perceptions of academic dishonesty in these collaborative writing scenarios.

Results

Table 5 presents the results from these studies exploring AI and academic dishonesty.

Table 5

Academic Dishonesty: Results

Title	Results	Author
Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence	# "Teachers unanimously agreed on the negative influence of AI on students' commitment to academic honesty, perceiving it as enabling dishonesty and hindering skill development. Furthermore, the interviews brought to light that the failure of teachers to identify AI-generated assignments is another contributing factor to students engaging in academic dishonesty." # "The second part of the questionnaire investigated the impact of using AI to be academically dishonest. Unquestionably, teachers unanimously concurred regarding the adverse influence of AI on the academic integrity of their students. They held the belief that AI had amplified the accessibility and allure of academic dishonesty for students, impeding the cultivation of fundamental general and transferable skills." # "The second part of the questionnaire investigated the impact of using AI to be academically dishonest. Unquestionably, teachers unanimously concurred regarding the adverse influence of AI on the academic integrity of their students. They held the belief that AI had amplified the accessibility and allure of academic dishonesty for students, impeding the cultivation of fundamental general and transferable skills." # "Strategies identified include problem-solving activities, plagiarism detection tools, and integration of AI in teaching practices. Teachers expressed divergent opinions and presented a variety of strategies to	Mohammadkarimi (2023)

Title	Results	Author
	effectively tackle and identify instances of academic dishonesty arising from AI implementation. These strategies encompassed incorporating problem-solving, critical-thinking activities, real-life and personal examples, encouraging students to express their own ideas, using plagiarism checker software such as Turnitin, using AI to write the assignment and comparing with students' writings, incorporating AI in teaching, and writing the first draft in the classroom."	
Comparing Student and Writing Instructor Perceptions of Academic Dishonesty When Collaborators Are Artificial Intelligence or Human	# "While some teachers acknowledged challenges in detecting AI-related academic dishonesty, the study underscores the importance of comprehensive training and support for teachers to utilize AI effectively while preserving academic integrity." # "Results suggest similar perception patterns among students and instructors, with both populations expressing significant differences in perceived dishonesty between AI and human collaborators in some scenarios." # "When compared to human collaborators, AI agents of collaboration are perceived as more academically dishonest for both writing instructors and students when the AI produces text."	Gallagher and Wagner (2024)

In the first study (Mohammadkarimi, 2023), teachers unanimously express concerns about AI's negative impact on academic integrity. They believe AI facilitates dishonesty, hinders skills development, and challenges their ability to detect AI-generated work. The strategies they suggest employing include activities like problem-solving and critical thinking, the use of AI detection tools, and the integration of AI into teaching practices. In the second study (Gallagher & Wagner, 2024), the results show that teachers need training and support and have similar perceptions related to the use of AI as a collaborator in promoting academic dishonesty. Also, in the second study (Gallagher & Wagner, 2024), participants perceive AI collaborators as more likely to engage in academic dishonesty than human collaborators, especially when the AI generates text. Despite this, the study emphasizes the importance of training and support for teachers to utilize AI to preserve academic integrity.

Implications

Table 6 presents the implications of these studies. The studies reveal the need for comprehensive training and support for teachers.

Table 6

Academic Dishonesty: Implications

Title	Implications	Author
Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence	# "the recognition of teachers' perceptions regarding the advantages of AI technologies highlights the need for incorporating these tools effectively into the educational process. Educators should be provided with training and support to harness the potential benefits of AI while also addressing concerns related to academic dishonesty." # "the unanimous agreement among teachers on the negative impact of AI on academic integrity emphasizes the urgency of addressing this issue. Institutions and policymakers should prioritize developing ethical guidelines and policies that promote responsible and ethical use of AI technologies in education." # "...the recognition of teachers' responsibility in identifying AI-generated assignments as a contributing factor to academic dishonesty emphasizes the significance of enhancing teachers' awareness and proficiency in identifying and addressing instances of plagiarism facilitated by AI. It is crucial to develop professional development initiatives and resources aimed at equipping teachers with the requisite skills and knowledge to effectively identify and prevent academic dishonesty stemming from AI." # "...the research identifies a number of techniques, such as the incorporation of problem-solving and critical-thinking activities, utilization of plagiarism detection tools, and integration of AI in teaching practices, which offer practical approaches to tackling and mitigating AI-related academic dishonesty. Educators and educational institutions could explore using these tactics, customizing them to their individual settings and needs."	Moham madkari mi (2023)
Comparing Student and Writing Instructor Perceptions of Academic Dishonesty When Collaborators Are Artificial Intelligence or Human	"... We suggest that separate training programs may not need to be developed for writing instructors and students about AIWT or non-AIWT with respect to academic dishonesty. For policy and teaching practices, then, we suggest being explicit about AIWT use when it comes to the production of text. Policy makers and writing instructors might address situations that involve the production of text in concrete ways. We believe that policies about AIWT may not need to be substantially different than current policies involving human collaboration. We suggest, the need for addressing challenge of academic dishonesty in tangible, real-world contexts."	Gallaghe r and Wagner (2024)

In these two studies, Mohammadkarimi (2023) emphasizes the importance of integrating AI technologies into education while addressing the challenges of academic dishonesty. They highlight the need for teacher training and support to effectively use AI, along with the development of ethical guidelines to mitigate the

negative impact of AI on academic integrity. Furthermore, they stress the role of educators in identifying AI-generated assignments and suggest incorporating problem-solving and critical-thinking activities, as well as plagiarism detection tools, to combat AI-related dishonesty. Similarly, Gallagher and Wagner (2024) argue that while AI and human collaborations in writing do not require separate training programs, explicit policies are necessary to address academic dishonesty in both AI and non-AI writing contexts. They advocate for concrete, real-world approaches to policy-making and teaching practices to manage the challenges posed by AI in academic settings.

Academic Integrity

Purpose

Six of the articles out of 103 articles aim at exploring academic integrity.

Table 7

Academic Integrity: Purpose

Title	Purpose	Research Question	Author
Algorithmically-driven writing and academic integrity: exploring educators' practices, perceptions, and policies in AI era	“to respond to calls for empirical investigation concerning the affordances and encumbrances of ADWTs, and their implications for academic integrity.”	“1. What are the perceptions of educators regarding the use of these tools and their implications for academic integrity? 2. What are the current practices of educators in using ChatGPT, DWAs, MTs, and APTs? 3. What are the policies provided by their institutions and their own suggested policies for the meaningful and ethical use of these tools?”	Gustilo et al. (2024)
Not quite eye to A.I.: student and teacher perspectives on the use of generative artificial intelligence in the writing process	“This study reports the results of a survey comparing educators' and university students' perceptions on the appropriate use of GenAI in the writing process.”	“1. What are undergraduate students' and teachers' perspectives on using GenAI in the writing process (brainstorming, outlining, writing, revising, evaluation, and feedback)? 2. How do student and teacher perspectives on the use of GenAI in the writing process compare?”	Barrett and Pack (2023)
Evaluating the efficacy of AI content detection	“The proliferation of artificial intelligence (AI)-generated	Not stated.	Elkhatat (2023)

Title	Purpose	Research Question	Author
tools in differentiating between human and AI-generated text	content, particularly from models like ChatGPT, presents potential challenges to academic integrity and raises concerns about plagiarism. The present study aims to investigate the capabilities of several recently launched AI content detectors and classifier tools in discerning human-written and AI-generated content.”		
Detecting AI-Generated Essays: The ChatGPT Challenge	“to discuss a method and necessary strategies to confront the challenge of maintaining academic integrity in all educational settings”	Not stated.	Cingillioglu (2023)
How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing?	“This study aimed to evaluate the integrity and accuracy of the responses to the question it was asked from the ChatGPT software program.”	Not stated.	Dashti et al. (2023)
Comparing Student and Generative Artificial Intelligence Chatbot Responses to Organic Chemistry Writing-to-Learn Assignments	“The emergence of generative artificial intelligence (AI) technology, such as chatbots ChatGPT and Bard, raises concerns regarding the value of open-ended writing tasks in the classroom; one concern involves academic integrity and whether students will use these chatbots to produce sufficient responses to open-ended writing tasks. The present study	“1. How do generative artificial intelligence chatbots exhibit evidence of mechanistic reasoning in their responses to organic chemistry writing-to-learn assignments? 2. How do generative artificial intelligence chatbot responses compare to authentic student responses for organic chemistry writing-to-learn assignments?”	Watts et al. (2023)

Title	Purpose	Research Question	Author
	investigates the degree to which generative AI chatbots exhibit mechanistic reasoning in response to organic chemistry WTL assignments.”		

The first two studies (Gustilo et al., 2024; Barrett & Pack, 2023) seek to explore the views of how AI writing tools are perceived, used, and regulated in academic environments, with a focus on balancing their potential benefits with concerns about academic integrity.

The third study (Elkhataat et al., 2023) investigates the capabilities of AI content detectors and classifier tools, while the fourth study (Cingillioglu, 2023) discusses a method to face the challenge of maintaining academic integrity in educational settings. The fifth and sixth studies (Dashti et al., 2023; Watts et al., 2023) evaluate ChatGPT responses for academic integrity.

Results

Table 8 presents the results from these studies that focus on academic integrity.

Table 8

Academic Integrity: Results

Title	Results	Author
Algorithmically-driven writing and academic integrity: exploring educators' practices, perceptions, and policies in AI era	# “The study found that teacher respondents highly value the diverse ways ADWTs can support their educational goals (perceived usefulness). However, they must overcome their barrier threshold such as limited access to these tools (perception of external control), a perceived lack of knowledge on their use (computer self-efficacy), and concerns about ADWTs' impact on academic integrity, creativity, and more (output quality).”	Gustilo et al. (2024)
Not quite eye to A.I.: student and teacher perspectives on the	# “...minor disagreement between students and teachers on the acceptable use of GenAI tools in the writing process. For all the steps of the writing process, students and	

Title	Results	Author
use of generative artificial intelligence in the writing process	teachers generally agreed that using GenAI to brainstorm ideas or model answers, or as a form of cognitive offloading for tasks that the user is already competent in, is acceptable. Both students and teachers perceived GenAI use to be more acceptable in the early stages of the writing process (i.e., brainstorming and outlining) than in later stages.” # “Conversely, using GenAI to complete writing task assignments, with or without disclosing the use of GenAI, is unacceptable. The use of GenAI for writing purposes is viewed as more acceptable when it is fulfilling a supportive role focused on idea generation and organization rather than when leveraged as an automatic writing completion tool.”	Barrett and Pack (2023)
Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text	“Findings reveal that the AI detection tools were more accurate in identifying content generated by GPT 3.5 than GPT 4. However, when applied to human-written control responses, the tools exhibited inconsistencies, producing false positives and uncertain classifications. Looking at the GPT 3.5 results, the OpenAI Classifier displayed the highest sensitivity, with a score of 100%, implying that it correctly identified all AI-generated content. However, its specificity and NPV were the lowest, at 0%, indicating a limitation in correctly identifying human-generated content and giving pessimistic predictions when it was genuinely human-generated. GPTZero exhibited a balanced performance, with a sensitivity of 93% and specificity of 80%, while Writer and Copyleaks struggled with sensitivity. The results for GPT 4 were generally lower, with Copyleaks having the highest sensitivity, 93%, and CrossPlag maintaining 100% specificity. The OpenAI Classifier demonstrated substantial sensitivity and NPV but no specificity.”	Elkhatat et al. (2023)
Detecting AI-Generated Essays: The ChatGPT Challenge	“Via support vector machine (SVM) algorithm 100% accuracy was recorded for identifying human generated essays. The author discussed the key use of Recall and F2 score for measuring classification performance and the importance of eliminating False Negatives and making sure that no actual human generated essays are incorrectly classified as AI generated.”	Cingillioglu (2023)
How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing?	“The results indicated that writing credit or references to such content or references in prestigious academic journals are not yet appropriate. This affects scientists, researchers, scholarship proposals, publications, and peer-review process and shows the need for disclaimer statements, education related to the evaluation of scientific evidence, and complete attention to academic integrity.”	Dashti et al. (2023)
Comparing Student and Generative Artificial Intelligence Chatbot Responses to Organic Chemistry Writing-to-Learn Assignments	“Findings indicate that chatbots respond differently to different WTL assignments. Additionally, the chatbots rarely incorporated the discussion of electron movement, a key feature of mechanistic reasoning. Furthermore, the chatbots, in general, do not engage in mechanistic reasoning at the same level as students. We contextualize the results by considering academic integrity with the assumption that students’ intentions are to engage in	Watts et al. (2023)

Title	Results	Author
	academically honest behavior, and we focus on understanding the ethical uses of generative AI for classroom assignments.”	

The first study (Gustilo et al., 2024) concludes that while teachers highly value the diverse support ADWTs provide for educational goals, they face significant barriers such as limited access, lack of knowledge, and concerns about academic integrity and output quality. The second study (Barrett & Pack, 2023) claims there is a minor disagreement between students and teachers on the use of GenAI, with both groups generally agreeing it is acceptable for brainstorming and outlining but not for completing writing tasks without disclosure.

The third study (Elkhataat et al., 2023) finds that AI detection tools were more accurate with GPT 3.5 than GPT 4, but exhibited inconsistencies with human-written content, showing varied sensitivity and specificity across different tools. The fourth study (Cingillioglu, 2023) demonstrates that a support vector machine (SVM) algorithm achieved 100% accuracy in identifying human-generated essays, highlighting the importance of eliminating false negatives to ensure accurate classification.

The fifth study (Dashti et al., 2023) indicates that AI-generated content is not yet suitable for writing credit or references in prestigious journals, underscoring the need for disclaimer statements and education on evaluating scientific evidence to maintain academic integrity. Lastly, the sixth study (Watts et al., 2023) reveals that chatbots respond differently to various assignments and generally do not engage in mechanistic reasoning at the same level as students, which raises concerns about the ethical use of AI in classroom assignments.

Implications

Table 9 presents the implications of these studies. The studies provide valuable insights for academic integrity.

Table 9

Academic Integrity: Implications

Title	Implications	Author
Algorithmically-driven writing and academic integrity: exploring educators' practices, perceptions, and policies in AI era	# “AI in education, in particular, ADWTs, calls for critical awareness of ethical protocols and requires collaborative efforts by AI experts, researchers, and educators to initiate extensive, longitudinal research (Zhang & Aslan 2021) that could generate more ethical guidelines for all stakeholders and spur new transformations and possibilities in education.”	Gustilo et al. (2024)
Not quite eye to A.I.: student and teacher perspectives on the use of generative artificial intelligence in the writing process	“The findings of this study underscore the importance of students and teachers being transparent in their use of GenAI tools. Using AI without disclosure by both students and teachers, was considered the least acceptable practice. If educators are transparent and clearly state and model acceptable uses of GenAI, then students may follow. While there is a nascent body of literature offering practical suggestions and advice for using GenAI in education by educators, learners, and researchers, an evidence-centered framework for leveraging GenAI in writing and in higher education in general needs further attention.”	Barrett and Pack (2023)
Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text	“This study underscores the need for further as AI text generation evolves, so must the tools designed to detect it. This necessitates continuous development and regular evaluation to ensure their efficacy and reliability. Furthermore, a balanced approach involving AI tools and traditional methods best upholds academic integrity in an ever-evolving digital landscape.”	Elkhatat et al. (2023)
Detecting AI-Generated Essays: The ChatGPT Challenge	“AI-generated essays submitted by students can be detected by teachers and educational designers using the proposed language model and machine learning (ML) classifier at a high accuracy. Human (student)-generated essays can and must be correctly identified with 100% accuracy even if the overall classification accuracy performance is slightly reduced. This is the first and only study that used an n-gram bag-of-words (BOWs) discrepancy language model as input for a classifier to make such prediction and compared the classification results of other AI-generated text detection software in an empirical way.”	Cingillio glu (2023)
How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing?	“1. Artificial intelligence is a leading technology in all fields of science, but using it without fact checking can lead to false and unethical results: AI-generated content cannot provide or analyze the same information as an author or researcher. 2.The results indicated that writing credit or references to such content or references in prestigious academic journals are not yet appropriate. This affects scientists, researchers, scholarship proposals, publications, and peer-review process and shows the need for disclaimer statements, education related to the evaluation of scientific evidence, and	Dashti et al. (2023)

	complete attention to academic integrity. 3. Scientific writing is only valid when performed manually by researchers. This study found that researchers must write scientific papers with care when using the ChatGPT software program or other AI tools and should not rely solely on such tools.”	
Comparing Student and Generative Artificial Intelligence Chatbot Responses to Organic Chemistry Writing-to-Learn Assignments	# “To enhance instruction, educators should transparently clarify policies regarding academic integrity and the ethical use of generative AI in the classroom. They should develop activities where students generate chatbot responses and critique them for content errors and lack of depth. It's important to communicate to students that generative AI cannot replace their knowledge and skills.” # “Discussions should be held on the ethical use of AI for certain tasks, such as generating essay ideas, versus tasks that require demonstrating understanding, depending on the learning goals. Educators should provide evidence that chatbots underperform compared to students on specific tasks and implement assignments, like Writing to Learn (WTL), with clear explanations of their learning objectives and pedagogical value.”	Watts et al. (2023)

The implications of the first study (Gustilo et al., 2024) conclude that educators must develop a critical awareness of ethical protocols for AI in education and promote transparency in using generative AI tools to maintain academic integrity. Based on the implications from the second study (Barrett & Pack, 2023), the research on student and teacher perspectives highlights the importance of transparent use of generative AI in education, advocating for clear policies and modeling acceptable practices to guide students. The third study (Elkhatat et al., 2023) proposes that further development and regular evaluation of AI detection tools are necessary to ensure their efficacy in differentiating between human and AI-generated text.

The fourth study (Cingillioglu, 2023) shares that AI-generated essays can be detected with high accuracy using a combination of language models and machine learning classifiers, emphasizing the need for reliable detection methods. According to the fifth study (Dashti et al., 2023), AI in scientific writing must be carefully fact-checked to avoid false results, and researchers should not only rely on AI tools, keeping manual oversight for credibility. The last study (Watts et al., 2023) points

out that educators should transparently clarify policies on the ethical use of generative AI in the classroom and develop activities that help students critically evaluate AI-generated content.

Plagiarism

Purpose

Two studies focus on plagiarism. While the first study (Ibrahim, 2023) tries to measure the effectiveness of AI detection tools and offer ways to develop stronger detection tools, the aim of the second study (Alexander et al., 2023) is to understand the difficulties of teachers trying to detect AI-generated texts and point out the skills and resources for better identification.

Table 10

Plagiarism: Purpose

Title	Purpose	Research Question	Author
Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines”	“to (a) assess the effectiveness of available AI-detection platforms to help L2 educators identify reliable platforms they can use to control AI-assisted plagiarism; (b) offer an in-depth investigation of this new, disruptive, and unexplored research territory to guide educators in dealing with AI-assisted plagiarism; and (c) evaluate underlying classifier type, design, and training mechanisms to contribute to the development of more robust machine-detection classifiers”	“(a) How effective is GPT-2 Output Detector Demo in detecting machine-generated essays? (b) How effective is Crossplag AI Content Detector in identifying machine-generated essays? (c) What is the difference in classification effectiveness between GPT-2 Output Detector Demo and Crossplag Detector? (d) What does the comparison of two RoBERTa-based detectors suggest about the robustness of AI-detection of AI-generated texts?”	Ibrahim (2023)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“to shed light on the challenges faced by English as a Second Language (ESL) lecturers in identifying AI-generated texts, and highlighting the skills and resources needed to enhance their detection capabilities”	“(1) How effective are current plagiarism detectors in providing a reliable measure of detection for identifying originality in academic writing and distinguishing human-generated text from AI-generated text? (2) What criteria are employed by ELS lecturers to assess academic	Alexander et al. (2023)

Title	Purpose	Research Question	Author
		writing?; and (3) How effective are these criteria in distinguishing human-generated text from AI-generated text?"	

Results

The results of the studies that aim at investigating plagiarism are shared in Table 11.

Table 11

Plagiarism: Results

Title	Results	Author
Using AI-based detectors to control AI-assisted plagiarism in ESL writing: "The Terminator Versus the Machines"	# "GPT-2 Output Detector Demo and Crossplag AI Content Detector achieved high accuracy in discriminating between machine-generated and human-written texts with an average accuracy of 89% despite being trained on a smaller and earlier version of GPT." # "GPT-2 Output Detector Demo was more sensitive to human-written texts and could detect them more accurately. At the same time, Crossplag Detector was more sensitive to machine-generated texts and had fewer false negatives." # "...these findings demonstrate that AI-based classifiers could offer viable resources for detecting machine-generated texts and controlling AI-assisted plagiarism, even if they need further development to achieve higher detection accuracy."	Ibrahim (2023)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	# "While the tested AI detectors were accurate in evaluating text fully written by a human or fully produced by ChatGPT, their accuracy dropped significantly when assessing text generated fully by AI but using a human outline (e.g., Essay II) and text which was a mix of human and AI-generated text (Essay III)." # "Participants tended to exploit a deficit model of assessment that focuses on error as an indicator of L2 writing output, with high levels of technical and grammatical accuracy and sophisticated language use as indicators of AI-generated text. They tended to assess human writing based on technical accuracy and error and AI writing based on verbal complexity." # "...limited awareness of the characteristics and metrics used by ChatGPT, as well as lack of attention to the veracity of facts and references generated by ChatGPT. Findings suggest that whilst analysing the texts, the participants paid attention to features of language ranging from spelling mistakes to the ways sentences were formed and the way the subjects were investigated; however, none of the participants reported on fact-checking. These hallucinated references were neither identified nor reported in participants' evaluation of the essay samples."	Alexander et al. (2023)

Title	Results	Author
	# “This study identifies the growing challenges encountered by ESL lecturers and underlines the need for digital literacy training, targeted professional development, the use of advanced detection tools and a review of assessment policies and practices in relation to AI.”	

The results from the studies indicate that AI detection tools are possible resources to deter plagiarism, but for high accuracy, they need to be developed more. Another finding is that when identifying AI-generated texts, participants concentrate on the language features but do not incorporate fact-checking, leading to hallucinated references not being identified. The second study (Alexander et al., 2023) illustrates the importance of training and professional development and the need for reviewed assessment policies and practices.

Implications

Table 12 provides the implications suggested by these studies. The studies highlight the importance of double-checking the results from AI detectors and the training for teachers to prevent plagiarism.

Table 12

Plagiarism: Implications

Title	Implications	Author
Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines”	# “...as GPT Output and Crossplag detectors could identify AI texts with a relatively high degree of accuracy, L2 educators can rely on them to investigate potential cases of AI-assisted plagiarism.” # “To overcome the limitations of each detector, it might be helpful to run a flagged text in one detector and then into another to double-check the results or upload the text to GLTR to get a visual report that can support initial results from an AI-based detector.” # “... since available classifiers can misevaluate human-written texts as machine-generated (i.e., false positives), L2 educators should not base their decisions about academic integrity solely on detectors’ analysis and use other traditional approaches to ensure the academic integrity of students’ work. For instance, educators can interview students about the content of their work or analyze the linguistic complexity of flagged work compared to students’ in-class work.” # “... since fine-tuned classifiers, especially RoBERTa-based classifiers appear to be relatively effective in detecting AI-	Ibrahim (2023)

Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	assisted plagiarism, major plagiarism detection platforms should integrate such classifiers into their systems rather than train their baseline classifiers from scratch.” “...as the findings demonstrate that model size substantially impacts the detection accuracy of classifiers, plagiarism detection platforms should train their detection classifiers on datasets from larger language models to improve their detection accuracy.” # “... to improve practice and to safeguard academic integrity, there is a need for digital training for educators with respect to the use of AI tools in the ESL classroom and the use of advanced detection tools.” # “... lecturers’ tendency to confuse student-written essays for AI-generated text points out the need for lecturers to be trained on the features of AI-created texts and the differences between them and humanwritten essays.” # “There is also a need for revision of assessment policies and procedures and the development of suitable rubrics and assessment criteria.”	Alexander et al. (2023)

Overall, the implications from the first study (Ibrahim, 2023) emphasize that educators should not only depend on AI detectors for their decisions but also use traditional methods like interviewing students or comparing the in-class work of students to submitted texts. Another important implication displays the need for AI detection tools to be trained on large data sets. On the other hand, the second study (Alexander et al., 2023) highlights the importance of training teachers to understand the textual features of AI-generated texts. The final implication suggests revising the policies and procedures, rubrics, and assessment criteria.

AI-Generated Text Authorship and Ownership

Purpose

Out of the 103 research articles, four studies focus on AI-generated text authorship and ownership. The Table 13 shares the purposes of these studies. Three of these studies aim at exploring only authorship, and one examines both authorship and ownership.

Table 13*AI-Generated Text Authorship and Ownership: Purpose*

Title	Purpose	Research Question	Author
Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game	“to investigate 16 education academics' perceptions of two pairs of texts written by either ChatGPT or a human by applying Turing's Imitation Game”	“RQ1: How readily can academics correctly identify texts written by either ChatGPT or humans? RQ2: What are the textual factors that academics perceive as being useful in identifying whether texts are written by ChatGPT or humans? RQ3: How does the application of these factors in decision-making align with accurate or inaccurate identification of text authorship?”	Matthews and Volpe (2023)
Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature	“to compare and contrast an essay written by one second-year English student for a course on English literature with a similar essay produced by ChatGPT / to clarify whether artificial intelligence can satisfy the formal requirements of academic writing and maintain the distinctive voice inherent in human-authored content.”	“1. To what degree can ChatGPT generate text that conforms to the formal requirements inherent in academic essays, including structural elements, citation protocols, and other essential components that define scholarly writing standards? 2. How identifiable and unique is the voice manifested in AI-generated text, particularly in the context of ChatGPT? 3. To what extent does the author contribute to the process of generating text through ChatGPT?”	Amirjalili et al. (2024)
The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors	“to investigate authorship and ownership in human-AI collaboration for personalized language generation”	“RQ1: Does the sense of ownership match the declaration of authorship for personalized AI-generated texts? RQ2: Does the level of influence in human-AI interaction affect the sense of ownership? RQ3: Does the sense of ownership depend on the quality of personalization? RQ4: In what ways does the AI Ghostwriter Effect differ from human ghostwriters?”	Draxler et al. (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“to investigate how successful advanced students are in determining whether an essay was written by AI tools or by a human and how much AI assistance they need to summarize, generate text and write from prompts when trained to use an AI-assisted word processor”	“1. How well do advanced students distinguish human-generated from AI-generated essays? What criteria do they think are decisive about Artificial Intelligence authorship? 2. When trained in the use of an AI-assisted word processor, how do advanced students make use of	Krajka and Olszak (2024)

Title	Purpose	Research Question	Author
		summarising, generating text and writing from prompts? 3. How do students' essays differ linguistically after training in the use of AI tools from those written before treatment? Do they show differences in burstiness and perplexity?"	

The first study (Matthews & Volpe, 2023) investigates how well academics can distinguish between texts written by ChatGPT and humans, and the factors influencing their decisions. The second study (Amirjalili et al., 2024) compares AI-generated and human-written essays in English literature to evaluate ChatGPT's ability to meet academic writing standards and maintain a unique voice. The third study (Draxler et al., 2023) examines the sense of ownership and authorship in human-AI collaborative writing, particularly with a focus on the AI Ghostwriter Effect. The fourth study (Krajka & Olszak, 2024) explores how advanced students identify AI-generated essays and how their writing practices and text characteristics change after being trained to use AI-assisted writing tools.

Results

The results from these studies are shown in Table 14.

Table 14

AI-Generated Text Authorship and Ownership: Results

Title	Results	Author
Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game	"Results indicated academics were only able to identify AI-generated texts half of the time, highlighting the sophistication of contemporary generative AI technology. Academics perceived the following categories as important for their decision-making: voice, word usage, structure, task achievement and flow. All five categories of decision-making were variously used to rationalise both accurate and inaccurate decisions about text authorship."	Matthews and Volpe (2023)
Exploring the boundaries of authorship: a	"While ChatGPT presented an ability to generate contextually relevant content, the results highlighted its need for support in guaranteeing factual accuracy and capturing	Amirjalili et al. (2024)

Title	Results	Author
comparative analysis of AI-generated text and human academic writing in English literature	the complex aspects of authorship that are common in human writing.” “...when compared to human-generated text, the AI-generated text was deficient in terms of specificity, depth, and accurate source referencing.”	
The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors	“...people often do not disclose AI as an author when using personalized AI-generated texts in writing, although they attribute ownership to the AI. This effect was independent of interaction methods and reduced when switching from open-ended author declaration fields to predefined response suggestion.” “The study also showed that the sense of control and leadership were positively correlated with the level of influence over the text generation, which ranged from receiving a non-editable AI text (Getting) to writing the text without AI support.”	Draxler et al. (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“The findings of this study indicate the high linguistic sensitivity of the research participants to factors regarding language and layout, which allows them to distinguish human authors from AI-powered texts. The study results reveal some linguistic differences in students’ reports before and after the treatment. However, the marked variations are rather spontaneous and not obvious. In some AI-enhanced reports, there are even some features of human writing, like high burstiness. As regards specific linguistic differences between students’ written assignments before and after the treatment, the four aspects of sentence length, sentence complexity, word frequency and the ratio of unique words inserted in the texts were used as objective measures when comparing students’ texts. Generally, when analysing the sentence length of the students’ reports after the quasi experimental treatment, most of the study participants formed slightly longer sentences in the post texts (67%, 10 out of 15 students), while only some students shortened their sentences (33%, 5 out of 15 students). It can be assumed that the application of the AI tools in the writing practice of the quasi experimental group of students helped to improve and enrich students’ writing skills in general.”	Krajka and Olszak (2024)

The results of the first study (Matthews & Volpe, 2023) conclude that academics can identify AI-generated texts only half of the time, relying on categories like voice, word usage, structure, task achievement, and flow, leading to some accurate and inaccurate identifications. The finding from the second study (Amirjalili et al., 2024) suggest that ChatGPT-generated content is contextually relevant but lacks specificity, depth, and accurate source referencing compared to human-written essays, which highlights the need for improved factual accuracy and understanding

of human authorship nuances. According to the results from the third study (Draxler et al., 2023), individuals often do not disclose AI's contribution to personalized texts, attributing ownership to the AI, with this effect being consistent across interaction methods and lessened by predefined response suggestions. The last study (Krajka & Olszak, 2024) reveals that advanced students, when trained in AI-assisted writing tools, exhibit high sensitivity to linguistic and layout factors, which enables them to distinguish AI-generated texts from human-authored ones, with remarkable changes in their writing's burstiness and perplexity with the help of post-training.

Implications

Table 15 provides the implications suggested by these studies.

Table 15

AI-Generated Text Authorship and Ownership: Implications

Title	Implications	Author
Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game	# "Experienced academics may be unable to distinguish between texts written by contemporary generative AI technology and humans. # "Academics are uncertain about the current capabilities of generative AI and need support in redesigning assessments that succeed in providing robust evidence of student achievement of learning outcome." # "Institutions must assess the adequacy of their assessment designs, AI use policies, and AI-related procedures to enhance students' capacity for effective and ethical use of generative AI technology."	Matthews and Volpe (2023)
Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature	# "As AI-generated text becomes more prevalent, the study emphasizes the centrality of discussions around authorship, plagiarism, and originality. The findings caution against assuming that AI-generated content, while appearing cohesive, necessarily upholds the standards of nuanced and authentic human writing. This has implications for the evolving discourse on the ethical use of AI in academic contexts."	Amirjalili et al. (2024)

The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors	# “Not only objective control, e.g., as induced by different Interaction Methods, but the subjective experience of control in human-AI interaction is essential to understanding how users relate to AI-generated content in terms of ownership.” # “Human-centered design for AI writing tools should consider (perceived) authorship as a design dimension. Adapting frameworks like CRediT to suit the contributions of algorithms could enrich the debate from a human-centered perspective grounded in common practices.”	Draxler et al. (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“There is little doubt that the fields of language teaching and learning could be significantly enhanced by the ongoing development of Artificial Intelligence capabilities. According to the study’s findings, the students were not very successful at distinguishing between AL and human-generated writing with accuracy rates between 40-80%. They tended to assess human writing based on technical accuracy and error and AI writing based on verbal complexity. Applying AI-powered tools could help students improve their writing competence and bring about changes to the way academic writing is taught as a whole. However, as the study’s findings show, they can also hinder the process of improving writing abilities because they lack imagination, use straightforward and sometimes repetitious vocabulary, respond slowly, or make minor grammatical errors. The research reveals that even highly proficient, almost bilingual, academic students encounter considerable difficulties when attempting to differentiate human writing from AI-assisted written texts. These challenges arise when analyzing formal and informal texts.”	Krajka and Olszak (2024)

The implications from these studies provide several important key insights.

The first study (Matthews & Volpe, 2023) highlights that it is often difficult for experienced academics to distinguish between AI-generated and human-written texts, and that indicates the advanced capabilities of current AI technologies and the need for redesigning assessments to correctly reflect student learning outcomes. In the study of Amirjalili et al. (2024), the comparison between AI-generated and human academic writing reveals that while AI can produce cohesive content, it often lacks the nuance and authenticity of human writing, which could raise important questions about authorship and academic ethical use. Additionally, Draxler et al. (2023) suggest that users frequently do not disclose AI involvement in their writing, suggesting a complex relationship between perceived ownership and declared authorship. The last study (Krajka & Olszak, 2024) proposes that although AI-

powered tools can boost students' writing competence, they also pose challenges by sometimes hindering the development of writing skills due to their limitations in imagination, vocabulary, and grammar.

Transparency and Disclosure

Purpose

There is one study that investigates this issue of transparency and disclosure in writing.

Table 16

Transparency and Disclosure: Purpose

Title	Purpose	Research Question	Author
The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing	“Our study aims to investigate whether nursing academic research journals require authors to declare their use of generative AI in their manuscripts. Examining whether nursing academic research journals require explicit declaration regarding the use of generative AI can offer insights into the level of transparency within the nursing research community.”	Not stated.	Tang et al. (2023)

The purpose of this study is to investigate whether nursing academic research journals ask for an explicit declaration of AI use.

Results

The results from this study are shown in Table 17.

Table 17

Transparency and Disclosure: Results

Title	Results	Author
The importance of transparency: Declaring the use of	# “47 out of 125 (37.6%) journals provided explicit statements about the use of generative AI tools in the writing process. 47 journals (37.6%) explicitly require declaration of the use of	Tang et al. (2023)

Title	Results	Author
generative artificial intelligence (AI) in academic writing	generative AI tools or AI-assisted technologies in the writing process, whereas 46 journals (36.8%) explicitly state that generative AI tools or AI-assisted technologies should not be listed as author or co-author.” # “The mean (SD) impact factors of those journals with and without explicit statement about the use of generative AI tools in the writing process were 2.24 (0.82) and 2.15 (1.07), respectively; the mean difference (0.09, 95% CI from -0.27 to 0.45) was not statistically significant ($p = 0.611$).” # “There is no statistically significant association ($p = 0.520$) between the themes of the journal.”	

The results from this study (Tang et al., 2023) indicate that 37.6% of journals explicitly require the declaration of generative AI usage in the writing process, while another 36.8% state that AI tools should not be listed as authors or co-authors. The mean impact factors of journals with and without such explicit statements show no statistically significant difference, and there is no significant association between journal themes and the presence of AI usage declarations.

Implications

Table 18 provides the implications. The implications mainly focus on the transparent declaration of generative AI usage.

Table 18

Transparency and Disclosure: Implications

Title	Implications	Author
The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing	# “We are advocating for active participation from nursing researchers in these initiatives to ensure that our perspectives are included.” # “Just as the declaration of IRB approval became a cornerstone of ethical research practice, the transparent acknowledgment of generative AI usage could similarly become a vital aspect of responsible academic conduct.” # “Another important aspect to consider is the declaration of the use of generative AI tools by journal reviewers. While authors are required to declare their utilization of generative AI tools in their submitted manuscripts, it is equally relevant to impose the same requirement on the reviewers.” # “Reviewers contribute to quality assessment, helping to differentiate reputable journals from predatory ones. This commitment to transparency reinforces the credibility of the review process and prevents inadvertent support for predatory journals. Additionally, the disclosure invites enhanced scrutiny and accountability, as predatory journals often lack robust peer review processes.”	Tang et al. (2023)

“...the disclosure invites enhanced scrutiny and accountability, as predatory journals often lack robust peer review processes. Furthermore, explicit declaration helps expose deceptive claims of AI usage by predatory journals, safeguarding against manipulation and promoting transparency in publishing decisions for researchers.”

The main implication from this study (Tang et al., 2023) is that transparent declaration of generative AI usage by both authors and reviewers in academic writing is essential for maintaining ethical research practices, enhancing scrutiny and accountability, and protecting against the deceptive practices of some predatory journals.

Ethical Use

Purpose

Two studies focus on the ethical use of ChatGPT by students for their writing assignments.

Table 19

Ethical Use: Purpose

Title	Purpose	Research Question	Author
Writing with AI: University Students' Use of ChatGPT	“to investigate the prevalence of use of ChatGPT among university students for written assignments, explore the ways students utilize the tool, and examine students' perspectives on the ethical aspects of its use”	“RQ1. How prevalent is the use of ChatGPT among university students? RQ2. How do university students use ChatGPT? RQ3. What are the perceptions of university students regarding the ethical aspects of using ChatGPT?”	Črček and Patekar (2023)
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	“to investigate and assess the impact of integrating ChatGPT, a generative artificial intelligence tool, into an inorganic chemistry classroom for an 8-page writing assignment on selected elements from the periodic table / to explore student experiences, perceptions, and the utility of ChatGPT across different stages of the writing process, considering its role in idea	“How can ChatGPT be effectively utilized as a writing tool? What are the potential benefits associated with its use? How does it impact the learning outcomes and overall writing proficiency of students? How do students use it if given a free range? What do students think of using generative AI for writing assignments? Can we learn about student opinions on the ethical use of ChatGPT? Does it produce verifiable information?”	Rojas (2024)

Title	Purpose	Research Question	Author
	generation, topic verification, proofreading, and editing while also investigating student opinions on the ethics of using ChatGPT in academic writing”		

Both studies focus on examining how university students use ChatGPT and their ethical considerations, but the first study (Črček & Patekar, 2023) has a broader focus on general written assignments, while the second (Rojas, 2024) specifically addresses a scientific writing assignment in an inorganic chemistry context.

Results

The results of the studies that aim at investigating the ethical use of AI are shared in Table 20.

Table 20

Ethical Use: Results

Title	Results	Author
Writing with AI: University Students' Use of ChatGPT	# “The results show that more than half of the participants use ChatGPT for written assignments, that most use it to generate ideas, while many use it to summarize, paraphrase, proofread, but also to write a part of the assignment for them.” # “According to the participants, the most ethically acceptable use of ChatGPT is for generating ideas, while other uses are perceived by many as being unethical; this, however, has not prevented some students from engaging in behaviors they deem unethical.”	Črček and Patekar (2023)
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	# “This study's results reveal positive student experiences and perceptions of integrating ChatGPT for a writing-intensive assignment, showcasing its effectiveness in idea generation, topic verification, proofreading, and editing.” # “The integration of ChatGPT into the inorganic chemistry classroom has yielded valuable insights into its multifaceted utility for writing assignments. A preassignment survey revealed a diverse range of prior experiences among students, with a majority expressing prior curiosity-driven exploration of ChatGPT. For the effectiveness of ChatGPT in the first assignment, focused on topic generation and verification, student responses emphasized the ease and efficiency brought to the writing process, highlighting ChatGPT's role as a facilitator for idea development.” # “Subsequent surveys further illustrated students' positive perceptions, with the second survey indicating widespread	Rojas (2024)

Title	Results	Author
	agreement on the usefulness of ChatGPT for scientific writing. Ethical considerations emerged, as students unanimously expressed reservations about using ChatGPT to generate entire text passages, emphasizing the importance of maintaining active student input.” # “The final survey underscored ChatGPT’s pivotal roles in generating ideas and facilitating proofreading and editing, showcasing its significance as a supplementary tool. Looking ahead, students exhibited varied inclinations toward future ChatGPT use, reflecting a nuanced approach shaped by their experiences in this chemistry class.”	

Both studies find that students widely use ChatGPT at various stages of the writing process, including idea generation, drafting, proofreading, and editing. Both studies also highlight mixed perspectives on the ethical implications of using ChatGPT, with concerns about academic integrity and originality. While the first study (Črček & Patekar, 2023) focuses on general written assignments, the second study (Rojas, 2024) specifically examines the use of ChatGPT in a scientific writing assignment, finding very similar positive usefulness and ethical concerns.

Implications

Table 21 provides the implications suggested by these studies.

Table 21

Ethical Use: Implications

Title	Implications	Author
Writing with AI: University Students' Use of ChatGPT	# “We firmly believe that universities and instructors must take a proactive stance, not through prohibition, but by embracing ChatGPT and other AI-powered tools while providing clear guidelines to students. With proper guidance, these tools can be utilized in ethical and creative ways, ultimately – and hopefully – leading to a reduction in academic dishonesty offenses. The responsibility lies with educational institutions to lead the way in shaping the ethical use and application of ChatGPT, fostering a culture of integrity and responsible technology use among students.” # “By acknowledging the potential pitfalls and challenges associated with ChatGPT, and by actively addressing them through comprehensive guidelines and ongoing support, universities can harness the power of AI to enhance education while maintaining academic integrity. This approach will pave the way for a more ethical and productive integration of AI	Črček and Patekar (2023)

Title	Implications	Author
	technologies in the learning environment, benefiting both students and instructors alike.”	
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	# “For amateurs or those who lack extensive subject knowledge, ChatGPT can quickly summarize a vast array of content and condense the material quickly for students. Importantly, the content generated by ChatGPT proved to be reliable, at least in these 53 student examples. Overall, student writing was improved through this process largely because students were provided a variety of mechanisms to revise their work. This process is achievable with writing centers, peer review, instructor feedback, etc., and ChatGPT appears to reliably contribute to that array of resources for editing writing. As a supplementary tool, ChatGPT can provide additional support and resources to facilitate writing assignments.” # “However, it will be crucial for educators to address ethical considerations associated with the use of generative AI tools such as ChatGPT. By emphasizing the importance of active student involvement in the writing process and discouraging the use of ChatGPT to generate entire text passages, educators can promote academic integrity to uphold standards of originality.” # “In STEM disciplines such as chemistry, where writing assignments play an important role in developing critical thinking and communication skills, integrating ChatGPT can offer students novel opportunities to engage with course material and express their understanding in innovative ways.” # “Importantly, the use of generative AI in the writing process poses the risk of bypassing critical thinking and inhibiting some learning opportunities. By relying solely on AI-generated content, students may miss the chance to develop their own creative writing skills. Additionally, without facing the challenge of making mistakes and learning from them, students may struggle to improve their writing abilities independently.” # “Educators can tailor ChatGPT-based activities to align with specific learning objectives and foster deeper learning experiences. Incorporating ChatGPT into writing assignments exposes students to emerging technologies and enhances their technological literacy.” # “Educators can use ChatGPT as a catalyst for discussions on the ethical implications, limitations, and potential applications of AI in academic and professional contexts, empowering students to critically evaluate and responsibly use AI tools in their future endeavors.”	Rojas (2024)

The implications of the first study (Črček & Patekar, 2023) suggest that addressing barriers such as limited access, lack of knowledge, and concerns about academic integrity is essential for educators to fully use ADWTs in for their educational goals. The implications of the second study (Rojas, 2024) highlight the

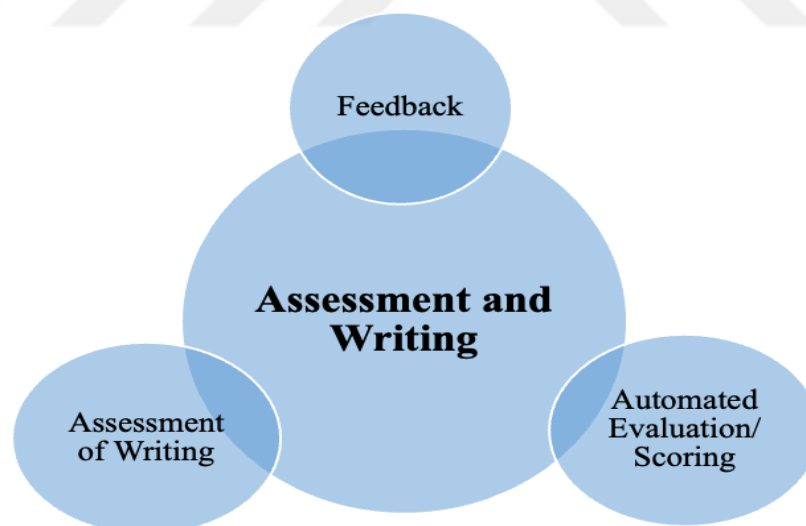
importance of transparency in using GenAI tools, with both students and teachers agreeing on the acceptability of these tools for brainstorming and outlining but not for completing tasks if they do not disclose the use of these tools. The implications of these studies suggest that there is a shared emphasis on the need for clear policies and education to ensure the ethical and effective use of AI in education. Both studies stress that overcoming access and knowledge barriers and promoting transparent use are important steps to integrate AI tools responsibly and maintain academic integrity.

Assessment and Writing

The assessment was another broader category that was identified after the analysis. The subcategories of this category include *feedback*, *assessment of writing*, and *automated evaluation or scoring* as seen on Figure 6.

Figure 6

Assessment and Writing



Feedback

Purpose

12 of 103 articles are related to feedback. Some of the studies compare AI feedback with human-generated feedback, while others look at the interaction between AI feedback and students.

Table 22*Feedback: Purpose*

Title	Purpose	Research Question	Author
Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study	“to present a study on the impact of ChatGPT as a formative feedback tool on the writing skills of undergraduate ESL students”	“1. In an intensive academic writing course, when the instructional hours and tasks are held constant, does the employment of ChatGPT as a feedback tool have any significant impact on undergraduate ESL students' writing skills? 2. How do the experimental group students perceive the impact of ChatGPT as a feedback tool on their writing skills?”	Mahapatra (2024)
AI-generated feedback on writing: insights into efficacy and ENL student preference	“to report on two longitudinal studies / to examine the efficacy of ChatGPT and GPT-4 as an AWE tool in terms of language improvement and student perceptions in the ENL population.”	“1. Does the application of AI-generated feedback result in superior linguistic progress among ENL students compared to those who receive feedback from a human tutor? 2. Does the preference for AI-generated feedback surpass that for human tutor-generated feedback among ENL students?”	Escalante et al. (2023)
Feedback sources in essay writing: peer-generated or AI-generated feedback?	“to use ChatGPT as a source of feedback for students' argumentative essay writing tasks and to compare the quality of ChatGPT-generated feedback with peer feedback”	“RQ1. To what extent does the quality of peer-generated and ChatGPT-generated feedback differ in the context of essay writing? RQ2. Does a relationship exist between the quality of essay writing performance and the quality of feedback generated by peers and ChatGPT?”	Banihashem et al. (2024)
Can AI-assisted essay assessment support teachers? A cross-sectional mixed-methods research conducted at the university of Montenegro	“The main aim of this paper is to investigate if this AI language model can be a viable teachers' assistance tool that provides immediate and meaningful feedback to teachers and students.”	Not given.	Ivanović (2023)
AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts	“to develop a Smart RoamLingo app to help writing meaningful content in authentic contexts with sample sentences based on texts generated by recognition technologies named	“1. What are the differences in the learning achievement among the three groups, the experimental group (EG) with two recognition technologies (i.e., ITR and STR) and AI support (i.e., AI-SS and AI-WF), the control group one (CG1) with two recognition technologies (i.e., ITR and STR), and the control group two (CG2) with one recognition technology only (i.e., STR)?	Hwang et al. (2023)

Title	Purpose	Research Question	Author
	as AI-Sample Sentences (AI-SS) and writing qualities with revision suggestions based on personal original writings named as AI-Writing Feedback (AI-WF)”	2. What are the correlations of the learning behaviors in the practice and assignment activities with the learning achievement of EG and the influence of learning behaviors in both practice and assignment activities on learning achievement? 3. What are the affordances of AI and its limitations in EFL writing? 4. What are the students’ perceptions towards the proposed Smart RoamLingo?”	
Defining the Role of Artificial Intelligence in Improving English Writing Skills Among Indonesian Students	“to explore the improvement of English writing skills through the integration of AI into reading practices and provision of adequate feedback for skills enhancement	Not given.	Kaharudin et al. (2024)
Learner Interaction with, and Response to, AI-Programmed Automated Writing Evaluation Feedback in EFL Writing: An Exploratory Study	“to examine in detail how EFL students interacted with the feedback of Pigai, the largest AI-programmed AWE in China by using a small data set arising from an analysis of five Chinese university-level English as a foreign language (EFL) students’ submissions”	“1. What are the salient features of iterative feedback by Pigai over a period of multiple resubmissions? 2. How do learners interact with Pigai feedback through the process of multiple revisions?”	Yang et al. (2024)
Using artificial intelligence to foster students’ writing feedback literacy, engagement, and outcome: A case of wordtune application	“to describe how AI is now used to provide feedback in the writing classroom and to suggest potential locations for further research on technology-driven feedback practices”	“(1) Does AI-enhanced classroom have any significant effects on L2 learners’ writing outcome? (2) Does AI-enhanced classroom have any significant effects on L2 learners’ writing feedback literacy and engagement? (3) What perceptions did L2 learners have towards employing AI-enhanced classroom in writing instruction?”	Rad et al. (2023)
An Empirical Study on the Artificial Intelligence Writing Evaluation System in China CET	“taking Juku as an example, to find whether it is really effective to help both the English teachers and the students in English teaching	“RQ1: Whether AWE is an effective platform to help students with their English writing? RQ2: What are the attitudes that teachers and students hold to AWE? RQ3: Whether there is a great difference between the traditional feedback and the one given by the computer?”	Lu (2019)

Title	Purpose	Research Question	Author
	and writing, their use of automated writing evaluation (AWE), their attitude toward AWE, and so on”		
Comparing the quality of human and ChatGPT feedback of students' writing	“to examine the ability of generative AI (ChatGPT) to provide formative feedback”	“(RQ1): How does the quality of formative feedback provided by ChatGPT compare to the quality of formative feedback provided by human evaluators? (RQ2): Does the quality of formative feedback provided by ChatGPT or human evaluators differ for essays judged as low, medium, and high quality? (RQ3): Does the quality of formative feedback provided by ChatGPT or human evaluators differ for English-speaking and EL students?”	Steiss et al. (2024)
Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?	“To address these [uncertainties regarding how ChatGPT complements teacher assessment and the overall effectiveness of this combined approach], we conducted a mixed-methods investigation involving 46 undergraduate students from a research university in southern China, engaging them in a Chinese academic writing task”	“RQ1: What are the features of scores and feedback provided by teachers and ChatGPT in the assessment of students' writing? RQ2: How do students perceive teacher and ChatGPT feedback? RQ3: How do students implement teacher and ChatGPT feedback to facilitate their revisions?”	Lu et al. (2024)
To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing	“to examine the possible role of ChatGPT in facilitating the teaching and learning of writing English as a Foreign Language (EFL)”	“RQ1: What are the differences between ChatGPT- and teacher-generated feedback on EFL students' writing? RQ2: How do EFL teachers perceive ChatGPT-generated feedback on EFL students' writing?”	Guo and Wang (2023)

Overall, the main purpose of these articles is to look at the impact of AI tools on feedback. Specifically, they want to explore the impact and efficacy of AI, specifically ChatGPT and other AI writing tools, on writing instruction in various

contexts. They investigate how AI-generated feedback affects writing skills and outcomes among ESL, ENL, and EFL students, comparing AI feedback with human and peer feedback. The studies also examine student and teacher perceptions of AI tools, their preferences for AI versus human feedback, and the potential of AI to complement traditional teaching methods. Additionally, some studies focus on the ethical considerations, cognitive-affective dimensions, and practical applications of AI in supporting personalized and contextualized writing practices.

Results

The results of the studies that are related to feedback are in Table 23

Table 23

Feedback: Results

Title	Results	Author
Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study	# "The findings of the study add to the literature on AWE, especially the use of generative AI. They strengthen and further theories of feedback as a dialogic tool and ChatGPT as a formative feedback tool that can be harmoniously integrated into large writing classes. The findings of the study demonstrate that interactions facilitated by ChatGPT during the process of writing has a direct positive impact on student learning. It positively influences how students seek feedback, how they engage with it and how they make improvements in their academic writing. It enables students to overcome the anxiety involved in asking for and receiving the desired kind of feedback."	Mahapatra (2024)
AI-generated feedback on writing: insights into efficacy and ENL student preference	# "Study 1 compared human tutor and AI-generated feedback to see if one would influence linguistic gains more than the other. The results indicate that AI-generated feedback did not result in superior linguistic progress among ENL students compared to those who received feedback from a human tutor. The between-subject variable of group did not have a significant effect on writing scores, suggesting that one method of feedback was not better than another in terms of scores." # "Study 2 investigated which form of feedback ENL students preferred and why. We found about half the students preferred receiving feedback from a human tutor, and half preferred AI-generated feedback. Those that preferred sitting down and discussing their feedback with a tutor cited the face-to-face interaction as having affective benefits, such as increasing engagement, as well as benefits for developing their speaking abilities. Those that	Escalante et al. (2023)

Title	Results	Author
	preferred AI-generated feedback primarily cited the clarity and specificity of the feedback as being useful for improving their writing. This echoes the findings of Dai et al. (2023), namely that AI-generated feedback was found to be more readable and detailed than feedback from an instructor.”	
Feedback sources in essay writing: peer-generated or AI-generated feedback?	# “The results showed a significant difference between feedback generated by ChatGPT and peers. While ChatGPT provided more descriptive feedback including information about how the essay is written, peers provided feedback including information about identification of the problem in the essay. The overarching look at the results suggests a potential complementary role for ChatGPT and students in the feedback process. For the first research question, the results revealed a disparity in feedback quality when comparing peer-generated feedback to feedback generated by ChatGPT. Peer feedback demonstrated higher quality compared to ChatGPT-generated feedback. ChatGPT tended to provide more descriptive feedback, often including elements such as summarizing the content of the essay. This inclination towards descriptive feedback could be related to ChatGPT’s capacity to analyze and synthesize textual information effectively. In contrast, students’ feedback content entailed high quality regarding identifying specific issues and areas for improvement. Peers outperformance compared to ChatGPT in identifying problems within the essays could be related to humans’ potential in cognitive skills, critical thinking abilities, and contextual understanding (e.g., Korteling et al., 2021; Lamb et al., 2019). This means that students, with their contextual knowledge and critical thinking skills, may be better equipped to identify issues within the essays that ChatGPT may overlook.” # “Regarding the relationship between the quality of essays and the quality of the feedback provided by ChatGPT and peers, we found no overall significant relationship.”	Banihashem et al. (2024)
Can AI-assisted essay assessment support teachers? A cross-sectional mixed-methods research conducted at the university of Montenegro	“The average time for ChatGPT to render an essay assessment was consistently below 30 seconds, while the teachers reported they needed 5 to 15 minutes to thoroughly read an essay and an additional 5 to 15 minutes to grade it. Thus, an essay assessment can take anywhere between 10 and 30 minutes per essay. ChatGPT can analyse the essays based on predefined criteria. Thus, professors’ feedback can be tailored to their needs and target specific essay deficiencies which need to be corrected. What has positively surprised us, was the level of detail that GPT was able to extract from the essays and how granular its analysis was. Each analysis was very detailed with a sound and logical foundation. Even though 78 essays have many things in common, we felt ChatGPT paid individual attention to each essay, rather than resorting to some generic statements with little value in terms of deeper insight. Our research has shown that ChatGPT is a reliable essay assessment tool, especially when it comes to analysing large amounts of data in a short	Ivanović (2023)

Title	Results	Author
	amount of time. ChatGPT can assess essays at a much faster rate than humans can. It can evaluate essays in a matter of seconds, while it may take humans hours or even days to read and evaluate many essays. While assessing those essays, ChatGPT assesses them objectively, based on the patterns it has learned from the training data it was trained on. When ChatGPT assesses an essay, just like humans, it analyses various aspects of the text, including the coherence of ideas, the strength of arguments, the use of evidence, the organisation and structure of the text, and the quality of language and grammar.”	
AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts	“EG significantly outperformed the other groups in the post-test. In EG, AI-SS can help the contextualization to produce comprehensive content actively while AI-WF can help the personalization to improve the writing quality through several revisions. Hence, the cohesion and consistency of their writing can be enhanced. Further, the total number of revisions in practices and the assignment score can significantly predict the post-test. Moreover, EG students revealed that Smart RoamLingo was useful for EFL writing. Therefore, it is strongly suggested to use Smart RoamLingo with AI-SS and AI-WF for EFL writing in authentic contexts.”	Hwang et al. (2023)
Defining the Role of Artificial Intelligence in Improving English Writing Skills Among Indonesian Students	# “The results showed that (1) Students' attitude toward the use of AI played a mediating role in explaining the effect of reading and feedback on writing skills, (2) Reading had an indirect effect through attitude on writing skills improvement, and (3) Lecturers feedback had an indirect effect through attitude on writing skills.” # “This study proved that virtual teaching of reading and feedback with a positive attitude had a significant effect on writing skills improvement. Improving writing skills requires good mastery of grammar and vocabulary, which can be developed by reading and feedback from lecturers.”	Kaharuddin et al. (2024)
Learner Interaction with, and Response to, AI-Programmed Automated Writing Evaluation Feedback in EFL Writing: An Exploratory Study	“Our detailed analysis showed that Pigai provided various types of feedback. The majority of error corrective feedback focused on local, language-related errors. At the same time, Pigai feedback provided a significant amount of non-error feedback items, increasing through multiple submissions but lacking examples and contextual information. The detailed analysis of all five students' revisions and resubmissions showed certain patterns via their responses to different types of feedback, initially with error-corrective feedback, then with non-error corrective feedback with trial and errors, and general feedback. It could be concluded that sustained engagement with Pigai feedback could facilitate writing improvement and develop students' autonomy in using various writing strategies.”	Yang et al. (2024)
Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application	“By using Wordtune, an AI-based application, the experimental group was able to significantly improve their writing outcomes, engagement, and feedback literacy when compared to the control group. Furthermore, the Wordtune application received positive feedback from students interviewed about its ability to improve their writing outcomes, engagement, and feedback literacy.	Rad et al. (2023)

Title	Results	Author
	These findings will contribute to our understanding of AI in L2 learning and inform instructional considerations for Wordtune's application.”	
An Empirical Study on the Artificial Intelligence Writing Evaluation System in China CET	“The author finds that: (1) Using AWE does effectively help the students with their English writing; (2) Both teachers and students have positive attitude to the use of AWE in terms of immediate and clear feedback, time-saving, and arousing interests in English writing; and (3) AWE still needs to be perfected as it cannot provide proper evaluation on the text structure, content logic, and coherence. So both teachers and students should take the score from AWE objectively.”	Lu (2019)
Comparing the quality of human and ChatGPT feedback of students' writing	“Human raters were better at providing high-quality feedback to students in all categories other than criteria-based. AI and humans showed differences in feedback quality based on essay quality. Feedback did not vary by language status for humans or AI. Well-trained evaluators provided higher quality feedback than ChatGPT. Considering the ease of generating feedback through ChatGPT and its overall quality, generative AI may still be useful in some contexts, particularly in formative early drafts or instances where a well-trained educator is unavailable. • Human feedback was better than ChatGPT feedback for 4/5 elements of formative feedback. • Differences between ChatGPT and humans were modest when considering the overall quality of feedback and time-savings • ChatGPT and human feedback varied in quality based on the score of the targeted essay. • ChatGPT has potential as an evaluative tool given tradeoffs between quality and time.”	Steiss et al. (2024)
Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?	“The intraclass correlation coefficient results revealed ChatGPT's efficiency in scoring students' writing, showing moderate to good consistency with teacher evaluations. A paired sample t-test unveiled significant differences in feedback quantity and types between ChatGPT and teacher assessments. Drawing from both interview data and quantitative findings, the study uncovers three ways in which ChatGPT complements teacher assessment, benefiting students with various writing proficiency levels: (1) fostering deeper comprehension of teacher assessments among students, (2) encouraging students to make judgments regarding feedback, and (3) promoting independent thinking about revisions.”	Lu et al. (2024)
To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing	“The results showed that ChatGPT produced a significantly larger amount of feedback than teachers and that compared with teacher feedback, which mainly focused on content-related and language-related issues, ChatGPT distributed its attention relatively equally among the three feedback foci (i.e., content, organisation, and language). Our results also indicated that ChatGPT and teachers displayed tendencies towards using different feedback types when evaluating different aspects of students' writing. Additionally, we examined EFL teachers' perceptions of using ChatGPT-generated feedback to support their own feedback. The five teachers reported	Guo and Wang (2023)

Title	Results	Author
	both positive and negative perceptions of the features of ChatGPT feedback and the relation between ChatGPT and teacher feedback. To foster EFL students' writing skills, we suggest that teachers collaborate with ChatGPT in generating feedback on student writing.”	

The findings indicate that AI tools can have a positive influence on students' writing abilities and engagement with feedback. Mahapatra (2024) found that ChatGPT positively influences how students seek and engage with feedback, leading to improvements in academic writing. Similarly, Rad et al. (2023) reported that AI-assisted tools like Wordtune can significantly improve students' writing outcomes, engagement, and feedback literacy. Lu (2019) noted that Automated Writing Evaluation (AWE) systems effectively help students with English writing, though they have limitations in evaluating text structure and coherence.

When comparing AI-generated feedback to human feedback, the results are mixed. Escalante et al. (2023) found that AI-generated feedback did not result in superior linguistic progress compared to human tutor feedback for ENL students. Steiss et al. (2024) reported that human raters provided higher-quality feedback than ChatGPT in most categories, except for criteria-based feedback. However, Lu et al. (2024) observed that ChatGPT showed moderate to good consistency with teacher evaluations in scoring students' writing. The research suggests that AI can play a complementary role to human feedback. Banihashem et al. (2024) found that ChatGPT and peer feedback may have complementary strengths, with ChatGPT providing more descriptive feedback and peers better at identifying specific issues. Lu et al. (2024) noted that ChatGPT can complement teacher assessment by fostering deeper comprehension, encouraging judgment, and promoting independent thinking

about revisions. Hwang et al. (2023) highlighted that AI tools can support personalization and contextualization in authentic writing contexts.

One of the most significant advantages of AI-based feedback is its efficiency. Ivanović (2023) reported that ChatGPT can assess essays much faster than humans, typically in under 30 seconds compared to 10-30 minutes for teachers. Steiss et al. (2024) emphasized the time-saving aspect of AI feedback, even when considering the modest differences in overall quality between AI and human feedback.

Perceptions of AI-based writing tools are generally positive among both students and teachers. Lu (2019) and Rad et al. (2023) found that students and teachers have positive attitudes toward AI-based writing evaluation tools. However, Guo and Wang (2023) noted that EFL teachers reported both positive and negative perceptions of using ChatGPT-generated feedback to support their own feedback.

Based on these findings, researchers have made several recommendations. Hwang et al. (2023) suggested using AI tools like Smart RoamLingo with AI-based support systems for EFL writing in authentic contexts. Guo and Wang (2023) recommended that teachers collaborate with ChatGPT in generating feedback on student writing to foster EFL students' writing skills. Steiss et al. (2024) proposed that AI feedback may be particularly useful for formative early drafts or when well-trained educators are unavailable.

In conclusion, these findings highlight the potential of AI tools in supporting ESL writing instruction while also acknowledging their current limitations and the continued importance of human involvement in the feedback process. As AI technology continues to evolve, it will likely play an increasingly significant role in ESL writing instruction, complementing human teaching and potentially improving learning outcomes for students.

Implications

Table 24 provides the implications suggested by these studies on feedback.

Table 24

Feedback: Implications

Title	Implications	Author
Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study	# "The study strengthens and advances theories of feedback as a dialogic tool and ChatGPT as a reliable writing tool, and has practical implications. With proper student training, ChatGPT can be a good feedback tool in large-size writing classes. Through the empirical evidence on the utility of ChatGPT as a formative feedback tool in the academic writing classroom, this study establishes that the integration of ChatGPT into the academic writing instruction can yield positive impacts when the instructor is aware of the affordances of ChatGPT and knows how to guide students about its utilities. It also proves that the dialogic nature of ChatGPT can be fully put to use when students are adequately prepared."	Mahapatra (2024)
AI-generated feedback on writing: insights into efficacy and ENL student preference	# "While the study found that AI-generated feedback did not lead to superior linguistic progress among ENL students compared to human tutor feedback, it is important to consider the potential time-saving benefits offered by AI-generated feedback for educators. Utilizing AI for providing feedback can potentially significantly reduce the time teachers spend on reviewing and responding to each student's assignment, thereby freeing up valuable time for other tasks." # "Furthermore, the time efficiency of AI-generated feedback can be particularly advantageous in large classes where providing individualized feedback by the instructor is logistically challenging and time-consuming. In light of the major findings highlighted above, a mixed approach to providing feedback may be most beneficial for both language educators and students. By utilizing GenAI, language educators may be able to produce more detailed feedback in a shorter amount of time for each individual learner. Providing opportunities for students to discuss AI-generated feedback with a human tutor and ask follow-up questions affords students the benefits of each modality: the clarity and specificity of the AI-generated feedback, and the engagement and the ability to practice speaking that comes with interacting with another human."	Escalante et al. (2023)
Feedback sources in essay writing: peer-generated or AI-generated feedback?	# "The implications of this study are valuable, shedding light on the prospective use of ChatGPT as a feedback source, particularly for complex tasks like argumentative essay writing. From a practical perspective of higher education, the study's findings offer insights into the potential integration of ChatGPT as a feedback source within higher education contexts. The discovery that ChatGPT's feedback quality could potentially complement peer feedback highlights its applicability for enhancing feedback practices in higher	Banihashem et al. (2024)

Title	Implications	Author
Can AI-assisted essay assessment support teachers? A cross-sectional mixed-methods research conducted at the university of Montenegro	education. This holds particular promise for courses with substantial enrolments and essay-writing components, providing teachers with a feasible alternative for delivering constructive feedback to a larger number of students.” # “From an educational standpoint, our research findings advocate for the integration of ChatGPT as a feedback resource with peer feedback within higher education environments for essay writing tasks since there is a complementary role potential for pee-generated and ChatGPT-generated feedback. This approach holds the potential to alleviate the workload burden on teachers, particularly in the context of online courses with a significant number of students.” # “To reach its full potential and become even more useful for teachers, ChatGPT needs to be integrated into teachers’ work. Integrating AI tools into teachers’ work in a supportive manner involves considering the specific needs and context of teachers and their students. AI tools can provide comprehensive training and professional development programmes for teachers to familiarise them with AI tools, their capabilities, and their limitations. Teachers should receive guidance on how to effectively use AI tools to enhance their teaching practices, streamline administrative tasks, and support student learning.” # “Furthermore, we have to define clear pedagogical objectives for integrating AI tools. We must identify specific areas where AI can provide value, such as personalised learning, formative assessment, or administrative tasks.” # “Moreover, we need to align the use of AI tools with the goals of the curriculum and teaching methodologies to ensure they enhance the teaching and learning experience. After this, we should involve teachers in the development and selection of AI tools. Engage them in collaborative discussions, seek their feedback, and encourage the co-creation of AI-based solutions that address their unique needs. This collaboration ensures that the tools are relevant, practical, and tailored to the specific teaching context. AI tools have to be introduced gradually and incrementally, allowing teachers to gain familiarity and confidence. We could start with pilot projects or small-scale implementations, gather feedback, and refine the integration based on teacher and student experiences. Gradual implementation minimizes disruption and allows for continuous improvement.” # “Institutions must ensure that the necessary infrastructure, technical support, and resources are in place to support the integration of AI tools. By considering these factors and providing ongoing support, training, and collaboration opportunities, teachers can effectively integrate AI tools into their work in a manner that enhances their teaching practices and benefits student learning.”	Ivanović (2023)
AI and Recognition Technologies to Facilitate English as	# “Based on the findings of this study, we summarize that recognition technologies such as STR and ITR can help generate vocabulary and phrases related to authentic	Hwang et al. (2023)

Title	Implications	Author
Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts	contexts surrounding learners and openAI can help make meaningful sentences to guide students how to describe their taken pictures.” # “Furthermore, openAI can provide students with more meaningful writing feedback based on their original writing to improve the writing quality including writing fluency with suggested conjunction words and giving more related description sentences. The integration of these recognition technologies and AI can work very well to support EFL writing.”	
Defining the Role of Artificial Intelligence in Improving English Writing Skills Among Indonesian Students	# “This study proved that virtual teaching of reading and feedback with a positive attitude had a significant effect on writing skills improvement. Improving writing skills requires good mastery of grammar and vocabulary, which can be developed by reading and feedback from lecturers.” # “Furthermore, electronic devices can facilitate this process with the application of AI. In conclusion, students' skills in learning English using AI needed to be improved. This was because students' attitude was shown to have a mediating role in explaining the effect of reading and teacher feedback on writing skills development. In this case, feedback had no direct influence on the development of writing skills. Therefore, building a positive attitude towards the use of AI in learning needs to be prioritized. Vocabulary development from reading and remedial exercises based on lecturers' feedback improved students' writing skills.”	Kaharuddin et al. (2024)
Learner Interaction with, and Response to, AI-Programmed Automated Writing Evaluation Feedback in EFL Writing: An Exploratory Study	# “There are also implications for the future of AWE to build more personalized learning models based on learners' engagement pattern with Pigai feedback. As learners' responses to feedback impact the effectiveness of their learning, it is important to embed the capacity for meaningful engagement in the development of learning technology (Zhang & Zhang, 2018). The content of the feedback could highlight the focus and relevance. Recent research in deep learning and AI-supported learning could be incorporated to enhance AWE systems in providing more meaningful and personalised feedback (Shi & Aryadoust, 2022). Research also showed that not all errors were picked in Pigai feedback, which implied that current Pigai feedback is mainly based on data retrieval and matching, therefore tailoring the content and providing more appropriate feedback for optimal learning effectiveness may increase the reliability of future AWE design.” # “In addition, future AI-programmed apps may be designed to give feedback for conceptual ideas and more genre types.”	Yang et al. (2024)
Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application	“This study elucidates ideas that might improve second-language learners' writing feedback literacy, writing engagement, and writing outcomes and contextualizes them within the perspective of AI. The suggested concepts might benefit the field on both a broad and a focused level. This study might, on the surface, be used as a guide to enhance feedback efficacy and its perceived value for the efficient use of AI to enhance writing feedback literacy,	Rad et al. (2023)

Title	Implications	Author
	thereby reducing the tension between instructors' feedback practices and students' expectations through personalized feedback. Instructors should adopt dialogic teaching, where student-teacher interaction is less emphasized to establish tolerance and acceptance with students as active members of the process instead of boosting themselves to a higher position with feedback as a unilateral process, to best promote and sustain student writing feedback literacy. In sum, Wordtune is well regarded for supporting students in their understanding of such information during the writing process and benefits from new data displayed in AI."	
An Empirical Study on the Artificial Intelligence Writing Evaluation System in China CET	# "Of course, there are still many aspects to be improved in the intelligent correction system of English composition. It is just a tool to assist teaching, and teachers' guidance should be combined with online learning." # "In addition, AWE system can only comment on the grammar errors and basic word collocations. It cannot meet the requirements of the evaluation for the composition of the text structure, content logic, and coherence. So the writing scores should be taken objectively. The intelligent correction system of composition can not only improve teachers' working efficiency, let teachers have more time to improve classroom teaching and research work, but also more importantly create opportunities for students to write as much as possible and improve their English writing ability." # "Finally, as the environmental conditions of using Internet and computer are changing, and the teachers scoring composition and students writing composition change with time, college English teachers should try to find good ways for marking essays and give effective feedback for students so that teachers can be liberated from heavy work of marking students' essays while the students will get effective feedback and move forward quickly with their good English composition."	Lu (2019)
Comparing the quality of human and ChatGPT feedback of students' writing	"Formative feedback is also a way for teachers to gain important knowledge about student writing development and to design evidence-based instruction (Graham, Harris, & Chambers, 2016). Even if the present study shows feedback provided by AI as relatively high quality given the time it requires, we cannot measure the potentially negative effect of not engaging with early student drafts on teachers' knowledge of their students and subsequent evidence-based instructional planning. Using generative AI, instead of teachers, to frequently provide feedback would be especially problematic if teachers do not oversee the process, looking over students' shoulders at times to see how they are engaging with the AI feedback, asking students' perceptions of the feedback, and considering the local context for best use cases and limitations. The pedagogical implications for both students and teaching must be considered carefully."	Steiss et al. (2024)
Can ChatGPT effectively	"It is recommended that teachers consider incorporating ChatGPT as a supplementary tool for writing assessment.	Lu et al. (2024)

Title	Implications	Author
complement teacher assessment of undergraduate students' academic writing?	They can utilize ChatGPT's scoring and feedback as a reference for their own assessments, thereby reducing the workload and leveraging the strengths of both approaches. Teachers should provide students with specific guidance on how to effectively utilize ChatGPT and teacher assessment. This guidance can include teaching students the skills needed to conduct comparisons between their text, assessment criteria, and the different feedback sources to evaluate the applicability of diverse feedback and identify potential revision strategies. Tailored assistance should be offered to students at various writing proficiency levels. Low-level students may benefit from instruction on how to respond when ChatGPT provides general, incorrect, or nonsensical feedback. In contrast, medium and high-level students should be encouraged to engage in reflective revision, exploring opportunities for extra revisions. This approach can help cultivate their independence and autonomy as learners and thinkers."	
To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing	# "Our findings suggest the potential of using LLM-based tools such as ChatGPT for evaluating student writing and supporting teacher feedback. We recommend that EFL teachers integrate ChatGPT feedback and their own feedback on student essays. During the process, teachers may make use of the advantages of ChatGPT feedback while paying attention to its limitations and addressing them using their own strengths. For example, ChatGPT is able to produce a large quantity of feedback in a few seconds, enabling teachers to be more selective about transmitting that feedback. They may examine, select, and adopt ChatGPT's comments as they see fit for their students. During their selection, teachers may rely on their knowledge of the class and their students. For instance, for students who are less skilled writers, teachers may choose fewer of ChatGPT's comments and use the comments that are the most actionable." # "Moreover, ChatGPT pays equal attention to different aspects of student writing; importantly, it appears competent in evaluating higher-level issues in student writing. Accordingly, the conventional division of labour, in which machines provide feedback on surface-level issues and teachers offer feedback on deep-level issues, should be changed. Teachers may add further comments based on ChatGPT's comments, if necessary." # "Additionally, teachers may annotate student essays to clarify the location of the problems pointed out by ChatGPT. Although it seems time efficient to provide students with Chat- GPT feedback directly, we suggest that teachers should first carefully examine the feedback generated by ChatGPT, modify it if necessary, and incorporate their own feedback into it before providing the final feedback to their students."	Guo and Wang (2023)

The integration of AI-generated feedback, particularly from tools like ChatGPT, into writing instruction has significant implications for both educators and

students. Mahapatra (2024) suggests that with proper student training, ChatGPT can be an effective feedback tool in large writing classes, highlighting the importance of instructor awareness of ChatGPT's capabilities and appropriate guidance for students. This aligns with Escalante et al.'s (2023) recommendation for a mixed approach to feedback, combining AI-generated feedback with human tutor interaction to maximize benefits.

Banihashem et al. (2024) emphasize the potential of ChatGPT as a complementary feedback source in higher education, particularly for courses with substantial enrollments and essay-writing components. They suggest that integrating ChatGPT with peer feedback could alleviate teachers' workload, especially in online courses with large student numbers. Ivanović (2023) further elaborates on this, recommending a gradual and incremental introduction of AI tools, with clear pedagogical objectives and teacher involvement in the development and selection process.

The implementation of AI in writing instruction requires careful consideration of infrastructure and support. Ivanović (2023) stresses the need for institutions to ensure necessary technical support and resources are in place. Hwang et al. (2023) highlight the potential of integrating recognition technologies with AI to support English as a Foreign Language (EFL) writing in authentic contexts. Kaharuddin et al. (2024) emphasize the importance of building a positive attitude towards AI use in learning, as it mediates the effect of reading and teacher feedback on writing skills development.

Several researchers point out areas for improvement and future development of AI writing tools. Yang et al. (2024) suggest enhancing Automated Writing Evaluation (AWE) systems to provide more meaningful and personalized feedback,

potentially incorporating deep learning and AI-supported learning. They also recommend designing future AI-programmed apps to give feedback on conceptual ideas and more genre types. Rad et al. (2023) propose using AI tools like Wordtune to enhance writing feedback literacy and reduce tension between instructors' feedback practices and students' expectations.

However, researchers also caution against over-reliance on AI-generated feedback. Lu (2019) emphasizes that AI writing evaluation systems should be used as tools to assist teaching, not replace teacher guidance. Steiss et al. (2024) warn about the potential negative effects of not engaging with early student drafts on teachers' knowledge of their students and subsequent evidence-based instructional planning. They stress the importance of teachers overseeing the AI feedback process and considering the local context for best use cases and limitations.

Lu et al. (2024) recommend incorporating ChatGPT as a supplementary tool for writing assessment, using it as a reference for teachers' own assessments. They suggest providing students with specific guidance on effectively utilizing both ChatGPT and teacher assessment, tailoring assistance to students at various writing proficiency levels. Guo and Wang (2023) propose a new approach where teachers examine, select, and adopt ChatGPT's comments based on their knowledge of the class and students, moving away from the conventional division of labor between machines and teachers in providing feedback.

In conclusion, these studies suggest that while AI-generated feedback tools offer significant potential to enhance writing instruction and assessment, their effective implementation requires careful consideration of pedagogical objectives, gradual integration, ongoing teacher involvement, and a balanced approach that combines the strengths of both AI and human feedback.

Assessment of Writing

Purpose

There are four studies that explore the role of AI tools in the *assessment of writing*.

Table 25

Assessment of Writing: Purpose

Title	Purpose	Research Question	Author
Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning	“This study examined the robustness and efficiency of four large language models (LLMs), GPT-4, GPT-3.5, iFLYTEK and Baidu Cloud, in assessing the writing accuracy of the Chinese language.”	“1. How robust are LLMs (GPT-4, GPT-3.5, iFLYTEK Open Platform, and Baidu AI Cloud) in assessing writing accuracy in general at the T-unit level and sentence level? 2. How efficient are LLMs (GPT-4, GPT-3.5, iFLYTEK Open Platform, and Baidu AI Cloud) in assessing writing accuracy in general at the T-unit level and sentence level? 3. What types of discrepancies exist between human rating results and AI rating results?”	Jiang et al. (2023)
The Death of the Short-Form Physics Essay in the Coming AI Revolution	“to present evidence of AI-generated short-form essays achieving First-Class grades on an essay writing assessment from an accredited, current university Physics module.”	Not stated.	Yeadon et al. (2022)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“to shed light on the challenges faced by English as a Second Language (ESL) lecturers in identifying AI-generated texts, and highlighting the skills and resources needed to enhance their detection capabilities”	“(1) How effective are current plagiarism detectors in providing a reliable measure of detection for identifying originality in academic writing and distinguishing human-generated text from AI-generated text? (2) What criteria are employed by ELS lecturers to assess academic writing?; and (3) How effective are these criteria in distinguishing human-generated text from AIgenerated text?”	Alexander et al. (2023)
Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?	“To address these [uncertainties regarding how ChatGPT complements teacher assessment and the overall effectiveness of this combined approach], we conducted a mixed-methods	“RQ1: What are the features of scores and feedback provided by teachers and ChatGPT in the assessment of students' writing? RQ2: How do students perceive teacher and ChatGPT feedback?	Lu et al. (2024)

Title	Purpose	Research Question	Author
	investigation involving 46 undergraduate students from a research university in southern China, engaging them in a Chinese academic writing task”	RQ3: How do students implement teacher and ChatGPT feedback to facilitate their revisions?”	

The first study (Jiang et al., 2023) evaluates the performance of various AI models in assessing writing accuracy in a second language context, while the second study (Yeadon et al., 2022) assesses the ability of AI-generated essays to achieve high academic grades in Physics. The third study (Alexander et al., 2023) examines the detection of AI-generated texts and the effectiveness of plagiarism detectors and assessment criteria in higher education. Finally, the fourth study (Lu et al., 2024) investigates the complementary role of ChatGPT in teacher assessments of student writing.

Results

The results from the studies that investigate the *assessment of writing* are shown in Table 26.

Table 26

Assessment of Writing: Results

Title	Results	Author
Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning	“The results indicate that GPT models and iFLYTEK achieved similar accuracy scores, with GPT-4 excelling in precision. These findings provide insights into the potential of LLMs in supporting the assessment of writing accuracy for language learners. The present study explores the effectiveness of GPT-4, GPT-3.5, iFLYTEK and Baidu in assessing writing accuracy. The analysis focused on metrics such as accuracy, precision, recall and F1 score. Each metric offers unique insights into different aspects of model performance. Although GPT-4 had more FNs, which led to a lower recall score, it had the lowest number of FPs at the same time. iFLYTEK and GPT-3.5 had higher recall and a lower precision score than GPT-4. Baidu, on the other hand, had a very low rate of FNs, and its FP rate was the highest. The discrepancy between a high recall score and lower accuracy and precisionscores suggests a potential trade-off between correctly identifying positive instances and	Jiang et al. (2023)

Title	Results	Author
	the presence of false positives. This trade-off highlights the need for a balanced evaluation and the consideration of the study context. Human raters had higher tolerance for variations in writing quality and can gain a more nuanced understanding of the writing by taking into account the context and overall effectiveness of communication. On the other hand, models like GPT-4 relied on predefined language patterns and tended to make more conservative judgments. As a result, these models were more prone to identify correct samples as incorrect, particularly when the expression used varied from the standardized norm. Adherence to a strict language pattern contributed to a high rate of FNs.”	
The Death of the Short-Form Physics Essay in the Coming AI Revolution	“Fifty AI answers were generated to create ten submissions that were independently marked by five separate markers. The AI generated submissions achieved an average mark of $71 \pm 2\%$, in strong agreement with the current module average of $71 \pm 5\%$. A typical AI submission would therefore most-likely be awarded a First Class, the highest classification available at UK universities. Plagiarism detection software returned a plagiarism score between $2 \pm 1\%$ (Grammarly) and $7 \pm 2\%$ (Turnitin).”	Yeadon et al. (2022)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“Findings reveal that participants tended to exploit a deficit model of assessment that focuses on error as an indicator of L2 writing output, with high levels of technical and grammatical accuracy and sophisticated language use as indicators of AI-generated text. Moreover, findings suggest that limited awareness of the characteristics and metrics used by ChatGPT, as well as lack of attention to the veracity of facts and references generated by ChatGPT, were features of participants' evaluations. This study identifies the growing challenges encountered by ESL lecturers and underlines the need for digital literacy training, targeted professional development, the use of advanced detection tools and a review of assessment policies and practices in relation to AI. Additionally, this study highlights the importance of reviewing and reinforcing institutional policies and practices that safeguard academic integrity and ensure quality higher education.”	Alexander et al. (2023)
Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?	“The intraclass correlation coefficient results revealed ChatGPT's efficiency in scoring students' writing, showing moderate to good consistency with teacher evaluations. A paired sample t-test unveiled significant differences in feedback quantity and types between ChatGPT and teacher assessments. Drawing from both interview data and quantitative findings, the study uncovers three ways in which ChatGPT complements teacher assessment, benefiting students with various writing proficiency levels: (1) fostering deeper comprehension of teacher assessments among students, (2) encouraging students to make judgments regarding feedback, and (3) promoting independent thinking about revisions. This study contributes to a more comprehensive understanding of the role of ChatGPT within the context of a combined assessment approach. It underscores that certain inherent weaknesses in ChatGPT's functioning can paradoxically lead to favorable outcomes. By shedding light on the synergy between ChatGPT and teacher	Lu et al. (2024)

Title	Results	Author
	assessments, this research seeks to inform and enhance writing instruction in higher education.”	

The first study (Jiang et al., 2023) on AI in writing assessment finds that GPT-4 excels in precision but has higher false negatives compared to other models like iFLYTEK, and this highlights a trade-off between accuracy and recall, whereas human raters demonstrate better contextual understanding. The second study (Yeadon et al., 2022) concludes that AI-generated essays in a Physics course receive first-class grades with low plagiarism scores, raising concerns about academic integrity. The third study (Alexander et al., 2023) concludes that ESL lecturers struggle to identify AI-generated texts due to too much focus on technical accuracy and a lack of awareness of AI metrics, and this emphasizes the need for enhanced digital literacy and detection tools. The last study (Lu et al., 2024) shows that ChatGPT exhibits moderate to good consistency with teacher evaluations, complementing teacher feedback by fostering student comprehension and independent thinking, which suggests a synergistic approach to writing assessments.

Implications

Table 27 provides the implications of these studies.

Table 27

Assessment of Writing: Implications

Title	Implications	Author
Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning	“These findings provide insights into the potential of LLMs in supporting the assessment of writing accuracy for language learners. Areas for improvement in future AI-facilitated writing assessment programs can target grammar structure component detection and redundancies, as well as word usages. For now, although AI models can offer valuable assistance in detecting errors and providing suggestions, human evaluations remain crucial for nuanced language analysis, context specific interpretations and making systematic assessments of writing accuracy.”	Jiang et al. (2023)

Title	Implications	Author
The Death of the Short-Form Physics Essay in the Coming AI Revolution	“The results of this short study suggest modern AI systems such as davinci-003 and ChatGPT are capable of writing high quality Physics essays, capable of achieving a first-class score, for free and within a few seconds. These tools essentially represent a wide spread democratization of paid essay writing services whereby anyone can have davinci-003 or ChatGPT write an essay for them. We could argue that, in contrast to merely sounding a death-knell for certain kinds of assessments, this technology will once again force us to re-think assessment, and will confer far greater benefits in the end than the comparatively small drawback of having to redesign existing assessments. Obviating the abuse of these technologies to undermine assessments could be as simple as requiring the students to produce the work within an invigilated setting. In such cases, the moral of the story is that while the threat to authentic assessment is profound, simple and practical solutions likely already exist that won't necessarily require a total upheaval of current assessments themselves, but may force a change of context at least. Besides assessment, this technology allows users to generate innumerable, original examples, which can be shown by teachers within writing workshops. Similarly, as technologies such as ChatGPT mature we could see their use as one-on-one tutors. Although, the availability of knowledge is effectively already widely democratized. The full course contents for most degrees at Stanford and MIT have been available on their respective MOOC platforms for a decade. Yet enrolling as a student at both institutions is still very competitive and a typical online MOOC can expect a completion rate of 10% [12] which is far below that of a typical university degree. Thus it is not often clear in advance how new technology will impact Higher Education.”	Yeadon et al. (2022)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“Concerning the implications for policy and practice, the results demonstrate that to improve practice and to safeguard academic integrity, there is a need for digital training for educators with respect to the use of AI tools in the ESL classroom and the use of advanced detection tools. There is also a need for revision of assessment policies and procedures and the development of suitable rubrics and assessment criteria. Lastly, lecturers’ tendency to confuse student-written essays for AI-generated text points out the need for lecturers to be trained on the features of AI-created texts and the differences between them and humanwritten essays.”	Alexander et al. (2023)
Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?	“It is recommended that teachers consider incorporating ChatGPT as a supplementary tool for writing assessment. They can utilize ChatGPT’s scoring and feedback as a reference for their own assessments, thereby reducing the workload and leveraging the strengths of both approaches. Teachers should provide students with specific guidance on how to effectively utilize ChatGPT and teacher assessment. This guidance can include teaching students the skills needed to conduct comparisons between their text, assessment criteria, and the different feedback sources to evaluate the applicability of diverse feedback and identify potential revision strategies. Tailored assistance should be offered to students at various writing proficiency levels. Low-level students may benefit from instruction on how to respond when ChatGPT provides	Lu et al. (2024)

Title	Implications	Author
	general, incorrect, or nonsensical feedback. In contrast, medium and high-level students should be encouraged to engage in reflective revision, exploring opportunities for extra revisions. This approach can help cultivate their independence and autonomy as learners and thinkers.”	

The integration of AI, particularly Large Language Models (LLMs) like ChatGPT, into writing assessment and education has significant implications for both educators and students. While AI shows promise in supporting writing assessment, there are important considerations and limitations to address.

Jiang et al. (2023) highlight the potential of LLMs in supporting the assessment of writing accuracy for language learners. However, they emphasize that human evaluations remain crucial for nuanced language analysis, context-specific interpretations, and systematic assessments of writing accuracy. They suggest that future AI-facilitated writing assessment programs should focus on improving grammar structure component detection, addressing redundancies, and enhancing word usage analysis.

The capabilities of AI in generating high-quality essays raise concerns about academic integrity. Cutts et al. (2023) demonstrate that modern AI systems can produce first-class Physics essays within seconds, essentially democratizing paid essay writing services. They argue that while this technology poses a threat to authentic assessment, practical solutions may already exist, such as requiring students to produce work in invigilated settings. The authors also suggest that this technology could be beneficial for generating examples for writing workshops and potentially serving as one-on-one tutors.

Addressing the challenges posed by AI-generated writing, Hayat et al. (2023) emphasize the need for digital training for educators in using AI tools in the ESL classroom and employing advanced detection tools. They call for a revision of

assessment policies and procedures, as well as the development of suitable rubrics and assessment criteria. Importantly, they highlight the need for lecturers to be trained in recognizing the features of AI-created texts and distinguishing them from human-written essays.

Lu et al. (2024) recommend incorporating ChatGPT as a supplementary tool for writing assessment, using it as a reference for teachers' own assessments. They suggest providing students with specific guidance on effectively utilizing both ChatGPT and teacher assessment. This includes teaching students to compare their text with assessment criteria and different feedback sources, evaluating the applicability of diverse feedback, and identifying potential revision strategies. The authors emphasize the importance of tailoring assistance to students at various writing proficiency levels, with different approaches for low-level and medium to high-level students.

In conclusion, while AI tools offer significant potential to enhance writing assessment and instruction, their effective implementation requires careful consideration of their limitations, ongoing teacher training, and a balanced approach that combines the strengths of both AI and human assessment. The implications of these studies suggest a need for a thoughtful integration of AI in writing education, with a focus on maintaining academic integrity, improving assessment practices, and fostering student autonomy in the writing process.

Automated Evaluation or Scoring

Purpose

There are three studies that focus on the role of AI tools for *automated evaluation and scoring*.

Table 28*Automated Evaluation or Scoring: Purpose*

Title	Purpose	Research Question	Author
Computer-assisted EFL writing and evaluations based on artificial intelligence: a case from a college reading and writing course	“to explore students' expectations and perceived effectiveness of computer-assisted review tools, and the differences in reliability and validity between human evaluation and automatic evaluation, to find a way to improve students' English writing ability”	“1. what are the respondents' expectations and subjective evaluations of AEE? 2. what are the respondents perceived effectiveness of AEE in their English writing improvement? 3. what are the advantages and disadvantages of AEE in compassion with the human rating?”	Wang (2020)
ChatGPT for Automated Writing Evaluation in Scholarly Writing Instruction	“to explore the suitability of ChatGPT for AWE in writing instruction”	Not stated.	Parker et al. (2023)
An intelligent scoring system for English writing based on artificial intelligence and machine learning	“to establish an intelligent English writing scoring system based on Artificial Intelligence (AI) and explore the application of multi-task learning algorithm models in machine learning”	Not stated.	Huang (2023)

The first study (Wang, 2020) aims to explore students' expectations and perceived effectiveness of computer-assisted review tools in improving their English writing skills and to compare the reliability and validity of human versus automatic evaluations. The second study (Parker et al, 2023) seeks to investigate the suitability of ChatGPT for Automated Writing Evaluation (AWE) in scholarly writing instruction. The third study (Huang, 2023) focuses on developing an intelligent English writing scoring system using AI and machine learning and examines the application of multi-task learning algorithm models. These studies collectively address the integration of AI tools in writing instruction and assessment, highlighting their potential benefits and challenges.

Results

The results from these three studies are shown in Table 29.

Table 29*Automated Evaluation or Scoring: Results*

Title	Results	Author
Computer-assisted EFL writing and evaluations based on artificial intelligence: a case from a college reading and writing course	“The experiment revealed that the respondents held high expectations for the computer-assisted evaluation tools, and the effectiveness of computer scoring feedback on students was higher than that of teacher scoring feedback. Moreover, at the end of the writing project, the students' independent learning ability and English writing ability were significantly improved. Besides, there was a positive correlation between students' initial expectations of computer-assisted learning tools and the final evaluation of learning results.”	Wang (2020)
ChatGPT for Automated Writing Evaluation in Scholarly Writing Instruction	“ChatGPT demonstrated utility as an AWE tool. Its scoring performance demonstrated stricter grading than human raters, related feedback to macro-level writing features, and supported multiple submissions and learner autonomy.”	Parker et al. (2023)
An intelligent scoring system for English writing based on artificial intelligence and machine learning	“The experimental results show that the multi-task learning model performs very well. There are 49,594 syntax errors in the training set, including 11,076 redundant word errors, 13,246 missing word errors, 21,898-word selection errors and 3374-word order errors. The test set has 11,085 syntax errors, including 2406 redundant word errors, 2973 missing word errors, 4860 vocabulary selection errors and 846-word order errors. Finally, the accuracy of the scoring system is evaluated, which is 91.5%. The accuracy of the algorithm can reach 97%. By training the model, the multi-task learning algorithm model can save a lot of time and labor costs for manually correcting essays.”	Huang (2023)

The results of these studies show the potential and effectiveness of AI and computer-assisted tools in writing evaluation and instruction. The findings from the first study (Wang, 2020) suggest that students have high expectations for computer-assisted evaluation tools, which are more effective in improving students' English writing and independent learning abilities than traditional teacher feedback. Additionally, there is a positive correlation between students' initial expectations and their final learning outcomes. The second study (Parker et al, 2023) demonstrates that ChatGPT is a useful AWE tool, providing stricter grading than human raters, offering feedback on macro-level writing features, and supporting learner autonomy through multiple submissions. The third study (Huang, 2023) shows that the

intelligent scoring system using a multi-task learning model performs exceptionally well, identifying various syntax errors with high accuracy, reaching 97% in the algorithm's performance, and significantly reducing the time and labor costs associated with manual essay correction.

Implications

Table 30 shares the implications of these studies.

Table 30

Automated Evaluation or Scoring: Implications

Title	Implications	Author
Computer-assisted EFL writing and evaluations based on artificial intelligence: a case from a college reading and writing course	“The AEE system can adequately compensate for the failure to give untimely corrective feedback in traditional teaching, help students improve their English writing performance and promote the improvement of students’ writing ability. Moreover, it stimulates students’ motivation and self-efficacy, cultivates students’ interest and confidence in writing and help build up the ability to learn independently. Furthermore, the evolution of AI will transform the writing teachers’ role from merely giving corrective feedback to constructing a learner-friendly social engagement context. Investing in intelligent computing technology and facilitating its use in classrooms is of significant policy implications.”	Wang (2020)
ChatGPT for Automated Writing Evaluation in Scholarly Writing Instruction	“ChatGPT is a new technology, and there is much to learn about its applications in nursing education. When conducting the exploratory assessment for this paper, it became immediately apparent that users must know how to ask the right questions to obtain the desired feedback. Prompt construction has received growing attention within the literature, as generative AIs must be supplied with optimized prompts to produce appropriate content (Nachson et al., 2023; Sanmarchi et al., 2023). Thus, it will be important to monitor emerging research on how to prompt ChatGPT correctly and educate faculty and students on how to best use ChatGPT for AWE. We also confirmed that ChatGPT could provide individualized feedback quickly. Emerging research on ChatGPT's capabilities has described its personalized interactions with users and deemed it a "powerful self-learning" tool (Sallam, 2023). Thus, by using GAI tools such as ChatGPT for feedback, student learning can be self-driven and individualized. Though GAI is a rapidly advancing technology, these exploratory applications in providing AWE-style feedback are a step forward in improving the efficiency and effectiveness of writing instruction in nursing education.”	Parker et al. (2023)
An intelligent scoring system for English writing based on artificial intelligence and machine learning	“Consequently, students can learn autonomously without teachers’ guidance. In the future, proofreading algorithms can be incorporated into the system to enhance their error correction training and achieve the integration of error detection and correction. In addition, machine translation	Huang (2023)

Title	Implications	Author
	is applied to the grammar error correction process of the algorithm to improve its ability to detect and correct syntax error in English writing, which will significantly promote students' autonomous learning and reduce their dependence on teachers.”	

The implications of these studies highlight the transformative potential of AI in writing instruction and evaluation. The first study (Wang, 2020) suggests that AI-enabled evaluation systems can provide timely corrective feedback, significantly improving students' English writing performance, motivation, self-efficacy, and independent learning abilities. This evolution shifts the role of writing teachers from giving corrective feedback to fostering a supportive and engaging learning environment, emphasizing the importance of investing in intelligent computing technologies for educational policy. The second study (Parker et al, 2023) underscores the need for proper training in prompt construction to effectively use ChatGPT for Automated Writing Evaluation (AWE). As ChatGPT can provide quick, individualized feedback and facilitate self-driven learning, it is crucial for educators and students to understand how to optimize prompts to maximize the tool's benefits. This emphasizes the importance of monitoring emerging research and educating faculty and students on effective usage, particularly in specialized fields like nursing education. The third study (Huang, 2023) implies that an intelligent scoring system with advanced AI and machine learning algorithms can enable students to learn autonomously without constant teacher guidance. The incorporation of proofreading algorithms and machine translation into the system can enhance its error detection and correction capabilities, further promoting autonomous learning and reducing student dependency on teachers. This points to a future where AI plays a critical role in personalized and efficient writing instruction, allowing teachers to focus on more complex educational tasks.

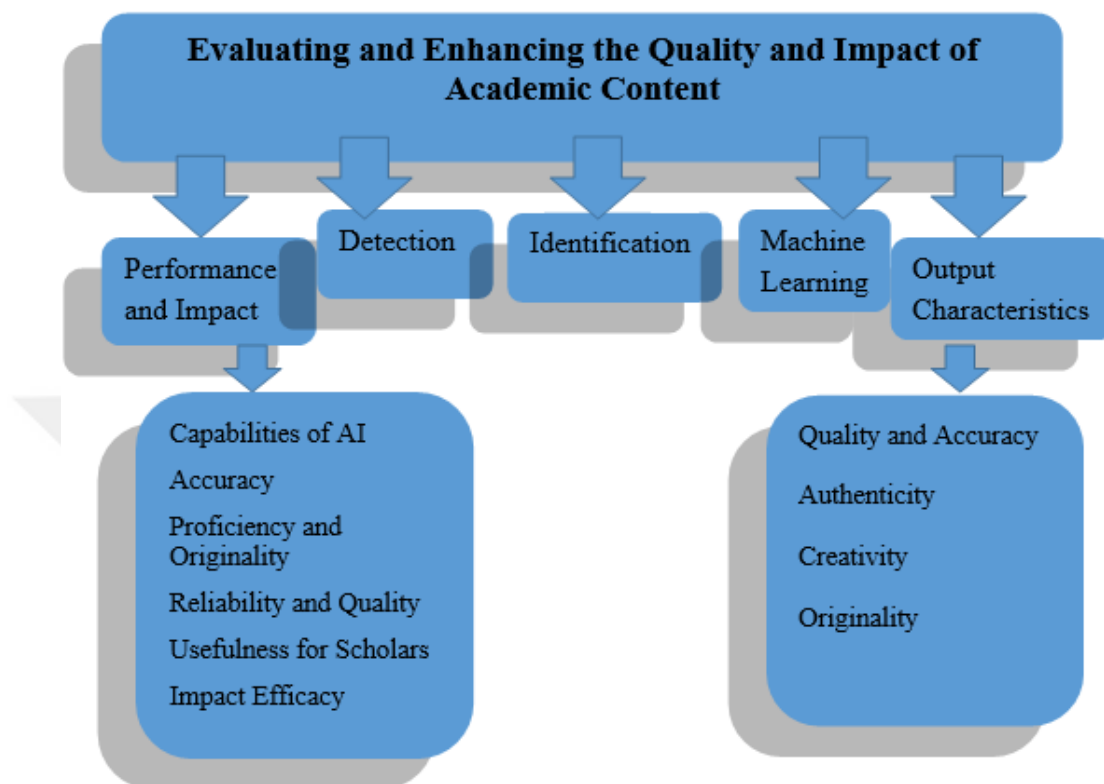
Evaluating and Enhancing the Quality and Impact of Academic Content

In a comprehensive review of 103 studies, 41 were found to focus primarily on *evaluating and enhancing the quality and impact of academic content*. These studies encompass a diverse range of research objectives, including the evaluation of AI writing tools' performance and the analysis of characteristics inherent to AI-generated content. Additionally, a subset of these studies investigates the efficacy of detection methodologies and the variables influencing the identification of AI-authored texts.

The overarching core concept of AI Writing Systems can be further delineated into several subcategories: *performance and impact, detection, identification, machine learning, and output characteristics*. Within the *performance and impact* subcategory, researchers have explored various aspects such as *capabilities, accuracy, proficiency and originality, reliability and quality, usefulness for scholars, and impact efficacy*. Similarly, the *output characteristics* subcategory encompasses studies examining the *quality and accuracy, authenticity, creativity, and originality* of AI-generated content (Figure 7).

Figure 7

Evaluating and Enhancing the Quality and Impact of Academic Content



Performance and Impact: Capabilities of AI

Purpose

There are six studies that aim at understanding the ability of AI tools to generate content.

Table 31

Capabilities of AI: Purpose

Title	Purpose	Research Question	Author
Strengths and Weaknesses of ChatGPT Models for Scientific Writing About Medical Vitamin B12: Mixed Methods Study	“to evaluate the ability of GPT-4 to generate scientific content and assist in scientific writing using medical vitamin B12 as the topic. Furthermore, the study will compare the performance of GPT-4 to its predecessor, GPT-3.5.”	Not stated.	Abuyaman (2023)

Title	Purpose	Research Question	Author
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities	"to provide a comprehensive analysis of the capabilities and limitations of six AI chatbots in scholarly writing in the humanities and archaeology"	Not stated.	Lozic and Štular (2023)
Clause complexing in research-article abstracts: Comparing human- and AI-generated texts	"to compare 50 abstracts of scientific research articles from Nature with generated versions produced by Bard, ChatGPT, and Poe Assistant by using a modified version of Halliday's clause-complexing framework"	"how do the occurrence rates of clauses, tactic relations, and logicosemantic relations in science abstracts compare with the same in AI-generated abstracts?"	Leong (2023)
Clinical Research With Large Language Models Generated Writing-Clinical Research with AI-assisted Writing (CRAW) Study	to assess GPT-3.5's ability to craft the background section of critical care clinical research questions compared to medical researchers with H-indices of 22 and 13.	Not stated.	Huespe et al. (2023)
GPT-4 Artificial Intelligence Model Outperforms ChatGPT, Medical Students, and Neurosurgery Residents on Neurosurgery Written Board-Like Questions	"to examine GPT-4's competence on neurosurgical board-style questions, comparing its performance with medical students and residents, to explore its potential in medical education and clinical decision-making"	Not stated.	Guerra et al. (2023)
ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners	"to investigate the effectiveness of the Chat Generative Pre-trained Transformer (ChatGPT) in detecting English as a foreign language (EFL) learners' writing errors compared to human instructors"	"1. Can the ChatGPT tool identify all the writing errors of the EFL learners and the underlying causes of these errors? 2. What are the capabilities of ChatGPT in analyzing EFL errors compared to instructors of English as a foreign language? 3. What are the discrepancies of the tool in accounting for the errors?"	Algaraady and Mahyoob (2023)

The six studies aim to evaluate the effectiveness of AI models like GPT-4 and GPT-3.5 in various aspects of scientific and academic writing. They compare

AI-generated content to human-generated content in fields ranging from medical writing and humanities to clinical research and education, highlighting strengths and weaknesses in accuracy, proficiency, and error detection. These studies collectively seek to understand the potential of AI in enhancing writing quality and educational outcomes, while also identifying limitations and areas for improvement in AI-assisted writing tools.

Results

The results from these six studies can be found in Table 32.

Table 32

Capabilities of AI: Results

Title	Results	Author
Strengths and Weaknesses of ChatGPT Models for Scientific Writing About Medical Vitamin B12: Mixed Methods Study	“The results indicated that GPT-4 can potentially streamline scientific writing through its ability to edit language and write abstracts, keywords, and abbreviation lists. However, significant limitations of ChatGPT were revealed, including its inability to identify and address bias, inability to include recent information, lack of transparency, and inclusion of inaccurate information. Additionally, it cannot check for plagiarism or provide proper references. The accuracy of GPT-4’s answers was found to be superior to that of GPT-3.5. ChatGPT can be considered a helpful assistant in the writing process but not a replacement for a scientist’s expertise.”	Abuyaman (2023)
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots’ Proficiency and Originality in Scientific Writing for Humanities	“In the quantitative test, ChatGPT-4 scored near the passing grade (−5), whereas ChatGPT-3.5 (−18), Bing (−21), and Bard (−31) were not far behind. Claude 2 (−75) and Aria (−80) scored much lower. In the qualitative test, all AI chatbots, but especially ChatGPT-4, demonstrated proficiency in recombining existing knowledge, but all failed to generate original scientific content. As a side note, our results suggest that with ChatGPT-4, the size of large language models has reached a plateau. Furthermore, this paper underscores the intricate and recursive nature of human research. This process of transforming raw data into refined knowledge is computationally irreducible, highlighting the challenges AI chatbots face in emulating human originality in scientific writing. Our results apply to the state of affairs in the third quarter of 2023. In conclusion, while large language models have revolutionised content generation, their ability to produce original scientific contributions in the humanities remains limited.”	Lozic and Štular (2023)
Clause complexing in research-article	# “(a) The occurrence rates of main clauses (4.59) and subordinate clauses (1.71) in Nature abstracts occupied a	Leong (2023)

Title	Results	Author
abstracts: Comparing human- and AI-generated texts	middle position relative to other abstract groups. The rate of embedded clauses for Nature abstracts was 3.35, differing only from that for Bard abstracts. (b) With the exception of Bard abstracts, the rate of simplexes in the other three abstract groups was about 2. The proportion of simplexes to main clauses in Nature, ChatGPT, and Poe abstracts was between 44% and 52%. (c) Parataxis and paratactic logico-semantic relations were uncommon, and the differences among the abstract groups were statistically insignificant. (d) Hypotactic logico-semantic relations in Nature abstracts were more common among ranking clauses (1.64) than embedded clauses (0.49)." # "The low occurrence rate of embedded hypotactic clauses (<1) was also generally true for the other abstract groups. Hence, even though embedded clauses occurred more often than subordinate clauses, it is the latter that carried the bulk of logico-semantic relations. The most common hypotactic logico-semantic relation involving ranking clauses in Nature and ChatGPT abstracts was enhancement. While ChatGPT abstracts relied mostly on non-finite clauses to express circumstantial information, Nature abstracts also used finite adverbial clauses. None of the chatbots matched Nature in all categories. Poe abstracts, in fact, differed from Nature abstracts in every category (with a significant ANOVA result) except the use of embedded clauses. The chatbot that came closest to matching Nature abstracts was ChatGPT; it differed in the use of subordinate, elaborating, and finite adverbial clauses. Even so, distinct differences were detected—ChatGPT was twice as likely to use elaborating clauses, but only one sixth as likely to use finite adverbial clauses. The statistics for Nature abstracts offer a possible means by which original texts can be separated from AI-generated ones. Rather than content, the focus here is on the building blocks that construct such content. The work of Levin et al. (2023) has demonstrated one way to do this using Grammarly; this present work proposes another using clause complexing."	
Clinical Research With Large Language Models Generated Writing-Clinical Research with AI-assisted Writing (CRAW) Study	"A total of 80 researchers completed the survey. The median H-index was 3 (interquartile range, 1-7.25) and most (36%) researchers were from the Critical Care specialty. When compared with researchers with an H-index of 22 and 13, GPT-3.5 was marked high on the Likert scale ranking on main background components (median 4.5 vs. 3.82 vs. 3.6 vs. 4.5, respectively; $p < 0.001$). The sensitivity and specificity to detect researchers writing versus GPT-3.5 writing were poor, 22.4% and 57.6%, respectively. Conclusions and Relevance: GPT-3.5 could create background research content indistinguishable from the writing of a medical researcher. It was marked higher compared with medical researchers with an H-index of 22 and 13 in writing the background section of a critical care clinical research question."	Huespe et al. (2023)
GPT-4 Artificial Intelligence Model Outperforms	"GPT-4 attempted 91.9% of Congress of Neurological Surgeons SANS questions and achieved 76.6% accuracy. The model's accuracy increased to 79.0% for text-only	Guerra et al. (2023)

Title	Results	Author
ChatGPT, Medical Students, and Neurosurgery Residents on Neurosurgery Written Board-Like Questions	questions. GPT-4 outperformed Chat Generative pre-trained transformer ($P < 0.001$) and scored highest in pain/peripheral nerve (84%) and lowest in spine (73%) categories. It exceeded the performance of medical students (26.3%), neurosurgery residents (61.5%), and the national average of SANS users (69.3%) across all categories.”	
ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners	“Our results revealed that ChatGPT successfully identified most surface-level errors but could not detect writing errors related to deep structure and pragmatics. Conversely, human teachers could spot most of these issues. These findings suggest that while ChatGPT can be a valuable tool in identifying surface-level errors, it cannot replace human instructors' expertise and nuanced understanding in detecting errors related to the more complex aspects of writing. The writing error types (data) are statistically analyzed. The descriptive analysis displays valuable insights into the reliability of the data and its potential implications, where the F-score, which measures the statistical model accuracy, is found to be 1.5. In the meantime, the p-value score, which shows the probability of obtaining results as extreme as the detected data, is calculated to be 0.23. The results suggest that the collected data is statistically significant, and further analysis may yield valuable insights.”	Algaraady and Mahyoob (2023)

According to the results from the first study (Abuyaman, 2023), GPT-4 shows potential in streamlining scientific writing tasks such as editing language and creating abstracts, but it also demonstrated significant limitations, including bias, outdated information, lack of transparency, and inaccuracies. Despite being a useful assistant, GPT-4 cannot replace the expertise of a scientist. The results from the second study (Lozic and Štular, 2023) suggest that ChatGPT-4 performed better than other chatbots in recombining existing knowledge but failed to generate original scientific content, highlighting the limitations of AI in producing truly innovative research. The study concludes that while large language models have advanced content generation, their ability to create original contributions in humanities remains limited. The third study (Leong, 2023) concludes that ChatGPT-generated abstracts came closest to matching those from Nature in terms of clause complexity but still showed distinct differences, such as higher use of elaborating clauses and lower use

of finite adverbial clauses. This suggests that while AI can mimic human writing structures to some extent, significant differences remain in how content is constructed.

Based on the results of the fourth study (Huespe et al., 2023), GPT-3.5 creates background research content that is indistinguishable from that written by medical researchers and is rated higher than the work of researchers with H-indices of 22 and 13. However, the sensitivity and specificity for detecting AI-generated content are low, indicating the difficulty in differentiating between human and AI writing. The fifth study (Guerra et al., 2023) shows that GPT-4 achieves a 76.6% accuracy on neurosurgical board-style questions, outperforming ChatGPT, medical students, and neurosurgery residents. It scores highest in pain/peripheral nerve questions and lowest in spine questions, surpassing the national average of SANS users. The findings from the last study (Algaraady & Mahyoob, 2023) suggest that ChatGPT effectively identifies surface-level writing errors but struggles with errors related to deep structure and pragmatics, which human teachers can detect more accurately. These findings suggest that while ChatGPT is useful for surface-level error detection, it cannot replace the comprehensive expertise of human instructors for more complex writing issues.

Implications

Table 33 provides the implications the studies aiming to understand the ability of AI tools.

Table 33

Capabilities of AI: Implications

Title	Implications	Author
Strengths and Weaknesses of ChatGPT Models for	“Through prompts related to vitamin B12, potential was found in the GPT-4 model of ChatGPT for scientific writing. This tool could particularly benefit non-English speaking scientists by	Abuyaman (2023)

Scientific Writing About Medical Vitamin B12: Mixed Methods Study	helping them streamline the language in their publications. This study revealed critical limitations of ChatGPT, mainly in relation to accuracy, bias, and references. The shortcomings warrant judicious use and avoiding its use as a substitute for human expertise. GPT-4 showed improvements compared to GPT-3.5, indicating promise for the further development of the tool. Overall, ChatGPT has the potential to be a useful tool for scientific writing, but researchers must be aware of its limitations and use it appropriately.”	
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots’ Proficiency and Originality in Scientific Writing for Humanities	“Additional shortcomings of the AI-generated content that we found in our test include reasoning errors, hallucinations, and biases. Reasoning errors refer to the chatbot’s inability to critically evaluate and link cause-and-effect relationships, as evidenced by several historical inaccuracies regarding the migration patterns and historical periods of the Slavs. Hallucinations denote confident but unsubstantiated claims by the AI, such as invented references and inaccurate dates. Our test also reveals significant biases in the content generated by the AI. These biases manifest as language bias, favouring English sources over other relevant languages; neo-colonial bias, displaying a preference for Western authors; and citation bias, skewing towards older and more highly cited publications. These findings highlight that, despite their technological prowess, AI chatbots remain reliant on their training data, echoing and even amplifying existing biases and knowledge gaps. Given that they veer towards past data, including, in our test, a book originally published in 1895, they are likely to be too conservative in their recommendations. Since the listed deficiencies are relatively inconspicuous compared to, for example, gender or racial biases, it is unlikely that they will be remedied in the foreseeable future by resourceintensive reinforcement learning from human feedback. These biases are among the key concerns with the use of AI chatbots in scientific writing, as most of them are less likely to be highlighted in the review processes.”	Lozic and Štular (2023)
Clause complexing in research-article abstracts: Comparing human- and AI-generated texts	“Scholars (and students) using chatbots to generate content for publication remains a vexing problem. While generative AI has its use in facilitating personalized learning and automating common, tedious tasks, it is merely a tool, but not an author; it has no independent thought, and so cannot take responsibility for its own generated content. Several journals, such as the PNAS (Proceedings of the National Academy of Science), Sage, and Science families of journals, have already banned the listing of chatbots as a ‘coauthor’ in articles; generated content in the manuscript must also be clearly attributed. The recent voluntary commitments by OpenAI, Google, and Meta to watermark AI-generated content is a welcome move (Bartz & Hu, 2023). But detection programs continue to be needed, given how quickly chatbot texts can proliferate, not all of which may be watermarked.	Leong (2023)
Clinical Research With Large Language Models Generated Writing-Clinical Research with AI-assisted Writing (CRAW) Study	“... the use of GPT-3.5 as a tool for generating scientific backgrounds can be particularly beneficial for non-native English speakers, as it provides assistance in generating fluency and coherence in writing.”	Huespe et al. (2023)

GPT-4 Artificial Intelligence Model Outperforms ChatGPT, Medical Students, and Neurosurgery Residents on Neurosurgery Written Board-Like Questions	“AI’s ability to synthesize clinical information for patient care must be continuously evaluated. As Open AI’s multimodal model continues to develop, we must be ready to test and adopt this robust tool and harness its strengths while limiting its weaknesses as clinicians to optimize clinical training and patient care. Open AI’s rapid improvement in multimodal language models signifies a promising future of integration into patient care. The potential for AI-integrated patient care is highlighted by the impressive performance of GPT-4 when compared to medical students, neurosurgery residents, and the nation’s average SANS user scores. As OpenAI begins its beta testing of ChatGPT plugins as of March 25, 2023, which allows internet integration and thirdparty web plugins, these developments will undoubtedly open the door for further AI-based tools designed specifically for the neurosurgeon.”	Guerra et al. (2023)
ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners	“The implications of our research are significant for both educators and students. By incorporating ChatGPT, educators can benefit from a more efficient and accurate analysis method, improving language proficiency and writing skills. This finding is consistent with the studies by Chen (2016). & Yan (2023), which emphasized the role of LLMs in providing personalized and immediate feedback to learners. Additionally, students can receive timely feedback and guidance in their writing, as highlighted by the findings of Lin et al. (2022), who emphasized the positive impact of LLM-based feedback on the development of EFL learners' writing skills.”	Algaraady and Mahyoob (2023)

The study on ChatGPT models for scientific writing about vitamin B12

(Abuyaman, 2023) implies that while GPT-4 can assist non-English speaking scientists by streamlining language, its limitations in accuracy, bias, and references mean it should not replace human expertise. The analysis of AI chatbots’ proficiency in the humanities (Lozic & Štular, 2023) reveals that, despite their content generation capabilities, they still suffer from reasoning errors, hallucinations, and biases, emphasizing the need for cautious use and further development to address these issues. The comparison of AI-generated and human-written research article abstracts (Leong, 2023) highlights the ongoing necessity for detection programs and clear attribution of AI-generated content, as AI cannot assume authorship responsibility. The clinical research (Huespe et al., 2023) study suggests that GPT-3.5 can significantly aid non-native English speakers by enhancing fluency and coherence in scientific writing. The evaluation of GPT-4’s performance on neurosurgery board-

like questions (Guerra et al., 2023) underscores the potential for AI to integrate into clinical training and patient care, provided its strengths are harnessed and weaknesses mitigated. Finally, the investigation into ChatGPT's capabilities in spotting EFL writing errors (Algaraady & Mahyoob, 2023) indicates that it can provide efficient and accurate feedback, thereby benefiting both educators and students in improving language proficiency and writing skills.

Performance and Impact: Accuracy of AI

Purpose

There are five studies related to accuracy of AI.

Table 34

Accuracy of AI: Purpose

Title	Purpose	Research Question	Author
Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning	"This study examined the robustness and efficiency of four large language models (LLMs), GPT-4, GPT-3.5, iFLYTEK and Baidu Cloud, in assessing the writing accuracy of the Chinese language."	"1. How robust are LLMs (GPT-4, GPT-3.5, iFLYTEK Open Platform, and Baidu AI Cloud) in assessing writing accuracy in general at the T-unit level and sentence level? 2. How efficient are LLMs (GPT-4, GPT-3.5, iFLYTEK Open Platform, and Baidu AI Cloud) in assessing writing accuracy in general at the T-unit level and sentence level? 3. What types of discrepancies exist between human rating results and AI rating results?"	Jiang et al. (2023)
A Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing	"In this study, we aimed to explore the accuracy and authenticity of ChatGPT3's academic writing and its similarity to natural human texts."	"Firstly, to what extent the tool's academic writing is accurate and authentic. Secondly, what is the degree of similarity between the generated virtual texts and natural human texts? Finally, what are the differences between the tool's generated texts and natural human language in the context of academics?"	Mahyoob et al. (2023)
A comparison of ChatGPT-generated articles with human-written articles	"to compare the accuracy and quality of several ChatGPT-generated	Not stated.	Ariyaratne et al. (2023)

Title	Purpose	Research Question	Author
ChatGPT in head and neck scientific writing: A precautionary anecdote	academic articles with those written by human authors.” “To evaluate the accuracy of ChatGPT references in scientific writing relevant to head and neck surgery”	Not stated.	Wu and Dang (2023)
How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing?	“This study aimed to evaluate the integrity and accuracy of the responses to the question it was asked from the ChatGPT software program.”	Not stated.	Dashti et al. (2023)

The studies collectively aim to evaluate the capabilities and limitations of AI language models, particularly ChatGPT, in academic and scientific writing. They assess the robustness, efficiency, accuracy, and authenticity of AI-generated texts compared to human-written texts, exploring discrepancies and potential biases. Additionally, these studies investigate the reliability of AI tools in specific fields, such as head and neck surgery, and their overall integrity in providing accurate and trustworthy scientific content.

Results

The results from these studies are shown in Table 35.

Table 35

Accuracy of AI: Results

Title	Results	Author
Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning	# “The results indicate that GPT models and iFLYTEK achieved similar accuracy scores, with GPT-4 excelling in precision. These findings provide insights into the potential of LLMs in supporting the assessment of writing accuracy for language learners. The present study explores the effectiveness of GPT-4, GPT-3.5, iFLYTEK and Baidu in assessing writing accuracy. The analysis focused on metrics such as accuracy, precision recall and F1 score. Each metric offers unique insights into different aspects of model performance. Although GPT-4 had more FNs, which led to a lower recall score, it had the lowest number of FPs at the same time. iFLYTEK and GPT-3.5 had higher recall and a lower precision score than GPT-4. Baidu, on the other hand, had a very low rate of FNs, and its FP rate was the highest.	Jiang et al. (2023)

Title	Results	Author
A Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing	The discrepancy between a high recall score and lower accuracy and precision scores suggests a potential trade-off between correctly identifying positive instances and the presence of false positives. This trade-off highlights the need for a balanced evaluation and the consideration of the study context.” # “Human raters had a higher tolerance for variations in writing quality and can gain a more nuanced understanding of the writing by taking into account the context and overall effectiveness of communication. On the other hand, models like GPT-4 relied on predefined language patterns and tended to make more conservative judgments. As a result, these models were more prone to identify correct samples as incorrect, particularly when the expression used varied from the standardized norm. Adherence to a strict language pattern contributed to a high rate of FNs.” “The results show that despite ChatGPT's exceptional capabilities, its serious defects are evident, as many issues in academic writing are raised. The major issues include information repetition, nonfactual inferences, illogical reasoning, fake references, hallucination, and lack of pragmatic interpretation. The multi-model principles used in this framework are based on a thorough analysis of the language patterns used in academic texts. By examining the academic language's structure, vocabulary, semantics, and other factors, we can better understand how ChatGPT produces language partially suitable for academic purposes. This paper provides a framework and multi-model principles for understanding the characteristics of academic language produced by ChatGPT, and these principles can be combined to produce a more complete and nuanced analysis.” # “The findings indicate that ChatGPT failed to produce well-formed academic texts and articles with authentic and reliable information. The tool can generate human-like language with grammatically correct sentences and well-formed structure due to the large, trained corpus. However, professional academic analysis and writing are still inadequate, mainly if there are issues related to pragmatics and social norms and values. The tool can be an assistive medium to guide and provide information about the various genres. However, researchers should not depend on the tool to write articles; it can be identified easily as (natfficial) a fabricated article and nonhuman language. This study recommended that, when using ChatGPT in academics, it is necessary to proceed cautiously.”	Mahyoob et al. (2023)
A comparison of ChatGPT-generated articles with human-written articles	“In total, 4 of the 5 articles written by ChatGPT were significantly inaccurate with fictitious references. One of the papers was well written, with a good introduction and discussion; however, all references were fictitious. ChatGPT is able to generate coherent research articles, which on initial review may closely resemble authentic articles published by academic researchers. However, all of the articles we assessed were factually inaccurate and had fictitious references. It is worth noting, however, that the articles generated may appear authentic to an untrained reader.”	Ariyaratne et al. (2023)

Title	Results	Author
ChatGPT in head and neck scientific writing: A precautionary anecdote	“Among 50 total references provided by ChatGPT, only five (10 %) were found to have the correct title, journal, authors, year of publication, and DOI. Merely 14 % of the presented references had correct DOI. References regarding free flap reconstruction for oral cancer were the least accurate from all the five categories, with no correct DOI. Complete inter-rater agreement was noted while evaluating the citations.”	Wu and Dang (2023)
How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing?	“When the author tested the ChatGPT software program's ability to locate articles in the JPD and Google Scholar using a set of keywords, the results did not match the papers that the ChatGPT software program had generated with the help of the AI tool. Consequently, all 75 articles provided by the ChatGPT software program were not accurately located in the JPD or Google Scholar databases and had to be added manually to ensure the accuracy of the relevant references.”	Dashti et al. (2023)

The first study (Jiang et al., 2023) concludes that GPT models and iFLYTEK achieve similar accuracy in assessing writing, with GPT-4 showing high precision but lower recall, indicating a trade-off between accurately identifying correct instances and the presence of false positives, while human raters provide more nuanced evaluations. Second study (Mahyoob et al., 2023) shows ChatGPT exhibits serious defects in academic writing, including information repetition, nonfactual inferences, illogical reasoning, and fake references, suggesting it can generate human-like text but falls short in producing reliable and well-formed academic articles. The third study (Ariyaratne et al., 2023) suggests ChatGPT-generated articles are coherent and closely resembled authentic research articles, but they contain significant factual inaccuracies and fictitious references, making them unreliable compared to human-written articles. According to the fourth study (Wu & Dang, 2023), among the 50 references provided by ChatGPT, only 10% are accurate, with most references being incorrect in titles, journals, authors, publication years, or DOIs, and this highlights significant inaccuracies in AI-generated citations for scientific writing. The last study (Dashti et al., 2023) shows that The ChatGPT software program fails to accurately locate articles in the JPD and Google Scholar

databases, with all 75 generated articles not matching the actual papers, necessitating manual addition of accurate references.

Implications

Table 36 provides the implications from these studies.

Table 36

Accuracy of AI: Implications

Title	Implications	Author
Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning	“These findings provide insights into the potential of LLMs in supporting the assessment of writing accuracy for language learners. Areas for improvement in future AI-facilitated writing assessment programs can target grammar structure component detection and redundancies, as well as word usages. For now, although AI models can offer valuable assistance in detecting errors and providing suggestions, human evaluations remain crucial for nuanced language analysis, context specific interpretations and making systematic assessments of writing accuracy.”	Jiang et al. (2023)
A Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing	“The findings indicate that ChatGPT failed to produce well-formed academic texts and articles with authentic and reliable information. The tool can generate human-like language with grammatically correct sentences and well-formed structure due to the large, trained corpus. However, professional academic analysis and writing are still inadequate, mainly if there are issues related to pragmatics and social norms and values. The tool can be an assistive medium to guide and provide information about the various genres. However, researchers should not depend on the tool to write articles; it can be identified easily as (natficial) a fabricated article and nonhuman language. This study recommended that, when using ChatGPT in academics, it is necessary to proceed cautiously.”	Mahyoob et al. (2023)
A comparison of ChatGPT-generated articles with human-written articles	“It is worth noting that the articles generated may appear authentic to an untrained reader, including early-career researchers. Clinicians unaware of this might believe what was written and might use this in their clinical practice or even teach the trainees. This has the potential to cascade and spread disinformation. If used in clinical practice, it can result in increased patient. Despite these concerns, there may be potential benefits to using ChatGPT in scientific research, particularly in the generation of well-worded and comprehensible text, provided factual data is entered into it by the user. ChatGPT’s ability to quickly process and analyse large volumes of data could prove useful in certain fields of study. The format of articles generated by ChatGPT can be used as a draft template to write an expanded version of the article. The language, grammar, and spelling checks can be done for the articles simultaneously. However, it is important to ensure that any findings generated by ChatGPT are verified by human experts to ensure accuracy and reliability. Overall, while ChatGPT may hold some promise in the future for certain applications in scientific research, it is	Ariyaratne et al. (2023)

	important to carefully consider its limitations and potential pitfalls, particularly in its current state. Future research should continue to investigate the role of ChatGPT in scientific publication and explore ways to mitigate its limitations and maximise its potential benefits while also maintaining the ethical standards and quality of academic research. It is crucial that researchers exercise caution when utilising Chat-GPT as a tool for scientific writing and that any generated papers are carefully scrutinised by human experts before publication if it is to be used.”	
ChatGPT in head and neck scientific writing: A precautionary anecdote	“When evaluating accuracy of chatGPT in producing references for head and neck surgery, only 10 % of the articles provided with the selected keywords were correct. A high degree of academic hallucination was noted. We would like to apprise the head and neck surgical community for an acute need to thoroughly cross verify references provided by chatGPT to prevent dissemination of fabricated/erogenous scientific data, which may negatively impact patient care/education.”	Wu and Dang (2023)
How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing?	“1. Artificial intelligence is a leading technology in all fields of science, but using it without fact checking can lead to false and unethical results: AI-generated content cannot provide or analyze the same information as an author or researcher. 2.The results indicated that writing credit or references to such content or references in prestigious academic journals are not yet appropriate. This affects scientists, researchers, scholarship proposals, publications, and peer-review process and shows the need for disclaimer statements, education related to the evaluation of scientific evidence, and complete attention to academic integrity. 3. Scientific writing is only valid when performed manually by researchers. This study found that researchers must write scientific papers with care when using the ChatGPT software program or other AI tools and should not rely solely on such tools.”	Dashti et al. (2023)

The implications from these studies highlight that while AI models like GPT-4 can assist in assessing writing accuracy and generating human-like language, they still require significant improvements in grammar detection, pragmatic interpretation, and ensuring factual accuracy. ChatGPT can generate coherent and well-structured text, but its outputs often contain inaccuracies, fictitious references, and hallucinations, necessitating cautious use and thorough human verification. The use of ChatGPT in scientific writing can potentially spread misinformation if not carefully checked by experts, indicating that AI-generated content should not replace human expertise. For head and neck scientific writing, the high rate of incorrect references generated by ChatGPT emphasizes the need for meticulous cross-

verification to avoid disseminating false information. Overall, these studies suggest that while AI tools can offer valuable assistance, they cannot yet replace the nuanced understanding and reliability provided by human researchers and must be used with caution and proper oversight.

Performance and Impact: Proficiency and Originality

Purpose

There is one study (Lozic & Štular, 2023) that investigates this issue of proficiency of ChatGPT and other AI tools for scientific writing.

Table 37

Proficiency and Originality: Purpose

Title	Purpose	Research Question	Author
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities	“With this in mind, the aim of this article was to design and conduct a test of AI chatbots’ abilities in scientific writing in the humanities. First, we were interested in their ability to generate correct answers to complex scientific questions. Second, we tested their capacity to generate original scientific contributions in humanities research.”	Not stated.	Lozic and Štular (2023)

Lozić and Štular (2023) aim to design and conduct a test to evaluate AI chatbots' proficiency and originality in scientific writing within the humanities. The study focuses on two main objectives: assessing the chatbots' ability to generate accurate answers to complex scientific questions and evaluating their capacity to produce original scientific contributions in humanities research.

Results

The results from this study are shown in Table 38.

Table 38*Proficiency and Originality: Results*

Title	Results	Author
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities	"In the quantitative test, ChatGPT-4 scored near the passing grade (−5), whereas ChatGPT-3.5 (−18), Bing (−21), and Bard (−31) were not far behind. Claude 2 (−75) and Aria (−80) scored much lower. In the qualitative test, all AI chatbots, but especially ChatGPT-4, demonstrated proficiency in recombining existing knowledge, but all failed to generate original scientific content. As a side note, our results suggest that with ChatGPT-4, the size of large language models has reached a plateau. Furthermore, this paper underscores the intricate and recursive nature of human research. This process of transforming raw data into refined knowledge is computationally irreducible, highlighting the challenges AI chatbots face in emulating human originality in scientific writing. Our results apply to the state of affairs in the third quarter of 2023. In conclusion, while large language models have revolutionised content generation, their ability to produce original scientific contributions in the humanities remains limited."	Lozic and Štular (2023)

The study (Lozic & Štular, 2023) finds that while ChatGPT-4 and other AI chatbots demonstrate proficiency in recombining existing knowledge, they fail to generate original scientific content, and this shows their ability to produce truly innovative research in the humanities is limited.

Implications

Table 39 provides the implications of this study.

Table 39*Proficiency and Originality: Implications*

Title	Implications	Author
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities	"Additional shortcomings of the AI-generated content that we found in our test include reasoning errors, hallucinations, and biases. Reasoning errors refer to the chatbot's inability to critically evaluate and link cause-and-effect relationships, as evidenced by several historical inaccuracies regarding the migration patterns and historical periods of the Slavs. Hallucinations denote confident but unsubstantiated claims by the AI, such as invented references and inaccurate dates. Our test also reveals significant biases in the content generated by the AI. These biases manifest as language bias, favouring English sources over other relevant languages; neo-colonial bias, displaying a preference for Western authors; and citation bias, skewing towards older and more highly cited	Lozic and Štular (2023)

Title	Implications	Author
	publications. These findings highlight that, despite their technological prowess, AI chatbots remain reliant on their training data, echoing and even amplifying existing biases and knowledge gaps. Given that they veer towards past data, including, in our test, a book originally published in 1895, they are likely to be too conservative in their recommendations. Since the listed deficiencies are relatively inconspicuous compared to, for example, gender or racial biases, it is unlikely that they will be remedied in the foreseeable future by resourceintensive reinforcement learning from human feedback. These biases are among the key concerns with the use of AI chatbots in scientific writing, as most of them are less likely to be highlighted in the review processes.”	

The study (Lozic & Štular, 2023) implies that despite their advanced capabilities, AI chatbots like ChatGPT exhibit significant reasoning errors, hallucinations, and biases, and their reliance on training data and amplifying existing knowledge gaps, which presents a major concern for their use in scientific writing is a serious issue.

Performance and Impact: Reliability and Quality

Purpose

There is one study (Dağcı et al., 2023) that investigates the reliability and quality of ChatGPT-generated texts.

Table 40

Reliability and Quality: Purpose

Title	Purpose	Research Question	Author
Reliability and Quality of the Nursing Care Planning Texts Generated by ChatGPT	“to examine the quality, authenticity, and reliability of the nursing care planning texts produced using ChatGPT”	“What is the quality subscale score of texts about nursing diagnosis generated by ChatGPT? • What is the reliability subscale score of texts about nursing diagnosis generated by ChatGPT? • What is the similarity rate of texts about nursing diagnosis generated by ChatGPT? • Is there a relationship between independent research variables (number of references, word count, similarity rate, etc) and the quality and reliability subscale scores of	Dağcı et al. (2023)

Title	Purpose	Research Question	Author
		texts about nursing diagnosis generated by ChatGPT?"	

Results

The results from this study are shown in Table 41.

Table 41

Reliability and Quality: Results

Title	Results	Author
Reliability and Quality of the Nursing Care Planning Texts Generated by ChatGPT	“The texts were evaluated by using descriptive criteria form and DISCERN tool to evaluate health information. The texts created without specifying the length were short, with a mean word count of 366.95 ± 57.98 (266- 506). The accessibility and accuracy of the references were checked. The majority were identified to be fictional and not found on either PubMed or Google. The similarity rate of the texts was low when checked through iThenticate and Turnitin; however, the Turnitin, which integrates AI detection technology, confirmed that all the texts were 100% written by AI (Table 1). The quality and content of the texts were evaluated using DISCERN. Their reliability, quality of information on nursing care, and overall quality were moderate (Table 2). In addition, 100% of the 40 ChatGPT texts had moderate reliability, with 39 (97.5%) providing moderate quality of information on nursing care. Regarding the overall quality of the texts, only 2 (5%) had good quality, 30 (75%) had moderate quality, and 8 (20%) had poor quality (Table 3). A statistically significant relationship was found between the number of accessible references and the reliability ($r = 0.408$) and quality ($r = 0.379$) of the texts ($P < .05$). Furthermore, a statistically significant relationship was found between the accessibility of web preferences and the reliability of the texts ($r = 0.351$; $P < .05$) (Table 4).”	Dağcı et al. (2023)

The study (Dağcı et al., 2023) finds that while the ChatGPT-generated texts have moderate quality and reliability, most references are fictional and not found on PubMed or Google, and all texts are confirmed to be 100% written by AI, with a significant relationship between the number of accessible references and the texts' reliability and quality.

Implications

Table 42 provides the implications.

Table 42*Reliability and Quality: Implications*

Title	Implications	Author
Reliability and Quality of the Nursing Care Planning Texts Generated by ChatGPT	“It is important to carefully evaluate the information provided by ChatGPT before using it in the treatment and care of patients by health care professionals. These results demonstrate the significant potential of ChatGPT for knowledge generation, text generation, and care planning in nursing. However, caution must be exercised, particularly in the clinical use of ChatGPT-generated texts. Nurses should thoroughly evaluate the accuracy and feasibility of these texts before incorporating them into patient treatment and care. Patient safety should always be paramount, and it should be remembered that such computer programs have limitations. We suggest that future studies be conducted that would compare texts generated by different AI algorithms and evaluate those using different tools to provide more comprehensive results.”	Dağcı et al. (2023)

The main implication from this study (Dağcı et al., 2023) is that while ChatGPT shows significant potential for generating knowledge and text for nursing care, healthcare professionals must carefully evaluate its accuracy and feasibility to ensure patient safety, acknowledging the limitations of AI-generated content.

Performance and Impact: Usefulness for Scholars***Purpose***

There are two studies one of which focuses on the publication challenges and the usefulness of AI-Based Writing Technologies and the other on the use of an artificial intelligence (AI) chatbot in academic writing.

Table 43*Usefulness for Scholars: Purpose*

Title	Purpose	Research Question	Author
Using Nominal Group Technique to Explore Publication Challenges and the Usefulness of AI-Based Writing Technologies: Insights From Indonesian Scholars	“to elicit the obstacles encountered by Indonesian scholars in gaining international publication, and the extent to which the advancement of AI-based writing technologies can provide solution”	Not stated.	Hamamah et al. (2023)

Title	Purpose	Research Question	Author
Artificial intelligence in scientific writing: a friend or a foe?	“to discuss the use of this artificial intelligence (AI) chatbot in academic writing by presenting one simplified example”	Not stated.	Altmäe et al. (2023)

Results

The results from these studies can be found in Table 44.

Table 44

Usefulness for Scholars: Results

Title	Results	Author
Using Nominal Group Technique to Explore Publication Challenges and the Usefulness of AI-Based Writing Technologies: Insights From Indonesian Scholars	“Another significant obstacle that academics shared in the NGT discussion, which was also anticipated, is the nonexistence portion of time to improve their English skills due to the teaching responsibilities and the lack of encouragement from their institutions, making the process of learning English even more challenging and labor-intensive with little financial prospects. Scholars eventually sought out fast and cost-effective tools to assist them in writing research papers in English. They turned to digital, AI-based writing and typing tools for this purpose. The participants acknowledged that using Google Translate was the most practical solution for not only translating their research from Indonesian to English, but also for translating papers in the opposite direction. This helped minimize the difficulty of synthesizing research from international, English-language journals. The second most popular program used by scholars for word choice, collocation, and conjugation exploration was WordHippo.”	Hamamah et al. (2023)
Artificial intelligence in scientific writing: a friend or a foe?	“Overall, the paper appears to be well-structured and clearly written, with appropriate citation of relevant literature. However, it's important to note that the decision to submit a paper to a specific journal depends on many factors, including the journal's scope, editorial policies, and the perceived fit of the paper with the journal's readership. It's always a good idea to carefully read the journal's guidelines and review previously published papers in the journal to ensure that the paper is a good fit before submitting it for consideration.”	Altmäe et al. (2023)

The results of the two studies highlight different aspects of AI's role in scientific writing. The first study (Hamamah et al., 2023) found that Indonesian scholars face significant obstacles in improving their English skills due to teaching responsibilities and lack of institutional support, leading them to rely on AI-based

tools like Google Translate and WordHippo to assist with writing and translating research papers. In contrast, the second study (Altmäe et al., 2023) indicates that while AI can help create well-structured and clearly written papers with appropriate citations, successful journal submission also depends on understanding the journal's scope, editorial policies, and previously published papers, suggesting a more complex role of AI in ensuring the paper's fit for publication.

Implications

Table 45 provides the implications.

Table 45

Usefulness for Scholars: Implications

Title	Implications	Author
Using Nominal Group Technique to Explore Publication Challenges and the Usefulness of AI-Based Writing Technologies: Insights from Indonesian Scholars	“The results of this study can inform policies and strategies aimed at supporting Indonesian academics in their efforts to publish in internationally recognized journals, and contribute to the global discourse on inclusivity and fairness in academic publishing. Further research and initiatives are needed to address the challenges faced by NNEs scholars and promote equitable participation in the global scientific community.”	Hamamah et al. (2023)
Artificial intelligence in scientific writing: a friend or a foe?	“When using ChatGPT in scientific writing, we want to highlight that if papers are not properly revised, there is a high risk of presenting incorrect information and non-existent references, especially among writers without domain expertise in the topic. Furthermore, once a paper has been submitted to a journal, the editors and referees will not have the time to check its accuracy phrase by phrase, reference. We sincerely hope that ChatGPT (and other similar LLMs) will not turn out to be a ‘tool of misinformation’. We consider the inaccuracy and misinformation to be the biggest concern today in using ChatGPT. When using the ChatGPT, it creates surprisingly intelligent sounding text. We would like to stress the word ‘sounding’ here as, based on our own (and others) experience, the text at first seems acceptable but, on closer inspection, can be seen to contain misinformation and inaccuracies. In short, we would like authors to be aware that it is an excellent tool to help but not to replace their work.”	Altmäe et al. (2023)

The implications of the two studies underscore distinct considerations for AI's role in academic writing. The first study (Hamamah et al., 2023) suggests that

policies and strategies should be developed to support Indonesian academics in publishing internationally, addressing challenges faced by non-native English-speaking (NNES) scholars, and promoting inclusivity and fairness in academic publishing. On the other hand, the second study by Altmäe et al., 2023 warns that improper use of AI tools like ChatGPT in scientific writing can lead to the dissemination of incorrect information and non-existent references, emphasizing the importance of careful revision and domain expertise to avoid misinformation, highlighting that AI should assist but not replace human authorship.

Performance: Impact Efficacy

Purpose

There is one study that examines the efficacy of an AI writing tool.

Table 46

Impact Efficacy: Purpose

Title	Purpose	Research Question	Author
Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial	“to examine the efficacy of a group format of an Artificial Intelligence (AI) powered writing tool for English second postgraduate students in the English academic writing context”	Not Stated.	Nazari et al. (2021)

This research aims to evaluate the effectiveness of an Artificial Intelligence (AI) powered writing tool when used in a group setting by postgraduate students who are non-native English speakers. The study (Nazari et al.,2021) focuses on assessing how this AI writing assistant impacts the students' performance in English academic writing contexts within higher education.

Results

The results from this study are shown in Table 47.

Table 47*Impact Efficacy: Results*

Title	Results	Author
Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial	“The results of the parametric test of analyzing of covariance revealed that at post-intervention, students who participated in the AI intervention group demonstrated statistically significant improvement in the scores, of the behavioral engagement, of the emotional engagement, of the cognitive engagement, of the self-efficacy for writing, of the positive emotions, and of the negative emotions, compared with NEAI. students. The results indicated the AI effectiveness in improving self-efficacy, engagement, and academic emotion at post-intervention compared with NEAI. Grammarly can improve students' EFL writing skill. The results suggest that AI-powered writing tools could be an efficient tool to promote learning behavior and attitudinal technology acceptance through formative feedback and assessment for non-native postgraduate students in English academic writing.”	Nazari et al. (2021)

The study (Nazari et al.,2021) reveals that students in the AI intervention group show statistically significant improvements in behavioral, emotional, and cognitive engagement, self-efficacy for writing, and both positive and negative emotions compared to those in the non-AI group, and this means that AI-powered writing tools like Grammarly can effectively enhance EFL writing skills and promote learning behavior and attitudinal technology acceptance through formative feedback and assessment.

Implications

Table 48 provides the implications of this study.

Table 48*Impact Efficacy: Implications*

Title	Implications	Author
Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial	“Overall, our findings the utility of applying an AI-powered writing tool to improve self-efficacy, engagement and emotion, emotions in the EFL context. Results suggest that Grammarly may affect cognitive, non-cognitive, and emotional domains of learning. The results indicate that AI application could be an efficient assist-writing tool in English academic writing for non-native postgraduate students.”	Nazari et al. (2021)

The study (Nazari et al., 2021) finds that using an AI-powered writing tool like Grammarly significantly improved self-efficacy, engagement, and emotional aspects of learning in the EFL context, and this shows its effectiveness as an assistive tool for English academic writing among non-native postgraduate students.

Detection

Purpose

Ten studies out of 103 focus on the detection of AI-generated content.

Table 49

Detection: Purpose

Title	Purpose	Research Question	Author
Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines”	“to (a) assess the effectiveness of available AI-detection platforms to help L2 educators identify reliable platforms they can use to control AI-assisted plagiarism; (b) offer an in-depth investigation of this new, disruptive, and unexplored research territory to guide educators in dealing with AI-assisted plagiarism; and (c) evaluate underlying classifier type, design, and training mechanisms to contribute to the development of more robust machine-detection classifiers”	“(a) How effective is GPT-2 Output Detector Demo in detecting machine-generated essays? (b) How effective is Crossplag AI Content Detector in identifying machine-generated essays? (c) What is the difference in classification effectiveness between GPT-2 Output Detector Demo and Crossplag Detector? (d) What does the comparison of two RoBERTa-based detectors suggest about the robustness of AI-detection of AI-generated texts?”	Ibrahim (2023)
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“to analyze the differences between medical texts written by human experts and those generated by ChatGPT and to design machine learning workflows to effectively detect and differentiate medical texts generated by ChatGPT”	“(1) What is the difference between medical content written by humans and that generated by ChatGPT? (2) Can we use machine learning methods to detect whether medical content is written by human experts or ChatGPT?”	Liao et al. (2023)
Evaluating the efficacy of AI content detection tools in differentiating	“to investigate the capabilities of various AI content detection tools in discerning human and AI-authored content”	Not stated.	Elkhatat et al. (2023)

Title	Purpose	Research Question	Author
between human and AI-generated text			
AI vs academia: Experimental study on AI text detectors' accuracy in behavioral health academic writing	“Two key aims were addressed: 1) determine the accuracy of AI detectors in distinguishing human versus AI generated content in real-world academic writing samples from the behavioral health domain, and 2) assess their effectiveness in recognizing texts produced by newer generative models not yet incorporated into detector training.”	Not stated.	Popkov and Barrett (2024)
Detecting AI-Generated Essays: The ChatGPT Challenge	“The greatest challenge, however, is the accuracy of detection. More importantly, genuinely human-generated essays should not be falsely classified as AI-generated for ethical, legal and professional reasons. This paper aims to address these challenges by proposing an AI-led detection method and a prediction (i.e. made by an ML classifier) performance measure.”	Not stated.	Cingillioglu (2023)
Human-Written vs AI-Generated Texts in Orthopedic Academic Literature: Comparative Qualitative Analysis	“The aim of this qualitative analysis is to determine the possibility that human researchers and AI-detection platforms can detect AI-generated texts.”	Not stated.	Hakam et al. (2024)
It takes one to know one—Machine learning for identifying OBGYN abstracts written by ChatGPT	“To use machine learning to optimize the detection of obstetrics and gynecology (OBGYN) Chat Generative Pre-Trained Transformer (ChatGPT) -written abstracts of all OBGYN journals.”	Not stated.	Levin et al. (2024)
Non-English academics face inequality via AI-generated essays and countermeasure tools	“to examine the detection limits of readily available countermeasure tools to detect human-written essays, ChatGPT-generated essays, and partially AI-aided essays by checking AI procedurally generated essays in English and three other languages (Malay, Mandarin, Japanese, and their AI-translated essays)”	Not stated.	Wee Hin and Reimer (2023)

Title	Purpose	Research Question	Author
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“to shed light on the challenges faced by English as a Second Language (ESL) lecturers in identifying AI-generated texts, and highlighting the skills and resources needed to enhance their detection capabilities”	“(1) How effective are current plagiarism detectors in providing a reliable measure of detection for identifying originality in academic writing and distinguishing human-generated text from AI-generated text? (2) What criteria are employed by ELS lecturers to assess academic writing? and (3) How effective are these criteria in distinguishing human-generated text from AIgenerated text?”	Alexander et al. (2023)
Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4	“to examine the stylistic features of case study, business correspondence, and academic writing ChatGPT-4-generated responses by exploring sentence length, paragraph structure, word choice, mood, tense, voice, pronouns, keywords density, lexical density, lexical diversity, and reading ease”	“1. Can Turnitin detect ChatGPT 4-generated responses? 2. Are Turnitin AI reports accurate? 3. What are the stylistic features of ChatGPT-4?”	Alafnan and Mohdzuki (2023)

Ibrahim (2023) aims to assess the effectiveness of AI-detection platforms in identifying AI-generated essays and to guide educators in dealing with AI-assisted plagiarism. Liao et al. (2023) seek to analyze differences between human-written and ChatGPT-generated medical texts and design machine learning workflows to detect AI-generated medical content. Elkhatat et al. (2023) investigate the capabilities of various AI content detection tools in distinguishing between human and AI-authored content. Popkov and Barrett (2024) aim to determine the accuracy of AI detectors in distinguishing human versus AI-generated content in behavioral health academic writing samples.

Cingillioglu (2023) proposes an AI-led detection method and a prediction performance measure to address challenges in detecting AI-generated essays. Hakam et al. (2024) seek to determine the ability of human researchers and AI detection

platforms to identify AI-generated texts in orthopedic academic literature. Levin et al. (2024) aim to use machine learning to optimize the detection of ChatGPT-written abstracts in obstetrics and gynecology journals. Wee Hin and Reimer (2023) examine the detection limits of readily available countermeasure tools for identifying human-written, ChatGPT-generated, and partially AI-aided essays in multiple languages.

Alexander et al. (2023) investigate the challenges faced by ESL lecturers in identifying AI-generated texts and highlight skills and resources needed to enhance detection capabilities. Alafnan and Mohdzuki (2023) examine the stylistic features of ChatGPT-4-generated responses in various writing genres and explore the accuracy of Turnitin in detecting AI-generated content.

Results

The results from these ten studies are shown in Table 50.

Table 50

Detection: Results

Title	Results	Author
Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines”	“Data analysis revealed that both platforms could identify AI-generated texts, but their detection accuracy was inconsistent across the dataset. The study’s findings revealed that GPT-2 Output Detector Demo and Crossplag AI Content Detector achieved high accuracy in discriminating between machine-generated and human-written texts with an average accuracy of 89% despite being trained on a smaller and earlier version of GPT. Also, the findings indicated that GPT-2 Output Detector Demo was more sensitive to human-written texts and could detect them more accurately. At the same time, Crossplag Detector was more sensitive to machine-generated texts and had fewer false negatives. These findings demonstrate that AI-based classifiers could offer viable resources for detecting machine-generated texts and controlling AI-assisted plagiarism, even if they need further development to achieve higher detection accuracy.”	Ibrahim (2023)
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“Medical texts written by humans were more concrete, more diverse, and typically contained more useful information, while medical texts generated by ChatGPT paid more attention to fluency and logic and usually expressed general terminologies rather than effective information specific to the context of the problem. A bidirectional encoder representation from transformers–	Liao et al. (2023)

Title	Results	Author
	based model effectively detected medical texts generated by ChatGPT, and the F 1 score exceeded 95%. Conclusions: Although text generated by ChatGPT is grammatically perfect and human-like, the linguistic characteristics of generated medical texts were different from those written by human experts. Medical text generated by ChatGPT could be effectively detected by the proposed machine learning algorithms. This study provides a pathway toward trustworthy and accountable use of large language models in medicine.”	
Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text	“Findings reveal that the AI detection tools were more accurate in identifying content generated by GPT 3.5 than GPT 4. However, when applied to human-written control responses, the tools exhibited inconsistencies, producing false positives and uncertain classifications. Looking at the GPT 3.5 results, the OpenAI Classifier displayed the highest sensitivity, with a score of 100%, implying that it correctly identified all AI-generated content. However, its specificity and NPV were the lowest, at 0%, indicating a limitation in correctly identifying human-generated content and giving pessimistic predictions when it was genuinely human-generated. GPTZero exhibited a balanced performance, with a sensitivity of 93% and specificity of 80%, while Writer and Copyleaks struggled with sensitivity. The results for GPT 4 were generally lower, with Copyleaks having the highest sensitivity, 93%, and CrossPlag maintaining 100% specificity. The OpenAI Classifier demonstrated substantial sensitivity and NPV but no specificity.”	Elkhatat et al. (2023)
AI vs academia: Experimental study on AI text detectors’ accuracy in behavioral health academic writing	“The free AI detector showed a median of 27.2% for the proportion of academic text identified as AI-generated, while commercial software Originality.AI demonstrated better performance but still had limitations, especially in detecting texts generated by Claude. The current findings explore the efficacy of two AI content detection tools, raising fundamental questions about the ability to detect AI-generated content in academic writing. The identified flaws of the tested AI detectors raise doubts about solely relying on these technologies to enforce strict policies around AI text generation in behavioral health publications, risking unfair penalties for scholars or failing to detect AI content. Our study suggests fieldspecific validation of AI detection tools before their implementation in academic settings. More research is needed on validating improved AI detection methods and exploring the ethical implications and practical aspects of implementing AI policies in academic writing. This includes assessing AI detection across different academic fields, comparing the linguistic features that produce higher accuracy, understanding the impact of AI detection on editorial decisions, and evaluating the ongoing evolution of AI technologies. Such comprehensive research is essential before AI detection tools can be considered fully reliable for policy enforcement.”	Popkov and Barrett (2024)
Detecting AI-Generated Essays:	“Via support vector machine (SVM) algorithm 100% accuracy was recorded for identifying human generated	Cingillioglu (2023)

Title	Results	Author
The ChatGPT Challenge	essays. The author discussed the key use of Recall and F2 score for measuring classification performance and the importance of eliminating False Negatives and making sure that no actual human generated essays are incorrectly classified as AI generated. The results of the proposed model's classification algorithms were compared to those of AI-generated text detection software developed by OpenAI, GPTZero and Copyleaks. / we described a language model to meet the challenge of achieving high accuracy in distinguishing ChatGPT-generated essays from human written essays with a particular focus on not falsely classifying genuinely humanwritten essays as AI-generated. We developed an n-gram BOWsdiscrepancy language model which we then used as input to a ML classifier which we trained to predict whether an essay was human or ChatGPT (AI)-generated. We discussed the importance of F2 score for measuring classification performance and totally eliminating False Negatives. We suggest using SVM algorithm upon the proposed model as we recorded 100% accuracy for identifying human generated essays and 92.7% overall accuracy. Although we attained higher overall accuracies (e.g., 94%) via other algorithms and other AI-generated text detection software introduced by OpenAI, GPTZero and Copyleaks yielded superior overall accuracy performances (95.3–96.7%), only SVM provided the perfect Recall with the highest F2 score (97.2%) and no False Negatives ensuring all genuinely human written essays were correctly classified as human generated. Whether a fad or a tech revolution, with its sophisticated AI-led natural language text generation capability, ChatGPT has already started to raise serious concerns over student cheating and as it keeps evolving, it will indisputably keep changing the way we hunt down and defeat academic deceit.”	
Human-Written vs AI-Generated Texts in Orthopedic Academic Literature: Comparative Qualitative Analysis	“Neither the researchers nor the AI-detection software could successfully identify the AI-generated texts. Furthermore, the criteria previously suggested in the literature did not correlate with whether the researchers deemed a text to be AI generated or whether they judged the article correctly based on these parameters. /The primary results of the study indicate that neither AI-detection software nor human critical appraisal can reliably distinguish AI-generated texts from human-written work. Regarding human detection of AI-generated texts, neither clinical experience nor area of expertise played a role in the evaluation of the presented material. The secondary results of the study indicate that criteria suggested by prior research, such as originality, style, and nuance, did not correlate with whether the researchers identified a text correctly or not. Furthermore, none of the criteria correlated with whether researchers judged a text as human written or AI generated. The qualitative analysis of the written answers did not provide any new insights on the subject in question.”	Hakam et al. (2024)
It takes one to know one—Machine	“Overall, 75 abstracts from 12 different OBGYN journals were randomly selected. There were seven (58%) Q1	Levin et al. (2024)

Title	Results	Author
learning for identifying OBGYN abstracts written by ChatGPT	journals, one (8%) Q2 journal, two (17%) Q3 journals, and two (17%) Q4 journals. Use of mixed dialects of English, absence of comma-misuse, absence of incorrect verb forms, and improper formatting were important prediction variables of ChatGPT-written abstracts. The deep-learning model had the highest predictive performance of all examined models. This model achieved the following performance: accuracy 0.90, precision 0.92, recall 0.85, area under the curve 0.95. Machine-learning- based tools reach high accuracy in identifying ChatGPT-written OBGYN abstracts.”	
Non-English academics face inequality via AI-generated essays and countermeasure tools	“AI modules such as ChatGPT are powerful tools for helping to streamline and increase the efficiency of human knowledge acquisition and dissemination. However, AI also poses a threat because its overuse raises authenticity issues, perhaps even of science itself. With the “emergence of countermeasure detection tools, it is hoped the system has checks and balances. Nevertheless, our study shows that the current tools are unable to reliably detect authenticity for non-English writing, indicating a potentially large issue for academia. Furthermore, the use of AI-aided translation and paraphrasing tools, which, on the surface, could help increase one’s accuracy in English writing, resulted in AI countermeasure tools flagging works as AI-aided essays, putting non-English academics at a disadvantage. Questions on the ethical limits of AI use in assisting scientific writing need to be collectively addressed.”	Wee Hin and Reimer (2023)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“Findings reveal that participants tended to exploit a deficit model of assessment that focuses on error as an indicator of L2 writing output, with high levels of technical and grammatical accuracy and sophisticated language use as indicators of AI-generated text. Moreover, findings suggest that limited awareness of the characteristics and metrics used by ChatGPT, as well as lack of attention to the veracity of facts and references generated by ChatGPT, were features of participants' evaluations. This study identifies the growing challenges encountered by ESL lecturers and underlines the need for digital literacy training, targeted professional development, the use of advanced detection tools and a review of assessment policies and practices in relation to AI. Additionally, this study highlights the importance of reviewing and reinforcing institutional policies and practices that safeguard academic integrity and ensure quality higher education.”	Alexander et al. (2023)
Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4	“The study revealed that ChatGPT-4 case study-generated responses are produced in paragraphs of 2 to 3 sentences of 16 to 18 words each. The sentences are mainly formed in the imperative mood. The use of the second-person pronoun “you” and the second-person possessive determiner “your” is prevalent. Keywords and lexical density are relatively low, lexical diversity is average, and the reading ease is relatively high. The study also found that ChatGPT-4 business correspondence responses are generated in paragraphs of 2 to 3 sentences of 16 to 20	Alafnan and Mohdzuki (2023)

Title	Results	Author
	words each. The sentences are mainly generated in declarative mood thru simple present tense in active voice using third-person singular pronouns. Technical words and abbreviations are used without outlining what they stand for. The keywords density, lexical density, and lexical diversity are high, and the reading ease is low. The study also revealed that ChatGPT-4 academic writing-generated responses are provided in paragraphs of 3 to 4 sentences of 16 to 19 words each. The sentences are mainly generated in declarative mood using active voice, agentless passive in times, with diverse present tenses. Keywords and lexical densities are high, and the lexical diversity is low, which makes the reading ease average difficulty, except for the undefined abbreviations. Noticeably, ChatGPT-4 supports the transgender movement by intentionally using the third-person plural pronoun “they” to refer to a singular.”	

The first study (Ibrahim, 2023) on AI-assisted plagiarism in ESL writing found that while both GPT-2 Output Detector Demo and Crossplag AI Content Detector could identify AI-generated texts, their detection accuracy varied, with GPT-2 being more sensitive to human-written texts and Crossplag to machine-generated texts. The second study (Liao et al., 2023) differentiating ChatGPT-generated and human-written medical texts revealed that ChatGPT's texts focused more on fluency and logic but were less context-specific, and the proposed machine learning model effectively detected these texts with an F1 score exceeding 95%. According to the third study (Elkhatat et al., 2023), the evaluation of AI content detection tools showed that while the tools were more accurate with GPT-3.5 generated content, they exhibited inconsistencies with human-written texts, with GPTZero showing the best balanced performance. The fourth study (Popkov & Barrett, 2024) on AI text detectors in behavioral health academic writing highlighted that both free and commercial detectors had limitations, particularly with newer generative models, indicating the need for field-specific validation of these tools.

Results from fifth study (Cingillioglu, 2023) revealed that the ChatGPT challenge study achieved 100% accuracy in identifying human-generated essays

using a support vector machine algorithm, emphasizing the importance of eliminating false negatives and ensuring genuinely human essays are not misclassified. The sixth study (Hakam et al., 2024) on orthopedic academic literature found that neither researchers nor AI-detection software could reliably distinguish AI-generated texts from human-written ones, indicating the limitations of both human and AI detection methods. The seventh study (Levin et al., 2024) on identifying OBGYN abstracts written by ChatGPT showed that a deep-learning model achieved high predictive performance with accuracy, precision, recall, and AUC scores of 0.90, 0.92, 0.85, and 0.95, respectively.

The eighth study (Wee Hin & Reimer, 2023) on non-English academics highlighted that current AI countermeasure tools struggle to detect authenticity in non-English writing, putting non-English academics at a disadvantage and raising ethical concerns about AI use in scientific writing. The ninth study (Alexander et al., 2023) on detecting AI-generated writing in ESL education revealed that educators often misidentify AI-generated texts due to limited awareness of AI characteristics, underscoring the need for digital literacy training and updated assessment policies. The last study (Alafnan & Mohdzuki, 2023) on the stylistic features of ChatGPT-4 found distinct patterns in different types of writing, such as using the imperative mood and second-person pronouns in case studies and high keyword density with low reading ease in business correspondence, revealing ChatGPT-4's varied writing styles.

Implications

Table 51 provides the implications of the studies focusing on detection.

Table 51*Detection: Implications*

Title	Implications	Author
Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines”	“The present study has several practical implications. First, as GPT Output and Crossplag detectors could identify AI texts with a relatively high degree of accuracy, L2 educators can rely on them to investigate potential cases of AI-assisted plagiarism. To overcome the limitations of each detector, it might be helpful to run a flagged text in one detector and then into another to double-check the results or upload the text to GLTR to get a visual report that can support initial results from an AI-based detector. Second, since available classifiers can misevaluate human-written texts as machine-generated (i.e., false positives), L2 educators should not base their decisions about academic integrity solely on detectors’ analysis and use other traditional approaches to ensure the academic integrity of students’ work. For instance, educators can interview students about the content of their work or analyze the linguistic complexity of flagged work compared to students’ in-class work. Third, since fine-tuned classifiers, especially RoBERTa-based classifiers appear to be relatively effective in detecting AI-assisted plagiarism, major plagiarism detection platforms should integrate such classifiers into their systems rather than train their baseline classifiers from scratch. Fourth, as the findings demonstrate that model size substantially impacts the detection accuracy of classifiers, plagiarism detection platforms should train their detection classifiers on datasets from larger language models to improve their detection accuracy.”	Ibrahim (2023)
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“In general, for artificial intelligence (AI) to realize its full potential in medicine, we should not rush into its implementation but advocate for its careful introduction and open debate about its risks and benefits. First, human medical writers will be responsible for ensuring the accuracy and completeness of the information communicated and for complying with ethical and regulatory guidelines. However, ChatGPT cannot be held responsible. Second, training an LLM requires a huge amount of data, but the quality of the data is difficult to guarantee, so the trained ChatGPT is biased. For example, ChatGPT can provide biased output and perpetuate sexist stereotypes [52]. Third, use of ChatGPT may lead to private information leakage. This may be because the LLM remembers personal privacy information in the training set [53]. What is more, the legal framework must be considered. Who shall be held accountable when an AI doctor makes an inevitable mistake? ChatGPT cannot be held accountable for its work, and there is no legal framework to determine who owns the rights to AI-generated work [15].”	Liao et al. (2023)
Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text	“This study underscores the need for further as AI text generation evolves, so must the tools designed to detect it. This necessitates continuous development and regular evaluation to ensure their efficacy and reliability. Furthermore, a balanced approach involving AI tools and traditional methods best upholds academic integrity in an ever-evolving digital landscape.”	Elkhatat et al. (2023)

AI vs academia: Experimental study on AI text detectors' accuracy in behavioral health academic writing	“Our study indicates that even tools with high performance in detecting true negatives, such as Originality.AI, should be carefully considered before implementation in editorial decision-making. Even a small percentage of incorrectly identified AI content could potentially undermine the perceived integrity of genuine academic work written solely by human authors. The small number of false positives for Originality.AI highlights that lower performance of the tool can be only identified on a larger sample set. Therefore, it is crucial to determine what level of accuracy makes these tools reasonable and to understand how they can be supported by human judgment to prevent errors and false decisions about the integrity of academic work.”	Popkov and Barrett (2024)
Detecting AI- Generated Essays: The ChatGPT Challenge	“AI-generated essays submitted by students can be detected by teachers and educational designers using the proposed language model and machine learning (ML) classifier at a high accuracy. Human (student)-generated essays can and must be correctly identified with 100% accuracy even if the overall classification accuracy performance is slightly reduced. This is the first and only study that used an n-gram bag-of-words (BOWs) discrepancy language model as input for a classifier to make such prediction and compared the classification results of other AI-generated text detection software in an empirical way.”	Cingillioglu (2023)
Human-Written vs AI- Generated Texts in Orthopedic Academic Literature: Comparative Qualitative Analysis	“The statistical and qualitative analysis of the presented material showed that researchers were unable to differentiate human-written from AI-generated texts. Furthermore, the secondary finding of this study was that previously suggested criteria, such as originality and comprehension, did not aid in the differentiation of human-written and LLM-generated texts. Both findings show that humans and AI-detection software currently fail to properly identify the use of LLMs in the academic literature. Furthermore, one can only speculate about the amount of undisclosed AI use in the academic literature. However, with the ever-increasing sophistication of LLMs, the integrity of future projects will be entirely dependent on scientists' attitudes, as AI can serve as a facilitator and accelerator in publishing but can also be used with malicious intent.”	Hakam et al. (2024)
It takes one to know one—Machine learning for identifying OBGYN abstracts written by ChatGPT	“Therefore, it is important to study means for identifying ChatGPT-generated abstracts. We here introduce a validated, very efficient deep-learning algorithm for identifying these abstracts. Moreover, as the tools used here are free, it is possible that future data will accumulate to further improve the prediction model. By publicly publishing this model, more relevant data can be analyzed, and models can be optimized to fit many other journals from other specialties and help with the task of identifying ChatGPT-written scientific abstracts in an efficient and easy manner.”	Levin et al. (2024)
Non-English academics face inequality via AI- generated essays and countermeasure tools	“The AI-aided para- phrased GPT-generated essays had lowered authenticity results. Therefore, users should be aware of the risks of relying on AI-aided paraphrasing tools, particularly if countermeasure tools are incorporated into plagiarism-checking tools. The recent increase in adoption of AI-aided translation and paraphrasing tools in writing could	Wee Hin and Reimer (2023)

	lead to many potential problems in a very near future if blanket GPT countermeasure tools are adopted, posing ethical concerns regarding the degree of reliance on AI tools in writing. The academic world needs to collectively discuss in detail the degrees of reasonable involvement of AI tools in writing. In this way, common standards can be incorporated into plagiarism tools, which play an important role in protecting the integrity of academia. Furthermore, aside from detection issues, another ethical issue in writing is information accuracy, especially regarding misinformation. The recent culture of open information has been often corrupted and overshadowed by false information. Since ChatGPT and perhaps future competitors use the open Internet in training, we should not limit our discussion on the limitations of using AI in writing but should also discuss using AI to detect misinformation.”	
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“Concerning the implications for policy and practice, the results demonstrate that to improve practice and to safeguard academic integrity, there is a need for digital training for educators with respect to the use of AI tools in the ESL classroom and the use of advanced detection tools. There is also a need for revision of assessment policies and procedures and the development of suitable rubrics and assessment criteria. Lastly, lecturers’ tendency to confuse student-written essays for AI-generated text points out the need for lecturers to be trained on the features of AI-created texts and the differences between them and humanwritten essays.”	Alexander et al. (2023)
Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4	“This study found that even though Turnitin’s AI report can detect AI-generated responses, their accuracy cannot be taken into consideration for punitive purposes for a couple of reason. Firstly, Turnitin mentions that these AI reports shall not be used for misconduct purposes. Secondly, after testing the accuracy of these reports, this study found an inconsistency in their outcome. The researchers gradually added human-generated material to the AI (ChatGPT-4)-generated material but the AI report remained 100%. This proves Turnitin’s claim that these reports shall not be used for misconduct purposes as their accuracy is way below 98%, the percentage claimed by Turnitin, accurate.”	Alafnan and Mohdzuki (2023)

Ibrahim (2023) suggests several practical implications for detecting AI-assisted plagiarism in ESL writing. The study recommends using multiple AI detectors to cross-check results and improve accuracy. However, it cautions against relying solely on these detectors, advising educators to use traditional approaches like student interviews to ensure academic integrity. The research also suggests that plagiarism detection platforms should integrate fine-tuned classifiers, especially RoBERTa-based ones, and train their detection classifiers on datasets from larger

language models to improve detection accuracy. Liao et al. (2023) highlight several important implications for the use of AI in medical writing. They emphasize the need for careful implementation of AI in medicine, advocating for open debate about its risks and benefits. The study points out that human medical writers remain responsible for ensuring accuracy and ethical compliance, as ChatGPT cannot be held accountable. It also raises concerns about potential biases in AI-generated content due to training data quality issues. Additionally, the research highlights the risk of privacy breaches and the lack of a legal framework for determining accountability and ownership of AI-generated work in medical contexts.

Elkhatat et al. (2023) underscore the need for continuous development and regular evaluation of AI content detection tools. As AI text generation evolves, they argue that detection tools must keep pace to ensure their efficacy and reliability. The study advocates for a balanced approach that combines AI tools with traditional methods to uphold academic integrity in the digital age. Popkov and Barrett (2024) caution against the uncritical implementation of AI text detectors in editorial decision-making, even for tools with high performance. They highlight the potential risks of false positives, which could undermine the perceived integrity of genuine academic work. The study emphasizes the importance of determining acceptable accuracy levels for these tools and suggests that human judgment should support AI detection to prevent errors in assessing academic integrity.

Cingillioglu (2023) presents a novel approach to detecting AI-generated essays using an n-gram bag-of-words discrepancy language model. The study emphasizes the importance of correctly identifying human-generated essays with 100% accuracy, even if it slightly reduces overall classification accuracy. This approach offers a new tool for educators and educational designers to detect AI-

generated content while prioritizing the accurate identification of genuine student work. Hakam et al. (2024) reveal concerning implications for the detection of AI-generated texts in academic literature. Their findings show that both human researchers and AI-detection software currently fail to properly identify the use of Large Language Models (LLMs) in academic writing. This raises questions about the potential prevalence of undisclosed AI use in academic literature and emphasizes that the integrity of future academic work will heavily depend on scientists' ethical attitudes toward AI use.

Levin et al. (2024) introduce a validated deep-learning algorithm for identifying ChatGPT-generated abstracts in OBGYN journals. The study highlights the importance of developing such tools to maintain academic integrity. By making their model publicly available, the authors aim to encourage further data accumulation and model optimization across various medical specialties, potentially improving the detection of AI-generated scientific abstracts. Wee Hin and Reimer (2023) highlight the risks associated with AI-aided paraphrasing tools, particularly for non-English academics. They call for collective discussion within academia to establish common standards for the reasonable involvement of AI tools in writing. The study also broadens the scope of concern beyond detection issues to include the accuracy of information and the potential for misinformation in AI-generated content.

Alexander et al. (2023) emphasize the need for digital training for educators in using AI tools and advanced detection methods in ESL classrooms. They call for revisions to assessment policies and procedures, including the development of new rubrics and assessment criteria. The study also highlights the importance of training lecturers to distinguish between AI-created and human-written texts, addressing the

tendency to confuse student-written essays with AI-generated content. Alafnan and Mohdzuki (2023) caution against using Turnitin's AI reports for punitive purposes due to inconsistencies in their accuracy. The study demonstrates that even when human-generated content is gradually added to AI-generated material, Turnitin's AI report remains at 100%, contradicting the company's claimed accuracy rate. This finding supports Turnitin's own disclaimer that these reports should not be used for misconduct purposes.

Identification

Purpose

Out of 103 studies, seven aim at investigating the identification of AI-generated content by humans.

Table 52

Identification: Purpose

Title	Purpose	Research Question	Author
Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse	“to explore the capability of academic staff assisted by the Turnitin Artificial Intelligence (AI) detection tool to identify the use of AI-generated content in university assessments”	Not stated.	Perkins et al. (2023)
Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students	“to test the performance of high school teachers and students on an AI Identification Test that assessed their ability to distinguish between short English essays written by ChatGPT versus high school students / to examine possible predictors of the ability to make this distinction, such as prior experience with ChatGPT / to measure their broader attitudes about the use of ChatGPT in academic settings, including their evaluations of its	Not stated.	Waltzer et al. (2023)

Title	Purpose	Research Question	Author
	impacts on education and what uses should be considered acceptable by using an AI Attitude Assessment”		
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“to present a comparative study, introducing a novel dataset of human-written and LLM-generated texts in different genres: essays, stories, poetry and Python code”	Not stated.	Hayawi et al. (2023)
Clinical Research With Large Language Models Generated Writing-Clinical Research with AI-assisted Writing (CRAW) Study	“Our primary aim was to investigate whether a representative sample of medical researchers can accurately differentiate between a LLMs generated (GPT-3.5) background and one generated by two expert medical researchers. Additionally, we aimed to compare the quality of the LLMs generated background with that produced by the expert researchers.”	Not stated.	Huespe et al. (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“to investigate how successful advanced students are in determining whether an essay was written by AI tools or by a human and how much AI assistance they need to summarize, generate text and write from prompts when trained to use an AI-assisted word processor”	“1. How well do advanced students distinguish human-generated from AI-generated essays? What criteria do they think are decisive about Artificial Intelligence authorship? 2. When trained in the use of an AI-assisted word processor, how do advanced students make use of summarising, generating text and writing from prompts? 3. How do students’ essays differ linguistically after training in the use of AI tools from those written before treatment? Do they show differences in burstiness and perplexity?”	Krajka and Olszak (2024)
Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry	“to assess behavioral reactions to the state-of-the-art Natural Language Generation algorithm GPT-2”	“1. Do people prefer creative text written by fellow humans to that generated by algorithms? 2. Does the information about the respective origin – being a human or an algorithm – sway this preference for the creative-text output? 3. Does human involvement (or lack of) in the selection of the text output generated by the algorithm make a difference?”	Köbis and Mossink (2020)

Title	Purpose	Research Question	Author
Association of reviewer experience with discriminating human-written versus ChatGPT-written abstracts	“Our objective was to determine if the seniority of the reviewer has an impact on the ability to discriminate between fabricated human and ChatGPT-written abstracts. We also aimed to explore the reviewers’ own perception of their accuracy and level of confidence in their decision.”	Not stated.	Levin et al. (2024)

These seven studies aim to evaluate the effectiveness of AI detection tools and human judgment in distinguishing AI-generated content from human-written texts in various academic and creative contexts. They explore the capabilities of teachers, students, and medical researchers in identifying AI-generated essays, stories, poetry, and medical research backgrounds while examining predictors of success and broader attitudes towards AI use in education. Additionally, they investigate how training in AI-assisted writing tools influences students' writing, the impact of AI-generated content on creative text preferences, and the association between reviewer experience and the ability to discriminate between human-written and AI-generated abstracts.

Results

The results from these studies about identification are shown in Table 53.

Table 53

Identification: Results

Title	Results	Author
Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse	“Although the AI detection tool identified 91% of the experimental submissions as containing AI-generated content, only 54.8% of the content was identified as AI-generated, underscoring the challenges of detecting AI content when advanced prompting techniques are used. When academic staff members marked the experimental submissions, only 54.5% were reported to the academic misconduct process, emphasising the need for greater awareness of how the results of AI detectors may be interpreted. Similar performance in	Perkins et al. (2023)

Title	Results	Author
	grades was obtained between student submissions and AI-generated content (AI mean grade: 52.3, Student mean grade: 54.4), showing the capabilities of AI tools in producing human-like responses in real-life assessment situations.”	
Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students	“Participants (69 high school teachers, 140 high school students, total N = 209) took an AI Identification Test in which they read pairs of essays—one written by a high school student and the other by ChatGPT—and guessed which was generated by the chatbot. Accuracy was only 70% for teachers, and it was slightly worse for students (62%). Self-reported confidence did not predict accuracy, nor did experience with ChatGPT or subject-matter expertise. Well-written student essays were especially hard to differentiate from the ChatGPT texts. In another set of measures, students reported greater optimism than their teachers did about the future role of ChatGPT in education. Students expressed disapproval of submitting ChatGPT-generated essays as one's own but rated this and other possible academic integrity violations involving ChatGPT less negatively than teachers did. These results form an empirical basis for further work on the relationship between AI and academic integrity.”	Waltzer et al. (2023)
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“We employ several machine learning models to classify the texts. The results indicate that the models exhibit superior performance in binary classification tasks, such as distinguishing human-generated text from a specific LLM, compared with the more complex multiclass tasks that involve discerning among human-generated and multiple LLMs. / This research marks a step forward in the field of LLMs, specifically in distinguishing human-written text from AI-generated content. A key contribution of this work is the introduction of a novel dataset for evaluating LLM-generated text from GPT and BARD across multiple genres, including stories and poetry. Our results indicate that while LLMs have made remarkable progress, it remains possible to discern human-written texts from those generated by LLMs using ML and DL models with a small training set. We observed high performance when differentiating BARD-generated texts from human-written texts. However, the task becomes more challenging with GPT-generated text, especially in story writing. By creating and leveraging this unique dataset, we have not only deepened our understanding of the nuanced differences between LLM and human writing but also provided a valuable resource for future research and development in this emerging field.”	Hayawi et al. (2023)
Clinical Research With Large Language Models Generated Writing-Clinical Research with AI-assisted Writing (CRAW) Study	“A total of 80 researchers completed the survey. The median H-index was 3 (interquartile range, 1-7.25) and most (36%) researchers were from the Critical Care specialty. When compared with researchers with an H-index of 22 and 13, GPT-3.5 was marked high on the Likert scale ranking on main background components (median 4.5 vs. 3.82 vs. 3.6 vs. 4.5, respectively; $p < 0.001$). The sensitivity and specificity to detect researchers writing versus GPT-3.5 writing were poor, 22.4% and 57.6%, respectively. Conclusions and Relevance: GPT-3.5 could create background research content indistinguishable from the writing of a medical researcher. It was marked higher compared with medical researchers with	Huespe et al. (2023)

Title	Results	Author
	an H-index of 22 and 13 in writing the background section of a critical care clinical research question.”	
“AI, will you help?” How learners use Artificial Intelligence when writing	“The findings of this study indicate the high linguistic sensitivity of the research participants to factors regarding language and layout, which allows them to distinguish human authors from AI-powered texts. The current investigation possesses potential advantages for educators in the realm of foreign language acquisition and instruction, as they contemplate the strengths of their bilingual language learners within academic writing instruction.”	Krajka and Olszak (2024)
Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI- generated from human-written poetry	“In a new incentivized version of the Turing Test, participants failed to reliably detect the algorithmically generated poems in the Human-in-the-loop treatment, yet succeeded in the Human-out-of-the-loop treatment. Further, people reveal a slight aversion to algorithm-generated poetry, independent on whether participants were informed about the algorithmic origin of the poem (Transparency) or not (Opacity). We discuss what these results convey about the performance of NLG algorithms to produce human-like text and propose methodologies to study such learning algorithms in human-agent experimental settings.”	Köbis and Mossink (2020)
Association of reviewer experience with discriminating human-written versus ChatGPT-written abstracts	“The 30 reviewers discriminated 20 abstracts, giving a total of 600 abstract evaluations. The reviewers were able to correctly identify 300/600 (50%) of the abstracts: 139/300 (46.3%) of the ChatGPT-generated abstracts and 161/300 (53.7%) of the human-written abstracts ($p=0.07$). Human-written abstracts had a higher rate of correct identification (median (IQR) 56.7% (49.2-64.1%) vs 45.0% (43.2-48.3%), $p=0.023$). Senior reviewers had a higher correct identification rate (60%) than junior reviewers and residents (45% each; $p=0.043$ and $p=0.002$, respectively). In a linear regression model including the experience level of the reviewers, familiarity with artificial intelligence (AI) and the country in which the majority of medical training was achieved (English speaking vs non-English speaking), the experience of the reviewer ($\beta=10.2$ (95% CI 1.8 to 18.7)) and familiarity with AI ($\beta=7.78$ (95% CI 0.6 to 15.0)) were independently associated with the correct identification rate ($p=0.019$ and $p=0.035$, respectively). In a correlation analysis the number of publications by the reviewer was positively correlated with the correct identification rate ($r_{28}=0.61$, $p<0.001$). Conclusion A total of 46.3% of abstracts written by ChatGPT were detected by reviewers. The correct identification rate increased with reviewer and publication experience.”	Levin et al. (2024)

The results from the first study by Perkins et al. (2023) report that an AI detection tool identifies 91% of experimental submissions as containing AI-generated content, but only 54.8% of the content is recognized as AI-generated. This highlights the challenges of detecting AI content, especially with advanced

prompting techniques. When academic staff marked the submissions, only 54.5% were reported for academic misconduct, emphasizing the need for greater awareness of AI detector results. The study finds similar performance in grades between student submissions and AI-generated content, demonstrating AI tools' capability to produce human-like responses in assessments. The second study by Waltzer et al. (2023) examine the ability of teachers and students to differentiate between essays written by high school students and ChatGPT. The accuracy rate is 70% for teachers and 62% for students. Confidence levels do not predict accuracy, nor does experience with ChatGPT or subject-matter expertise. Well-written student essays are particularly hard to distinguish from ChatGPT texts. Students are more optimistic than teachers about ChatGPT's future role in education, although both groups disapprove of submitting ChatGPT-generated essays as one's own. These findings provide an empirical basis for further research on AI and academic integrity.

The third study by Hayawi et al. (2023) uses machine learning models to classify texts, finding superior performance in binary classification tasks (human vs. specific LLM) compared to multiclass tasks (human vs. multiple LLMs). They introduce a novel dataset for evaluating LLM-generated text from GPT and BARD across various genres, showing that it is possible to discern human-written texts from LLM-generated texts using small training sets. The study notes high performance in differentiating BARD-generated texts but finds GPT-generated text, especially in story writing, more challenging to identify. The fourth study by Huespe et al. (2023) finds that GPT-3.5 can create background research content that is indistinguishable from medical researchers' writing. Researchers with an H-index of 22 and 13 rate GPT-3.5 higher in writing background sections of critical care clinical research. The sensitivity and specificity to detect researcher writing versus GPT-3.5 writing are

poor, at 22.4% and 57.6%, respectively. The study concludes that GPT-3.5 can produce high-quality research content comparable to human researchers.

According to the fifth study by Krajka and Olszak (2024), the findings suggest that research participants show high linguistic sensitivity to language and layout, allowing them to distinguish human authors from AI-powered texts. This investigation has potential benefits for educators in foreign language acquisition and instruction, as it highlights the strengths of bilingual language learners in academic writing. In the sixth study, Köbis and Mossink (2020) conduct an incentivized Turing Test and find that participants fail to reliably detect algorithmically generated poems in the Human-in-the-loop treatment but succeed in the Human-out-of-the-loop treatment. Participants exhibit a slight aversion to algorithm-generated poetry, regardless of whether they are informed about the algorithmic origin. The study suggests that NLG algorithms can produce human-like text, but methodologies are needed to study these algorithms in human-agent experimental settings.

The last study by Levin et al. (2024) reports that reviewers correctly identify 50% of the abstracts, with 46.3% of ChatGPT-generated abstracts and 53.7% of human-written abstracts correctly identified. Senior reviewers have a higher correct identification rate (60%) compared to junior reviewers and residents (45% each). Reviewer experience and familiarity with AI are independently associated with the correct identification rate. The number of publications by the reviewer is positively correlated with the correct identification rate. The study concludes that reviewer and publication experience improve the detection of AI-generated abstracts.

Implications

Table 54 provides the implications of these studies.

Table 54*Identification: Implications*

Title	Implications	Author
Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse	“It is a challenge for educators to distinguish AI-generated essays from genuinely human (student) generated essays. Therefore, school administrators and leaders must implement AI-generated text detection software to uphold academic integrity standards. AI generated essays submitted by students can be detected by teachers and educational designers using the proposed language model and ML classifier at a high accuracy. Human (student) generated essays can and must be correctly identified with 100% accuracy even if the overall classification accuracy performance is slightly reduced.”	Perkins et al. (2023)
Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students	# “Our research shows that distinguishing student writing from writing produced by chatbots such as ChatGPT is a significant challenge already, and it is likely to get even more difficult as AI continues to improve. Our findings also suggest that content area expertise and familiarity with ChatGPT are not sufficient to overcome this problem. We suspect that total bans on ChatGPT will be ineffective at best and that a more fruitful approach will be to find ways to understand and make use of the benefits that ChatGPT offers while minimizing its risks. Teachers who are already overburdened should not be asked to solve this problem on their own. Instead, researchers and educators will need to work together to develop best practices and effective strategies for disseminating them. Conversations that relate to this challenge are already underway, with much of the focus on using ChatGPT as a tool to achieve other goals, such as supporting language learners and providing personalized feedback on writing assignments [48].” # “Given that learning to work with AI systems will be a valuable skill for the foreseeable future, it will also be important to focus on teaching students how to use chatbots effectively while also thinking critically about the output they generates. As noted by New York Times tech columnist Kevin Roose, students “need to know their way around these tools — their strengths and weaknesses, their hallmarks and blind spots—in order to work alongside them” [42]. Understanding how people think about the role of ChatGPT in educational contexts is important not only due to its implications for student learning but for broader moral questions as well, including fairness. As one student commented, “I’m concerned because people who don’t use AI, like me, will see their education and grades suffer if it becomes normal.” These kinds of risks, along with the potential benefits, require creative problem-solving to effectively address. Empirical evidence and vigorous debate will play a critical role in navigating this new educational landscape.”	Waltzer et al. (2023)
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“This research has several significant implications, particularly in differentiating between human-generated and AI-generated texts. First, the relatively high performance achieved across most experiments, especially SVM, in distinguishing between human and GPT-generated or BARD-generated text indicates that with the right features and model configuration, it is feasible to detect artificially generated text. The only exception is the poor performance of the models in detecting GPT-generated stories	Hayawi et al. (2023)

	from human-written ones. The results could extend to various applications, such as detecting deep fake text in news or social media or validating the authenticity of content. The results also suggest that ML can successfully differentiate between the nuances in human writing and algorithmic output, even with limited data size. However, the differences between the performance of the classifiers when identifying GPT or BARD versus human text, suggest that different AI writing models might have unique ‘fingerprints’ or characteristics that these classifiers can detect. This could lead to the development of more specialised detection tools for different types of LLMs. Finally, the study also contributes to raising ethical implications in the context of LLMs in academic integrity. As AI text generation improves, the potential for misuse increases, necessitating the importance of effective detection tools. It also emphasises the need for a discussion about the responsibility of AI developers, researchers and users in ensuring that these technologies are used ethically and safely.”	
Clinical Research with Large Language Models Generated Writing-Clinical Research with AI-assisted Writing (CRAW) Study	“The use of GPT-3.5 as a tool for generating scientific backgrounds can be particularly beneficial for non-native English speakers, as it provides assistance in generating fluency and coherence in writing.”	Huespe et al. (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“There is little doubt that the fields of language teaching and learning could be significantly enhanced by the ongoing development of Artificial Intelligence capabilities. According to the study’s findings, the students were not very successful at distinguishing between AL and human-generated writing with accuracy rates between 40-80%. They tended to assess human writing based on technical accuracy and error and AI writing based on verbal complexity. Applying AI-powered tools could help students improve their writing competence and bring about changes to the way academic writing is taught as a whole. However, as the study’s findings show, they can also hinder the process of improving writing abilities because they lack imagination, use straightforward and sometimes repetitious vocabulary, respond slowly, or make minor grammatical errors. The research reveals that even highly proficient, almost bilingual, academic students encounter considerable difficulties when attempting to differentiate human writing from AI-assisted written texts. These challenges arise when analyzing formal and informal texts.”	Krajka and Olszak (2024)
Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry	“Algorithms that generate text resembling human language are becoming ever more widely accessible. Not only novelists with writer’s block can make use of freely available algorithms like GPT-2. Understanding humans’ behavioral reactions to such algorithms helps shape policies to ensure artificial intelligence remains beneficial (Crawford & Calo, 2016). As a step in that direction, the present set of studies adopts a behavioral science approach to examine creative artificial intelligence. We hope more studies follow suit to inform policies of disclosure of algorithmic presence and provide new behavioral insights into human versus (creative) artificial intelligence.”	Köbis and Mossink (2020)

Association of reviewer experience with discriminating human-written versus ChatGPT-written abstracts	“Our study shows that the capability of human reviewers to detect abstracts authored by ChatGPT is limited, with a success rate that mirrors the randomness of a coin flip. This detection rate, however, shows a marked positive correlation with the reviewer’s academic experience and familiarity with the literature. These findings are crucial for journal editorial teams as they underscore the need for heightened awareness and potentially revised editorial processes to maintain the integrity of scientific publishing in the era of AI.”	Levin et al. (2024)

First, the study by Perkins et al. (2023) implies that educators face significant challenges in distinguishing AI-generated essays from student-written ones, suggesting the need for AI detection software to maintain academic integrity standards.

Second, according to Waltzer et al. (2023), differentiating between student writing and AI-generated text is difficult and likely to become more challenging. It suggests that banning AI tools may be ineffective, and instead, educators should focus on finding ways to use AI beneficially while minimizing risks.

Third, a study by Hayawi et al. (2023) suggests that machine learning models can effectively detect AI-generated text in most cases, which has implications for identifying deep fake text and validating content authenticity. It also raises ethical concerns about the use of AI in academic settings.

Fourth, research by (Huespe et al.) implies that AI language models like GPT-3.5 can be particularly beneficial for non-native English speakers in generating scientific writing, potentially enhancing fluency and coherence. Fifth, a study by Krajka & Olszak (2024) suggests that students struggle to distinguish between AI and human-generated writing, which has implications for how academic writing is taught and assessed. It also indicates that AI tools could both help and hinder writing skill development.

Sixth, research by Köbis & Mossink (2020) implies that people have difficulty differentiating between AI-generated and human-written poetry, which

raises questions about the need for disclosure policies regarding AI-generated creative content. Seventh, a study Levin et al. (2024) suggests that human reviewers have limited ability to detect AI-written abstracts, with detection rates improving with academic experience. This has significant implications for maintaining the integrity of scientific publishing in the age of AI.

Machine Learning

Purpose

There are four studies related to machine learning.

Table 55

Machine Learning: Purpose

Title	Purpose	Research Question	Author
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“to analyze the differences between medical texts written by human experts and those generated by ChatGPT and to design machine learning workflows to effectively detect and differentiate medical texts generated by ChatGPT”	“(1) What is the difference between medical content written by humans and that generated by ChatGPT? (2) Can we use machine learning methods to detect whether medical content is written by human experts or ChatGPT?”	Liao et al. (2023)
It takes one to know one—Machine learning for identifying OBGYN abstracts written by ChatGPT	“To use machine learning to optimize the detection of obstetrics and gynecology (OBGYN) Chat Generative Pre-Trained Transformer (ChatGPT) -written abstracts of all OBGYN journals.”	Not stated.	Levin et al. (2024)
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“...distinguishing between human-written and AI-generated text has become a significant task. This article presents a comparative study, introducing a novel dataset of human-written and LLM-generated texts in different genres: essays, stories, poetry and Python code. We employ several machine learning models to classify the texts”	Not stated.	Hayawi et al. (2023)
An intelligent scoring system for English writing	“to establish an intelligent English writing scoring system based on Artificial	Not stated.	Huang (2023)

Title	Purpose	Research Question	Author
based on artificial intelligence and machine learning	Intelligence (AI) and explore the application of multi-task learning algorithm models in machine learning”		

The four studies aim to explore and enhance the detection and differentiation of AI-generated texts from human-written content using various machine-learning techniques. The first study (Liao et al., 2023) focuses on analyzing the differences between human-written and ChatGPT-generated medical texts and developing machine learning workflows to effectively identify AI-generated content. The second study (Levin et al., 2024) aims to optimize machine learning methods to detect ChatGPT-written abstracts in obstetrics and gynecology journals. The third study (Hayawi et al., 2023) presents a comparative analysis of human and AI-generated texts across multiple genres, employing several machine-learning models for classification. The fourth study (Huang, 2023) aims to create an intelligent English writing scoring system using AI and investigate the application of multi-task learning algorithm models in machine learning.

Results

The results from these studies are shown in Table 56.

Table 56

Machine Learning: Results

Title	Results	Author
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“Medical texts written by humans were more concrete, more diverse, and typically contained more useful information, while medical texts generated by ChatGPT paid more attention to fluency and logic and usually expressed general terminologies rather than effective information specific to the context of the problem. A bidirectional encoder representation from transformers-based model effectively detected medical texts generated by ChatGPT, and the F 1 score exceeded 95%. Conclusions: Although text generated by ChatGPT is grammatically perfect and human-like, the linguistic characteristics of generated medical texts were different from	Liao et al. (2023)

Title	Results	Author
	those written by human experts. Medical text generated by ChatGPT could be effectively detected by the proposed machine learning algorithms. This study provides a pathway toward trustworthy and accountable use of large language models in medicine.”	
It takes one to know one—Machine learning for identifying OBGYN abstracts written by ChatGPT	“Overall, 75 abstracts from 12 different OBGYN journals were randomly selected. There were seven (58%) Q1 journals, one (8%) Q2 journal, two (17%) Q3 journals, and two (17%) Q4 journals. Use of mixed dialects of English, absence of comma-misuse, absence of incorrect verb forms, and improper formatting were important prediction variables of ChatGPT-written abstracts. The deep-learning model had the highest predictive performance of all examined models. This model achieved the following performance: accuracy 0.90, precision 0.92, recall 0.85, area under the curve 0.95. Machine-learning-based tools reach high accuracy in identifying ChatGPT-written OBGYN abstracts.”	Levin et al. (2024)
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“We employ several machine learning models to classify the texts. The results indicate that the models exhibit superior performance in binary classification tasks, such as distinguishing human-generated text from a specific LLM, compared with the more complex multiclass tasks that involve discerning among human-generated and multiple LLMs. / This research marks a step forward in the field of LLMs, specifically in distinguishing human-written text from AI-generated content. A key contribution of this work is the introduction of a novel dataset for evaluating LLM-generated text from GPT and BARD across multiple genres, including stories and poetry. Our results indicate that while LLMs have made remarkable progress, it remains possible to discern human-written texts from those generated by LLMs using ML and DL models with a small training set. We observed high performance when differentiating BARD-generated texts from human-written texts. However, the task becomes more challenging with GPT-generated text, especially in story writing. By creating and leveraging this unique dataset, we have not only deepened our understanding of the nuanced differences between LLM and human writing but also provided a valuable resource for future research and development in this emerging field.”	Hayawi et al. (2023)
An intelligent scoring system for English writing based on artificial intelligence and machine learning	“The experimental results show that the multi-task learning model performs very well. There are 49,594 syntax errors in the training set, including 11,076 redundant word errors, 13,246 missing word errors, 21,898-word selection errors and 3374-word order errors. The test set has 11,085 syntax errors, including 2406 redundant word errors, 2973 missing word errors, 4860 vocabulary selection errors and 846-word order errors. Finally, the accuracy of the scoring system is evaluated, which is 91.5%. The accuracy of the algorithm can reach 97%. By training the model, the multi-task learning algorithm model can save a lot of time and labor costs for manually correcting essays.”	Huang (2023)

The studies reveal that medical texts generated by ChatGPT are less concrete and context-specific compared to human-written texts but can be effectively detected using machine learning models with an F1 score exceeding 95%. In the case of OBGYN abstracts, the deep-learning model achieves high accuracy (0.90), precision (0.92), recall (0.85), and an area under the curve (0.95) in identifying AI-generated content, which highlights the importance of specific linguistic features for prediction. The research on distinguishing human and AI-generated texts across genres, including essays and poetry, shows that machine learning models perform better in binary classification tasks and identify unique 'fingerprints' of different LLMs despite the challenges posed by GPT-generated texts. Lastly, the intelligent scoring system for English writing based on a multi-task learning model demonstrates a high accuracy rate (91.5% to 97%) in correcting syntax errors, which can significantly reduce time and labor costs for manual corrections.

Implications

Table 57 provides the implications of the studies on machine learning.

Table 57

Machine Learning: Implications

Title	Implications	Author
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“In general, for artificial intelligence (AI) to realize its full potential in medicine, we should not rush into its implementation but advocate for its careful introduction and open debate about its risks and benefits. First, human medical writers will be responsible for ensuring the accuracy and completeness of the information communicated and for complying with ethical and regulatory guidelines. However, ChatGPT cannot be held responsible. Second, training an LLM requires a huge amount of data, but the quality of the data is difficult to guarantee, so the trained ChatGPT is biased. For example, ChatGPT can provide biased output and perpetuate sexist stereotypes [52]. Third, use of ChatGPT may lead to private information leakage. This may be because the LLM remembers personal privacy information in the training set [53]. What is more, the legal framework must be considered. Who shall be held accountable when an AI doctor makes an inevitable mistake? ChatGPT cannot be held	Liao et al. (2023)

Title	Implications	Author
	accountable for its work, and there is no legal framework to determine who owns the rights to AI-generated work [15].”	
It takes one to know one—Machine learning for identifying OBGYN abstracts written by ChatGPT	“It may be possible to identify artificial intelligence-generated scientific text. Some special characteristics may be considered. First, the text is likely to exhibit a formal and structured style, following the conventions of academic or technical writing. This includes proper grammar, precise terminology, and a clear progression of ideas with appropriate headings and subheadings. Additionally, the content will often feature citations, references to established research, and adherence to a specific citation style. Artificial intelligence-generated scientific text will typically avoid colloquial language, emotive expressions, or overly informal tone, striving for an objective and informative approach. The accuracy and relevance of the information presented in the text, as well as its alignment with established scientific knowledge, are crucial indicators of its authenticity. Lastly, careful examination may reveal occasional errors or inaccuracies that are not in line with established scientific facts, potentially hinting at the artificial intelligence origin of the text (“hallucination”). ^{8,9} However, as previously described, human reviewers could not accurately differentiate ChatGPT-generated abstracts.”	Levin et al. (2024)
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“This research has several significant implications, particularly in differentiating between human-generated and AI-generated texts. First, the relatively high performance achieved across most experiments, especially SVM, in distinguishing between human and GPT-generated or BARD-generated text indicates that with the right features and model configuration, it is feasible to detect artificially generated text. The only exception is the poor performance of the models in detecting GPT-generated stories from human-written ones. The results could extend to various applications, such as detecting deep fake text in news or social media or validating the authenticity of content. The results also suggest that ML can successfully differentiate between the nuances in human writing and algorithmic output, even with limited data size. However, the differences between the performance of the classifiers when identifying GPT or BARD versus human text, suggest that different AI writing models might have unique ‘fingerprints’ or characteristics that these classifiers can detect. This could lead to the development of more specialised detection tools for different types of LLMs. Finally, the study also contributes to raising ethical implications in the context of LLMs in academic integrity. As AI text generation improves, the potential for misuse increases, necessitating the importance of effective detection tools. It also emphasises the need for a discussion about the responsibility of AI developers, researchers and users in ensuring that these technologies are used ethically and safely.”	Hayawi et al. (2023)
An intelligent scoring system for English writing based on artificial intelligence and machine learning	“Consequently, students can learn autonomously without teachers’ guidance. In the future, proofreading algorithms can be incorporated into the system to enhance their error correction training and achieve the integration of error detection and correction. In addition, machine translation is applied to the grammar error correction process of the	Huang (2023)

Title	Implications	Author
	algorithm to improve its ability to detect and correct syntax error in English writing, which will significantly promote students’ autonomous learning and reduce their dependence on teachers.”	

The implications of these studies underscore the critical need for careful integration and oversight of AI tools in various domains to ensure accuracy, ethical use, and effective learning outcomes. In medicine, the cautious introduction of AI, like ChatGPT, is essential due to potential biases, privacy concerns, and the lack of a legal framework for accountability. The ability to identify AI-generated scientific text, such as OBGYN abstracts, is crucial, though human reviewers currently struggle with accurate differentiation, highlighting the need for specialized detection tools. The research on differentiating human and AI-generated texts across genres indicates that with proper features and models, it is feasible to detect AI content, but ethical considerations and the risk of misuse remain significant concerns. The development of an intelligent scoring system for English writing shows promise in promoting autonomous learning and reducing teacher dependency, though future improvements in error correction and syntax detection are necessary to enhance its effectiveness. Overall, these studies call for ongoing dialogue, ethical considerations, and advancements in detection technologies to harness AI's potential while mitigating its risks.

Output Characteristics: Quality and Accuracy of Academic Content

Purpose

There are eight studies that investigate this issue of quality in AI-generated content.

Table 58*The Quality and Accuracy of Academic Content: Purpose*

Title	Purpose	Research Question	Author
Strengths and Weaknesses of ChatGPT Models for Scientific Writing About Medical Vitamin B12: Mixed Methods Study	“to evaluate the ability of GPT-4 to generate scientific content and assist in scientific writing using medical vitamin B12 as the topic. Furthermore, the study will compare the performance of GPT-4 to its predecessor, GPT-3.5.”	Not stated.	Abuyaman (2023)
Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study	“to assess the applicability of an artificial intelligence (AI) model in generating abstracts for basic preclinical research.”	Not stated.	Cheng et al. (2023)
A large-scale comparison of human-written versus ChatGPT-generated essays	“to systematically assess the quality of AI-generated content through a large-scale study comparing human-written versus ChatGPT-generated argumentative student essays.”	“RQ1: How good is ChatGPT based on GPT-3 and GPT-4 at writing argumentative student essays? RQ2: How do AI-generated essays compare to essays written by students? RQ3: What are linguistic devices that are characteristic of student versus AI-generated content?”	Herbold et al. (2023)
A comparison of ChatGPT-generated articles with human-written articles	“to compare the accuracy and quality of several ChatGPT-generated academic articles with those written by human authors.”	Not stated.	Ariyaratne et al. (2023)
Reliability and Quality of the Nursing Care Planning Texts Generated by ChatGPT	“to examine the quality, authenticity, and reliability of the nursing care planning texts produced using ChatGPT”	• What is the quality subscale score of texts about nursing diagnosis generated by ChatGPT? • What is the reliability subscale score of texts about nursing diagnosis generated by ChatGPT? • What is the similarity rate of texts about nursing diagnosis generated by ChatGPT? • Is there a relationship between independent research variables (number of references, word count, similarity rate, etc) and the quality and reliability subscale scores of texts about nursing diagnosis generated by ChatGPT?”	Dağcı et al. (2023)

Title	Purpose	Research Question	Author
The Death of the Short-Form Physics Essay in the Coming AI Revolution	“to present evidence of AI-generated short-form essays achieving First-Class grades on an essay writing assessment from an accredited, current university Physics module.”	Not stated.	Yeadon et al. (2022)
Can ChatGPT assist authors with abstract writing in medical journals? Evaluating the quality of scientific abstracts generated by ChatGPT and original abstracts	“the present study aimed to evaluate ChatGPT’s potential application in academic writing, with a specific focus on generating scientific abstracts from full RCT reports. The objectives of the study were to compare the adherence of ChatGPT-generated and original abstracts to the CONSORT-A statement and explore factors that influence the reporting quality.”	Not stated.	Hwang et al. (2024)
Generative artificial intelligence: Can ChatGPT write a quality abstract?	to investigate if the freely available version of ChatGPT can produce a quality conference abstract using a fictitious but accurately calculated data table as applied by a non-medically trained person	Not stated.	Babl and Babl (2023)

The eight studies collectively aim to assess the quality of various types of scientific content that ChatGPT-generated. They explore ChatGPT's performance in creating different text formats, such as abstracts, essays, and nursing care plans, and compare the quality and accuracy of AI-generated content to human-produced work. Additionally, the studies investigate factors influencing ChatGPT's output, such as the complexity of the topic, the specific version of the model, and the length of the text. Overall, these studies aimed at providing insights into the potential applications and limitations of ChatGPT in scientific writing.

Results

The results from the studies on quality are shown in Table 59.

Table 59*The Quality and Accuracy of Academic Content: Results*

Title	Results	Author
Strengths and Weaknesses of ChatGPT Models for Scientific Writing About Medical Vitamin B12: Mixed Methods Study	“The results indicated that GPT-4 can potentially streamline scientific writing through its ability to edit language and write abstracts, keywords, and abbreviation lists. However, significant limitations of ChatGPT were revealed, including its inability to identify and address bias, inability to include recent information, lack of transparency, and inclusion of inaccurate information. Additionally, it cannot check for plagiarism or provide proper references. The accuracy of GPT-4’s answers was found to be superior to that of GPT-3.5. ChatGPT can be considered a helpful assistant in the writing process but not a replacement for a scientist’s expertise.”	Abuyaman (2023)
Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study	“The quality of ChatGPT-generated abstracts was lower than that of the actual abstracts (10-point Likert scale: mean 4.72, SD 2.09 vs mean 8.09, SD 1.03; $P < .001$). The difference in quality was significant in the unstructured format (mean difference -4.33 ; 95% CI -4.79 to -3.86 ; $P < .001$) but minimal in the 4-subheading structured format (mean difference -2.33 ; 95% CI -2.79 to -1.86). Among the 30 ChatGPT-generated abstracts, 3 showed wrong conclusions, and 10 were identified as AI content. The mean percentage of similarity between the original and the generated abstracts was not high (2.10%-4.40%). The blinded reviewers achieved a 93% (224/240) accuracy rate in guessing which abstracts were written using ChatGPT.”	Cheng et al. (2023)
A large-scale comparison of human-written versus ChatGPT-generated essays	“Our results demonstrate that ChatGPT generates essays that are rated higher regarding quality than human-written essays. The writing style of the AI models exhibits linguistic characteristics that are different from those of the human-written essays. Our results provide clear answers to the first two research questions that consider the quality of the generated essays: ChatGPT performs well at writing argumentative student essays and outperforms the quality of the human-written essays significantly. The ChatGPT-4 model has (at least) a large effect and is on average about one point better than humans on a seven-point Likert scale. Regarding the third research question, we find that there are significant linguistic differences between humans and AI-generated content. The AI-generated essays are highly structured, which for instance is reflected by the identical beginnings of the concluding sections of all ChatGPT essays (‘In conclusion, [...]’). The initial sentences of each essay are also very similar starting with a general statement using the main concepts of the essay topics. Although this corresponds to the general structure that is sought after for argumentative essays, it is striking to see that the ChatGPT models are so rigid in realizing this, whereas the human-written essays are looser in representing the guideline on the linguistic surface. Moreover, the linguistic fingerprint has the counter-intuitive property that the use of discourse markers is negatively correlated with logical coherence.”	Herbold et al. (2023)

Title	Results	Author
A comparison of ChatGPT-generated articles with human-written articles	“In total, 4 of the 5 articles written by ChatGPT were significantly inaccurate with fictitious references. One of the papers was well written, with a good introduction and discussion; however, all references were fictitious. ChatGPT is able to generate coherent research articles, which on initial review may closely resemble authentic articles published by academic researchers. However, all of the articles we assessed were factually inaccurate and had fictitious references. It is worth noting, however, that the articles generated may appear authentic to an untrained reader.”	Ariyaratne et al. (2023)
Reliability and Quality of the Nursing Care Planning Texts Generated by ChatGPT	“The texts were evaluated by using descriptive criteria form and DISCERN tool to evaluate health information. The texts created without specifying the length were short, with a mean word count of 366.95 ± 57.98 (266- 506). The accessibility and accuracy of the references were checked. The majority were identified to be fictional and not found on either PubMed or Google. The similarity rate of the texts was low when checked through iThenticate and Turnitin; however, the Turnitin, which integrates AI detection technology, confirmed that all the texts were 100% written by AI (Table 1). The quality and content of the texts were evaluated using DISCERN. Their reliability, quality of information on nursing care, and overall quality were moderate (Table 2). In addition, 100% of the 40 ChatGPT texts had moderate reliability, with 39 (97.5%) providing moderate quality of information on nursing care. Regarding the overall quality of the texts, only 2 (5%) had good quality, 30 (75%) had moderate quality, and 8 (20%) had poor quality (Table 3). A statistically significant relationship was found between the number of accessible references and the reliability ($r = 0.408$) and quality ($r = 0.379$) of the texts ($P < .05$). Furthermore, a statistically significant relationship was found between the accessibility of web references and the reliability of the texts ($r = 0.351$; $P < .05$) (Table 4).”	Dağcı et al. (2023)
The Death of the Short-Form Physics Essay in the Coming AI Revolution	“Fifty AI answers were generated to create ten submissions that were independently marked by five separate markers. The AI generated submissions achieved an average mark of $71 \pm 2\%$, in strong agreement with the current module average of $71 \pm 5\%$. A typical AI submission would therefore most-likely be awarded a First Class, the highest classification available at UK universities. Plagiarism detection software returned a plagiarism score between $2 \pm 1\%$ (Grammarly) and $7 \pm 2\%$ (Turnitin).”	Yeadon et al. (2022)
Can ChatGPT assist authors with abstract writing in medical journals? Evaluating the quality of scientific abstracts generated by ChatGPT and original abstracts	“Original abstracts achieved a mean OQS of 11.89 (95% CI: 11.23-12.54), outperforming GPT 3.5 (7.89; 95% CI: 7.32-8.46) and GPT 4 (5.18; 95% CI: 4.64-5.71). Compared to GPT 3.5 and 4 outputs, original abstracts were more adherent with 10 and 14 CONSORT-A items, respectively. In blind assessments, GPT 3.5-generated abstracts were deemed most readable in 62.22% of cases which was significantly greater than the original (31.11%; $P = 0.003$) and GPT 4-generated (6.67%; $P < 0.001$) abstracts. Moreover, ChatGPT 3.5 exhibited a hallucination rate of 0.03 items per abstract compared to 1.13 by GPT 4. No determinants for improved reporting quality were identified	Hwang et al. (2024)

Title	Results	Author
	for GPT-generated abstracts. Conclusions: While ChatGPT could generate more readable abstracts, their overall quality was inferior to the original abstracts. Yet, its proficiency to concisely relay key information with minimal error holds promise for medical research and warrants further investigations to fully ascertain the LLM's applicability in this domain. Main Outcomes and Measures: In this study, we generated a background section of a critical care clinical research question on "acute kidney injury in sepsis" using three different methods: researcher with H-index greater than 20, researcher with H-index greater than 10, and GPT-3.5. The three background sections were presented in a blinded survey to researchers with an H-index range between 1 and 96. First, the researchers evaluated the main components of the background using a 5-point Likert scale. Second, they were asked to identify which background was written by humans only or with large language model-generated tools."	
Generative artificial intelligence: Can ChatGPT write a quality abstract?	"The resulting abstract was well written without obvious errors and followed the abstract instructions. One of the references was fictitious, known as 'hallucination'. ChatGPT or similar programmes, with careful review of the product by authors, may become a valuable scientific writing tool. The scientific and medical use of generative artificial intelligence, however, raises many questions. ChatGPT or similar programmes when used as a writing tool with careful review by scientific authors may become a valuable tools to speed a first draft of papers and grants, summarise or simplify long articles, reduce duplication, polish and streamline existing documents or assist researchers with limited English skills (although it also works in other languages). It does not seem obvious why standard methodology or even results sections of a paper could not be written with the assistance of generative AI when large sections of research papers are proscribed in some detail – often with suggested wording by various EQUATOR reporting guidelines. ⁷ EQUATOR now includes guidance for studies including AI and machine learning."	Babl and Babl (2023)

Abuyaman (2023) finds that GPT-4 can streamline scientific writing but has limitations in addressing bias and providing accurate information. Cheng et al. (2023) report that ChatGPT-generated abstracts are generally lower quality than human-written ones, especially in unstructured formats. Herbold et al. (2023) observe that ChatGPT-generated essays are rated higher in quality than human-written essays, albeit with distinct linguistic characteristics. Ariyaratne et al. (2023) note that most ChatGPT-generated articles contain significant inaccuracies and fictitious references. Dağcı et al. (2023) find AI-generated nursing care texts to be of

moderate quality with fictional references. Yeadon et al. (2022) report that AI-generated physics essays achieve marks comparable to human averages with low plagiarism scores. Hwang et al. (2024) observe that while GPT-3.5 generates more readable medical abstracts, they are of lower overall quality than human-written ones. Babl and Babl (2023) suggest that ChatGPT could be a valuable scientific writing tool but emphasize the need for careful review by authors. Across these studies, a common core concept emerges: AI-generated content can be of high quality and readability, but it often lacks accuracy, proper referencing, and the nuanced understanding required for scientific writing, highlighting the continued importance of human expertise and oversight in academic writing.

Implications

Table 60 provides the implications of the studies on the quality and accuracy of academic content.

Table 60

The Quality and Accuracy of Academic Content: Implications

Title	Implications	Author
Strengths and Weaknesses of ChatGPT Models for Scientific Writing About Medical Vitamin B12: Mixed Methods Study	“Through prompts related to vitamin B12, potential was found in the GPT-4 model of ChatGPT for scientific writing. This tool could particularly benefit non-English speaking scientists by helping them streamline the language in their publications. This study revealed critical limitations of ChatGPT, mainly in relation to accuracy, bias, and references. The shortcomings warrant judicious use and avoiding its use as a substitute for human expertise. GPT-4 showed improvements compared to GPT-3.5, indicating promise for the further development of the tool. Overall, ChatGPT has the potential to be a useful tool for scientific writing, but researchers must be aware of its limitations and use it appropriately.”	Abuyaman (2023)
Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study	“In conclusion, AI-generated content can be identified by both human experts and computer-based analyses. AI may be capable of deceiving human experts in common or less specialized subjects. However, the limitations of AI have become more apparent in professional domains. AI-driven output detectors based on linguistic analyses are promising in computer science. Nevertheless, we anticipate ongoing competition between AI output detectors and writers.”	Cheng et al. (2023)

Title	Implications	Author
A large-scale comparison of human-written versus ChatGPT-generated essays	# “Our results provide a strong indication that the fear many teaching professionals have is warranted: the way students do homework and teachers assess it needs to change in a world of generative AI models. For nonnative speakers, our results show that when students want to maximize their essay grades, they could easily do so by relying on results from AI models like ChatGPT. The very strong performance of the AI models indicates that this might also be the case for native speakers, though the difference in language skills is probably smaller. However, this is not and cannot be the goal of education.” # “Consequently, educators need to change how they approach homework. Instead of just assigning and grading essays, we need to reflect more on the output of AI tools regarding their reasoning and correctness. AI models need to be seen as an integral part of education, but one which requires careful reflection and training of critical thinking skills. Furthermore, teachers need to adapt strategies for teaching writing skills: as with the use of calculators, it is necessary to critically reflect with the students on when and how to use those tools. For instance, constructivists⁶² argue that learning is enhanced by the active design and creation of unique artifacts by students themselves. In the present case this means that, in the long term, educational objectives may need to be adjusted. This is analogous to teaching good arithmetic skills to younger students and then allowing and encouraging students to use calculators freely in later stages of education. Similarly, once a sound level of literacy has been achieved, strongly integrating AI models in lesson plans may no longer run counter to reasonable learning goals.”	Herbold et al. (2023)
A comparison of ChatGPT-generated articles with human-written articles	“It is worth noting that the articles generated may appear authentic to an untrained reader, including early-career researchers. Clinicians unaware of this might believe what was written and might use this in their clinical practice or even teach the trainees. This has the potential to cascade and spread disinformation. If used in clinical practice, it can result in increased patient. Despite these concerns, there may be potential benefits to using ChatGPT in scientific research, particularly in the generation of well-worded and comprehensible text, provided factual data is entered into it by the user. ChatGPT’s ability to quickly process and analyse large volumes of data could prove useful in certain fields of study. The format of articles generated by ChatGPT can be used as a draft template to write an expanded version of the article. The language, grammar, and spelling checks can be done for the articles simultaneously. However, it is important to ensure that any findings generated by ChatGPT are verified by human experts to ensure accuracy and reliability. Overall, while ChatGPT may hold some promise in the future for certain applications in scientific research, it is important to carefully consider its limitations and potential pitfalls, particularly in its current state. Future research should continue to investigate the role of ChatGPT in scientific publication and explore ways to mitigate its limitations and maximise its potential benefits while also maintaining the ethical	Ariyaratne et al. (2023)

Title	Implications	Author
	standards and quality of academic research. It is crucial that researchers exercise caution when utilising Chat- GPT as a tool for scientific writing and that any generated papers are carefully scrutinised by human experts before publication if it is to be used.”	
Reliability and Quality of the Nursing Care Planning Texts Generated by ChatGPT	“it is important to carefully evaluate the information provided by ChatGPT before using it in the treatment and care of patients by health care professionals. These results demonstrate the significant potential of ChatGPT for knowledge generation, text generation, and care planning in nursing. However, caution must be exercised, particularly in the clinical use of ChatGPT-generated texts. Nurses should thoroughly evaluate the accuracy and feasibility of these texts before incorporating them into patient treatment and care. Patient safety should always be paramount, and it should be remembered that such computer programs have limitations. We suggest that future studies be conducted that would compare texts generated by different AI algorithms and evaluate those using different tools to provide more comprehensive results.”	Dağcı et al. (2023)
The Death of the Short-Form Physics Essay in the Coming AI Revolution	“The results of this short study suggest modern AI systems such as davinci-003 and ChatGPT are capable of writing high quality Physics essays, capable of achieving a first-class score, for free and within a few seconds. These tools essentially represent a wide spread democratization of paid essay writing services whereby anyone can have davinci-003 or ChatGPT write an essay for them. We could argue that, in contrast to merely sounding a death-knell for certain kinds of assessments, this technology will once again force us to re-think assessment, and will confer far greater benefits in the end than the comparatively small drawback of having to redesign existing assessments. Obviating the abuse of these technologies to undermine assessments could be as simple as requiring the students to produce the work within an invigilated setting. In such cases, the moral of the story is that while the threat to authentic assessment is profound, simple and practical solutions likely already exist that won't necessarily require a total upheaval of current assessments themselves, but may force a change of context at least. Besides assessment, this technology allows users to generate innumerable, original examples, which can be shown by teachers within writing workshops. Similarly, as technologies such as ChatGPT mature we could see their use as one-on-one tutors. Although, the availability of knowledge is effectively already widely democratized. The full course contents for most degrees at Stanford and MIT have been available on their respective MOOC platforms for a decade. Yet enrolling as a student at both institutions is still very competitive and a typical online MOOC can expect a completion rate of 10% [12] which is far below that of a typical university degree. Thus it is not often clear in advance how new technology will impact Higher Education.”	Yeadon et al. (2022)
Can ChatGPT assist authors with abstract writing in medical	“ChatGPT performed inferiorly to the authors in generating scientific abstracts of CONSORT-A standards. However, it is commendable that ChatGPT produced authentic-looking	Hwang et al. (2024)

Title	Implications	Author
journals? Evaluating the quality of scientific abstracts generated by ChatGPT and original abstracts	abstracts with an appropriate structure and summary of the main report with minimal guidance and error. Given the exploratory nature of this study, definitive conclusion regarding GPT's efficacy and applicability in abstract writing remain yet to be established. Therefore, further investigations employing objective evaluation measures will be imperative to ascertain the true relevance and potential of ChatGPT in academic writing and medical research."	
Generative artificial intelligence: Can ChatGPT write a quality abstract?	"Advantages of AI in medicine will need to be balanced with possible legal, ethical and patient safety challenges. It seems likely that AI, which is constantly evolving and improving, will be widely incorporated into medicine in general and scientific writing in particular. Researchers, journals and funding agencies will need to rapidly come to terms with it."	Babl and Babl (2023)

Abuyaman (2023) suggests that ChatGPT has potential as a tool for scientific writing, particularly for non-English speaking scientists, but emphasizes the need for judicious use due to limitations in accuracy and bias. Cheng et al. (2023) highlight the ongoing competition between AI output detectors and writers, noting that AI may deceive in common subjects but shows limitations in specialized domains. Herbold et al. (2023) argue that the strong performance of AI models necessitates changes in education, proposing that AI tools should be integrated into teaching while emphasizing critical thinking. Ariyaratne et al. (2023) warn about the potential spread of misinformation through AI-generated content, but also see potential benefits in data processing if properly verified by human experts. Dağcı et al. (2023) stress the importance of careful evaluation of ChatGPT-generated information in healthcare, emphasizing patient safety. Yeadon et al. (2022) predict that AI will force a rethinking of assessment methods in education, suggesting that practical solutions like invigilated settings may be necessary. They also note AI's potential as a tutoring tool but caution that its impact on higher education remains uncertain. Hwang et al. (2024) call for further investigations to determine ChatGPT's true relevance and potential in academic writing and medical research. Babl and Babl (2023) anticipate

the widespread incorporation of AI in medicine and scientific writing, emphasizing the need for researchers, journals, and funding agencies to rapidly adapt while balancing advantages with legal, ethical, and patient safety challenges.

Output Characteristics: Authenticity

Purpose

There are three studies that investigate the authenticity of AI-generated content.

Table 61

Authenticity: Purpose

Title	Purpose	Research Question	Author
A Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing	“we aimed to explore the accuracy and authenticity of ChatGPT3’s academic writing and its similarity to natural human texts.”	“Firstly, to what extent the tool’s academic writing is accurate and authentic? Secondly, what is the degree of similarity between the generated virtual texts and natural human texts? Finally, what are the differences between the tool’s generated texts and natural human language in the context of academics?”	Mahyoob et al. (2023)
Perceptions of Professionalism and Authenticity in AI-Assisted Writing	“Our study had several purposes: (a) explore the degree to which AI-generated emails are perceived as having appropriate tone and style; (b) explore the degree to which AI-generated emails are perceived as effective; and (c)explore attitudes toward AI-assisted writing in the workplace.”	Not stated.	Coman and Cardon (2024)
How to cheat on your final paper: Assigning AI for student writing	“to share results from a pedagogical experiment that assigns undergraduates to cheat on a final class essay by requiring their use of text-generating AI software”	“What are the ethics of using AI this way? What counts as plagiarism? What are the conditions, if any, we should place on AI assistance for student writing? And how might working with AI change the way we think about writing, authenticity, and creativity?”	Fyfe (2022)

The three studies aim to investigate the capabilities and implications of AI-generated text, specifically ChatGPT, in academic and professional contexts. The first study (Mahyoob et al., 2023) explores the accuracy, authenticity, and human likeness of ChatGPT's academic writing. The second study (Coman & Cardon, 2024) assesses perceptions of AI-generated emails in terms of professionalism, effectiveness, and overall attitudes toward AI-assisted writing. The third study (Fyfe, 2022) experimentally examines the ethical and pedagogical implications of using AI for student writing, exploring concepts of plagiarism, AI assistance, and the potential impact on writing processes.

Results

The results from these studies are shown in Table 62.

Table 62

Authenticity: Results

Title	Results	Author
A Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing	“The results show that despite ChatGPT's exceptional capabilities, its serious defects are evident, as many issues in academic writing are raised. The major issues include information repetition, nonfactual inferences, illogical reasoning, fake references, hallucination, and lack of pragmatic interpretation. This paper provides a framework and multi-model principles for understanding the characteristics of academic language produced by ChatGPT, and these principles can be combined to produce a more complete and nuanced analysis. The findings indicate that ChatGPT failed to produce well-formed academic texts and articles with authentic and reliable information. The tool can generate human-like language with grammatically correct sentences and well-formed structure due to the large, trained corpus. However, professional academic analysis and writing are still inadequate, mainly if there are issues related to pragmatics and social norms and values. The tool can be an assistive medium to guide and provide information about the various genres. However, researchers should not depend on the tool to write articles; it can be identified easily as (natficial) a fabricated article and nonhuman language. This study recommended that, when using ChatGPT in academics, it is necessary to proceed cautiously.”	Mahyoob et al. (2023)
Perceptions of Professionalism and Authenticity in AI-Assisted Writing	“Across all messages, professionals consistently rated the AI-generated messages as professional, effective, efficient, confident, and direct. They rate sincerity and caring slightly lower in some disclosed conditions, particularly for	Coman and Cardon (2024)

Title	Results	Author
	ChatGPT-generated messages, suggesting the importance of tool selection when using generative AI for workplace writing. Those professionals who use AI more frequently for work are more likely to view AI-assisted writing as authentic, effective, and confidence-building. Implications for teaching business communication, including the need to address AI literacy, and suggestions for future research are provided.”	
How to cheat on your final paper: Assigning AI for student writing	“Ten students overall answered that they had not cheated. Some of them hedged that while they didn’t technically plagiarize, it still felt wrong. But about a quarter of the class had no need of such nuances. Intriguingly, they remained confident in their status as authors of their papers, and firmly defended their work as “not cheating” even within the conventional framework of plagiarism. They claimed to have done original, intellectual work—just not necessarily by writing. Instead, they explained their effort more in terms of assembly and editing. The AI could generate grammatical text, but until that text was sifted, organized, and stitched into an essay, the AI was not writing, <i>per se</i> .”	Fyfe (2022)

According to the first study (Mahyoob et al., 2023), ChatGPT demonstrated significant limitations in producing accurate and authentic academic writing, characterized by issues such as factual errors, illogical reasoning, and fabricated references, despite its ability to generate grammatically correct and structurally sound text. The second study (Coman & Cardon, 2024) revealed that professionals perceived AI-generated emails as professional and effective, but sincerity and caring were rated lower in some cases, particularly for ChatGPT-generated content. Frequent AI users were more accepting of AI-assisted writing. The results from the last study (Fyfe, 2022) suggest that students assigned to use AI for a final paper exhibited varying responses, with some feeling unethical, while others confidently defended their work as original despite acknowledging significant AI involvement in the writing process.

Implications

Table 63 provides the implications of the studies on authenticity.

Table 63*Authenticity: Implications*

Title	Implications	Author
A Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing	“The findings indicate that ChatGPT failed to produce well-formed academic texts and articles with authentic and reliable information. The tool can generate human-like language with grammatically correct sentences and well-formed structure due to the large, trained corpus. However, professional academic analysis and writing are still inadequate, mainly if there are issues related to pragmatics and social norms and values. The tool can be an assistive medium to guide and provide information about the various genres. However, researchers should not depend on the tool to write articles; it can be identified easily as (natficial) a fabricated article and nonhuman language. This study recommended that, when using ChatGPT in academics, it is necessary to proceed cautiously.”	Mahyoob et al. (2023)
Perceptions of Professionalism and Authenticity in AI-Assisted Writing	“Our study also carries implications for teachers of business communication. As generative AI is more widely adopted for even routine workplace communication, teachers should consider how to best prepare students to thrive in an environment where generative AI use is the norm. One key implication is that teachers should reevaluate the role of grammar instruction in business communication curriculum. Good grammar is a primary marker of professionalism and other positive attributes, making grammar instruction a core concern of professional writing. However, producing grammatically correct prose is one of the greatest strengths of generative AI tools. These tools are far more powerful than spell checkers of the past, and they are soon to be native across writing platforms. Widespread adoption of these tools suggests the need to reassess the role of grammar instruction in the classroom. Prior studies have shown that generative AI is a powerful skill-leveler, so teachers have an opportunity to help students leverage these tools to improve their grammar and therefore their perceived professionalism. Secondly, if professionals have access to multiple tools that can dynamically produce and edit text for correctness, the differentiating factor for business communicators may be the ability to identify context and adjust generated text to match the purpose and needs of the audience. We found that AI-generated messages scored relatively lower for relationship-oriented aspects of writing like sincerity and caring. This research suggests that instructors should explore ways to emphasize these aspects in their communication classes. In a world where everyone produces grammatically correct writing, EQ and empathy may be the differentiating skills that business communication students need to stand out. A third implication for business communication instructors is the need to emphasize AI literacy in their courses. An important part of AI literacy is the ability to understand various AI platforms and the value of their output. This study suggests that there is difference between two of the platforms—ChatGPT and Google—in the context of routine but relationally significant emails. Instructors should find ways to allow students to compare and contrast the output of	Coman and Cardon (2024)

Title	Implications	Author
	various platforms. This is particularly important because most university students have not experimented with more than one platform. A year into the gAI era, a November 2023 survey of university students across a dozen universities found that while 83% had used ChatGPT, only 11% had used Google Bard, 3% had tried Google’s Help Me Write, 2% had used Dall-E, 2% had used Claude, 1% had tried MidJourney, and 11% had tried another platform.”	
How to cheat on your final paper: Assigning AI for student writing	“Whether or not writing with AI becomes cheating is a much more complicated question than software purveyors and institutional plagiarism policies would have us believe. And humanities scholars can play an important role in helping explain and navigate those complexities. For starters, we should reframe that originating question in a few different ways. How do institutionalized definitions of plagiarism block the possibilities of writing with computational assistance or even partnership? What kinds of creative or critical frameworks do plagiarism policies impose or preclude, especially compared to what we really value for students? How are these values supported (or not) by the learning activities that AI seems likely to impact (like writing essays)? How might we modify these activities or emphasize different outcomes? How do plagiarism regimes or AI platforms differently affect populations of students? Are there new modes of creativity, critical reasoning, rhetoric, assemblage, and expression that computational assistance helps us identify? How might AI-assisted writing exercises align with frameworks for teaching AI literacy? ¹⁶ What are the risks and harms of these technologies that we need to factor in? And how can we bring different disciplinary perspectives on these questions to bear upon the development and implementation of AI?”	Fyfe (2022)

The first study (Mahyoob et al., 2023), highlights the significant limitations of ChatGPT in producing reliable and authentic academic content, emphasizing the need for cautious use and human oversight. The second study (Coman & Cardon, 2024) reveals that while AI-generated text can be perceived as professional and effective, human qualities like empathy and contextual understanding remain crucial for effective communication. Consequently, educators must adapt to this new reality by focusing on AI literacy, critical thinking, and human-centered communication skills. The third study (Fyfe, 2022) underscores the complex ethical and pedagogical challenges posed by AI in academic writing, questioning traditional notions of

plagiarism and authorship and calling for a re-evaluation of writing assessment and instruction.

Output Characteristics: Creativity

Purpose

There are two studies that investigate this issue of creativity.

Table 64

Creativity: Purpose

Title	Purpose	Research Question	Author
A Proposed Framework for Human-like Language Processing of ChatGPT in Academic Writing	• To find out the level of creativity indicators (fluency, flexibility, and narrative originality) of AI systems and students, based on the sample of the Creative Imagination Test for Adults (PIC-A) by Artola et al. (2012). • To compare the creativity of the AI systems with that of the students, based on the indicators of fluency, flexibility, and narrative originality. • To compare the scores of the creativity indicators obtained by the students in the two phases of application of the tests, the first without any type of aid and the second using an AI system (ChatGPT)."	Not stated.	Vicente-Yagüe Jara et al. (2023)
How to cheat on your final paper: Assigning AI for student writing	"to share results from a pedagogical experiment that assigns undergraduates to cheat on a final class essay by requiring their use of text-generating AI software"	"What are the ethics of using AI this way? What counts as plagiarism? What are the conditions, if any, we should place on AI assistance for student writing? And how might working with AI change the way we think about writing, authenticity, and creativity?"	Fyfe (2022)

The first study (Vicente-Yagüe Jara et al., 2023) aims to compare the creativity of AI systems and humans, using specific indicators, while the second

study (Fyfe, 2022) explores the ethical implications of students using AI for essay writing and the potential impact on writing practices.

Results

The results from these two studies are shown in Table 65.

Table 65

Creativity: Results

Title	Results	Author
Writing, creativity, and artificial intelligence. ChatGPT in the university context	“The findings reveal statistically significant differences between the AIs and the students in the indicators of fluency, flexibility, and narrative originality in Game 2. Furthermore, significant differences are found between students' pre-test and post-test scores in fluency, flexibility, and narrative originality in Game 2 and in fluency in Game 3. Finally, the assistance provided by AI in writing tasks and verbal creativity is highlighted, and this should be considered in language teaching; in any case, AI cannot replace human intelligence and creativity. Firstly, it is noticeable that the means obtained by the AIs in the three indicators of game 2 are higher than the means indicated in the instrument manual, while, on the contrary, the means achieved in the indicators of game 3 are lower. In addition, after comparing the creativity of the AI systems with that of the students, the AI scores on the indicators in game 2 significantly outperformed the students' scores, while in game 3 the creativity indicators showed similar levels in both groups. This result is consistent with the above, as AI scores were higher in game 2. With respect to game 3, the students outperform, although not significantly, the AI in narrative originality. Furthermore, the significant improvement in the student's written productions after the use of ChatGPT has been highlighted, specifically in all the indicators of game 2 and the fluency of game 3. The potential of human-AI collaboration in the writing process is highlighted, as “AI is a means, not an end” (Breton, 2021) and it has been a tool that has worked as an assistant to the students in game 2, which is aligned with the productive possibilities of AI technology, whereas the flexibility and originality of game 3 has not, as it has already been shown that the AIs have not been able to imagine the effects of a fantasy situation.”	Vicente-Yagüe Jara et al. (2023)
How to cheat on your final paper: Assigning AI for student writing	“Ten students overall answered that they had not cheated. Some of them hedged that while they didn't technically plagiarize, it still felt wrong. But about a quarter of the class had no need of such nuances. Intriguingly, they remained confident in their status as authors of their papers, and firmly defended their work as “not cheating” even within the conventional framework of plagiarism. They claimed to have done original, intellectual work—just not necessarily by writing. Instead, they explained their effort more in terms of assembly and editing. The AI could generate grammatical	Fyfe (2022)

Title	Results	Author
	text, but until that text was sifted, organized, and stitched into an essay, the AI was not writing, per se.”	

Borras-Comes et al. (2023) find that AI systems outperform students in certain creativity indicators, particularly in Game 2 of their study, showing higher fluency, flexibility, and narrative originality. However, in Game 3, which involves more imaginative tasks, students and AI perform similarly, with students slightly outperforming AI in narrative originality. The study also reveals significant improvements in students' written productions after using ChatGPT, especially in all indicators of Game 2 and fluency in Game 3. The authors highlight the potential of human-AI collaboration in the writing process, viewing AI as a tool that can assist students, particularly in structured tasks.

Watkins (2023) reports that in a study of students using AI for writing assignments, about three-quarters of the students admit to using AI, with varying interpretations of whether this constitutes cheating. Interestingly, even those who used AI remain confident in their authorship, defending their work as original. These students describe their effort in terms of assembly and editing rather than traditional writing, suggesting that they view their role as curating and organizing AI-generated content rather than producing the text from scratch. This indicates a shift in how students perceive the writing process and their role as authors in the age of AI.

Implications

Table 66 provides the implications of the studies on creativity.

Table 66

Creativity: Implications

Title	Implications	Author
Writing, creativity, and artificial	“...in language teaching; in any case, AI cannot replace human intelligence and creativity. the use of AI should be	Vicente-Yagüe

Title	Implications	Author
intelligence. ChatGPT in the university context	aimed at empowering teachers and not replacing or displacing them (UNESS0, 2022b). An AI Literacy plan is required to train teachers in both technical skills and ethical-philosophical debates (Flores-Vívar G García-Iñalvo, 2023; UNESS0, 2022a). Similarly, changes in educational practice in the coming years will be determined by AI developments and will need to rely on research in Tull collaboration with teachers, educational leaders, and learners to ensure that appropriate educational policies are put in place.”	Jara et al. (2023)
How to cheat on your final paper: Assigning AI for student writing	“Whether or not writing with AI becomes cheating is a much more complicated question than software purveyors and institutional plagiarism policies would have us believe. And humanities scholars can play an important role in helping explain and navigate those complexities. For starters, we should reframe that originating question in a few different ways. How do institutionalized definitions of plagiarism block the possibilities of writing with computational assistance or even partnership? What kinds of creative or critical frameworks do plagiarism policies impose or preclude, especially compared to what we really value for students? How are these values supported (or not) by the learning activities that AI seems likely to impact (like writing essays)? How might we modify these activities or emphasize different outcomes? How do plagiarism regimes or AI platforms differently affect populations of students? Are there new modes of creativity, critical reasoning, rhetoric, assemblage, and expression that computational assistance helps us identify? How might AI-assisted writing exercises align with frameworks for teaching AI literacy? ¹⁶ What are the risks and harms of these technologies that we need to factor in? And how can we bring different disciplinary perspectives on these questions to bear upon the development and implementation of AI?”	Fyfe (2022)

Vicente-Yagüe Jara et al. (2023) emphasize that while AI is a powerful tool in language teaching, it cannot replace human intelligence and creativity. They advocate for an AI Literacy plan to train teachers in both technical skills and ethical-philosophical debates. The authors stress that AI should empower teachers rather than replace them. They predict that educational practices will be significantly influenced by AI developments in the coming years, calling for collaborative research involving teachers, educational leaders, and learners to ensure appropriate educational policies are implemented.

Fyfe (2022) argues that the question of whether writing with AI constitutes cheating is more complex than current institutional plagiarism policies suggest. He

proposes that humanities scholars should play a crucial role in navigating these complexities. Fyfe raises several important questions, including how current plagiarism definitions might hinder the potential of AI-assisted writing, what creative or critical frameworks are imposed or precluded by these policies, and how AI might affect different student populations. He suggests exploring new modes of creativity, critical reasoning, and expression that AI assistance could facilitate. Fyfe also emphasizes the need to consider the risks and harms of these technologies and to incorporate diverse disciplinary perspectives in developing and implementing AI in education.

Output Characteristics: Originality

Purpose

There are two studies that investigate this issue of originality.

Table 67

Originality: Purpose

Title	Purpose	Research Question	Author
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots’ Proficiency and Originality in Scientific Writing for Humanities	“With this in mind, the aim of this article was to design and conduct a test of AI chatbots’ abilities in scientific writing in the humanities. First, we were interested in their ability to generate correct answers to complex scientific questions. Second, we tested their capacity to generate original scientific contributions in humanities research. Since AI chatbots are developing at a staggering pace, our results apply to the state of affairs in the third quarter of 2023 (23Q3).”	Not stated.	Lozic and Štular (2023)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“to shed light on the challenges faced by English as a Second Language (ESL) lecturers in identifying AI-generated texts, and highlighting the skills and	“(1) How effective are current plagiarism detectors in providing a reliable measure of detection for identifying originality in academic writing and distinguishing human-generated text from AI-generated	Alexander et al. (2023)

Title	Purpose	Research Question	Author
	resources needed to enhance their detection capabilities”	text?; (2) What criteria are employed by ELS lecturers to assess academic writing?; and (3) How effective are these criteria in distinguishing human-generated text from AI-generated text?”	

The first study (Lozic & Štular, 2023) aims to evaluate AI chatbots' abilities to generate correct answers to complex scientific questions and produce original scientific contributions in humanities research as of the third quarter of 2023. The second study (Alexander et al., 2023) investigates the challenges faced by ESL lecturers in identifying AI-generated texts and explores the effectiveness of current plagiarism detectors and the criteria used by lecturers in distinguishing human-generated from AI-generated academic writing.

Results

The results from these two studies are shown in Table 68.

Table 68

Originality: Results

Title	Results	Author
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots’ Proficiency and Originality in Scientific Writing for Humanities	“In the quantitative test, ChatGPT-4 scored near the passing grade (–5), whereas ChatGPT-3.5 (–18), Bing (–21), and Bard (–31) were not far behind. Claude 2 (–75) and Aria (–80) scored much lower. In the qualitative test, all AI chatbots, but especially ChatGPT-4, demonstrated proficiency in recombining existing knowledge, but all failed to generate original scientific content. As a side note, our results suggest that with ChatGPT-4, the size of large language models has reached a plateau. Furthermore, this paper underscores the intricate and recursive nature of human research. This process of transforming raw data into refined knowledge is computationally irreducible, highlighting the challenges AI chatbots face in emulating human originality in scientific writing. Our results apply to the state of affairs in the third quarter of 2023. In conclusion, while large language models have revolutionised content generation, their ability to produce original scientific contributions in the humanities remains limited.”	Lozic and Štular (2023)
Who Wrote This Essay? Detecting AI-Generated Writing in	“Findings reveal that participants tended to exploit a deficit model of assessment that focuses on error as an indicator of L2 writing output, with high levels of technical and	Alexander et al. (2023)

Title	Results	Author
Second Language Education in Higher Education	grammatical accuracy and sophisticated language use as indicators of AI-generated text. Moreover, findings suggest that limited awareness of the characteristics and metrics used by ChatGPT, as well as lack of attention to the veracity of facts and references generated by ChatGPT, were features of participants' evaluations. This study identifies the growing challenges encountered by ESL lecturers and underlines the need for digital literacy training, targeted professional development, the use of advanced detection tools and a review of assessment policies and practices in relation to AI. Additionally, this study highlights the importance of reviewing and reinforcing institutional policies and practices that safeguard academic integrity and ensure quality higher education.”	

Lozic and Štular (2023) find that in a quantitative test of AI chatbots' scientific writing abilities for humanities, ChatGPT-4 scores near the passing grade, while ChatGPT-3.5, Bing, and Bard perform slightly worse. Claude 2 and Aria score significantly lower. In qualitative testing, all AI chatbots, particularly ChatGPT-4, show proficiency in recombining existing knowledge but fail to generate original scientific content. The authors note that the size of large language models may have reached a plateau with ChatGPT-4. They emphasize that the process of transforming raw data into refined knowledge is computationally irreducible, highlighting AI's limitations in replicating human originality in scientific writing.

Alexander et al. (2023) report that ESL lecturers tend to use a deficit model of assessment when evaluating AI-generated writing, focusing on errors as indicators of L2 writing output. They find that high levels of technical and grammatical accuracy, along with sophisticated language use, are perceived as indicators of AI-generated text. The study reveals that participants have limited awareness of ChatGPT's characteristics and metrics, and pay little attention to the veracity of facts and references generated by AI. The authors highlight the growing challenges faced by ESL lecturers in distinguishing AI-generated text from student writing.

Implications

Table 69 provides the implications of these two studies.

Table 69

Originality: Implications

Title	Implications	Author
Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities	“Additional shortcomings of the AI-generated content that we found in our test include reasoning errors, hallucinations, and biases. Reasoning errors refer to the chatbot’s inability to critically evaluate and link cause-and-effect relationships, as evidenced by several historical inaccuracies regarding the migration patterns and historical periods of the Slavs. Hallucinations denote confident but unsubstantiated claims by the AI, such as invented references and inaccurate dates. Our test also reveals significant biases in the content generated by the AI. These biases manifest as language bias, favouring English sources over other relevant languages; neo-colonial bias, displaying a preference for Western authors; and citation bias, skewing towards older and more highly cited publications. These findings highlight that, despite their technological prowess, AI chatbots remain reliant on their training data, echoing and even amplifying existing biases and knowledge gaps. Given that they veer towards past data, including, in our test, a book originally published in 1895, they are likely to be too conservative in their recommendations. Since the listed deficiencies are relatively inconspicuous compared to, for example, gender or racial biases, it is unlikely that they will be remedied in the foreseeable future by resourceintensive reinforcement learning from human feedback. These biases are among the key concerns with the use of AI chatbots in scientific writing, as most of them are less likely to be highlighted in the review processes.”	Lozic and Štular (2023)
Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education	“Concerning the implications for policy and practice, the results demonstrate that to improve practice and to safeguard academic integrity, there is a need for digital training for educators with respect to the use of AI tools in the ESL classroom and the use of advanced detection tools. There is also a need for revision of assessment policies and procedures and the development of suitable rubrics and assessment criteria. Lastly, lecturers’ tendency to confuse student-written essays for AI-generated text points out the need for lecturers to be trained on the features of AI-created texts and the differences between them and humanwritten essays.”	Alexander et al. (2023)

The study reveals significant biases in AI-generated content, including language bias favoring English sources, neo-colonial bias preferring Western authors, and citation bias skewing towards older and more highly cited publications.

Lozic and Štular (2023) emphasize that these biases reflect and potentially amplify existing knowledge gaps and biases in the AI's training data. They caution that AI chatbots may provide overly conservative recommendations due to their reliance on past data.

Importantly, Lozic and Štular (2023) argue that these deficiencies are relatively inconspicuous compared to more obvious biases like gender or racial biases. As a result, they suggest that these issues are unlikely to be addressed in the near future through resource-intensive reinforcement learning from human feedback. This raises concerns about the use of AI chatbots in scientific writing, as these biases may not be easily detected in review processes.

Alexander et al. (2023) focus on the implications of AI-generated writing in second language education in higher education. They emphasize the need for digital training for educators in using AI tools in the ESL classroom and employing advanced detection tools to safeguard academic integrity. The authors call for a revision of assessment policies and procedures, as well as the development of suitable rubrics and assessment criteria to address the challenges posed by AI-generated content.

A crucial finding from Alexander et al. (2023) is the tendency of lecturers to confuse student-written essays with AI-generated text. This underscores the need for specific training for lecturers to recognize the features of AI-created texts and distinguish them from human-written essays. Such training is essential to ensure fair assessment and maintain the integrity of academic writing in second language education.

In conclusion, both studies highlight the complex challenges posed by AI in academic writing and assessment. They emphasize the need for ongoing education

and training for both educators and students, as well as the development of new policies and assessment criteria to address the unique challenges presented by AI-generated content in academic contexts.

User Behavior and Pedagogical Implications of AI-Assisted Writing

The section includes a range of interconnected human factors, including the *acceptance and adoption* of AI tools, *users and uses*, *attitudes and challenges* towards AI assistance, *trust* in AI systems, writers' *self-efficacy* and *motivation*, and the balance between AI support and writer *independence and proficiency*, and all of these studies collectively shape the cognitive and affective dimensions of AI-enhanced writing experiences (Figure 8).

Figure 8

User Behavior and Pedagogical Implications of AI-Assisted Writing



Acceptance

Purpose

There are three studies that investigate the concept of acceptance.

Table 70*Acceptance: Purpose*

Title	Purpose	Research Question	Author
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“(1) to capture the collective wisdom of business communication instructors / (2) to develop an expert-driven framework of AI literacy for communicators”	“1. What concerns do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 2. To what degree do business communication instructors experience anxiety towards teaching or integrating AI-assisted writing into their courses? 3. What perceived benefits do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 4. To what degree do business communication professors perceive AI-assisted writing (e.g., ChatGPT) will be accepted in the near future for workplace writing?”	Cardon et al. (2023)
To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model	“to investigate doctoral students' acceptance toward ChatGPT in writing and the factors that influence it”	“(1) how is the doctoral students' acceptance intention to ChatGPT in writing? (2) what factors may influence doctoral students' acceptance intention to ChatGPT in writing?”	Zou and Huang (2023)
A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning	“to investigate the use and acceptance of ChatGPT-4 in students' EFL writing”	“1. Are there any significant differences in the writing performance of the experimental group compared with the control group after the use of ChatGPT-4? 2. Are there any significant differences in the writing performance before and after the employment of ChatGPT-4? 3. Would students' perceptions toward ChatGPT-4 use significantly change after the implementation of ChatGPT-4?”	Boudouaia et al. (2024)

Cardon et al. (2023) seeks to capture the collective wisdom of business communication instructors and develop an expert-driven framework of AI literacy for communicators. Zou and Huang (2023) investigate doctoral students' acceptance of ChatGPT in writing and the factors influencing it. Boudouaia et al. (2024)

examines the use and acceptance of ChatGPT-4 in students' EFL writing, focusing on its impact on writing performance and students' perceptions.

Results

The results from this study are shown in Table 71.

Table 71

Acceptance: Results

Title	Results	Author
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“This study of 343 communication instructors revealed a collective view that AI-assisted writing will be widely adopted in the workplace and will require significant changes to instruction. Key perceived challenges include less critical thinking and authenticity in writing. Key perceived benefits include more efficiency and better idea generation in writing. Students will need to develop AI literacy--composed of application, authenticity, accountability, and agency--to succeed in the workplace.”	Cardon et al. (2023)
To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model	“The findings further indicated that attitude was a significant predictor of behavioural intention to use ChatGPT in writing and mediated the impacts of perceived usefulness and perceived ease of use on it. Perceived ease of ChatGPT use was in turn influenced by students' past ChatGPT use experience. This study provides powerful evidence for the applicability of Technology Acceptance Model in the acceptance of ChatGPT in writing. The results have significant implications for leveraging ChatGPT for writing in higher education.”	Zou and Huang (2023)
A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning	“The results showed that the experimental group (EG) outperformed the control group (CG). Besides, the findings revealed that students in the EG in post-test outperformed their pre-test scores. The findings also revealed substantial improvements in the EG's views of perceived usefulness, perceived ease of use, attitudes, and behavioral intention. According to the results, ChatGPT-4 helped boost students' EFL writing skills, which ultimately led to their acceptance. Students appear particularly interested in ChatGPT-4 because of its potential usefulness in putting what they learn about EFL writing into practice. Some suggestions and recommendations were provided.”	Boudouaia et al. (2024)

Cardon et al. (2023) reveals that communication instructors anticipate widespread adoption of AI-assisted writing in the workplace, necessitating changes in instruction and the development of AI literacy among students. Zou and Huang (2023) demonstrate that doctoral students' attitude significantly predicts their

intention to use ChatGPT in writing, with perceived usefulness and ease of use as mediating factors. Boudouaia et al. (2024) reports that the experimental group using ChatGPT-4 outperforms the control group in EFL writing, showing improvements in both writing performance and perceptions towards AI use in writing.

Implications

Table 72 provides the implications suggested by these studies.

Table 72

Acceptance: Implications

Title	Implications	Author
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“Develop AI literacy. Oversee creative and innovative learning and teaching. Operate in communities of practice. Solicit cross-disciplinary input. Set community-based vision and policy. Provide resources for innovation. Pause institutional constraints and build culture.”	Cardon et al. (2023)
To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model	“...technology developers can keep simplifying and optimizing the operation of ChatGPT based on user feedback and provide comprehensible instructions or use cases regarding how to apply ChatGPT to write more effectively and ethically. Instead of prohibiting the use of ChatGPT in writing, policy makers need to take into Practically, the results of the study could also generate useful implications for technology developers, policy-makers, writing teachers, and doctoral students to leverage ChatGPT for the teaching and learning of writing. Doctoral students' strong intention to use ChatGPT in writing suggest that ChatGPT may augment its function as an educational tool for writing in higher education. Considering the significant and strong effect of attitude on students' behavioural intentions to use ChatGPT in writing, it is of necessity for educational institutions, writing teachers, and technology developers to be aware of students' attitudes and increase their positive evaluation of and affective reactions towards using ChatGPT in writing. For instance, technology developer can make the usage of ChatGPT more innovative, enjoyable and interesting so as to create more positive attitudes and boost learners' intrinsic motivation to use ChatGPT in writing. Given the increasing concerns for information, ethical and learning risks associated with ChatGPT (e.g., Barrot, 2023; Dwivedi et al., 2023) and doctoral students' strong intention to use ChatGPT for writing, measures must be taken to mitigate such negative impacts of ChatGPT on doctoral students. For example, technology developers can strengthen the quality control of generated responses. Similarly, writing teachers need to provide trainings on effective, ethical and responsible use of ChatGPT in writing. Besides, perceived ease of use and perceived usefulness are found to have a	Zou and Huang (2023)

Title	Implications	Author
A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning	significant influence on students' attitude, which could further exert an effect on students' intentions to use ChatGPT in writing. The sequential and circular influential relationship among the variables implies a need for technology developers to increase the usefulness and ease of using ChatGPT in writing to make it more functional and user-friendly. For consideration the students' voice and align their educational needs with the AI tool (EDUCAUSE, 2023). For writing teachers and institutional administrators, efforts to integrate ChatGPT in writing courses or training programs are needed to capitalize on ChatGPT's affordances for writing and improve students' ability to use ChatGPT as an effective writing assistant tool. Given the significant effect of past ChatGPT experience on perceived ease of ease, instructing doctoral students to increase their use of ChatGPT, and reflect upon and communicate the skills for utilizing ChatGPT to promote writing performance could be an effective way to develop their expertise in ChatGPT. Also, doctoral students can experiment with ChatGPT in a conscious manner, and record their hands-on experience to continuously improve the capability for effective and ethical use of ChatGPT for writing." "EFL teachers are encouraged to incorporate ChatGPT-4 into their lessons to provide students an opportunity to participate. Teachers should model the ethical use of ChatGPT-4 and give students the skills they need to confidently adopt the tool in the classroom. It is possible that holding workshops, providing training, and rolling out ChatGPT-4 in stages could all help achieve this goal. Teachers should encourage their students to have good intentions. Potentially enhanced by demonstrating its usefulness and advantages, the use of ChatGPT-4 to study EFL writing has the potential to be more effective. Furthermore, it is presumed that course or exercise designers in language curricula can emphasize and promote students' positive attitudes, sense of ease, and ability to implement via the use of ChatGPT-4. This allows for greater linguistic and technical fluency in the language generally. Courses should be flexible so that students can take charge of their own learning experiences. One approach would be to offer students the background and resources they need to effectively use ChatGPT-4 in and out of the classroom."	Boudouaia et al. (2024)

Cardon et al. (2023) emphasize developing AI literacy and fostering innovative teaching practices. They advocate for creating communities of practice, seeking cross-disciplinary input, setting community-based policies, and providing resources for innovation while temporarily suspending institutional constraints to build a supportive culture.

Zou and Huang (2023) recommend that technology developers simplify ChatGPT's operation and provide clear instructions for effective and ethical use. They advise policymakers to consider students' needs when integrating AI tools. For writing teachers and administrators, they suggest incorporating ChatGPT into courses and training programs to leverage its benefits. They also encourage doctoral students to experiment with ChatGPT consciously and reflect on their experiences to improve their skills in using it for writing.

Boudouaia et al. (2024) encourage EFL teachers to incorporate ChatGPT-4 into their lessons, model its ethical use, and provide students with the necessary skills. They suggest conducting workshops and implementing ChatGPT-4 in stages. They recommend that course designers emphasize promoting positive attitudes and ease of use among students. They also advise creating flexible courses that allow students to take charge of their learning experiences and provide resources for effective use of ChatGPT-4 both in and outside the classroom.

Adoption

Purpose

There are three studies that aim at understanding the adoption of AI tools by the users from various fields.

Table 73

Adoption: Purpose

Title	Purpose	Research Question	Author
Embrace or resist? Drivers of artificial intelligence writing software adoption in academic and non-academic contexts	“to investigate the perspectives of using artificial intelligence writing software (AIWS) in professional contexts, focusing on academic and non-academic writers”	“What factors impact the academic and non-academic groups’ readiness to integrate these tools into their writing practices, and what ethical concerns or implications might arise from such incorporation?”	Papakonstantinidis et al. (2024)

Title	Purpose	Research Question	Author
Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective	“This research aims to understand students’ viewpoints on their use or avoidance of AI in composing academic essays, seeking to reveal the reasons behind their choice to embrace or forgo AI tools for writing tasks. Additionally, the study identifies the prevalent AI applications students typically use in the process of crafting academic essays.”	Not stated.	Malik et al. (2023)
Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing	“to explore the students' experience with ChatGPT in providing scaffolding for writing essays. It also discloses students' appraisals of utilising ChatGPT”	Not stated.	Nugroho et al. (2024)

These studies aim to understand the adoption and use of artificial intelligence writing software (AIWS) in different contexts. The first study (Papakonstantinidis et al., 2024) focuses on factors influencing the integration of AIWS in both academic and non-academic writing, while the second and third studies (Malik et al., 2023; Nugroho et al., 2024) specifically examine students' perspectives and experiences with AI in academic essay writing.

Results

The results from these studies are shared in Table 74.

Table 74

Adoption: Results

Title	Results	Author
Embrace or resist? Drivers of artificial intelligence writing software adoption in academic and non-academic contexts	“Findings yield insights into non-academic writers’ readiness and implications of AIWS adoption. Business, non-academic professionals view AIWS as a tool for efficiency and content quality, while writers in academic contexts express concerns about biases, manipulation, and job displacement. The study contributes to AIWS understanding, benefiting developers, educational institutions, and content creators, and elucidates differing attitudes and age	Papakonstantinidis et al. (2024)

Title	Results	Author
	dynamics between academics and professionals. The research underscores the multifaceted influence of AIWS, providing a foundation for future exploration in this emerging domain, as well as practical applications for industries and educational institutions. / These findings highlight the complex interplay between cognitive factors, behavioral outcomes, and age in the acceptance and utilization of AI writing software. The stronger direct effect observed among academics suggests that their acceptance of AI writing tools is primarily driven by cognitive factors. On the other hand, professionals' acceptance is influenced by affective factors mediated by cognitive factors. Additionally, the moderation effect of age further emphasizes the need to consider age-related differences in the adoption of AI writing software, particularly among professionals. Understanding these nuanced relationships contributes to a comprehensive understanding of the factors influencing the acceptance and usage of AI writing software in different professional contexts."	
Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective	"Findings indicated a positive reception of AI-powered writing tools, with students acknowledging their benefits in grammar checks, plagiarism detection, language translation, and essay outlines. AI was found to enhance students' writing abilities, self-efficacy, and understanding of academic integrity. However, some students expressed concerns about potential impacts on creativity, critical thinking, and ethical writing practices. The study emphasized a balanced approach to AI integration, where AI collaborates with human authors, and also identified popular AI tools used by Indonesian students. As such, the research highlighted AI's significant role in supporting academic writing while preserving human creativity and critical thinking, underscoring the importance of maintaining a balanced integration to sustain human ingenuity and critical thought in academic discourse."	Malik et al. (2023)
Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing	"The findings revealed that students derived advantages from using ChatGPT for translation, writing accuracy, writing efficiency, idea generation, and practical use. This study also reveals students' criticism in terms of inaccurate information generated as well as the possibility of students committing academic dishonesty when using ChatGPT. This study indicates that although students use ChatGPT to help them write their essays, they have the awareness to fact-check the information to avoid committing academic dishonesty."	Nugroho et al. (2024)

The first study (Papakonstantinidis et al., 2024) explores attitudes toward AI writing software in academic and non-academic settings, finding varied acceptance

based on profession and age. The second study (Malik et al., 2023) reveals that students generally hold positive views about AI for writing tasks and recognize its benefits for grammar, plagiarism, and idea generation while expressing concerns about potential negative impacts. The third study (Nugroho et al., 2024) focuses on students' experiences with ChatGPT and identifies both advantages, like improved writing efficiency and accuracy, as well as potential drawbacks, such as the generation of inaccurate information and ethical concerns.

Implications

Table 75 provides the implications of these studies.

Table 75

Adoption: Implications

Title	Implications	Author
Embrace or resist? Drivers of artificial intelligence writing software adoption in academic and non-academic contexts	“Professional acceptance of AI writing software has significant practical implications for a variety of industries. To design AI writing software that meets the needs and preferences of its users, organizations and developers need to understand professionals’ attitudes and levels of acceptance towards these tools. It is possible for developers to improve the functionality, usability, and effectiveness of AI writing tools by incorporating user feedback and addressing ethical issues. As a result, productivity, content quality, and writing processes can be improved. Organizations can also use the findings of this study to determine how AI writing tools should be integrated into their workflows to achieve the best results. By assessing the potential benefits and challenges of AI writing software, they can make informed decisions regarding resource allocation, training programs, and workflow restructuring. The multifaceted nature of acceptance and utilization can be addressed by stakeholders by creating interventions, training, and supportive environments that take cognition, affective, and aging factors into account. This is why a comprehensive approach to AIWS could maximize its benefits and contribute to improved writing performance across a variety of settings. Future research can build upon this study’s findings by employing qualitative research methods to explore professionals’ attitudes and perceptions in greater depth. In-depth interviews, focus group discussions, and case studies can provide rich insights into the specific concerns, challenges, and opportunities professionals encounter when	Papakonstantinidis et al. (2024)

Title	Implications	Author
	incorporating AI writing tools into their careers. Such studies and taxonomies (Gkinko & Elbanna, 2022) can explore the experiences of professionals across different industries and job roles, uncovering valuable nuances and identifying best practices for successful implementation.”	
Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective	“As such, the practical implication of this study lies in illuminating the transformative impact and valued utility of AI-powered tools in academic writing, providing educators, developers, and students with insights and actionable guidance on the balanced integration and optimal utilization of AI technologies to enhance writing processes while preserving human creativity and critical thinking.”	Malik et al. (2023)
Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing	“Students should be taught how to use the ChatGPT to support their writing (Dwivedi et al., 2023). Rather than banning ChatGPT in language classrooms, instructors should find ways to integrate AI to ensure its mindful use in education (Barrot, 2023). Rather than treating ChatGPT as a threat, instructors should consider it as a tool for language learners' intellectual intensification (Dwivedi et al., 2023). To achieve this, instructors should consider assessments and policies regarding the use of ChatGPT in writing courses to ensure academic integrity (Imran & Almusharraf, 2023). Educators should prepare for a time when ChatGPT or similar advanced technological tools become more capable, accurate, and less reliant on prompt quality. This advancement will not only revolutionise the way students learn and acquire knowledge but will also change the current assessment method. Hence, educators should begin to consider innovative assessments and policies for the AI era.”	Nugroho et al. (2024)

The implications of these studies underscore the complex and multifaceted nature of AI writing software adoption. The first study (Papakonstantinidis et al., 2024) highlights the need for tailored AI tools that address specific professional needs and concerns, emphasizing the importance of user-centered design and ethical considerations. The second study (Malik et al., 2023) emphasizes the potential of AI to enhance students' writing abilities while preserving human creativity and critical thinking, suggesting a collaborative approach to AI integration in education. The third study (Nugroho et al., 2024) underscores the need for careful integration of AI tools in education, including the development of AI literacy, ethical guidelines, and

innovative assessment methods to ensure responsible and effective use of AI in writing.

Attitudes and Challenges

Purpose

There are three studies that focus on the attitudes and challenges towards AI.

Table 76

Attitudes and Challenges: Purpose

Title	Purpose	Research Question	Author
Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students	“to test the performance of high school teachers and students on an AI Identification Test that assessed their ability to distinguish between short English essays written by ChatGPT versus high school students / to examine possible predictors of the ability to make this distinction, such as prior experience with ChatGPT / to measure their broader attitudes about the use of ChatGPT in academic settings, including their evaluations of its impacts on education and what uses should be considered acceptable by using an AI Attitude Assessment”	Not stated.	Waltzer et al. (2023)
Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive?	“to map perception, obstacles, and recommendations for optimizing use of AI in teaching academic writing in Indonesian”	“RQ1. How do Indonesian students perceive the usefulness of AI technology in their academic writing class? RQ2. How do Indonesian students perceive the ease of using AI technology in their academic writing class? RQ3. What do Indonesian students’ attitudes toward the usage of AI technology in their academic writing class?”	Utami et al. (2023)
The impact of ChatGPT on L2 writing and expected responses: Voice	“to examine doctoral students’ views on the impact of using ChatGPT on	“(1) what potential benefits does ChatGPT present for L2 writing practice? (2) what are the drawbacks of using ChatGPT in L2	Zou and Huang (2023)

Title	Purpose	Research Question	Author
from doctoral students	L2 writing and their expected responses”	writing? (3) how do doctoral students expect the educational sector to respond to ChatGPT?”	

The first study (Waltzer et al., 2023) aims to assess the ability of humans to distinguish between AI-generated and human-written text, exploring factors influencing this ability and attitudes towards AI in education. The second study (Utami et al., 2023) investigates Indonesian students' perceptions of AI in academic writing, focusing on usefulness, ease of use, and overall attitudes. The third study (Zou & Huang, 2023) examines doctoral students' views on the impact of ChatGPT on L2 writing, exploring potential benefits, drawbacks, and expectations for the future of AI in education.

Results

The results of the studies on attitude are in Table 77.

Table 77

Attitudes and Challenges: Results

Title	Results	Author
Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students	“Participants (69 high school teachers, 140 high school students, total N = 209) took an AI Identification Test in which they read pairs of essays—one written by a high school student and the other by ChatGPT—and guessed which was generated by the chatbot. Accuracy was only 70% for teachers, and it was slightly worse for students (62%). Self-reported confidence did not predict accuracy, nor did experience with ChatGPT or subject-matter expertise. Well-written student essays were especially hard to differentiate from the ChatGPT texts. In another set of measures, students reported greater optimism than their teachers did about the future role of ChatGPT in education. Students expressed disapproval of submitting ChatGPT-generated essays as one's own but rated this and other possible academic integrity violations involving ChatGPT less negatively than teachers did. These results form an empirical basis for further work on the relationship between AI and academic integrity.”	Waltzer et al. (2023)
Utilization of artificial intelligence technology in an academic writing class: How do	“Findings show that (1) AI-based learning tools help students to do academic research, especially in the planning step, to identify and develop the topics, as well as in the drafting step, to develop a paper draft, (2) AI-based learning tools are deemed flexible in accessibility despite not being able to cover	Utami et al. (2023)

Title	Results	Author
Indonesian students perceive?	all necessities required by students in writing process, (3) students are interested in using AI technology in academic writing class so that learning process will not be boring. Although AI has been used in academic writing classes, tools have not positively impacted quality of students' academic papers in all indicators. There are several obstacles to using AI, namely (1) need for more available feature, especially in editing Indonesian text, and (2) in contrast, the features still need to be optimized."	
The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students	"Thematic analysis revealed that ChatGPT could support writers at the pre-writing, during-writing and post-writing stages and serve as a self-learning tool for writing and thinking development with its human and non-human features. Nonetheless, its generative nature also gave rise to concerns for learning loss, authorial voice, unintelligent texts, academic integrity as well as social and safety risks. Based on the benefits and drawbacks, the doctoral students expected the education sector to make concerted efforts for the effective, ethical and responsible use of ChatGPT in L2 writing. Suggestions are accordingly provided for future considerations in teaching and research to leverage ChatGPT for L2 writing. / The findings overall illustrate that ChatGPT serves as a double-edged sword for students in L2 writing classrooms. While ChatGPT could support the writing process and promote self-directed learning with its human and non-human features, it may lead to a series of learning, ethical, social and safety risks, such as learning loss and insufficient authorial voice. Hence, the doctoral students advocated for concerted efforts for the effective, ethical and responsible use of ChatGPT in L2 writing."	Zou and Huang (2023)

The first study (Waltzer et al., 2023) suggests that teachers and students struggle to accurately differentiate between AI-generated and human-written essays, with overall accuracy rates below 70%. This highlights the challenge of detecting AI-generated content. Additionally, the study finds differing attitudes between teachers and students regarding the potential impact of AI on education and academic integrity. The second study (Utami et al., 2023) reveals that Indonesian students perceive AI-based tools as helpful for academic research, particularly in the planning and drafting stages. However, the quality of student papers is not uniformly improved, and there are identified limitations in the available AI tools, such as a lack of features for editing Indonesian text. The final study (Zou & Huang, 2023) suggests that doctoral students recognize both the potential benefits and drawbacks

of using ChatGPT for L2 writing. While ChatGPT can support the writing process and facilitate self-learning, concerns about learning loss, compromised authorial voice, and academic integrity are also raised. The study emphasizes the need for a balanced and responsible approach to ChatGPT integration in L2 writing education.

Implications

Table 78 provides the implications of these studies.

Table 78

Attitudes and Challenges: Implications

Title	Implications	Author
Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students	“Our research shows that distinguishing student writing from writing produced by chatbots such as ChatGPT is a significant challenge already, and it is likely to get even more difficult as AI continues to improve. Our findings also suggest that content area expertise and familiarity with ChatGPT are not sufficient to overcome this problem. We suspect that total bans on ChatGPT will be ineffective at best and that a more fruitful approach will be to find ways to understand and make use of the benefits that ChatGPT offers while minimizing its risks. Teachers who are already overburdened should not be asked to solve this problem on their own. Instead, researchers and educators will need to work together to develop best practices and effective strategies for disseminating them. Conversations that relate to this challenge are already underway, with much of the focus on using ChatGPT as a tool to achieve other goals, such as supporting language learners and providing personalized feedback on writing assignments [48]. Given that learning to work with AI systems will be a valuable skill for the foreseeable future, it will also be important to focus on teaching students how to use chatbots effectively while also thinking critically about the output they generate. As noted by New York Times tech columnist Kevin Roose, students “need to know their way around these tools — their strengths and weaknesses, their hallmarks and blind spots—in order to work alongside them” [42]. Understanding how people think about the role of ChatGPT in educational contexts is important not only due to its implications for student learning but for broader moral questions as well, including fairness. As one student commented, “I’m concerned because people who don’t use AI, like me, will see their education and grades suffer if it becomes normal.” These kinds of risks, along with the potential benefits, require creative problem-solving to effectively address. Empirical evidence and vigorous debate will play a critical role in navigating this new educational landscape.”	Waltzer et al. (2023)
Utilization of artificial intelligence technology in an	“These are the recommendations for the optimization of AI-based learning tools, which are (1) adding features to edit Indonesian text, including spelling, diction, and sentence	Utami et al. (2023)

Title	Implications	Author
academic writing class: How do Indonesian students perceive?	structure, and (2) enhancing AI literacy to be able to explore and leverage the existing features optimally.”	
The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students	“The findings can generate useful implications for different stakeholders to leverage ChatGPT for L2 writing education. Given its facilitative role in different stages of writing, writing teaches need to highlight the process approach to writing. They can consciously integrate ChatGPT in the brainstorming, outlining, revision, and proofreading process. For example, the teachers can make some short demonstration lessons (Su et al., 2023) to illustrate how to use ChatGPT to effective support the writing process. Equally important is the need for writing teachers to explore effective collaborative writing models and develop robust pedagogical guidelines for students and ChatGPT to coordinate their respective strengths to enhance writing efficiency and quality. To mitigate the potential learning and ethical risks induced by ChatGPT, it is vital to guide students exercise critical thinking throughout the writing process and encourage them to develop a distinctive authorial voice in L2 writing. For instance, teachers can concentrate more on criticality and original insights in the writing process, and instruct students to use a repertoire of discursive options and strategies (e.g., hedges, attitude markers, self-mentions, and code glosses) to construct their voice in L2 writing (Hyland, 2008). By the same token, student writers need to raise their awareness of the affordances and constraints, and explore effective strategies for human-AI collaborative writing with care and creativity. Since ChatGPT is a powerful self-learning tool, student writers can also use ChatGPT as a personal tutor for timely feedback and utilize its rich learning resources to boost critical reflection and self-regulated learning of writing. Particularly, for unfamiliar genres, students may utilize ChatGPT to generate several examples and compare those with their own writing in detail, with a view to noticing the gaps, reflecting on the noticed features, and consciously incorporating them in the following writing practices. In addition, the affordances and drawbacks also make it necessary for policymakers and educators to reformulate the regulatory policies of plagiarism and develop a protocol of ChatGPT use to ensure its ethical and responsible application in L2 writing. Considering the ethical, social and safety concerns raised by the students, technology developers can also take measures to enhance transparency and misinformation governance. Since applying ChatGPT in L2 writing involves a “concerted endeavor,” opportunities should be provided for students, teachers, policymakers, technology developers and other related stakeholders to engage in an audit and collaboratively formulate strategies to realize the educational use of ChatGPT in L2 writing (UNESCO, 2023).”	Zou and Huang (2023)

The implications of these studies emphasize the complex challenges and opportunities posed by AI in education. The first study (Waltzer et al., 2023)

highlights the difficulty of distinguishing between human and AI-generated text, emphasizing the need for collaborative efforts among educators, researchers, and policymakers to develop effective strategies for AI integration. The second study (Utami et al., 2023) points to the potential of AI as a tool for enhancing student learning while acknowledging the need for further development and optimization of AI tools. The third study (Zou & Huang, 2023) emphasizes the dual nature of AI as both a support and a potential challenge for L2 writing, highlighting the importance of responsible AI integration, critical thinking, and the development of AI literacy. Overall, these studies collectively suggest a need for a balanced and nuanced approach to AI in education, focusing on maximizing benefits while mitigating risks.

Self-Efficacy

Purpose

There are four studies that focused on self-efficacy.

Table 79

Self-Efficacy: Purpose

Title	Purpose	Research Question	Author
The Effects of Generative AI Platforms on Undergraduates' Narrative Intelligence and Writing Self-Efficacy	“to investigate whether digital story creation tasks on web-based platforms can influence the narrative intelligence and writing self-efficacy of undergraduate students / to explore the consequences of digital story creation tasks on the levels of narrative intelligence and writing self-efficacy among undergraduates majoring in Primary Education using several web-based storytelling platforms.”	“RQ1. Does creating digital stories using generative AI platforms have a significant impact on the narrative intelligence of undergraduate students? RQ2. Does creating digital stories with generative AI platforms have a significant effect on the writing self-efficacy of undergraduate students?”	Pellas (2023)
Incorporating a Reflective Thinking Promoting Mechanism into Artificial Intelligence-Supported English	“to propose a reflective thinking promotion mechanism-based AI-supported English writing (RTP-AIEW) approach to deepen learners' thinking	“1. In terms of academic performance, is the RTP-AIEW approach better than the Conventional AI-supported English writing (C-AIEW) approach?	Liu et al. (2021)

Title	Purpose	Research Question	Author
Writing Environments	and improve their EFL writing quality”	2. Is the RTP-AIEW approach better than the traditional C-AIEW approach in terms of self-efficacy? 3. Is the RTP-AIEW approach better than the traditional C-AIEW approach in terms of self-regulated learning?”	
Using AI to Evaluate a Competency-Based Online Writing Course in Nursing	“to investigate the impact of a 14-week fully online competency-based writing course on students' self-efficacy, task value, and writing performance”	“How does a discipline-specific online writing course affect students' scholarly writing performance, self-efficacy, and task value?”	Wolf and Wolf (2023)
From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill	“to investigate the impact of an AI-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill”	•• Research question 1: How does ChatGPT-based writing instruction influence EFL teachers' self-efficacy? •• Research question 2: How does ChatGPT-based writing instruction influence EFL learners' writing skills? •• Research question 3: How does ChatGPT-based writing instruction influence the retention of EFL learners' writing skills?”	Ghafouri et al. (2024)

These studies aim to understand the implications of AI in various educational contexts. The first study (Pellas, 2023) investigates the ability to detect AI-generated text, the second (Liu et al., 2021) explores student perceptions of AI in academic writing, the third (Wolf & Wolf, 2023) examines the impact of ChatGPT on L2 writing, and the fourth (Ghafouri et al., 2024) analyzes the relationship between AI literacy and students' writing performance.

Results

Table 80 shows the results of these studies.

Table 80

Self-Efficacy: Results

Title	Results	Author
The Effects of Generative AI Platforms on Undergraduates'	“The findings indicate that the utilization of generative AI platforms in the context of story creation tasks can substantially enhance both narrative intelligence scores and writing self-efficacy when compared to conventional	Pellas (2023)

Title	Results	Author
Narrative Intelligence and Writing Self-Efficacy	platforms. Nonetheless, there was no significant difference in the creative identity factor. The findings reveal a noteworthy difference in favor of the experimental group. In line with previous studies [8,21], this study's findings indicate that the levels of narrative intelligence among undergraduate students exposed to the experimental AI-supported instructional method were significantly higher than those of the control group. In relation to RQ2, the analysis of writing self-efficacy, measured through the SAWSES, was conducted using a repeated measurements two-way ANOVA. Consistent with previous studies' findings [4,19], the evidence from this comparative analysis allowed us to explore whether the proposed treatment had a significant impact on the writing self-efficacy of undergraduate students. The analysis also indicated a significant difference in favor of the experimental group in the "Writing essentials" factor. This suggests that the proposed treatment was effective in improving students' understanding of fundamental writing principles, further supporting the positive impact of the experimental approach. In the "Reflective writing" factor, a significant difference was observed in favor of the experimental group. In the "Reflective writing" category, ANOVA results similarly indicated a substantial impact of the intervention. Nonetheless, it is worth noting that there was no significant difference in the "Creative identity" factor. This implies that the proposed treatment did not have a significant impact on students' perceptions of their creative identity in writing."	
Incorporating a Reflective Thinking Promoting Mechanism into Artificial Intelligence-Supported English Writing Environments	"The results indicated that the proposed approach not only significantly improved the experimental group students' English writing performance, but also improved their self-efficacy and self-regulated learning, and significantly reduced their cognitive load. In addition, students' learning experience and perceptions are also discussed. / research question (1) has been answered: The RTP-AIEW learning approach was effective in terms of improving students' writing performance. research question (2), the results of the study showed that students in the experimental group who learned using the RTP-AIEW approach had significantly higher self-efficacy than students in the control group who learned using the C-AIEW approach. research question (3), the current study found that there was a significant difference in the self-regulated learning of the students in the experimental group in this study compared to that of the control group. That is, the RTP-AIEW learning approach significantly improved the learners' self-regulated learning compared with C-AIEW. research question (4), this study confirmed that the RTP-AIEW approach is effective in terms of reducing students' cognitive load. Students in the experimental group who learned with RTP-AIEW had significantly lower cognitive load than those who learned using the C-AIEW approach. research question (5), interviews with learners revealed that in the field of EFL writing, the RTP-AIEW approach was perceived to be effective in terms of improving teaching efficiency and reducing teachers' burden of correcting assignments. Besides, most students in the study said that this approach improved the efficiency of learning EFL writing, leaving more time for reflection, so they were more motivated	Liu et al. (2021)

Title	Results	Author
	to learn and more willing to continue using this approach, and recommended it to other students and teachers in the future. In addition, several students also pointed out that the rubrics for evaluating the essay need to be clarified for students, which would be beneficial for improving their EFL writing performance.”	
Using AI to Evaluate a Competency-Based Online Writing Course in Nursing	“The results showed statistically significant gains in self-efficacy and writing performance with large effect sizes. This study addresses the gap in nursing education regarding the assessment of online, research-based writing interventions on students' scholarly writing capacity. this study indicates that using automated essay scoring enabled by artificial intelligence can provide an efficient, reliable, and valid way to evaluate students' writing performance for educational research.”	Wolf and Wolf (2023)
From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill	“The Independent Samples t-test results showed that CGWIP significantly enhanced teachers' self-efficacy compared to the control group. Also, the results of One Way ANCOVA revealed that CGWIP significantly improved learners' writing skills and that these effects persisted over time. The study implied that the protocol can nurture teachers' efficiency by helping them in various aspects of L2 writing instruction, including brainstorming, revising, providing feedback, and assessment, which in turn, improves learners' writing skills.”	Ghafari et al. (2024)

According to the first study (Pellas, 2023), the use of generative AI platforms in story creation tasks significantly enhanced undergraduate students' narrative intelligence and writing self-efficacy compared to traditional methods. However, there was no significant impact on creative identity. The second study (Liu et al., 2021) shares that incorporating a reflective thinking promoting mechanism into AI-supported English writing environments positively influenced students' writing performance, self-efficacy, self-regulated learning, and reduced cognitive load. Students reported positive learning experiences and perceived benefits for both students and teachers.

The third study (Wolf & Wolf, 2023) concludes that the use of AI-enabled automated essay scoring in a competency-based online nursing writing course resulted in significant gains in student self-efficacy and writing performance, which demonstrates the potential of AI for efficient and reliable writing assessment. The

last study (Ghafouri et al., 2024) reveals that the ChatGPT-based writing instruction protocol significantly improved both teachers' self-efficacy and students' writing skills. The protocol demonstrated effectiveness in various aspects of writing instruction and had a lasting impact on students' writing abilities.

Implications

Table 81 provides the implications of these studies.

Table 81

Self-Efficacy: Implications

Title	Implications	Author
The Effects of Generative AI Platforms on Undergraduates' Narrative Intelligence and Writing Self-Efficacy	“The results of this research carry significance not only for the academic community but also for educational practitioners and technology developers. Understanding whether digital story creation activities influence narrative intelligence and writing self-efficacy can offer valuable insights into the design and implementation of generative AI platforms to enhance, rather than impede, individuals’ storytelling skills and confidence in the digital age. Enhancement of narrative intelligence and writing self-efficacy: This study’s results provide evidence that the use of generative AI platforms can significantly improve undergraduate students’ narrative intelligence scores and writing self-efficacy. This finding aligns with the growing recognition of the potential benefits of technologyassisted learning in enhancing core competencies. Educators and institutions can leverage these tools to empower students with the skills and confidence necessary for effective communication; 2. Effects of generative AI platforms on storytelling: The positive outcomes observed in the experimental group highlight the evolving role of AI technologies in education. Generative AI platforms can augment traditional teaching methods by offering personalized feedback, suggesting improvements, and facilitating the writing process. The study suggests that these technologies can serve as valuable educational aids, supporting students in their journey to become proficient writers; 3. Human creativity vs. generative AI assistance: Educators and curriculum designers should consider incorporating generative AI platforms into writing instruction. On the one hand, while generative AI demonstrated clear benefits in terms of narrative intelligence and writing self-efficacy, the absence of a significant difference in the “Creative identity” factor suggests that human creativity remains a distinctive and irreplaceable aspect of storytelling. This finding underscores the importance of striking a balance between generative AI assistance and human creativity, particularly in fields where originality and creative expression are highly valued. On the other hand, it is crucial to do so thoughtfully, recognizing the strengths and limitations of AI tools. These platforms can be	Pellas (2023)

Title	Implications	Author
	especially useful for tasks involving grammar, structure, and organization, leaving space for students to focus on the creative aspects of their writing.”	
Incorporating a Reflective Thinking Promoting Mechanism into Artificial Intelligence-Supported English Writing Environments	“Theoretically, the current study empirically examined the effects of double-loop learning theory in the learning process, and the findings also expanded our understanding of the role of AI-supported automated feedback in EFL learning, particularly in writing. Pedagogically, the current study provides possible pedagogical scaffolding for EFL writing courses with the support of AI-supported automated feedback. Specifically, it has been pointed out that the benefits of AI-supported automated feedback for teaching and learning somehow depends on how the technology is integrated into the pedagogical practice. Methodologically, this study contributes to the conclusiveness of findings regarding the efficacy of RTP-AIEW in facilitating EFL writing, self-efficacy, and self-regulated learning, and reducing EFL cognitive load. These findings are valuable since previous studies mainly focused on the effects of automated rating tools on learning performance rather than on learning perceptions (Barrot, 2021; Guo et al., 2021).”	Liu et al. (2021)
Using AI to Evaluate a Competency-Based Online Writing Course in Nursing	“There are four important implications for practice based on the reviewed literature and the results of this study. First, nursing faculty should offer students required writing and writing-intensive courses based on cognitive apprenticeship principles. Second, this study affirms Wornock’s (2009) assertion that the migration of writing pedagogy to an online context is effective. Faculty can use best practices from research on composition and writing instruction to inform their online instruction. Third, a competency-based approach that provides students with a flexible path to mastery can operate within a traditional time-based semester online. Instructors who support the development of nursing students’ writing and thinking skills can empower students from a wide range of educational backgrounds. The scale of help is a promising strategy for students from underrepresented and low socioeconomic groups to enter the profession and enable their ongoing success in graduate school and as future leaders, scholars, faculty, and researchers. Fourth, this study indicates that using automated essay scoring enabled by artificial intelligence can provide an efficient, reliable, and valid way to evaluate students’ writing performance for educational research. This has significant implications for evaluating writing, not only allowing this study to be replicated but also extended into longitudinal studies that track the development of writing performance across a curriculum. Additionally, this instrument may help researchers further explore the relationship between writing performance, critical thinking, clinical judgment, and decision-making. The administration of automated essay scoring at the beginning and end of a course or program can also provide students with data points to reflect upon their writing progress over time.”	Wolf and Wolf (2023)
From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based	“Several theoretical and pedagogical implications can be induced from the present study. The results implied that ChatGPT-based instruction would remind the teachers about their capabilities to overcome challenges by influencing their	Ghafouri et al. (2024)

Title	Implications	Author
writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill	students' outcomes while triggering a sense of self-efficacy. Further, the study contributes to the current positive psychology-laden trends of L2 research by suggesting that if applied in a well-designed and controlled manner, ChatGPT can benefit EFL teachers by increasing their self-efficacy, which is correlated with other notions such as reduced stress and burnout (Jerusalem & Schwarzer, 1992; Kim & Burić, 2020; Pfitzner-Eden, 2016; Skaalvik & Skaalvik, 2010), creative teaching methods (Thurlings et al., 2015), perseverance (Heng & Chu, 2023; Milner & Hoy, 2003; Shao, 2023), job satisfaction (Demir, 2018), and similar psycho-emotional entities the combination of which promotes teachers' wellbeing (von Muenchhausen et al., 2021; Xiyun et al., 2022). Additionally, based on the results and in line with recent AI-driven L2 studies, we argue that ChatGPT is beneficial for L2 writing through features such as automated essay scoring, intelligent tutoring systems, or natural language processing (Ali et al., 2023; H.S. Kim et al., 2021; Kohnke et al., 2023; Lin & Mubarak, 2021). ChatGPT can also adapt to the users' needs, interests, and proficiency levels and provide immediate, personalized, relevant feedback and support. ChatGPT can also interact with the users in a human-like manner, using natural and coherent language, and engage them in a dialogue. These advantages make ChatGPT a more appealing and effective tool for L2 writing instruction, as it can cater to teachers' and students' diverse and dynamic needs."	

These studies collectively demonstrate the potential of generative AI to significantly enhance educational outcomes, particularly in writing and creativity. The first study (Pellas, 2023) highlights the positive impact of AI on narrative intelligence and writing self-efficacy, suggesting that AI can be a valuable tool for developing these skills. The second study (Liu et al., 2021) emphasizes the importance of thoughtful AI integration, demonstrating how reflective thinking mechanisms can optimize learning experiences and reduce cognitive load. The third study (Wolf & Wolf, 2023) shows the potential of AI in providing efficient and reliable assessment, enabling personalized learning and improved writing performance. The fourth study (Ghafouri et al., 2024) underlines the role of AI in empowering both teachers and students, enhancing instructional effectiveness and student engagement. Overall, these findings suggest that AI has the potential to

revolutionize education, but its successful implementation requires careful consideration of pedagogical, ethical, and technological factors.

Users and Uses

Purpose

There are five studies that investigate usage behavior.

Table 82

Users and Uses: Purpose

Title	Purpose	Research Question	Author
'brave new world' or not?: A mixed-methods study of the relationship between second language writing learners' perceptions of chatgpt, behaviors of using chatgpt, and writing proficiency	“to investigate the relationship between second or foreign language (L2) writing learners' perception of ChatGPT, their usage behavior, and L2 writing proficiency, including complexity, accuracy, and fluency.”	Not stated.	Dong (2024)
Research writing with ChatGPT: A descriptive embodied practice framework	“to offer a framework of descriptive codes that identify practices early adopters enact when integrating these technologies into research writing processes based on a qualitative analysis of 35 publicly available videos depicting the use of ChatGPT and Bing”	“1. In terms of academic we describe the interactions between humans and text generation software technologies as they are enacted in the embodied practice of research writing? 2. What implications do these embodied practices have for our understandings about written invention in research writing?”	Pigg (2024)
Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective	“This research aims to understand students' viewpoints on their use or avoidance of AI in composing academic essays, seeking to reveal the reasons behind their choice to embrace or forgo AI tools for writing tasks. Additionally, the study identifies the prevalent AI applications students typically use in the process of crafting academic essays.”	Not stated.	Malik et al. (2023)
Human-AI collaboration patterns in AI-	“to examine the nature of human-AI interactions in academic writing,	RQ1. What are the strategies and patterns doctoral students employ in their academic writing processes	Nguyen et al. (2024)

Title	Purpose	Research Question	Author
assisted academic writing	specifically investigating the strategies doctoral students employ when collaborating with a GAI-powered assisting tool”	when collaborating with GAI-assisted writing tools? o RQ1a) What are the key hidden states in GAI-assisted academic writing? o RQ1b) What are the predominant patterns in GAI-assisted academic writing? RQ2. How is the effectiveness of each identified strategic pattern in GAI-assisted academic writing characterised by its specific nature and underlying attributes?”	
Graduate Students' ChatGPT Experience and Perspectives during Thesis Writing	“to analyze the perspectives of 80 male and female postgraduate students in various academic fields at Jadara University regarding the use of ChatGPT in writing their theses”	“1. What are the perceptions of postgraduate students at Jadara University regarding the use of ChatGPT in writing their theses? 2. Are there any gender differences in the perceptions of postgraduate students at Jadara University regarding the use of ChatGPT in writing their theses?”	Rababah et al. (2024)

Dong (2024) aims to investigate the relationship between second language (L2) writing learners' perceptions of ChatGPT, their usage behavior, and L2 writing proficiency, focusing on complexity, accuracy, and fluency. Pigg (2024) seek to develop a framework of descriptive codes identifying practices early adopters use when integrating AI technologies into research writing processes. They analyze 35 publicly available videos showing the use of ChatGPT and Bing to understand human-AI interactions in research writing and their implications for written invention. Malik et al. (2023) aim to understand higher education students' perspectives on using or avoiding AI in composing academic essays, exploring the reasons behind their choices and identifying the most common AI applications used in academic writing tasks.

Nguyen et al. (2024) examine the nature of human-AI interactions in academic writing, focusing on the strategies doctoral students employ when collaborating with AI-powered writing tools. They investigate key hidden states,

predominant patterns, and the effectiveness of strategic patterns in AI-assisted academic writing. Rababah et al. (2024) analyze the perspectives of 80 postgraduate students at Jadara University regarding the use of ChatGPT in writing their theses, considering various academic fields and exploring potential gender differences in perceptions.

Results

The results from these studies are shown in Table 83.

Table 83

Users and Uses: Results

Title	Results	Author
'brave new world' or not?: A mixed-methods study of the relationship between second language writing learners' perceptions of chatgpt, behaviors of using chatgpt, and writing proficiency	“Results indicated that performance expectancy and effort expectancy significantly influenced students’ intended behaviors, while the effect of social influence was non-significant. Facilitating conditions significantly predicted actual behaviors. Furthermore, the study found that ChatGPT usage significantly predicted only complexity. During the focus group discussion, students expressed the need for support from teachers. The study highlights the importance of human agency in writing instruction, including clear principles, critical evaluation of ChatGPT feedback, and constant monitoring to reinforce human roles in writing. These findings provide insights into students’ perceptions and behaviors toward AI-based writing tools and inform effective writing instruction in the AI era. / First, we found that students’ perceptions of performance expectancy and effort expectancy are significant predictors of their use of ChatGPT for writing consultation and essay revision.our study revealed that facilitating conditions has a positive, significant impact on use behavior, consistent with previous research. the coefficient of the path from behavioral intention to use behavior is significant but relatively small, and the focus group discussion provides an explanation for this finding. The discussion indicates that students have concerns and even fear regarding the use of ChatGPT, which may be attributed to the criticisms surrounding this tool.”	Dong (2024)
Research writing with ChatGPT: A descriptive embodied practice framework	“Overall, the framework offered in this article provides a descriptive starting point that future researchers can reference when generating interview, survey, or observational protocols to further study the effects of these technologies. For teachers and students, this framework offers an accessible way to name and describe moves that students may already be using privately, so that they can discuss issues of ethics, invention, originality, and productivity related to these writing practices. And finally, this article contributes one of the first descriptive analyses of ChatGPT-supported research writing practices,	Pigg (2024)

Title	Results	Author
	while also demonstrating the importance of continuing and further refining our research into embodied writing practice.”	
Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective	“Findings indicated a positive reception of AI-powered writing tools, with students acknowledging their benefits in grammar checks, plagiarism detection, language translation, and essay outlines. AI was found to enhance students' writing abilities, self-efficacy, and understanding of academic integrity. However, some students expressed concerns about potential impacts on creativity, critical thinking, and ethical writing practices. The study emphasized a balanced approach to AI integration, where AI collaborates with human authors, and also identified popular AI tools used by Indonesian students. As such, the research highlighted AI's significant role in supporting academic writing while preserving human creativity and critical thinking, underscoring the importance of maintaining a balanced integration to sustain human ingenuity and critical thought in academic discourse.”	Malik et al. (2023)
Human-AI collaboration patterns in AI-assisted academic writing	“Findings indicate that doctoral students engaging in iterative, highly interactive processes with the GAI-powered assisting tool generally achieve better performance in the writing task. In contrast, those who use GAI merely as a supplementary information source, maintaining a linear writing approach, tend to get lower writing performance. This study points to the need for further investigations into human-AI collaboration in learning in higher education, with implications for tailored educational strategies and solutions.”	Nguyen et al. (2024)
Graduate Students' ChatGPT Experience and Perspectives during Thesis Writing	“The outcomes of this research provide insightful information on the perspectives of postgraduate students at Jadara University regarding the use of ChatGPT in writing their theses. The findings suggest that students view ChatGPT as a beneficial tool that enhances the writing process, writing quality, knowledge retrieval, and the generation of new ideas. Students were aware of the significance of responsible use and cross-referencing AI-generated information, despite some expressing concerns the potential for plagiarism.”	Rababah et al. (2024)

The first study (Dong, 2024) reveals a complex relationship between L2 learners' perceptions, ChatGPT usage, and writing proficiency. While perceptions influence intended usage, actual usage is more strongly linked to facilitating conditions. The study emphasizes the need for human guidance in AI tool utilization. The second study (Pigg, 2024) develops a framework for understanding human-AI interactions in research writing, and this highlights the importance of embodied practices in shaping writing processes. This framework provides a foundation for future research into the ethical and creative dimensions of AI-assisted writing.

The third study (Malik et al., 2023) finds a generally positive perception of AI tools among students, with benefits seen in grammar, plagiarism, and idea generation. However, concerns about creativity and critical thinking are also expressed, emphasizing the need for a balanced approach to AI integration. The fourth study (Nguyen et al., 2024) identifies interactive and iterative processes as key to successful AI-assisted academic writing. It highlights the importance of tailored educational strategies to support effective human-AI collaboration and emphasizes the need for further research into this area. The fifth study (Rababah et al., 2024) reveals positive student perceptions of ChatGPT, with benefits seen in writing quality, idea generation, and knowledge retrieval. However, concerns about plagiarism and responsible use are also expressed, and this underscores the need for guidelines and education on AI tool utilization.

Implications

Table 84 provides the implications of these studies.

Table 84

Users and Uses: Implications

Title	Implications	Author
'brave new world' or not?: A mixed-methods study of the relationship between second language writing learners' perceptions of chatgpt, behaviors of using chatgpt, and writing proficiency	“The article suggests providing guidance to students on how to use ChatGPT effectively and ethically rather than banning it. Social influence had non-significant predictive effects on using ChatGPT, and teachers must provide clear guidance on appropriate uses and consequences for unethical use from the beginning. The second suggestion is to encourage students to be critical of the feedback provided by ChatGPT and identify false sources or arguments. A longitudinal and constant monitoring approach is necessary to monitor students' progress and provide appropriate guidance. Universities should provide support to teachers, including establishing writing guidelines, monitoring writing progress, and strategic positioning of students. Finally, the article emphasizes the need for educators to enhance their writing assessment literacy to provide meaningful feedback and support to students who engage with AI-based writing tools.”	Dong (2024)
Research writing with ChatGPT: A	“Many research writers are looking to ChatGPT in one of two scenarios. In one scenario, they are turning to this tool when they are stuck or struggling to offer solutions to ongoing	Pigg (2024)

Title	Implications	Author
descriptive embodied practice framework	problems with reading comprehension, writer's block, and/or understanding genre conventions. In another scenario, they are turning to the tool to meet the ever-increasing demands of productivity that require we do more with limited time. In both cases, generative AI technology for some offers a short-term solution that feels "game changing" to students and professionals who shared their sessions in public forums. In the long-term, however, it remains to be seen whether using these tools to overcome writing struggles will translate into longer term writing learning and development, as well as whether these technologies will merely amp up increasing expectations for writing productivity."	
Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective	"As such, the practical implication of this study lies in illuminating the transformative impact and valued utility of AI-powered tools in academic writing, providing educators, developers, and students with insights and actionable guidance on the balanced integration and optimal utilization of AI technologies to enhance writing processes while preserving human creativity and critical thinking."	Malik et al. (2023)
Human-AI collaboration patterns in AI-assisted academic writing	"This study offers several implications for different stakeholders in higher education concerning the integration of AI tools like the GAI-powered tool in academic writing. For educators, the data reveal the specific cognitive and metacognitive skills contributing to effective collaboration with AI. This information is vital for developing targeted pedagogical strategies and interventions to enhance students' skill sets. For instructional designers and developers in the EdTech industry, the study identifies key functionalities particularly useful to students in the writing process. The research suggests that AI-enabled writing tools should promote adaptive coordination among users, as this approach correlates with higher writing performance. In terms of policymaking, the findings offer a substantial foundation for formulating policies on the inclusion of AI technologies in academic settings. Policymakers must consider the academic benefits and ethical considerations such as data privacy and potential misuse. Accordingly, comprehensive frameworks must be developed to govern the responsible use of these technologies. Finally, for students, particularly at the doctoral level, the study provides invaluable guidance for optimising their use of AI tools in academic writing. The research underscores the need for a proactive, dynamic engagement with the AI tool, as this mode of interaction has been shown to result in higher-quality academic work."	Nguyen et al. (2024)
Graduate Students' ChatGPT Experience and Perspectives during Thesis Writing	"The results have several repercussions, especially for academic institutions, educators, and researchers. To begin, incorporating AI technologies such as ChatGPT into academic writing may provide students with beneficial assistance and help them enhance their writing abilities. Writing support services offered by institutions may benefit from the integration of AI-powered tools. Additionally, these institutions offer workshops on how to effectively and	Rababah et al. (2024)

Title	Implications	Author
	ethically utilize such tools. In addition, teachers can guide students through ChatGPT to enhance the quality of their writing process and encourage critical examination of the material produced. The positive feedback from students regarding ChatGPT underscores the potential for further research and development of writing support tools utilizing AI. Researchers may continue their investigations into the effectiveness of AI tools in different writing scenarios and explore innovative methods to tackle issues related to plagiarism while promoting academic integrity. In addition, future research could explore the potential long-term effects of integrating AI technologies on students' writing skills and academic achievements.”	

These studies collectively highlight the complex and evolving relationship between humans and AI in writing contexts. The first study (Dong, 2024) emphasizes the importance of understanding the factors influencing AI tool adoption and usage, suggesting a need for balanced guidance and support for students. The second study (Pigg, 2024) underscores the significance of embodied practices in AI-assisted writing, emphasizing the need for further research into the impact of these tools on writing processes. The third study (Malik et al., 2023) reaffirms the potential of AI to enhance writing skills while emphasizing the importance of maintaining human creativity and critical thinking. The fourth study (Nguyen et al., 2024) highlights the importance of interactive and iterative processes in AI-assisted writing, emphasizing the need for tailored educational strategies and policies. Finally, the fifth study (Rababah et al., 2024) underscores the potential benefits of AI for writing support while emphasizing the need for responsible use and further research into AI's long-term impact on writing skills and academic achievement. Overall, these studies collectively point to the need for a nuanced and balanced approach to AI integration in writing, with a focus on human agency, critical thinking, and ethical considerations.

Trust

Purpose

The study on trust (Tossell et al., 2024) aims to understand how students perceive and use ChatGPT in a college essay writing context, focusing on their experiences, learning outcomes, trust in the tool, and implications for grading.

Table 85

Trust: Purpose

Title	Purpose	Research Question	Author
Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence	“to examine student experiences before and after an essay writing assignment that required the use of ChatGPT within an undergraduate engineering course”	“RQ1: What are students’ perceptions of an assignment requiring the use of ChatGPT? RQ2: What are student perceptions of ChatGPT to support their learning and are they comfortable taking responsibility for the content it produces? RQ3: Do students trust ChatGPT and how does this impact their intent to rely on it for future assignments and for grading?”	Tossell et al. (2024)

Results

The results from this study are in Table 86.

Table 86

Trust: Results

Title	Results	Author
Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence	“Our quantitative and thematic analyses uncovered that ChatGPT did not simplify the writing process. Instead, the tool transformed the student learning experience yielding mixed responses. Participants reported finding ChatGPT valuable for learning, and their comfort with its ethical and benevolent aspects increased postuse. Concerns with ChatGPT included poor accuracy and limited feedback on the confidence of its output. Students preferred instructors to use ChatGPT to help grade their assignments, with appropriate oversight. They did not trust ChatGPT to grade by itself. Student views of ChatGPT evolved from a perceived cheating tool to a collaborative resource that requires human oversight and calibrated trust. Implications for writing, education, and trust in artificial intelligence are discussed.”	Tossell et al. (2024)

The study (Tossell et al., 2024) observes that while ChatGPT do not simplify the writing process, it does impact student learning experiences, with both positive and negative perceptions. Students value ChatGPT as a learning tool but express concerns about its accuracy and limitations. They prefer human oversight in grading and demonstrate a shift in perspective from viewing ChatGPT as a cheating tool to a collaborative resource requiring trust and careful monitoring.

Implications

Table 87 provides the implications of this study.

Table 87

Trust: Implications

Title	Implications	Author
Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence	“AI technology, such as ChatGPT, can enhance student learning in STEM and non-STEM fields, but educators should approach it with caution. Providing clear practical guidelines and instructions for students can help optimize their interactions with AI tools like ChatGPT. Instructors should experiment with integrating ChatGPT intentionally in the writing process, considering its iterative and dynamic nature. AI developers should continue refining and improving language models to ensure effectiveness and reliability, addressing common issues like repetitive language generation and inadequate essay production. Human oversight is crucial in grading, as students prefer instructor involvement in high-stakes situations, although they also see potential benefits in using AI tools like ChatGPT as an assistive tool in grading. Encouraging students to use AI-generated content as a starting point and iteratively refining and enhancing their essays can foster a deeper understanding of the writing process and enable skill development. The integration of AI tools in traditional schools and platforms can enhance personalized learning experiences and supplement existing course materials.”	Tossell et al. (2024)

The study (Tossell et al., 2024) shows the potential of AI tools like ChatGPT to enhance student learning while emphasizing the need for careful integration and human oversight. It highlights the importance of clear guidelines, ongoing support, and critical evaluation of AI-generated content. The study also emphasizes the role

of educators in fostering student development of AI literacy and ethical use of these tools.

Motivation

Purpose

This study aims to investigate the effectiveness of AI-assisted language learning in enhancing EFL students' writing skills and motivation.

Table 88

Motivation: Purpose

Title	Purpose	Research Question	Author
Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students	“to evaluate the impact of AI-assisted language learning on Chinese English as a Foreign Language (EFL) students' writing skills and writing motivation”	“1. To what extent does AI-assisted language learning instruction contribute to the enhancement of academic writing skills in EFL learners compared to non-AI-assisted instruction? 2. To what extent does AI-assisted language learning instruction contribute to the enhancement of writing motivation in EFL learners compared to non-AI-assisted instruction? 3. What are the perceptions of EFL learners toward the impact of AI-assisted language learning on their academic writing skills and writing motivation?”	Song and Song (2023)

Results

The results from this study are shown in Table 89.

Table 89

Motivation: Results

Title	Results	Author
Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students	“Quantitative analysis reveals significant improvements in both writing skills and motivation among students who received AI-assisted instruction compared to the control group. The experimental group demonstrates enhanced proficiency in various aspects of writing, including organization, coherence, grammar, and vocabulary. Qualitative findings showcase diverse perspectives, ranging from recognition of AI's innovative instructional role and its positive influence on writing skills and motivation to concerns	Song and Song (2023)

Title	Results	Author
	about contextual accuracy and over-reliance. Participants also reflect on the long-term impact and sustainability of AI-assisted instruction, emphasizing the need for ongoing development and adaptation of AI tools.”	

The study (Song & Song, 2023) finds that AI-assisted language learning significantly improves EFL students' writing skills and motivation compared to traditional instruction, while also revealing diverse student perceptions regarding the strengths and limitations of AI tools.

Implications

Table 90 provides the implications.

Table 90

Motivation: Implications

Title	Implications	Author
Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students	“The findings of this study carry important implications for the field of EFL, particularly in relation to the teaching of writing. The integration of AI tools into EFL writing courses is recommended as part of a technology-integrated language learning approach, as it aligns with innovative pedagogical practices and has the potential to significantly enhance students’ academic writing skills and writing motivation. To enhance EFL learners’ academic writing skills and motivation, it is highly encouraged for learners to actively incorporate AI tools into their writing activities. By leveraging the capabilities of AI, learners can receive immediate and targeted feedback, enabling them to identify areas for improvement and enhance their writing performance accordingly. This integration of AI into their writing practice has the potential to yield more efficient and effective learning outcomes. To provide engaging and motivating writing activities supported by AI, EFL educators and teachers should consider establishing dedicated AI-supported language learning classes. These classes can serve as platforms to train EFL learners on how to effectively utilize AI tools to improve their writing skills. Through providing guidance and support in the use of AI tools, educators can empower learners to take ownership of their learning process and make meaningful progress in their writing abilities. Moreover, organizing an AI-supported language learning class specifically tailored to writing activities can bring significant benefits to EFL learners. Within such a class, learners can receive AI-generated writing feedback and comments addressing various aspects of their written texts, including language usage, organization, and content. This comprehensive feedback can provide learners with specific guidance to address their writing challenges and refine their skills across different areas.”	Song and Song (2023)

The study (Song & Song, 2023) highlights the potential of AI-assisted language learning to significantly enhance EFL students' writing skills and motivation. It recommends integrating AI tools into writing courses, providing learners with training on effective AI utilization, and creating dedicated AI-supported language learning environments. By doing so, educators can empower students to take ownership of their learning and achieve greater writing proficiency.

Independence and Proficiency

Purpose

This study aims to examine how AI-based writing tools can foster student independence, improve writing proficiency, and increase writing engagement through the lens of self-determination theory.

Table 91

Independence and Proficiency: Purpose

Title	Purpose	Research Question	Author
Artificial Intelligence in writing instruction: A self-determination theory perspective	“The primary goal of the research was to identify the ways in which writing tools powered by artificial intelligence can improve students' independence, proficiency, and writing-relatedness. This study aimed to delve into the multifaceted role of artificial intelligence (AI) in writing instruction, with a particular emphasis on AI-based writing tools like Grammarly and QuillBot.”	Not stated.	Rahayu et al. (2024)

Results

The results from this study are shown in Table 92.

Table 92

Independence and Proficiency: Results

Title	Results	Author
Artificial Intelligence in writing instruction: A self-determination theory perspective	“The main findings of the study revealed that AI-based writing tools enhance writing competence and fulfil psychological needs such as autonomy, relatedness, and integrity in writing but risks of relying too much on AI were also pointed out. For example, AI could kill creativity, make writing styles more similar, and make people feel like they have no control over their writing. While the study acknowledges the significant potential of AI to enhance the writing process. Moreover, it highlights the importance of being aware of the potential dangers that it poses. The results emphasize the need for a well-rounded strategy that capitalizes on the advantages of AI in writing education while guaranteeing that it enhances and complements unique writing approaches, as opposed to dictating them. / The study revealed that AI tools' ability to analyze and predict data is crucial in enhancing individuals' self-efficacy and meeting their psychological need to feel competent in their interactions with their environment. Moreover, it has been discovered that AI's ability to offer tailored experiences fosters a feeling of honesty and sincerity, allowing individuals to express themselves genuinely in a way that corresponds with their own values and identity.”	Rahayu et al. (2024)

The study (Rahayu et al., 2024) reveals that AI-based writing tools can enhance writing competence and fulfill psychological needs, but excessive reliance on AI may hinder creativity and autonomy. The results emphasize the need for a balanced approach that leverages AI's benefits while preserving human agency and individuality in writing.

Implications

Table 93 provides the implications of the study.

Table 93

Independence and Proficiency: Implications

Title	Implications	Author
Artificial Intelligence in writing instruction: A self-determination theory perspective	“The study suggests that AI-based writing tools can significantly enhance the writing process, catering to technical and psychological needs. It's important to be aware of the possible dangers of relying too much on AI, though. Educators and individuals who regularly use AI-based writing tools should strive for a balanced approach, leveraging the benefits	Rahayu et al. (2024)

Title	Implications	Author
	of AI while ensuring it serves as an enhancement tool rather than a determinant of writing style. These findings provide a foundation for further research into the integration of AI in educational practices, contributing to the ongoing discourse on the role of AI in education.”	

The study (Rahayu et al., 2024) emphasizes the potential of AI to enhance writing instruction while highlighting the importance of maintaining human agency and critical thinking in the writing process. It recommends a balanced approach that leverages AI as a supportive tool rather than a replacement for human expertise.

AI-Integrated Writing Instruction and Outcomes

The analysis of the studies yielded several common core concepts related to *AI-Integrated Writing Instruction and Outcomes*. Related to writing practices and outcomes, there were several building blocks: *writing skills*, *writing performances*, *writing process*, *AI-enhanced classroom*, *writing quality*, *scaffolding* and *AI-assisted writing* (Figure 9).

Figure 9

AI-Integrated Writing Instruction and Outcomes



Writing Skills

Purpose

There are 11 studies that aimed at understanding the impact of AI on writing skills.

Table 94

Writing Skills: Purpose

Title	Purpose	Research Question	Author
Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study	“to address this research gap [re: register] by examining the impact of ChatGPT, an AI-powered chatbot, on the acquisition of register knowledge across various writing tasks.”	“1. Can the use of ChatGPT as a learning assistant help students improve self-editing their writing? 2. What are the students’ opinions and suggestions regarding using ChatGPT as a learning assistant?”	Punar Özçelik and Yangın Ekşi (2024)
Empowering Chinese Language Learners from Low-Income Families to Improve Their Chinese Writing with ChatGPT’s Assistance Afterschool	“to explore the potential of ChatGPT in supporting and empowering Chinese language learners (CLLs) whose first language is English to enhance their writing skills”	“1. Is there a functional relation between Chinese language learners from low-income families using ChatGPT after school twice a week and improvements in their Chinese writing?”	Li et. Al (2023)
Defining the Role of Artificial Intelligence in Improving English Writing Skills Among Indonesian Students	“to explore the improvement of English writing skills through the integration of AI into reading practices and provision of adequate feedback for skills enhancement”	Not stated.	Kaharuddin et al. (2024)
Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students	“to evaluate the impact of AI-assisted language learning on Chinese English as a Foreign Language (EFL) students' writing skills and writing motivation”	“1. To what extent does AI-assisted language learning instruction contribute to the enhancement of academic writing skills in EFL learners compared to non-AI-assisted instruction? 2. To what extent does AI-assisted language learning instruction contribute to the enhancement of writing motivation in EFL learners compared to non-AI-assisted instruction? 3. What are the perceptions of EFL learners toward the impact of AI-assisted language learning on their academic	Song and Song (2023)

Title	Purpose	Research Question	Author
Investigating EFL Students' Writing Skills Through Artificial Intelligence: Wordtune Application as a Tool	“to fill this gap [few studies have explored how AI-driven applications impact Saudi English as a Foreign Language (EFL) writing] by investigating whether and to what extent Wordtune facilitates Saudi students' writing”	writing skills and writing motivation?” “1. To what extent does Wordtune enhance student writing? 2. What are the specific improvements in Wordtune-mediated student writing? 3. Are there any statistical differences in students' writing skills based on their sex?”	Mahmud (2023)
Enhancing efl reading and writing through ai-powered tools: Design, implementation, and evaluation of an online course	“to design, implement, and evaluate a synchronous online course on AI-powered reading and writing tools for senior high school students learning how to read and write English”	“RQ1: What did a synchronous online course in which AI-powered reading and writing tools were taught look like? RQ2: In this synchronous online course, did students perceive two main learning activities (i.e. student-centered presentations and teacher-centered lectures) differently? RQ3: In this synchronous online course, what were the learner background variables (i.e. grade level, gender, language proficiency, and school location) that predicted learning performance?”	Hsiao and Chang (2023)
Impact of ChatGPT on Learners in a L2 Writing Practicum: An Exploratory Investigation	“to examine the impact of ChatGPT's automatic text generation on undergraduate students' behavior and perceptions in learning L2 writing skills”	“RQ1: How do participants in the practicum develop skills in applying ChatGPT to complete undergraduate L2 writing tasks? RQ2: How do participants in the practicum perceive the application of ChatGPT in L2 Writing?”	Yan (2023)
Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application	“to describe how AI is now used to provide feedback in the writing classroom and to suggest potential locations for further research on technology-driven feedback practices”	“(1) Does AI-enhanced classroom have any significant effects on L2 learners' writing outcome? (2) Does AI-enhanced classroom have any significant effects on L2 learners' writing feedback literacy and engagement? (3) What perceptions did L2 learners have towards employing AI-enhanced classroom in writing instruction?”	Rad et al. (2023)
University students' perceptions of artificial intelligence-	“to explore the perceptions of Korean university students regarding artificial	“1. How do Korean university students perceive using AI-	Lee et al. (2024)

Title	Purpose	Research Question	Author
based tools for English writing courses	intelligence (AI)-based writing tools that include tools guided by machine learning, such as Google Translate and Naver Papago, and generative AI tools, such as Grammarly”	based tools in liberal arts English classes? 2. What are the advantages and disadvantages of using AI-based tools in liberal arts English classes?”	
From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill	“to investigate the impact of an AI-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill”	“•• Research question 1: How does ChatGPT-based writing instruction influence EFL teachers' self-efficacy? •• Research question 2: How does ChatGPT-based writing instruction influence EFL learners' writing skills? •• Research question 3: How does ChatGPT-based writing instruction influence the retention of EFL learners' writing skills?”	Ghafouri et al. (2024)

All studies focus on integrating AI tools, specifically ChatGPT or similar applications, into writing instruction to enhance writing skills. Each study aims to explore how AI can improve various aspects of writing, including self-editing, register knowledge, overall writing skills, and specific language writing skills (e.g., Chinese or English). Furthermore, while all studies share the common goal of leveraging AI tools to enhance writing skills, each has unique aspects such as target populations (e.g., low-income families, EFL students), specific AI applications (e.g., Wordtune, Google Translate), and additional factors like motivation, feedback literacy, and teacher self-efficacy.

Results

The results from the studies on writing skills are shown in Table 95.

Table 95

Writing Skills: Results

Title	Results	Author
Cultivating writing skills: the role of ChatGPT as a learning	“Preliminary results indicate that students found ChatGPT beneficial for acquiring formal register knowledge but perceived it as unnecessary for informal writing.	Punar Özçelik

Title	Results	Author
assistant—a case study	Additionally, the effectiveness of ChatGPT in teaching neutral register was questioned by the participants. The results of the study showed that ChatGPT has the potential to assist students in developing their writing abilities, particularly in formal register. Students used ChatGPT enthusiastically and actively for their writing tasks. They benefited from its suggestions and corrections to enhance the formal aspects of their writings. The study noted several difficulties, though, including technical difficulties and limitations in interpreting informal and neutral registers. Despite these drawbacks, students' varied viewpoints and insightful ideas emphasized the need for significant functional improvements to ChatGPT to make it a more useful learning tool for self-editing. ChatGPT can provide significant assistance to students in their writing tasks with careful evaluation and modifications. Students had perceived benefits of using ChatGPT for formal text corrections, while they reported its lack of didactic nature. They had some trustworthiness and reliability concerns. Students stated that ChatGPT's limitations in understanding neutral and informal registers posed potential challenges as a learning assistant, making it suitable only for specific types of texts. They reported the need for improvement in ChatGPT's algorithms and suggestions. While they had negative opinions on its impact on writing levels and meaning changes, its usefulness for timesaving and shortening long texts were noted as positive. This research contributes to the existing literature by shedding new light on the effects of AI-generated chatbots in register learning during the writing process, offering insights into their potential as learning assistants."	and Yangın Ekşi (2024)
Empowering Chinese Language Learners from Low-Income Families to Improve Their Chinese Writing with ChatGPT's Assistance Afterschool	"1. ChatGPT's effectiveness in improving writing skills: The findings revealed that Chat- GPT can assist Chinese language learners in enhancing their writing skills by correcting erroneous characters and developing well-structured sentences. 2. Empowerment through the use of ChatGPT: Participants expressed a sense of empowerment in using ChatGPT at home, as it provided them with valuable assistance when they encountered difficulties or felt lost in their language learning journey. 3. Writing score improvement after using ChatGPT: All of the participants showed an increase in their writing scores during the intervention. Participants with lower scores at baseline demonstrated immediate and significant improvements in their Chinese writing scores, while participants with higher scores at baseline presented a gradual and more moderate improvement in their writing scores."	Li et al. (2023)
Defining the Role of Artificial Intelligence in Improving English Writing Skills Among Indonesian Students	"The results showed that (1) Students' attitude toward the use of AI played a mediating role in explaining the effect of reading and feedback on writing skills, (2) Reading had an indirect effect through attitude on writing skills improvement, and (3) Lecturers feedback had an indirect effect through attitude on writing skills. / This study found that virtual teaching through AI with a positive attitude had	Kaharuddin et al. (2024)

Title	Results	Author
	an indirect significant effect on writing skills improvement. The total indirect effect of reading on writing was 1.140, which showed reading contributed 11.40% indirect effect to the writing skills improvement. In this case, the teaching of reading did not significantly affect writing skills without a positive attitude of students. The improvement should be boosted by a positive attitude toward the process. Feedback had indirect effects through attitude on writing. In this case, it cannot significantly affect writing skills without positive attitude. This study proved that virtual teaching of reading and feedback with a positive attitude had a significant effect on writing skills improvement. Improving writing skills requires good mastery of grammar and vocabulary, which can be developed by reading and feedback from lecturers. Furthermore, electronic devices can facilitate this process with the application of AI.”	
Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students	“Quantitative analysis reveals significant improvements in both writing skills and motivation among students who received AI-assisted instruction compared to the control group. The experimental group demonstrates enhanced proficiency in various aspects of writing, including organization, coherence, grammar, and vocabulary. Qualitative findings showcase diverse perspectives, ranging from recognition of AI's innovative instructional role and its positive influence on writing skills and motivation to concerns about contextual accuracy and over-reliance. Participants also reflect on the long-term impact and sustainability of AI-assisted instruction, emphasizing the need for ongoing development and adaptation of AI tools.”	Song and Song (2023)
Investigating EFL Students' Writing Skills Through Artificial Intelligence: Wordtune Application as a Tool	“The results showed that using Wordtune, students in the experimental group surpassed those in the control group. Quantitative findings demonstrated that using Wordtune, the experimental group upgraded their writing and outperformed the control group in the final writing exam. Qualitative findings indicate that using Wordtune, the experimental group made modest writing gains at the lexical and syntactic levels. While lexical gains included more concrete nouns, vivid adjectives, and precise verbs, sentence-structure gains included the increasing presence of complex phrases and complex and compound sentences. Finally, the impact of Wordtune on writing quality was the same for both male and female participants.”	Mahmud (2023)
Enhancing efl reading and writing through ai-powered tools: Design, implementation, and evaluation of an online course	“Results showed that a learning loop was created to connect online learning with offline practice; students had better optimal experience in student-centered presentations than in teacher-centered lectures; and language proficiency predicted semester grade and assignment quantity. This study has both theoretical and practical value; it serves as an example of how to design a quality synchronous online course on AI tools for EFL reading and writing based on theoretical approaches. / By incorporating AIpowered tools into real-world scenarios, students can see the practical application and benefits of these tools in their lives; by presenting their experiences with these tools to the class, they are engaged in interpersonal interaction, which results	Hsiao and Chang (2023)

Title	Results	Author
	in the highest level of engagement according to ICAP theory (Chi, 2009) and immediate feedback that fosters optimal experience according to flow theory. Second, the learning of any new tools (AI-powered tools in our case) requires a short and clear explanation of the rationale, a step-by-step demonstration, practice in and after class, reflecting on and sharing user experiences with the tools, and timely/direct feedback from teachers and peers.”	
Impact of ChatGPT on Learners in a L2 Writing Practicum: An Exploratory Investigation	“The findings revealed the affordance and potential applicability of the tool in L2 writing pedagogy. Additionally, the tool also showcased an automatic workflow that could maximize the efficiency in composing writing. However, participants generally expressed their concern with its threats to academic honesty and educational equity. The study impelled the reconceptualization of plagiarism in the new era, development of regulatory policies and pedagogical guidance to regulate proper utilization of the tool. Being a pioneering effort, the study accentuated future research directions for more insights into the application of ChatGPT in L2 learning, and the establishment of corresponding pedagogical adjustments. / students could easily grasp the basic skills to use ChatGPT in writing and improved their proficiency and capabilities through collaborative activities; students acknowledge the strength of ChatGPT and expressed more concern for educational equity.”	Yan (2023)
Using artificial intelligence to foster students’ writing feedback literacy, engagement, and outcome: A case of wordtune application	“By using Wordtune, an AI-based application, the experimental group was able to significantly improve their writing outcomes, engagement, and feedback literacy when compared to the control group. Furthermore, the Wordtune application received positive feedback from students interviewed about its ability to improve their writing outcomes, engagement, and feedback literacy. These findings will contribute to our understanding of AI in L2 learning and inform instructional considerations for Wordtune's application.”	Rad et al. (2023)
University students’ perceptions of artificial intelligence-based tools for English writing courses	“The study results indicate that these AI-based writing tools could improve English language learners (ELLs) writing skills. ELLs also noted the strengths and weaknesses of each AI-based tool, including the accessibility of translation machine learning and the error-checking capabilities of generative AI. However, interview data analysis indicates that the excessive use of AI-based writing tools could interfere with ELLs’ English writing process.”	Lee et al. (2024)
From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill	“The Independent Samples t-test results showed that CGWIP significantly enhanced teachers' self-efficacy compared to the control group. Also, the results of One Way ANCOVA revealed that CGWIP significantly improved learners' writing skills and that these effects persisted over time. The study implied that the protocol can nurture teachers' efficiency by helping them in various aspects of L2 writing instruction, including brainstorming,	Ghafouri et al. (2024)

Title	Results	Author
	revising, providing feedback, and assessment, which in turn, improves learners' writing skills.”	

Punar Özçelik and Yangın Ekşi (2024) find that students consider ChatGPT beneficial for formal register writing but unnecessary for informal writing. Students use ChatGPT enthusiastically for formal writing tasks, benefiting from its suggestions and corrections. However, they note limitations in interpreting informal and neutral registers and express concerns about trustworthiness and reliability. Li et al. (2023) report that ChatGPT effectively assists Chinese language learners in improving their writing skills, particularly in correcting characters and structuring sentences. Participants feel empowered using ChatGPT at home and show improved writing scores, with lower-scoring students demonstrating more significant improvements. Kaharuddin et al. (2024) demonstrate that students' attitudes toward AI mediate the effect of reading and feedback on writing skills improvement. Virtual teaching through AI, combined with positive student attitudes, indirectly but significantly affects writing skills improvement.

Song and Song (2023) reveal significant improvements in writing skills and motivation among students receiving AI-assisted instruction compared to a control group. Qualitative findings show diverse perspectives, including recognition of AI's innovative role and concerns about contextual accuracy and over-reliance. Mahmud (2023) finds that students using Wordtune outperform the control group in writing exams, with modest gains at lexical and syntactic levels. The impact is consistent across genders. Hsiao and Chang (2023) show that AI-powered tools create a learning loop connecting online learning with offline practice. Student-centered presentations lead to better optimal experiences than teacher-centered lectures, and language proficiency predicts semester grades and assignment quantity.

Yan (2023) reports that students easily grasp ChatGPT's basic use in writing and improve their proficiency through collaborative activities. Students acknowledge ChatGPT's strengths but express concerns about educational equity and academic honesty. Rad et al. (2023) demonstrate that the Wordtune application significantly improves students' writing outcomes, engagement, and feedback literacy compared to a control group.

Lee et al. (2024) indicate that AI-based writing tools can improve English language learners' writing skills, but excessive use might interfere with the writing process. Ghafouri et al. (2024) show that a ChatGPT-based writing instruction protocol significantly enhances teachers' self-efficacy and improves learners' writing skills, with effects persisting over time.

Implications

Table 96 provides the implications from these nine studies.

Table 96

Writing Skills: Implications

Title	Implications	Author
Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study	“Students’ varied viewpoints and insightful ideas emphasized the need for significant functional improvements to ChatGPT to make it a more useful learning tool for self-editing. ChatGPT can provide significant assistance to students in their writing tasks with careful evaluation and modifications.”	Punar Özçelik and Yangın Ekşi (2024)
Empowering Chinese Language Learners from Low-Income Families to Improve Their Chinese Writing with ChatGPT’s Assistance Afterschool	“These findings highlight that ChatGPT shows promise as a supportive tool for CLLs from low-income families, reducing educational inequality and promoting equitable access to language learning opportunities.	Li et al. (2023)
Defining the Role of Artificial Intelligence in Improving English Writing Skills Among Indonesian Students	“This study proved that virtual teaching of reading and feedback with a positive attitude had a significant effect on writing skills improvement. Improving writing skills requires good mastery of grammar and vocabulary, which can be developed by reading and feedback from lecturers. Furthermore, electronic devices can facilitate this process with the application of AI. In conclusion, students' skills in learning English using AI needed to be improved. This was	Kaharuddin et al. (2024)

Title	Implications	Author
	because students' attitude was shown to have a mediating role in explaining the effect of reading and teacher feedback on writing skills development. In this case, feedback had no direct influence on the development of writing skills. Therefore, building a positive attitude towards the use of AI in learning needs to be prioritized. Vocabulary development from reading and remedial exercises based on lecturers' feedback improved students' writing skills.”	
Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students	“The findings of this study carry important implications for the field of EFL, particularly in relation to the teaching of writing. The integration of AI tools into EFL writing courses is recommended as part of a technology-integrated language learning approach, as it aligns with innovative pedagogical practices and has the potential to significantly enhance students’ academic writing skills and writing motivation. To enhance EFL learners’ academic writing skills and motivation, it is highly encouraged for learners to actively incorporate AI tools into their writing activities. By leveraging the capabilities of AI, learners can receive immediate and targeted feedback, enabling them to identify areas for improvement and enhance their writing performance accordingly. This integration of AI into their writing practice has the potential to yield more efficient and effective learning outcomes. To provide engaging and motivating writing activities supported by AI, EFL educators and teachers should consider establishing dedicated AI-supported language learning classes. These classes can serve as platforms to train EFL learners on how to effectively utilize AI tools to improve their writing skills. Through providing guidance and support in the use of AI tools, educators can empower learners to take ownership of their learning process and make meaningful progress in their writing abilities. Moreover, organizing an AI-supported language learning class specifically tailored to writing activities can bring significant benefits to EFL learners. Within such a class, learners can receive AI-generated writing feedback and comments addressing various aspects of their written texts, including language usage, organization, and content. This comprehensive feedback can provide learners with specific guidance to address their writing challenges and refine their skills across different areas.”	Song and Song (2023)
Investigating EFL Students' Writing Skills Through Artificial Intelligence: Wordtune Application as a Tool	“The following recommendations are made to stakeholders. First, instructors should consider integrating AI-powered digital writing tools, such as Wordtune, into their writing classrooms. By explaining to students the merits of Wordtune rewrites over the original input sentences, instructors would get students to think and learn about the factors that distinguish strong sentences from weak sentences, and this consciousness-raising will lead to better writing. Second, students should be encouraged to use Wordtune rigorously in their out-of-class writing practice. When they see multiple and better rewrites of their input sentences, students naturally attempt to understand what made the rewrites better than the original; in this process, they cultivate good writing habits and adopt	Mahmud (2023)

Title	Implications	Author
Enhancing efl reading and writing through ai-powered tools: Design, implementation, and evaluation of an online course	effective writing techniques. Finally, instructors should inform students about the limitations of Wordtune because an uncritical reliance on a digital writing tool might put students on the path of becoming blind followers of writing technology, which would be counterproductive to their writing.” “Our results suggest the following implications for synchronous course design in this post-pandemic era. First, basing the design of a brand-new course on learning theories is beneficial, because it helps ensure that the course is effective in promoting meaningful learning and achieving intended outcomes. Second, the learning of any new tools (AI-powered tools in our case) requires a short and clear explanation of the rationale, a step-by-step demonstration, practice in and after class, reflecting on and sharing user experiences with the tools, and timely/direct feedback from teachers and peers. Third, striking a balance between teacher-centered and student-centered teaching approaches is recommended. An integrated view of organizing both teacher-centered and studentcentered activities (Mascolo, 2009) seems preferable to achieve meaningful online learning experiences with adequate teaching presence (e.g. clear goals and scaffolding), cognitive presence (e.g. discussion on challenging questions), and social presence (e.g. teacher-student and student-student interaction) (Garrison et al., 1999), and to foster flow experiences in which optimal and immersive engagement occurs throughout the learning activities (Csikszentmihalyi, 1990). Finally, it is necessary to recognize, admit, and address difficulties students face in synchronous online course, especially in teacher-centered lectures, such as overwhelming course contents and feelings of fatigue. Teachers could split course contents into smaller parts, provide more scaffolding and support, ask for genuine feedback from students, and reach out to those whose English language proficiency lags behind.”	Hsiao and Chang (2023)
Impact of ChatGPT on Learners in a L2 Writing Practicum: An Exploratory Investigation	“For pedagogy, extra attention should be paid to the application of tools such as Chat- GPT in authentic learning environment. Pedagogical guidance for proper utilization of state-of-the-art technologies should be implemented, regulatory policies in educational institutions should be reformulated, and plagiarism detection literacy of lecturers should be reinforced. For an undergraduate EFL program, ChatGPT’s performance is comparable to that of an average or above-average student in composing writing for a given topic. The research suggests educators to adopt a rational perspective towards the emergence of ChatGPT and follow-up applications taking advantage of the powerful GPT-3.5 language model. Instead of obstructing students’ exposure and access to such tools, lecturers should develop and implement a protocol of application to make proper use of the technology.”	Yan (2023)
Using artificial intelligence to foster students’ writing feedback literacy,	“This study elucidates ideas that might improve second-language learners’ writing feedback literacy, writing engagement, and writing outcomes and contextualizes them within the perspective of AI. The suggested concepts	Rad et al. (2023)

Title	Implications	Author
engagement, and outcome: A case of wordtune application	might benefit the field on both a broad and a focused level. This study might, on the surface, be used as a guide to enhance feedback efficacy and its perceived value for the efficient use of AI to enhance writing feedback literacy, thereby reducing the tension between instructors' feedback practices and students' expectations through personalized feedback. Instructors should adopt dialogic teaching, where student-teacher interaction is less emphasized to establish tolerance and acceptance with students as active members of the process instead of boosting themselves to a higher position with feedback as a unilateral process, to best promote and sustain student writing feedback literacy. In sum, Wordtune is well regarded for supporting students in their understanding of such information during the writing process and benefits from new data displayed in AI."	
University students' perceptions of artificial intelligence-based tools for English writing courses	"These findings suggest that using AI-based learning tools in university English writing classes requires careful consideration. Rather than relying solely on these AI-based tools, university English classes should be designed to incorporate them effectively. Specifically, proper guidance for integrating AI-based English tools can positively impact the learning process, enhancing university students' English writing without hampering their language development (Gayed et al., 2022; Klekovkina & Denié-Higney, 2022)."	S2
From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill	"Several theoretical and pedagogical implications can be induced from the present study. The results implied that ChatGPT-based instruction would remind the teachers about their capabilities to overcome challenges by influencing their students' outcomes while triggering a sense of self-efficacy. Further, the study contributes to the current positive psychology-laden trends of L2 research by suggesting that if applied in a well designed and controlled manner, ChatGPT can benefit EFL teachers by increasing their self-efficacy, which is correlated with other notions such as reduced stress and burnout (Jerusalem & Schwarzer, 1992; Kim & Burić, 2020; Pfitzner-Eden, 2016; Skaalvik & Skaalvik, 2010), creative teaching methods (Thurlings et al., 2015), perseverance (Heng & Chu, 2023; Milner & Hoy, 2003; Shao, 2023), job satisfaction (Demir, 2018), and similar psycho-emotional entities the combination of which promotes teachers' wellbeing (von Muenchhausen et al., 2021; Xiyun et al., 2022). Additionally, based on the results and in line with recent AI-driven L2 studies, we argue that ChatGPT is beneficial for L2 writing through features such as automated essay scoring, intelligent tutoring systems, or natural language processing (Ali et al., 2023; H.S. Kim et al., 2021; Kohnke et al., 2023; Lin & Mubarak, 2021). ChatGPT can also adapt to the users' needs, interests, and proficiency levels and provide immediate, personalized, relevant feedback and support. ChatGPT can also interact with the users in a human-like manner, using natural and coherent language, and engage them in a dialogue. These advantages make ChatGPT a more appealing and effective tool for L2 writing instruction, as it can cater to teachers' and students' diverse and dynamic needs."	Ghafouri et al. (2024)

Punar Özçelik and Yangın Ekşi (2024) emphasize the need for significant functional improvements to ChatGPT to make it a more useful self-editing tool. They suggest that with careful evaluation and modifications, ChatGPT can provide significant assistance in writing tasks. Li et al. (2023) highlight the potential of large language models, like ChatGPT, in supporting writing accuracy assessment for language learners. They suggest improvements in grammar structure detection and word usage, emphasizing that these tools can help reduce educational inequality by offering equitable access to language learning opportunities for CLLs from low-income families. However, they also stress that human evaluations remain crucial for nuanced language analysis and context-specific interpretations. Kaharuddin et al. (2024) prove that virtual teaching of reading and feedback, combined with a positive attitude, significantly improves writing skills. They emphasize the importance of building a positive attitude towards AI use in learning and prioritizing vocabulary development and remedial exercises based on feedback.

Song and Song (2023) recommend integrating AI tools into EFL writing courses as part of technology-integrated language learning. They suggest establishing AI-supported language learning classes to train learners in effective AI tool use and provide comprehensive AI-generated feedback on various aspects of writing. Mahmud (2023) recommends integrating AI-powered digital writing tools like Wordtune into writing classrooms. They suggest encouraging students to use such tools rigorously in out-of-class practice while also informing them about the tools' limitations to prevent uncritical reliance.

Hsiao and Chang (2023) emphasize the importance of basing course design on learning theories, providing clear explanations and demonstrations for new tools, balancing teacher-centered and student-centered approaches, and addressing

difficulties in synchronous online courses. Yan (2023) calls for implementing pedagogical guidance for proper utilization of technologies like ChatGPT, reformulating regulatory policies, and reinforcing plagiarism detection literacy among lecturers. They suggest adopting a rational perspective towards ChatGPT and developing application protocols for its proper use. Rad et al. (2023) suggest using AI to enhance writing feedback literacy and reduce tension between instructors' feedback practices and students' expectations. They recommend adopting dialogic teaching to promote student writing feedback literacy.

Lee et al. (2024) emphasize the need for careful consideration when using AI-based learning tools in university English writing classes. They suggest designing classes to incorporate these tools effectively with proper guidance to enhance writing without hampering language development. Ghafouri et al. (2024) highlight the potential of ChatGPT-based instruction to increase teachers' self-efficacy, which correlates with reduced stress, creative teaching methods, and job satisfaction. They argue that ChatGPT can benefit L2 writing through features like automated essay scoring and personalized feedback, making it an appealing and effective tool for L2 writing instruction.

Writing Performance

Purpose

There are four studies that focus on the essay writing performance of students using AI tools.

Table 97

Writing Performance: Purpose

Title	Purpose	Research Question	Author
ChatGPT-3.5 as writing assistance in students' essays	“this research investigated whether ChatGPT would improve students’ essay grades, reduce writing time,	Not stated.	Bašić et al. (2023)

Title	Purpose	Research Question	Author
	and affect text authenticity. / to examine students' essay-writing performances with or without ChatGPT as an essay-writing assistance tool"		
Feedback sources in essay writing: peer-generated or AI-generated feedback?	"to use ChatGPT as a source of feedback for students' argumentative essay writing tasks and to compare the quality of ChatGPT-generated feedback with peer feedback"	"RQ1. To what extent does the quality of peer-generated and ChatGPT-generated feedback differ in the context of essay writing? RQ2. Does a relationship exist between the quality of essay writing performance and the quality of feedback generated by peers and ChatGPT?"	Banihashem et al. (2024)
Using AI to Evaluate a Competency-Based Online Writing Course in Nursing	"to investigate the impact of a 14-week fully online competency-based writing course on students' self-efficacy, task value, and writing performance"	"How does a discipline-specific online writing course affect students' scholarly writing performance, self-efficacy, and task value?"	Wolf and Wolf (2023)
A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning	"to investigate the use and acceptance of ChatGPT-4 in students' EFL writing"	"1. Are there any significant differences in the writing performance of the experimental group compared with the control group after the use of ChatGPT-4? 2. Are there any significant differences in the writing performance before and after the employment of ChatGPT-4? 3. Would students' perceptions toward ChatGPT-4 use significantly change after the implementation of ChatGPT-4?"	Boudouaia et al. (2024)

The first study (Bašić et al., 2023) aims to examine ChatGPT's impact on students' acquisition of register knowledge across various writing tasks. The second study (Banihashem et al., 2024) explores how ChatGPT can support and empower Chinese language learners from low-income families in enhancing their writing skills. The third study (Wolf & Wolf, 2023) investigates the improvement of English writing skills through the integration of AI into reading practices and feedback

provision. The fourth study (Boudouaia et al., 2024) evaluates the impact of AI-assisted language learning on Chinese EFL students' writing skills and motivation.

Results

Table 98 shares the results of studies that focused on essay writing performance.

Table 98

Writing Performance: Results

Title	Results	Author
ChatGPT-3.5 as writing assistance in students' essay	"The average essay grade was C for both control (traditional essay-writing, n = 9) and experimental (ChatGPT-assisted essay-writing, n = 9) groups. None of the predictors affected essay scores: group, writing duration, study module, and GPA. The text unauthenticity was slightly higher in the experimental group, but the similarity among essays was generally low in the overall sample. In the experimental group, the AI classifier recognized more potential AI-generated texts. Our results demonstrate that the ChatGPT group did not perform better in either of the indicators; the students did not deliver higher quality content, did not write faster, nor had a higher degree of authentic text."	Bašić et al. (2023)
Feedback sources in essay writing: peer-generated or AI-generated feedback?	"The results showed a significant difference between feedback generated by ChatGPT and peers. While ChatGPT provided more descriptive feedback including information about how the essay is written, peers provided feedback including information about identification of the problem in the essay. The overarching look at the results suggests a potential complementary role for ChatGPT and students in the feedback process. For the first research question, the results revealed a disparity in feedback quality when comparing peer-generated feedback to feedback generated by ChatGPT. Peer feedback demonstrated higher quality compared to ChatGPT-generated feedback. This discrepancy is attributed primarily to variations in the descriptive and problem-identification features of the feedback. ChatGPT tended to provide more descriptive feedback, often including elements such as summarizing the content of the essay. This inclination towards descriptive feedback could be related to ChatGPT's capacity to analyze and synthesize textual information effectively. In contrast, students' feedback content entailed high quality regarding identifying specific issues and areas for improvement. Peers outperformance compared to ChatGPT in identifying problems within the essays could be related to humans' potential in cognitive skills, critical thinking abilities, and contextual understanding (e.g., Korteling et al., 2021; Lamb et al., 2019). This means that students, with their contextual knowledge and critical thinking skills, may be	Banihashem et al. (2024)

Title	Results	Author
	better equipped to identify issues within the essays that ChatGPT may overlook. Regarding the relationship between the quality of essays and the quality of the feedback provided by ChatGPT and peers, we found no overall significant relationship.”	
Using AI to Evaluate a Competency-Based Online Writing Course in Nursing	“The results showed statistically significant gains in self-efficacy and writing performance with large effect sizes. This study addresses the gap in nursing education regarding the assessment of online, research-based writing interventions on students' scholarly writing capacity. this study indicates that using automated essay scoring enabled by artificial intelligence can provide an efficient, reliable, and valid way to evaluate students' writing performance for educational research.”	Wolf and Wolf (2023)
A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning	“The results showed that the experimental group (EG) outperformed the control group (CG). Besides, the findings revealed that students in the EG in post-test outperformed their pre-test scores. The findings also revealed substantial improvements in the EG's views of perceived usefulness, perceived ease of use, attitudes, and behavioral intention. According to the results, ChatGPT-4 helped boost students' EFL writing skills, which ultimately led to their acceptance. Students appear particularly interested in ChatGPT-4 because of its potential usefulness in putting what they learn about EFL writing into practice. Some suggestions and recommendations were provided.”	Boudouaia et al. (2024)

The first study (Bašić et al., 2023) found no significant difference in essay quality, writing speed, or text authenticity between the control group and the ChatGPT-assisted group, with both averaging a grade of C. In the second study (Banihashem et al., 2024), it was shown that while ChatGPT provided more descriptive feedback, peer feedback was of higher quality in identifying specific problems, suggesting a complementary role for both sources in the feedback process. The third study (Wolf & Wolf, 2023) revealed statistically significant improvements in students' self-efficacy and writing performance, demonstrating the effectiveness of AI-enabled automated essay scoring in evaluating writing performance. Lastly, the fourth study (Boudouaia et al., 2024) showed that the experimental group using ChatGPT-4 outperformed the control group, with notable improvements in writing

skills and positive perceptions of the tool's usefulness, ease of use, and overall acceptance.

Implications

The implications from these studies are in Table 99.

Table 99

Writing Performance: Implications

Title	Implications	Author
ChatGPT-3.5 as writing assistance in students' essay	"We anticipate that these results can relieve some concerns about this tool's usage in academic writing. ChatGPT-assisted writing could depend on the previous knowledge and skills of the user, which might, in certain instances, lead to confusion in inexperienced users and result in poorer essay writing performance. The academia and media concern about this tool might be unjustified, as, in our example, the ChatGPT was found to perform similarly to any web-based search: the more you know—the more you will find. In some ways, instead of providing structure and facilitating writing, it could distract students and make them underperform."	Bašić et al. (2023)
Feedback sources in essay writing: peer-generated or AI-generated feedback?	"The implications of this study are valuable, shedding light on the prospective use of ChatGPT as a feedback source, particularly for complex tasks like argumentative essay writing. From a practical perspective of higher education, the study's findings offer insights into the potential integration of ChatGPT as a feedback source within higher education contexts. The discovery that ChatGPT's feedback quality could potentially complement peer feedback highlights its applicability for enhancing feedback practices in higher education. This holds particular promise for courses with substantial enrolments and essay-writing components, providing teachers with a feasible alternative for delivering constructive feedback to a larger number of students. From an educational standpoint, our research findings advocate for the integration of ChatGPT as a feedback resource with peer feedback within higher education environments for essay writing tasks since there is a complementary role potential for peer-generated and ChatGPT-generated feedback. This approach holds the potential to alleviate the workload burden on teachers, particularly in the context of online courses with a significant number of students."	Banihashem et al. (2024)
Using AI to Evaluate a Competency-Based Online Writing Course in Nursing	"There are four important implications for practice based on the reviewed literature and the results of this study. First, nursing faculty should offer students required writing and writing-intensive courses based on cognitive apprenticeship principles. Second, this study affirms Wornock's (2009) assertion that the migration of writing pedagogy to an online context is effective. Faculty can use best practices from research on composition and writing	Wolf and Wolf (2023)

Title	Implications	Author
	instruction to inform their online instruction. Third, a competency-based approach that provides students with a flexible path to mastery can operate within a traditional time-based semester online. Instructors who support the development of nursing students' writing and thinking skills can empower students from a wide range of educational backgrounds. The scale of help is a promising strategy for students from underrepresented and low socioeconomic groups to enter the profession and enable their ongoing success in graduate school and as future leaders, scholars, faculty, and researchers. Fourth, this study indicates that using automated essay scoring enabled by artificial intelligence can provide an efficient, reliable, and valid way to evaluate students' writing performance for educational research. This has significant implications for evaluating writing, not only allowing this study to be replicated but also extended into longitudinal studies that track the development of writing performance across a curriculum. Additionally, this instrument may help researchers further explore the relationship between writing performance, critical thinking, clinical judgment, and decision-making. The administration of automated essay scoring at the beginning and end of a course or program can also provide students with data points to reflect upon their writing progress over time."	
A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning	"EFL teachers are encouraged to incorporate ChatGPT-4 into their lessons to provide students an opportunity to participate. Teachers should model the ethical use of ChatGPT-4 and give students the skills they need to confidently adopt the tool in the classroom. It is possible that holding workshops, providing training, and rolling out ChatGPT-4 in stages could all help achieve this goal. Teachers should encourage their students to have good intentions. Potentially enhanced by demonstrating its usefulness and advantages, the use of ChatGPT-4 to study EFL writing has the potential to be more effective. Furthermore, it is presumed that course or exercise designers in language curricula can emphasize and promote students' positive attitudes, sense of ease, and ability to implement via the use of ChatGPT-4. This allows for greater linguistic and technical fluency in the language generally. Courses should be flexible so that students can take charge of their own learning experiences. One approach would be to offer students the background and resources they need to effectively use ChatGPT-4 in and out of the classroom."	Boudouaia et al. (2024)

The implications of these studies highlight the potential and challenges of integrating AI tools like ChatGPT into educational settings. The study on ChatGPT-3.5 (Bašić et al., 2023) suggests that while concerns about its use in academic writing may be overstated, it may confuse inexperienced users and lead to poorer

performance, indicating a need for more structured implementation. The research comparing peer-generated and AI-generated feedback (Banihashem et al., 2024), reveals that ChatGPT can complement peer feedback in higher education, offering a scalable solution for providing constructive feedback, especially in large classes.

The study on AI evaluation in nursing education (Wolf & Wolf, 2023) emphasizes the effectiveness of online, competency-based writing courses and the reliability of automated essay scoring, which can aid in longitudinal studies and support diverse student populations. Lastly, the study on ChatGPT-4 in EFL learning (Boudouaia et al., 2024) encourages the incorporation of AI tools in language instruction, recommending staged implementation, ethical usage training, and flexible course designs to enhance students' writing skills and positive attitudes towards AI-assisted learning.

Writing Process

Purpose

Four studies focused on the writing process.

Table 100

Writing Process: Purpose

Title	Purpose	Research Question	Author
Not quite eye to A.I.: student and teacher perspectives on the use of generative artificial intelligence in the writing process	“With the goal of identifying some common expectations, the purpose of this study is to explore student and teacher perspectives of using GenAI for various tasks in the writing process, including brainstorming, outlining, writing, and revising done by students, and evaluation and feedback done by teachers. This study reports the results of a survey comparing educators’ (n = 68) and university students’ (n = 158) perceptions on the	“1. What are undergraduate students’ and teachers’ perspectives on using GenAI in the writing process (brainstorming, outlining, writing, revising, evaluation, and feedback)? 2. How do student and teacher perspectives on the use of GenAI in the writing process compare?”	Barrett and Pack (2023)

Title	Purpose	Research Question	Author
	appropriate use of GenAI in the writing process.”		
“AI, will you help?” How learners use Artificial Intelligence when writing	“to investigate how successful advanced students are in determining whether an essay was written by AI tools or by a human and how much AI assistance they need to summarize, generate text and write from prompts when trained to use an AI-assisted word processor”	“1. How well do advanced students distinguish human-generated from AI-generated essays? What criteria do they think are decisive about Artificial Intelligence authorship? 2. When trained in the use of an AI-assisted word processor, how do advanced students make use of summarising, generating text and writing from prompts? 3. How do students’ essays differ linguistically after training in the use of AI tools from those written before treatment? Do they show differences in burstiness and perplexity?”	Krajka and Olszak (2024)
Human-AI collaboration patterns in AI-assisted academic writing	“to examine the nature of human-AI interactions in academic writing, specifically investigating the strategies doctoral students employ when collaborating with a GAI-powered assisting tool”	“RQ1. What are the strategies and patterns doctoral students employ in their academic writing processes when collaborating with GAI-assisted writing tools? RQ1a) What are the key hidden states in GAI-assisted academic writing? RQ1b) What are the predominant patterns in GAI-assisted academic writing? RQ2. How is the effectiveness of each identified strategic pattern in GAI-assisted academic writing characterised by its specific nature and underlying attributes?”	Nguyen et al. (2024)
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	“to investigate and assess the impact of integrating ChatGPT, a generative artificial intelligence tool, into an inorganic chemistry classroom for an 8-page writing assignment on selected elements from the periodic table / to explore student experiences, perceptions, and the utility of ChatGPT across different stages of the writing process, considering its role in idea generation, topic verification, proofreading, and editing while also investigating student opinions on the ethics of using ChatGPT in academic writing”	“How can ChatGPT be effectively utilized as a writing tool? What are the potential benefits associated with its use? How does it impact the learning outcomes and overall writing proficiency of students? How do students use it if given a free range? What do students think of using generative AI for writing assignments? Can we learn about student opinions on the ethical use of ChatGPT? Does it produce verifiable information?”	Rojas (2024)

Barrett and Pack (2023) aim to explore and compare student and teacher perspectives on using generative AI (GenAI) for various tasks in the writing process, including brainstorming, outlining, writing, revising, evaluation, and feedback. Their goal is to identify common expectations regarding the appropriate use of GenAI in writing. Krajka and Olszak (2024) seek to investigate how successfully advanced students can distinguish between AI-generated and human-written essays, and examine how students use AI assistance for summarizing, generating text, and writing from prompts when trained to use an AI-assisted word processor. They also aim to analyze linguistic differences in students' essays before and after AI tool training. Nguyen et al. (2024) aim to examine the nature of human-AI interactions in academic writing, focusing on the strategies doctoral students employ when collaborating with AI-assisted writing tools. They investigate key hidden states and predominant patterns in AI-assisted academic writing, as well as the effectiveness of identified strategic patterns. Rojas (2024) seeks to investigate and assess the impact of integrating ChatGPT into an inorganic chemistry classroom for a writing assignment. The study aims to explore student experiences, perceptions, and the utility of ChatGPT across different stages of the writing process, including idea generation, topic verification, proofreading, and editing. It also investigates student opinions on the ethics of using ChatGPT in academic writing and its impact on learning outcomes and writing proficiency.

Results

The results of the studies on the writing process are in Table 101.

Table 101*Writing Process: Results*

Title	Results	Author
Not quite eye to A.I.: student and teacher perspectives on the use of generative artificial intelligence in the writing process	“Results indicate minor disagreement between students and teachers on the acceptable use of GenAI tools in the writing process, as well as classroom and institutional-level lack of preparedness for GenAI. The prevailing narrative in our results demonstrated that for all the steps of the writing process, students and teachers generally agreed that using GenAI to brainstorm ideas or model answers, or as a form of cognitive offloading for tasks that the user is already competent in, is acceptable. Conversely, using GenAI to complete writing task assignments, with or without disclosing the use of GenAI, is unacceptable. Both students and teachers perceived GenAI use to be more acceptable in the early stages of the writing process (i.e., brainstorming and outlining) than in later stages. These results suggest that use of GenAI for writing purposes is viewed as more acceptable when it is fulfilling a supportive role focused on idea generation and organization rather than when leveraged as an automatic writing completion tool. As to differences in perspectives, students tended to disagree less than teachers that using GenAI without disclosure was appropriate, and teachers tended to disagree less than students about using GenAI to model feedback or for evaluation of student writing.”	Barrett and Pack (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“The findings of this study indicate the high linguistic sensitivity of the research participants to factors regarding language and layout, which allows them to distinguish human authors from AI-powered texts. The study results reveal some linguistic differences in students’ reports before and after the treatment. However, the marked variations are rather spontaneous and not obvious. In some AI-enhanced reports, there are even some features of human writing, like high burstiness. As regards specific linguistic differences between students’ written assignments before and after the treatment, the four aspects of sentence length, sentence complexity, word frequency and the ratio of unique words inserted in the texts were used as objective measures when comparing students’ texts. Generally, when analysing the sentence length of the students’ reports after the quasi experimental treatment, most of the study participants formed slightly longer sentences in the post texts (67%, 10 out of 15 students), while only some students shortened their sentences (33%, 5 out of 15 students). It can be assumed that the application of the AI tools in the writing practice of the quasi experimental group of students helped to improve and enrich students’ writing skills in general.”	Krajka and Olszak (2024)
Human-AI collaboration patterns in AI-assisted academic writing	“Findings indicate that doctoral students engaging in iterative, highly interactive processes with the GAI-powered assisting tool generally achieve better performance in the writing task. In contrast, those who use GAI merely as a supplementary information source, maintaining a linear writing approach, tend to get lower writing performance. This study points to the need for further investigations into human-AI collaboration in	Nguyen et al. (2024)

Title	Results	Author
	learning in higher education, with implications for tailored educational strategies and solutions.”	
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	“This study's results reveal positive student experiences and perceptions of integrating ChatGPT for a writing-intensive assignment, showcasing its effectiveness in idea generation, topic verification, proofreading, and editing. / The integration of ChatGPT into the inorganic chemistry classroom has yielded valuable insights into its multifaceted utility for writing assignments. A pre-assignment survey revealed a diverse range of prior experiences among students, with a majority expressing prior curiosity-driven exploration of ChatGPT. For the effectiveness of ChatGPT in the first assignment, focused on topic generation and verification, student responses emphasized the ease and efficiency brought to the writing process, highlighting ChatGPT’s role as a facilitator for idea development. Subsequent surveys further illustrated students’ positive perceptions, with the second survey indicating widespread agreement on the usefulness of ChatGPT for scientific writing. Ethical considerations emerged, as students unanimously expressed reservations about using ChatGPT to generate entire text passages, emphasizing the importance of maintaining active student input. The final survey underscored ChatGPT’s pivotal roles in generating ideas and facilitating proofreading and editing, showcasing its significance as a supplementary tool. Looking ahead, students exhibited varied inclinations toward future ChatGPT use, reflecting a nuanced approach shaped by their experiences in this chemistry class.”	Rojas (2024)

Barrett and Pack (2023) find minor disagreements between students and teachers on the acceptable use of GenAI tools in the writing process. Both groups generally agree that using GenAI for brainstorming, modeling answers, or cognitive offloading is acceptable while using it to complete writing tasks is unacceptable. GenAI use is viewed as more acceptable in the early stages of writing (brainstorming and outlining) than in the later stages. Students tend to be more accepting of undisclosed GenAI use, while teachers are more open to using GenAI for feedback and evaluation.

Krajka and Olszak (2024) report that participants in their study show high linguistic sensitivity in distinguishing human-authored from AI-generated texts. After treatment with AI tools, the results show that most students (67%) formed slightly longer sentences in their writing. The use of AI tools appears to have

improved and enriched students' writing skills generally, though linguistic differences before and after treatment were not always obvious.

Nguyen et al. (2024) find that doctoral students who engage in iterative, highly interactive processes with AI-powered writing tools generally achieve better performance in writing tasks. In contrast, those who use AI merely as a supplementary information source, maintaining a linear writing approach, tend to perform lower in writing tasks.

Rojas (2024) reveals positive student experiences and perceptions of integrating ChatGPT for a writing-intensive assignment in inorganic chemistry. Students found ChatGPT effective for idea generation, topic verification, proofreading, and editing. They emphasized its role in facilitating idea development and improving writing efficiency. However, students expressed ethical reservations about using ChatGPT to generate entire text passages, stressing the importance of maintaining active student input. The study also had results which demonstrate that there are varied student inclinations toward the use of ChatGPT for their future works.

Implications

Table 102 provides the implications of these studies.

Table 102

Writing Process: Implications

Title	Implications	Author
Not quite eye to A.I.: student and teacher perspectives on the use of generative artificial intelligence in the writing process	“The findings of this study underscore the importance of students and teachers being transparent in their use of GenAI tools. Using AI without disclosure by both students and teachers, was considered the least acceptable practice. Yeo (2023) suggests that educators “accept and befriend [GenAI] by showing learners how to use AI authoring ethically and gainfully to achieve their learning intentions and goals” (p. 10). If educators are transparent and clearly state and model acceptable uses of GenAI, then students may follow. While there is a nascent body of literature offering practical suggestions and advice for using GenAI in education by	Barrett and Pack (2023)

Title	Implications	Author
	educators, learners, and researchers, an evidence-centered framework for leveraging GenAI in writing and in higher education in general needs further attention.”	
“AI, will you help?” How learners use Artificial Intelligence when writing	“There is little doubt that the fields of language teaching and learning could be significantly enhanced by the ongoing development of Artificial Intelligence capabilities. According to the study’s findings, the students were not very successful at distinguishing between AL and human-generated writing with accuracy rates between 40-80%. They tended to assess human writing based on technical accuracy and error and AI writing based on verbal complexity. Applying AI-powered tools could help students improve their writing competence and bring about changes to the way academic writing is taught as a whole. However, as the study’s findings show, they can also hinder the process of improving writing abilities because they lack imagination, use straightforward and sometimes repetitious vocabulary, respond slowly, or make minor grammatical errors. The research reveals that even highly proficient, almost bilingual, academic students encounter considerable difficulties when attempting to differentiate human writing from AI-assisted written texts. These challenges arise when analyzing formal and informal texts.”	Krajka and Olszak (2024)
Human-AI collaboration patterns in AI-assisted academic writing	“This study offers several implications for different stakeholders in higher education concerning the integration of AI tools like the GAI-powered tool in academic writing. For educators, the data reveal the specific cognitive and metacognitive skills contributing to effective collaboration with AI. This information is vital for developing targeted pedagogical strategies and interventions to enhance students’ skill sets. For instructional designers and developers in the EdTech industry, the study identifies key functionalities particularly useful to students in the writing process. The research suggests that AI-enabled writing tools should promote adaptive coordination among users, as this approach correlates with higher writing performance. In terms of policymaking, the findings offer a substantial foundation for formulating policies on the inclusion of AI technologies in academic settings. Policymakers must consider the academic benefits and ethical considerations such as data privacy and potential misuse. Accordingly, comprehensive frameworks must be developed to govern the responsible use of these technologies. Finally, for students, particularly at the doctoral level, the study provides invaluable guidance for optimising their use of AI tools in academic writing. The research underscores the need for a proactive, dynamic engagement with the AI tool, as this mode of interaction has been shown to result in higher-quality academic work.”	Nguyen et al. (2024)
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	“Based on the findings of this study, several implications for practice emerge for educators seeking to integrate generative AI tools such as ChatGPT into college-level writing assignments. Primarily, students favored topic generation as a useful mechanism for beginning their writing assignments. For amateurs or those who lack extensive subject knowledge, ChatGPT can quickly summarize a vast array of content and	Rojas (2024)

Title	Implications	Author
	condense the material quickly for students. Importantly, the content generated by ChatGPT proved to be reliable, at least in these 53 student examples. Overall, student writing was improved through this process largely because students were provided a variety of mechanisms to revise their work. This process is achievable with writing centers, peer review, instructor feedback, etc., and ChatGPT appears to reliably contribute to that array of resources for editing writing. As a supplementary tool, ChatGPT can provide additional support and resources to facilitate writing assignments. However, it will be crucial for educators to address ethical considerations associated with the use of generative AI tools such as ChatGPT. By emphasizing the importance of active student involvement in the writing process and discouraging the use of ChatGPT to generate entire text passages, educators can promote academic integrity to uphold standards of originality. In STEM disciplines such as chemistry, where writing assignments play an important role in developing critical thinking and communication skills, integrating ChatGPT can offer students novel opportunities to engage with course material and express their understanding in innovative ways. Importantly, the use of generative AI in the writing process poses the risk of bypassing critical thinking and inhibiting some learning opportunities. By relying solely on AI-generated content, students may miss the chance to develop their own creative writing skills. Additionally, without facing the challenge of making mistakes and learning from them, students may struggle to improve their writing abilities independently. Educators can tailor ChatGPT-based activities to align with specific learning objectives and foster deeper learning experiences. Incorporating ChatGPT into writing assignments exposes students to emerging technologies and enhances their technological literacy. Educators can use ChatGPT as a catalyst for discussions on the ethical implications, limitations, and potential applications of AI in academic and professional contexts, empowering students to critically evaluate and responsibly use AI tools in their future endeavors.”	

Barrett and Pack (2023) emphasize the importance of transparency in using generative AI (GenAI) tools for both students and teachers. They suggest that educators should openly demonstrate acceptable uses of GenAI, which may encourage students to follow suit. The authors point out that while there are emerging suggestions for using GenAI in education, there's a need for a more comprehensive, evidence-based framework for leveraging GenAI in writing and higher education. Krajka and Olszak (2024) indicate that AI-powered tools can potentially enhance students' writing competence and transform the teaching of

academic writing. However, they also note that these tools can hinder writing skill development due to limitations such as lack of imagination, repetitive vocabulary, slow responses, and minor grammatical errors. The authors highlight that even highly proficient students struggle to differentiate between human and AI-assisted writing, suggesting implications for how writing is assessed and taught.

Nguyen et al. (2024) present implications for various stakeholders in higher education. For educators, they suggest developing targeted strategies to enhance students' cognitive and metacognitive skills for effective AI collaboration. They recommend that EdTech developers focus on creating tools that promote adaptive coordination. For policymakers, the study provides a basis for developing comprehensive frameworks governing the responsible use of AI in academic settings. For students, particularly at the doctoral level, the research offers guidance on optimizing AI tool use in academic writing. Rojas (2024) suggests that integrating ChatGPT into college-level writing assignments can be beneficial, particularly for topic generation and revision. The author recommends that educators address ethical considerations and emphasize active student involvement to maintain academic integrity. Rojas notes that while ChatGPT can enhance engagement with course material, there's a risk of bypassing critical thinking and inhibiting learning opportunities. The study suggests using ChatGPT as a catalyst for discussions on AI's ethical implications and applications in academic and professional contexts.

AI-Enhanced Classroom

Purpose

There are six studies that aim to understand the features of AI-enhanced classrooms.

Table 103*AI-Enhanced Classroom: Purpose*

Title	Purpose	Research Question	Author
Writing, creativity, and artificial intelligence. ChatGPT in the university context	“to study the creative potential of Artificial Intelligence (AI) for writing skills in an educational context / to provide evidence on the use of AI and contribute to its integration in the classroom as support for the teaching-learning process.”	“• To find out the level of creativity indicators (fluency, flexibility, and narrative originality) of AI systems and students, based on the sample of the Creative Imagination Test for Adults (PIC-A) by Artola et al. (2012). • To compare the creativity of the AI systems with that of the students, based on the indicators of fluency, flexibility, and narrative originality. • To compare the scores of the creativity indicators obtained by the students in the two phases of application of the tests, the first without any type of aid and the second using an AI system (ChatGPT).”	Vicente-Yagüe Jara et al. (2023)
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“(1) to capture the collective wisdom of business communication instructors / (2) to develop an expert-driven framework of AI literacy for communicators.”	“1. What concerns do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 2. To what degree do business communication instructors experience anxiety towards teaching or integrating AI-assisted writing into their courses? 3. What perceived benefits do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 4. To what degree do business communication professors perceive AI-assisted writing (e.g., ChatGPT) will be accepted in the near future for workplace writing?”	Cardon et al. (2023)
Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive?	“to map perception, obstacles, and recommendations for optimizing use of AI in teaching academic writing in Indonesian”	“RQ1. How do Indonesian students perceive the usefulness of AI technology in their academic writing class? RQ2. How do Indonesian students perceive the ease of using AI technology in their academic writing class? RQ3. What do Indonesian students’ attitudes toward the usage of AI technology in their academic writing class?”	Utami et al. (2023)

Title	Purpose	Research Question	Author
Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application	"to describe how AI is now used to provide feedback in the writing classroom and to suggest potential locations for further research on technology-driven feedback practices"	"(1) Does AI-enhanced classroom have any significant effects on L2 learners' writing outcome? (2) Does AI-enhanced classroom have any significant effects on L2 learners' writing feedback literacy and engagement? (3) What perceptions did L2 learners have towards employing AI-enhanced classroom in writing instruction?"	Rad et al. (2023)
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	"to explore student experiences, perceptions, and the utility of ChatGPT across different stages of the writing process, considering its role in idea generation, topic verification, proofreading, and editing while also investigating student opinions on the ethics of using ChatGPT in academic writing"	"How can ChatGPT be effectively utilized as a writing tool? What are the potential benefits associated with its use? How does it impact the learning outcomes and overall writing proficiency of students? How do students use it if given a free range? What do students think of using generative AI for writing assignments? Can we learn about student opinions on the ethical use of ChatGPT? Does it produce verifiable information?"	Rojas (2024)
The Transformative Power of AI Writing Technologies: Enhancing EFL Writing Instruction through the Integrative Use of Writerly and Google Docs	"(1) to investigate the impacts of integrating Writerly and Google Docs to enhance EFL writing instruction (2) to assess students' perceptions towards using these AI technologies"	"(1) What are the impacts of using integrative AI writing technologies on enhancing writing instruction? (2) What are students' perceptions towards using AI writing technologies?"	Wale and Kassahun (2024)

Overall, these studies collectively aim to understand and optimize the integration of AI tools in the writing process, enhancing creativity, feedback, and instructional effectiveness while addressing ethical concerns and user perceptions.

Results

The results from these studies are shown in Table 104.

Table 104

AI Enhanced Classroom: Results

Title	Results	Author
Writing, creativity, and artificial intelligence. ChatGPT	"The findings reveal statistically significant differences between the AIs and the students in the indicators of fluency, flexibility, and narrative originality in Game 2. Furthermore,	Vicente-Yagüe

Title	Results	Author
in the university context	significant differences are found between students' pre-test and post-test scores in fluency, flexibility, and narrative originality in Game 2 and in fluency in Game 3. Finally, the assistance provided by AI in writing tasks and verbal creativity is highlighted, and this should be considered in language teaching; in any case, AI cannot replace human intelligence and creativity. Firstly, it is noticeable that the means obtained by the AIs in the three indicators of game 2 are higher than the means indicated in the instrument manual, while, on the contrary, the means achieved in the indicators of game 3 are lower. In addition, after comparing the creativity of the AI systems with that of the students, the AI scores on the indicators in game 2 significantly outperformed the students' scores, while in game 3 the creativity indicators showed similar levels in both groups. This result is consistent with the above, as AI scores were higher in game 2. With respect to game 3, the students outperform, although not significantly, the AI in narrative originality. Furthermore, the significant improvement in the student's written productions after the use of ChatGPT has been highlighted, specifically in all the indicators of game 2 and the fluency of game 3. The potential of human-AI collaboration in the writing process is highlighted, as "AI is a means, not an end" (Breton, 2021) and it has been a tool that has worked as an assistant to the students in game 2, which is aligned with the productive possibilities of AI technology, whereas the flexibility and originality of game 3 has not, as it has already been shown that the AIs have not been able to imagine the effects of a fantasy situation."	Jara et al. (2023)
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	"This study of 343 communication instructors revealed a collective view that AI-assisted writing will be widely adopted in the workplace and will require significant changes to instruction. Key perceived challenges include less critical thinking and authenticity in writing. Key perceived benefits include more efficiency and better idea generation in writing. Students will need to develop AI literacy--composed of application, authenticity, accountability, and agency--to succeed in the workplace."	Cardon et al. (2023)
Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive?	"Findings show that (1) AI-based learning tools help students to do academic research, especially in the planning step, to identify and develop the topics, as well as in the drafting step, to develop a paper draft, (2) AI-based learning tools are deemed flexible in accessibility despite not being able to cover all necessities required by students in writing process, (3) students are interested in using AI technology in academic writing class so that learning process will not be boring. Although AI has been used in academic writing classes, tools have not positively impacted quality of students' academic papers in all indicators. There are several obstacles to using AI, namely (1) need for more available feature, especially in editing Indonesian text, and (2) in contrast, the features still need to be optimized. "	Utami et al. (2023)
Using artificial intelligence to foster students' writing feedback literacy,	"By using Wordtune, an AI-based application, the experimental group was able to significantly improve their writing outcomes, engagement, and feedback literacy when compared to the control group. Furthermore, the Wordtune	Rad et al. (2023)

Title	Results	Author
engagement, and outcome: A case of wordtune application	application received positive feedback from students interviewed about its ability to improve their writing outcomes, engagement, and feedback literacy. These findings will contribute to our understanding of AI in L2 learning and inform instructional considerations for Wordtune's application.”	
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	“This study's results reveal positive student experiences and perceptions of integrating ChatGPT for a writing-intensive assignment, showcasing its effectiveness in idea generation, topic verification, proofreading, and editing. / The integration of ChatGPT into the inorganic chemistry classroom has yielded valuable insights into its multifaceted utility for writing assignments. A preassignment survey revealed a diverse range of prior experiences among students, with a majority expressing prior curiosity-driven exploration of ChatGPT. For the effectiveness of ChatGPT in the first assignment, focused on topic generation and verification, student responses emphasized the ease and efficiency brought to the writing process, highlighting ChatGPT’s role as a facilitator for idea development. Subsequent surveys further illustrated students’ positive perceptions, with the second survey indicating widespread agreement on the usefulness of ChatGPT for scientific writing. Ethical considerations emerged, as students unanimously expressed reservations about using ChatGPT to generate entire text passages, emphasizing the importance of maintaining active student input. The final survey underscored ChatGPT’s pivotal roles in generating ideas and facilitating proofreading and editing, showcasing its significance as a supplementary tool. Looking ahead, students exhibited varied inclinations toward future ChatGPT use, reflecting a nuanced approach shaped by their experiences in this chemistry class.”	Rojas (2024)
The Transformative Power of AI Writing Technologies: Enhancing EFL Writing Instruction through the Integrative Use of Writerly and Google Docs	“When the quantitative data were analyzed through independent samples T-test and descriptive statistics, the qualitative data were analyzed thematically. The results confirmed that the integration of Writerly and Google Docs AI technologies, significantly improved EFL writing instruction, as evidenced by a statistically significant difference in writing performance between the experimental and control groups. Hence, students who learned through the integration of Writerly and Google Docs showed improved writing performance as they were able to produce essays that effectively addressed task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy, whereas those who learned through the conventional method were less effective in producing quality essays. The findings also revealed that the experimental group students had positive perceptions towards integrating Writerly and Google Docs because they found these AI writing technologies interesting, effective, goal-oriented, and supportive. Consequently, this study recommends researchers, curriculum designers, material designers, teachers, and students pay due attention to Writerly and Google Docs.”	Wale and Kassahun (2024)

The results of these studies reveal significant insights into the integration of AI tools in the writing classes across various educational contexts. The first study (Vicente-Yagüe Jara et al., 2023) shows that AI outperforms students in certain creativity indicators, highlighting AI's role in enhancing writing tasks while acknowledging that AI cannot replace human creativity entirely. Results from the second study (Cardon et al., 2023) indicates that business communication instructors see both benefits, such as efficiency and idea generation, and challenges, such as reduced critical thinking, in AI-assisted writing, emphasizing the need for AI literacy. The third study (Utami et al., 2023) finds that AI tools assist in academic research and drafting but face obstacles in fully meeting students' needs, especially in editing Indonesian text.

The fourth study (Rad et al., 2023) shows significant improvements in writing outcomes and feedback literacy among L2 learners using Wordtune, with positive student feedback on the tool's effectiveness. The fifth (Rojas, 2024) reveals positive student experiences with ChatGPT for idea generation, topic verification, proofreading, and editing, though ethical concerns about generating entire text passages were noted. Finally, the last study (Wale & Kassahun, 2024) demonstrates significant improvements in writing performance and positive student perceptions, suggesting the effectiveness of integrating AI tools like Writerly and Google Docs in EFL writing instruction.

Implications

Table 105 shared the implications of these studies.

Table 105*AI Enhanced Classroom: Implications*

Title	Results	Author
Writing, creativity, and artificial intelligence. ChatGPT in the university context	“in language teaching; in any case, AI cannot replace human intelligence and creativity. the use of AI should be aimed at empowering teachers and not replacing or displacing them (UNESS0, 2022b). An AI Literacy plan is required to train teachers in both technical skills and ethical-philosophical debates (Flores-Vívar G García-Iñalvo, 2023; UNESS0, 2022a). Similarly, changes in educational practice in the coming years will be determined by AI developments and will need to rely on research in Tull collaboration with teachers, educational leaders, and learners to ensure that appropriate educational policies are put in place.”	Vicente-Yagüe Jara et al. (2023)
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“Develop AI literacy. Oversee creative and innovative learning and teaching. Operate in communities of practice. Solicit cross-disciplinary input. Set community-based vision and policy. Provide resources for innovation. Pause institutional constraints and build culture.”	Cardon et al. (2023)
Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive?	“These are the recommendations for the optimization of AI-based learning tools, which are (1) adding features to edit Indonesian text, including spelling, diction, and sentence structure, and (2) enhancing AI literacy to be able to explore and leverage the existing features optimally.”	Utami et al. (2023)
Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application	“This study elucidates ideas that might improve second-language learners' writing feedback literacy, writing engagement, and writing outcomes and contextualizes them within the perspective of AI. The suggested concepts might benefit the field on both a broad and a focused level. This study might, on the surface, be used as a guide to enhance feedback efficacy and its perceived value for the efficient use of AI to enhance writing feedback literacy, thereby reducing the tension between instructors' feedback practices and students' expectations through personalized feedback. Instructors should adopt dialogic teaching, where student-teacher interaction is less emphasized to establish tolerance and acceptance with students as active members of the process instead of boosting themselves to a higher position with feedback as a unilateral process, to best promote and sustain student writing feedback literacy. In sum, Wordtune is well regarded for supporting students in their understanding of such information during the writing process and benefits from new data displayed in AI.”	Rad et al. (2023)
An Investigation into ChatGPT's Application for a Scientific Writing Assignment	“Based on the findings of this study, several implications for practice emerge for educators seeking to integrate generative AI tools such as ChatGPT into college-level writing assignments. Primarily, students favored topic generation as a useful mechanism for beginning their writing assignments.	Rojas (2024)

Title	Results	Author
The Transformative Power of AI Writing Technologies: Enhancing EFL Writing Instruction through the Integrative Use of Writerly and Google Docs	For amateurs or those who lack extensive subject knowledge, ChatGPT can quickly summarize a vast array of content and condense the material quickly for students. Importantly, the content generated by ChatGPT proved to be reliable, at least in these 53 student examples. Overall, student writing was improved through this process largely because students were provided a variety of mechanisms to revise their work. This process is achievable with writing centers, peer review, instructor feedback, etc., and ChatGPT appears to reliably contribute to that array of resources for editing writing. As a supplementary tool, ChatGPT can provide additional support and resources to facilitate writing assignments. However, it will be crucial for educators to address ethical considerations associated with the use of generative AI tools such as ChatGPT. By emphasizing the importance of active student involvement in the writing process and discouraging the use of ChatGPT to generate entire text passages, educators can promote academic integrity to uphold standards of originality. In STEM disciplines such as chemistry, where writing assignments play an important role in developing critical thinking and communication skills, integrating ChatGPT can offer students novel opportunities to engage with course material and express their understanding in innovative ways. Importantly, the use of generative AI in the writing process poses the risk of bypassing critical thinking and inhibiting some learning opportunities. By relying solely on AI-generated content, students may miss the chance to develop their own creative writing skills. Additionally, without facing the challenge of making mistakes and learning from them, students may struggle to improve their writing abilities independently. Educators can tailor ChatGPT-based activities to align with specific learning objectives and foster deeper learning experiences. Incorporating ChatGPT into writing assignments exposes students to emerging technologies and enhances their technological literacy. Educators can use ChatGPT as a catalyst for discussions on the ethical implications, limitations, and potential applications of AI in academic and professional contexts, empowering students to critically evaluate and responsibly use AI tools in their future endeavors.” “Thus, using integrated AI writing technologies in writing instruction is recommended to enhance EFL writing instruction. Hence, it implies that it is significant to instigate using integrated AI writing technologies to enhance writing instruction and enable students to enhance their writing performance. Writing teachers need to use Writerly and Google Docs in their writing instruction to enhance their students’ writing performance. Correspondingly, writing course material developers should reconsider these integrated AI writing technologies while developing writing course instructional materials. Moreover, students have to use these integrated AI writing technologies to enhance their writing performance since writing is required in their academic journey and future careers.”	Wale and Kassahun (2024)

The first study (Vicente-Yagüe Jara et al., 2023) concludes that AI should empower rather than replace teachers, suggesting the necessity of an AI literacy plan to train educators in both technical skills and ethical-philosophical debates. The second study (Cardon et al., 2023) suggests that developing AI literacy, fostering innovative teaching, and creating community-based policies are essential to effectively harness AI's potential in educational settings. The third study (Utami et al., 2023) demonstrates that optimizing AI tools by adding features for editing Indonesian text and enhancing AI literacy is crucial to maximizing their benefits in academic writing classes. The fourth study (Rad et al., 2023) implies that AI has the potential to improve writing feedback literacy, engagement, and outcomes, recommending a dialogic teaching approach to better align feedback practices with student expectations.

The fifth study (Rojas, 2024) finds that ChatGPT is valuable for facilitating various writing stages, emphasizing the importance of maintaining academic integrity and encouraging active student involvement in the writing process. The sixth study (Wale & Kassahun, 2024) concludes that integrating AI technologies like Writerly and Google Docs significantly enhances EFL writing instruction, advocating for their inclusion in writing course materials to improve student writing performance. Collectively, these studies advocate for the integration of AI tools to enhance the writing process while ensuring that ethical considerations and educational goals are met.

Writing Quality

Purpose

There are two studies that investigate writing quality.

Table 106*Writing Quality: Purpose*

Title	Purpose	Research Question	Author
The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective	“to examine the range of available Artificial Intelligence (AI) writing tools and assess their influence on student writing, particularly in terms of content and organization, as perceived by English as a Foreign Language (EFL) teachers.”	“(1) What types of AI writing tools are used by EFL teachers to improve students' writing quality, particularly in terms of content and organization? (2) How do EFL teachers perceive the impact of AI writing tools on the content and organization aspects of students' writing?”	Marzuki et al. (2023)
AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts	“to develop a Smart RoamLingo app to help writing meaningful content in authentic contexts with sample sentences based on texts generated by recognition technologies named as AI-Sample Sentences (AI-SS) and writing qualities with revision suggestions based on personal original writings named as AI-Writing Feedback (AI-WF)”	“1. What are the differences in the learning achievement among the three groups, the experimental group (EG) with two recognition technologies (i.e., ITR and STR) and AI support (i.e., AI-SS and AI-WF), the control group one (CG1) with two recognition technologies (i.e., ITR and STR), and the control group two (CG2) with one recognition technology only (i.e., STR)? 2. What are the correlations of the learning behaviors in the practice and assignment activities with the learning achievement of EG and the influence of learning behaviors in both practice and assignment activities on learning achievement? 3. What are the affordances of AI and its limitations in EFL writing? 4. What are the students' perceptions towards the proposed Smart RoamLingo?”	Hwang et al. (2023)

The purpose of the first study (Marzuki et al., 2023) is to examine the range of AI writing tools and their perceived impact on the content and organization of students' writing from the perspective of EFL teachers, while the second study (Hwang et al., 2023) aims to develop and evaluate a Smart RoamLingo app to enhance EFL writing through AI-generated sample sentences and personalized feedback.

Results

The results from these two studies are shown in Table 107.

Table 107*Writing Quality: Results*

Title	Results	Author
The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective	“The findings of this study imply that integrating AI writing tools can prove beneficial in elevating the quality of EFL student writing. This study aimed to examine various types of AI writing tools and their influence on the quality of student writing, focusing on content and organization, from the perspective of EFL teachers. The findings revealed a diverse array of AI writing tools employed in EFL teaching and shed light on the strategies teachers use to integrate these tools to address students' writing challenges. In particular, the tools, including Quillbot, Jenni, Chat-GPT, WordTune, Copy.ai, Paperpal, and Essay writer, were found to foster a comprehensive learning environment and enrich students' overall academic performance. The study also found unanimous agreement among the teachers about the positive role of AI writing tools in enhancing the clarity and logical progression of students' writing. However, concerns were expressed about the possibility of students becoming overly reliant on these tools, which could inhibit their critical thinking and problem-solving skills. The study's implications suggest the integration of AI writing tools can be beneficial in enhancing the quality of EFL student writing.”	Marzuki et al. (2023)
AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts	“EG significantly outperformed the other groups in the post-test. In EG, AI-SS can help the contextualization to produce comprehensive content actively while AI-WF can help the personalization to improve the writing quality through several revisions. Hence, the cohesion and consistency of their writing can be enhanced. Further, the total number of revisions in practices and the assignment score can significantly predict the post-test. Moreover, EG students revealed that Smart RoamLingo was useful for EFL writing. Therefore, it is strongly suggested to use Smart RoamLingo with AI-SS and AI-WF for EFL writing in authentic contexts.”	Hwang et al. (2023)

The first study (Marzuki et al., 2023) found that integrating AI writing tools, such as Quillbot, Jenni, Chat-GPT, WordTune, Copy.ai, Paperpal, and Essay writer, positively influences the quality of EFL student writing by enhancing clarity and logical progression. EFL teachers unanimously agreed on the benefits of these tools in fostering a comprehensive learning environment and improving overall academic performance, though they expressed concerns about potential over-reliance on AI inhibiting critical thinking and problem-solving skills. Meanwhile, the second study (Hwang et al., 2023) demonstrated that the experimental group (EG) using the Smart

RoamLingo app, which incorporates AI-generated sample sentences (AI-SS) and personalized writing feedback (AI-WF), significantly outperformed other groups in writing cohesion, consistency, and quality. EG students found the app useful for EFL writing, indicating that Smart RoamLingo effectively supports contextualization and personalization in writing, suggesting its strong potential for enhancing EFL writing in authentic contexts.

Implications

Table 108 shows the implications.

Table 108

Writing Quality: Implications

Title	Results	Author
The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective	"The study's implications suggest the integration of AI writing tools can be beneficial in enhancing the quality of EFL student writing. "	Marzuki et al. (2023)
AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts	"Based on the findings of this study, we summarize that recognition technologies such as STR and ITR can help generate vocabulary and phrases related to authentic contexts surrounding learners and openAI can help make meaningful sentences to guide students how to describe their taken pictures. Furthermore, openAI can provide students with more meaningful writing feedback based on their original writing to improve the writing quality including writing fluency with suggested conjunction words and giving more related description sentences. The integration of these recognition technologies and AI can work very well to support EFL writing."	Hwang et al. (2023)

The study on the impact of AI writing tools (Marzuki et al., 2023) suggests that integrating these tools can significantly enhance the quality of EFL student writing. Similarly, the study on AI and recognition technologies (Hwang et al., 2023) indicates that using tools like STR, ITR, and OpenAI can support EFL writing by

generating relevant vocabulary, providing meaningful feedback, and improving writing fluency and quality in authentic contexts.

Scaffolding

Purpose

Two studies out of 103 focused on scaffolding.

Table 109

Scaffolding: Purpose

Title	Purpose	Research Question	Author
Teacher’s Perceptions of Using an Artificial Intelligence-Based Educational Tool for Scientific Writing	“to investigate how teachers perceived an AI-enhanced scaffolding system developed to support students’ scientific writing for STEM education”	“1) How do teachers perceive AI-based scaffolding for scientific writing? 2) What can be potential issues of AI utilization in the classroom from teachers’ perspectives?”	Kim and Kim (2022)
Students’ appraisals post-ChatGPT use: Students’ narrative after using ChatGPT for writing	“to explore the students’ experience with ChatGPT in providing scaffolding for writing essays. It also discloses students’ appraisals of utilising ChatGPT”	Not stated.	Nugroho et al. (2024)

The first study (Kim & Kim, 2022) aims to investigate teachers' perceptions of an AI-enhanced scaffolding system for supporting students' scientific writing in STEM education, while the second study (Nugroho et al., 2024) seeks to explore students' experiences and appraisals of using ChatGPT for essay writing scaffolding.

Results

The results of these studies are in Table 110.

Table 110

Scaffolding: Results

Title	Results	Author
Teacher’s Perceptions of Using an Artificial Intelligence-Based Educational Tool for Scientific Writing	“Results revealed that most STEM teachers positively experienced AI as a source for superior scaffolding. On the other hand, they also raised the possibility of several issues caused by using AI such as the change in the role played by the teachers in the classroom and the transparency of the decisions made by the AI system. These results can be used as	Kim and Kim (2022)

Title	Results	Author
	a foundation for which to create guidelines for the future integration of AI with STEM education in schools, since it reports teachers’ experiences utilizing the system and various considerations regarding its implementation.”	
Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing	“The findings revealed that students derived advantages from using ChatGPT for translation, writing accuracy, writing efficiency, idea generation, and practical use. This study also reveals students' criticism in terms of inaccurate information generated as well as the possibility of students committing academic dishonesty when using ChatGPT. This study indicates that although students use ChatGPT to help them write their essays, they have the awareness to fact-check the information to avoid committing academic dishonesty.”	Nugroho et al. (2024)

The first study (Kim & Kim, 2022) finds that most STEM teachers had positive experiences with AI-enhanced scaffolding for scientific writing, though they raised concerns about changing teacher roles and AI transparency, suggesting a need for guidelines for AI integration in STEM education. The second study (Nugroho et al., 2024) reveals that students benefited from using ChatGPT for translation, writing accuracy, efficiency, and idea generation, but they also noted issues with inaccurate information and potential academic dishonesty, highlighting the importance of fact-checking.

Implications

Table 111 provides the implications.

Table 111

Scaffolding: Implications

Title	Implications	Author
Teacher’s Perceptions of Using an Artificial Intelligence-Based Educational Tool for Scientific Writing	“In this regard, this study brings great significance to the field in revealing STEM teachers’ overall positive perception regarding this innovative AI-based scaffolding and opportunities for future improvements. In addition, results of teachers’ experiences using the systems and their considerations of its implementation from this study can be used as a foundation for developing guidelines for the future integration of AI into school curricula, particularly in STEM education.”	Kim and Kim (2022)

Title	Implications	Author
Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing	“Students should be taught how to use the ChatGPT to support their writing (Dwivedi et al., 2023). Rather than banning ChatGPT in language classrooms, instructors should find ways to integrate AI to ensure its mindful use in education (Barrot, 2023). Rather than treating ChatGPT as a threat, instructors should consider it as a tool for language learners’ intellectual intensification (Dwivedi et al., 2023). To achieve this, instructors should consider assessments and policies regarding the use of ChatGPT in writing courses to ensure academic integrity (Imran & Almusharraf, 2023). Educators should prepare for a time when ChatGPT or similar advanced technological tools become more capable, accurate, and less reliant on prompt quality. This advancement will not only revolutionise the way students learn and acquire knowledge but will also change the current assessment method. Hence, educators should begin to consider innovative assessments and policies for the AI era.”	Nugroho et al. (2024)

The implications of the first study (Kim & Kim, 2022) suggest that the positive perceptions of STEM teachers toward AI-based scaffolding systems can inform the development of guidelines for integrating AI into school curricula, particularly in STEM education, highlighting opportunities for future improvements. The second study (Nugroho et al., 2024) implies that instead of banning tools like ChatGPT, educators should integrate them into writing courses to enhance learning, ensure mindful use, and maintain academic integrity. This includes preparing for advancements in AI technology, which will necessitate innovative assessments and policies to support the evolving educational landscape.

AI-Assisted Writing

Purpose

There are two studies that investigate AI-assisted writing.

Table 112

AI Assisted Writing: Purpose

Title	Purpose	Research Question	Author
ChatGPT and the digitisation of writing	“to uncover how students’ practices of writing in higher education are being	“(1) How were postgraduate students using ChatGPT and other	Zhao et al. (2024)

Title	Purpose	Research Question	Author
	impacted by ChatGPT. / to propose a model of generative AI literacy to assess their capabilities in doing so.”	digital writing tools for writing tasks in the summer of 2023? (2) What do students consider the benefits and problems of ChatGPT’s use? (3) What are the strengths and weaknesses in student generative AI literacy?”	
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“(1) to capture the collective wisdom of business communication instructors / (2) to develop an expert-driven framework of AI literacy for communicators”	“1. What concerns do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 2. To what degree do business communication instructors experience anxiety towards teaching or integrating AI-assisted writing into their courses? 3. What perceived benefits do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 4. To what degree do business communication professors perceive AI-assisted writing (e.g., ChatGPT) will be accepted in the near future for workplace writing?”	Cardon et al. (2023)

The first study (Zhao et al., 2024) aims to understand how students are using ChatGPT and other digital tools in their writing processes, focusing on their practices, perceptions, and the implications for generative AI literacy. The second study (Cardon et al., 2023) seeks to identify business communication instructors' perspectives on AI-assisted writing, including their concerns, anxieties, benefits, and predictions for future workplace adoption.

Results

The results from these studies are shown in Table 113.

Table 113

AI Assisted Writing: Results

Title	Results	Author
ChatGPT and the digitisation of writing	“The study confirms previous research, that students are keen to use generative AI and see it as part of a general trend in technology development (Chan and Hu, 2023). Our	Zhao et al. (2024)

Title	Results	Author
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	respondents tended to use discourses around time-saving, efficiency, and stress avoidance as justification for using the tools. ChatGPT was used rather intensively and iteratively but in highly individualistic ways. This differential use may reflect that there was little input at this time from the institution to help students how to use and where to draw the line in terms of appropriate use. Students wanted guidance from the institution on ChatGPT, echoing Attewell's (2023) findings. ChatGPT has brought advanced functions to digital writing but also intensified a sense of controversy in this area. For students, the worries focussed particularly on the unreliability of information it produced, fear of being accused of plagiarism if they used it and a concern about growing dependence on technology. While often its use was claimed to be justified for its time and stress reduction this may have underestimated the overall impact on learning. Saving time on learning tasks may unintentionally remove significant opportunities to learn. Many of the deeper ethical and societal issues such as around the exploitative way ChatGPT was developed were not fully understood. Yet it was clear that how writing was done digitally had become controversial. This could be seen as a benefit of ChatGPT, in that a gradual infiltration of digital tools into writing was made more visible institutionally and the controversial dimensions of technology use in education brought to the fore. Using the model developed above we can point to strengths and weaknesses in students' emergent generative AI literacy (Table 3). Overall our student interviewees showed significant generative AI literacy in most areas, particularly when considering it was early days with their use of it and almost none of them had received support from their teachers or from their institution in understanding how to use it. The weakest area of development was probably appreciation of the societal impacts of generative AI. As instructors catch up with students in understanding AI hopefully they can help students build up a more systematic understanding of pragmatic use of AI, a more reflective approach and a much more critical awareness of the social implications of AI." "This study of 343 communication instructors revealed a collective view that AI-assisted writing will be widely adopted in the workplace and will require significant changes to instruction. Key perceived challenges include less critical thinking and authenticity in writing. Key perceived benefits include more efficiency and better idea generation in writing. Students will need to develop AI literacy--composed of application, authenticity, accountability, and agency--to succeed in the workplace."	Cardon et al. (2023)

The first study (Zhao et al., 2024) indicates that students are rapidly adopting ChatGPT but often lack guidance in its effective and ethical use. While students perceive time-saving benefits, concerns about reliability, plagiarism, and

overreliance on technology are also prevalent. The second study (Cardon et al., 2023) emphasizes the need for AI literacy among educators and students to navigate the challenges and opportunities presented by AI-assisted writing. It highlights the potential for increased efficiency and idea generation while emphasizing the importance of critical thinking and authenticity in writing.

Implications

Table 114 provides the implications.

Table 114

AI Assisted Writing: Implications

Title	Implications	Author
ChatGPT and the digitisation of writing	“The social controversy around ChatGPT can be seen as a useful opportunity to engage students in a discussion about the digitisation of writing and promote AI literacy in this context. It has also developed a framework of generative AI literacy as a way of assessing their use. This can be used to plan both institutional policy and instructor support by identifying gaps in AI literacy that need to be filled. For example, educational developers could use the framework to facilitate discussions with teaching staff, aiding in the development of their AI literacy and enhancing their ability to teach AI literacy to students. In addition, universities could apply the dimensions of the AI framework to formulate policies and provide concrete examples that guide learning and teaching practices. The framework also has the potential in evaluating student AI literacy”	Zhao et al. (2024)
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“Develop AI literacy. Oversee creative and innovative learning and teaching. Operate in communities of practice. Solicit cross-disciplinary input. Set community-based vision and policy. Provide resources for innovation. Pause institutional constraints and build culture.”	Cardon et al. (2023)

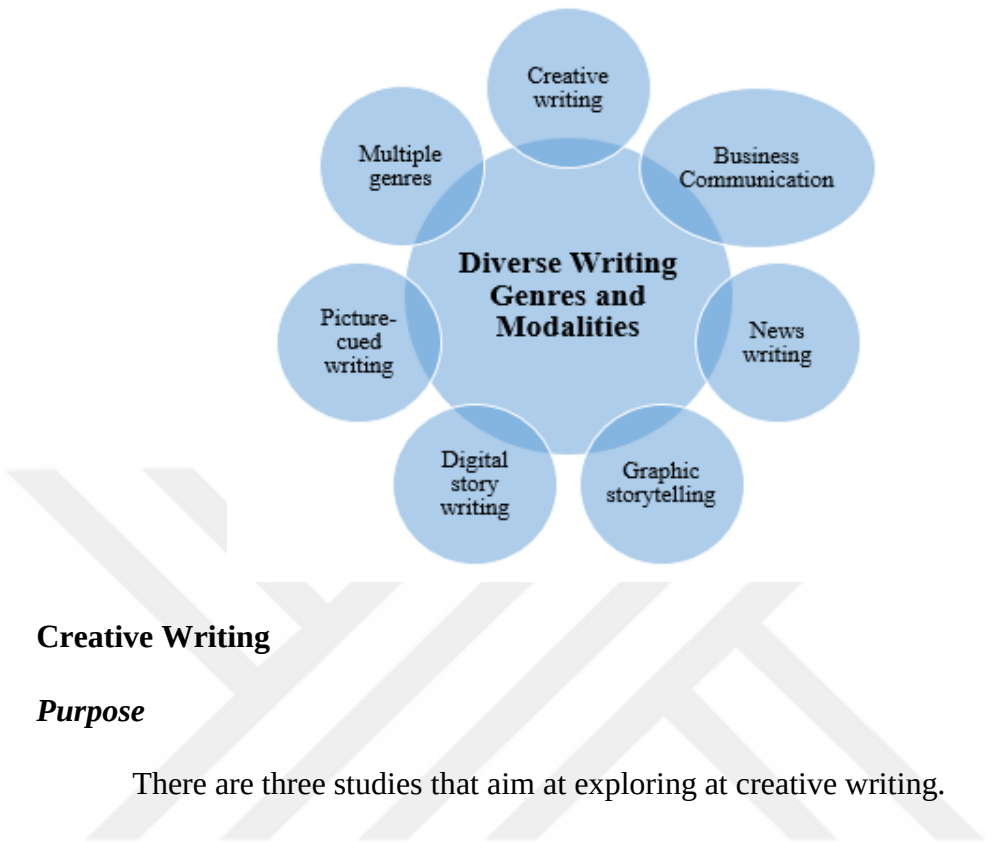
Dong (2024) aims to investigate the relationship between second language (L2) writing learners' perceptions of ChatGPT, their usage behavior, and L2 writing proficiency, focusing on complexity, accuracy, and fluency. Pigg (2024) seek to develop a framework of descriptive codes identifying practices early adopters use when integrating AI technologies into research writing processes. They analyze 35 publicly available videos showing the use of ChatGPT and Bing to understand

human-AI interactions in research writing and their implications for written invention. Malik et al. (2023) aim to understand higher education students' perspectives on using or avoiding AI in composing academic essays, exploring the reasons behind their choices and identifying the most common AI applications used in academic writing tasks. Nguyen et al. (2024) examine the nature of human-AI interactions in academic writing by focusing on the strategies doctoral students employ when collaborating with AI-powered writing tools. They investigate key hidden states, predominant patterns, and the effectiveness of strategic patterns in AI-assisted academic writing. Rababah et al. (2024) analyze the perspectives of 80 postgraduate students at Jadara University regarding the use of ChatGPT in writing their theses while considering various academic fields and exploring potential gender differences in perceptions.

Diverse Writing Genres and Modalities

The analysis and synthesis of the existing body of research revealed that 11 of the 103 studies focused on the use of AI tools in *Diverse Genres and Modalities*. These genres were *creative writing, business communication, news writing, graphic storytelling, digital story writing, picture-cued writing* and one study with *multiple genres*. Figure 10 shows these genres.

Figure 10
Diverse Writing Genres and Modalities



Creative Writing

Purpose

There are three studies that aim at exploring at creative writing.

Table 115

Creative Writing: Purpose

Title	Purpose	Research Question	Author
Writing, creativity, and artificial intelligence. ChatGPT in the university context	• To find out the level of creativity indicators (fluency, flexibility, and narrative originality) of AI systems and students, based on the sample of the Creative Imagination Test for Adults (PIC-A) by Artola et al. (2012). • To compare the creativity of the AI systems with that of the students, based on the indicators of fluency, flexibility, and narrative originality. • To compare the scores of the creativity indicators obtained by the students in the two phases of application of the tests, the first without any type of aid and the second using an AI system (ChatGPT)."	Not stated.	Vicente-Yagüe Jara et al. (2023)

Title	Purpose	Research Question	Author
Is Chatgpt a menace for creative writing ability? An experiment	“to quantitatively investigate whether ChatGPT's use negatively affects university students' creative writing abilities, focusing on originality, content presentation, accuracy, and elaboration in essays”	“RQ. Does ChatGPT affect the creative writing ability of tertiary level students?”	Niloy et al. (2024)
Beyond the author: Artificial intelligence, creative writing and intellectual emancipation	“to explore university students' engagement with Generative Artificial Intelligence (GenAI) tools for creative writing and graphic storytelling, drawing on Jacques Rancière's philosophy of intellectual equality and emancipation”	Not stated.	Tsao and Nogues (2024)

The first study (Vicente-Yagüe Jara et al., 2023) aims to evaluate the creativity indicators of AI systems and students, comparing fluency, flexibility, and narrative originality between the two groups, and assessing the impact of AI assistance on students' creativity scores. The second study (Niloy et al., 2024) seeks to quantitatively determine whether ChatGPT negatively affects university students' creative writing abilities, focusing on originality, content presentation, accuracy, and elaboration in essays. The third study (Tsao & Nogues, 2024) explores university students' engagement with Generative Artificial Intelligence (GenAI) tools for creative writing and graphic storytelling, informed by Jacques Rancière's philosophy of intellectual equality and emancipation.

Results

The results from these studies are shown in Table 116.

Table 116

Creative Writing: Results

Title	Results	Author
Writing, creativity, and artificial intelligence. ChatGPT in the university context	“The findings reveal statistically significant differences between the AIs and the students in the indicators of fluency, flexibility, and narrative originality in Game 2. Furthermore, significant differences are found between students' pre-test and post-test scores in fluency, flexibility, and narrative originality in Game 2 and in fluency in Game 3. Finally, the	Vicente-Yagüe Jara et al. (2023)

Title	Results	Author
	assistance provided by AI in writing tasks and verbal creativity is highlighted, and this should be considered in language teaching; in any case, AI cannot replace human intelligence and creativity. Firstly, it is noticeable that the means obtained by the AIs in the three indicators of game 2 are higher than the means indicated in the instrument manual, while, on the contrary, the means achieved in the indicators of game 3 are lower. In addition, after comparing the creativity of the AI systems with that of the students, the AI scores on the indicators in game 2 significantly outperformed the students' scores, while in game 3 the creativity indicators showed similar levels in both groups. This result is consistent with the above, as AI scores were higher in game 2. With respect to game 3, the students outperform, although not significantly, the AI in narrative originality. Furthermore, the significant improvement in the student's written productions after the use of ChatGPT has been highlighted, specifically in all the indicators of game 2 and the fluency of game 3. The potential of human-AI collaboration in the writing process is highlighted, as "AI is a means, not an end" (Breton, 2021) and it has been a tool that has worked as an assistant to the students in game 2, which is aligned with the productive possibilities of AI technology, whereas the flexibility and originality of game 3 has not, as it has already been shown that the AIs have not been able to imagine the effects of a fantasy situation."	
Is Chatgpt a menace for creative writing ability? An experiment	"The EGp incorporates ChatGPT in their creative writing process as an intervention. The study evaluates originality, content presentation, accuracy, and elaboration, utilizing the Wilcoxon Signed-Rank Test for analysis. The findings reveal a detrimental association between ChatGPT use and university students' creative writing abilities. Analysing both machine-based and human-based assessments substantiates earlier qualitative observations regarding ChatGPT's adverse impact on creative writing. This study highlights the necessity of approaching AI integration, particularly in creative writing disciplines, with caution. While AI tools have merits, their integration should be thoughtful, considering the potential drawbacks. / The findings of this experiment highlight a significant decline in students' creative writing ability as a result of using ChatGPT. Notably, the scores of the EGp displayed a substantial difference in the posttest compared to the pre-test. As this disparity was not evident in the pre-test, the result signified that the introduction of ChatGPT led to a notable impact on students' creative writing capabilities. This experiment also underscores the moderate role that ChatGPT plays in causing such disparities between the Control Group and the EGp."	Niloy et al. (2024)
Beyond the author: Artificial intelligence, creative writing and intellectual emancipation	"Qualitative data analysis from a co-curricular creative writing programme, including reflections, surveys, and focus-group interviews, reveals emerging artificial intelligence literacies and students' improvisational aptitudes for interpreting, subverting, and transforming notions of authorship. Students decentred authorial attribution through the pragmatic adoption of the technology as a creative catalyst, negotiated creative conventions by adopting non-conventional communication strategies, and reconceptualised creativity as distributed across	Tsao and Nogues (2024)

Title	Results	Author
	human and non-human agents. Our approach of student-driven learning for autonomous exploration, sense-making, and criticality with GenAI indicates the potential for promoting conditions for students to exercise intellectual equality and emancipation. The findings contribute to the understanding of authorship and creativity; begin to contour emerging GenAI literacies and competencies; and suggest that creative collaborations with GenAI may be a promising way to foster emancipatory practices in the classroom, while nurturing creative and critical skills.”	

The results of the first study (Vicente-Yagüe Jara et al., 2023) indicate significant differences in creativity indicators, with AI outperforming students in fluency, flexibility, and narrative originality in certain tasks, but students showed notable improvement in their scores after using AI assistance. The second study (Niloy et al., 2024) reveals that the use of ChatGPT negatively impacted university students' creative writing abilities, with a decline in originality, content presentation, accuracy, and elaboration, and suggesting the need for cautious integration of AI in creative writing. The findings from the third study (Tsao & Nogues, 2024) show that students developed emerging AI literacies and creative improvisational skills, rethinking authorship and creativity by engaging with AI as a creative facilitator and indicating the potential of GenAI to promote intellectual equality and emancipation in educational settings.

Implications

Table 117 provides the implications.

Table 117

Creative Writing: Implications

Title	Implications	Author
Writing, creativity, and artificial intelligence. ChatGPT in the university context	“in language teaching; in any case, AI cannot replace human intelligence and creativity. the use of AI should be aimed at empowering teachers and not replacing or displacing them (UNESS0, 2022b). An AI Literacy plan is required to train teachers in both technical skills and ethical-philosophical debates (Flores-Vívar G García-Iñalvo, 2023; UNESS0,	Vicente-Yagüe Jara et al. (2023)

Title	Implications	Author
	2022a). Similarly, changes in educational practice in the coming years will be determined by AI developments and will need to rely on research in Tull collaboration with teachers, educational leaders, and learners to ensure that appropriate educational policies are put in place.”	
Is Chatgpt a menace for creative writing ability? An experiment	“Educators may implement changes in assigning tasks to students compared to their earlier practices, based on the identified factors that are being affected negatively, to ensure ChatGPT does not hinder a student's creativity at a greater extentThe extent of using ChatGPT should be limited to self-learning as positive effect was experienced through the experiment Policymakers may use the findings of the study to impose strict policies in academia for ensuring academic integrity (Example: must use of plagiarism detecting software for checking scripts, assigning tasks to students which require more analytical abilities, providing tasks which are not properly readable by LLM's like ChatGPT such as image-based questions, case studies etc.”	Niloy et al. (2024)
Beyond the author: Artificial intelligence, creative writing and intellectual emancipation	“Educators in higher education must acknowledge our current coexistence with AI. They bear the responsibility for creating conditions that enhance students’ critical awareness, aiming to reduce structural inequalities and inspire imaginative thinking for a better future. Human participation is crucial in directing a technology that generates “new” content by drawing upon our digitised past.”	Tsao and Nogues (2024)

The implications of these studies underscore the nuanced role of AI in educational contexts, particularly in writing and creativity. The first study (Vicente-Yagüe Jara et al., 2023) emphasizes that while AI can enhance language teaching, it cannot replace human intelligence and creativity. Instead, AI should empower teachers, necessitating an AI literacy plan to train educators in technical skills and ethical-philosophical debates to ensure the development of appropriate educational policies. The second study (Niloy et al., 2024) suggests that educators should adjust task assignments to prevent ChatGPT from hindering student creativity, recommending limited use of ChatGPT for self-learning and the implementation of strict academic integrity policies, such as using plagiarism detection software and designing tasks that require higher analytical abilities. The third study (Tsao & Nogues, 2024) highlights the importance of integrating AI in higher education

responsibly, encouraging educators to foster students' critical awareness and imaginative thinking to address structural inequalities and promote a collaborative and innovative learning environment.

Business Communication

Purpose

Two studies focused on workplace writing.

Table 118

Business Communication: Purpose

Title	Purpose	Research Question	Author
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“(1) to capture the collective wisdom of business communication instructors / (2) to develop an expert-driven framework of AI literacy for communicators”	“1. What concerns do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 2. To what degree do business communication instructors experience anxiety towards teaching or integrating AI-assisted writing into their courses? 3. What perceived benefits do business communication and writing professors hold about AI-assisted writing (e.g., Chat GPT)? 4. To what degree do business communication professors perceive AI-assisted writing (e.g., ChatGPT) will be accepted in the near future for workplace writing?”	Cardon et al. (2023)
Perceptions of Professionalism and Authenticity in AI-Assisted Writing	“Our study had several purposes: (a) explore the degree to which AI-generated emails are perceived as having appropriate tone and style; (b) explore the degree to which AI-generated emails are perceived as effective; and (c)explore attitudes toward AI-assisted writing in the workplace.”	Not stated.	Coman and Cardon (2024)

The purposes of these studies are to capture the collective wisdom of business communication instructors and develop an expert-driven framework of AI literacy,

and to explore the perceptions of professionalism and authenticity in AI-assisted writing, specifically examining the tone, style, effectiveness, and attitudes towards AI-generated emails in the workplace.

Results

The results from these two studies are shown in Table 119.

Table 119

Business Communication: Results

Title	Results	Author
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“This study of 343 communication instructors revealed a collective view that AI-assisted writing will be widely adopted in the workplace and will require significant changes to instruction. Key perceived challenges include less critical thinking and authenticity in writing. Key perceived benefits include more efficiency and better idea generation in writing. Students will need to develop AI literacy--composed of application, authenticity, accountability, and agency--to succeed in the workplace.”	Cardon et al. (2023)
Perceptions of Professionalism and Authenticity in AI-Assisted Writing	“Across all messages, professionals consistently rated the AI-generated messages as professional, effective, efficient, confident, and direct. They rate sincerity and caring slightly lower in some disclosed conditions, particularly for ChatGPT-generated messages, suggesting the importance of tool selection when using generative AI for workplace writing. Those professionals who use AI more frequently for work are more likely to view AI-assisted writing as authentic, effective, and confidence-building. Implications for teaching business communication, including the need to address AI literacy, and suggestions for future research are provided.”	Coman and Cardon (2024)

The first study (Cardon et al., 2023) reveals that 343 communication instructors collectively view AI-assisted writing as a growing trend in the workplace, necessitating significant instructional changes. The perceived challenges include diminished critical thinking and authenticity, while the benefits are increased efficiency and idea generation. Developing AI literacy—encompassing application, authenticity, accountability, and agency—is essential for students' future workplace success. Meanwhile, the second study (Coman & Cardon, 2024) finds that professionals generally rate AI-generated messages as professional, effective,

efficient, confident, and direct, though sincerity and caring are rated slightly lower for ChatGPT-generated messages. Frequent AI users are more likely to view AI-assisted writing as authentic and effective, highlighting the importance of tool selection and the need to integrate AI literacy into business communication teaching.

Implications

Table 120 provides the implications.

Table 120

Business Communication: Implications

Title	Results	Author
The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	“Develop AI literacy. Oversee creative and innovative learning and teaching. Operate in communities of practice. Solicit cross-disciplinary input. Set community-based vision and policy. Provide resources for innovation. Pause institutional constraints and build culture.”	Cardon et al. (2023)
Perceptions of Professionalism and Authenticity in AI-Assisted Writing	“Our study also carries implications for teachers of business communication. As generative AI is more widely adopted for even routine workplace communication, teachers should consider how to best prepare students to thrive in an environment where generative AI use is the norm. One key implication is that teachers should reevaluate the role of grammar instruction in business communication curriculum. Good grammar is a primary marker of professionalism and other positive attributes, making grammar instruction a core concern of professional writing. However, producing grammatically correct prose is one of the greatest strengths of generative AI tools. These tools are far more powerful than spell checkers of the past, and they are soon to be native across writing platforms. Widespread adoption of these tools suggests the need to reassess the role of grammar instruction in the classroom. Prior studies have shown that generative AI is a powerful skill-leveler, so teachers have an opportunity to help students leverage these tools to improve their grammar and therefore their perceived professionalism. Secondly, if professionals have access to multiple tools that can dynamically produce and edit text for correctness, the differentiating factor for business communicators may be the ability to identify context and adjust generated text to match the purpose and needs of the audience. We found that AI-generated messages scored relatively lower for relationship-oriented aspects of writing like sincerity and caring. This research suggests that instructors should explore ways to emphasize these aspects in their communication classes. In a world where everyone produces grammatically correct writing, EQ and empathy may be the differentiating skills that business communication students need to stand out. A third implication for business communication instructors is the need to emphasize AI	Coman and Cardon (2024)

Title	Results	Author
	literacy in their courses. An important part of AI literacy is the ability to understand various AI platforms and the value of their output. This study suggests that there is difference between two of the platforms—ChatGPT and Google—in the context of routine but relationally significant emails. Instructors should find ways to allow students to compare and contrast the output of various platforms. This is particularly important because most university students have not experimented with more than one platform. A year into the gAI era, a November 2023 survey of university students across a dozen universities found that while 83% had used ChatGPT, only 11% had used Google Bard, 3% had tried Google’s Help Me Write, 2% had used Dall-E, 2% had used Claude, 1% had tried MidJourney, and 11% had tried another platform.”	

The implications of the studies emphasize the critical need for developing AI literacy to adapt to the increasing use of AI-assisted writing. First study suggests that it is crucial to oversee innovative learning, foster communities of practice, solicit cross-disciplinary input, and create community-based policies while providing resources for innovation and easing institutional constraints. The second study (Coman & Cardon, 2024) highlights that teachers of business communication must prepare students to thrive in an AI-integrated environment by re-evaluating the role of grammar instruction, emphasizing emotional intelligence and empathy, and comparing different AI platforms' outputs. As generative AI becomes a norm in workplace communication, the ability to produce contextually appropriate, sincere, and empathetic communication will differentiate successful business communicators. Educators should therefore integrate AI literacy into their curricula, enabling students to leverage AI tools effectively while maintaining professionalism and relational aspects in their writing.

Multimodal Writing

Purpose

There is one study that aims to explore the differences in composing processes between EFL students using generative AI-assisted digital multimodal composing and traditional writing, and to examine how students create images in the AI-assisted composing process.

Table 121

Multimodal Writing: Purpose

Title	Purpose	Research Question	Author
Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing	"to explore the generative AI-assisted composing processes of two groups of English as a foreign language (EFL) writers over two weeks in this qualitative study"	"1) Do students behave differently in creating texts in generative AI-assisted digital multimodal composing and traditional writing? And if they do, what are these differences? 2) How do students create images in generative AI-assisted digital multimodal composing?"	Liu et al. (2024)

Results

The results from this study are shown in Table 122.

Table 122

Multimodal Writing: Results

Title	Results	Author
Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing	"To answer research question one about whether and how students would behave differently in creating texts in generative AI-assisted digital multimodal composing and traditional writing, our research showed that students behaved differently in creating texts in generative AI-assisted digital multimodal composing and traditional writing. Students created more transitional texts and examples and tended to use summarized search results from Bing Chat. To answer research question two regarding how students would create images in generative AI-assisted digital multimodal composing, our results illustrated that students benefited greatly from prompts provided by ChatGPT and Bing Chat. Their evaluation benchmarks of AI images also reflected their critical thinking of utilizing generative AI technologies."	Liu et al. (2024)

The study (Liu et al., 2024) concludes that EFL students behaved differently when creating texts with generative AI-assisted digital multimodal composing compared to traditional writing, using more transitional texts and summarized search

results, and they benefited significantly from AI-generated prompts for creating images, demonstrating critical thinking in evaluating AI-generated content.

Implications

Table 123 provides the implications of this study.

Table 123

Multimodal Writing: Implications

Title	Implications	Author
Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing	“almost all participants reflected that the draft generated by ChatGPT could function as a model text to improve their English writing skills relating to organizational structures, language and content and AI images could make their digital multimodal composing process more enjoyable. They were also quite impressed by the personalized experience provided by generative AI technologies, which echoed the claim that AI can be applied in education to offer individualized instruction that recognizes learner individual differences and needs (Yilmaz et al., 2022). On top of that, the participants deemed that ChatGPT could improve the efficiency of writing, which is in accord with a previous finding that AI-powered tools helped students get more efficient code writing experience (Yilmaz & Yilmaz, 2023a). However, AI texts were still difficult to accurately capture several features of human writing quality, such as emotional depth, writer voice and identity, and rhetorical flexibility (Barrot, 2023). Additionally, AI images sometimes deviated from what the students meant and failed to meet their expectations. In practice, students need to scrutinize the text generated by ChatGPT, adjusting them carefully to adapt to a new rhetorical situation, adding their own opinions from their stance, thus taking full advantage of ChatGPT to sharpen their writing skills. It is incumbent on students to evaluate AI images more carefully and optimize them to produce images of higher quality for their digital multimodal texts. In a nutshell, the positive influence of generative AI on English writing has been justified by this study. However, the use of generative AI in English writing may still be constrained in the absence of effective teaching methods. Students commonly reflected that generative AI could make users too lazy to think on their own, which may have the risk of making students become totally dependent on it (Yilmaz & Yilmaz, 2023b). Relying on generative AI can also render the students to lose interest in exploring problems and analyzing problems, which deviates from the essence of inquiry-based learning. In our interviews, some students shared their ideas about how to teach students to use generative AI properly to write. For example, Cindy proposed that students needed to write drafts on their own before they should ask ChatGPT to generate a draft on the same topic. Thus, pedagogically, teachers may require students to write a review report to compare the two drafts and illustrate what they have learned from ChatGPT to enhance their critical	Liu et al. (2024)

Title	Implications	Author
	thinking skills. Alice proposed that students should be notified that they could use ChatGPT to help with their course essays. Nonetheless, if they chose to totally depend on it, they would fail this course. Students also commonly supported using AntiGPT to conduct efficient detection of AI-generated texts to ensure academic integrity of writing.”	

The study (Liu et al., 2024) suggests that generative AI, like ChatGPT, can enhance English writing skills by providing model texts and personalized experiences, improving writing efficiency, and making the composing process more enjoyable, though it struggles with capturing emotional depth, writer voice, and rhetorical flexibility. It is important for students to critically evaluate and adjust AI-generated texts and images, and educators should implement effective teaching methods to prevent over-reliance on AI, such as having students compare their own drafts with AI-generated ones to enhance critical thinking and ensure academic integrity.

News Writing

Purpose

The aim of this study is to examine how the nature of a story's author (human or AI news bot journalist) affects audiences' assessments of the news story and its writer.

Table 124

News Writing: Purpose

Title	Purpose	Research Question	Author
Can AI Become Walter Cronkite? Testing the Machine Heuristic, the Hostile Media Effect, and Political News Written by Artificial Intelligence	“to examine how the nature of a story's author (human or news bot journalist) affects how audiences assess a news story and its writer”	Not stated.	Hong et al. (2024)

Results

The results from this study are shown in Table 125.

Table 125

News Writing: Results

Title	Results	Author
Can AI Become Walter Cronkite? Testing the Machine Heuristic, the Hostile Media Effect, and Political News Written by Artificial Intelligence	“Results reveal that people express more positive attitudes-and trust-toward human journalists than toward AI journalists, but they perceive the latter as less biased. Also, story and outlet credibility have a negative relationship with ideological distance from outlets.”	Hong et al. (2024)

The results of the study (Hong et al. (2024) reveal that people have more positive attitudes and trust toward human journalists compared to AI journalists, but perceive AI journalists as less biased; additionally, story and outlet credibility decrease as the ideological distance from the outlets increases.

Implications

Table 126 provides the implications.

Table 126

News Writing: Implications

Title	Implications	Author
Can AI Become Walter Cronkite? Testing the Machine Heuristic, the Hostile Media Effect, and Political News Written by Artificial Intelligence	“News organizations should acknowledge the variety of factors that can influence perceptions of news credibility and bias before they proclaim that the goal of automated journalism is to create unbiased AI news writers.”	Hong et al. (2024)

The study (Hong et al. (2024) suggests that news organizations should recognize the multiple factors influencing perceptions of news credibility and bias before aiming to create unbiased AI news writers through automated journalism.

Narrative Writing

Purpose

The purpose of this study (Zhou et al., 2023) is to compare the narrative writing performance of ChatGPT and Chinese intermediate English learners to identify the chatbot's advantages and disadvantages in writing, and to determine how updating commands can improve ChatGPT's writing quality.

Table 127

Narrative Writing: Purpose

Title	Purpose	Research Question	Author
Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing	“To investigate these issues [ChatGPT is a publicly available chatbot that can quickly generate texts on given topics, but it remains unknown whether the chatbot is as good as or even superior to human writers in all aspects of writing, and whether the chatbot's writing quality can be significantly improved by updating commands], this study compared writing performance on a narrative topic by ChatGPT and Chinese intermediate English (CIE) learners so as to reveal the chatbot's advantages and disadvantages in writing”	“(1) How do ChatGPT and CIE learners perform in narrative English writing from the perspective of discourse components? (2) Which of the components in a revised version can be improved by ChatGPT when user responses are updated for revision? (3) How are the discourse components of the texts produced by ChatGPT and CIE learners correlated?”	Zhou et al. (2023)

Results

The results from this study are shown in Table 128.

Table 128

Narrative Writing: Results

Title	Results	Author
Chinese intermediate English learners outdid ChatGPT in deep cohesion:	“The data were analyzed by virtue of the five discourse components using Coh-Metrix (a special instrument for analyzing language discourse), and the results revealed that ChatGPT performed better than human writers in narrativity, word concreteness, and referential cohesion, but worse in	Zhou et al. (2023)

Title	Results	Author
Evidence from English narrative writing	syntactic simplicity and deep cohesion in its initial version. After its commands were updated, the resulting version was better in its syntactic simplicity, but it still lagged far behind the CIE learners' writing in deep cohesion. In addition, a correlation analysis of the discourse components suggests that narrativity was correlated with referential cohesion for both ChatGPT and the human writers, but the correlations varied within each group."	

The results from the study (Zhou et al., 2023) reveals that ChatGPT outperforms human writers in narrativity, word concreteness, and referential cohesion, but performs worse in syntactic simplicity and deep cohesion; even after updating commands, ChatGPT still lags behind Chinese intermediate English learners in deep cohesion.

Implications

Table 129 provides the implications.

Table 129

Narrative Writing: Implications

Title	Implications	Author
Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing	"..the findings suggest that ChatGPT with advanced ToM capabilities could potentially enhance their writing performance, especially in terms of referential cohesion. This study suggests that AI tools like ChatGPT are really powerful in text writing but also lags far behind human writers particularly in the use of advanced inference like deep cohesion and underscores the need for continued research and development to optimize the performance of AI-powered writing tools."	Zhou et al. (2023)

The findings imply that while AI tools like ChatGPT are powerful in text writing, they lag behind human writers in advanced inference tasks such as deep cohesion, highlighting the need for continued research and development to optimize AI-powered writing tools, particularly in enhancing their Theory of Mind (ToM) capabilities.

Graphic Storytelling

Purpose

There is one study (Tsao & Nogues, 2024) that aims to explore university students' engagement with Generative Artificial Intelligence (GenAI) tools for creative writing and graphic storytelling, using Jacques Rancière's philosophy of intellectual equality and emancipation as a framework.

Table 130

Graphic Storytelling: Purpose

Title	Purpose	Research Question	Author
Beyond the author: Artificial intelligence, creative writing and intellectual emancipation	“to explore university students' engagement with Generative Artificial Intelligence (GenAI) tools for creative writing and graphic storytelling, drawing on Jacques Rancière's philosophy of intellectual equality and emancipation”	Not stated.	Tsao and Nogues (2024)

Results

The results from this study are shown in Table 131.

Table 131

Graphic Storytelling: Results

Title	Results	Author
Beyond the author: Artificial intelligence, creative writing and intellectual emancipation	“Qualitative data analysis from a co-curricular creative writing programme, including reflections, surveys, and focus-group interviews, reveals emerging artificial intelligence literacies and students' improvisational aptitudes for interpreting, subverting, and transforming notions of authorship. Students decentred authorial attribution through the pragmatic adoption of the technology as a creative catalyst, negotiated creative conventions by adopting non-conventional communication strategies, and reconceptualised creativity as distributed across human and non-human agents. Our approach of student-driven learning for autonomous exploration, sense-making, and criticality with GenAI indicates the potential for promoting conditions for students to exercise intellectual equality and emancipation. The findings contribute to the understanding of authorship and creativity; begin to contour emerging GenAI literacies and competencies; and suggest that creative	Tsao and Nogues (2024)

Title	Results	Author
	collaborations with GenAI may be a promising way to foster emancipatory practices in the classroom, while nurturing creative and critical skills.”	

The results reveal that students developed emerging AI literacies and improvisational skills, using GenAI tools to interpret, subvert, and transform traditional notions of authorship. They pragmatically adopted technology as a creative catalyst, negotiated creative conventions with non-conventional communication strategies, and reconceptualized creativity as a collaboration between human and non-human agents, suggesting that GenAI can foster intellectual equality and emancipation in the classroom while nurturing creative and critical skills.

Implications

Table 132 provides the implications.

Table 132

Graphic Storytelling: Implications

Title	Implications	Author
Beyond the author: Artificial intelligence, creative writing and intellectual emancipation	“Educators in higher education must acknowledge our current coexistence with AI. They bear the responsibility for creating conditions that enhance students’ critical awareness, aiming to reduce structural inequalities and inspire imaginative thinking for a better future. Human participation is crucial in directing a technology that generates “new” content by drawing upon our digitised past.”	Tsao and Nogues (2024)

The study (Tsao & Nogues, 2024) implies that educators in higher education must create conditions that enhance students’ critical awareness and imaginative thinking to reduce structural inequalities, emphasizing the importance of human participation in directing AI technology to generate new content.

Digital Stories

Purpose

The purpose of this study (Pellas, 2023) is to investigate whether creating digital stories using generative AI platforms can influence the narrative intelligence and writing self-efficacy of undergraduate students majoring in Primary Education.

Table 133

Digital Stories: Purpose

Title	Purpose	Research Question	Author
The Effects of Generative AI Platforms on Undergraduates' Narrative Intelligence and Writing Self-Efficacy	“to investigate whether digital story creation tasks on web-based platforms can influence the narrative intelligence and writing self-efficacy of undergraduate students / to explore the consequences of digital story creation tasks on the levels of narrative intelligence and writing self-efficacy among undergraduates majoring in Primary Education using several web-based storytelling platforms.”	“RQ1. Does creating digital stories using generative AI platforms have a significant impact on the narrative intelligence of undergraduate students? RQ2. Does creating digital stories with generative AI platforms have a significant effect on the writing self-efficacy of undergraduate students?”	Pellas (2023)

Results

The results from this study are shown in Table 134.

Table 134

Digital Stories: Results

Title	Results	Author
The Effects of Generative AI Platforms on Undergraduates' Narrative Intelligence and Writing Self-Efficacy	“The findings indicate that the utilization of generative AI platforms in the context of story creation tasks can substantially enhance both narrative intelligence scores and writing self-efficacy when compared to conventional platforms. Nonetheless, there was no significant difference in the creative identity factor. The findings reveal a noteworthy difference in favor of the experimental group. In line with previous studies [8,21], this study's findings indicate that the levels of narrative intelligence among undergraduate students exposed to the experimental AI-supported instructional method were significantly higher than those of the control group. In relation to RQ2, the analysis of writing self-efficacy,	Pellas (2023)

Title	Results	Author
	measured through the SAWSES, was conducted using a repeated measurements two-way ANOVA. Consistent with previous studies’ findings [4,19], the evidence from this comparative analysis allowed us to explore whether the proposed treatment had a significant impact on the writing self-efficacy of undergraduate students. The analysis also indicated a significant difference in favor of the experimental group in the “Writing essentials” factor. This suggests that the proposed treatment was effective in improving students’ understanding of fundamental writing principles, further supporting the positive impact of the experimental approach. In the “Reflective writing” factor, a significant difference was observed in favor of the experimental group. In the “Reflective writing” category, ANOVA results similarly indicated a substantial impact of the intervention. Nonetheless, it is worth noting that there was no significant difference in the “Creative identity” factor. This implies that the proposed treatment did not have a significant impact on students’ perceptions of their creative identity in writing.”	

The findings from this study (Pellas, 2023) indicate that utilizing generative AI platforms for story creation tasks significantly enhances both narrative intelligence and writing self-efficacy among undergraduate students compared to conventional platforms, though there was no significant difference in the creative identity factor. Specifically, the experimental group showed higher scores in narrative intelligence, writing essentials, and reflective writing, highlighting the positive impact of the AI-supported instructional method on these aspects of writing.

Implications

Table 135 provides the implications.

Table 135

Digital Stories: Implications

Title	Implications	Author
The Effects of Generative AI Platforms on Undergraduates’ Narrative Intelligence and Writing Self-Efficacy	“The results of this research carry significance not only for the academic community but also for educational practitioners and technology developers. Understanding whether digital story creation activities influence narrative intelligence and writing self-efficacy can offer valuable insights into the design and implementation of generative AI platforms to enhance, rather than impede, individuals’ storytelling skills and	Pellas (2023)

Title	Implications	Author
	confidence in the digital age. Enhancement of narrative intelligence and writing self-efficacy: This study’s results provide evidence that the use of generative AI platforms can significantly improve undergraduate students’ narrative intelligence scores and writing self-efficacy. This finding aligns with the growing recognition of the potential benefits of technology assisted learning in enhancing core competencies. Educators and institutions can leverage these tools to empower students with the skills and confidence necessary for effective communication; 2. Effects of generative AI platforms on storytelling: The positive outcomes observed in the experimental group highlight the evolving role of AI technologies in education. Generative AI platforms can augment traditional teaching methods by offering personalized feedback, suggesting improvements, and facilitating the writing process. The study suggests that these technologies can serve as valuable educational aids, supporting students in their journey to become proficient writers; 3. Human creativity vs. generative AI assistance: Educators and curriculum designers should consider incorporating generative AI platforms into writing instruction. On the one hand, while generative AI demonstrated clear benefits in terms of narrative intelligence and writing self-efficacy, the absence of a significant difference in the “Creative identity” factor suggests that human creativity remains a distinctive and irreplaceable aspect of storytelling. This finding underscores the importance of striking a balance between generative AI assistance and human creativity, particularly in fields where originality and creative expression are highly valued. On the other hand, it is crucial to do so thoughtfully, recognizing the strengths and limitations of AI tools. These platforms can be especially useful for tasks involving grammar, structure, and organization, leaving space for students to focus on the creative aspects of their writing.”	

The results of this research (Pellas, 2023) have significant implications for the academic community, educational practitioners, and technology developers by demonstrating that generative AI platforms can enhance narrative intelligence and writing self-efficacy among undergraduate students. The positive outcomes suggest that integrating AI technologies into education can support traditional teaching methods, providing personalized feedback and facilitating the writing process, thereby empowering students with essential communication skills. However, the lack of significant impact on creative identity highlights the importance of balancing AI

assistance with human creativity, ensuring that while AI aids in grammar, structure, and organization, students still engage deeply in the creative aspects of their writing.

Picture-Cued Writing

Purpose

The purpose of this study is to propose a novel method for automatically scoring student responses to picture-cued writing tasks using cross-modal deep learning technology and NLP algorithms.

Table 136

Picture-Cued Writing: Purpose

Title	Purpose	Research Question	Author
AI-Assisted Automated Scoring of Picture-Cued Writing Tasks for Language Assessment	“to propose a novel method for automatically scoring student responses to picture-cued writing tasks”	“RQ1. How effectively can the cross-modal deep learning technology be used to measure the matching degrees between visual pictures and their responses written by k12 students? RQ2. Can an AI-assisted tool be further developed for scoring student responses to picture-cued writing tasks automatically by incorporating cross-modal AI technology and NLP algorithms?”	Zhao et al. (2022)

Results

The results from this study are shown in Table 137.

Table 137

Picture-Cued Writing: Results

Title	Results	Author
AI-Assisted Automated Scoring of Picture-Cued Writing Tasks for Language Assessment	“we first designed a picture-cued writing test and collected nearly 4 k responses from 279 K12 students. Based on these responses, we then developed an AI scoring model by incorporating the emerging cross-modal matching technology and some NLP algorithms. The performance of the model was evaluated carefully with six popular measures and was found to demonstrate accurate scoring results with a small mean absolute error of 0.479 and a high adjacent-agreement rate of 90.64%. We believe this method could reduce the subjective elements inherent in human grading and save teachers' time from the mundane task of grading to other valuable endeavors	Zhao et al. (2022)

Title	Results	Author
	such as designing teaching plans based on AI-generated diagnosis of student progress.”	

The results indicate that the AI-assisted scoring model, developed using cross-modal matching technology and NLP algorithms, accurately scored nearly 4,000 picture-cued writing responses from 279 K-12 students, with a small mean absolute error of 0.479 and a high adjacent-agreement rate of 90.64%, suggesting its potential to reduce subjective grading elements and save teachers' time.

Implications

Table 138 provides the implications.

Table 138

Picture-Cued Writing: Implications

Title	Implications	Author
AI-Assisted Automated Scoring of Picture-Cued Writing Tasks for Language Assessment	“The findings from this study provided preliminary validity evidence to support the use of the method in language instruction, which not only can reduce teachers’ grading workload but also provide learners with timely feedback. Furthermore, since multimodal data have been widely used in education, this study provides some inspiration for handling similar tasks by using the emerging cross-modal AI technology.”	Zhao et al. (2022)

The findings suggest that the AI-assisted scoring method can effectively reduce teachers' grading workload and provide timely feedback to learners, offering valuable insights for utilizing cross-modal AI technology in language instruction and similar educational tasks.

Multiple Genres

Purpose

The purpose of the study is to present a comparative analysis by introducing a novel dataset of human-written and LLM-generated texts across various genres, including essays, stories, poetry, and Python code.

Table 139*Multiple Genres: Purpose*

Title	Purpose	Research Question	Author
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“to present a comparative study, introducing a novel dataset of human-written and LLM-generated texts in different genres: essays, stories, poetry and Python code”	Not stated.	Hayawi et al. (2023)

Results

The results from this study are shown in Table 140.

Table 140*Multiple Genres: Results*

Title	Results	Author
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“We employ several machine learning models to classify the texts. The results indicate that the models exhibit superior performance in binary classification tasks, such as distinguishing human-generated text from a specific LLM, compared with the more complex multiclass tasks that involve discerning among human-generated and multiple LLMs. / This research marks a step forward in the field of LLMs, specifically in distinguishing human-written text from AI-generated content. A key contribution of this work is the introduction of a novel dataset for evaluating LLM-generated text from GPT and BARD across multiple genres, including stories and poetry. Our results indicate that while LLMs have made remarkable progress, it remains possible to discern human-written texts from those generated by LLMs using ML and DL models with a small training set. We observed high performance when differentiating BARD-generated texts from human-written texts. However, the task becomes more challenging with GPT-generated text, especially in story writing. By creating and leveraging this unique dataset, we have not only deepened our understanding of the nuanced differences between LLM and human writing but also provided a valuable resource for future research and development in this emerging field.”	Hayawi et al. (2023)

The study (Hayawi et al., 2023) concludes that machine learning models perform better in binary classification tasks, distinguishing between human-generated text and text from a specific LLM, compared to more complex multiclass tasks involving multiple LLMs. The introduction of a novel dataset for evaluating

texts generated by GPT and BARD across various genres reveal that it is still possible to discern human-written texts from those generated by LLMs, although distinguishing GPT-generated texts, especially in story writing, proves more challenging.

Implications

Table 141 provides the implications.

Table 141

Multiple Genres: Implications

Title	Implications	Author
The imitation game: Detecting human and AI-generated texts in the era of ChatGPT and BARD	“This research has several significant implications, particularly in differentiating between human-generated and AI-generated texts. First, the relatively high performance achieved across most experiments, especially SVM, in distinguishing between human and GPT-generated or BARD-generated text indicates that with the right features and model configuration, it is feasible to detect artificially generated text. The only exception is the poor performance of the models in detecting GPT-generated stories from human-written ones. The results could extend to various applications, such as detecting deep fake text in news or social media or validating the authenticity of content. The results also suggest that ML can successfully differentiate between the nuances in human writing and algorithmic output, even with limited data size. However, the differences between the performance of the classifiers when identifying GPT or BARD versus human text, suggest that different AI writing models might have unique ‘fingerprints’ or characteristics that these classifiers can detect. This could lead to the development of more specialised detection tools for different types of LLMs. Finally, the study also contributes to raising ethical implications in the context of LLMs in academic integrity. As AI text generation improves, the potential for misuse increases, necessitating the importance of effective detection tools. It also emphasises the need for a discussion about the responsibility of AI developers, researchers and users in ensuring that these technologies are used ethically and safely.”	Hayawi et al. (2023)

The implications of this research (Hayawi et al., 2023) are significant, particularly for differentiating between human-generated and AI-generated texts. The high performance of machine learning models, especially SVM, in distinguishing texts generated by GPT and BARD from human-written ones, suggests that with the

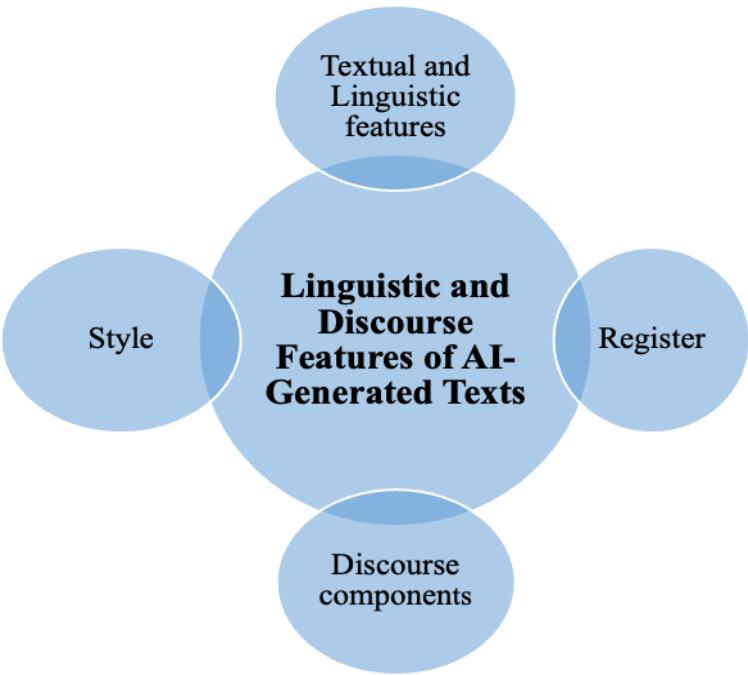
right features and model configuration, it is feasible to detect artificially generated text, though detecting GPT-generated stories remains challenging. These findings have applications in detecting deep fake text and validating content authenticity, indicating that different AI writing models have unique "fingerprints" that classifiers can detect, potentially leading to more specialized detection tools. Additionally, the study highlights ethical implications related to AI text generation, emphasizing the need for effective detection tools and a discussion about the responsibility of AI developers, researchers, and users in ensuring the ethical and safe use of these technologies.

Linguistic and Discourse Features of AI-Generated Texts

Nine of the studies are related to *the linguistic and discourse features of AI-generated texts*. Among these features, several subcategories are identified. These subcategories include *textual and linguistic features*, *register knowledge*, *discourse components*, and *style* (Figure 11).

Figure 11

Linguistic and Discourse Features of AI-Generated Texts



Textual and Linguistic Features

Purpose

There are five studies related to textual and linguistic features.

Table 142

Textual and Linguistic Features: Purpose

Title	Purpose	Research Question	Author
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“to analyze the differences between medical texts written by human experts and those generated by ChatGPT and to design machine learning workflows to effectively detect and differentiate medical texts generated by ChatGPT”	“(1) What is the difference between medical content written by humans and that generated by ChatGPT? (2) Can we use machine learning methods to detect whether medical content is written by human experts or ChatGPT?”	Liao et al. (2023)
A large-scale comparison of human-written versus ChatGPT-generated essays	“to systematically assess the quality of AI-generated content through a large-scale study comparing human-written versus ChatGPT-generated argumentative student essays.”	“RQ1: How good is ChatGPT based on GPT-3 and GPT-4 at writing argumentative student essays? RQ2: How do AI-generated essays compare to essays written by students? RQ3: What are linguistic devices that are characteristic of student versus AI-generated content?”	Herbold et al. (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“to investigate how successful advanced students are in determining whether an essay was written by AI tools or by a human and how much AI assistance they need to summarize, generate text and write from prompts when trained to use an AI-assisted word processor”	“1. How well do advanced students distinguish human-generated from AI-generated essays? What criteria do they think are decisive about Artificial Intelligence authorship? 2. When trained in the use of an AI-assisted word processor, how do advanced students make use of summarising, generating text and writing from prompts? 3. How do students’ essays differ linguistically after training in the use of AI tools from those written before treatment? Do they show differences in burstiness and perplexity?”	Krajka and Olszak (2024)
Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game	“to investigate 16 education academics' perceptions of two pairs of texts written by either ChatGPT or a human by applying Turing's Imitation Game”	“RQ1: How readily can academics correctly identify texts written by either ChatGPT or humans? RQ2: What are the textual factors that academics perceive as being useful in identifying whether texts are written by ChatGPT or humans?”	Matthews and Volpe (2023)

Title	Purpose	Research Question	Author
Clause complexing in research-article abstracts: Comparing human- and AI-generated texts	“to compare 50 abstracts of scientific research articles from Nature with generated versions produced by Bard, ChatGPT, and Poe Assistant by using a modified version of Halliday's clause-complexing framework”	RQ3: How does the application of these factors in decision-making align with accurate or inaccurate identification of text authorship?” “how do the occurrence rates of clauses, tactic relations, and logicosemantic relations in science abstracts compare with the same in Aigenerated abstracts?”	Leong (2023)

The purpose of the first study (Liao et al., 2023) is to analyze the differences between medical texts written by human experts and those generated by ChatGPT, and to design machine learning workflows to effectively detect and differentiate between the two. The second study (Herbold et al., 2023) aims to systematically assess the quality of AI-generated content by comparing human-written and ChatGPT-generated argumentative student essays on a large scale. The purpose of the third study (Krajka & Olszak, 2024) is to investigate how successfully advanced students can distinguish between human-generated and AI-generated essays, and how they use AI assistance for summarizing, generating text, and writing from prompts after being trained with an AI-assisted word processor. The study by Matthews and Volpe (2023) investigates education academics' perceptions of ChatGPT-generated versus human-written texts using Turing's Imitation Game, focusing on their ability to identify the text authorship and the textual factors influencing their decisions. There is another study (Leong, 2023) that aims to compare the occurrence rates of clauses, tactic relations, and logico-semantic relations in scientific research article abstracts from Nature with those generated by Bard, ChatGPT, and Poe Assistant, using a modified version of Halliday's clause-complexing framework.

Results

The results of these studies are in Table 143.

Table 143

Textual and Linguistic Features: Results

Title	Results	Author
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“Medical texts written by humans were more concrete, more diverse, and typically contained more useful information, while medical texts generated by ChatGPT paid more attention to fluency and logic and usually expressed general terminologies rather than effective information specific to the context of the problem. A bidirectional encoder representation from transformers-based model effectively detected medical texts generated by ChatGPT, and the F1 score exceeded 95%. Conclusions: Although text generated by ChatGPT is grammatically perfect and human-like, the linguistic characteristics of generated medical texts were different from those written by human experts. Medical text generated by ChatGPT could be effectively detected by the proposed machine learning algorithms. This study provides a pathway toward trustworthy and accountable use of large language models in medicine.”	Liao et al. (2023)
A large-scale comparison of human-written versus ChatGPT-generated essays	“Our results demonstrate that ChatGPT generates essays that are rated higher regarding quality than human-written essays. The writing style of the AI models exhibits linguistic characteristics that are different from those of the human-written essays. Our results provide clear answers to the first two research questions that consider the quality of the generated essays: ChatGPT performs well at writing argumentative student essays and outperforms the quality of the human-written essays significantly. The ChatGPT-4 model has (at least) a large effect and is on average about one point better than humans on a seven-point Likert scale. Regarding the third research question, we find that there are significant linguistic differences between humans and AI-generated content. The AI-generated essays are highly structured, which for instance is reflected by the identical beginnings of the concluding sections of all ChatGPT essays (‘In conclusion, [...]’). The initial sentences of each essay are also very similar starting with a general statement using the main concepts of the essay topics. Although this corresponds to the general structure that is sought after for argumentative essays, it is striking to see that the ChatGPT models are so rigid in realizing this, whereas the human-written essays are looser in representing the guideline on the linguistic surface. Moreover, the linguistic fingerprint has the counter-intuitive property that the use of discourse markers is negatively correlated with logical coherence.”	Herbold et al. (2023)
“AI, will you help?” How learners use Artificial Intelligence when writing	“The findings of this study indicate the high linguistic sensitivity of the research participants to factors regarding language and layout, which allows them to distinguish human authors from AI-powered texts. The study results reveal some linguistic differences in students’ reports before	Krajka and Olszak (2024)

Title	Results	Author
	and after the treatment. However, the marked variations are rather spontaneous and not obvious. In some AI-enhanced reports, there are even some features of human writing, like high burstiness. As regards specific linguistic differences between students' written assignments before and after the treatment, the four aspects of sentence length, sentence complexity, word frequency and the ratio of unique words inserted in the texts were used as objective measures when comparing students' texts. Generally, when analysing the sentence length of the students' reports after the quasi experimental treatment, most of the study participants formed slightly longer sentences in the post texts (67%, 10 out of 15 students), while only some students shortened their sentences (33%, 5 out of 15 students). It can be assumed that the application of the AI tools in the writing practice of the quasi experimental group of students helped to improve and enrich students' writing skills in general."	
Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game	"Results indicated academics were only able to identify AI-generated texts half of the time, highlighting the sophistication of contemporary generative AI technology. Academics perceived the following categories as important for their decision-making: voice, word usage, structure, task achievement and flow. All five categories of decision-making were variously used to rationalise both accurate and inaccurate decisions about text authorship."	Matthews and Volpe (2023)
Clause complexing in research-article abstracts: Comparing human- and AI-generated texts	# "(a) The occurrence rates of main clauses (4.59) and subordinate clauses (1.71) in Nature abstracts occupied a middle position relative to other abstract groups. The rate of embedded clauses for Nature abstracts was 3.35, differing only from that for Bard abstracts. (b) With the exception of Bard abstracts, the rate of simplexes in the other three abstract groups was about 2. The proportion of simplexes to main clauses in Nature, ChatGPT, and Poe abstracts was between 44% and 52%. (c) Parataxis and paratactic logico-semantic relations were uncommon, and the differences among the abstract groups were statistically insignificant. (d) Hypotactic logico-semantic relations in Nature abstracts were more common among ranking clauses (1.64) than embedded clauses (0.49)." # "The low occurrence rate of embedded hypotactic clauses (<1) was also generally true for the other abstract groups. Hence, even though embedded clauses occurred more often than subordinate clauses, it is the latter that carried the bulk of logico-semantic relations. The most common hypotactic logico-semantic relation involving ranking clauses in Nature and ChatGPT abstracts was enhancement. While ChatGPT abstracts relied mostly on non-finite clauses to express circumstantial information, Nature abstracts also used finite adverbial clauses. None of the chatbots matched Nature in all categories. Poe abstracts, in fact, differed from Nature abstracts in every category (with a significant ANOVA result) except the use of embedded clauses. The chatbot that came closest to matching Nature abstracts was ChatGPT; it differed in the use of subordinate, elaborating, and finite adverbial clauses. Even so, distinct differences were detected—ChatGPT was twice as likely to use elaborating	Leong (2023)

Title	Results	Author
	clauses, but only one sixth as likely to use finite adverbial clauses. The statistics for Nature abstracts offer a possible means by which original texts can be separated from AI-generated ones. Rather than content, the focus here is on the building blocks that construct such content. The work of Levin et al. (2023) has demonstrated one way to do this using Grammarly; this present work proposes another using clause complexing.”	

Liao et al. (2023) find that human-written medical texts are more concrete, diverse, and contain more useful information, while ChatGPT-generated medical texts focus more on fluency and logic, often using general terminology rather than context-specific information. They report that a BERT-based model effectively detects ChatGPT-generated medical texts with an F1 score exceeding 95%. The authors conclude that despite ChatGPT's grammatically perfect and human-like text, the linguistic characteristics of AI-generated medical texts differ from those written by human experts. Herbold et al. (2023) demonstrate that ChatGPT generates essays rated higher in quality than human-written essays. They find that ChatGPT-4 outperforms human-written essays by about one point on a seven-point Likert scale, with at least a large effect size. The study reveals significant linguistic differences between human and AI-generated content. AI-generated essays are highly structured, with identical beginnings in concluding sections and similar initial sentences. The authors note that while this structure aligns with argumentative essay guidelines, ChatGPT models are more rigid in their approach compared to the looser structure of human-written essays. Interestingly, they find that the use of discourse markers is negatively correlated with logical coherence in AI-generated texts.

Krajka and Olszak (2024) report that research participants demonstrate high linguistic sensitivity, allowing them to distinguish between human-authored and AI-powered texts. They observe some linguistic differences in students' reports before

and after using AI tools, although these variations are subtle and not always obvious. In some AI-enhanced reports, they note features typically associated with human writing, such as high burstiness. Analyzing specific linguistic differences, they find that most students (67%) form slightly longer sentences after using AI tools, while 33% shorten their sentences. The authors suggest that the application of AI tools in the quasi-experimental group's writing practice generally helps to improve and enrich students' writing skills.

The results of the study by Matthews and Volpe (2023) suggest that academics can only identify AI-generated texts half of the time, which indicates the advanced sophistication of generative AI; key decision-making factors included voice, word usage, structure, task achievement, and flow, which influences both correct and incorrect identifications of text authorship.

The study (Leong, 2023) reveals that Nature abstracts occupy a middle position in terms of the occurrence rates of main and subordinate clauses compared to AI-generated abstracts, with ChatGPT abstracts coming closest to matching Nature abstracts. However, distinct differences are noted, such as ChatGPT's higher use of elaborating clauses and lower use of finite adverbial clauses. Parataxis and paratactic logico-semantic relations are rare and show no significant differences among the groups, while hypotactic logico-semantic relations are more common in ranking clauses than in embedded clauses across all groups. The results suggest that the structural elements of abstracts, rather than their content, can be used to distinguish between human-written and AI-generated texts.

Implications

Table 144 provides the implications.

Table 144*Textual and Linguistic Features: Implications*

Title	Implications	Author
Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study	“In general, for artificial intelligence (AI) to realize its full potential in medicine, we should not rush into its implementation but advocate for its careful introduction and open debate about its risks and benefits. First, human medical writers will be responsible for ensuring the accuracy and completeness of the information communicated and for complying with ethical and regulatory guidelines. However, ChatGPT cannot be held responsible. Second, training an LLM requires a huge amount of data, but the quality of the data is difficult to guarantee, so the trained ChatGPT is biased. For example, ChatGPT can provide biased output and perpetuate sexist stereotypes [52]. Third, use of ChatGPT may lead to private information leakage. This may be because the LLM remembers personal privacy information in the training set [53]. What is more, the legal framework must be considered. Who shall be held accountable when an AI doctor makes an inevitable mistake? ChatGPT cannot be held accountable for its work, and there is no legal framework to determine who owns the rights to AI-generated work [15].”	Liao et al. (2023)
A large-scale comparison of human-written versus ChatGPT-generated essays	# “Our results provide a strong indication that the fear many teaching professionals have is warranted: the way students do homework and teachers assess it needs to change in a world of generative AI models. For nonnative speakers, our results show that when students want to maximize their essay grades, they could easily do so by relying on results from AI models like ChatGPT. The very strong performance of the AI models indicates that this might also be the case for native speakers, though the difference in language skills is probably smaller. However, this is not and cannot be the goal of education.” # “Consequently, educators need to change how they approach homework. Instead of just assigning and grading essays, we need to reflect more on the output of AI tools regarding their reasoning and correctness. AI models need to be seen as an integral part of education, but one which requires careful reflection and training of critical thinking skills. Furthermore, teachers need to adapt strategies for teaching writing skills: as with the use of calculators, it is necessary to critically reflect with the students on when and how to use those tools. For instance, constructivists⁶² argue that learning is enhanced by the active design and creation of unique artifacts by students themselves. In the present case this means that, in the long term, educational objectives may need to be adjusted. This is analogous to teaching good arithmetic skills to younger students and then allowing and encouraging students to use calculators freely in later stages of education. Similarly, once a sound level of literacy has been achieved, strongly integrating AI models in lesson plans may no longer run counter to reasonable learning goals.”	Herbold et al. (2023)

Title	Implications	Author
“AI, will you help?” How learners use Artificial Intelligence when writing	“There is little doubt that the fields of language teaching and learning could be significantly enhanced by the ongoing development of Artificial Intelligence capabilities. According to the study’s findings, the students were not very successful at distinguishing between AL and human-generated writing with accuracy rates between 40-80%. They tended to assess human writing based on technical accuracy and error and AI writing based on verbal complexity. Applying AI-powered tools could help students improve their writing competence and bring about changes to the way academic writing is taught as a whole. However, as the study’s findings show, they can also hinder the process of improving writing abilities because they lack imagination, use straightforward and sometimes repetitious vocabulary, respond slowly, or make minor grammatical errors. The research reveals that even highly proficient, almost bilingual, academic students encounter considerable difficulties when attempting to differentiate human writing from AI-assisted written texts. These challenges arise when analyzing formal and informal texts.”	Krajka and Olszak (2024)
Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game	# “Experienced academics may be unable to distinguish between texts written by contemporary generative AI technology and humans. # “Academics are uncertain about the current capabilities of generative AI and need support in redesigning assessments that succeed in providing robust evidence of student achievement of learning outcome.” # “Institutions must assess the adequacy of their assessment designs, AI use policies, and AI-related procedures to enhance students’ capacity for effective and ethical use of generative AI technology.”	Matthews and Volpe (2023)
Clause complexing in research-article abstracts: Comparing human- and AI- generated texts	“Scholars (and students) using chatbots to generate content for publication remains a vexing problem. While generative AI has its use in facilitating personalized learning and automating common, tedious tasks, it is merely a tool, but not an author; it has no independent thought, and so cannot take responsibility for its own generated content. Several journals, such as the PNAS (Proceedings of the National Academy of Science), Sage, and Science families of journals, have already banned the listing of chatbots as a ‘coauthor’ in articles; generated content in the manuscript must also be clearly attributed. The recent voluntary commitments by OpenAI, Google, and Meta to watermark AI-generated content is a welcome move (Bartz & Hu, 2023). But detection programs continue to be needed, given how quickly chatbot texts can proliferate, not all of which may be watermarked.”	Leong (2023)

Liao et al. (2023) suggest that for AI to fully realize its potential in medicine, it should be introduced carefully and with open debate about its risks and benefits. They emphasize that human medical writers must ensure accuracy and compliance

with ethical guidelines, as ChatGPT cannot be held accountable. They highlight concerns about data quality, biases in ChatGPT outputs, privacy issues, and the lack of a legal framework for accountability when AI makes mistakes. Herbold et al. (2023) indicate that the strong performance of AI models like ChatGPT in generating essays necessitates changes in how educators assign and assess homework. They argue that AI should be integrated into education to enhance critical thinking and writing skills. Teachers should adapt strategies, reflecting on the use of AI tools and adjusting educational objectives, similar to how calculators are integrated into arithmetic learning.

Krajka and Olszak (2024) find that AI tools can significantly enhance language teaching and learning. However, their study shows that students often struggle to distinguish between AI and human-generated writing. While AI can improve writing competence, it may also hinder creativity and vocabulary use. They reveal that even proficient students find it challenging to differentiate between AI-assisted and human texts, suggesting that AI tools should be used with caution in academic writing instruction.

The implications from the study by Matthews and Volpe (2023) suggest that experienced academics struggle to distinguish between AI-generated and human-written texts, and this indicates a need for support in redesigning assessments to accurately measure student learning outcomes; institutions must evaluate their assessment designs, AI use policies, and procedures to enhance students' effective and ethical use of generative AI technology.

The study by Leong (2023) highlights the problematic nature of scholars and students using chatbots for generating publication content, emphasizing that while AI can aid in learning and task automation, it lacks independent thought and

responsibility. Journals like PNAS, Sage, and Science have banned listing chatbots as coauthors and require clear attribution of AI-generated content. The recent commitment by AI companies to watermark AI-generated content is a positive step, but detection programs are still necessary to manage the rapid production of unmarked chatbot-generated texts.

Discourse Components

Purpose

There is one study that investigates discourse components of texts by learners and ChatGPT.

Table 145

Discourse Components: Purpose

Title	Purpose	Research Question	Author
Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing	“To investigate these issues [ChatGPT is a publicly available chatbot that can quickly generate texts on given topics, but it remains unknown whether the chatbot is as good as or even superior to human writers in all aspects of writing, and whether the chatbot's writing quality can be significantly improved by updating commands], this study compared writing performance on a narrative topic by ChatGPT and Chinese intermediate English (CIE) learners so as to reveal the chatbot's advantages and disadvantages in writing”	“(1) How do ChatGPT and CIE learners perform in narrative English writing from the perspective of discourse components? (2) Which of the components in a revised version can be improved by ChatGPT when user responses are updated for revision? (3) How are the discourse components of the texts produced by ChatGPT and CIE learners correlated?”	Zhou et al. (2023)

The study (Zhou et al., 2023) aims to compare the narrative writing performance of ChatGPT and Chinese intermediate English learners in terms of discourse components to identify the chatbot's strengths and weaknesses in writing and assess the potential improvements through updated user commands.

Results

The results from this study are shown in Table 146.

Table 146

Discourse Components: Results

Title	Results	Author
Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing	“The data were analyzed by virtue of the five discourse components using Coh-Metrix (a special instrument for analyzing language discourse), and the results revealed that ChatGPT performed better than human writers in narrativity, word concreteness, and referential cohesion, but worse in syntactic simplicity and deep cohesion in its initial version. After its commands were updated, the resulting version was better in its syntactic simplicity, but it still lagged far behind the CIE learners' writing in deep cohesion. In addition, a correlation analysis of the discourse components suggests that narrativity was correlated with referential cohesion for both ChatGPT and the human writers, but the correlations varied within each group.”	Zhou et al. (2023)

The study (Zhou et al., 2023) reveals that ChatGPT outperforms Chinese intermediate English learners in narrativity, word concreteness, and referential cohesion, but performs worse in syntactic simplicity and deep cohesion, with updated commands improving syntactic simplicity but still falling short in deep cohesion compared to the human writers; additionally, narrativity is correlated with referential cohesion for both groups, though the specific correlations differed.

Implications

Table 147 provides the implications.

Table 147

Discourse Components: Implications

Title	Implications	Author
Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing	“..the findings suggest that ChatGPT with advanced ToM capabilities could potentially enhance their writing performance, especially in terms of referential cohesion. This study suggests that AI tools like ChatGPT are really powerful in text writing but also lags far behind human writers particularly in the use of advanced inference like deep cohesion and underscores the need for continued research and development to optimize the performance of AI-powered writing tools.”	Zhou et al. (2023)

The findings imply that while ChatGPT shows promise in referential cohesion, it falls behind human writers in advanced inference like deep cohesion, and this highlights the need for ongoing research and development to enhance AI writing tools.

Style

Purpose

There are two studies that focus on style. The first study (Alafnan & Mohdzuki, 2023) examines the stylistic features of ChatGPT-4-generated responses in case studies, business correspondence, and academic writing by analyzing various stylistic features and assessing the accuracy of Turnitin's detection of AI-generated content. The second study (Zaitsu & Jin, 2023) aims to identify and analyze the distinctive stylistic characteristics in Japanese texts produced by AI language models (specifically ChatGPT versions using GPT-3.5 and GPT-4) compared to those written by human authors.

Table 148

Style: Purpose

Title	Purpose	Research Question	Author
Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4	“to examine the stylistic features of case study, business correspondence, and academic writing ChatGPT-4-generated responses by exploring sentence length, paragraph structure, word choice, mood, tense, voice, pronouns, keywords density, lexical density, lexical diversity, and reading ease”	“1. Can Turnitin detect ChatGPT 4-generated responses? 2. Are Turnitin AI reports accurate? 3. What are the stylistic features of ChatGPT-4?”	Alafnan and Mohdzuki (2023)
Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis	“to compare Japanese stylometric features of texts generated by ChatGPT, equipped with GPT-3.5 and GPT-4, and those written by humans.”	Not stated.	Zaitsu and Jin (2023)

Results

The results from this study are shown in Table 149.

Table 149

Style: Results

Title	Results	Author
Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4	“The study revealed that ChatGPT-4 case study-generated responses are produced in paragraphs of 2 to 3 sentences of 16 to 18 words each. The sentences are mainly formed in the imperative mood. The use of the second-person pronoun “you” and the second-person possessive determiner “your” is prevalent. Keywords and lexical density are relatively low, lexical diversity is average, and the reading ease is relatively high. The study also found that ChatGPT-4 business correspondence responses are generated in paragraphs of 2 to 3 sentences of 16 to 20 words each. The sentences are mainly generated in declarative mood thru simple present tense in active voice using third-person singular pronouns. Technical words and abbreviations are used without outlining what they stand for. The keywords density, lexical density, and lexical diversity are high, and the reading ease is low. The study also revealed that ChatGPT-4 academic writing-generated responses are provided in paragraphs of 3 to 4 sentences of 16 to 19 words each. The sentences are mainly generated in declarative mood using active voice, agentless passive in times, with diverse present tenses. Keywords and lexical densities are high, and the lexical diversity is low, which makes the reading ease average difficulty, except for the undefined abbreviations. Noticeably, ChatGPT-4 supports the transgender movement by intentionally using the third-person plural pronoun “they” to refer to a singular.”	Alafnan and Mohdzuki (2023)
Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis	“MDS revealed distinct distributions at each stylometric feature of GPT (3.5 and 4) and human. Although GPT-4 is more powerful than GPT-3.5 because it has more parameters, both GPT (3.5 and 4) distributions are overlapping. These results indicate that although the number of parameters may increase in the future, GPT-generated texts may not be close to that written by humans in terms of stylometric features. Second, we verified the classification performance of random forest (RF) classifier for two classes (GPT and human) focusing on Japanese stylometric features. This study revealed the high performance of RF in each stylometric feature: The RF classifier focusing on the rate of function words achieved 98.1% accuracy. Furthermore, the RF classifier focusing on all stylometric features reached 100% in terms of all performance indexes (accuracy, recall, precision, and F1 score). This study concluded that at this stage we human discriminate ChatGPT from human limited to Japanese language.”	Zaitso and Jin (2023)

The study (Alafnan & Mohdzuki, 2023) findings demonstrate that ChatGPT-4-generated responses exhibit distinct stylistic features across different types of writing: case studies use short paragraphs and imperative mood with prevalent second-person pronouns; business correspondence employs declarative mood in the active voice with technical terms and low reading ease; academic writing features a mix of active and passive voice with high keyword density and average reading ease, with notable use of "they" as a singular pronoun.

The second study (Zaitzu & Jin, 2023) finds that while GPT-4 has more parameters than GPT-3.5, both versions produce overlapping distributions in stylometric features, indicating that increased parameters do not necessarily make AI-generated texts more human-like. A random forest (RF) classifier focusing on Japanese stylometric features demonstrates high performance, with 98.1% accuracy for function word rates and 100% accuracy, recall, precision, and F1 score when considering all features, suggesting effective discrimination between ChatGPT-generated and human-written texts in Japanese.

Implications

Table 150 provides the implications.

Table 150

Style: Implications

Title	Implications	Author
Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4	“This study found that even though Turnitin’s AI report can detect AI-generated responses, their accuracy cannot be taken into consideration for punitive purposes for a couple of reason. Firstly, Turnitin mentions that these AI reports shall not be used for misconduct purposes. Secondly, after testing the accuracy of these reports, this study found an inconsistency in their outcome. The researchers gradually added human-generated material to the AI (ChatGPT-4)-generated material but the AI report remained 100%. This proves Turnitin’s claim that these reports shall not be used for misconduct purposes as their accuracy is way below 98%, the percentage claimed by Turnitin, accurate.”	Alafnan and Mohdzuki (2023)

Title	Implications	Author
Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis	“The conclusions of this study are as follows: (1) Stylometric features in Japanese were distinct between texts generated by ChatGPT (GPT-3.5 and -4) and texts written by humans. (2) Distribution of GPT-4 generated-texts overlapped the distribution of GPT-3.5 in stylometric features and were not close to those of humans. (3) Currently, we can classify AI-generated and human texts in the case of Japanese, with performance levels (accuracy, recall, precision, and F1 score) of 100%.”	Zaitsu and Jin (2023)

The first study's implications (Alafnan & Mohdzuki, 2023) highlight that Turnitin's AI detection reports, while capable of identifying AI-generated responses, lack the accuracy required for punitive measures, evidenced by inconsistencies in detection when human-generated material is mixed with AI content, thereby emphasizing Turnitin's advisory against using these reports for misconduct purposes.

The implications from the second study (Zaitsu & Jin, 2023) indicate that Japanese stylometric features distinctly differentiate between ChatGPT-generated (both GPT-3.5 and GPT-4) and human-written texts, with AI-generated texts not closely resembling human writing. Despite the increased parameters in GPT-4, its outputs remain distinguishable from human texts, and current classification methods using stylometric features can achieve perfect accuracy, recall, precision, and F1 scores in identifying AI-generated content in Japanese.

Register

Purpose

There is one study (Punar Özçelik & Yangın Ekşi, 2024) that aims to examine the impact of using ChatGPT as a learning assistant on students' acquisition of register knowledge across various writing tasks, specifically investigating whether it helps students improve their self-editing skills and gathering students' opinions and suggestions on its use.

Table 151*Register: Purpose*

Title	Purpose	Research Question	Author
Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study	“to address this research gap [re: register] by examining the impact of ChatGPT, an AI-powered chatbot, on the acquisition of register knowledge across various writing tasks.”	“1. Can the use of ChatGPT as a learning assistant help students improve self-editing their writing? 2. What are the students’ opinions and suggestions regarding using ChatGPT as a learning assistant?”	Punar Özçelik and Yangın Ekşi (2024)

Results

The results from this study are shown in Table 152.

Table 152*Register: Results*

Title	Results	Author
Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study	“Preliminary results indicate that students found ChatGPT beneficial for acquiring formal register knowledge but perceived it as unnecessary for informal writing. Additionally, the effectiveness of ChatGPT in teaching neutral register was questioned by the participants. The results of the study showed that ChatGPT has the potential to assist students in developing their writing abilities, particularly in formal register. Students used ChatGPT enthusiastically and actively for their writing tasks. They benefited from its suggestions and corrections to enhance the formal aspects of their writings. The study noted several difficulties, though, including technical difficulties and limitations in interpreting informal and neutral registers. Despite these drawbacks, students’ varied viewpoints and insightful ideas emphasized the need for significant functional improvements to ChatGPT to make it a more useful learning tool for self-editing. ChatGPT can provide significant assistance to students in their writing tasks with careful evaluation and modifications. Students had perceived benefits of using ChatGPT for formal text corrections, while they reported its lack of didactic nature. They had some trustworthiness and reliability concerns. Students stated that ChatGPT’s limitations in understanding neutral and informal registers posed potential challenges as a learning assistant, making it suitable only for specific types of texts. They reported the need for improvement in ChatGPT’s algorithms and suggestions. While they had negative opinions on its impact on writing levels and meaning changes, its usefulness for timesaving and shortening long texts were noted as positive. This research contributes to the existing literature by shedding new light on the effects of AI-generated chatbots in register learning during the writing process, offering insights into their potential as learning assistants.”	Punar Özçelik and Yangın Ekşi (2024)

The study (Punar Özçelik & Yangın Ekşi, 2024) concludes that students perceive ChatGPT as beneficial for acquiring formal register knowledge but less useful for informal writing and question its effectiveness in teaching neutral register. While ChatGPT helps improve students' formal writing by providing suggestions and corrections, technical issues and limitations in interpreting informal and neutral registers are noted. Students appreciate ChatGPT's ability to save time and shorten texts but have concerns about its reliability and impact on writing quality. The study highlights the need for improvements in ChatGPT's algorithms to enhance its utility as a learning tool for self-editing, emphasizing its potential in specific contexts.

Implications

Table 153 provides the implications of this study.

Table 153

Register: Implications

Title	Implications	Author
Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study	“Students’ varied viewpoints and insightful ideas emphasized the need for significant functional improvements to ChatGPT to make it a more useful learning tool for self-editing. ChatGPT can provide significant assistance to students in their writing tasks with careful evaluation and modifications.”	Punar Özçelik and Yangın Ekşi (2024)

The implications of this study highlight the need for substantial functional improvements to ChatGPT to enhance its effectiveness as a self-editing learning tool, and this emphasizes the fact that with careful evaluation and modifications, ChatGPT can significantly assist students in their writing tasks.

CHAPTER 5: DISCUSSION

Introduction

This chapter discusses the key findings of the study. It begins with an overview of the study, followed by a discussion of the major findings. The chapter then explores the implications for practice and suggests directions for future research. Finally, it addresses the study's limitations.

Overview of the Study

The rapid advancement of AI-driven writing tools and platforms has made it essential for EFL instructors to follow guidelines for effective writing instruction. Despite the significant potential these technologies hold for enhancing writing instruction, many educators currently lack the necessary skills to effectively integrate these tools into their teaching practices. In response to this need, this interpretive meta-synthesis study was conducted to lay the groundwork for comprehensive guidelines for writing instruction aimed at helping EFL writing instructors manage the challenges and opportunities presented by AI-informed writing instruction. The research involved a systematic analysis of recent relevant literature to identify and categorize recurring concepts related to the application of AI in writing pedagogy, leading to the development of core areas for inclusion in the guidelines that will help develop a competency framework.

These guidelines are designed to address the evolving needs of EFL instructors, enabling them to function effectively in the AI-enhanced educational environment while harnessing the potential of AI to improve student outcomes. They

provide practical approaches for EFL writing instructors, addressing both the benefits and challenges of AI integration in teaching writing. The goal of guidelines is to guide instructors for using AI tools effectively while maintaining academic integrity and promoting student engagement.

By examining the interrelationships among the identified concepts, the study provided a comprehensive understanding of the current landscape of AI research in writing instruction. The synthesis of these interconnected insights into *Core categories* served as the basis for formulating guidelines for effective writing instruction. These guidelines intend for instructors to incorporate AI-driven techniques and tools into their writing instruction, fostering improved teaching practices and better learning outcomes.

Developed through a comprehensive meta-interpretive study focused on AI and writing, these guidelines provide a structured approach to developing and applying essential instructional strategies for EFL writing instructors in AI-integrated classrooms. The guidelines are organized into seven core areas, each addressing a critical aspect of AI integration in EFL writing instruction:

1. *Good Academic Conduct in Writing*
2. *Assessment and Writing*
3. *Evaluating and Enhancing the Quality and Impact of Academic Content*
4. *User Behavior and Pedagogical Implications of AI-Assisted Writing*
5. *AI-Integrated Writing Instruction and Outcomes*
6. *Diverse Writing Genres and Modalities*
7. *Linguistic and Discourse Features of AI-Generated Texts*

For each area, the guidelines present statements under the following sections,

offering practical approaches for effective instruction in AI-integrated environments:

- *Knowledge of*
- *Instructional Design*
- *Process of Teaching and Student Engagement*
- *Technology Integration*
- *Communication & Collaboration*
- *Assessment and Evaluation*
- *Professional Growth and Development*

Each core area begins with an introduction outlining its significance, followed by statements in the form of guidelines. Additionally, each area includes a discussion alignment with and potential gaps in existing discussion in relation to alignment with, and potential gaps in instructional practices informed by existing competency frameworks. This section provides insights into potential improvements for current practices and their policy implications for educators, researchers, policymakers, and developers.

These guidelines aim to promote ethical, effective, and innovative educational practices through the development of essential guidelines for AI-enhanced writing instruction.

Guidelines for Effective Writing Instruction for EFL Writing Instructors in AI-Enhanced Classes

Guideline Area 1: Good Academic Conduct in Writing

In this guideline area, the emerged relational core concepts include *academic dishonesty, academic integrity, plagiarism, AI-generated text authorship and ownership, transparency and disclosure, and ethical use*. Related to each core category, guidelines stemming from the studies are presented section by section as

follows:

Knowledge of:

- Demonstrates understanding of ethical considerations related to AI use in academic writing, including issues of plagiarism, intellectual property, and academic integrity (Alexander et al., 2023; Amirjalili et al., 2024; Črček & Patekar, 2023; Gustilo et al., 2024; Ibrahim, 2023; Mahapatra, 2024; Mohammadkarimi, 2023; Rojas, 2024; Yan, 2023).
- Articulates the potential impacts of AI writing tools on academic honesty and the development of critical thinking skills (Herbold et al., 2023; Krajka & Olszak, 2024; Malik et al., 2023; Popkov & Barrett, 2024; Watts et al., 2023; Waltzer et al., 2023).
- Understands the capabilities and limitations of AI detection tools in identifying potential instances of academic misconduct (Alafnan & Mohdzuki, 2023; Elkhatat et al., 2023; Hayawi et al., 2023; Ibrahim, 2023; Perkins et al., 2023; Popkov & Barrett, 2024).

Instructional Design:

- Designs assignments that promote original thinking and make it challenging for AI to generate adequate responses without significant human input and critical thought (Herbold et al., 2023; Mohammadkarimi, 2023; Pellas, 2023; Rojas, 2024; Yeadon et al., 2022).
- Develops curriculum elements that explicitly address ethical AI use in academic writing, including proper attribution and citation of AI-generated content (Barrett & Pack, 2023; Lee et al., 2024; Zou & Huang, 2023).

- Creates learning activities that help students distinguish between appropriate AI assistance and academic dishonesty (Ivanović, 2023; Liu et al., 2024; Watts et al., 2023).

Process of Teaching and Student Engagement:

- Guides students in the responsible and ethical use of AI tools in their writing process, emphasizing the importance of maintaining academic integrity (Barrett & Pack, 2023; Dong, 2024; Gustilo et al., 2024; Zou & Huang, 2023).
- Facilitates discussions on the ethical implications of AI use in academic writing, encouraging students to develop their own ethical framework (Dong, 2024; Gallagher & Wagner, 2024; Watts et al., 2023).
- Engages students in exercises that demonstrate the differences between AI-generated and original human-written content, fostering critical evaluation skills (Amirjalili et al., 2024; Herbold et al., 2023; Krajka & Olszak, 2024; Mahyoob et al., 2023; Vicente-Yagüe Jara et al., 2023; Watts et al., 2023; Zhou et al., 2023).

Technology Integration:

- Implements AI detection tools cautiously to identify potential academic misconduct while recognizing and explaining the limitations of such tools to students (Alafnan & Mohdzuki, 2023; Cingillioglu, 2023; Ibrahim, 2023; Perkins et al., 2023).
- Utilizes AI writing tools in controlled classroom settings to demonstrate both their capabilities and the ethical considerations of their use (Coman & Cardon, 2024; Lee et al., 2024; Liu et al., 2024; Rojas, 2024; Tossell et al., 2024; Watts et al., 2023).

Communication & Collaboration:

- Collaborates with colleagues to develop and implement clear, consistent policies on acceptable AI use in academic writing tasks (Črček & Patekar, 2023; Gallagher & Wagner, 2024; Mohammadkarimi, 2023; Ivanović, 2023; Matthews & Volpe, 2023; Nguyen et al., 2024).
- Engages in open dialogue with students about the evolving nature of academic integrity in the age of AI (Barrett & Pack, 2023; Gustilo et al., 2024; Pellas, 2023).

Assessment and Evaluation:

- Develops assessment strategies that effectively evaluate student writing for originality and critical thinking, even when AI tools have been used as writing aids (Liu et al., 2024; Niloy et al., 2024; Yeadon et al., 2022;)
- Creates rubrics that include criteria for ethical AI use and proper attribution of AI-generated content (Alexander et al., 2023; Niloy et al., 2024; Yeadon et al., 2022).

Professional Growth and Development:

- Stays informed about emerging AI technologies and their potential impacts on academic writing and integrity (Cardon et al., 2023; Mohammadkarimi, 2023).
- Participates in professional development opportunities focused on ethical considerations and best practices for maintaining academic integrity in an AI-integrated writing environment (Alexander et al., 2023; Gustilo et al., 2024; Mohammadkarimi, 2023).

Alignment and Gaps. This guideline area is particularly important in the

context of AI-enhanced EFL writing instruction because it addresses the ethical challenges and considerations that arise from integrating AI tools in academic writing. The statements focus on maintaining academic integrity, preventing plagiarism, and fostering responsible AI use among students. The guideline statements in this area ensure that teachers are equipped to navigate the complex terrain of AI-assisted writing while upholding academic standards and fostering critical thinking skills.

When compared to existing competency statements, these guidelines in this study align with the ISTE Standards for Educators, particularly Standard 3 (Citizen) and Standard 7 (Analyst). For example, the focus on ethical considerations and academic integrity in AI use aligns with ISTE's emphasis on promoting and modeling digital citizenship, while the emphasis on understanding AI capabilities and limitations reflects ISTE's call for educators to understand and use data to drive instruction and support students (ISTE, 2024).

These guidelines also align closely with the TESOL International Association's Standards, especially Domain 3 (Instructing) and Domain 8 (Commitment and Professionalism). The ones related to designing assignments that promote original thinking and guiding students in responsible AI use directly reflect TESOL's standards for creating supportive learning environments and using appropriate technology to enhance instruction (TESOL International Association, 2008).

The emerging guidelines align closely with several aspects of the UNESCO framework (UNESCO, 2024). The "Ethics of AI" aspect is well-represented, mirroring the progression from understanding ethical principles to guiding responsible use and co-creating ethical rules. While the "AI Foundations &

Applications" aspect is less explicit in the emerging guidelines, there is some alignment in understanding AI capabilities. The "AI pedagogy" aspect is reflected in the instructional design and teaching process competencies, focusing on ethical AI integration in teaching. The "AI for professional development" aspect aligns with the professional growth competencies, both emphasizing continuous learning about AI technologies and their implications in education. While the UNESCO framework provides a broader overview of AI competencies for teachers, these emerging guidelines offer more specific guidance for maintaining academic integrity in writing instruction. They address unique challenges in academic writing, such as plagiarism and proper attribution of AI-generated content, which are not explicitly covered in the UNESCO framework.

These emerging guidelines could actually complement other competency frameworks such as British Council Continuing Professional Development (CPD) Framework, the TESOL International Association's Standards for ESL/EFL Teachers of Adults, the European Profiling Grid (EPG), the Cambridge English Teaching Framework, the Equals Framework for Language Teacher Training & Development, the BALEAP Competency Framework for Teachers of English for Academic Purposes (EAP)] by adding a specific focus on AI ethics and integration. This alignment would enhance the existing emphasis on teachers' responsibility to engage with new developments in language teaching methodology and technology adoption, specifically in writing instruction.

The competency frameworks highlighted in the document reveal several gaps when aligned with the guideline statements provided, particularly concerning AI's role in academic integrity and writing instruction. For instance, while frameworks like the British Council's CPD and the TESOL Standards acknowledge the

importance of technology integration, they fall short of addressing AI's specific challenges and ethical considerations, such as those outlined in these statements on plagiarism, AI detection tools, and the ethical use of AI-generated content. This gap is especially pronounced in frameworks like the European Profiling Grid (EPG) and Cambridge English Teaching Framework, where AI-related competencies are only indirectly referenced under broader digital literacy or technology integration categories.

Moreover, the process of teaching and student engagement, as well as assessment strategies, are underdeveloped in these frameworks concerning AI's impact. The statements in this study's framework emphasize the need for explicit instruction and evaluation of AI-assisted writing, something that the existing frameworks barely touch upon. Also, the competencies around creating rubrics for ethical AI use and guiding students in distinguishing between human and AI-generated content are crucial areas that require more direct focus. The lack of AI-specific training and professional development opportunities in these frameworks also highlights a significant gap and emphasizes the need for updated guidelines that reflect the evolving technological landscape in education.

Policy Impact and Future Directions. Institutionally, these guidelines highlight the need for clear policies on AI use in academic writing, potentially requiring revisions to existing academic integrity guidelines (Alexander et al., 2023; Fyfe, 2022; Mohammadkarimi, 2023; Köbis & Mossink, 2020; Matthews & Volpe, 2023; Nguyen et al., 2024; Niloy et al., 2024; Nugroho et al., 2024; Watts et al., 2023; Yan, 2023; Zhao et al., 2024; Zou & Huang, 2023). They also suggest a need for institutions to provide comprehensive training and professional development opportunities in AI ethics and integration (Alexander et al., 2023; Boudouaia et al.,

2024; Gallagher & Wagner, 2024; Mohammadkarimi, 2023; Ivanović, 2023; Papakonstantinidis et al., 2024).

Future research directions for promoting good academic conduct in AI-assisted writing cover various critical areas. One key direction is empirically validating the effectiveness of these guidelines in maintaining academic integrity in AI-integrated writing environments, possibly through longitudinal studies tracking changes in student writing practices and academic honesty (Barrett & Pack, 2023; Boudouaia et al., 2024; Malik et al., 2023; Papakonstantinidis et al., 2024; Pellas, 2023; Yan, 2023; Zou & Huang, 2023). Additionally, several studies (Alexander et al., 2023; Barrett & Pack, 2023; Gustilo et al., 2024; Matthews & Volpe, 2023) point to the need for broader investigations into stakeholder perceptions and the use of AI tools across different educational levels and cultural contexts. Several other studies suggest focusing on detection. The development of robust AI detection methods and their integration with human judgment is a key focus (Amirjalili et al., 2024; Cingillioglu, 2023; Popkov & Barrett, 2024), as is the exploration of new areas in intelligent computer-assisted language learning to address AI-assisted plagiarism (Ibrahim, 2023). Many researchers emphasize the importance of creating and evaluating ethical guidelines and policies for AI use in academic settings (Mohammadkarimi, 2023; Rojas, 2024; Tang et al., 2023), with some proposing a shift toward a more positive framing of academic integrity issues (Gallagher & Wagner, 2024). The challenges of authorship and transparency in AI-assisted writing are also highlighted, with recommendations for designing better authorship declarations (Draxler et al., 2023) and investigating the impact of AI disclosure on research perception (Tang et al., 2023). Additionally, there is a call for using diverse data sources to understand students' AI use patterns (Črček & Patekar, 2023).

Collectively, these directions underscore the need for comprehensive research to address the complex challenges that AI presents to traditional notions of academic integrity and writing practices.

Guideline Area 2: Assessment and Writing

Within this core category, the relational core concepts that emerged are *Feedback, Automated Evaluation/ Scoring, and Assessment of Writing*.

The guideline statements stemming from the studies are categorized as follows:

Knowledge of:

- Demonstrates understanding of AI-generated feedback characteristics, including its readability, detail, and efficiency compared to human feedback (Banihashem et al., 2024; Escalante et al., 2023; Guo & Wang, 2023; Rad et al., 2023; Steiss et al., 2024; Yang et al., 2024).
- Articulates the complementary nature of AI and human feedback, recognizing the strengths and limitations of each in the assessment process (Escalante et al., 2023; Guo & Wang, 2023; Lu, 2019; Steiss et al., 2024).
- Understands the potential biases and limitations of AI-enabled automated essay scoring systems (Lu, 2019; Wolf & Wolf, 2023; Yang et al., 2024; Zhao et al., 2022).

Instructional Design:

- Designs writing instruction that integrates AI feedback tools to improve students' writing outcomes, engagement, and feedback literacy (Guo & Wang, 2023; Rad et al., 2023)
- Develops a mixed approach to feedback while combining AI-generated feedback with human-tutor interaction to maximize benefits and address the

limitations of each (Escalante et al., 2023; Guo & Wang, 2023; Steiss et al., 2024).

- Creates assessment tasks that effectively evaluate student writing skills in the context of AI assistance, with a focus on critical thinking and originality (Liu et al., 2024; Herbold et al., 2023; Rojas, 2024; Zou & Huang, 2023).

Process of Teaching and Student Engagement:

- Guides students in effectively utilizing both AI and teacher assessments, with tailored assistance to various writing proficiency levels (Boudouaia et al., 2024; Lu et al., 2024; Song & Song, 2023;).
- Teaches students to critically evaluate and apply AI-generated feedback, fostering their ability to make informed revisions (Liu et al., 2024; Dong, 2024).
- Engages students in discussions about the role of AI in assessment, promoting understanding of its benefits and limitations (Lozic & Štular, 2023; Semrl et al., 2023; Watts et al., 2023;).

Technology Integration:

- Implements AI writing assessment tools gradually and incrementally, with clear pedagogical objectives and teacher oversight (Boudouaia et al., 2024; Rojas, 2024; Watts et al., 2023).
- Utilizes AI-enabled automated essay scoring systems to provide efficient and reliable preliminary writing assessment, with human oversight for final evaluation (Guo & Wang, 2023; Ivanović, 2023; Zhao et al., 2022).
- Integrates AI-based applications to enhance students' writing outcomes and feedback literacy in order to complement rather than replace human assessment (Watts et al., 2023)

Communication & Collaboration:

- Collaborates with colleagues to develop best practices for integrating AI tools in writing assessments (Cardon et al., 2023; Matthews & Volpe, 2023).
- Engages in interdisciplinary collaboration to enhance AI writing evaluation systems for more meaningful and personalized feedback (Ivanović, 2023; Vicente-Yagüe Jara et al., 2023)
- Communicates clearly with students about the role of AI in the assessment process, setting appropriate expectations and boundaries (Dong, 2024; Nugroho et al., 2024).

Assessment and Evaluation:

- Develops assessment strategies that effectively evaluate student writing in the context of AI assistance, focusing on higher-order skills such as critical thinking and creativity (Niloy et al., 2024; Rojas, 2024; Yeadon et al., 2022).
- Creates rubrics that account for the use of AI tools in the writing process while still emphasizing original thought and proper attribution (Alexander et al., 2023; Niloy et al., 2024; Yeadon et al., 2022).
- Implements evaluation methods that consider both the benefits and risks of AI use in writing, adapting criteria to address new challenges in assessment (Alexander et al., 2023; Lu, 2019; Nugroho et al., 2024).

Professional Growth and Development:

- Pursues ongoing training in recognizing features of AI-created texts and distinguishing them from human-written essays for effective assessment (Alexander et al., 2023; Mohammadkarimi, 2023).
- Develops proficiency in prompt construction for effective use of AI tools in Automated Writing Evaluation (AWE) (Nugroho et al., 2024; Parker et al.,

2023).

- Stays informed about advancements in AI-assisted writing assessment technologies and their implications for educational practices (Perkins et al., 2023; Rojas, 2024).

Alignment and Gaps. The guideline area of Assessment and Writing is highly significant in the context of AI integration in EFL writing instruction since it could address the challenges in the evolving landscape where AI tools are increasingly used to provide feedback, evaluate essays, and support the overall assessment process. This guideline area aims to ensure that teachers can effectively leverage AI technologies while balancing and maintaining the critical role of human judgment and pedagogical expertise.

The guideline statements in this area align well with several existing frameworks, such as the British Council Continuing Professional Development (CPD) Framework, which emphasizes "Assessing learning" as a professional practice. These competencies extend this practice by specifically addressing AI integration in assessment (British Council, 2015). Similarly, they provide a more detailed and AI-specific extension of the TESOL International Association's Standards for ESL/EFL Teachers of Adults, which includes a domain on "Assessing" that emphasizes using appropriate technology to enhance assessment (TESOL International Association, 2008). While the European Profiling Grid (EPG) does not explicitly mention AI, its emphasis on digital media use and assessment strategies aligns with the competencies outlined here, particularly in the "Technology Integration" category (North et al, 2017). The BALEAP Competency Framework for Teachers of English for Academic Purposes (EAP) emphasizes assessment practices in writing instruction, and the guidelines outlined here complement this by

addressing the specific challenges and opportunities presented by AI in writing assessment (BALEAP, 2008).

The ISTE Standards for Educators, particularly the Analyst standard (2.7), align closely with the guidelines outlined in the Assessment and Writing area for AI integration in EFL writing instruction (Crompton, 2023). This standard emphasizes the importance of educators understanding and using data to drive instruction and support student achievement, and this part resonates strongly with the AI-assisted assessment competencies. The sub-standard 2.7.a, which calls for offering alternative assessments, is particularly relevant in the context of AI-enabled writing evaluation because it encourages educators to explore diverse assessment methods that may include AI-assisted tools. This aligns with the guideline statements that focus on designing writing instruction that integrates AI feedback tools and developing mixed approaches to feedback. The 2.7.b sub-standard, which promotes the use of technology to create assessments, directly supports the guidelines related to implementing AI writing assessment tools and utilizing AI-enabled automated essay scoring systems. This underscores the importance of technological integration in modern assessment practices. Finally, the 2.7.c sub-standard, which emphasizes using data to guide progress, corresponds with the guidelines that focus on utilizing AI-generated feedback and automated scoring to inform instructional decisions and student progress. This connection highlights the potential of AI to provide rich, data-driven insights that can enhance the assessment process and guide personalized instruction in EFL writing contexts.

When comparing the guideline statements on assessment and writing with other frameworks, such as the European Profiling Grid (EPG), the Cambridge English Teaching Framework, and the UNESCO Draft AI Competency Frameworks,

several gaps become apparent. The EPG, for instance, offers a structured approach to teaching competencies but lacks specificity regarding the integration of AI in writing assessment. While it covers digital media competencies at a high level, it does not delve into the characteristics of AI-generated feedback or the complementary nature of AI and human feedback. The framework's focus on traditional assessment methods misses out on the complexities introduced by AI, such as biases in automated essay scoring systems, which are critical considerations highlighted in these guideline statements.

The Cambridge English Teaching Framework similarly touches on the use of digital resources and assessment practices but does not extend this to include AI's role in writing evaluation. The framework's emphasis on traditional feedback mechanisms does not account for the integration of AI tools in instructional design, which is essential for improving students' writing outcomes and engagement, as noted in this guideline area. Additionally, the framework does not address the gradual and increasing implementation of AI tools with clear pedagogical objectives, a key area that these statements emphasize for effective technology integration in education.

In contrast, the UNESCO Draft AI Competency Frameworks begin to address some aspects of AI's impact on education but are still in development and primarily focus on broader AI literacy rather than the specific nuances of AI in writing assessment. While UNESCO's framework introduces concepts such as AI pedagogy and the ethical use of AI, it lacks detailed guidance on developing rubrics that account for AI use or on creating assessment strategies that balance AI's efficiency with human oversight. This highlights a broader gap across existing frameworks, where the rapid evolution of AI in educational contexts demands more specialized

competencies, particularly in assessment and writing, to ensure educators are equipped to navigate these new challenges effectively.

Policy Impact and Future Directions. These guidelines have significant directions for institutional policies. To start with, institutions may need to revise their teacher training curricula to include modules on AI-assisted writing assessment (Ivanović, 2023). Teachers need to be prepared to implement these technologies effectively. Also, new policies regarding the use of AI in writing assessments or redesigned, innovative assessment tools may need to be developed (Alexander et al., 2023; Dong, 2024; Lu et al., 2024; Matthews & Volpe, 2023; Nguyen et al., 2024; Yeadon et al., 2022). They must address issues of fairness, transparency, and academic integrity. Additionally, ongoing professional development programs should be implemented to keep teachers updated on the latest AI technologies in writing assessments. To address issues of bias, privacy, and data protection, clear institutional guidelines on the ethical use of AI in writing assessments are necessary. Finally, institutions may also need to invest in AI-enabled writing assessment tools for proper infrastructure to support their implementation.

To further develop this area, several follow-up studies and research directions are suggested. To further develop this area, several key research directions are proposed. Jiang et al. (2023) suggest exploring broader language patterns in LLMs, expanding the analysis to more writing genres and larger samples, and improving AI model recall while maintaining precision. Punar Özçelik and Yangın Ekşi (2024) recommend studies with diverse participants, varied writing assessments, and more structured prompt engineering. Alexander et al. (2023) call for research on reviewing and revising assessment policies in light of AI advancements, as well as large-scale studies examining AI's impact on academic integrity across higher education. They

also emphasize investigating stakeholder attitudes towards AI in education. Lu et al. (2024) advocate for research with extensive samples across various genres and experimenting with different prompts to improve AI-generated feedback in writing assessments. These directions aim to enhance our understanding of AI-assisted writing assessment and inform more effective integration of AI in EFL writing instruction.

Guideline Area 3: Evaluating and Enhancing the Quality and Impact of Academic Content

Related to this core category, the relational core concepts that emerged are

Performance and Impact

- *Capabilities*
- *Accuracy*
- *Proficiency and Originality*
- *Reliability and Quality*
- *Usefulness for Scholars*
- *Impact Efficacy*

Output Characteristics

- *Quality and Accuracy of Academic Content*
- *Authenticity*
- *Creativity*
- *Originality*

Detection, Identification, and Machine Learning.

The guideline statements stemming from the studies are categorized as follows:

Knowledge of:

- Understands the criteria for evaluating the quality of academic content, including AI-generated or AI-assisted work (Altmäe et al., 2023; Lozic & Štular, 2023).
- Recognizes the potential of AI tools to enhance the reach and impact of academic content through improved accessibility and translation (Abuyaman, 2023; Alexander et al., 2023; Huespe et al., 2023; Tang et al., 2023; Utami et al., 2023).
- Demonstrates awareness of the challenges in maintaining content quality when integrating AI tools into the academic writing process (Liu et al., 2024; Parker et al., 2023; Steiss et al., 2024).
- Understands the capabilities and limitations of AI in generating and enhancing academic content across various disciplines (Dashti et al., 2023; Gustilo et al., 2024; Hwang et al., 2024; Wu & Dang, 2023).
- Recognizes the differences in accuracy, reliability, and quality between AI-generated and human-written academic content (Ariyaratne et al., 2023; Babl & Babl, 2023; Cheng et al., 2023; Dağcı et al., 2023; Hakam et al., 2024; Hwang et al., 2024; Steiss et al., 2024).

Instructional Design:

- Develops curricula that teach students to critically evaluate the quality and credibility of both AI-generated and human-written academic content (Cardon et al., 2023; Coman & Cardon, 2024; Gallagher & Wagner, 2024; Utami et al., 2023; Vicente-Yagüe Jara et al., 2023; Zhao et al., 2024).
- Designs curricula that teach students about AI detection methods and their implications for academic integrity (Cingillioglu, 2023; Elkhatat et al., 2023; Ibrahim, 2023; Levin et al., 2024; Liao et al., 2023; Popkov & Barrett, 2024).

- Creates learning activities that leverage AI tools to enhance the clarity, coherence, and impact of students' academic writing (Pellas, 2023; Utami et al., 2023; Zou & Huang, 2023).
- Designs assignments that require students to use AI tools for content enhancement while maintaining the integrity and originality of their work (Mohammadkarimi, 2023; Rojas, 2024; Watts et al., 2023).

Process of Teaching and Student Engagement:

- Guides students in using AI tools to improve the structure, clarity, and argumentation in their academic writing (Bašić et al., 2023; Malik et al., 2023; Mahapatra, 2024; Krajka & Olszak, 2024; Semrl et al., 2023; Song & Song, 2023).
- Facilitates peer review processes that incorporate AI-assisted evaluation to enhance the quality of student feedback (Rojas, 2024; Tang et al., 2023).
- Guides students in critically comparing the authenticity and creativity of AI-generated content with human-written work (Coman & Cardon, 2024; Mahyoob et al., 2023)
- Guides students in developing strategies to mitigate potential biases in AI-enhanced academic content (Abuyaman, 2023; Liao et al., 2023; Lozic & Štular, 2023).
- Engages students in discussions about the potential impact of AI on the dissemination and reception of academic content in their field (Mohammadkarimi, 2023; Marzuki et al., 2023; Wu & Dang, 2023).

Technology Integration:

- Implements AI-powered tools for content analysis to help students identify areas for improvement in their academic writing (Li et al., 2023; Lu, 2019;

Hwang et al., 2023).

- Utilizes AI-driven plagiarism detection tools to ensure the originality and integrity of academic content (Ibrahim, 2023; Mohammadkarimi, 2023; Yan, 2023).
- Utilizes machine learning algorithms to analyze and improve the quality and impact of academic content (Levin et al., 2024; Liao et al., 2023; Hayawi et al., 2023; Huang, 2023).
- Incorporates AI-based citation and referencing tools to improve the accuracy and completeness of academic works (Barrett & Pack, 2023; Lee et al., 2024; Zou & Huang, 2023).

Communication & Collaboration:

- Communicates with students about the importance of balancing AI assistance with human expertise in producing high-quality academic content (Elkhatat et al., 2023; Malik et al., 2023; Pellas, 2023; Rahayu et al., 2024).
- Collaborates with AI developers and researchers to improve the capabilities and accuracy of AI tools in academic content creation and enhancement (Papakonstantinidis et al., 2024; Tossell et al., 2024; Zou & Huang, 2023).
- Engages in interdisciplinary discussions about the impact of AI on scholarly communication and knowledge dissemination (Ivanović, 2023; Vicente-Yagüe Jara et al., 2023).

Assessment and Evaluation:

- Develops criteria for evaluating the quality and impact of AI-enhanced academic content in student work (Altmäe et al., 2023; Lozic & Štular, 2023).
- Creates assessment methods that measure students' ability to effectively use AI tools to improve their academic writing while maintaining originality (Liu

et al., 2024; Niloy et al., 2024; Yeadon et al., 2022).

- Implements peer and self-assessment strategies that incorporate AI-assisted evaluation to enhance students' critical analysis skills (Coman & Cardon, 2024; Mahyoob et al., 2023).
- Creates evaluation methods that can differentiate between AI-enhanced and purely human-generated academic content (Cheng et al., 2023; Leong, 2023; Liao et al., 2023; Hayawi et al., 2023; Zaitso & Jin, 2023).
- Develops assessment criteria that specifically evaluate the authenticity, creativity, and originality of student work in an AI-integrated academic environment (Amirjalili et al., 2024; Coman & Cardon, 2024; Draxler et al., 2023; Köbis & Mossink, 2020; Niloy et al., 2024; Tsao & Nogues, 2024).

Professional Growth and Development:

- Stays informed about emerging AI technologies that have the potential to enhance the quality and impact of academic content (Cardon et al., 2023; Mohammadkarimi, 2023).
- Participates in workshops and training sessions focused on best practices for integrating AI in academic writing and publishing (Alexander et al., 2023; Boudouaia et al., 2024; Gallagher & Wagner, 2024; Mohammadkarimi, 2023; Ivanović, 2023; Papakonstantinidis et al., 2024).
- Engages in research on the long-term effects of AI integration on the quality and impact of academic content in various disciplines (Rababah et al., 2024).
- Participates in training on advanced machine learning techniques relevant to academic content creation and evaluation (Levin et al., 2024; Liao et al., 2023; Hayawi et al., 2023; Huang, 2023).
- Stays informed about developments in AI detection technologies and their

implications for academic publishing and integrity (Hakam et al., 2024; Hamamah et al., 2023; Levin et al., 2024; Tang et al., 2023).

Alignment and Gaps. This guideline area addresses the challenges and opportunities presented by AI tools in creating, evaluating, and enhancing academic content. It ensures that educators can effectively leverage AI technologies to improve the quality of academic writing while maintaining the integrity and originality of student work.

The guidelines outlined in this area align well with several existing frameworks and standards, particularly the ISTE Standards for Educators (Crompton, 2023). The ISTE Standards for Educators, especially the Designer (Standard 5) and Analyst (Standard 7) roles, closely align with these guidelines. Standard 5.2 emphasizes designing authentic learning activities that recognize and accommodate learner variability, which resonates with the instructional design competencies focused on creating learning activities that leverage AI tools to enhance academic writing. Standard 7.1, which calls for providing alternative ways for students to demonstrate competency and reflect on their learning using technology, aligns with the assessment and evaluation competencies that incorporate AI-assisted evaluation methods.

The ISTE Standards for Educators provide additional relevant connections. Standard 3 (Citizen) emphasizes teaching students to critically examine online sources, which aligns with the guidelines focused on evaluating AI-generated content. Standard 4 (Collaborator) supports the communication and collaboration competencies, particularly in engaging with interdisciplinary discussions about AI's impact on scholarly communication. Standard 6 (Facilitator) resonates with the process of teaching and student engagement competencies, especially in guiding

students to use AI tools effectively while maintaining critical thinking skills.

The AI Competency Framework for Teachers (AI CFT) by UNESCO includes aspects that align closely with the guidelines outlined in this area. For instance, under the "AI Foundations & Applications" aspect, the framework emphasizes the progression from basic understanding to advanced application and creation of AI. This aligns with the competencies related to understanding AI capabilities in generating and enhancing academic content, and implementing AI-powered tools for content analysis. The "AI Pedagogy" aspect of the UNESCO framework is particularly relevant. It covers the integration of AI-assisted teaching methods and AI-enhanced pedagogical transformation, which resonates with the instructional design competencies in our framework. For example, the guideline for designing curricula that teach students to evaluate AI-generated content critically aligns with UNESCO's emphasis on AI-pedagogy integration. The "Human-centered Mindset" aspect of the UNESCO framework, which focuses on human agency, accountability, and social responsibility in the context of AI, connects well with the competencies related to maintaining academic integrity and originality in an AI-integrated environment. This is particularly relevant to the assessment and evaluation competencies that aim to differentiate between AI-enhanced and purely human-generated academic content.

Furthermore, the UNESCO framework's "AI for Professional Development" aspect aligns with our Professional Growth and Development competencies. Both emphasize the importance of continuous learning about AI advancements and their implications for educational practices. In the context of writing instruction, the UNESCO framework's emphasis on leveraging AI to improve teaching methods and educational outcomes is particularly relevant. This aligns with the guidelines focused

on using AI tools to enhance the clarity, coherence, and impact of students' academic writing.

These guidelines extend beyond existing frameworks by specifically addressing the unique challenges of AI in academic content creation. For instance, while the British Council CPD Framework and the TESOL Standards emphasize technology integration, the guidelines outlined here provide more detailed guidance on evaluating AI-generated content and using AI tools to enhance academic writing. The European Profiling Grid (EPG) and BALEAP Competency Framework, while not explicitly mentioning AI, have emphases on digital literacy and academic writing that these competencies build upon and extend into the realm of AI integration.

In comparing the statements in this guideline area on evaluating and enhancing the quality and impact of academic content with frameworks like the European Profiling Grid (EPG), the BALEAP Competency Framework for EAP Teachers, and the EdTech Masterplan 2030, several crucial gaps become evident. The EPG, for example, primarily focuses on traditional teaching competencies and digital media use, but it does not fully address the complexities introduced by AI in enhancing academic content's quality, originality, and impact. The statements shared in this part highlight the importance of understanding AI's role in generating academic content and its potential biases, yet the EPG does not cover these aspects, leaving a significant gap in preparing educators to critically evaluate AI-generated work.

Similarly, the BALEAP Competency Framework for EAP Teachers emphasizes language teaching and academic literacy but lacks detailed guidance on leveraging AI to enhance academic content. While it touches on text production and assessment, it does not address AI's role in improving the clarity, coherence, and impact of academic writing, which these statements highlight as critical competencies.

Furthermore, the need for educators to understand AI detection methods and their implications for academic integrity is not adequately covered in the BALEAP framework, reflecting a broader gap in addressing AI's influence on academic standards and practices.

The EdTech Masterplan 2030, while more forward-looking in integrating technology into education, still does not fully account for the specific competencies required to evaluate AI-enhanced academic content. The plan emphasizes the role of technology in teaching and learning but falls short in detailing how AI can be used to maintain and enhance content authenticity, creativity, and impact. These competency statements stress the need for educators to guide students in using AI responsibly while preserving the originality of their work, which is a critical area that the EdTech Masterplan does not sufficiently explore. This gap underscores the need for more robust frameworks that equip educators with the skills to navigate the challenges of AI integration in academic content creation and evaluation.

Policy Impact and Future Directions. The institutional policy directions of these guidelines are significant. Institutions may need to revise their academic integrity policies and create clear guidelines to address the use of AI in content creation and evaluation (Črček & Patekar, 2023; Levin et al., 2024; Matthews & Volpe, 2023; Mohammadkarimi, 2023; Zhao et al., 2024). Moreover, professional development programs should be implemented to ensure educators are proficient in using and teaching about AI tools for academic writing. These programs are seen as crucial for helping teachers understand, utilize, and teach about AI effectively and ethically. Zhao et al. (2024) recommend facilitating discussions to develop AI literacy. Matthews and Volpe (2023) and Ivanović (2023) suggest providing training on AI capabilities and limitations. Mohammadkarimi (2023) emphasize equipping

teachers with skills to identify and prevent AI-related academic dishonesty. Črček and Patekar (2023) and Nugroho et al. (2024) highlight the need for preparing educators to redesign assessments and curricula for the AI era. Coman and Cardon (2024) suggest that such programs should cover the ethical use of AI and the integration of AI tools into teaching practices. Ghafouri et al. (2024) and Boudouaia et al. (2024) stress strategies for guiding students in responsible AI use. Overall, there is a strong emphasis on the need for ongoing support and education for teachers to keep pace with rapidly evolving AI technologies and their implications for education.

Institutions may also need to invest in AI-powered content analysis and plagiarism detection tools. By addressing these guidelines and aligning them with established standards like ISTE, institutions can ensure that their educators are well-prepared to navigate the complex landscape of AI-enhanced academic content creation and evaluation, which could also balance the benefits of AI technology with the essential human elements of creativity, critical thinking, and academic integrity in EFL writing instruction.

Research directions regarding AI's performance, impact, and output characteristics in academic writing encompass a broad spectrum of areas. Several studies (Bašić et al., 2023; Mahapatra, 2024; Punar Özçelik & Yangın Ekşi, 2024) advocate for expanded research involving larger, more diverse participant groups and varied writing assessments to validate AI's impact on writing skills. Investigations into AI's capabilities and limitations in scientific writing are recommended (Abuyaman, 2023); Hwang et al., 2024; Jiang et al., 202), alongside explorations of stylometric analysis and linguistic features to differentiate AI-generated from human-written texts (Cheng et al., 2023; Leong, 2023; Liao et al., 2023; Hayawi et

al., 2023; Zaitso & Jin, 2023). The development and improvement of AI detection tools is a significant focus (Elkhatat et al., 2023; Levin et al., 2024; Perkins et al., 2023), with emphasis on their efficacy across various AI models. Continued research into AI's role in a scientific publication is suggested (Ariyaratne et al., 2023; Huespe et al., 2023), as well as studies on AI performance in different writing domains and audience perceptions of AI-generated content (Guerra et al., 2023; Köbis & Mossink, 2020; Hong et al., 2024). The relationship between reviewer experience and the ability to detect AI-generated content is another area of interest (Levin et al., 2024), along with AI's capabilities in error analysis (Algaraady & Mahyoob, 2023) and performance in specific writing aspects like cohesion (Zhou et al., 2023). Collectively, these future research directions underscore the need for comprehensive research into AI's capabilities, limitations, and impacts across various facets of academic writing. They also highlight the importance of developing reliable detection methods and investigating the authenticity, creativity, and originality of AI-generated academic content.

Guideline Area 4: User Behavior and Pedagogical Implications of AI-Assisted Writing

The relational core concepts that emerged under this core category are *Acceptance, Adoption, Attitudes and Challenges, Self-efficacy, Users and Uses, Trust, Motivation* and *Independence and Proficiency*.

The guideline statements derived from the studies are organized into the following categories:

Knowledge of:

- Demonstrates understanding of varied acceptance and adoption rates of AI

writing tools across different student populations and disciplines (Lee et al., 2024; Papakonstantinidis et al., 2024; Rad et al., 2023; Zou & Huang, 2023)

- Articulates the complex relationship between users' perceptions, intentions, and actual usage of AI writing tools (Nugroho et al., 2024; Papakonstantinidis et al., 2024; Rababah et al., 2024; Zou & Huang, 2023).
- Recognizes the potential impacts of AI-assisted writing on students' cognitive processes, writing strategies, and self-efficacy (Liu et al., 2024; Nazari et al., 2021; Nguyen et al., 2024).

Instructional Design:

- Develops curricula that address the diverse ways students interact with and utilize AI writing tools (Cardon et al., 2023; Gustilo et al., 2024; Hwang et al., 2023; Liu et al., 2024).
- Creates learning activities that promote responsible AI use while fostering students' independent critical thinking and writing skills (Nguyen et al., 2024; Zou & Huang, 2023).
- Designs flexible courses that allow for personalized learning experiences with AI tools while maintaining instructional oversight (Leong, 2023; Yang et al., 2024; Tossell et al., 2024).

Process of Teaching and Student Engagement:

- Adapts teaching strategies to address varying levels of AI familiarity, acceptance, and reliance among students (Boudouaia et al., 2024; Lu et al., 2024; Song & Song, 2023).
- Facilitates discussions about the implications of AI use in students' future academic and professional writing contexts (Marzuki et al., 2023; Mohammadkarimi, 2023; Wu & Dang, 2023).

- Guides students in developing metacognitive strategies for effectively managing their interaction with AI writing tools (Nguyen et al., 2024).

Technology Integration:

- Implements AI writing tools incrementally, with clear pedagogical objectives and ongoing evaluation of their impact on student learning behaviors (Boudouaia et al., 2024; Rojas, 2024; Watts et al., 2023).
- Utilizes learning analytics to track and understand patterns of AI tool usage among students, informing instructional decisions (Lu et al., 2024; Zou & Huang, 2023).
- Integrates AI tools in ways that complement rather than replace traditional writing instruction methods (Watts et al., 2023).

Communication & Collaboration:

- Engages with students to understand and address their perceptions, concerns, and experiences with AI-assisted writing (Dong, 2024; Lee et al., 2024; Nugroho et al., 2024; Tossell et al., 2024; Zou & Huang, 2023).
- Collaborates with colleagues to share insights and develop best practices for managing student behaviors around AI tool use (Črček & Patekar, 2023; Gallagher & Wagner, 2024; Ivanović, 2023; Matthews & Volpe, 2023; Mohammadkarimi, 2023; Nguyen et al., 2024).
- Communicates clear expectations and guidelines for AI use in writing assignments, promoting transparency and ethical use (Mohammadkarimi, 2023; Rojas, 2024; Tang et al., 2023),

Assessment and Evaluation:

- Develops assessment methods that consider the impact of AI use on students' writing processes and outputs (Liu et al., 2024; Parker et al., 2023; Steiss et

al., 2024).

- Creates evaluation criteria that value both effective AI tool use and independent writing skills (Rojas, 2024; Wang, 2020).
- Implements strategies to monitor and evaluate changes in student writing behavior resulting from AI tool integration (Dong, 2024).

Professional Growth and Development:

- Stays informed about emerging research on the psychological and cognitive impacts of AI-assisted writing on learners' content (Cardon et al., 2023; Liu et al., 2024; Nazari et al., 2021; Nguyen et al., 2024; Mohammadkarimi, 2023).
- Participates in professional development opportunities focused on adaptive teaching strategies for AI-integrated writing environments (Ivanović, 2023).
- Engages in action research to better understand and respond to student behaviors and learning outcomes in AI-assisted writing contexts (Wolf & Wolf, 2023; Zhou et al., 2023).

Alignment and Gaps. The guideline area of *User Behavior and Pedagogical Implications of AI-assisted writing* is crucial for AI integration in EFL writing instruction. This area is related to the complex interactions between students and AI writing tools by focusing on acceptance, adoption, attitudes, and the impact on learning behaviors. With these competency improvements, educators can learn how to navigate the challenges and opportunities presented by AI-assisted writing while maintaining pedagogical effectiveness and fostering students' independent writing skills.

This guideline area aligns well with several aspects of the ISTE Standards for Educators. For instance, ISTE Standard 5 (Designer) emphasizes designing authentic

learning experiences that leverage technology. This resonates with the instructional design competencies in our framework, particularly in creating learning activities that promote responsible AI use while fostering critical thinking skills. Also, ISTE Standard 6 (Facilitator) aligns with our process of teaching and student engagement competencies, especially in adapting teaching strategies to address varying levels of AI familiarity among students.

The UNESCO Draft AI Competency Frameworks for Teachers and School Students also provide relevant connections. The "Human-centered Mindset" aspect of the UNESCO framework aligns with the guidelines outlined related to understanding the complex relationship between users' perceptions and actual usage of AI writing tools. The "AI Pedagogy" aspect resonates with our instructional design competencies, particularly in developing curricula that address diverse ways students interact with AI tools.

While frameworks like the British Council CPD and TESOL Standards emphasize technology integration, the guidelines in this study provide more detailed guidance on managing student behaviors and learning outcomes in AI-assisted writing contexts. Also, the European Profiling Grid (EPG) and BALEAP Competency Framework, while not explicitly mentioning AI, have emphases on learner autonomy and digital literacy that the competencies suggested build upon and extend into the realm of AI integration. Overall, although there are some parts that are in alignment with some parts of the competency frameworks, the guidelines shared in this study extend beyond existing frameworks by specifically addressing the unique challenges of AI in writing instruction.

When comparing these guideline statements on user behavior and the pedagogical implications of AI-assisted writing with other frameworks, such as the

UNESCO Draft AI Competency Frameworks, the Cambridge English Teaching Framework, and the Eaquals Framework for Language Teacher Training & Development, several notable gaps become apparent. The UNESCO Draft AI Competency Frameworks, while progressive in their approach to AI literacy, do not specifically address the varied acceptance and adoption rates of AI tools across different student populations, nor do they delve into the psychological and cognitive impacts of AI-assisted writing on self-efficacy and writing strategies. These statements emphasize these areas as critical for understanding how students interact with AI tools, which are not represented in the UNESCO framework.

Similarly, the Cambridge English Teaching Framework provides a solid foundation for integrating digital resources into teaching but lacks the depth needed to address the nuanced relationship between user perceptions and actual usage of AI tools, as highlighted in these guideline statements. The framework also does not explore the importance of designing curricula that cater to the diverse ways students utilize AI, a key aspect of these instructional design competencies. This omission suggests a gap in preparing educators to navigate the complex dynamics of AI tool usage in modern writing environments.

The Eaquals Framework for Language Teacher Training & Development, while comprehensive in its coverage of teaching competencies, does not explicitly focus on the pedagogical implications of AI tools in writing instruction. These guideline statements highlight the need for flexible courses and personalized learning experiences with AI tools, areas where the Eaquals framework falls short. Additionally, the emphasis on guiding students in developing metacognitive strategies for interacting with AI tools, which these statements underscore, is not sufficiently covered in the Eaquals framework. This gap points to a broader need for

frameworks to evolve in response to the growing influence of AI on student behavior, motivation, and proficiency in writing.

These comparisons reveal that while existing frameworks provide a strong foundation in traditional teaching competencies, they often lack the specific guidance needed to address the pedagogical challenges and opportunities presented by AI-assisted writing. This shows the importance of developing more targeted guidelines that reflect the evolving landscape of AI in education.

Policy Impact and Future Directions. The institutional policy directions for *User Behavior and Pedagogical Implications of AI-assisted writing* are multidimensional and require careful consideration. Several articles address institutional policy directions related to user behavior and pedagogical aspects of AI-assisted writing. Zou and Huang (2023) highlight the need for policies that consider user acceptance and adoption of AI tools in academic writing. Additionally, Rahayu et al. (2024) suggest institutional guidelines should address motivation and independence when integrating AI-based writing tools. Nguyen et al. (2024) imply that policies should consider human-AI collaboration patterns and their impact on user proficiency. Also, Lee et al. (2024) and Krajka and Olszak (2024) indicate the importance of policies that account for students' attitudes and challenges in using AI writing tools. Furthermore, Nugroho et al. (2024) suggest that institutions should develop policies that address students' evolving perceptions and use patterns of AI tools. Moreover, Tossell et al. (2024) highlight the need for policies that consider trust issues in AI-assisted writing assignments. Also, Coman and Cardon (2024) imply that policies should address perceptions of professionalism in AI-assisted writing contexts. Lastly, Ghafouri et al. (2024) suggest that institutions should develop policies that support teachers' self-efficacy in using AI writing tools.

Overall, these studies emphasize the need for institutional policies that are responsive to the complex interplay of user behavior, attitudes, and pedagogical implications in AI-assisted writing, ensuring that policies evolve alongside users' changing relationships with these technologies.

Research directions concerning user behavior and pedagogical aspects of AI-assisted writing highlight the need for comprehensive investigations across various dimensions. Two studies suggest exploring the acceptance and adoption patterns of AI writing tools among different user groups, and also longitudinal research to track attitude changes over time (Papakonstantinidis et al., 2024; Zou & Huang, 2023). The impact of these tools on motivation, independence, and human-AI collaboration in academic writing warrants further examination, according to Nguyen et al., 2024 and Rahayu et al., 2024. In addition to these areas, researchers recommend investigating students' perceptions, use patterns, and challenges in utilizing AI writing tools (Krajka & Olszak, 2024; Lee et al., 2024; Nugroho et al., 2024), as well as their trust in AI-generated content and perceptions of professionalism in AI-assisted writing (Coman & Cardon, 2024; Tossell et al., 2024). The effects of AI-assisted writing on teachers' self-efficacy and students' writing skills are areas of interest (Ghafouri et al., 2024; Song & Song, 2023), along with the long-term impacts on writing proficiency and learning behaviors (Mahmud, 2023; Yan, 2023). Studies also emphasize the need to explore how AI tools influence students' creative writing abilities and writing feedback literacy (Niloy et al., 2024; Rad et al., 2023). In brief, these future research directions underscore the importance of diverse research approaches to gain a deeper understanding of AI's role in academic writing, its impact on users, and its pedagogical implications.

Guideline Area 5: AI-Integrated Writing Instruction and Outcomes

The relational core concepts that emerged within this core category include *Writing Skills*, *Writing Performances*, *Writing Process*, *AI-enhanced classroom*, *Writing Quality*, *Scaffolding*, and *AI-assisted Writing*.

The guideline statements resulting from the studies are classified into the following categories:

Knowledge of:

- Understands the potential of AI tools to enhance various aspects of writing instruction, including grammar, style, and content development (Ariyaratne et al., 2023; Coman & Cardon, 2024; Kaharuddin et al., 2024; Pellas, 2023).
- Recognizes the impact of AI-integrated writing instruction on students' learning outcomes across different proficiency levels and writing genres (Matthews & Volpe, 2023; Song & Song, 2023).
- Demonstrates awareness of the challenges and opportunities in balancing AI-assisted instruction with traditional writing pedagogy (Elkhatat et al., 2023; Malik et al., 2023; Pellas, 2023; Rahayu et al., 2024).

Instructional Design:

- Develops AI-supported writing curricula that systematically integrate AI tools to address specific learning objectives (Liu et al., 2021; Song & Song, 2023; Yang et al., 2024).
- Creates scaffolded learning experiences that gradually introduce AI tools while fostering students' independent writing skills (Liu et al., 2021; Hsiao & Chang, 2023; Kim & Kim, 2022).
- Designs differentiated instruction strategies that leverage AI to meet diverse student needs and learning styles in writing (Lu et al., 2024; Ivanović, 2023; Song & Song, 2023).

Process of Teaching and Student Engagement:

- Implements dialogic teaching approaches that promote student engagement with both AI tools and human instruction in the writing process s (Liu et al., 2024; Parker et al., 2023; Steiss et al., 2024).
- Guides students in developing metacognitive strategies for effectively utilizing AI assistance in their writing development (Nguyen et al., 2024).
- Facilitates collaborative writing projects that incorporate AI tools to enhance peer learning and feedback processes (Ivanović, 2023; Nguyen et al., 2024; Zou & Huang, 2023).

Technology Integration:

- Selects and implements appropriate AI writing tools based on specific instructional goals and student needs (Liu et al., 2024; Ivanović, 2023).
- Integrates AI-generated writing samples and prompts to stimulate student creativity and critical thinking in writing tasks (Niloy et al., 2024; Parker et al., 2023; Tsao & Nogues, 2024; Vicente-Yagüe Jara et al., 2023).
- Utilizes AI-powered writing assistants to provide real-time, personalized support during in-class writing activities (Cardon et al., 2023; Fyfe, 2022).

Communication & Collaboration:

- Engages in professional learning communities to share best practices and challenges in AI-integrated writing instruction (Cardon et al., 2023)
- Collaborates with educational technology specialists to optimize the implementation of AI writing tools in the curriculum (Papakonstantinidis et al., 2024; Tossell et al., 2024; Zou & Huang, 2023).
- Communicates effectively with students about the role of AI in their writing development, setting clear expectations for its use (Dong, 2024; Nugroho et

al., 2024).

Assessment and Evaluation:

- Develops assessment strategies that measure the effectiveness of AI-integrated writing instruction on student learning outcomes (Wolf & Wolf, 2023).
- Creates rubrics that evaluate students' ability to effectively utilize AI tools in their writing process while maintaining originality and critical thinking (Alexander et al., 2023; Niloy et al., 2024; Yeadon et al., 2022).
- Implements formative assessment techniques that leverage AI to provide timely and targeted feedback on student writing (Alexander et al., 2023; Niloy et al., 2024; Yeadon et al., 2022).

Professional Growth and Development:

- Stays informed about emerging AI technologies and their potential applications in writing instruction (Cardon et al., 2023; Mohammadkarimi, 2023).
- Participates in professional development opportunities focused on innovative pedagogies for AI-integrated writing instruction (Alexander et al., 2023; Gustilo et al., 2024; Mohammadkarimi, 2023).
- Conducts action research to evaluate the impact of AI-integrated writing instruction on student outcomes in specific contexts (Wolf & Wolf, 2023; Zhou et al., 2023).

Alignment and Gaps. The guideline area of AI-Integrated Writing Instruction and Outcomes is crucial for the effective integration of AI in EFL writing instruction. This area addresses the complex interplay between AI tools and traditional writing

pedagogy, focusing on enhancing writing skills, improving writing performance, and optimizing the writing process through AI-enhanced classroom environments. These guidelines ensure that educators can effectively leverage AI technologies to improve writing instruction while maintaining pedagogical effectiveness and fostering students' independent writing skills.

This guideline area aligns well with several aspects of the ISTE Standards for Educators. ISTE Standard 5 (Designer) emphasizes designing authentic learning experiences that leverage technology, which resonates with our instructional design guidelines, particularly in developing AI-supported writing curricula that systematically integrate AI tools to address specific learning objectives. ISTE Standard 6 (Facilitator) aligns with the process of teaching and student engagement guidelines, especially in implementing dialogic teaching approaches that promote student engagement with both AI tools and human instruction in the writing process.

The UNESCO Draft AI Competency Frameworks for Teachers and School Students also provides relevant connections. The "AI Pedagogy" aspect of the UNESCO framework aligns with the guidelines related to integrating AI-generated writing samples and prompts to stimulate student creativity and critical thinking in writing tasks. The "AI Foundations & Applications" aspect resonates with our technology integration competencies, particularly in selecting and implementing appropriate AI writing tools based on specific instructional goals and student needs.

These guidelines, however, extend beyond existing frameworks by specifically addressing the unique challenges of AI in writing instruction. While frameworks like the British Council CPD and TESOL Standards emphasize technology integration, the guidelines in this study provide more detailed directions on leveraging AI to enhance specific aspects of writing instruction, such as grammar,

style, and content development. The European Profiling Grid (EPG) and BALEAP Competency Framework, while not explicitly mentioning AI, have emphases on innovative teaching methods and digital literacy that these guidelines build upon and extend into the realm of AI integration in writing instruction.

When the guideline statements on AI-integrated writing instruction and outcomes are compared with frameworks like the European Profiling Grid (EPG), the BALEAP Competency Framework for EAP Teachers, and the UNESCO Draft AI Competency Frameworks, it becomes clear that there are several critical gaps. The EPG, for instance, focuses heavily on traditional teaching methodologies and digital media use but does not sufficiently address the integration of AI in writing instruction. The statements in this part emphasize the need for a balanced approach that leverages AI while maintaining traditional pedagogy, a complex dynamic that the EPG does not explore in depth. This omission leaves a gap in preparing educators to effectively blend AI tools with established instructional practices to enhance writing skills and outcomes.

Similarly, the BALEAP Competency Framework for EAP Teachers is comprehensive in addressing academic writing and language skills but lacks explicit guidance on the use of AI to enhance writing quality and performance. While the framework covers text production and scaffolding, it does not delve into how AI can be systematically integrated into writing curricula to support specific learning objectives. These guideline statements highlight the importance of scaffolded learning experiences that gradually introduce AI tools, a crucial aspect that is missing from the BALEAP framework. This gap indicates a need for more targeted guidelines that address the unique challenges and opportunities presented by AI in writing instruction.

The UNESCO Draft AI Competency Frameworks, although forward-thinking in promoting AI literacy, do not fully capture the intricacies of AI-integrated writing instruction. These statements point to the need for professional development that equips educators with the skills to select and implement appropriate AI tools based on instructional goals. This is a consideration that the UNESCO frameworks do not sufficiently address. Moreover, the emphasis on creating assessment strategies that measure the effectiveness of AI integration in writing instruction in this guideline area underscores a significant gap in existing frameworks, which often overlook the need for specialized rubrics and formative assessment techniques that account for AI's role in the writing process.

These comparisons reveal that while current frameworks provide a solid foundation in traditional and digital competencies, they fall short of addressing the specific needs of AI-integrated writing instruction. This underscores the importance of developing more competencies that reflect the evolving landscape of AI in education, ensuring that educators are fully equipped to harness the potential of AI to enhance writing instruction and outcomes.

Policy Impact and Future Directions. Several studies suggest important institutional policy directions for AI-assisted writing instruction and outcomes. To begin with, Punar Özçelik and Yangın Ekşi (2024) and Mahapatra (2024) highlight the need for policies that leverage AI tools to enhance students' writing skills and performance. Building on this, Hwang et al. (2023) and Zhao et al. (2022) indicate that institutions should consider policies for integrating AI-powered automated scoring systems to support writing assessments. Furthermore, Huang (2023) and Lu et al. (2024) suggest policies for incorporating AI feedback mechanisms to improve writing quality and scaffolding. In a similar way, Song and Song (2023) and Wolf

and Wolf (2023) emphasize the importance of policies and guidance that could create AI-enhanced writing classrooms. While Rojas (2024) focuses specifically on the need for AI-based classroom activities throughout the writing process, from brainstorming to revision, Algaraady and Mahyoob (2023) complement this by highlighting the positive impact of AI tools on improving student writing performance. Additionally, Guo and Wang (2023) suggest policies should address how to combine human and AI feedback in writing instruction effectively. Lastly, Altmäe et al. (2023) and Wale and Kassahun (2024) highlight the potential for policies that encourage the use of integrated AI writing technologies to enhance overall writing instruction. Taken together, these studies underscore the multifaceted nature of AI's role in writing education and the need for comprehensive, thoughtful policy approaches by educators, policymakers, and institutions.

Research directions for AI in writing instruction and outcomes cover a wide range of areas. Studies by Mahapatra, 2024 and Punar Özçelik and Yangın Ekşi, 2024 suggest investigating AI's impact on specific writing genres, micro-aspects of writing, and its relationship with language proficiency. There is also a call for exploring AI's role in personalized writing support and improving AI-assisted automated scoring methods (Hwang et al., 2023; Zhao et al., 2022). Additionally, researchers recommend examining the integration of AI feedback mechanisms to enhance writing quality and scaffolding (Huang, 2023; Lu et al., 2024), as well as investigating the long-term effects of AI-assisted learning on writing skills and motivation (Song & Song, 2023; Wolf & Wolf, 2023). The potential of AI in STEM writing assignments and its capabilities in error analysis are areas of interest (Algaraady & Mahyoob, 2023; Rojas, 2024). Studies also emphasize the need to explore effective combinations of human and AI feedback in writing instruction

(Fyfe, 2022; Guo & Wang, 2023) and the potential of integrated AI writing technologies to enhance overall writing instruction (Altmäe et al., 2023; Wale & Kassahun, 2024). In the context of EFL writing, research is needed on the impact of AI-powered tools on various aspects, including reading comprehension and feedback literacy (Hsiao & Chang, 2023; Lu, 2019; Wang, 2020). Furthermore, investigations into how AI affects the research writing process and comparisons of human and AI feedback quality are recommended (Pigg, 2024; Steiss et al., 2024). In summary, these future research directions collectively highlight the need for comprehensive research to understand and optimize AI's role in enhancing writing skills, processes, and instruction across various educational contexts.

Guideline Area 6: Diverse Writing Genres and Modalities

The relational core concepts identified within this core category are *Creative Writing, Business Communication, Multimodal Writing, News Writing, Narrative English Writing, Graphic Storytelling, Digital Stories, Picture-cued Writing, and Multiple Genres (Essays, Stories, Poetry, and Python code)*.

The guideline statements identified in the studies are grouped into the following categories:

Knowledge of:

- Demonstrates understanding of how AI tools can be applied to various writing genres (e.g., academic, creative, professional, technical) and modalities (e.g., text, multimedia, digital storytelling) (Hayawi et al. (2023;
- Recognizes the unique challenges and opportunities AI presents in different writing contexts and for various audience types (Cardon et al., 2023; Papakonstantinidis et al., 2024).
- Understands the potential of AI to enhance or limit creativity and originality

across different writing genres (Niloy et al., 2024; Tsao & Nogues, 2024 Vicente-Yagüe Jara et al., 2023).

Instructional Design:

- Develops genre-specific writing assignments that leverage AI tools while maintaining the integrity and conventions of each genre.
- Creates multimodal writing projects that integrate AI-assisted text generation with other media forms (e.g., images, audio, video).
- Designs cross-genre writing activities that use AI to help students understand and apply different genre conventions and styles.

Process of Teaching and Student Engagement:

- Guides students in adapting their use of AI tools for different writing genres and audience expectations (Bašić et al., 2023; Malik et al., 2023; Mahapatra, 2024; Krajka & Olszak, 2024; Semrl et al., 2023; Song & Song, 2023).
- Facilitates workshops on using AI to enhance creativity in various forms of writing, from poetry to technical documentation (Boudouaia et al., 2024; Rababah et al., 2024; Yeadon et al., 2022).
- Engages students in critical discussions about the impact of AI on evolving genre conventions and new forms of digital writing (Lozic & Štular, 2023; Semrl et al., 2023; Watts et al., 2023).

Technology Integration:

- Implements AI tools that support genre-specific writing tasks, such as narrative generation for creative writing or data interpretation for technical writing (Pellas, 2023; Zhou et al., 2023).
- Utilizes AI-powered language models to demonstrate style variations across different genres and modalities (Hayawi et al., 2023).

- Integrates AI-assisted tools for multimodal composition, helping students combine text with other media elements effectively (Liu et al., 2024).

Communication & Collaboration:

- Collaborates with subject matter experts to ensure appropriate AI integration in discipline-specific writing genres (Levin et al., 2024).
- Engages with professional writers and industry representatives to understand AI's role in various professional writing contexts (Cardon et al., 2023).
- Facilitates peer collaboration in AI-assisted writing projects across different genres and modalities (Rojas, 2024; Tang et al., 2023).

Assessment and Evaluation:

- Develops assessment criteria that consider both AI assistance and genre-specific requirements in evaluating student writing (Matthews & Volpe, 2023; Song & Song, 2023).
- Creates rubrics that assess students' ability to effectively use AI tools while maintaining genre integrity and audience awareness (Alexander et al., 2023; Niloy et al., 2024; Yeadon et al., 2022).
- Implements portfolio assessments that showcase students' AI-assisted writing across multiple genres and modalities (Liu et al., 2024).

Professional Growth and Development:

- Stays informed about emerging AI applications in various writing genres and new digital writing modalities (Cardon et al., 2023; Mohammadkarimi, 2023).
- Participates in workshops and training sessions on AI's impact on specific writing genres and multimodal composition (Alexander et al., 2023; Boudouaia et al., 2024; Gallagher & Wagner, 2024; Mohammadkarimi, 2023;

Ivanović, 2023; Papakonstantinidis et al., 2024).

- Engages in action research to evaluate the effectiveness of AI integration in teaching diverse writing genres and modalities (Zou & Huang, 2023; Pigg, 2024).

Alignment and Gaps. The guideline area of Diverse Writing Genres and Modalities addresses the unique challenges and opportunities presented by AI in various writing contexts related to various genres. This area recognizes the expanding scope of writing in the digital age and AI's potential to transform traditional genre boundaries. It aligns closely with the evolving nature of 21st-century literacy, as reflected in the ISTE Standards for Educators, particularly Standard 6 (Creative Communicator), which emphasizes the importance of students communicating clearly and creatively across various platforms and formats. The UNESCO Draft AI Competency Frameworks for Teachers and School Students also provides relevant connections, with its emphasis on AI-enhanced pedagogical transformation resonating with competencies related to implementing AI tools for genre-specific writing tasks.

There are several important gaps that emerge when the competency statements in this guideline area are compared to the existing competency frameworks. For example, The EPG, although addresses traditional language teaching competencies, it does not delve into the complexities of integrating AI into different writing genres, such as academic, creative, professional, and technical writing. These statements, however, emphasize the importance of understanding how AI can enhance or limit creativity across these genres, which is absent in the EPG. This gap suggests a need for more specialized competencies that address the unique challenges AI presents in diverse writing contexts and audience types.

The Cambridge English Teaching Framework similarly provides a strong foundation for language teaching but lacks explicit guidance on using AI tools for genre-specific writing tasks. These guideline statements highlight the value of developing genre-specific assignments and cross-genre activities that leverage AI while maintaining genre integrity. Additionally, the framework does not address the integration of AI in multimodal composition, which is increasingly relevant in today's digital writing landscape. This omission indicates a broader need for frameworks to evolve in order to address the growing role of AI in supporting multimodal and genre-specific writing instruction.

The UNESCO Draft AI Competency Frameworks, while progressive in promoting AI literacy, do not fully address the application of AI to various writing genres and modalities. These statements point to the need for educators to implement AI tools that support genre-specific tasks and to develop rubrics that assess students' ability to use AI effectively while maintaining genre conventions and audience awareness. The lack of detailed guidance on these aspects in the UNESCO frameworks highlights a significant gap in preparing educators to navigate the evolving landscape of AI-assisted writing. Furthermore, the emphasis on professional growth and development, particularly in staying informed about emerging AI applications in different writing genres, highlights the need for continuous learning opportunities that are not sufficiently covered in current frameworks.

In short, these comparisons reveal that while existing frameworks provide a solid basis for traditional and digital competencies, they often lack the specificity required to address the nuanced application of AI across various writing genres and modalities.

Policy Impact and Future Directions. From an institutional policy perspective, this guideline area may necessitate a reevaluation of writing curricula across disciplines. The integration of AI in diverse writing genres presents challenges and opportunities for institutions, educators, and policymakers. To begin with, in the realm of creative writing, Köbis and Mossink (2020) suggest that institutions should develop guidelines for ethical AI use that preserve human creativity while leveraging AI's potential. Similarly, for business communication, Coman and Cardon (2024) emphasize the need for policies that address the use of AI tools in professional writing contexts. Moving to multimodal writing, Liu et al. (2024) highlight the importance of institutional support for incorporating AI in digital storytelling and graphic narratives. Furthermore, Hwang et al. (2023) suggest that educators should develop strategies for using AI to enhance narrative English writing instruction. Additionally, Zhao et al. (2022) propose that policymakers should support the development of AI-assisted assessment tools for picture-cued writing tasks. Lastly, regarding multiple genres, Köbis and Mossink (2020) and Levin et al. (2024) collectively imply that institutions should establish flexible policies that address AI use across various writing forms, including essays, stories, poetry, and even code writing. Overall, these studies underscore the need for a comprehensive and nuanced approach to AI integration in writing education, one that respects the unique characteristics of each genre while harnessing the potential of AI to enhance creativity, efficiency, and learning outcomes across diverse writing genres.

Research directions for AI in specialized writing genres encompass a diverse range of areas. Studies suggest exploring AI's role in digital multimodal composing and its impact on students' cognitive processes (Liu et al., 2024), as well as examining AI's capabilities in generating longer creative texts and poems (Köbis &

Mossink, 2020). Further research on audience perceptions of AI-generated news content is recommended (Hong et al., 2024), alongside investigations into AI's impact on business communication writing tasks and emerging notions of professionalism in this context (Coman & Cardon, 2024). The long-term effects of AI on narrative intelligence and digital storytelling need further exploration (Pellas, 2023), as does AI's potential in scientific report writing and analytical tasks in STEM fields (E15). Improving AI-assisted methods for specialized tasks like picture-cued writing through larger datasets is also suggested by Zhao et al., 2022. Collectively, these directions highlight the need for genre-specific research to comprehensively understand AI's potential and limitations across diverse writing contexts, from creative and narrative writing to business communication and scientific reporting. Briefly, such research could provide valuable insights into how AI technologies can be effectively integrated into various specialized writing processes while maintaining the unique characteristics and requirements of each genre.

Guideline Area 7: Linguistic and Discourse Features of AI-Generated Texts

The relational core concepts associated with this core category are *Textual and Linguistic Features, Discourse Components, Style, and Register*.

The guideline statements generated from the studies are sorted into the following categories:

Knowledge of:

- Demonstrates deep understanding of the linguistic characteristics typical of AI-generated texts, including patterns in syntax, lexical choice, and discourse structure (Cheng et al., 2023; Perkins et al., 2023).
- Recognizes the strengths and limitations of AI in producing various linguistic features, such as cohesion, coherence, and register appropriateness (Hakam et

al., 2024; Zhou et al., 2023).

- Understands how different AI models and prompts can influence the linguistic and discourse features of generated texts (Niloy et al., 2024; Parker et al., 2023; Tsao & Nogues, 2024; Vicente-Yagüe Jara et al., 2023).

Instructional Design:

- Develops curriculum units that teach students to identify and analyze linguistic features specific to AI-generated texts (Herbold et al., 2023; Krajka & Olszak, 2024; Liao et al., 2023).
- Creates comparative analysis activities that highlight differences between AI-generated and human-written texts across various linguistic dimensions (Amirjalili et al., 2024; Herbold et al., 2023; Krajka & Olszak, 2024; Mahyooob et al., 2023; Vicente-Yagüe Jara et al., 2023; Watts et al., 2023; Zhou et al., 2023).
- Designs exercises that challenge students to improve upon AI-generated texts by enhancing specific linguistic and discourse features (Liu et al., 2024).

Process of Teaching and Student Engagement:

- Guides students in critical analysis of AI-generated texts, focusing on linguistic patterns, discourse structure, and stylistic elements (Bašić et al., 2023; Malik et al., 2023; Mahapatra, 2024; Krajka & Olszak, 2024; Semrl et al., 2023; Song & Song, 2023).
- Facilitates discussions on the implications of AI-specific linguistic features for various writing contexts and purposes (Lozic & Štular, 2023; Semrl et al., 2023; Watts et al., 2023).

Technology Integration:

- Utilizes text analysis tools to help students visualize and quantify linguistic

features of AI-generated texts (Cheng et al., 2023; Hayawi et al., 2023; Levin et al., 2024).

- Implements AI writing assistants in classroom activities to demonstrate real-time generation of linguistic features (Bašić et al., 2023; Liu et al., 2024).
- Integrates corpus linguistics tools to compare large-scale patterns between AI-generated and human-written texts (Alafnan & Mohdzuki, 2023; Cingillioglu, 2023; Hayawi et al., 2023; Zaitsu & Jin, 2023).

Communication & Collaboration:

- Collaborates with linguists and NLP experts to stay updated on research regarding the linguistic properties of AI-generated text (Gustilo et al., 2024; Zhao et al. (2022)
- Engages in interdisciplinary discussions on the impact of AI-generated text features on various fields (e.g., journalism, literature, academic publishing) (Ivanović, 2023; Vicente-Yagüe Jara et al., 2023).
- Communicates findings and insights about AI text features to students, colleagues, and the broader academic community (Herbold et al., 2023)

Assessment and Evaluation:

- Develops assessment methods that evaluate students' ability to identify and analyze linguistic features specific to AI-generated texts (Herbold et al., 2023; Krajka & Olszak, 2024; Liao et al., 2023).
- Implements projects that require students to compare and contrast the linguistic features of texts generated by different AI models (Cheng et al., 2023; Leong, 2023; Liao et al., 2023; Hayawi et al., 2023; Zaitsu & Jin, 2023).

Professional Growth and Development:

- Stays informed about advancements in AI language models and their implications for text generation features (Cardon et al., 2023; Gustilo et al., 2024; Utami et al., 2023).
- Participates in workshops and conferences focused on the linguistic analysis of AI-generated text (Cardon et al., 2023; Ivanović, 2023).
- Conducts or engages with research on the evolving linguistic characteristics of AI-generated text across different domains and genres (Levin et al., 2024; Lozic & Štular, 2023; Herbold et al., 2023; Liao et al., 2023; Krajka & Olszak, 2024)

Alignment and Gaps. The guideline area of *Linguistic and Discourse Features of AI-Generated Texts* addresses a critical aspect of AI integration in EFL writing instruction. This area focuses on the unique linguistic characteristics of AI-generated content. Thus, educators need to develop a deep understanding of how AI models produce text and how these texts differ from human-written compositions due to the growing need for critical digital literacy in the age of AI, and these guideline statements extend beyond basic technology use to include sophisticated and comprehensive analysis of AI-generated language.

This guideline area reflects several aspects of the ISTE Standards for Educators, particularly Standard 3 (Citizen), which emphasizes critical examination of online sources. In the context of AI-generated texts, this standard takes on new significance, requiring educators to guide students in discerning and analyzing the linguistic features that may indicate AI authorship. The UNESCO Draft AI Competency Frameworks also provide relevant connections, especially in their emphasis on AI literacy and critical thinking skills, which are crucial for understanding and evaluating AI-generated language.

The comparison of the guideline statements on the linguistic and discourse features of AI-generated texts with frameworks like the European Profiling Grid (EPG), the BALEAP Competency Framework for EAP Teachers, and the UNESCO Draft AI Competency Frameworks reveals several gaps. The EPG, for example, offers a broad overview of language teaching competencies but does not delve into the specific linguistic characteristics of AI-generated texts. These statements highlight the importance of understanding how AI impacts syntax, lexical choices, and discourse structures. This omission indicates a need for more specialized competencies that equip educators to critically analyze and teach the nuances of AI-generated language.

Similarly, the BALEAP Competency Framework for EAP Teachers focuses on developing academic writing skills but does not explicitly consider the unique challenges AI presents in producing linguistic features like cohesion, coherence, and register appropriateness. These guideline statements emphasize the need for educators to guide students in comparing AI-generated texts with human-written ones, a critical skill that many existing frameworks overlook. This gap points to the necessity of updating such frameworks to include competencies that reflect the realities of AI's influence on discourse and linguistic features.

The UNESCO Draft AI Competency Frameworks, while more adjusted to AI's role in education, still do not provide detailed guidance on analyzing the linguistic and discourse features of AI-generated texts. These competencies, however, focus on developing curriculum units that teach students to identify and improve upon AI-generated linguistic elements. This is an area where the UNESCO framework is less explicit. Additionally, in these guideline statements, there is an emphasis on using AI and text analysis tools in classroom activities to enhance

linguistic understanding. The UNESCO frameworks touch on this aspect only broadly.

These comparisons reveal that existing frameworks offer a foundation in traditional and digital literacy, but they lack the specific guidance needed to address the linguistic and discourse features of AI-generated texts.

Policy Impact and Future Directions. The analysis of linguistic and discourse features in AI-generated texts presents significant directions for educational institutions, policymakers, and educators. Firstly, regarding textual and linguistic features, Zaitse and Jin (2023) suggest that institutions should develop policies for identifying and evaluating AI-generated content based on specific linguistic markers. Building on this, Levin et al. (2024) recommend that educators be trained to recognize the unique stylistic elements of AI-written texts. In terms of discourse components, Zhou et al. (2023) highlight the need for institutional guidelines on assessing cohesion and coherence in AI-generated writing. Furthermore, Alafnan and Mohdzuki (2023) emphasize the importance of policies that address the detection of AI-specific discourse patterns in academic submissions. When it comes to style, Lozic and Štular (2023) suggest that institutions should establish criteria for the acceptable use of AI in developing a writing style, particularly in academic contexts. Additionally, Amirjalili et al. (2024) propose that educators should be equipped to guide students in maintaining appropriate registers when using AI writing tools. Levin et al. (2024) further imply that policymakers should consider developing standards for evaluating the authenticity of AI-generated texts across different registers and genres. Collectively, these studies highlight the need for a comprehensive approach to understanding and managing the linguistic and discourse features of AI-generated texts in educational settings. Such an approach

should aim to maintain academic integrity while also leveraging the potential benefits of AI in language learning and writing instruction.

Research directions concerning textual and linguistic features, discourse components, style, and register in AI-generated writing emphasize the need for comprehensive analysis across various text types and languages. For example, Zaitsu and Jin, 2023 suggest expanding stylometric analysis to more diverse text formats, including informal writing like blogs and social media posts, and comparing texts generated by different AI systems. There is also a call for extending linguistic analysis to various genres and disciplines, incorporating genre-specific features into AI detection algorithms (Leong, 2023). Moreover, researchers recommend further investigation into methods for distinguishing AI-generated from human-written texts based on established linguistic and stylistic criteria (Amirjalili et al., 2024). The exploration of feature selection's influence on AI detection classifiers and the extension of analysis to non-English languages are also proposed by Hayawi et al., 2023. Additionally, there is a need for more in-depth research on the stylistic features of AI-generated text to enhance detection methods (Alafnan & Mohdzuki, 2023), as well as investigations into AI's performance in specific linguistic aspects like cohesion in narrative writing (Zhou et al., 2023). In sum, these directions collectively underscore the importance of nuanced, genre-specific linguistic analysis to better understand and identify the unique characteristics of AI-generated text across different writing contexts and languages.

Implications for Practice

This study has several important implications for practice in EFL writing instruction. First and foremost, it provides comprehensive instructional guidelines for EFL instructors to develop the necessary competencies for effectively integrating AI

technologies into their writing pedagogy. These guidelines can serve as a roadmap for professional development programs and teacher training initiatives and help educators adapt to the rapidly evolving landscape of AI-assisted writing instruction.

Another key implication is the need for a holistic approach to AI integration in writing classrooms. The guidelines encompass various aspects of teaching, from ethical considerations and assessment strategies to genre-specific applications and linguistic analysis. This suggests that educators should try for a balanced and multidimensional approach when incorporating AI tools into their practice.

The study also highlights the importance of ongoing professional development in this field. As AI technologies continue to advance, instructors must stay informed about new developments and their potential impacts on writing instruction. Institutions should consider implementing regular training sessions and workshops to keep their faculty up-to-date with the latest AI tools and best practices.

Furthermore, the guidelines emphasize the critical role of ethical considerations in AI-assisted writing instruction. This implies that educators and institutions should prioritize the development of clear guidelines and policies regarding the use of AI in academic writing. Such policies should comprehensively address issues of academic integrity, plagiarism, and the appropriate attribution of AI-generated content.

Another practical implication is the need for a more nuanced and innovative approach to assessment in AI-integrated writing environments. Instructors may need to revise their evaluation methods and rubrics to account for the use of AI tools while still emphasizing original thought and critical thinking skills. Institutions may also need to invest in AI-powered assessment tools, with proper training for educators to ensure fairness and integrity in the process of assessment.

Lastly, the study underscores the importance of fostering students' AI literacy alongside their writing skills. This suggests that writing curricula should incorporate elements that teach students how to effectively and responsibly use AI tools as part of their writing process. This could prepare them for a future where AI-assisted writing is likely to be commonplace.

Implications for Further Research

Future Study Implications from the Synthesized Studies

The field of AI-assisted writing is rapidly evolving, and researchers have identified several key areas for future investigation. The most pressing need, emphasized is for studies with expanded sample sizes and greater diversity. This includes larger participant groups, more diverse demographics, and research across multiple institutions and disciplines. Closely related to this is the call for methodological advancements. Researchers suggest employing mixed-methods approaches, conducting longitudinal studies, and designing more robust experimental studies to enhance the validity and generalizability of findings.

A significant focus is on exploring AI integration in various writing contexts. This includes investigating its application across different genres and disciplines, comparing its use in academic versus professional writing, and examining its impact across multiple languages and cultural contexts. Complementing this is the need to study the linguistic features of AI-generated text, including contrastive analysis with human-written text and examining stylometric features across different AI models.

Ethical considerations and academic integrity issues are another major concern. This encompasses developing more effective plagiarism detection methods, creating guidelines for appropriate AI use, and addressing questions of authorship and intellectual property in AI-assisted writing. Closely related to this are

suggestions for new assessment and evaluation strategies, which include the development of AI-assisted scoring systems, new rubrics for AI-integrated writing, and methods for measuring originality and critical thinking in this new context.

Researchers also emphasize the importance of studying the long-term effects of AI use on writing skills and investigating user behavior and perceptions. This includes examining the impact on creativity and critical thinking, potential over-reliance on AI tools, and how student and educator attitudes towards AI evolve over time. Further research could explore how collaboration between teachers and students around AI tools can enhance the learning process. Additionally, there is a call for research on AI literacy and pedagogical strategies, focusing on developing AI literacy curricula and best practices for integrating AI into writing instruction.

The technological aspects of AI in writing are not overlooked, calling for research on advancements in AI writing tools. This includes comparing different AI models, and developing more sophisticated AI writing assistants, and improving AI's understanding of context and nuance. Related to this is the need for better AI detection tools and methods to balance the ability to identify AI-generated text.

Finally, researchers highlight the importance of investigating the cognitive and metacognitive aspects of AI-assisted writing, discipline-specific applications, feedback and revision processes, and policy and institutional implications. These areas encompass studying the impact on writing processes, exploring AI applications in fields like STEM and creative writing, examining the effectiveness of AI-generated feedback, and developing institutional guidelines for AI use in academic settings.

This diverse range of research reflects the complex and multifaceted nature of AI's impact on writing. It emphasizes the need for a comprehensive, interdisciplinary

approach to fully understand and effectively leverage AI in writing instruction while addressing the challenges it presents.

Future Study Implications for Interpretive Meta-Synthesis Studies

In light of the rapid advancements in artificial intelligence, future interpretive meta-synthesis studies could expand their scope beyond the current focus on AI and writing to encompass a broader range of skills and emerging AI technologies.

Although this study has provided valuable insights into AI's impact on writing, there is now a pressing need to explore how AI integrates with and affects other skills, such as speaking, listening, and reading. To address this need, future researchers should consider conducting comprehensive meta-syntheses that examine the intersection of AI with multiple skill domains. By doing so, they can provide a more holistic view of AI's impact on education and professional development.

Furthermore, as AI continues to evolve, studies should also incorporate analyses of newer AI tools. For instance, researchers might explore multimodal AI systems, AI-powered virtual reality environments, and adaptive learning platforms since all of these have the potential to significantly influence skill development.

Additionally, future meta-syntheses could benefit significantly from adopting a longitudinal approach. Such an approach would allow researchers to track how perceptions and implementations of AI in skill development evolve over time, which could help them capture the dynamic nature of this rapidly advancing field.

Moreover, to ensure a comprehensive understanding of AI's impact, researchers should consider expanding the scope of their data sources. While academic literature remains crucial, the inclusion of industry reports, policy documents, and user-generated content could provide valuable insights into real-world applications and challenges.

In terms of methodological advancements, future studies could employ more sophisticated text analysis techniques and AI-assisted research tools in the meta-synthesis process itself. These advanced methods have the potential to uncover deeper insights and patterns across the synthesized studies. This could lead to more nuanced and robust findings.

An important direction for future research is the validation and refinement of the guidelines proposed in this study. Researchers could conduct empirical studies to test the effectiveness and applicability of the guidelines in real-world educational settings. This could involve longitudinal studies tracking the implementation of the guidelines in various EFL writing classrooms and its impact on student outcomes and teacher development or comparative studies examining the performance of instructors trained using these guidelines versus those using traditional professional development methods. Additionally, mixed-methods research could combine quantitative assessments of instructor competencies with qualitative insights into their experiences implementing AI-integrated writing instruction. Also, action research projects where instructors actively use and reflect on the guidelines could provide feedback for its iterative improvement. Lastly, cross-cultural studies could test the guidelines' adaptability and effectiveness across different educational contexts and cultural settings. Overall, these validation studies would not only provide evidence for the guidelines' effectiveness but also offer insights into its continuous improvement and adaptation to the evolving landscape of AI in education.

By broadening the focus, incorporating newer technologies, changing methodological approaches, and validating proposed frameworks future meta-synthesis studies can offer even more valuable guidance for educators, policymakers,

and technology experts. As the field continues to progress, such comprehensive and forward-looking research will play a crucial role in shaping effective strategies for integrating AI into educational contexts.

Limitations

While this meta-synthesis study has aimed to provide valuable insights into the intersection of AI and writing instruction, it is important to acknowledge several limitations that may have influenced its outcomes and scope.

Firstly, the study's inclusion of articles exclusively from Q1 and Q2 journals, with the aim of ensuring a high standard of academic rigor, may have excluded relevant research published in other reputable sources. This limitation potentially narrows the perspective of the synthesis, as emerging trends or innovative approaches might appear in journals outside these Q1 and Q2 journals. Additionally, the rapid pace of AI development in education means that significant findings may be published in various formats, including conference proceedings or preprint servers, which were not included in this study's scope.

Secondly, the focus on ChatGPT as the primary AI tool in the search criteria represents another limitation. While ChatGPT has undoubtedly been a significant player in the AI writing landscape, it is only one of many AI technologies influencing writing instruction. This narrowed-down focus may have led to the omission of studies that examine other AI writing tools or broader AI applications in education. As a result, the study's findings may not fully capture the broader scope of AI's impact on writing instruction and may be more reflective of the specific capabilities and limitations of ChatGPT.

Lastly, the decision to concentrate on developing guidelines for writing instruction for EFL teachers rather than including students represents a significant

limitation. This focus was necessary to maintain a manageable scope; however, the interaction between AI and writing is fundamentally a two-way process involving both educators and learners. By focusing solely on teacher competencies, the study may have missed important insights into how students engage with, learn from, and develop competencies related to AI-assisted writing. This limitation suggests a clear direction for future research: exploring the competencies and developing guidelines required by students in an AI-enhanced writing environment.



REFERENCES

- Abd-Elaal, E. S., Gamage, S. H. P. W., & Mills, J. E. (2022). Assisting academics to identify computer-generated writing. *European Journal of Engineering Education*, 47(5), 725–745. <https://doi.org/10.1080/03043797.2022.2046709>
- Afshan, B. (2024). Navigating the ethical landscape: AI integration in education. *KU Education Journal*, 30(6). <https://doi.org/10.53555/kuey.v30i6.5546>
- Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, 126, 334–345. <https://doi.org/10.1016/j.compedu.2018.07.021>
- Akşit, H. N. (1998). A case study on teacher perception of quality of work life within the context of human resource management in an educational institution. [Unpublished doctoral dissertation]. Middle East Technical University.
- Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, 100(8), 2965–2971. <https://doi.org/10.1021/acs.jchemed.3c00323>
- Alasadi, E., & Baiz, R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, 100(8), 2965–2971. <https://doi.org/10.1021/acs.jchemed.3c00323>
- Aldosemani, T., Assalahi, H., Lhothali, A., & Albsisi, M. (2023). Automated writing evaluation in EFL contexts. *International Journal of Computer-Assisted Language Learning and Teaching*, 13(1), 19. <https://doi.org/10.4018/ijcallt.329962>
- Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). The effects of artificial intelligence applications in educational settings: Challenges and

strategies. *Technological Forecasting and Social Change*, 199, article 123076. <https://doi.org/10.1016/j.techfore.2023.123076>

Alikaniotis, D., Yannakoudakis, H., & Rei, M. (2016). Automatic text scoring using neural networks. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 715–725. Association for Computational Linguistics. <https://doi.org/10.18653/v1/P16-1068>

Aljuaid, H. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal (AWEJ)*, Special Issue on ChatGPT, 26–55. <https://dx.doi.org/10.24093/awej/ChatGPT.2>

Allen, I. E., & Seaman, J. (2013). Changing course: Ten years of tracking online education in the United States. Sloan Consortium.

Allen, L. K., Öncel, P., & Flynn, L. (2024). Expanding AWE to incorporate reading and writing evaluation. *Routledge*. <https://doi.org/10.4324/9781003397618-33>

Alsaedi, N. (2024). ChatGPT and EFL/ESL writing: A systematic review of advantages and challenges. *English Language Teaching*, 17(5), 41. <https://doi.org/10.5539/elt.v17n5p41>

An, X., Li, Y., Wang, J., & Zhang, T. (2023). Modeling English teachers' behavioral intention to use artificial intelligence in middle schools. *Education and Information Technologies*, 28(5), 5187–5208. <https://doi.org/10.1007/s10639-022-11286-z>

- Anson, C. M., & Straume, I. (2022). Amazement and trepidation: Implications of AI-based natural language production for the teaching of writing. *Journal of Academic Writing*, 12(1), 1–9. <https://doi.org/10.18552/joaw.v12i1.820>
- Ateeq, A., Alaghbari, M., Alzoraiki, M., Milhem, M., & Beshr, B. (2024). Empowering academic success: Integrating AI tools in university teaching for enhanced assignment and thesis guidance. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSYS)* (pp. 297–301). IEEE.
<https://doi.org/10.1109/ICETSYS61505.2024.10459686>
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
<https://doi.org/10.5281/zenodo.3554657>
- Baker, T., & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Nesta.
https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf
- BALEAP. (2008). *Competency framework for teachers of English for academic purposes*. Retrieved July 28, 2024, from <https://www.baleap.org/wp-content/uploads/2016/04/teap-competency-framework.pdf>
- Balta, N. (2023). Ethical considerations in using AI in educational research. *Journal of Research in Innovative Education*, 12(3).
<https://doi.org/10.51853/jorids/14205>
- Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: A critical review. *BMC Medical Research Methodology*, 9, 59.
<https://doi.org/10.1186/1471-2288-9-59>

- Bawden, D. (2008). Origins and concepts of digital literacy. In C. Lankshear & M. Knobel (Eds.), *Digital literacies: Concepts, policies and practices* (pp. 17–32). Peter Lang.
- Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.
- Bergmann, J., & Sams, A. (2014). *Flipped learning: Gateway to student engagement*. International Society for Technology in Education.
- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. <https://doi.org/10.1007/s12528-013-9077-3>
- Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. *ASEE National Conference Proceedings* (Vol. 30, No. 9), 1–18.
- Bonk, C. J., & Graham, C. R. (Eds.). (2012). *The handbook of blended learning: Global perspectives, local designs*. John Wiley & Sons.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- British Council. (2015). Continuing professional development (CPD) framework for teachers. Retrieved July 28, 2024, from https://www.teachingenglish.org.uk/sites/teacheng/files/CPD%20framework%20for%20teachers_WEB.PDF
- Bruneault, F., Laflamme, A., & Mondoux, A. (2022). AI ethics training in higher education: Competency framework. *OSF Preprints*. <https://doi.org/10.31235/osf.io/x5nck>

- Bu, Q. (2022). Ethical risks in integrating artificial intelligence into education and potential countermeasures. *Science Insights*, 41(1), 561–566.
<https://doi.org/10.15354/si.22.re067>
- Cambridge Assessment English. (2018). *Cambridge English teaching framework*. Retrieved from <https://www.cambridgeenglish.org/Images/172992-full-level-descriptors-cambridge-english-teaching-framework.pdf>
- Campbell, K. A., Orr, E., Durepos, P., Nguyen, L., Li, L., Whitmore, C., Gehrke, P., Graham, L., & Jack, S. M. (2021). Reflexive thematic analysis for applied qualitative health research. *The Qualitative Report*, 26(6), 2011–2028.
<https://doi.org/10.46743/2160-3715/2021.5010>
- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11(4), 190–202. <https://doi.org/10.1109/TMMS.1970.299942>
- Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age. *Business and Professional Communication Quarterly*, 86(3), 257–295. <https://doi.org/10.1177/23294906231176517>
- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge University Press.
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278.
<https://doi.org/10.1109/ACCESS.2020.2988510>
- Chen, X., Binti, N. H., & Kamarudin, S. (2023). Preferred reporting items for a systematic review and meta-analysis (PRISMA): The relationship between social media use and political participation behavior. *International Journal of*

Academic Research in Business & Social Sciences, 13(4), 110-133.

<https://doi.org/10.6007/ijarbss/v13-i4/16646>

Chen, X., Liu, Y., Xiang, H., Jin, H., Boudouaia, A., Yao, J., & Li, Y. (2024).

Exploring differences between automated writing evaluation (AWE)

feedback and teacher feedback and their impact on college students' second

language (L2) writing practice. SSRN. <https://doi.org/10.2139/ssrn.4826248>

Chen, Y., Wang, Y., & Chen, N. (2014). Is FLIP enough? Or should we use the

FLIPPED model instead? *Computers & Education*, 79, 16–27.

<https://doi.org/10.1016/j.compedu.2014.07.004>

Cheng, X., Sun, J., & Zarifis, A. (2020). Artificial intelligence and deep learning in

educational technology research and practice. *British Journal of Educational*

Technology, 51, 1653–1656. <https://doi.org/10.1111/bjet.13018>

Chrastina, J. (2018). Meta-synthesis of qualitative studies: Background,

methodology, and applications. *Proceedings of the NORDSCI Conference*,

11. <https://doi.org/10.32008/NORDSCI2018/B1/V1/13>

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in education needs

interpretable machine learning: Lessons from open learner modelling. *arXiv*

Preprint. <https://doi.org/10.48550/arXiv.1807.00154>

Cota-Rivera, E. I., González Correa, M. E., Bernal Marín, L. A., Marquez

Montenegro, M. Y., Herrera, A. M., & Martinez Martinez, M. A. (2024).

Transforming education with the power of artificial intelligence: Case

studies. In Ş. Demir & M. Demir (Eds.), *Enhancing higher education and*

research with OpenAI models (pp. 113–140). IGI Global.

<https://doi.org/10.4018/979-8-3693-1666-5.ch006>

- Crompton, H. (2023). Evidence of the ISTE Standards for Educators leading to learning gains. *Journal of Digital Learning in Teacher Education*, 39(4), 201–219. <https://doi.org/10.1080/21532974.2023.2244089>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(22). <https://doi.org/10.1186/s41239-023-00392-8>
- Dašić, P., Dašić, J., Crvenković, B., & Serifi, V. (2016). A review of intelligent tutoring systems in e-learning. *Annals of the Oradea University – Fascicle of Management and Technological Engineering*, 15, 85–90. <https://doi.org/10.15660/AUOFMTE.2016-3.3276>
- Davison, R. M., Laumer, S., Tarafdar, M., & Wong, L. H. M. (2023). Pickled eggs: Generative AI as research assistant or co-author? *Information Systems Journal*, 33(5), 989–994. <https://doi.org/10.1111/isj.12455>
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1810.04805>
- Dickinson, P. (2018). Addressing current and future challenges in EAL writing with Universal Design for Learning. In P. Taalas, J. Jalkanen, L. Bradley, & S. Thouësny (Eds.), *Future-proof CALL: Language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 41–46). Research-publishing.net. <https://doi.org/10.14705/rpnet.2018.26.810>
- Dong, R., Fu, D., Zhou, X., & Wang, B. (2019). Writing instruction in China: Challenges and efforts. *International Journal of Information and Education Technology*, 9(10), 756–760. <https://doi.org/10.18178/ijiet.2019.9.10.1292>

- Dreon, O., Kerper, R. M., & Landis, J. (2011). Digital storytelling: A tool for teaching and learning in the YouTube generation. *Middle School Journal*, 42(5), 4–10. <https://doi.org/10.1080/00940771.2011.11461777>
- Drisko, J. (2019). Qualitative research synthesis: An appreciative and critical introduction. *Qualitative Social Work*, 19(3), 147332501984880. <https://doi.org/10.1177/1473325019848808>
- EDUCAUSE. (2022). 2022 EDUCAUSE horizon report: Teaching and learning edition. Retrieved from <https://library.educause.edu/resources/2022/4/2022-educause-horizon-report-teaching-and-learning-edition>
- Edwards, J., & Kaimal, G. (2016). Using meta-synthesis to support application of qualitative methods findings in practice: A discussion of meta-ethnography, narrative synthesis, and critical interpretive synthesis. *The Arts in Psychotherapy*, 51, 30–35. <https://doi.org/10.1016/j.aip.2016.07.003>
- Elali, F. R., & Rachid, L. N. (2023). AI-generated research paper fabrication and plagiarism in the scientific community. *Patterns*, 4(3), 100706. <https://doi.org/10.1016/j.patter.2023.100706>
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal of Educational Integrity*, 19, Article 17. <https://doi.org/10.1007/s40979-023-00140-5>
- El-Rashidy, M. A., Mohamed, R. G., El-Fishawy, N. A., & Shouman, M. A. (2022). Reliable plagiarism detection system based on deep learning approaches. *Neural Computing and Applications*, 34, 18837–18858. <https://doi.org/10.1007/s00521-022-07486-w>

- Engin, M. (2014). Extending the flipped classroom model: Developing second language writing skills through flipped learning. *International Journal of Instructional Technology and Distance Learning*, 11(12), 1–10.
<https://doi.org/10.14434/josotlv14i5.12829>
- European Commission, Directorate-General for Education, Youth, Sport and Culture. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2766/153756>
- European Commission. (2019). *Key competences for lifelong learning*. Publications Office of the European Union. Retrieved from https://ec.europa.eu/education/education-in-the-eu/council-recommendation-on-key-competences-for-lifelong-learning_en
- Eva, N. R., Azizah, S., & Virdyna, N. K. (2023). AI in academic writing: Assessing current usage and future implications. *Insania: Jurnal Pemikiran Alternatif Kependidikan*, 28(1). <https://doi.org/10.24090/insania.v28i1a.9278>
- Fang, X., Ng, D. T. K., & Leung, J. K. L. (2023). A systematic review of artificial intelligence technologies used for story writing. *Education and Information Technologies*, 28, 14361–14397. <https://doi.org/10.1007/s10639-023-11741-5>
- Farina, A., & Stevenson, C. N. (2024). Ethical navigations: Adaptable frameworks for responsible AI use in higher education. In A. Ara & A. Ara (Eds.), *Exploring the ethical implications of generative AI* (pp. 63–87). IGI Global. <https://doi.org/10.4018/979-8-3693-1565-1.ch005>
- Fingfeld-Connett, D. (2018). *A guide to qualitative meta-synthesis*: Routledge. <https://doi.org/10.4324/9781351212793>

- Florence, R., Sullivan, P., & Keith, K. (2019). Exploring the potential of natural language processing to support microgenetic analysis of collaborative learning discussions. *British Journal of Educational Technology*, 50(6), 3047-3063. <https://doi.org/10.1111/bjet.12875>
- Gaertner, M. N., & McClarty, K. L. (2015). Performance, perseverance, and the full picture of college readiness. *Educational Measurement: Issues and Practice*, 34(2), 20–33. <https://doi.org/10.1111/emip.12066>
- Garcia, J. M., Sehgal, J., Thompson, J. L., Woods, S. P., & Medina, L. D. (2024). The relationship between apathy and cognitive impairment among Hispanic/Latin Americans: A preferred reporting items for systematic reviews and meta-analyses systematic review. *Journal of Geriatric Psychiatry and Neurology*, 37(3), 175–193. <https://doi.org/10.1177/08919887231207640>
- Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integrating innovation. *The Modern Language Journal*, 93(4), 719–740. <https://doi.org/10.1111/j.1540-4781.2009.00969.x>
- Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. John Wiley & Sons.
- Gartner, S., & Krašna, M. (2023). Artificial intelligence in education - Ethical framework. *2023 12th Mediterranean Conference on Embedded Computing (MECO)*, 1–7. <https://doi.org/10.1109/MECO58584.2023.10155012>
- Gavina, M. A., & Ibay-Pamo, L. M. (2024). Effect of automated writing evaluation in higher education academic writing performance. *Edulangue*, 6(2), 118–137. <https://doi.org/10.20414/edulangue.v6i2.8300>

- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055.
<https://doi.org/10.1016/j.caeai.2022.100055>
- Godwin-Jones, R. (2017). Smartphones and language learning. *Language Learning & Technology*, 21(2), 3–17. <https://doi.org/10.10125/44607>
- Godwin-Jones, R. (2018). Second language writing online: An update. *Language Learning & Technology*, 22(1), 1–15. <https://doi.org/10.10125/44574>
- Goldshtein, M., Alhashim, A. G., & Roscoe, R. D. (2023). Exploring variance in student writing and syntax: Implications for automated writing evaluation. SSRN. <https://doi.org/10.2139/ssrn.4552191>
- Golwalker, A., Solanki, J., Choudhary, G., & Sarker, G. (2024). Narration with graphics: An AI generated story teller. *International Journal of Scientific Research in Engineering and Management*, 8(5), 1.
<https://doi.org/10.55041/IJSREM32119>
- Gonda, D., Luo, J., Wong, Y. L., & Lei, C. U. (2018). Evaluation of developing educational chatbots based on the seven principles for good teaching. *Proceedings of the 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 446–453. IEEE.
<https://doi.org/10.1109/TALE.2018.8615175>
- Gong, D., Yang, H. H., & Cai, J. (2020). Exploring the key influencing factors on college students' computational thinking skills through flipped-classroom instruction. *International Journal of Educational Technology in Higher Education*, 17, 19. <https://doi.org/10.1186/s41239-020-00196-0>

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *Advances in Neural Information Processing Systems*, 3. <https://doi.org/10.1145/3422622>
- Gozali, I., Ryan, A., Wijaya, T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. *The JALT CALL Journal*, 20(1). <https://doi.org/10.29140/jaltcall.v20n1.1200>
- Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *APA educational psychology handbook, Vol. 3. Applications to learning and teaching* (pp. 451–473). American Psychological Association. <https://doi.org/10.1037/13275-018>
- Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), *The handbook of blended learning: Global perspectives, local designs* (pp. 3–21). Pfeiffer.
- Gupta, B. (2023). Spotting the artificial intelligence mask: Detecting articles written by language models/ChatGPT. *Indian Journal of Anaesthesia*, 67(9), 835. https://doi.org/10.4103/ija.ija_497_23
- Guskey, T. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery." *Journal of Advanced Academics*, 19(1), 8–31. <https://doi.org/10.4219/jaa-2007-704>

- Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4). <https://doi.org/10.1177/0008125619864925>
- Harry, A. (2023). Role of AI in education. *Injury: Interdisciplinary Journal and Humanity*, 2(3), 260–268. <https://doi.org/10.58631/injury.v2i3.52>
- Heift, T., & Schulze, M. (2015). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. Routledge.
- Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A systematic review of deep learning approaches to educational data mining. *Complexity*, 2019, Article 1306039, 22 pages. <https://doi.org/10.1155/2019/1306039>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. <https://doi.org/10.1111/ejed.12533>
- Horn, M. B., & Staker, H. (2011). *The rise of K-12 blended learning*. Innosight Institute. <https://www.christenseninstitute.org/wp-content/uploads/2013/04/The-rise-of-K-12-blended-learning.pdf>
- Hossain, M., & Mustafa, H. A. (2023). Automated writing evaluation using sentence by sentence scoring model. *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)* (pp. 1–6). IEEE. <https://doi.org/10.1109/NCIM59001.2023.1021275>
- Huang, H. L., Chou, C. P., Leu, S., & others. (2020). Effects of a quasi-experimental study of using flipped classroom approach to teach evidence-based medicine

to medical technology students. *BMC Medical Education*, 20, Article 31.

<https://doi.org/10.1186/s12909-020-1946-7>

Hung, H. T. (2014). Flipping the classroom for English language learners to foster active learning. *Computer Assisted Language Learning*, 28(1), 81–96.

<https://doi.org/10.1080/09588221.2014.967701>

Hutson, J., Plate, D., & Berry, K. (2024). Embracing AI in English Composition: Insights and Innovations in Hybrid Pedagogical Practices. *International Journal of Changes in Education*, 1(1), 19–31.

<https://doi.org/10.47852/bonviewIJCE42022290>

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines”. *Language Testing in Asia*, 13(1), 1-28. <https://doi.org/10.1186/s40468-023-00260-2>

Jett, J. R. (2013). *Passage from pen and paper to keyboard and screen: An investigation of the evolution of writing instruction in one-to-one laptop settings* (Publication No. 3575630) [Doctoral dissertation, North Carolina State University]. ProQuest Dissertations & Theses Global.

Ji, S., & Yuan, T. (2022). Conversational intelligent tutoring systems for online learning: What do students and tutors say? *Proceedings of the EDUCON 2022*, 292–298. <https://doi.org/10.1109/EDUCON52537.2022.9766567>

Johnson, M., Fitzsimons, S., & Coleman, V. (2023). Development challenges in challenging contexts: A 3-stage curriculum framework design approach for education in emergencies. *Prospects*, 53(1), 43–56.

<https://doi.org/10.1007/s11125-022-09601-0>

- Kazim, E., & Koshiyama, A. (2020). A high-level overview of AI ethics. *SSRN Journal*. <https://doi.org/10.2139/ssrn.3609292>
- Kerr, P. (2016). Adaptive learning. *ELT Journal*, 70(1), 88–93.
<https://doi.org/10.1093/elt/ccv055>
- Kim, K., & Kwon, K. (2022). Exploring the AI competencies of elementary school teachers in South Korea. *Computers and Education: Artificial Intelligence*, 4, Article 100137. <https://doi.org/10.1016/j.caeai.2023.100137>
- Kim, M. K., Kim, N. J., & Heidari, A. (2022). Learner experience in artificial intelligence-scaffolded argumentation. *Assessment & Evaluation in Higher Education*, 47(8), 1301–1316.
<https://doi.org/10.1080/02602938.2022.2042792>
- Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1312.6114>
- Klein-Collins, R. (2013). *Sharpening our focus on learning: The rise of competency-based approaches to degree completion*. National Institute for Learning Outcomes Assessment.
<https://learningoutcomesassessment.org/documents/Klein%20Collins%20OP20.pdf>
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30–43.
- Kovanen, B., Ware, R., Mericle, M., Turnipseed, N., Coleman, J., Elliott, C., Popovics, J., Cooper, S., Gallagher, J., Prior, P., & Zilles, J. (2020). Implementing writing-as-process in engineering education. *Proceedings of*

the ASEE Annual Conference & Exposition, Paper ID #30364.

<https://doi.org/10.18260/1-2--34786>

Kress, G. (2003). *Literacy in the new media age*. Routledge.

Lang, S. (2018). Evolution of instructor response? Analysis of five years of feedback to students. *Journal of Writing Analytics*, 2, Article 2.

<https://doi.org/10.37514/JWA-J.2018.2.1.02>

Lankshear, C., & Knobel, M. (2011). *New literacies: Everyday practices and social learning* (3rd ed.). Open University Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221.

<https://doi.org/10.1016/j.caeai.2024.100221>

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2013). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 1150–1181). International Reading Association. <https://doi.org/10.1598/0710.42>

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.

Li, Z., Liang, C., Peng, J., & Yin, M. (2024). The value, benefits, and concerns of generative AI-powered assistance in writing. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, Article

1048, 1–25. Association for Computing Machinery.

<https://doi.org/10.1145/3613904.3642625>

Liddy, E. D. (2001). Natural language processing. In M. A. Drake (Ed.), *Encyclopedia of library and information science* (2nd ed.). Marcel Dekker.

Lin, C., Huang, A. Y., & Lu, O. H. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments*, 10(1), 1–22. <https://doi.org/10.1186/s40561-023-00260-y>

Lin, H. C., Ho, C. F., & Yang, H. (2022). Understanding the adoption of artificial intelligence-enabled language e-learning system: An empirical study of UTAUT model. *International Journal of Mobile Learning and Organisation*, 16(1), 74-69. <https://doi.org/10.1504/ijmlo.2022.119966>

Lin, Y., Luo, Q., & Qian, Y. (2023). Investigation of artificial intelligence algorithms in education. *Applied and Computational Engineering*, 16, 180-184. <https://doi.org/10.54254/2755-2721/16/20230886>

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Livberber, T., & Ayvaz, S. (2023). The impact of artificial intelligence in academia: Views of Turkish academics on ChatGPT. *Heliyon*, 9, e19688. <https://doi.org/10.1016/j.heliyon.2023.e19688>

Liz-Domínguez, M., Caeiro-Rodríguez, M., Llamas-Nistal, M., & Mikic-Fonte, F. A. (2019). Systematic literature review of predictive analysis tools in higher education. *Applied Sciences*, 9(24), 5569. <https://doi.org/10.3390/app9245569>

- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-16. ACM. <https://doi.org/10.1145/3313831.3376727>
- Luckin, R., & Holmes, W. (2017). *A.I. in education: The promises and the pitfalls*. OECD Education Working Papers, No. 159. OECD Publishing. <https://doi.org/10.1787/3f52c8ac-en>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.
- Luo, J., & Qiu, L. (2024). Review of artificial intelligence-based tools in EFL classroom. *The Arts, Culture, and Language Journal*, 1(7). <https://doi.org/10.61173/s266nr95>
- Maeda, Y., Caskurlu, S., Kenney, R. H., Kozan, K., & Richardson, J. C. (2022). Moving qualitative synthesis research forward in education: A methodological systematic review. *Educational Research Review*, 35, 100424. <https://doi.org/10.1016/j.edurev.2021.100424>
- Major, C. H., & Savin-Baden, M. (2012). *An introduction to qualitative research synthesis: Managing the information explosion in social science research*. Routledge. <https://doi.org/10.4324/9780203497555>
- Májovský, M., Černý, M., Netuka, D., & Mikolov, T. (2024). Perfect detection of computer-generated text faces fundamental challenges. *Cell Reports Physical Science*, 5(1), 101769. <https://doi.org/10.1016/j.xcrp.2023.101769>
- Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., & Darwis, A. (2023). Exploring artificial intelligence in academic essay: Higher education students' perspective. *International Journal of Educational Research Open*, 5, 100296. <https://doi.org/10.1016/j.ijedro.2023.100296>

- Mat, A. C. (2024). Automated writing evaluation: Users' perception and expectations. *International Journal of Information and Education Technology*, 14(2). 183–192. <https://doi.org/10.18178/ijiet.2024.14.2.2039>
- McCarthy, J. (2017). Enhancing feedback in higher education: Students' attitudes towards online and in-class formative assessment feedback models. *Active Learning in Higher Education*, 18(2), 127–141. <https://doi.org/10.1177/1469787417707615>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12–14. <https://doi.org/10.1609/aimag.v27i4.1904>
- McLellan, H. (2007). Digital storytelling in higher education. *Journal of Computing in Higher Education*, 19(1), 65–79. <https://doi.org/10.1007/BF03033420>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. U.S. Department of Education. <https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf>
- Mikeladze, T., Meijer, P. C., & Verhoeff, R. P. (2024). A comprehensive exploration of artificial intelligence competence frameworks for educators: A critical review. *European Journal of Education*, 59(3), e12663. <https://doi.org/10.1111/ejed.12663>

- Miles, M. B., Huberman, A. M., & Saldana, J. (2018). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097.
<https://doi.org/10.1371/journal.pmed.1000097>
- Mugo, D. G., Muthwii, S. M., & Gakuru, P. M. (2014). Tracing writing technologies through time: A historical reflection of writing systems, writing surfaces and writing implements. *Journal of Educational Research and Reviews*, 2(6), 83–88.
- Muttaqin, I. (2022). Necessary to increase teacher competency in facing the artificial intelligence era. *Al-Hayat: Journal of Islamic Education*, 6(2), 549–559.
<https://doi.org/10.35723/ajie.v6i2.460>
- Nasir, M., Hasan, M., Adlim, A., & Syukri, M. (2024). Utilizing artificial intelligence in education to enhance teaching effectiveness. *Proceedings of the International Conference on Education (PICE)*, 2, Article 1367.
<https://doi.org/10.32672/pice.v2i1.1367>
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence-powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), e07014.
<https://doi.org/10.1016/j.heliyon.2021.e07014>
- Nazim, M. (2024). Exploring EFL teachers' insights regarding artificial intelligence-driven tools in student-centered writing instructions. *International Journal of English Linguistics*, 14(3), 93–108. <https://doi.org/10.5539/ijel.v14n3p93>

Nguyen, N. (2023). Exploring the role of AI in education. *London Journal of Social Sciences*, pp. 6, 84–95. <https://doi.org/10.31039/ljss.2023.6.108>

North, B., Mateva, G., & Rossner, R. (2013). *The European Profiling Grid (EPG)*. European Centre for Modern Languages of the Council of Europe. Retrieved July 28, 2024, from https://www.eaquals.org/wp-content/uploads/The_EPG_-_PDF_publication_final.pdf

OpenAI. (2023). *About OpenAI*. Retrieved August 22, 2024, from <https://www.openai.com/about>

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27(6), 7893–7925. <https://doi.org/10.1007/s10639-022-10925-9>

Page, E. B. (1966). The imminence of... grading essays by computer. *The Phi Delta Kappan*, 47(5), 238–243. <http://www.jstor.org/stable/20371545>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

Pangrazio, L. (2016). Reconceptualizing critical digital literacy. *Discourse: Studies in the Cultural Politics of Education*, 37(2), 163–174. <https://doi.org/10.1080/01596306.2014.942836>

- Paschal, M. J., & Melly, I. K. (2023). Ethical guidelines on the use of AI in education. In J. Keengwe (Ed.), *Creative AI tools and ethical implications in teaching and learning* (pp. 230–245). IGI Global.
<https://doi.org/10.4018/979-8-3693-0205-7.ch013>
- Pennington, M. (2013). Evolutionary trends in writing pedagogy: An early 21st century view. *Writing & Pedagogy*, 5(1), 1–15.
<https://doi.org/10.1558/wap.v5i1.1>
- Perrotta, C., & Selwyn, N. (2019). Deep learning goes to school: Toward a relational understanding of AI in education. *PsyArXiv Preprints*.
<https://doi.org/10.31235/osf.io/48t7e>
- Phobun, P., & Vicheanpanya, J. (2010). Adaptive intelligent tutoring systems for e-learning systems. *Procedia - Social and Behavioral Sciences*, 2, 4064–4069.
<https://doi.org/10.1016/j.sbspro.2010.03.641>
- Pilán, I., Vajjala, S., & Volodina, E. (2016). A readable read: Automatic assessment of language learning materials based on linguistic complexity. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1603.08868>
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv Preprint*.
<https://doi.org/10.48550/arXiv.1910.10683>
- Rahayu, R., Sukardi, W., Muliati, M., & De Vega, N. (2024). Artificial intelligence in writing instruction: A self-determination theory perspective. *XLinguae*.
<https://doi.org/10.18355/xl.2024.17.01.16>

- Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674.
<https://doi.org/10.1080/09588221.2018.1428994>
- Rani, S., & Kumar, P. (2017). A sentiment analysis system to improve teaching and learning. *Computer*, 50, 36–43. <https://doi.org/10.1109/MC.2017.133>
- Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2016). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*, 46(2), 330–343.
<https://doi.org/10.1111/bjet.12198>
- Roe, J., Renandya, W. A., & Jacobs, G. M. (2023). A review of AI-powered writing tools and their implications for academic integrity in the language classroom. *Journal of Applied Linguistics*, 2(1). <https://doi.org/10.59588/2961-3094.1035>
- Roquet, H., Navarro Gil, N., & Nicolás-Conesa, F. (2024). The impact of adjunct instruction on EFL academic writing at university. *Journal of Immersion and Content-Based Language Education*, 12(2), 162–191.
<https://doi.org/10.1075/jicb.22005.roq>
- Rossner, R. (2017). *Language teaching competences*. Oxford University Press.
- Saaïda, M. (2023). AI-driven transformations in higher education: Opportunities and challenges. *International Journal of Educational Research and Studies*, 5(1), 29–36. <https://doi.org/10.5281/zenodo.8164414>
- Sadik, A. (2008). Digital storytelling: A meaningful technology-integrated approach for engaged student learning. *Educational Technology Research and Development*, 56, 487–506. <https://doi.org/10.1007/s11423-008-9091-8>

- Salloum, S. A. (2024). AI perils in education: Exploring ethical concerns. In A. Al-Marzouqi, S. A. Salloum, M. Al-Saidat, A. Aburayya, & B. Gupta (Eds.), *Artificial intelligence in education: The power and dangers of ChatGPT in the classroom* (Vol. 144, pp. 43). Springer, Cham.
https://doi.org/10.1007/978-3-031-52280-2_43
- Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Morano-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT challenges blended learning methodologies in engineering education: A case study in mathematics. *Applied Sciences*, 13(10), 6039. <https://doi.org/10.3390/app13106039>
- Scimago, L. (2006). Qualitative meta-synthesis: A question of dialoguing with texts. *Journal of Advanced Nursing*, 53(3), 311–318.
<https://doi.org/10.1111/j.1365-2648.2006.03721.x>
- Scimago. (n.d.). SJR: Scientific journal rankings. Retrieved from
<https://www.scimagojr.com>
- Sehgal, P. S., Zaidi, N. N., & Rana, S. S. (2024). Application of AI in the field of education: Dynamic course generation. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 12(5), 31–.
<https://doi.org/10.22214/ijraset.2024.61320>
- Selim, A. (2024). The transformative impact of AI-powered tools on academic writing: Perspectives of EFL university students. *International Journal of English Linguistics*, 14(1), 14–29. <https://doi.org/10.5539/ijel.v14n1p14>
- Seo, K., Tang, J., Roll, I., & others. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, 18, 54.
<https://doi.org/10.1186/s41239-021-00292-9>

- Sewell, K. A., Schellinger, J., & Bloss, J. E. (2023). Effect of PRISMA 2009 on reporting quality in systematic reviews and meta-analyses in high-impact dental medicine journals between 1993-2018. *PLoS ONE*, 18(12), e0295864. <https://doi.org/10.1371/journal.pone.0295864>
- Shen, R., & Chong, S. W. (2023). Learner engagement with written corrective feedback in ESL and EFL contexts: A qualitative research synthesis using a perception-based framework. *Assessment & Evaluation in Higher Education*, 48(3), 276–290. <https://doi.org/10.1080/02602938.2022.2072468>
- Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook of automated essay evaluation: Current applications and new directions*. Routledge/Taylor & Francis Group.
- Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, 36(2), 187–209. <https://doi.org/10.1017/S0958344023000265>
- Siddharthan, A. (2014). A survey of research on text simplification. *ITL - International Journal of Applied Linguistics*, 165(2), 259–298. <https://doi.org/10.1075/itl.165.2.06sid>
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46, 31–40.
- Sim, J., & Mengshoel, A. M. (2023). Metasynthesis: Issues of empirical and theoretical context. *Qualitative and Quantitative Methods*, 57, 3339–3361. <https://doi.org/10.1007/s11135-022-01502-w>
- Singh, S., & Jain, P. (2024). AI in special education: Emerging trends and challenges. In S. Misra, A. Jain, M. Kaushik, C. Banerjee, & Y. Singh (Eds.),

- Advanced intelligence systems and innovation in entrepreneurship* (pp. 56-75). IGI Global. <https://doi.org/10.4018/979-8-3693-0790-8.ch005>
- Smith, L. (2017). Creative spaces for developing independent writing with English teachers. In D. Stevens & K. Lockney (Eds.), *Students, places, and identities in English and the arts: Creative spaces in education* (pp. 42–54). Routledge. <https://doi.org/10.4324/9781315528014-4>
- Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *Cambodian Journal of Educational Research*, 3, 110–121. <https://doi.org/10.62037/cjer.2023.03.01.06>
- Son, H. J. (2023). Perceptions and attitudes of English learners towards the utilization of AI-based automated writing evaluation (AWE) programs and instructor feedback. *English Language Education*, 29(3), 91-112. <https://doi.org/10.35828/etak.2023.29.3.91>
- Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. John Wiley & Sons.
- Srinivasa, K. G., Kurni, M., & Saritha, K. (2022). Harnessing the power of AI to education. In K. G. Srinivasa, M. Kurni, & K. Saritha (Eds.), *Learning, teaching, and assessment methods for contemporary learners* (pp. 311-342). Springer. https://doi.org/10.1007/978-981-19-6734-4_13
- Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*, 131, 33-48. <https://doi.org/10.1016/j.compedu.2018.12.005>
- Sun, J. C. Y., & Wu, Y. T. (2016). Analysis of learning achievement and teacher-student interactions in flipped and conventional classrooms. *The*

- International Review of Research in Open and Distributed Learning*, 17, 79–99. <https://doi.org/10.19173/irrodl.v17i1.2116>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *arXiv*. <https://doi.org/10.48550/arXiv.1409.3215>
- Sylvester, R., & Greenidge, W. (2009). Digital storytelling: Extending the potential for struggling writers. *The Reading Teacher*, 63, 284–295. <https://doi.org/10.1598/RT.63.4.3>
- Sytnyk, L., & Podlinyayeva, O. (2024). AI in education: Main possibilities and challenges. *Scientific Collection InterConf+*, 45(201), 569-579. <https://doi.org/10.51582/interconf.19-20.05.2024.058>
- Tahiru, F. (2021). AI in education: A systematic literature review. *Journal of Cases on Information Technology*, 23(1), 1-20. <https://doi.org/10.4018/JCIT.2021010101>
- TESOL International Association. (2008). *Standards for ESL/EFL teachers of adults framework*. TESOL Press.
- The Institute for Ethical AI in Education. (2021). *The ethical framework for AI in education*. <https://fb77c667c4d6e21c1e06.b-cdn.net/wp-content/uploads/2021/03/The-Ethical-Framework-for-AI-in-Education-Institute-for-Ethical-AI-in-Education-Final-Report.pdf>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45. <https://doi.org/10.1186/1471-2288-8-45>
- Thorne, S., Jensen, L., Kearney, M. H., Noblit, G., & Sandelowski, M. (2004). Qualitative metasynthesis: Reflections on methodological orientation and

ideological agenda. *Qualitative Health Research*, 14(10), 1342–1365.

<https://doi.org/10.1177/1049732304269888>

Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ.

BMC Medical Research Methodology, 12, 181. <https://doi.org/10.1186/1471-2288-12-181>

Trần, T. T. H. (2024). AI tools in teaching and learning English academic writing skills. *Proceedings of the AsiaCALL International Conference*, 45-58.

AsiaCALL. <https://doi.org/10.54855/paic.23413>

Tseng, W. T., Liou, H. J., & Chu, H. C. (2019). Vocabulary learning in virtual environments: Learner autonomy and collaboration. *System*, p. 88, Article

102190. <https://doi.org/10.1016/j.system.2019.102190>

Tucker, B. (2012). The flipped classroom. *Education Next*, 12(1), 82–83.

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and future of teaching and learning: Insights and*

recommendations. <https://oet.wp.nnth.dev/ai-future-of-teaching-and-learning/>

UNESCO. (2023). *ChatGPT and artificial intelligence in higher education*.

<https://unesdoc.unesco.org/ark:/48223/pf0000385146>

UNESCO. (2023). *Guidance for generative AI in education and research*.

<https://unesdoc.unesco.org/ark:/48223/pf0000386693>

UNESCO. (2024). *Draft AI competency frameworks for teachers and school*

students. <https://unesdoc.unesco.org/ark:/48223/pf0000385796>

UNICEF. (2021). *Policy guidance on AI for children 2.0*.

<https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf>

- Valerie, A., & Storey, V. A. (2023). AI technology and academic writing. *International Journal of Adult Education and Technology*, 14(1), 1-15.
<https://doi.org/10.4018/IJAET.325795>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv Preprint*.
<https://doi.org/10.48550/arXiv.1706.03762>
- Vincent-Lancrin, S., & van der Vlies, R. (2020). Trustworthy artificial intelligence (AI) in education: Promises and challenges. *OECD Education Working Papers*, No. 218. OECD Publishing. <https://doi.org/10.1787/a6c90fa9-en>
- Wan, S., & Niu, Z. (2018). An e-learning recommendation approach based on the self-organization of learning resources. *Knowledge-Based Systems*, 160, 71–87. <https://doi.org/10.1016/j.knosys.2018.06.014>
- Wang, L. (2022). Proactive push research on personalized learning resources based on machine learning. *2022 IEEE International Conference on Unmanned Systems (ICUS)*, 986-991. IEEE.
<https://doi.org/10.1109/ICUS55513.2022.9987163>
- Wang, Y. H., & Liao, H. C. (2011). Data mining for adaptive learning in a TESL-based e-learning system. *Expert Systems with Applications*, 38, 6480–6485.
<https://doi.org/10.1016/j.eswa.2010.11.098>
- Warschauer, M. (2010). New tools for teaching writing. *Language Learning & Technology*, 14(1), 3–8. <http://doi.org/10.125/44196>

- Warschauer, M., & Grimes, D. (2019). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 14(2), 170–185.
<https://doi.org/10.1080/15544800701771580>
- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, 31(2), 57–71.
- Winkler, R., & Sollner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Proceedings of the 2018 Academy of Management Annual Meeting* (Vol. 2018), 15903.
<https://doi.org/10.5465/AMBPP.2018.15903abstract>
- Wongras, P., & Tanantong, T. (2023). An extended UTAUT model for analyzing users' acceptance factors for artificial intelligence adoption in human resource recruitment: A case study of Thailand. *Education and Information Technologies*, 3(7), 13–27. <https://doi.org/10.20944/preprints202311.1612.v1>
- Wu, T., & Yu, Z. (2023). Bibliometrically and systematically analyzing automated writing evaluation for English learning. *Forum for Linguistic Studies*, 5(3).
<https://doi.org/10.59400/fls.v5i3.1907>
- Wu, W., Zhang, B., Li, S., & Liu, H. (2022). Exploring factors of the willingness to accept AI-assisted learning environments: An empirical investigation based on the UTAUT model and perceived risk theory. *Frontiers in Psychology*, 13(3), 78-81. <https://doi.org/10.3389/fpsyg.2022.870777>
- Xiong, J., Yang, J., Yan, L., Awais, M., Khan, A. A., Alizadehsani, R., & Acharya, U. R. (2024). Efficient reinforcement learning-based method for plagiarism detection boosted by a population-based algorithm for pretraining weights. *Expert Systems with Applications*, 238(Part E), 122088.
<https://doi.org/10.1016/j.eswa.2023.122088>

- Yalcinalp, S., Türkoğlu, H., Engin Koc, S., & Ersoy, H. (2024). Utilization of artificial intelligence in education: A perspective on learning strategies. In S. Kadry (Ed.), *Artificial intelligence for quality education* [Working Title]. IntechOpen. <https://doi.org/10.5772/intechopen.1005086>
- Yang, J., Zhang, X. L., & Su, P. (2018). Deep-learning-based agile teaching framework of software development courses in computer science education. *Procedia Computer Science*, 154, 137–145. <https://doi.org/10.1016/j.procs.2019.06.021>
- Yeo, M. A. (2023). Academic integrity in the age of artificial intelligence (AI) authoring apps. *TESOL Journal*, 14(3), e716. <https://doi.org/10.1002/tesj.716>
- Yıldırım A., & Şimsek, H. (2021). *Sosyal bilimlerde nitel araştırma yöntemleri* (12. baskı). Seçkin.
- Yuan, X. (2023). The impact of automated evaluation feedback on students' writing revision. *International Journal of Languages, Literature and Linguistics*, 9(6), 464. <https://doi.org/10.18178/ijlll.2023.9.6.464>
- Zawacki-Richter, O., Marín, V. I., & Bond, M. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39). <https://doi.org/10.1186/s41239-019-0171-0>
- Zhai, X., Chu, X., Chai, C. S., Yung Jong, M. S., Istenic, A., Spector, M., Liu, B., Yuan, J., & Li, Y. (2020). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021(1), 8812542. <https://doi.org/10.1155/2021/8812542>

Zhang, K., & Aslan, A. B. (2020). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/j.caeai.2021.100025>



Appendix A

The Research Articles Selected for the Synthesis

- Abuyaman, O. (2023). Strengths and weaknesses of ChatGPT models for scientific writing about medical vitamin B12: Mixed methods study. *JMIR Formative Research*, 7, Article e49459. <https://doi.org/10.2196/49459>
- Alafnan, M. A., & Mohd Zuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial Intelligence and Technology*, 3(3), 85–94. <https://doi.org/10.37965/jait.2023.0267>
- Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English with Technology*, 23(2), 25–43. <https://doi.org/10.56297/BUKA4060/XHLD5365>
- Algaraady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journal*, 9, 3-17. <https://doi.org/10.24093/awej/call9.1>
- Altmäe, S., Sola-Leyva, A., & Salumets, A. (2023). Artificial intelligence in scientific writing: A friend or a foe? *Reproductive BioMedicine Online*, 47(1), 3-9. <https://doi.org/10.1016/j.rbmo.2023.04.009>
- Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*, 9, Article 1347421. <https://doi.org/10.3389/feduc.2024.1347421>
- Ariyaratne, S., Iyengar, K. P., Nischal, N., Chitti Babu, N., & Botchu, R. (2023). A comparison of ChatGPT-generated articles with human-written articles.

Skeletal Radiology, 52(9), 1755-1758. <https://doi.org/10.1007/s00256-023-04340-5>

Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract? *Emergency Medicine Australasia*, 35(5), 809-811. <https://doi.org/10.1111/1742-6723.14233>

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21(1), Article 23. <https://doi.org/10.1186/s41239-024-00455-4>

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20(1), 1-24. <https://doi.org/10.1186/s41239-023-00427-0>

Bašić, Ž., Banovac, A., Kružić, I., & Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Humanities and Social Sciences Communications*, 10(1), 1-5. <https://doi.org/10.1057/s41599-023-02269-7>

Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *Journal of Educational Computing Research*, 1-29. <https://doi.org/10.1177/07356331241247465>

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, 86(3), 257-295. <https://doi.org/10.1177/23294906231176517>

- Cheng, S. L., Tsai, S. J., Bai, Y. M., Ko, C. H., Hsu, C. W., Yang, F. C., Tsai, C. K., Tu, Y. K., Yang, S. N., Tseng, P. T., Hsu, T. W., Liang, C. S., & Su, K. P. (2023). Comparisons of quality, correctness, and similarity between ChatGPT-generated and human-written abstracts for basic research: Cross-sectional study. *Journal of Medical Internet Research*, 25, e51229. <https://doi.org/10.2196/51229>
- Cingillioglu, I. (2023). Detecting AI-generated essays: The ChatGPT challenge. *The International Journal of Information and Learning Technology*, 40(3), 259–268. <https://doi.org/10.1108/IJILT-03-2023-0043>
- Črček, N., & Patekar, J. (2023). Writing with AI: University students' use of ChatGPT. *Journal of Language and Education*, 9, 128-138. <https://doi.org/10.17323/jle.2023.17379>
- Dağcı, M., Çam, F., & Dost, A. (2023). Reliability and quality of the nursing care planning texts generated by ChatGPT. *Nurse Educator*, 49(3), e109–e114. <https://doi.org/10.1097/NNE.0000000000001566>
- Dashti, M., Londono, J., Ghasemi, S., & Moghaddasi, N. (2023). How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing? *Journal of Prosthetic Dentistry*. Advance online publication. <https://doi.org/10.1016/j.prosdent.2023.05.023>
- Dong, L. (2024). “Brave new world” or not?: A mixed-methods study of the relationship between second language writing learners' perceptions of ChatGPT, behaviors of using ChatGPT, and writing proficiency. *Current Psychology*, 43, 1-15. <https://doi.org/10.1007/s12144-024-05728-9>
- Draxler, F., Werner, A., Lehmann, F., Hoppe, M., Schmidt, A., Buschek, D., & Welsch, R. (2024). The AI ghostwriter effect: When users do not perceive

ownership of AI-generated text but self-declare as authors. *ACM Transactions on Computer-Human Interaction*, 31(2), 1–40.

<https://doi.org/10.1145/3637875>

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal of Educational Integrity*, 19, Article 17.

<https://doi.org/10.1007/s40979-023-00140-5>

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 1–20.

<https://doi.org/10.1186/s41239-023-00425-2>

Fyfe, P. (2022). How to cheat on your final paper: Assigning AI for student writing. *AI & Society*, 38, 1395–1405. <https://doi.org/10.1007/s00146-022-01397-z>

Gallagher, J., & Wagner, K. (2024). Comparing student and writing instructor perceptions of academic dishonesty when collaborators are artificial intelligence or human. *Journal of Business and Technical Communication*, 38(3), 266–288. <https://doi.org/10.1177/10506519241239937>

Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. Advance online publication.

<https://doi.org/10.1177/13621688241239764>

Guerra, G. A., Hofmann, H., Sobhani, S., Hofmann, G., Gomez, D., Soroudi, D., Hopkins, B. S., Dallas, J., Pangal, D. J., Cheok, S., Nguyen, V. N., Mack, W. J., & Zada, G. (2023). GPT-4 artificial intelligence model outperforms

ChatGPT, medical students, and neurosurgery residents on neurosurgery written board-like questions. *World Neurosurgery*, 179, e160-e165.

<https://doi.org/10.1016/j.wneu.2023.08.042>

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29, 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>

Gustilo, L., Ong, E., & Lapinid, M. R. (2024). Algorithmically-driven writing and academic integrity: Exploring educators' practices, perceptions, and policies in the AI era. *International Journal for Educational Integrity*, 20(3), 1–43. <https://doi.org/10.1007/s40979-024-00153-8>

Hakam, H., Prill, R., Korte, L., Lovreković, B., Ostojic, M., Ramadanov, N., & Mühlensiepen, F. (2024). Human-written vs AI-generated texts in orthopedic academic literature: Comparative qualitative analysis. *JMIR Formative Research*, 8, e52164. <https://doi.org/10.2196/52164>

Hamamah, H., Emaliana, I., Hapsari, Y., Degeng, P., & Fadillah, A. (2023). Using nominal group technique to explore publication challenges and the usefulness of AI-based writing technologies: Insights from Indonesian scholars. *Theory and Practice in Language Studies*, 13, 2038-2047. <https://doi.org/10.17507/tpls.1308.20>

Hayawi, K., Shahriar, S., & Mathew, S. (2023). The imitation game: Detecting human and AI-generated texts in the era of large language models. *Journal of Information Science*, 1–36. Advance online publication. <https://doi.org/10.1177/01655515241227531>

- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13(1), 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Hong, J. W., Chang, H. C. H., & Tewksbury, D. (2024). Can AI become Walter Cronkite? Testing the machine heuristic, the hostile media effect, and political news written by artificial intelligence. *Digital Journalism*, 1–24. <https://doi.org/10.1080/21670811.2024.2323000>
- Hsiao, J. C., & Chang, J. S. (2023). Enhancing EFL reading and writing through AI-powered tools: Design, implementation, and evaluation of an online course. *Interactive Learning Environments*, 1-16. <https://doi.org/10.1080/10494820.2023.2207187>
- Huang, Z. (2023). An intelligent scoring system for English writing based on artificial intelligence and machine learning. *International Journal of System Assurance Engineering and Management*. Advance online publication. <https://doi.org/10.1007/s13198-023-02105-w>
- Huespe, I. A., Echeverri, J., Khalid, A., Carboni Bisso, I., Musso, C. G., Surani, S., Bansal, V., & Kashyap, R. (2023). Clinical research with large language models generated writing-Clinical Research with AI-assisted Writing (CRAW) Study. *Critical Care Explorations*, 5(10), e0975. <https://doi.org/10.1097/CCE.0000000000000975>
- Hwang, T., Aggarwal, N., Khan, P., Roberts, T., Mahmood, A., Griffiths, M., Parsons, N., & Khan, S. (2024). Can ChatGPT assist authors with abstract writing in medical journals? Evaluating the quality of scientific abstracts generated by ChatGPT and original abstracts. *PLoS One*, 19, e0297701. <https://doi.org/10.1371/journal.pone.0297701>

- Hwang, W.-Y., Nurtantyana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., & Surjono, H. D. (2023). AI and recognition technologies to facilitate English as a foreign language writing for supporting personalization and contextualization in authentic contexts. *Journal of Educational Computing Research*, 61(5), 1008-1035. <https://doi.org/10.1177/07356331221137253>
- Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines.” *Language Testing in Asia*, 13, Article 46. <https://doi.org/10.1186/s40468-023-00260-2>
- Ivanović, I. (2023). Can AI-assisted essay assessment support teachers? A cross-sectional mixed-methods research conducted at the University of Montenegro. *Annales, Series Historia et Sociologia*, 33(3), Article 30. <https://doi.org/10.19233/ASHS.2023.30>
- Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. *Languages*, 8(4), Article 247. <https://doi.org/10.3390/languages8040247>
- Kaharuddin, A., Ahmad, D., Mardiana, Latif, I., Arafah, B., & Suryadi, R. (2024). Defining the role of artificial intelligence in improving English writing skills among Indonesian students. *Journal of Language Teaching and Research*, 15, 568-678. <https://doi.org/10.17507/jltr.1502.25>
- Kim, N. J., & Kim, M. K. (2022). Teacher’s perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in Education*, 7, Article 755914. <https://doi.org/10.3389/educ.2022.755914>
- Köbis, N. C., & Mossink, L. (2020). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from

human-written poetry. *Computers in Human Behavior*, 114, Article 106553.

<https://doi.org/10.1016/j.chb.2020.106553>

Krajka, J., & Olszak, I. (2024). “AI, will you help?” How learners use artificial intelligence when writing. *XLinguae*, 17, 34-48.

<https://doi.org/10.18355/XL.2024.17.01.03>

Lee, Y.-J., Davis, R., & Lee, S. (2024). University students' perceptions of artificial intelligence-based tools for English writing courses. *Online Journal of Communication and Media Technologies*, 14, Article 202412.

<https://doi.org/10.30935/ojcmt/14195>

Leong, A. (2023). Clause complexing in research-article abstracts: Comparing human- and AI-generated texts. *ExELL*, 11, 99-132.

<https://doi.org/10.2478/exell-2023-0008>

Levin, G., Meyer, R., Guigue, P. A., & Brezinov, Y. (2024). It takes one to know one: Machine learning for identifying OBGYN abstracts written by ChatGPT. *International Journal of Gynaecology & Obstetrics*, 165(3), 1257-1260.

<https://doi.org/10.1002/ijgo.15365>

Levin, G., Pareja, R., Viveros-Carreño, D., Sanchez Diaz, E., Yates, E. M., Zand, B., & Ramirez, P. T. (2024). Association of reviewer experience with discriminating human-written versus ChatGPT-written abstracts. *International Journal of Gynecological Cancer*, 34(5), 669-674.

<https://doi.org/10.1136/ijgc-2023-005162>

Li, X., Li, B., & Cho, S.-J. (2023). Empowering Chinese language learners from low-income families to improve their Chinese writing with ChatGPT's assistance afterschool. *Languages*, 8(4), 238.

<https://doi.org/10.3390/languages8040238>

- Liao, W., Liu, Z., Dai, H., Xu, S., Wu, Z., Zhang, Y., Huang, X., Zhu, D., Cai, H., Li, Q., Liu, T., & Li, X. (2023). Differentiating ChatGPT-generated and human-written medical texts: Quantitative study. *JMIR Medical Education*, 9, e48904. <https://doi.org/10.2196/48904>
- Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2021). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, 31, 1-19. <https://doi.org/10.1080/10494820.2021.2012812>
- Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, 211, Article 104977. <https://doi.org/10.1016/j.compedu.2023.104977>
- Lozic, E., & Štular, B. (2023). Fluent but not factual: A comparative analysis of ChatGPT and other AI chatbots' proficiency and originality in scientific writing for humanities. *Future Internet*, 15, 336. <https://doi.org/10.3390/fi15100336>
- Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment & Evaluation in Higher Education*, 49(5), 616–633. <https://doi.org/10.1080/02602938.2024.2301722>
- Lu, X. (2019). An empirical study on the artificial intelligence writing evaluation system in China CET. *Big Data*, 7(2), 121-129. <https://doi.org/10.1089/big.2018.0151>

- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mahmud, F. (2023). Investigating EFL students' writing skills through artificial intelligence: Wordtune application as a tool. *Journal of Language Teaching and Research*, 14(5), 1395-1404. <https://doi.org/10.17507/jltr.1405.28>
- Mahyoob, M., Algaraady, J., & Alblwi, A. (2023). A proposed framework for human-like language processing of ChatGPT in academic writing. *International Journal of Emerging Technologies in Learning (iJET)*, 18(14), 282-293. <https://doi.org/10.3991/ijet.v18i14.41725>
- Malik, A., Pratiwi, Y., Andajani, K., Numertayasa, I., Suharti, S., Darwis, A., & Marzuki, M. (2023). Exploring artificial intelligence in academic essay: Higher education student's perspective. *International Journal of Educational Research Open*, 5, 100296. <https://doi.org/10.1016/j.ijedro.2023.100296>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2). Article 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
- Matthews, J., & Volpe, C. (2023). Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game. *Australasian Journal of Educational Technology*, 39, 82-100. <https://doi.org/10.14742/ajet.8896>
- Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning & Teaching*, 6(2), 1–9. <https://doi.org/10.37074/jalt.2023.6.2.10>

- Nazari, N., Shabbir, M., & Setiawan, R. (2021). Application of artificial intelligence-powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49(5), 847–864. <https://doi.org/10.1080/03075079.2024.2323593>
- Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Ur Rahman, S. I. (2024). Is ChatGPT a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning*, 40(2), 919–930. <https://doi.org/10.1111/jcal.12929>
- Nugroho, A., Andriyanti, E., Widodo, P., & Mutiaraningrum, I. (2024). Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing. *Innovations in Education and Teaching International*, 1-13. <https://doi.org/10.1080/14703297.2024.2319184>
- Papakonstantinidis, S., Kwiatek, P., & Spathopoulou, F. (2024). Embrace or resist? Drivers of artificial intelligence writing software adoption in academic and non-academic contexts. *Contemporary Educational Technology*, 16(2), 495. <https://doi.org/10.30935/cedtech/14250>
- Parker, J. L., Becker, K., & Carroca, C. (2023). ChatGPT for automated writing evaluation in scholarly writing instruction. *Journal of Nursing Education*, 62(12), 721-727. <https://doi.org/10.3928/01484834-20231006-02>
- Pellas, N. (2023). The effects of generative AI platforms on undergraduates' narrative intelligence and writing self-efficacy. *Education Sciences*, 13(11), 1155. <https://doi.org/10.3390/educsci13111155>

- Perkins, M., Roe, J., Postma, D., McGaughan, J., & Hickerson, D. (2023). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics*, 22, 1-25. <https://doi.org/10.1007/s10805-023-09492-6>
- Pigg, S. (2024). Research writing with ChatGPT: A descriptive embodied practice framework. *Computers and Composition*, 71, Article 102830. <https://doi.org/10.1016/j.compcom.2024.102830>
- Popkov, A. A., & Barrett, T. S. (2024). AI vs academia: Experimental study on AI text detectors' accuracy in behavioral health academic writing. *Accountability in Research*, 1-17. <https://doi.org/10.1080/08989621.2024.2331757>
- Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—A case study. *Smart Learning Environments*, 11(1), 1-18. <https://doi.org/10.1186/s40561-024-00296-8>
- Rababah, L., Rababah, M., & Al-khawaldeh, N. (2024). Graduate students' ChatGPT experience and perspectives during thesis writing. *International Journal of Engineering Pedagogy (iJEP)*, 14(3), 22-35. <https://doi.org/10.3991/ijep.v14i3.48395>
- Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 1–21. <https://doi.org/10.1080/10494820.2023.2208170>
- Rahayu, A., Weda, S., Muliati, M., & Vega, N. (2024). Artificial intelligence in writing instruction: A self-determination theory perspective. *XLinguae*, 17, 234-244. <https://doi.org/10.18355/XL.2024.17.01.16>

- Rojas, A. J. (2024). An investigation into ChatGPT's application for a scientific writing assignment. *Journal of Chemical Education*, 101(5), 1959–1965. <https://doi.org/10.1021/acs.jchemed.4c00034>
- Semrl, N., Feigl, S., Taumberger, N., Bracic, T., Fluhr, H., Blockeel, C., & Kollmann, M. (2023). AI language models in human reproduction research: Exploring ChatGPT's potential to assist academic writing. *Human Reproduction*, 38(12), 2281–2288. <https://doi.org/10.1093/humrep/dead207>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, Article 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Tang, A., Li, K., Kwok, K. O., Cao, L., Luong, S., & Tam, W. (2024). The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing. *Journal of Nursing Scholarship*, 56(2), 314–318. <https://doi.org/10.1111/jnu.12938>
- Tossell, C. C., Tenhundfeld, N. L., Momen, A., Cooley, K., & de Visser, E. J. (2024). Student perceptions of ChatGPT use in a college essay assignment: Implications for learning, grading, and trust in artificial intelligence. *IEEE Transactions on Learning Technologies*, 17, 1069–1081. <https://doi.org/10.1109/TLT.2024.3355015>

- Tsao, J., & Nogues, C. (2024). Beyond the author: Artificial intelligence, creative writing, and intellectual emancipation. *Poetics*, 102, Article 101865. <https://doi.org/10.1016/j.poetic.2024.101865>
- Utami, S., Andayani, A., Winarni, R., & Sumarwati, S. (2023). Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive? *Contemporary Educational Technology*, 15(4), article ep450. <https://doi.org/10.30935/cedtech/13419>
- Vicente-Yagüe Jara, M. I., Martínez, O., Navarro, V., & Cuéllar Santiago, F. (2023). Escritura, creatividad e inteligencia artificial. ChatGPT en el contexto universitario [Writing, creativity, and artificial intelligence: ChatGPT in the university context]. *Comunicar*, 77, 47-57. <https://doi.org/10.3916/C77-2023-04>
- Wale, B. D., & Kassahun, Y. F. (2023). The transformative power of AI writing technologies: Enhancing EFL writing instruction through the integrative use of Writerly and Google Docs. *Human Behavior and Emerging Technologies*, 2024(1), Article 9221377. <https://doi.org/10.1155/2024/9221377>
- Waltzer, T., Cox, R., & Heyman, G. (2023). Testing the ability of teachers and students to differentiate between essays generated by ChatGPT and high school students. *Human Behavior and Emerging Technologies*, 2023, Article 1923981, 1–9. <https://doi.org/10.1155/2023/1923981>
- Wang, Y., Yasmin, F., & Akbar, A. (2023). Impact of the internet on English language learning among university students: Mediating role of academic self-efficacy. *Frontiers in Psychology*, 14, 1184185. <https://doi.org/10.3389/fpsyg.2023.1184185>

- Wang, Z. (2020). Computer-assisted EFL writing and evaluations based on artificial intelligence: A case from a college reading and writing course. *Library Hi Tech*, 40(1), 80–97. <https://doi.org/10.1108/LHT-05-2020-0113>
- Watts, F., Dood, A., Shultz, G., & Rodriguez, J.-M. (2023). Comparing student and generative artificial intelligence chatbot responses to organic chemistry writing-to-learn assignments. *Journal of Chemical Education*, 100(10), 3806–3817. <https://doi.org/10.1021/acs.jchemed.3c00664>
- Wee, H. B., & Reimer, J. (2023). Non-English academics face inequality via AI-generated essays and countermeasure tools: Countermeasure detection tool testing. *BioScience*, 73, 1-3. <https://doi.org/10.1093/biosci/biad034>
- Wolf, R., & Wolf, A. (2023). Using AI to evaluate a competency-based online writing course in nursing. *Online Learning*, 27(3), 41–69. <https://doi.org/10.24059/olj.v27i3.3974>
- Wu, R. T., & Dang, R. R. (2023). ChatGPT in head and neck scientific writing: A precautionary anecdote. *American Journal of Otolaryngology*, 44(6), 103980. <https://doi.org/10.1016/j.amjoto.2023.103980>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>
- Yang, H., Gao, C., & Shen, H. (2024). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. *Education and Information Technologies*, 29, 3837–3858. <https://doi.org/10.1007/s10639-023-11991-3>

- Yeadon, W., Inyang, O., Mizouri, A., Peach, A., & Testrow, C. (2022). The death of the short-form physics essay in the coming AI revolution. *Journal of Physics Education*, 10. <https://doi.org/10.48550/arXiv.2212.11661>
- Zaitso, W., & Jin, M. (2023). Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis. *PLoS One*, 18(8), e0288453. <https://doi.org/10.1371/journal.pone.0288453>
- Zhao, R., Zhuang, Y., Zou, D., Xie, Q., & Yu, P. (2022). AI-assisted automated scoring of picture-cued writing tasks for language assessment. *Education and Information Technologies*, 28, 7031–7063. <https://doi.org/10.1007/s10639-022-11473-y>
- Zhao, X., Cox, A., & Cai, L. (2024). ChatGPT and the digitisation of writing. *Humanities and Social Sciences Communications*, 11, Article 482. <https://doi.org/10.1057/s41599-024-02904-x>
- Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System*, 118, Article 103141. <https://doi.org/10.1016/j.system.2023.103141>
- Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology*, 14, 1259531. <https://doi.org/10.3389/fpsyg.2023.1259531>