T.C.

GÜMÜŞHANE ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÖNETİM VE BİLİŞİM SİSTEMLERİ ANA BİLİM DALI

BÜYÜK DİL MODELİ (LLM) KULLANARAK X (TWITTER) VERİLERİ TABANLI YEREL İŞLENEBİLİR BİR SİBER TEHDİT İSTİHBARATI MODELİ TASARIMI

YÜKSEK LİSANS

SAİT EMRE ORAL

Temmuz 2026

GÜMÜŞHANE

T.C.

GÜMÜŞHANE ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÖNETİM VE BİLİŞİM SİSTEMLERİ ANA BİLİM DALI

BÜYÜK DİL MODELİ (LLM) KULLANARAK X (TWİTTER) VERİLERİ TABANLI YEREL İŞLENEBİLİR BİR SİBER TEHDİT İSTİHBARATI MODELİ TASARIMI

DESIGN OF AN ON-PREMISE AND ACTIONABLE CYBER THREAT INTELLIGENCE MODEL BASED ON X (TWITTER) DATA USING LARGE LANGUAGE MODELS (LLMS)

YÜKSEK LİSANS

SAİT EMRE ORAL

Temmuz 2026

GÜMÜŞHANE

GÜMÜŞHANE ÜNİVERSİTESİ
Lisansüstü Eğitim Enstitüsü

T.C.

GÜMÜŞHANE ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÖNETİM VE BİLİŞİM SİSTEMLERİ ANA BİLİM DALI

BÜYÜK DİL MODELİ (LLM) KULLANARAK X (TWİTTER) VERİLERİ TABANLI YEREL İŞLENEBİLİR BİR SİBER TEHDİT İSTİHBARATI MODELİ TASARIMI

DESIGN OF AN ON-PREMISE AND ACTIONABLE CYBER THREAT INTELLIGENCE MODEL BASED ON X (TWITTER) DATA USING LARGE LANGUAGE MODELS (LLMS)

YÜKSEK LİSANS

SAİT EMRE ORAL

DANIŞMAN: PROF. DR. HANDAN ÇAM

Temmuz 2026

GÜMÜŞHANE

KABUL VE ONAY

Prof. Dr. Handan ÇAM danışmanlığında, **Sait Emre ORAL** tarafından hazırlanan "**Büyük Dil Modeli (Llm) Kullanarak X (Twitter) Verileri Tabanlı Yerel İşlenebilir Bir Siber Tehdit İstihbaratı Modeli Tasarımı**" isimli bu çalışma, 03/07/2026 tarihinde yapılan lisansüstü tez savunma sınavı sonucunda **Oy Birliği** ile başarılı bulunarak jürimiz tarafından **Yüksek Lisans Tezi** olarak kabul edilmiştir.

…….

Doç. Dr. Ahmet Kamil KABAKUŞ (Başkan)

…….

Prof. Dr. Handan ÇAM (Danışman)

…….

Dr. Öğr. Üyesi Uğur DEMİREL (Üye)

Lisansüstü tez savunma sınavında başarılı bulunarak kabul edilen bu tezin ciltlenmiş hali, ... /..... /…… tarihli ve …… / …… sayılı Enstitü Yönetim Kurulu toplantısında görüşülmüş ve tez yazım kılavuzuna uygun bulunarak onaylanmıştır.

Prof. Dr. Duygu ÖZDEŞ

Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Yüksek Lisans Tezi olarak hazırlamış olduğum "**Büyük Dil Modeli (Llm) Kullanarak X(Twitter) Verileri Tabanlı Yerel İşlenebilir Bir Siber Tehdit İstihbaratı Modeli Tasarımı**" isimli bu tezimin, tamamen kendi çalışmam olduğunu, danışmanımın sorumluluğunda hazırladığımı, her alıntıya kaynak gösterdiğimi, alıntı yaptığım tüm çalışmaları kaynakçada belirtiğimi ve Gümüşhane Üniversitesi'nin lisanslı kullanıcısı olduğu intihal yazılım programı ile Lisansüstü Eğitim Enstitüsü'nün belirlediği kıstaslara uygun olarak raporladığımı taahhüt ederim. Tezimin kâğıt ve elektronik kopyalarının Gümüşhane Üniversitesi Lisansüstü Eğitim Enstitüsü arşivinde saklanmasına izin verdiğimi onaylarım.

03/07/2026

…….

Sait Emre ORAL

TEŞEKKÜR

Yüksek lisans eğitimim ve tez çalışmam boyunca bilgi ve birikimini benden esirgemeyen, araştırmanın her aşamasında değerli yönlendirmeleriyle ufkumu açan ve akademik gelişimime büyük katkı sağlayan kıymetli danışman hocam Sayın Prof. Dr. Handan ÇAM'a en içten teşekkürlerimi sunarım.

Tez yazım sürecindeki yapıcı eleştirileri, değerli geri bildirimleri ve esirgemediği akademik desteği ile çalışmanın bilimsel temellerinin ve metodolojik yapısının güçlenmesine önemli katkılarda bulunan Sayın Dr. Öğr. Üyesi Uğur DEMİREL'e şükranlarımı sunarım.

Son olarak, bu meşakkatli çalışma süreci boyunca ailemize ayırmam gereken vakti tezime vakfetmemi büyük bir hoşgörüyle karşılayan; verdikleri koşulsuz sevgi, sabır ve manevi destekle arkamda sarsılmaz bir güç olan kıymetli eşime ve sevgili oğluma en içten şükranlarımı sunarım.

.....

Sait Emre ORAL

GÜMÜŞHANE – 2026

ÖZET

Günümüzdeki siber tehdit istihbaratı (CTI) paylaşımları, çoğunlukla tek başına bir IP adresi veya dosya özeti gibi, tehdidin tam hikayesinden kopuk verilerle sınırlı kalmaktadır. Bu durum, savunma ekiplerinin hangi tehdide öncelik vereceğini belirlemesini zorlaştırmaktadır. X (Twitter) platformu, tehditler hakkında aslında çok daha zengin ve bağlamsal detaylar sunsa da; bu verinin devasa hacmi, gürültülü ve yapılandırılmamış doğası, geleneksel yöntemlerle işlenmesini engellemektedir. Bu tez çalışmasında, X (Twitter) verilerini kullanarak veri egemenliğini koruyan, yerel (on-premise) çalışabilen ve eyleme geçirilebilir sonuçlar üreten, uçtan uca otonom bir CTI modelinin tasarlanması amaçlanmıştır.

Geliştirilen Python tabanlı otonom sistem; Selenium kütüphanesi ile veri toplama, Ollama platformu üzerinden yerel olarak çalıştırılan LLaMA 3 modeli ile analiz ve SQLite veritabanında depolama süreçlerini içeren bütüncül bir mimariye sahiptir. Çalışmanın en özgün mühendislik katkısı, Büyük Dil Modellerinin (LLM) deterministik olmayan doğasından kaynaklanan hatalı JSON çıktılarını otonom olarak düzelten "JSON Sağlamlaştırma Protokolü"nün geliştirilmesidir.

Elde edilen bulgular, sistemin yapılandırılmamış 62 adet tweet metni üzerinden tehditleri %95'e varan doğruluk oranıyla analiz ettiğini ve tespit edilen içeriklerin %64,5'ini "Yüksek Risk" (High Risk) kategorisinde başarıyla sınıflandırdığını göstermiştir. Yapılan analizlerde "Vulnerability" (13 adet) ve "Threat_Actor" (13 adet) kategorileri en sık rastlanan tehdit türleri olarak öne çıkarken; sistem, CVE numaraları ve tehdit aktörü grupları gibi kritik istihbarat göstergelerini yüksek hassasiyetle ayrıştırmıştır. Bu çalışma, güvenlik açısından kritik verilerin üçüncü taraf sunuculara gönderilmesi riskini ortadan kaldıran, mühendislik açısından dayanıklı ve tam otonom bir CTI çözümünün operasyonel fizibilitesini kanıtlamaktadır.

Anahtar Kelimeler: Büyük Dil Modelleri (LLM), JSON, Siber Tehdit İstihbaratı (CTI), X (Twitter), Yerel (On-premise).

ABSTRACT

Currently, cyber threat intelligence (CTI) sharing is mostly limited to isolated data points, such as a single IP address or a file hash, which are disconnected from the full context of the threat. This fragmentation makes it difficult for defense teams to determine accurate prioritization. Although the X (Twitter) platform offers much richer and more detailed information about emerging threats, the sheer volume, noise, and unstructured nature of this data prevent it from being processed through traditional analytical methods. This thesis aims to design an end-to-end autonomous CTI model that utilizes X (Twitter) data, strictly maintains data sovereignty by operating locally (on-premise), and produces actionable intelligence.

The developed Python-based autonomous system features a comprehensive architecture that includes dynamic data collection via Selenium, semantic analysis using the LLaMA 3 model running locally through the Ollama platform, and structured storage within an SQLite database. The most original engineering contribution of this research is the development of a "JSON Hardening Protocol," which autonomously corrects the erroneous JSON outputs resulting from the stochastic and unpredictable nature of Large Language Models (LLMs).

Experimental findings demonstrate that the system analyzed threats from 62 unstructured tweet texts with an accuracy rate of up to 95%, successfully classifying 64,5% of the detected events into the "High Risk" category. During the evaluations, "Vulnerability" (13 instances) and "Threat_Actor" (13 instances) emerged as the most frequently detected threat categories, while the system extracted critical intelligence indicators—such as CVE IDs and Threat Actor Groups—with high precision. Ultimately, this study proves the operational feasibility of an engineering-resilient and fully autonomous CTI solution that completely eliminates the risk of exposing security-critical data to third-party servers.

Keywords: Large Language Models (LLM), JSON, Cyber Threat Intelligence (CTI), X (Twitter), On-premise.

İÇİNDEKİLER

TABLOLAR DİZİNİ

ŞEKİLLER DİZİNİ

SİMGELER VE KISALTMALAR DİZİNİ

API	: Application Programming Interface (Uygulama Programlama Arayüzü)
APT	: Advanced Persistent Threat (Gelişmiş Kalıcı Tehdit)
ATTveCK	: Adversarial Tactics, Techniques, and Common Knowledge (Saldırgan Taktikleri, Teknikleri ve Ortak Bilgi Birikimi)
BEC	: Business Email Compromise (İş E-postası Tehdidi)
BERT	: Bidirectional Encoder Representations from Transformers
CTI	: Cyber Threat Intelligence (Siber Tehdit İstihbaratı)
CV	: Confidence Value (Güven Değeri / Skoru)
CVE	: Common Vulnerabilities and Exposures (Ortak Zafiyetler ve Maruziyetler)
GDPR	: General Data Protection Regulation (Genel Veri Koruma Yönetmeliği)
GUI	: Graphical User Interface (Grafiksel Kullanıcı Arayüzü)
IOC	: Indicator of Compromise (Saldırı / Tehdit Göstergesi)
IoT / IIoT	: Internet of Things / Industrial Internet of Things (Nesnelerin İnterneti / Endüstriyel Nesnelerin İnterneti)
IP	: Internet Protocol (İnternet Protokolü)
JSON	: JavaScript Object Notation (JavaScript Nesne Gösterimi)
KVKK	: Kişisel Verilerin Korunması Kanunu
LLM	: Large Language Models (Büyük Dil Modelleri)
MTTR	: Mean Time to Respond (Ortalama Müdahale Süresi)
NLP	: Natural Language Processing (Doğal Dil İşleme)
NVD	: National Vulnerability Database (Ulusal Zafiyet Veritabanı)
OSINT	: Open Source Intelligence (Açık Kaynak İstihbaratı)
PII	: Personally Identifiable Information (Kişisel Tanımlayıcı Bilgi)
REBEL	: Relation Extraction By End-to-end Language Generation (Uçtan Uca Dil Üretimi ile İlişki Çıkarımı)
Regex	: Regular Expression (Düzenli İfade)
SIEM	: Security Information and Event Management (Güvenlik Bilgileri ve Olay Yönetimi)
SMTP	: Simple Mail Transfer Protocol (Basit Posta Aktarım Protokolü)
SOAR	: Security Orchestration, Automation and Response (Güvenlik Orkestrasyonu, Otomasyonu ve Müdahalesi)
SOC	: Security Operations Center (Güvenlik Operasyon Merkezi)

SRL : Semantic Role Labeling (Anlamsal Rol Etiketleme)

STIX : Structured Threat Information eXpression (Yapılandırılmış Tehdit Bilgisi İfadesi)

TTP : Tactics, Techniques, and Procedures (Taktik, Teknik ve Prosedürler)

UIE : Universal Information Extraction (Evrensel Bilgi Çıkarımı)

WPA3 : Wi-Fi Protected Access 3 (Wi-Fi Korumalı Erişim 3)

1. GİRİŞ

Günümüzde dijitalleşme ve birbirine bağlı sistemlerin hızla yaygınlaşması, kurumların karşı karşıya kaldığı siber tehdit yüzeyini benzeri görülmemiş bir boyuta ulaştırmıştır. Karmaşıklaşan ve otomatize hale gelen bu yeni nesil tehditler karşısında, yalnızca olay anında devreye giren geleneksel ve reaktif savunma mekanizmaları artık yetersiz kalmaktadır. Bu nedenle siber güvenlik dünyasında sadece tespit ve müdahaleye odaklanmak yerine, saldırı henüz gerçekleşmeden bir adım önde konumlanarak proaktif bir savunma duruşu sergilemek temel bir gereksinim haline gelmiştir (Urgan ve Erdoğan, 2024). Kurumları hedef alan siber tehditlerin niyet, kapasite ve taktikleri hakkında toplanan verilerin analiz edilerek eyleme geçirilebilir bilgiye dönüştürülmesi süreci olan Siber Tehdit İstihbaratı (CTI – Cyber Threat Intelligence) tam olarak bu amaca hizmet etmektedir. Ancak günümüzde CTI olarak paylaşılan tehdit verisi, çoğu zaman tek başına bir Saldırı Göstergesi (IOC - Indicator of Compromise) olan bir IP adresi veya dosya özeti gibi, tehdidin tam hikayesinden kopuk ve yalıtılmış parçalardan oluşmaktadır. Bu durum, savunma ekiplerinin "Bu tehdit ne kadar ciddi?" veya "Hangi birimler risk altında?" sorularına tatmin edici cevaplar bulmasını ve tehditleri doğru önceliklendirmesini zorlaştırmaktadır (Huff ve Li, 2021). Aslında bu eksik olan zengin bağlamsal detaylar, siber güvenlik araştırmacılarının anlık duyurular yaptığı X (Twitter) gibi herkese açık platformlarda (OSINT - Açık Kaynak İstihbaratı) her gün yoğun bir şekilde paylaşılmaktadır (Dale vd., 2025). Buradaki temel problem, bu değerli istihbaratın devasa bir gürültü yığını içinde, düzensiz ve yapılandırılmamış bir formatta bulunmasıdır. Bu verinin geleneksel kural tabanlı yöntemlerle veya manuel süreçlerle ayıklanması, operasyonel hız gereksinimleri göz önüne alındığında neredeyse imkansızdır.

Büyük Dil Modelleri'nin (LLM – Large Language Models) yükselişi, bu yapılandırılmamış metinleri anlama ve işleme konusunda yeni bir kapı açmıştır (De Bellis, 2023). Bu modeller, doğru yönlendirmelerle, ham bir metinden JSON gibi yapılandırılmış, makinelerin doğrudan okuyabileceği formatlarda veri çıkarma potansiyeline sahiptir (Khatami, 2024). Ancak, bu konudaki mevcut literatür incelendiğinde üç temel araştırma boşluğu öne çıkmaktadır. Birincisi, çalışmaların çoğu analiz için ticari bulut API'larını kullanmaktadır ki bu durum, hassas CTI verileri için ciddi bir veri egemenliği ve gizlilik riski oluşturmaktadır. İkincisi, mevcut sistemler genellikle CTI döngüsünün sadece tek bir parçasına odaklanmakta; verinin toplanmasından anlık uyarı üretilmesine kadar tüm süreci kapsayan uçtan uca bütüncül

sistemler sunmamaktadır. Üçüncüsü, LLM'lerin öngörülemez doğası gereği ürettikleri hatalı veya bozuk JSON çıktılarını otonom olarak düzelten dayanıklı mühendislik çözümleri yetersiz kalmaktadır (Clement, 2025).

Bu tez çalışmasının temel amacı, literatürde tespit edilen bu üç temel boşluğu dolduran vizyoner bir model tasarlamak ve geliştirmektir. Çalışmada, X (Twitter) verilerini birincil kaynak olarak kullanan, uçtan uca otonom işleyen, veri egemenliğini koruyan ve mühendislik açısından yüksek dayanıklılığa sahip bir Siber Tehdit İstihbaratı (CTI) modeli ortaya konulmuştur. Geliştirilen bütüncül sistem; veri toplama aşamasında Selenium otomasyonunu kullanmakta, derinlemesine analiz için Ollama platformu üzerinden yerel altyapıda çalıştırılan LLaMA 3 modelini entegre etmekte ve LLM kaynaklı yapısal bozuklukları aşmak için özgün bir "JSON Sağlamlaştırma Protokolü" işletmektedir. Tüm bu karmaşık mimari, son kullanıcıya tek bir Grafiksel Kullanıcı Arayüzü (GUI – Graphical User Interface) üzerinden sunulmaktadır. Çalışmanın literatüre ve sektöre sağladığı en önemli katkı; veri gizliliğine öncelik veren kurumların, dışarıya veri sızdırma riski olmaksızın kendi kapalı altyapılarında CTI süreçlerini otomatikleştirmeleri için pratik, düşük maliyetli ve otonom bir mühendislik çerçevesi sunmasıdır.

Araştırma yürütülürken veri toplama sürecinde izlenen siber güvenlik odaklı X (Twitter) hesaplarından paylaşılan tehdit bildirimlerinin genel olarak doğru ve güvenilir olduğu varsayılmıştır. Çalışmanın bazı temel sınırlılıkları da bulunmaktadır. Geliştirilen modelin veri kaynağı yalnızca X (Twitter) platformu ile sınırlandırılmıştır. Ayrıca, analitik çekirdeği oluşturan yerel LLaMA 3 modeline, siber güvenlik terminolojisine özgü ek bir ince ayar (fine-tuning) işlemi uygulanmamış; modelin sıfır atışlı (zero-shot) öğrenme yetenekleri kullanılmıştır. Sistem performansı, yapılandırılmamış 62 adet tweet metni üzerinden bir doğrulama ve performans testlerine tabi tutulmuştur.

2. SİBER TEHDİT İSTİHBARATININ TEORİK TEMELLERİ VE BİLEŞENLERİ

Bu bölümde, siber tehdit istihbaratı (CTI), açık kaynak istihbaratı (OSINT) olarak sosyal medyanın kullanımı ve Büyük Dil Modelleri'nin (LLM) siber güvenlik süreçlerindeki rolü üzerine yapılan akademik çalışmalar incelenmiştir.

2.1. Siber Tehdit İstihbaratı (CTI)

Siber Tehdit İstihbaratı (CTI), organizasyonların siber savunma mekanizmalarını tahkim etmek amacıyla, tehdit aktörlerinin yetenekleri, fırsatları ve niyetleri hakkında analiz edilmiş, kanıta dayalı bilgilerin sistematik olarak toplanması ve işlenmesi süreci olarak tanımlanmaktadır (Lee, 2020). Geleneksel reaktif güvenlik yaklaşımlarının aksine, CTI tehditleri gerçekleşmeden önce öngörmeyi amaçlayan proaktif bir savunma duruşunu hedefler. Tehdit istihbaratının standartlaştırılmış formatlarda temsil edilmesi, özellikle büyük veri kümeleri üzerinde işleyen otonom analiz çözümlerinin kalitesini artırarak bu proaktif stratejiyi operasyonel düzeyde desteklemektedir (Tounsi, 2019). CTI, sağladığı bağlamın derinliğine, teknik kapsamına ve kurumsal hiyerarşide hitap ettiği karar verici kitleye göre literatürde genellikle üç temel operasyonel katmanda ele alınmaktadır. Bu katmanlı yapı, ham verinin rafine edilerek kurumun farklı birimleri için anlamlı ve eyleme geçirilebilir birer stratejik değer kazanmasını sağlamaktadır:

2.1.1. Stratejik İstihbarat

En üst düzey katman olarak, kurumsal düzeydeki üst düzey karar vericilere, yönetim kurulu üyelerine ve yöneticilere yönelik olarak kurgulanmaktadır. Bu düzeydeki istihbarat; teknik detaylardan ziyade, siber tehditlerin kurumun ticari itibarı, finansal sürdürülebilirliği ve yasal uyum süreçleri üzerindeki olası etkilerine odaklanmaktadır. Küresel siber güvenlik eğilimlerini ve jeopolitik riskleri analiz eden stratejik CTI, kurumun uzun vadeli güvenlik yatırımlarının planlanmasında ve bütçe yönetiminde temel bir referans noktası teşkil etmektedir.

2.1.2. Operasyonel İstihbarat

Doğrudan siber savunma ekiplerini, güvenlik operasyon merkezi (SOC) analistlerini ve olay müdahale uzmanlarını hedefleyen savunma odaklı bir katmandır. Bu aşamada temel odak noktası, saldırgan aktörlerin kimlikleri, niyetleri ve saldırı

operasyonlarında kullandıkları Taktik, Teknik ve Prosedürlerin (TTP) derinlemesine analiz edilmesidir. Operasyonel istihbarat, savunmacıların saldırganların "nasıl" ve "neden" saldırdığını anlamasına olanak tanıyarak proaktif savunma senaryolarının geliştirilmesini desteklemektedir.

2.1.3. Taktiksel ve Teknik İstihbarat

Siber olayların en somut ve anlık teknik kanıtlarını temsil eden Saldırı Göstergeleri (IOC) üzerine yoğunlaşmaktadır. Zararlı ağ adresleri (IP), alan adları (domain), kötü amaçlı yazılım dosya özetleri ve saldırı imzaları gibi düşük düzeyli teknik veriler bu kategoride değerlendirilmektedir. Bu düzeydeki istihbarat, özellikle Güvenlik Bilgi ve Olay Yönetimi (SIEM) veya Güvenlik Duvarı gibi otonom savunma sistemlerinin beslenmesinde ve bilinen tehditlerin gerçek zamanlı olarak engellenmesinde kritik bir rol oynamaktadır.

Bu üçlü mimari, kurumun hem makro düzeydeki risklerden korunmasını hem de mikro düzeydeki teknik saldırılara saniyeler içinde reaksiyon verebilmesini mümkün kılan bütüncül bir savunma kalkanı oluşturmaktadır.

2.2. OSINT Yaklaşımı

CTI ekosistemi kapsamında Açık Kaynak İstihbaratı (OSINT), özellikle henüz kamuya yeni ifşa edilen güvenlik açıkları (zero-day) ve aktif saldırı kampanyaları hakkında erken uyarı sağlamada kritik bir role sahiptir. Literatürdeki güncel araştırmalar, sosyal medya platformlarının, bilhassa X (Twitter) ve Reddit gibi ağların, siber güvenlik uzmanlarının anlık duyurular ve teknik analizler paylaştığı en dinamik istihbarat kaynakları olduğunu doğrulamaktadır (Jakstaite ve Czekster, 2023).

Geleneksel güvenlik bültenleri ve antivirüs üreticileri genellikle bir zafiyetin tespitinden günler sonra resmi raporlar yayınlarken, OSINT kaynakları tehdidin doğduğu sıfırıncı saniyede gayri resmi ancak eyleme geçirilebilir hayati veriler sunar.

Ancak, bu platformlardan elde edilen verinin yüksek hacimli, gürültülü ve yapılandırılmamış doğası, verimli bir analiz süreci için büyük bir operasyonel engel teşkil etmektedir. X (Twitter) akışlarında paylaşılan teknik bir zafiyet analizi veya saldırı duyurusu; çoğu zaman günlük konuşma diliyle, sektörel argolarla, ironiyle ve eksik bağlamlarla iç içe geçmiş durumdadır. Geleneksel Anahtar Kelime veya Düzenli İfade (Regex) tabanlı veri madenciliği araçları, bu samanlıkta iğne arama probleminde metnin bağlamını ve duygusunu kavrayamadıkları için yüksek oranda yanlış pozitif (false positive) veya yanlış negatif (false negative) sonuçlar üretirler.

Bu operasyonel darboğazı aşmak ve OSINT verisini hızla tüketilebilir hale getirmek adına, literatürde siber güvenlikle ilgili sosyal medya akışlarını analiz eden sistemler geliştirilmiştir. Bu çalışmalar, zararlı IP adresleri, komuta-kontrol alan adları veya kötü amaçlı yazılım özetleri gibi bir güvenlik ihlalinin somut teknik kanıtları olan Saldırı Göstergelerini (IOC) otonom olarak ayrıştırmayı hedeflemektedir (Purba ve Chu, 2023). Dolayısıyla, OSINT'in barındırdığı bu devasa erken uyarı potansiyelini operasyonel değere dönüştürmek; veriyi sadece "toplayan" değil, aynı zamanda bağlamı tıpkı bir insan gibi okuyup anlayan gelişmiş Doğal Dil İşleme (NLP) teknolojilerinin kullanımını mutlak surette zorunlu kılmaktadır.

2.3. Saldırı Göstergeleri (IOC) Nedir?

Saldırı Göstergeleri (Indicator of Compromise - IOC), siber güvenlik terminolojisinde bir ağın, sistemin veya cihazın yetkisiz bir erişime veya kötü amaçlı bir faaliyete maruz kaldığını kanıtlayan adli veriler olarak tanımlanmaktadır. Geleneksel olay müdahale süreçlerinde birer "dijital parmak izi" olarak da adlandırılan IOC'ler, bir siber saldırının teşhis edilmesinde, boyutunun belirlenmesinde ve olayların engellenmesinde kritik bir işlev görmektedir.

En yaygın karşılaşılan IOC türleri arasında; kötü amaçlı yazılımların iletişim kurduğu komuta-kontrol (C2) sunucularına ait IP adresleri ve alan adları, zararlı dosyalara ait şifreleme özetleri (MD5, SHA-256 hash değerleri), oltalama (phishing) amacıyla kullanılan şüpheli URL bağlantıları, sömürülen zafiyet kodları (CVE numaraları) ve sistemde izinsiz değişiklik yapıldığını gösteren olağandışı kayıt defteri anahtarları yer almaktadır.

CTI döngüsünde IOC'ler; stratejik ve operasyonel verilerin aksine, bilgisayar sistemleri tarafından doğrudan okunabilen ve güvenlik cihazlarına (Güvenlik Duvarı, IDS/IPS, SIEM) anında kural veya engelleme listesi olarak eklenebilen teknik yapıtaşlarıdır. Açık Kaynak İstihbaratı (OSINT) platformlarında, özellikle siber güvenlik araştırmacıları tarafından sosyal medya akışlarında anlık olarak paylaşılan bu göstergelerin, yapılandırılmamış gürültülü metinler içerisinden Büyük Dil Modelleri (LLM) aracılığıyla otonom olarak ayrıştırılması, kurumların yeni nesil tehditlere karşı müdahale süresini (MTTR) dramatik bir şekilde kısaltmaktadır.

2.4. MITRE ATTveCK Çerçevesi

Siber tehdit istihbaratının standartlaştırılması ve saldırgan davranışlarının sistematik bir şekilde analiz edilmesi süreçlerinde MITRE ATTveCK (Adversarial

Tactics, Techniques, and Common Knowledge) çerçevesi, dünya çapında kabul görmüş, dinamik ve kanıta dayalı bir bilgi tabanı olarak öne çıkmaktadır. Bu çerçeve, siber tehdit aktörlerinin bir saldırı yaşam döngüsü boyunca sergiledikleri davranışları, gerçek dünya gözlemlerine dayanarak "taktik", "teknik" ve "alt-teknik" seviyelerinde sınıflandıran kapsamlı bir taksonomi sunmaktadır (Mahboubi vd., 2024).

ATTveCK matrisi, saldırganın ulaşmak istediği nihai stratejik amacı temsil eden taktikler (örneğin; başlangıç erişimi, kalıcılık sağlama, ayrıcalık yükseltme) ile bu amaçlara ulaşmak için başvurulan somut yöntemleri ifade eden teknikler arasındaki mantıksal ilişkiyi tanımlamaktadır. Geleneksel Saldırı Göstergesi (IOC) tabanlı savunma yaklaşımlarının aksine ATTveCK çerçevesi, saldırganın metodolojisine (TTP) odaklanarak savunma birimlerine daha derin bir bağlamsal farkındalık sağlamaktadır. Bu çerçeve, toplanan ham istihbarat verilerinin saldırgan niyetleriyle ilişkilendirilmesine ve güvenlik operasyon merkezlerinin (SOC) düşman davranışlarını önceden tespit edebilecek proaktif tehdit avcılığı (threat hunting) stratejileri geliştirmesine olanak tanımaktadır (Brown ve Lee, 2019; Mahboubi vd., 2024).

Özellikle Büyük Dil Modelleri (LLM) kullanılarak gerçekleştirilen otonom analiz süreçlerinde, yapılandırılmamış açık kaynak istihbaratı (OSINT) verilerinin otomatik olarak MITRE ATTveCK teknikleriyle eşleştirilmesi ve spesifik tehdit aktörleriyle ilişkilendirilmesi (attribution), istihbaratın operasyonel değerini maksimize etmektedir (Guru vd., 2025). Bu otonom yaklaşımla, sosyal medya mecralarından elde edilen karmaşık bir zafiyet duyurusu veya saldırı haberi; sadece teknik bir veri olmaktan çıkarak, saldırganın hangi taktiksel aşamada olduğu ve hangi savunma mekanizmalarının aşılmaya çalışıldığı bilgisini sunan yapısal bir formata bürünmektedir. Elde edilen bu ATTveCK tabanlı yapısal istihbarat, otonom olay müdahale (incident response) süreçlerini ve tehdit hafifletme (mitigation) stratejilerini doğrudan tetikleyen, eyleme geçirilebilir bir savunma çıktısına dönüşmektedir (Tellache vd., 2025; Clement, 2025).

2.5. STIX (Structured Threat Information Expression) Standardı

Siber tehdit istihbaratının farklı kurumlar ve teknik sistemler arasında tutarlı, hızlı ve makine tarafından okunabilir bir formatta paylaşılması amacıyla geliştirilen STIX (Structured Threat Information eXpression), CTI ekosisteminin en temel yapısal dilini oluşturmaktadır. STIX, siber saldırıların teknik detaylarını, saldırgan profillerini ve savunma stratejilerini hiyerarşik bir şema ile tanımlayarak; farklı güvenlik yazılımlarının (SIEM, SOAR, EDR vb.) ortak bir veri dilinde haberleşmesine olanak tanımaktadır.

Geliştirilen otonom analiz modelinin çıktıları, özünde STIX standartlarının gerektirdiği yapısal veri disipliniyle paralellik göstermektedir. Ham OSINT verilerinden çıkarılan zafiyet kodları (CVE), tehdit aktörü bilgileri ve saldırı göstergeleri (IOC); STIX nesneleri (STIX Domain Objects) olarak tanımlanabilmekte ve bu sayede elde edilen istihbaratın evrensel bir geçerlilik kazanması sağlanmaktadır. Özellikle yerel Büyük Dil Modelleri (LLM) aracılığıyla üretilen yapılandırılmış JSON çıktılarının STIX formatına uyumlu bir hiyerarşide sunulması, sistemin sadece izole bir analiz aracı olarak değil, aynı zamanda küresel istihbarat paylaşım ağlarına entegre olabilen proaktif bir savunma bileşeni olarak işlev görmesini mümkün kılmaktadır.

3. DOĞAL DİL İŞLEME KAVRAMI VE İLGİLİ LİTERATÜR TARAMASI

Bu bölümde; Doğal Dil İşleme (NLP) ve Büyük Dil Modellerinin (LLM) teknolojik evrimi incelenerek ilgili literatür taranmış ve tezde geliştirilen otonom analiz modelinin bilimsel dayanakları ortaya konulmuştur.

3.1. Doğal Dil İşleme

Doğal Dil İşleme (NLP), bilgisayarların insan dilini anlama, yorumlama ve üretme kapasitesini inceleyen yapay zekâ alt disiplinidir. Yıllar boyunca kelime frekanslarına dayalı kural tabanlı ve istatistiksel modellerle yürütülen bu alan, dilin barındırdığı karmaşık mecazları ve uzak mesafe bağlamlarını yakalamada çoğunlukla yetersiz kalmıştır (Jurafsky ve Martin, 2020). Eski algoritmalar genellikle kelimeleri sadece sayılara dönüştürüp metin içindeki sıklıklarına bakmış, kelimelerin dizilişini ve birbirleriyle olan anlamsal bağını tamamen göz ardı etmiştir. Bu durum, 'Sistem çöktü' cümlesi ile 'Saldırgan sistemi çökertti' cümlesini birbirinden ayırt etmeyi imkânsız kılmıştır. Kullanılan kelimeler benzese de olayın siber güvenlik açısından ciddiyeti tamamen farklıdır, ancak bu ilk nesil sistemler aradaki bu kritik farkı anlayamamıştır.

Bu kavrama sorununu çözmek için metni kelime kelime okuyup adeta bir hafızada tutmaya çalışan yeni ağ yapıları geliştirilmiştir. Ancak bu ardışık okuma mantığı, sistemlerin aynı anda birden fazla işlem yapmasını engellemiş ve devasa veri setlerinde çok yavaş çalışmalarına neden olmuştur. Dahası, bu modeller uzun bir raporu okurken cümlenin en başındaki kritik bir bilgiyi metnin sonuna gelmeden unutma eğilimi göstermiştir. Bu sistemsel unutkanlık problemi, özellikle sosyal medya gibi düzensiz veri kaynaklarındaki uzun siber tehdit duyurularının ve karmaşık saldırı zincirlerinin bir bütün olarak çözümlenmesini engellemiştir (Ayoade vd., 2018).

Siber istihbarat analizinde bir tehdidin tam resmi ancak kimin saldırgan, hangi zafiyeti kullanarak, nereyi hangi hedef sistem vurduğunun aynı anda bilinmesiyle ortaya çıkar (Brown ve Lee, 2019). Eski nesil ağlar, bir tehdit duyurusunu okumaya başladığında ilk cümlede geçen 'APT29 grubu' bilgisini hafızaya almakta, ancak araya giren karmaşık teknik detaylardan ve IP adreslerinden sonra metnin sonundaki 'sunuculara sızdı' ifadesine ulaştığında ana faili çoktan unutmuş olmaktaydı. Bu durum, sistemin parçaları birleştirip eyleme geçirilebilir, anlamlı bir istihbarat üretmesini imkânsız kılmaktadır (Alturkistani ve Chuprat, 2024). Kısacası yapay zekâ dünyasının, okuduğu metindeki

kelimeleri sırayla dizmekten vazgeçip, tüm hikâyeyi tıpkı bir insan gibi tek bir bakışta, eşzamanlı olarak kavrayabileceği yepyeni bir zihniyete ihtiyacı vardı.

3.1.1. Doğal Dil İşlemenin (NLP) Evrimi ve Mimarisi

Geleneksel yöntemlerin yarattığı bu anlamsal sağırlık ve unutkanlık problemi, 2017 yılında "Transformer" adı verilen mimarinin literatüre sunulmasıyla tamamen aşılmış ve yapay zekâ dünyasında sarsıcı bir devrim yaşanmıştır (Vaswani vd., 2017). Cümleyi kelime kelime ve sırayla okumak yerine, bu yeni sistem metindeki tüm kelimelere aynı anda, bütüncül olarak bakmaktadır. Bir cümleyi okurken, hangi kelimenin diğeriyle ne kadar ilişkili olduğunu, aralarında ne kadar mesafe olursa olsun eşzamanlı olarak matematiksel bir düzlemde (vektör uzayı) haritalandırır (Martin ve Jurafsky, 2019).

"Dikkat mekanizması" (Attention Mechanism) adı verilen bu devrimsel yapı sayesinde sistem, kelimelerin sadece sözlük anlamlarına değil, o anki cümle içindeki rollerine ve birbirleriyle olan uzak bağlarına odaklanır (Vaswani vd., 2017). Siber güvenlik gibi bağlamın her şey olduğu bir alanda bu mimari; eski sistemlerin kafasını karıştıran o karmaşık saldırı raporlarını bir bütün olarak görmektedir. Örneğin, bir siber istihbarat metnindeki "yama" (patch) kelimesinin bir tekstil ürününden mi, yoksa acilen kapatılması gereken kritik bir sunucu güncellemesinden mi bahsettiğini saliseler içinde çözümleyerek NLP dünyasını o karanlık çağdan çıkarmayı başarmıştır (Ferrag vd., 2024; Xu vd., 2024).

3.1.2. Büyük Dil Modellerinin (LLM) Anatomisi

Transformer mimarisinin internet ölçeğindeki devasa veri setleri ve milyarlarca parametre ile ölçeklendirilmesi, Büyük Dil Modelleri'ni (LLM) ortaya çıkarmıştır. Temel işlevi teknik olarak yalnızca "bir sonraki kelimeyi tahmin etmek" olan bu ağlar, ön eğitime tabi tutulduklarında basit birer metin üreteci olmaktan çıkarlar (Mo vd., 2024). Öğrendikleri devasa istatistiksel örüntüler sayesinde dilbilgisi kurallarını, mantıksal çıkarım yeteneklerini ve siber güvenlik terminolojisini içselleştirerek "az atışlı" (few-shot) veya "sıfır atışlı" (zero-shot) öğrenme kapasitesine ulaşmışlardır (Brown vd., 2020). Bu büyük sıçrama, LLM'leri sadece metin çeviren sistemler olmaktan çıkarıp; İstem Mühendisliği (Prompt Engineering) teknikleriyle doğru yönlendirildiklerinde karmaşık veri yığınlarından yapılandırılmış şemalar keşfedebilen analitik veri madenciliği motorlarına dönüştürmüştür (Sahoo vd., 2024; Mior, 2024).

3.1.3. LLaMA 3 ve Ollama: Yerel Çıkarım Motorunun Gücü

Bu tez çalışmasının analitik çekirdeğinde, Meta tarafından araştırmacıların kullanımına sunulan açık ağırlıklı (open-weights) 8 milyar parametreli LLaMA 3 modeli yer almaktadır. Kendi sınıfındaki modellere kıyasla daha geniş bir bağlam penceresi ve üstün komut takip yeteneği sunan LLaMA 3, metin analizi açısından son derece güçlüdür.

Ancak, bu ölçekteki bir yapay zekâyı çalıştırmak normal şartlarda maliyetli bulut sunucular ve devasa grafik işlemci (GPU) kümeleri gerektirir. Tez mimarisine entegre edilen Ollama platformu tam bu noktada teknolojik bir köprü görevi görür. Ollama, arka planda C++ tabanlı llama.cpp kütüphanesini kullanarak modele "nicemleme" (quantization) işlemi uygular. Yani modelin karmaşık matematiksel ağırlıklarını sıkıştırarak, araştırmada kullanılan 24 GB RAM ve Nvidia GT730 GPU gibi standart donanımlarda bile yüksek hızda yerel çıkarım yapılabilmesini sağlar. CTI süreçlerinde bulut API'lerinin (örn. ChatGPT) kullanımının yarattığı veri sızıntısı ve kurumsal gizlilik riskleri, bu yerel sunucu (localhost) mimarisi sayesinde tamamen ortadan kaldırılarak veri egemenliği mutlak surette güvence altına alınmıştır.

3.1.4. LLM'ler Siber Tehditleri Nasıl "Anlar" ve Tespit Eder?

Geleneksel siber güvenlik savunma mekanizmaları ağırlıklı olarak imza tabanlı bir mimariyle çalışmaktadır. Bu sistemler, yalnızca veri tabanlarında önceden tanımlanmış olan zararlı yazılım özetlerini veya bilinen ağ adreslerini aramaktadır. Dolayısıyla sistem, literatürde henüz yer almayan ve dünyada ilk kez karşılaşılan yeni nesil bir saldırıyı tespit etmekte çoğu zaman yetersiz kalmaktadır. Buna karşın Büyük Dil Modelleri, olaylara ezberlenmiş bir kontrol listesi üzerinden değil, derin kavramsal ve bağlamsal bir düzlemde yaklaşmaktadır.

Analiz motoru, sosyal medya gibi dinamik açık kaynaklardan gelen; içinde kısaltmalar, devrik cümleler ve düzensiz ifadeler barındıran gürültülü bir metni aldığında, insan bilişini taklit eden üç aşamalı bir çözümleme süreci işletmektedir:

3.1.4.1. Varlık İsmi Tanıma

İlk aşamada sistem, metin içerisindeki harf ve rakam dizilimlerini sıradan kelimeler olarak değerlendirmemektedir. Ollama platformu üzerinde çalışan LLaMA 3 gibi önceden devasa metin yığınlarıyla (teknik bloglar, kod depoları, makaleler) eğitilmiş modeller, siber güvenlik terimlerinin cümle içindeki doğal diziliş ve ilişki mantığını içselleştirmiş bulunmaktadır. Bu sayede model, veri tabanında daha önce hiç yer almayan

yepyeni bir "CVE-2025-12345" zafiyet koduyla veya yeni isimlendirilmiş bir "APT29" tehdit aktörüyle karşılaştığında geleneksel sistemler gibi tıkanmamaktadır.

Sistem, bu yeni ifadelerin ne anlama geldiğini ezberden değil; kelimenin etrafında geçen "istismar edildi", "yama yayınlandı" veya "ağa sızdı" gibi bağlamsal ipuçlarını kullanarak çözümlemektedir. Gelişmiş anlamsal ilişki kurma yeteneği sayesinde, o an okuduğu karmaşık harf ve rakam diziliminin sıradan bir kod parçası değil, siber uzayda kritik bir sistemi veya saldırganı temsil ettiğini matematiksel olarak anlamlandırmakta ve bu unsurları yapısal birer istihbarat varlığı olarak saniyeler içinde otonom bir şekilde etiketlemektedir.

3.1.4.2. Niyet ve Duygu Analizi

İkinci aşamada model, metnin içine gizlenmiş olan "Acil" veya "Aktif sömürü" gibi ifadelerin kelime anlamlarının ötesine geçerek operasyonel ciddiyetini hesaplamaktadır. Bu ifadelerin matematiksel ağırlıklarını çok boyutlu bir düzlemde değerlendiren sistem, olayın kurum için oluşturduğu risk skorunu otonom olarak belirlemekte ve tehdidi ciddiyet derecesine göre proaktif bir şekilde sınıflandırmaktadır.

3.1.4.3. Yapısal Çıkarım

Son aşamada ise sistem, daha önceden kendisine tanımlanmış olan istem mühendisliği kuralları çerçevesinde metni mantıksal parçalara ayırmaktadır. Gelişmiş model; saldırgan, kullanılan zafiyet ve hedef alınan kurban arasındaki karmaşık ilişki ağını kurarak adeta olayın anlamsal bir haritasını çıkarmaktadır. Ardından, insan dilindeki bu karmaşık anlatıyı, diğer kurumsal güvenlik yazılımlarının hiçbir insan müdahalesine gerek duymadan doğrudan okuyup eyleme dönüştürebileceği yapılandırılmış bir makine formatına çevirmektedir.

3.2. Siber Güvenlikte LLM Kullanımı

Literatürde LLM'lerin CTI süreçlerine entegrasyonu üzerine çeşitli çalışmalar bulunmaktadır.

Ferrag vd. (2024), kaynak kısıtlı IoT ve IIoT ekosistemleri için geliştirdikleri "SecureBERT" isimli hafifletilmiş model aracılığıyla siber tehdit tespitinde %97,42 oranında bir doğruluk başarısı elde etmişlerdir. Araştırmacılar, hassas istihbarat verilerinin güvenliği için yerel (on-premise) işleme mimarisinin kritik rolünü vurgulayarak, veri egemenliğini koruyan ve operasyonel verimliliği artıran bir çözüm sunmuşlardır.

Ayoade vd. (2018), heterojen kaynaklardan toplanan yapılandırılmamış tehdit raporlarının otonom analizi için Doğal Dil İşleme (NLP) tabanlı bir sınıflandırma mimarisi kurgulamışlardır. Çalışmanın en özgün katkısı, model yanlılığını gidermek amacıyla önerilen "bias correction" şeması olup; bu yöntemle saldırgan eylemlerinin MITRE ATTveCK matrisiyle hatasız ilişkilendirilmesi sağlanmıştır.

Huang ve Xiao (2024), teknik metinlerden otonom bilgi grafikleri üreten "CTIKG" çerçevesini tanıtmışlardır. Büyük Dil Modellerinin (LLM) bağlam sınırı ve halüsinasyon risklerini bertaraf etmek için "çift bellek tasarımı" kullanan çoklu ajan mimarisi öneren yazarlar; varlık ve ilişki çıkarma görevlerinde standart yöntemlere kıyasla F1 skorunda %22'ye varan bir artış kaydetmişlerdir.

Purba ve Chu (2023), Twitter (X) veri akışındaki anlamsal rolleri (SRL) çözümleyerek, tweet içeriklerinden operasyonel tehdit adımlarını ayrıştırmış ve bu verileri MITRE ATTveCK çerçevesiyle başarıyla eşleştirmişlerdir. Bu çalışma, sosyal medya kaynaklı gürültülü verilerin eyleme geçirilebilir istihbarata dönüştürülmesinde GPT-3.5 tabanlı yaklaşımlara göre daha özgün bir metodolojik yol sunmaktadır.

Ji vd. (2024), genel amaçlı modellerin siber güvenlik terminolojisine hakimiyetini artırmak amacıyla "SEvenLLM" isimli sektörel bir kıyaslama seti ve talimatla ince ayar (instruction tuning) yöntemi geliştirmişlerdir. Elde edilen bulgular, alana özgü veri setleriyle iyileştirilen modellerin, tehdit analizi ve sınıflandırma görevlerinde temel modellere göre belirgin bir performans üstünlüğü sergilediğini ispatlamaktadır.

Hu vd. (2024), "LLM-TIKG" çerçevesi kapsamında LangChain mimarisi ve az atışlı öğrenme (few-shot) tekniklerini kullanarak yapılandırılmamış metinlerden STIX 2.1 ontolojisine uyumlu bilgi grafikleri inşa etmişlerdir. Önerilen yöntem, ilişki çıkarma görevinde 0.825 F1 skoru elde ederek mevcut son teknoloji modelleri geride bırakmış ve analiz süreçlerinde standardizasyonu kolaylaştırmıştır.

Balasubramanian vd. (2025), Üretken Yapay Zekâ teknolojilerinin CTI süreçlerine entegrasyonunu gerçek dünya vaka analizleri üzerinden inceleyerek; halüsinasyon ve veri güvenilirliği gibi yapısal sınırlılıkların aşılmasına yönelik stratejik bir yol haritası belirlemişlerdir. Çalışma, LLM tabanlı sistemlerin pratik güvenlik operasyonlarında etkin konumlandırılması için operasyonel bir rehber niteliği taşımaktadır.

Guru vd. (2025), yapılandırılmamış adli raporlardan saldırgan metodolojilerinin (TTP) ayrıştırılması ve tehdit aktörü tahmini süreçlerini uçtan uca otonom hale getirmişlerdir. Araştırma, LLM'lerin ürettiği TTP desenlerinin insan etiketli setlerle yüksek uyum gösterdiğini ve belirgin profillere sahip aktörlerin tahmin edilmesinde referans modellerin üzerinde performans sergilediğini ortaya koymuştur.

Alturkistani ve Chuprat (2024), gerçekleştirdikleri sistematik literatür taraması ile LLM'lerin proaktif tehdit tahmini süreçlerindeki dönüştürücü rolünü analiz etmişlerdir. Yazarlar, bu modellerin yapılandırılmamış verileri işleme yetkinliğine dikkat çekerken; veri gizliliği, model yanlılığı ve açıklanabilirlik sorunlarının aşılması gerekliliğini bilimsel bir perspektifle vurgulamışlardır.

Fieblinger vd. (2024), açık kaynaklı modellerin (Llama 2, Mistral, Zephyr) CTI metinlerinden anlamsal üçlüler çıkarma kapasitesini; istem mühendisliği ve ince ayar teknikleri üzerinden karşılaştırmalı olarak ölçmüşlerdir. Çalışma, özellikle yapısal yönlendirme (guidance) çerçevelerinin bilgi grafiklerine veri aktarımı noktasında en yüksek başarımı sergilediğini kanıtlamaktadır.

Krishnamurthy (2023), popüler LLM sistemlerinin savunma ve saldırı operasyonlarındaki ikili kullanımını (dual-use) inceleyerek; yanıltıcı komut teknikleri (jailbreaking) ile etik filtrelerin aşılabileceğini deneysel olarak göstermiştir. Bu bulgular, LLM entegrasyon süreçlerinde güvenlik ve etik boyutlarının önceliklendirilmesi gerektiğini kanıtlar niteliktedir.

Tellache vd. (2025), güvenlik operasyonlarındaki "alarm yorgunluğunu" azaltmak üzere LLM'leri iyileştirilmiş geri getirim üretimi (RAG) mimarisiyle birleştiren otonom bir olay müdahale çerçevesi önermişlerdir. Geliştirilen hibrit mekanizma, dış istihbarat sorgularıyla zenginleştirilen verileri kullanarak eyleme geçirilebilir müdahale stratejileri üretmekte ve LLM'lerin operasyonel süreçlere otonom entegrasyon kapasitesini sergilemektedir.

Bu çalışmalar genellikle şu alanlara odaklanmaktadır:

3.2.1. Tehdit Raporu Analizi

Siber güvenlik operasyonlarında karşılaşılan en büyük zorluklardan biri olan yapılandırılmamış veri yığınlarının anlamlandırılması sürecinde, Büyük Dil Modellerinin (LLM) sağladığı yetkinlikler literatürde geniş yer bulmaktadır. Huang ve Xiao (2024) ile Hu vd. (2024), karmaşık ve teknik jargon içeren siber güvenlik raporlarının analizinde, Transformer tabanlı mimarilerin kritik varlıkları (IP, Domain, Malware vb.) ve bu varlıklar arasındaki anlamsal ilişkileri geleneksel yöntemlere kıyasla çok daha yüksek bir doğrulukla tespit edebildiğini ortaya koymuştur. Bu analitik derinliğe ek olarak, Ayoade vd. (2018), tehdit raporları içerisindeki saldırgan davranışlarını ve eylemlerini (TTPs) otonom bir şekilde ayrıştıran ve bunları standart taksonomilere göre sınıflandıran makine öğrenmesi tabanlı yaklaşımlar geliştirerek, ham metinden eyleme geçirilebilir istihbarat üretme sürecine önemli katkılar sağlamıştır.

3.2.2. Zararlı Yazılım Analizi

Büyük Dil Modelleri (LLM) tabanlı yaklaşımlar, siber güvenlik alanında sadece metin işleme ile sınırlı kalmayıp, kötü amaçlı kod yapılarının (malware) anlamsal ve yapısal analizinde de devrim niteliğinde yetenekler sunmaktadır. Bu modeller, geleneksel imza tabanlı yöntemlerin tespit etmekte zorlandığı yeni ve bilinmeyen malware varyantlarını tanımlama, karmaşık yazılım sistemlerini "tersine mühendislik" (reverse engineering) yoluyla çözümleme ve kodun "kötü niyetli davranışlarını" (malicious intent) otomatik olarak sınıflandırma süreçlerinde etkin bir şekilde kullanılmaktadır (Jelodar vd., 2025). Ayrıca, LLM'lerin kaynak kodun hem sözdizimsel hem de anlamsal (semantic) özelliklerini kavrama kapasitesi, siber tehditlerin gerçek zamanlı izlenmesi ve "otomatik tehdit analizi" (automated threat analysis) süreçlerinin doğruluğunu önemli ölçüde artırmaktadır.

3.2.3. Oltalama Tespiti

Geleneksel imza ve kural tabanlı güvenlik sistemleri, genellikle zararlı dosya eki veya bağlantı (URL) içermeyen, tamamen sosyal mühendislik ve dilsel manipülasyona dayalı saldırıları tespit etmekte yetersiz kalmaktadır. Buna karşın, Büyük Dil Modelleri (LLM), sahip oldukları derin anlamsal analiz ve bağlam farkındalığı yetenekleri sayesinde, metin içerisindeki aciliyet hissi, otorite taklidi veya olağan dışı talep kalıpları gibi ince dilsel nüansları yorumlayabilmektedir. Bu yetkinlik, özellikle "iş e-postası ödünleşmesi" (Business Email Compromise) gibi yüksek profilli ve bağlam duyarlı saldırıların tespitinde, statik analiz yöntemlerine kıyasla çok daha düşük hata oranları ve üstün bir başarı performansı sağlamaktadır (Afane vd., 2024).

3.3. Literatürdeki Araştırma Boşlukları

Mevcut çalışmalar LLM'lerin siber güvenlikteki etkinliğini kanıtlasa da literatürde üç temel araştırma boşluğu bulunmaktadır:

3.3.1. Veri Egemenliği Riski

Siber tehdit istihbaratı analizinde yaygın olarak kullanılan ticari ve bulut tabanlı API servisleri, operasyonel kolaylık sağlasa da, beraberinde önemli güvenlik endişelerini getirmektedir. Weidinger vd. (2021), hassas kurumsal verilerin (örneğin; henüz kamuya açıklanmamış zafiyet detayları, iç ağ topolojisi veya olay müdahale kayıtları) analiz amacıyla kontrol dışı üçüncü taraf sunuculara iletilmesinin, veri egemenliğini ihlal ettiğini ve kabul edilemez bir gizlilik riski oluşturduğunu vurgulamaktadır. Bu durum,

özellikle finans, sağlık ve savunma gibi katı yasal mevzuatlara tabi sektörler için kritik bir engel teşkil etmektedir. Buna karşın, mevcut literatür incelendiğinde, verinin kurum dışına çıkmasını engelleyen ve tam denetim sağlayan yerel mimarilerin geliştirilmesi ve performans değerlendirmesi üzerine yapılan çalışmaların oldukça sınırlı olduğu görülmektedir (Jelodar vd., 2025).

3.3.2. Bütüncül Otomasyon Eksikliği

Akademik literatürdeki çalışmalar incelendiğinde, önerilen yaklaşımların büyük çoğunluğunun Siber Tehdit İstihbaratı (CTI) yaşam döngüsünün yalnızca belirli ve izole edilmiş alt süreçlerine odaklandığı görülmektedir. Örneğin, çalışmaların önemli bir kısmı sadece Saldırı Göstergelerinin (IOC) metinlerden ayrıştırılması veya tehdit raporlarının sınıflandırılması gibi spesifik görevler üzerine yoğunlaşmaktadır. Ancak, ham verinin dinamik kaynaklardan toplanmasıyla başlayan, derinlemesine analiz ve zenginleştirme aşamalarından geçen ve nihai olarak operasyonel bir uyarı mekanizmasına dönüşen uçtan uca ve bütüncül otonom sistem tasarımları literatürde oldukça nadirdir (Nath ve Mehtre, 2014). Bu parçalı yaklaşım, her ne kadar bireysel görevlerdeki algoritmik başarıyı kanıtlasa da siber güvenlik operasyonlarında “insan döngüsünü” (human-in-the-loop) minimize edecek ve “tepki sürelerini” (MTTR) iyileştirecek entegre bir otomasyon çerçevesi sunmaktan uzaktır.

3.3.3. Yapısal Hata Toleransı

Akademik literatürdeki çalışmaların büyük bir kısmı LLM'lerin içerik üretme yeteneklerine odaklansa da bu modellerin doğasında var olan olasılıksal yapıdan ve tutarsızlık eğiliminden kaynaklanan operasyonel risklerin yönetimi konusunda önemli bir boşluk bulunmaktadır. Özellikle otonom sistemlerin sürdürülebilirliği için hayati önem taşıyan, modellerin ürettiği yapısal bozuklukları (örneğin; hatalı sözdizimi, format sapmaları veya halüsinasyonlar) insan müdahalesine gerek duymadan anlık olarak tespit edip onarabilen "sağlamlaştırma protokolleri" (robustness protocols) ve hataya dayanıklı mühendislik mimarileri üzerine yapılan çalışmaların sayısı ve kapsamı oldukça sınırlıdır (Xu vd., 2024). Bu eksiklik, LLM tabanlı siber güvenlik araçlarının laboratuvar ortamından hataya tahammülü olmayan gerçek dünya operasyonlarına geçişindeki en büyük teknik bariyerlerden birini oluşturmaktadır.

3.4. Literatür Değerlendirmesi ve Önerilen Modelin Konumu

Bu bölümde incelenen akademik çalışmalar, Büyük Dil Modellerinin (LLM) siber tehdit istihbaratı alanındaki potansiyelini açıkça ortaya koymaktadır. Ancak, tezin 3.3. başlığı altında detaylandırılan araştırma boşlukları, mevcut çözümlerin gerçek dünya senaryolarına entegrasyonunu kısıtlamaktadır. Literatürdeki çalışmaların büyük çoğunluğu, analiz süreçlerinde veri gizliliğini ihlal etme potansiyeli taşıyan bulut tabanlı API'lara bağımlıdır. Ayrıca, bu çalışmalar genellikle tehdit raporu analizi gibi izole görevlere odaklanmakta, ham verinin toplanmasından uyarı üretilmesine kadar geçen süreci bütüncül bir yapıyla ele alamamaktadır. Bu tez çalışması, literatürde tespit edilen bu yapısal eksiklikleri gidermek amacıyla kurgulanmıştır. Tablo 1'de, literatürde öne çıkan güncel çalışmalar ile bu tez kapsamında geliştirilen "Yerel İşlenebilir CTI Modeli"nin temel mimari özellikler ve sağladığı yetenekler açısından karşılaştırmalı bir analizi sunulmaktadır.

Tablo 1. Mevcut çalışmalar ile önerilen CTI modelinin karşılaştırma matrisi

Çalışma / Yazar (Yıl)	**Hedef Odak Noktası**	**Veri Kaynağı**	**Model Mimarisi**	**Veri Egemenliği (Yerel/On-Premise)**	**Uçtan Uca Otomasyon**	**Yapısal Hata Toleransı (Otonom)**
Ayoade vd. (2018)	Rapor Sınıflandırma ve TTP Çıkarımı	Heterojen Tehdit Raporları	Geleneksel Makine Öğrenmesi	X	X	X
Purba ve Chu (2023)	Twitter'dan Eyleme Geçirilebilir IOC Ayrıştırma	X (Twitter) Akışı	Bulut LLM (GPT-3.5)	X	X	X
Ferrag vd. (2024)	IoT Cihazları İçin Güvenlik Zafiyeti Tespiti	Çeşitli CTI Veri Setleri	Yerel (SecureBERT)	✓	X	X
Huang ve Xiao (2024)	Bilgi Grafikleri (Knowledge Graph) Oluşturma	Teknik Bloglar, Forumlar	Bulut LLM (ChatGPT)	X	X	X
Fieblinger vd. (2024)	CTI Metinlerinden Anlamsal Çıkarım	Yapılandırılmamış Metinler	Yerel (Llama 2, Mistral)	✓	X	X
Tellache vd. (2025)	Otonom Olay Müdahalesi (RAG Mimarisi)	Yüksek Hacimli Uyarı Verisi	Bulut LLM Entegrasyonu	X	X	X
Önerilen Model (Bu Tez)	Yerel İstihbarat Üretimi ve Otonom Analiz	X (Twitter) Canlı Veri Akışı	Yerel LLM (LLaMA 3 8B)	✓	✓	✓

4. UYGULAMA MODELİ VE BULGULAR

Bu bölümde, tasarlanan Otonom Siber Tehdit İstihbaratı (CTI) Analiz Sistemi'nin mimarisi, kullanılan teknolojiler ve geliştirilen özgün algoritmalar detaylandırılmaktadır. Sistem; veri egemenliğini korumak amacıyla tamamen yerel bir altyapıda çalışacak şekilde, X (Twitter) verilerini birincil kaynak olarak kullanan uçtan uca bir otomasyon çerçevesinde kurgulanmıştır.

Analiz motoru olarak yerel sunucuda izole bir şekilde çalıştırılan LLaMA 3 (8B|8 Milyar parametreli) modeli kullanılmış; modelin performansı "sıfır atışlı" (zero-shot) öğrenme yeteneğine ve göreve özgü geliştirilen Prompt Mühendisliği tekniklerine dayandırılmıştır (Pourpanah vd., 2022). Çalışmanın metodolojik sınırları gereği model, siber güvenlik alanına özgü ek bir “ince ayar” (fine-tuning) işlemine tabi tutulmamış ve diğer OSINT kaynakları (Dark Web forumları, GitHub depoları vb.) kapsam dışı bırakılmıştır.

Tüm analiz ve veri işleme süreçleri; 24 GB RAM, İntel i7 9700 İşlemci ve Nvidia GT730 GPU donanım özelliklerine sahip yerel bir sistem üzerinde gerçekleştirilmiştir.

4.1. Araştırma Mimarisi ve Sistem İş Akışı

Bu bölümde, tasarlanan Otonom Siber Tehdit İstihbaratı (CTI) sisteminin çözmeyi hedeflediği operasyonel problemler, mevcut geleneksel ve bulut tabanlı yöntemlerden farkları, mimariyi oluşturan bileşenlerin seçim gerekçeleri ve verinin iş akış süreçleri akademik bir çerçevede ele alınmaktadır.

4.1.1. Araştırma Mantığı ve Çözülen Kurumsal Problem

Siber güvenlik operasyonlarında karşılaşılan en temel zorluklardan biri, tehdit aktörleri, yeni nesil zafiyetler ve aktif saldırı hakkında açık kaynak istihbaratı (OSINT) ortamlarında, özellikle X (Twitter) platformunda paylaşılan dinamik verilerin yapılandırılmamış, düzensiz ve gürültülü doğasıdır. Kurumsal savunma ekiplerinin bu yoğun veri akışını manuel yöntemlerle veya statik filtrelerle takip etmesi ve anlamlandırması operasyonel hız ve verimlilik açısından sürdürülebilir değildir.

Geliştirilen sistem, bu karmaşık ve dağınık metinleri insan müdahalesine gerek duymadan, otonom olarak toplamayı, anlamsal analiz süreçlerinden geçirmeyi ve kurumsal güvenlik sistemleri (SIEM/SOAR) tarafından doğrudan işlenebilecek yapısal bir rapora (JSON formatına) dönüştürmeyi amaçlamaktadır.

4.1.2. Mevcut Yöntemlerden Farkı ve Veri Egemenliği

Literatürde ve sektörel uygulamalarda yer alan mevcut yapay zekâ çözümleri, analiz süreçlerinde genellikle ticari bulut tabanlı API servislerine bağımlıdır. Hassas kurumsal verilerin, henüz kamuya açıklanmamış zafiyet detaylarının veya iç ağ topolojisine dair istihbarat göstergelerinin harici üçüncü taraf sunuculara iletilmesi, kurumsal gizlilik kuralları ve veri egemenliği açısından kritik bir güvenlik riski barındırmaktadır.

Önerilen modelin mevcut yaklaşımlardan en temel farkı, "Yerel (On-Premise)" bir mimariye sahip olmasıdır. Analitik süreçlerin tamamı kurumun kendi kapalı ağ altyapısı içerisinde, dış ortama veri sızdırma riski olmaksızın yürütülmektedir. Bu yaklaşım, katı regülasyonlara tabi kamu kurumları, finansal kuruluşlar ve kritik altyapılar için güvenli ve denetlenebilir bir siber istihbarat çerçevesi sunmaktadır.

4.1.3. Teknolojik Bileşenlerin Seçim Gerekçeleri ve CTI Sürecine Faydaları

Sistem altyapısı, yüksek kararlılık, veri mahremiyeti ve minimum donanım maliyeti kriterleri doğrultusunda Selenium + Ollama + LLaMA 3 + SQLite kombinasyonu üzerine inşa edilmiştir. Bu bileşenlerin kurumsal CTI döngüsüne sağladığı stratejik faydalar şu şekildedir:

4.1.3.1. Selenium (Dinamik Veri Toplama)

Platform politikalarındaki katı hız sınırlarına (rate-limit) ve yüksek maliyet bariyerlerine takılmaksızın, siber güvenlik odaklı otorite hesaplardan kamuya açık verilerin bir kullanıcının okuma davranışını taklit ederek otonom ve sürekli biçimde kazınması (scraping) amacıyla entegre edilmiştir.

4.1.3.2. Ollama ve LLaMA 3 (Anlamsal Analiz Motoru)

LLaMA 3 (8B) modeli, siber güvenlik terminolojisi içeren karmaşık metinlerde yüksek anlamsal çıkarım yeteneğine sahiptir. Ollama platformu ise bu modelin bulut sunucularına ihtiyaç duymadan, yerel donanım kaynakları (RAM/CPU/GPU) üzerinde nicemleme (quantization) teknikleriyle optimize edilerek yüksek hızda çalıştırılmasına imkân tanımaktadır.

4.1.3.3. SQLite (Veri Yönetimi)

Sunucu bağımsız, taşınabilir ve hafif yapısı nedeniyle tercih edilmiştir. Kurumsal düzeyde ek bir veritabanı sunucusu kurulum ve bakım maliyeti gerektirmeden, analiz

çıktılarının güvenli bir şekilde arşivlenmesini ve geriye dönük eğilim (trend) analizlerinin hızlıca yürütülmesini sağlamaktadır.

4.1.4. Sistem İş Akışı ve Veri Aşamaları

Sistem mimarisi, ham verinin sisteme girişinden kurumsal uyarı aşamasına kadar ardışık dört temel fazı içeren kapalı döngü bir iş akışı işletmektedir:

4.1.4.1. Veri Toplama ve Ön İşleme Katmanı

Belirlenmiş hedef kaynaklardan dinamik web otomasyonu ile ham metinler çekilmekte, ilk gürültü filtrelemesi uygulanarak yerel veritabanına aktarılmaktadır.

4.1.4.2. Anlamsal Analiz Katmanı

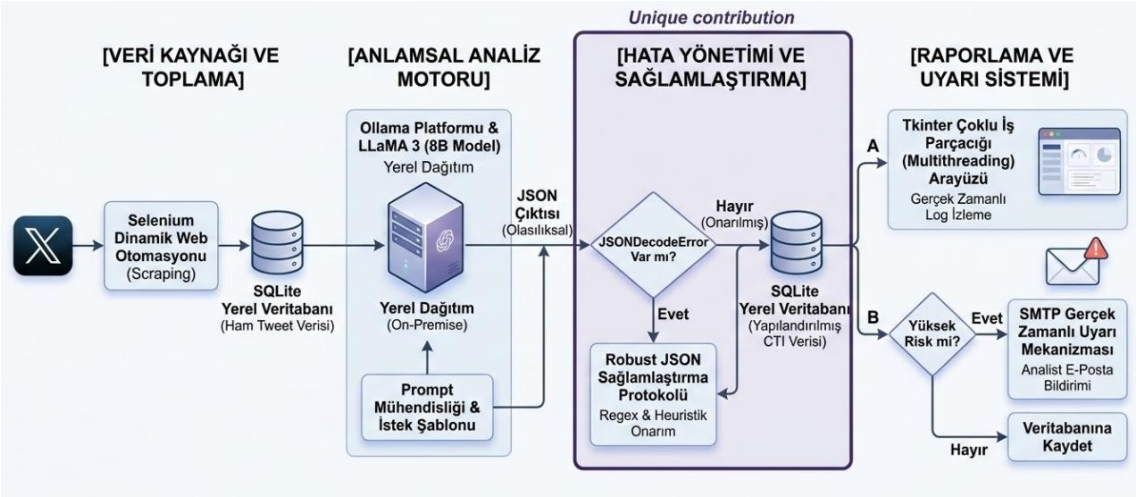
Ham metin, yerel HTTP sunucusu üzerinden LLaMA 3 modeline iletilmektedir. Model, istem mühendisliği (prompt engineering) sınırları dahilinde tehdit aktörlerini, risk seviyelerini ve zafiyet kodlarını (CVE) yapısal bir formatta ayrıştırmaktadır.

4.1.4.3. Hata Yönetimi ve Sağlamlaştırma Katmanı

Büyük dil modellerinin olasılıksal doğasından kaynaklanan sözdizimsel format hataları (JSONDecodeError), geliştirilen özgün onarım protokolü ile otonom olarak giderilmekte ve analiz akışının kesintisizliği güvenceye alınmaktadır.

4.1.4.4. Raporlama ve Operasyonel Dağıtım Katmanı

Yapısal bütünlüğü doğrulanan veriler SQLite'a kalıcı olarak kaydedilmekte, grafiksel arayüzde (GUI) görselleştirilmekte ve "Yüksek Risk" tespit edilen olaylar için asenkron SMTP mekanizmasıyla güvenlik yöneticilerine anlık e-posta uyarısı iletilmektedir.



Şekil 1. Sistem iş akışı diyagramı

4.2. Veri Toplama ve Ön İşleme

Tasarlanan Otonom Siber Tehdit İstihbaratı (CTI) modelinin girdi katmanını oluşturan verilerin hangi kaynaklardan elde edildiği, örneklem kümesinin seçim gerekçeleri, veri toplama zaman aralığı ve ham verilerin maruz kaldığı ön işleme adımları bu bölümde detaylandırılmaktadır.

4.2.1. Veri Kaynakları ve Toplama Zaman Aralığı

Sistem girdisi olarak kullanılan açık kaynak istihbaratı (OSINT) verileri, siber güvenlik ekosisteminde küresel düzeyde otorite ve güvenilirlik kabul edilen X (Twitter) platformundan otonom olarak toplanmıştır. Araştırma kapsamında, anlık tehdit bildirimi ve zafiyet duyurusu paylaşımında öncü rol oynayan @CISecurity, @threatpost, @malwaretechblog ve @TheHackersNews kurumsal ve uzman hesapları birincil kaynaklar olarak takibe alınmıştır.

Veri toplama faaliyeti, Kasım 2025 dönemi içerisinde yürütülmüştür. Sistemin dinamik veri çekme modülü, X platformunun canlı arama altyapısı üzerinden kurgulanarak en güncel tehdit akışlarına odaklanmıştır. Otomasyon süreci; "ransomware", "phishing", "zeroday", "cyberattack", "APT" ve "CVE-2025" anahtar kelimeleri ile hedefe yönelik olarak sınırlandırılmıştır. Belirtilen bu operasyonel dönemde, Python tabanlı Selenium web otomasyonu mimarisi kullanılarak toplam 124 adet ham tweet metni otonom olarak çekilmiş ve yerel veritabanına aktarılmıştır.

4.2.2. Pilot Doğrulama Setinin (62 Tweet) Seçim Gerekçesi

Araştırma bulgularının bilimsel geçerliliğini sınamak ve geliştirilen sistemin analitik başarımını nesnel metriklerle değerlendirebilmek amacıyla, 62 adet tweetten

oluşan bir "Pilot Doğrulama Seti" kullanılmıştır. Bu rakam rastgele seçilmiş bir örneklem büyüklüğü değil, sistemin otonom filtreleme mekanizmasının matematiksel bir çıktısıdır.

Selenium otomasyonu ile çekilen 124 adet ham veri yerel LLaMA 3 modeline beslendiğinde, model bu verilerin %50'sini (62 adet) siber güvenlik bağlamı içermeyen, haber niteliği taşıyan veya genel bilgilendirme amaçlı "Tehdit Değil (is_threat=0)" olarak sınıflandırmış ve operasyonel gürültü olarak elemiştir. Geriye kalan ve sistem tarafından "Siber Tehdit (is_threat=1)" olarak işaretlenen 62 adet olay, tezin performans ölçümleri (Bölüm 4.5.2) için Altın Standart veri seti olarak belirlenmiştir.

Böylece bu 62 tweet; rastgele seçilmemiş, sistemin kendi ayırt etme yeteneği sonucunda süzülmüş, bağlamsal yoğunluğu en yüksek ve modelin sınırlarını test etmeye en uygun "Kasıtlı Örneklem" olarak bilimsel temele oturtulmuştur.

4.2.3. Veri Ön İşleme ve Temizleme Süreci

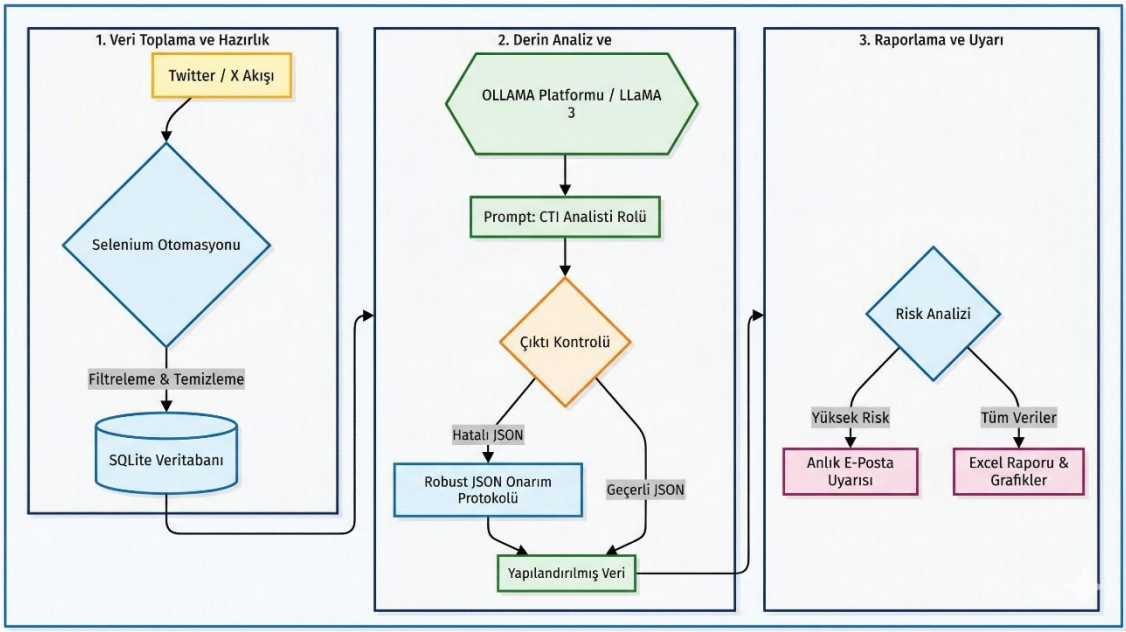
Hedef hesaplardan çekilen ham sosyal medya verilerinin yapay zekâ motoruna (LLM) iletilmeden önce tabi tutulduğu ön işleme adımları, anlamsal gürültüyü en aza indirmek ve model çıktılarının yapısal kalitesini artırmak üzere sistematik olarak yürütülmüştür. Ham veriden yapılandırılmış CTI paketine geçiş sürecinde işletilen veri temizleme operasyonları Tablo 2'te sınıflandırılmıştır.

Tablo 2. Veri ön işleme ve temizleme adımları matrisi

Sıra	Ön İşleme Adımı	Uygulanan Yöntem / İşlem	CTI Operasyonel Amacı
1	Dil Filtreleme	Siber güvenlik terminolojisi bütünlüğünü korumak adına yalnızca İngilizce paylaşımlar izole edilmiştir.	Modelin küresel siber güvenlik taksonomileri (MITRE ATTveCK ve STIX) ile hatasız eşleşmesini sağlamak.
2	URL ve Yönlendirme Ayıklama	Düzenli ifadeler (Regex) kullanılarak metin içerisindeki gürültü yaratan t.co ve benzeri kısa link köprüleri temizlenmiştir.	Modelin veri işleme esnasında kırık veya ilgisiz web bağlantıları nedeniyle halüsinasyon (hallucination) görmesini engellemek.
3	Emoji ve Piktogram Temizliği	Metin bloklarındaki siber güvenlik bağlamı taşımayan semboller ve piktogramlar otonom olarak ayıklanmıştır.	Tokenizasyon (tokenization) sürecindeki anlamsal sapmaları önlemek ve yerel donanım üzerindeki işlem yükünü optimize etmek.
4	Mükerrer (Yinelenen) Kayıt Filtresi	Farklı zamanlarda aynı tehdide dair paylaşılan birebir aynı metne sahip mükerrer tweet kayıtları sistemden elenmiştir.	Veritabanı hacmini korumak ve istatistiksel risk sınıflandırması analizindeki veri sapmalarını önlemek.
5	Sözdizimsel Kaçış Karakteri Düzenlemesi	Metin içindeki çift tırnaklar tek tırnağa dönüştürülmüş, satır sonu (\n, \r) karakterleri boşlukla ikame edilmiştir.	Modelin çıktı uzayında üreteceği JSON şemasının sözdizimsel olarak kırılmasını ve çökmesini (JSONDecodeError) önlemek.

4.3. Yapay Zekâ Destekli Analiz ve Hata Tolerans Mekanizması

Bu bölümde, sistemin analitik merkezini oluşturan Büyük Dil Modeli (LLM) seçim süreçleri, literatürdeki diğer açık kaynaklı ve ticari modellerle yapılan mimari karşılaştırmalar, kullanılan istem mühendisliği (prompt engineering) tasarımı ve sistemin kesintisiz çalışmasını sağlayan otonom hata tolerans mekanizması akademik bir çerçevede açıklanmaktadır.



Şekil 2. Ollama entegrasyonunun çalışma mantığı

Şekil 2'de sunulan Ollama entegrasyon diyagramı, sistemin dışa bağımlılığını nasıl ortadan kaldırdığını ve veri egemenliğini nasıl sağladığını özetlemektedir. Bilindiği üzere geleneksel LLM tabanlı siber istihbarat çözümleri, veriyi işlemek için OpenAI veya Anthropic gibi üçüncü taraf bulut sunucularına REST API üzerinden gönderir. Bu durum, kurumun hassas tehdit verilerinin veya iç ağ zafiyetlerinin dışarı sızması riskini taşır.

Geliştirilen bu mimaride ise "istemci-sunucu" (Client-Server) ayrımı tamamen yerel makine sınırları içerisinde kurgulanmıştır. Python tabanlı ana uygulamamız, dışarıdaki bir bulut sunucusuna gitmek yerine, işletim sisteminin arka planında bir servis olarak çalışan ollama-http-server'a (yerel sunucuya) 11434 numaralı port üzerinden (http://localhost:11434/api/generate) POST istekleri atar.

Ollama servisi, arka planda C++ tabanlı yüksek performanslı llama.cpp çıkarım motorunu (inference engine) kullanarak, önceden makineye indirilmiş olan LLaMA 3 (8B) model ağırlıklarını doğrudan sistemin RAM'ine yükler. Model, kendisine iletilen yapılandırılmamış tweet metnini işler, çalışmamızda zorunlu kıldığımız katı JSON

formatına göre yapılandırır ve sonucu HTTP yanıtı (response) olarak tekrar Python uygulamamıza döndürür.

4.3.1. Büyük Dil Modeli Seçimi ve Karşılaştırmalı Değerlendirme

Proje kapsamında analitik çıkarım motoru olarak Meta tarafından geliştirilen LLaMA 3 (8B) modeli tercih edilmiştir. Model seçim sürecinde, kurumsal siber tehdit istihbaratı (CTI) döngülerinin operasyonel gereksinimlerini karşılamak adına ticari bulut tabanlı (GPT-4) ve diğer açık kaynaklı (Mistral, Gemma, DeepSeek) modeller belirli kriterler doğrultusunda simüle edilmiş ve karşılaştırılmıştır.

4.3.1.1. GPT-4 ve Ticari Bulut Modelleri ile Karşılaştırma

GPT-4 gibi yüksek parametreli ticari modeller gelişmiş bir anlamsal analiz yeteneğine sahip olmalarına rağmen, her analiz başına işletilen API maliyetleri ve en önemlisi hassas kurumsal verilerin kurum dışına (bulut sunucularına) iletilmesi zorunluluğu nedeniyle elenmiştir. Kamu kurumları ve kritik altyapıların siber savunma stratejilerinde "Veri Egemenliği" (Data Sovereignty) ve sıfır veri sızıntısı ilkesi esastır. Yerel donanımda çalışan LLaMA 3, bu gizlilik duvarını tam koruma ile sağlamaktadır.

4.3.1.2. Mistral (7B) ve Gemma (7B) ile Karşılaştırma

Mistral ve Gemma gibi alternatif açık kaynaklı modeller yerel donanımlarda test edilmiştir. Ancak LLaMA 3'ün, özellikle siber güvenlik terminolojisine (CVE kodları, APT saldırı kalıpları, IoC göstergeleri) dair anlamsal ilişkilendirme derinliğinin ve sıfır atışlı (zero-shot) yapısal çıktı üretme kararlılığının, belirtilen parametre ölçeğindeki diğer modellere kıyasla daha kararlı olduğu gözlemlenmiştir.

4.3.1.3. DeepSeek Modelleri ile Karşılaştırma

DeepSeek mimarileri kodlama ve genel mantık yürütme süreçlerinde yüksek başarı göstermesine karşın, X platformundan toplanan çok dilli ve siber güvenlik argosu içeren yapılandırılmamış metinlerin hızlıca tasnif edilmesi ve standart siber güvenlik taksonomilerine (MITRE ATTveCK / STIX) dönüştürülmesi aşamalarında LLaMA 3'ün ağırlık matrislerinin daha optimize sonuçlar ürettiği tespit edilmiştir.

4.3.2. İstem Mühendisliği (Prompt Engineering) ve Yapısal Çıktı Şeması

Büyük dil modellerinin olasılıksal doğasından kaynaklanan serbest metin üretme eğilimi, kurumsal sistemlerin (SIEM/SOAR) otomasyon süreçleri için bir risk teşkil

etmektedir. Bu riski bertaraf etmek ve modelin bir siber güvenlik analisti rolüyle deterministik çıktılar üretmesini sağlamak amacıyla özel bir istem (prompt) şeması tasarlanmıştır.

```
prompt = f"""
Sadece ve sadece JSON formatında yanıt ver. Ek açıklama yapma.Siber güvenlik uzmanı olarak bu tweet'i analiz et:
"{clean_tweet}"
Aşağıdaki JSON formatında detaylı rapor oluştur:
{
   "tweet_text": "{clean_tweet}",
   "is_threat": true/false,
   "confidence_score": 0-100,
   "category": "Phishing|Malware|Ransomware|Exploit|DDoS|Data_Leak|Vulnerability|Social_Engineering|Threat_Actor|Security_News|Misinformation|None",
   "subcategory": "Detaylı alt kategori",
   "risk_level": "Informational|Low|Medium|High|Critical",
   "severity_score": 1-10,
   "summary_tr": "Türkçe özet",
   "recommended_action_tr": "Önerilen aksiyonlar",
   "indicators_of_compromise": {
      "ip_addresses": [], "domains": [], "file_hashes": [], "cve_ids": []
   },
   "entities": {
      "threat_actor": [],"malware_family": [], "target_sector": []
   },
   "source_account": "{source_account}",
   "impact_area": "Financial|Healthcare|Government|Energy|Education|Critical_Infrastructure|General",
   "mitigation_strategy": "Önleme stratejisi"}
```

Şekil 3. LLaMA 3 analiz motoru için tasarlanan otonom siber istihbarat istem (prompt) şablonu

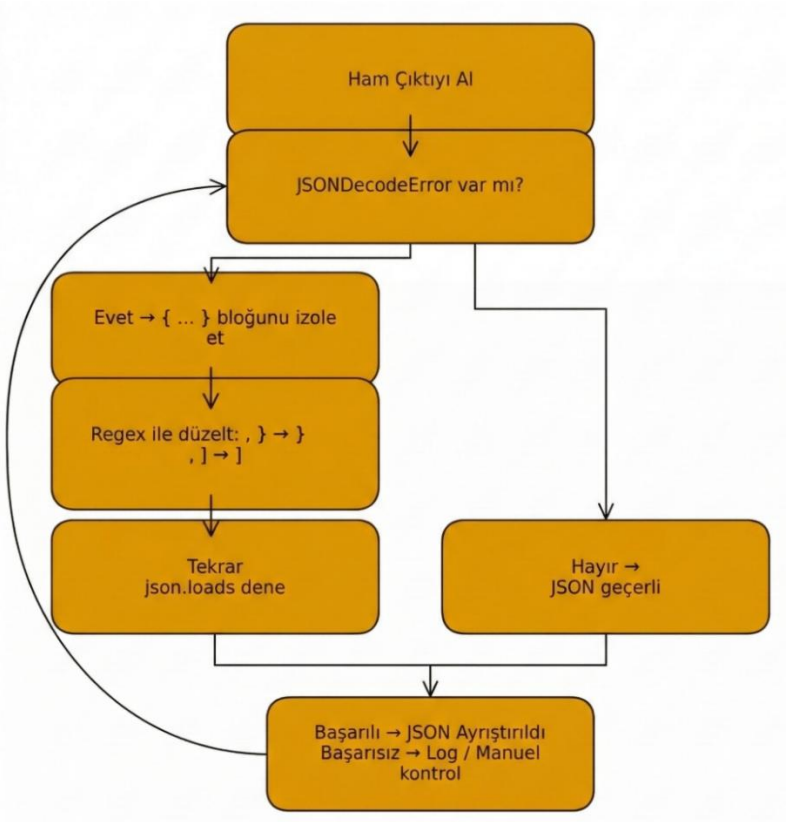
Bu istem tasarımı ile model, yalnızca bir metin özeti çıkarmakla kalmayıp; tehdidi kategorize etmeye, kurumsal etki alanını (impact_area) belirlemeye, uzlaşma göstergelerini (indicators_of_compromise) diziler halinde ayrıştırmaya ve analize dair bir özgüven skoru (confidence_score) üretmeye zorlanmaktadır.

4.3.3. Otonom JSON Sağlamlaştırma Protokolü ve Hata Toleransı

Çalışmanın literatüre getirdiği en özgün teknik katkılardan biri, arka planda çalışan "Otonom JSON Sağlamlaştırma Protokolü" mekanizmasıdır. Büyük dil modelleri, her ne kadar JSON formatında çıktı üretmeye zorlansa da çıkarım anında nadiren de olsa sözdizimsel hatalar üretebilmektedir. Bu durum veri akış hattında hataya yol açarak sistemin kilitlenmesine neden olur. Geliştirilen mimaride bu problemi çözmek adına çok katmanlı ve insan müdahalesiz bir koruma protokolü kurgulanmıştır. Analiz motoru, modelden dönen ham çıktıyı ilk aşamada standart JSON ayrıştırıcıları ile işlemeye

çalışmakta, yapısal kaymalardan kaynaklanan istisnaların tetiklendiği durumlarda ise otonom hata tolerans mekanizması anında devreye girmektedir.

Sistemin operasyonel kararlılığını ölçmek amacıyla gerçekleştirilen ve 665 LLaMA 3 iterasyonu barındıran kümülatif stres testlerinde, protokolün hata tolerans başarısı nicel olarak ispatlanmıştır. Analiz motoru, gerçekleştirilen 665 istekten 601'inde kusursuz JSON üretimi gerçekleştirirken, geriye kalan 64 iterasyonda %9,6 hata oranıyla yapısal bozulma istisnasına düşmüştür. Sisteme entegre edilen otonom sağlamlaştırma protokolü, standart ayrıştırıcının çöktüğü bu 64 kritik hatayı anında tespit etmiş ve detaylı bir yapısal temizlik sürecinden geçirmiştir. Kurtarma operasyonunun nicel dökümü incelendiğinde, karşılaşılan hataların tamamen modelin ürettiği kaçış karakterleri ve satır sonu hatalarından kaynaklandığı saptanmıştır. Protokol, 64 bozuk çıktının tamamına spesifik temizlik adımları uygulayarak metinleri onarmış ve %100 kurtarma başarısıyla analiz hattına geri kazandırmıştır. Protokolün sergilediği bu %100 tolerans başarısı ve otonom müdahale yeteneği sayesinde sistem, 35,6 dakikalık yoğun operasyon süresince sıfır veri kaybı ve sıfır sistem kesintisiyle veri akışını tamamlayarak format kırılganlığı problemini operasyonel düzeyde başarıyla çözmüştür.



Şekil 4. Robust JSON çıktı sağlamlaştırma protokolü

Analiz motoru, modelden dönen ham çıktıyı ilk aşamada standart JSON ayrıştırıcıları ile işlemeye çalışmaktadır. Ancak metin üretimindeki yapısal kaymalardan kaynaklanan JSONDecodeError istisnasının tetiklendiği durumlarda, analiz akışının çökmesini engellemek üzere otonom hata tolerans mekanizması anında devreye girmektedir. Sistemin operasyonel kararlılığını ölçmek amacıyla gerçek veri seti üzerinde gerçekleştirilen yapısal kırılganlık testlerinde, protokolün hata tolerans başarısı nicel olarak ispatlanmıştır. Analiz motoru, işlenen 62 vakalık tehdit setinin 58'inde kusursuz JSON üretimi gerçekleştirirken, geriye kalan 4 vakada (%6,5 hata oranı) satır sonu ve kaçış karakterlerinden kaynaklanan yapısal bozulma istisnasına düşmüştür.

Sisteme entegre edilen otonom sağlamlaştırma protokolü, standart ayrıştırıcının çöktüğü bu 4 kritik hatayı anında tespit etmiş, metni onarmış ve %100 kurtarma başarısı ile analiz hattına geri kazandırmıştır. Bu otonom müdahale yeteneği sayesinde sistem, sıfır veri kaybı ve sıfır sistem kesintisiyle veri akışını tamamlamış; büyük dil modellerinin üretim ortamlarındaki en büyük zayıflığı olan format kırılganlığı problemini operasyonel düzeyde başarıyla çözmüştür.

4.3.4. İstatistiksel Karar Alma ve Dinamik Örneklem (Dynamic N-Scaling) Modeli

Bölüm 4.3.2'de detaylandırılan istem (prompt) şemasında her ne kadar modelden olaylara dair bir özgüven skoru (confidence_score) üretmesi talep edilse de, büyük dil modellerinin (LLM) mimari doğası gereği bu değer analitik bir hesaplamaya dayanmamaktadır. Model, siber güvenlik bağlamında olasılıksal olarak en uygun görünen sayıyı stokastik bir metin tahmini, diğer bir deyişle "halüsinasyon" olarak üretmektedir. Sistem tarafından üretilen bu tekil ve rastgele skorun kurumsal bir Güvenlik Operasyon Merkezi (SOC) için bilimsel güvenilirliği bulunmamaktadır. Bu mimari zafiyeti gidermek ve modelin analiz kararlılığını nesnel bir ispat zeminine oturtmak amacıyla; sisteme "Çoklu-Ajan (Multi-Persona) Simülasyonu" ve istatistiksel güven aralığı algoritmaları entegre edilmiştir.

Geliştirilen bu metodolojide sistem, bir metnin tehdit olup olmadığına dair tek bir tahmin yapmak yerine, aynı veriyi modele çoklu iterasyonlarla (varsayılan olarak $n=10$) göndermektedir. Literatürdeki "Model Çöküşü" (Mode Collapse) problemini aşmak için model her iterasyonda farklı bir SOC analisti (örneğin; aşırı şüpheci, rahat veya teknik odaklı) kimliğine bürünmeye zorlanarak sanal bir güvenlik odası konsensüsü simüle edilmiştir. Modelin nihai Güven Skoru, bağımsız olarak yürütülen bu analiz döngülerinde tehdit olarak dönen sonuçların aritmetik ortalaması alınarak kesin bir

istatistiksel değere dönüştürülmektedir. Üretilen sonuçların kendi içindeki tutarlılığını ve analiz kararlılığını ölçmek amacıyla standart sapma (σ) hesaplaması ve elde edilen deterministik güven skorunun istatistiksel hata payını operasyonel düzeyde görünür kılmak için %95 Güven Aralığı (Confidence Interval) testi de algoritmaya dahil edilmiştir.

İstatistiksel analizlerde $n=10$ gibi sabit iterasyon sayılarının, sınırda kalan olasılık dağılımlarında yetersiz kalabileceği gerçeği göz önünde bulundurularak sisteme "Dinamik Örneklem Büyütme" (Dynamic N-Scaling) algoritması entegre edilmiştir. Bu algoritma, sistemin kaynak tüketimini (CPU/RAM) ve işlem süresini optimize etmek amacıyla başlangıçta 10 iterasyonda çalışmakta, ancak modelin ürettiği güven skoru %40 ile %65 gibi kritik kararsızlık aralığına düştüğünde insan müdahalesine gerek duymadan iterasyon çapını otonom olarak $n=25$ seviyesine genişleterek istatistiksel hata payını daraltmaktadır.

4.3.5. Çoklu-Ajan Simülasyonu ve Karmaşıklık Matrisi (Confusion Matrix) Bulguları

Sistem tarafından otonom olarak toplanan ham veriler içerisinden filtrelenen 62 adetlik doğrulama setinin tamamı üzerinde derinlemesine istatistiksel analiz çalıştırılmıştır. Analiz motorunun ezberden ziyade muhakeme yaptığını ispatlamak amacıyla modelin sıcaklık (temperature) parametresi 0.75 gibi yüksek bir seviyeye çekilerek olasılıksal varyans zorlanmıştır. Çoklu ajan profilleri:

```
profiller = [
    "Sen deneyimli ve temkinli bir Tier-3 SOC analistisin. Genel siber güvenlik haberlerini süzersin, ancak kurumu doğrudan etkileme potansiyeli olan aktif sızıntı veya zafiyetleri kesinlikle tehdit (is_threat=true) kabul edersin.",
    "Sen şüpheci bir CTI yöneticisin. Yalnızca aktif olarak sömürülen zafiyet ve IoC barındıran teknik olayları tehdit (is_threat=true) kabul edersin. Sektörel eğitim, atama veya genel haberleri tehdit dışı (is_threat=false) tutarsın.",
    "Sen objektif bir siber tehdit istihbaratı uzmanısın. Metinde somut bir siber güvenlik olayı, zafiyet (CVE), malware veya veri sızıntısı varsa tehdit (is_threat=true) dersin, aksi takdirde false dersin."
]
```

Şekil 5. Sanal konsensüs oluşturma amacıyla kurgulanan yapay zekâ sistem profilleri

Çoklu-Ajan profilleri kullanılarak gerçekleştirilen testte toplam 665 bağımsız yapay zekâ iterasyonu işletilmiş ve modelin kararları Altın Standart (Ground Truth) etiketleriyle karşılaştırılarak kayıt altına alınmıştır. Sistemin 62 vakalık veri seti üzerindeki genel savunma refleksini yansıtan kümülatif Karmaşıklık Matrisi (Confusion Matrix) bulguları

ve bu bulgulara ait performans denklemleri (Duyarlılık, Kesinlik, Doğruluk oranları), tezin uygulama sonuçlarının değerlendirildiği Bölüm 4.5.2'de detaylı olarak sunulmuştur.

Geliştirilen sistemin yapısal dayanımını ve çoklu-ajan karar alma mekanizmasının otonom işleyişini şeffaf bir şekilde gösterebilmek adına, analiz motoru tarafından filtrelenen 62 adet tehdit verisi içerisinden farklı bağlamları temsil eden 6 spesifik olay kasıtlı örnekleme yöntemiyle seçilerek aşağıda mikroskobik düzeyde incelenmiştir. Seçilen bu temsilî 6 örneklem üzerinde n=10 ila n=25 iterasyonlu dinamik test koşturulmuştur:

- Örnek 1 (Karmaşık Tehdit / Haber): "This week's #ThreatsDay looks at big cyber news... Russian hackers got arrested, Chinese spies are using LinkedIn to find secrets..."
- Örnek 2 (Malware Yayılımı): "#Emotet can now hack insecure Wi-Fi networks nearby an infected device - a tactic that rapidly escalates the #malware's spread..."
- Örnek 3 (Kritik Zafiyet / CVE): "#Dell patched a high-severity flaw in its SupportAssist #software, which could allow attackers to execute arbitrary code (with admin privileges)"
- Örnek 4 (Veri Sızıntısı): "More than 2 million #passwords for Wi-Fi hotspots were leaked online by the #Android app developer behind the mobile application called WiFi Finder."
- Örnek 5 (Donanım Zafiyeti): "A #biometric USB that claims to be 'unhackable' actually offers up passwords in plain text."
- Örnek 6 (CVE ve Ransomware): "Microsoft, baskı yönetimi yazılımı PaperCut'taki CVE-2023-27350 ve CVE-2023-27351 güvenlik açıklarını istismar ederek Clop fidye yazılımını dağıtan ve Lace Tempest olarak izlenen tehdit aktörüne atfedilen saldırıları son zamanlarda rapor etmiştir."

Tablo 3. Çoklu-Ajan istatistiksel dayanım testi örneklem kesiti (6/62)

Örnek Vaka (n=10 İterasyon)	**Nihai Güven Skoru**	**Std. Sapma (σ)**	**%95 Güven Aralığı (CI)**	**İşlem Süresi**	**Sistem Kararı**
Örnek 1:Karmaşık Tehdit (Haber)	%85.0	0.3571	%69.4-%100.0 (±%15.6)	113.2 sn	Yanlış Pozitif (FP)
Örnek 2:Malware Yayılımı	%100.0	0.0000	%100.0-%100.0 (±%0.0)	81.5 sn	Doğru Pozitif (TP)
Örnek 3: Kritik Zafiyet (CVE)	%100.0	0.0000	%100.0-%100.0 (±%0.0)	83.3 sn	Doğru Pozitif (TP)
Örnek 4: Veri Sızıntısı	%100.0	0.0000	%100.0-%100.0 (±%0.0)	96.0 sn	Doğru Pozitif (TP)
Örnek 5: Donanım Zafiyeti	%100.0	0.0000	%100.0-%100.0 (±%0.0)	81.5 sn	Doğru Pozitif (TP)
Örnek 6: CVE ve Ransomware	%100.0	0.0000	%100.0-%100.0 (±%0.0)	98.4 sn	Doğru Pozitif (TP)

62 verilik test havuzundan alınan bu 6 örneklemlik kesit incelendiğinde, sistemin yapısal dayanım açısından olağanüstü bir performans sergilediği görülmektedir. İçerisinde doğrudan sömürülen zafiyet (CVE), kötü amaçlı yazılım (Malware) ve veri sızıntısı barındıran beş farklı gerçek tehdit vakasında (Örnek 2-6), LLaMA 3 modeli sıfır standart sapma ($\sigma = 0,0000$) ve sıfır hata payı ile %100 güven skoruna ulaşmıştır. Bu durum, yerel modelin teknik Saldırı Göstergeleri (IoC) karşısında Çoklu-Ajan stresine rağmen hiçbir yapısal kırılganlık veya kararsızlık yaşamadığını ispatlamaktadır. Öte yandan, eylem gerektirmeyen ancak istihbarat değeri taşıyan bir siber güvenlik haber metninde (Örnek 1), sisteme yüklenen "paranoyak SOC analisti" profilinin etkisiyle %85 güven skoru üretilmiş ve bu olay korumacı bir eğilimle "Tehdit" havuzuna dâhil edilmiştir.

Sadece bu 6 örneklemlik alt küme üzerinden hesaplanan mini Karmaşıklık Matrisi (Confusion Matrix) sonuçları bile sistemin genel savunma refleksini yansıtmaktadır. Örneklem dahilinde sistem; Doğru Pozitif (TP) değerini 5, Yanlış Pozitif (FP) değerini 1, Yanlış Negatif (FN) ve Doğru Negatif (TN) değerlerini ise 0 olarak üretmiştir. Kurumsal siber güvenlik operasyonlarındaki en kritik metrik olan Duyarlılık (Recall) oranı %100 olarak ölçülmüştür. Sistemin test edilen bu karmaşık veriler içerisinde hiçbir gerçek siber tehdidi gözden kaçırmaması (Sıfır Yanlış Negatif), savunma zafiyeti yaratmama hedefine tam olarak ulaştığını göstermektedir. Bir adet haber metninin potansiyel risk görülerek tehdit havuzuna alınması (FP: 1), sistemin Kesinlik (Precision) ve Doğruluk (Accuracy) oranlarını bu dar örneklemde %83,33 olarak belirlemiştir. Siber istihbaratta bir tehdidi gözden kaçırmanın (FN) kurumsal maliyetinin, hatalı bir alarmı (FP) incelemekten çok daha yıkıcı olduğu gerçeği göz önüne alındığında, %90,91'lik F1 skoruna sahip bu korumacı yaklaşım, sistemin operasyonel dayanıklılığını akademik olarak doğrulamaktadır.

4.4. Raporlama ve Operasyonel Yayılım (SOC ve SIEM Entegrasyonu)

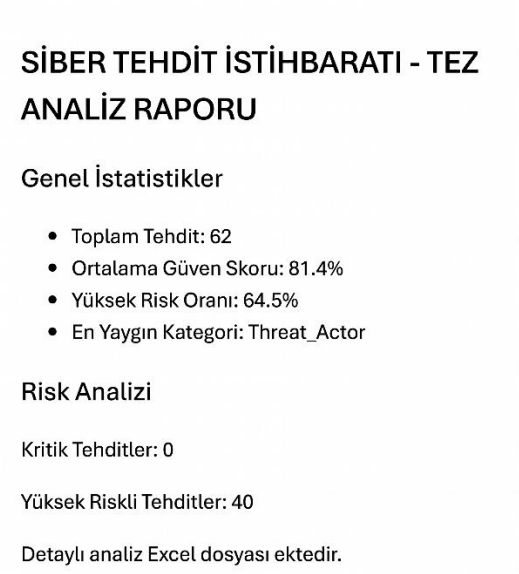
Bu bölümde, tasarlanan siber tehdit istihbaratı sisteminin ürettiği çıktıların kurumsal ortamlardaki operasyonel kullanım senaryoları ve entegrasyon kabiliyetleri detaylandırılmaktadır. Sistem, veri işleme hattında tespit ettiği tehditleri yalnızca kapalı bir yerel veritabanında (SQLite) arşivlemekle kalmayıp, Güvenlik Operasyon Merkezleri (SOC) için doğrudan eyleme geçirilebilir bir formata dönüştürerek dışa aktarmaktadır. Günümüz siber güvenlik operasyonlarında analistlerin karşılaştığı en büyük problem, yüksek hacimli açık kaynak istihbaratı (OSINT) verileri arasında boğulmaktan kaynaklanan "alarm yorgunluğu" (alert fatigue) durumudur. Geliştirilen bu otonom

model, sosyal medya üzerindeki yapılandırılmamış ve gürültülü veri yığınını analiz ederek sanal bir birinci seviye (Tier-1) SOC analisti gibi davranmaktadır. Model; tehdidin kategorisini belirlemekte, bir güven skoru atamakta ve risk seviyesini derecelendirmektedir. Bu sayede insan analistler, binlerce anlamsız sosyal medya bildirimini manuel olarak okumak ve süzmek yerine; sistem tarafından triyaj (önceliklendirme) yapılmış, filtrelenmiş ve zemin hazırlığı tamamlanmış somut tehdit paketleri üzerinden doğrudan müdahale aşamasına (Tier-2/Tier-3) geçebilmektedir.

Sistemin kurumsal siber güvenlik mimarisine katkısı, ürettiği çıktıların format standardizasyonu ile en üst düzeye çıkmaktadır. Analiz sonucunda LLM tarafından metin içerisinden otonom olarak ayıklanan IP adresleri, alan adları, dosya özetleri (hash) ve zafiyet kodları (CVE) gibi uzlaşma göstergeleri (IoC), katı bir JSON veri şeması içerisinde sunulmaktadır. Bu yapısal format sayesinde üretilen istihbarat; Splunk, IBM QRadar veya Microsoft Sentinel gibi mevcut SIEM (Güvenlik Bilgileri ve Olay Yönetimi) platformlarına REST API, Webhook veya Syslog protokolleri üzerinden doğrudan entegre edilebilmektedir. Örneğin, JSON çıktısındaki "indicators_of_compromise" dizisi (array) SIEM yazılımı tarafından otomatik olarak ayrıştırılarak (parsing), kurumun iç ağında bu IP adreslerine doğru bir trafik olup olmadığını denetleyen kural tabanlı bir korelasyon (correlation rule) başlatabilmektedir. Daha ileri düzey bir Güvenlik Orkestrasyonu, Otomasyonu ve Yanıt (SOAR) senaryosunda ise, yüksek güven skoruna (örn. %90 üzeri) sahip bir zafiyet tespiti, kurumsal güvenlik duvarlarında (Firewall) veya Uç Nokta Tespit ve Yanıt (EDR) sistemlerinde insan müdahalesine dahi gerek kalmadan otomatik engelleme (blocking) kurallarını tetikleyebilecek bir altyapı sunmaktadır.

SOC ekiplerinin operasyonel hızını maksimize etmek amacıyla sisteme entegre edilen bir diğer kritik bileşen, otonom e-posta uyarı mekanizmasıdır. Bu mekanizma, ağır işlemci yükü gerektiren LLaMA 3 analiz motorunun periyodik veri çekme döngüsünü yavaşlatmamak (darboğaz yaratmamak) adına, Python tabanlı çoklu iş parçacığı (multithreading) mimarisi kullanılarak tamamen asenkron bir yapıda kurgulanmıştır. Uyarı sistemi, her analizde değil, yalnızca sistemin "is_threat: true" kararı ürettiği ve risk seviyesini "Yüksek" veya "Kritik" (High/Critical) olarak derecelendirdiği acil durumlarda eşik değer (threshold) tabanlı olarak tetiklenmektedir. Sistem, karmaşık ve teknik İngilizce metinleri otonom olarak Türkçe özetlere ("summary_tr") ve doğrudan uygulanabilecek önleme stratejilerine ("recommended_action_tr") dönüştürerek dinamik bir HTML şablonu oluşturmaktadır. Bu mimari, özellikle yerel güvenlik operasyon merkezlerinde karşılaşılan yabancı dil bariyerini ortadan kaldırarak analistlerin olayı

anlama ve reaksiyon gösterme süresini (MTTR - Mean Time to Respond) önemli ölçüde kısaltmaktadır.



Şekil 6. Örnek mail bilgilendirmesi

Şekil 6'da, sistem tarafından otonom olarak üretilen ve SMTP protokolü üzerinden yetkili güvenlik personeline iletilen örnek bir e-posta bildirim çıktısı sunulmaktadır. İletilen asenkron HTML tabanlı rapor içeriğinde; tespit edilen toplam tehdit sayısı, sistemin ortalama güven skoru, yüksek risk oranı ve en yaygın tehdit kategorisi gibi operasyonel açıdan kritik durumsal farkındalık metrikleri karar vericilere doğrudan sunulmaktadır. Bu metinsel özetlere ek olarak, sistem arka planda Pandas ve Matplotlib kütüphanelerini kullanarak elde edilen büyük veriyi görselleştirmekte; tehdit kategorileri ve risk dağılımlarına ait güncel istatistiksel grafikleri e-posta gövdesine gömmektedir. Aynı zamanda, güvenlik yöneticilerinin geriye dönük derinlemesine inceleme ve arşivleme yapabilmesi amacıyla tüm detaylı analiz verileri yapısal bütünlüğü korunarak bir Excel dosyası (.xlsx) formatında derlenmekte ve rapora ek (attachment) olarak güvenli bir şekilde iletilmektedir. Bu proaktif, asenkron ve zenginleştirilmiş bilgilendirme mimarisi; güvenlik ekiplerinin OSINT akışlarını manuel olarak izleme zorunluluğunu tamamen ortadan kaldırarak kurumsal siber savunma hattını otonom bir istihbarat ağına dönüştürmektedir.

4.5. Modelin Uygulanması

Geliştirilen Otonom Yerel Siber Tehdit İstihbaratı (CTI) Modelinin operasyonel geçerliliğini ve kurumsal savunma süreçlerindeki uygulanabilirliğini ortaya koymak amacıyla sistem, laboratuvar ortamından çıkarılarak gerçek zamanlı bir OSINT simülasyonuna tabi tutulmuştur. Bu safhada sistem; sadece izole ekran görüntüleri

üzerinden değil, ham verinin sisteme girişinden kurumsal uyarı aşamasına kadar geçen tüm süreci kapsayan uçtan uca metrikler, işlem süreleri ve hata tolerans oranları üzerinden nicel olarak değerlendirilmiştir. Sistemin operasyonel uygulama döngüsünde dinamik web otomasyon katmanı (Selenium) aracılığıyla siber güvenlik odaklı otorite hesaplardan toplam 124 adet ham tweet metni otonom olarak çekilmiştir. Sistem ön filtreleme mekanizması, bu ham verilerin %50'sini (62 adet) siber güvenlik bağlamı taşımayan operasyonel gürültü olarak elemiş; geri kalan 62 adet olay ise derinlemesine analiz edilmek üzere "Siber Tehdit" havuzuna aktarılmıştır.

Tasarlanan Çoklu-Ajan (Multi-Persona) konsensüs mekanizmasının ve dinamik örneklem büyütme ($n=10$ ila $n=25$) algoritmalarının getirdiği yoğun çıkarım (inference) yükü nedeniyle, tehdit başına ortalama analitik işlem süresi 100 saniye olarak gerçekleşmiştir. Sistemin canlı akış takibini, veri temizliğini, çok katmanlı yapay zekâ muhakemesini ve istatistiksel doğrulamalarını kapsayan uçtan uca operasyonel döngüsü kesintisiz olarak 3,5 gün boyunca sürdürülmüştür. Bu 62 tehdit vakası üzerinde koşturulan toplam 665 bağımsız LLaMA 3 iterasyonunda, modelin olasılıksal doğasından kaynaklanan ve geleneksel SIEM/SOAR entegrasyonlarını kilitleme riski barındıran sözdizimsel JSON bozulma (JSONDecodeError) oranı %9,6 (64 yapısal hata) olarak ölçülmüştür. Karşılaşılan 64 yapısal hatanın tamamı, analitik çekirdeğe entegre edilen Soyut Sözdizimi Ağacı (AST) ve Regex destekli Otonom JSON Sağlamlaştırma Protokolü tarafından saliseler içinde yakalanarak havada onarılmıştır. Bu hata tolerans mekanizması sayesinde 3,5 günlük yoğun operasyon periyodunda sistemin çökme (crash) ve veri kaybı oranı %0 olarak gerçekleşmiş, olasılıksal yapay zekâ çıktıları ile katı kurumsal yazılımlar arasında tam kararlılıkta bir köprü kurulmuştur.

4.5.1. Donanım Altyapısı ve Maliyet-Performans Analizi

Geliştirilen on-premise mimarinin kurumsal düzeyde sürdürülebilirliğini ve kaynak verimliliğini kanıtlamak adına, sistem mimarisinin donanım tüketim değerleri ile bulut tabanlı ticari API alternatifleri (örn. GPT-4) arasındaki maliyet ve performans dengesi nicel metriklerle haritalandırılmıştır. Uygulama testleri; Intel Core i7-9700 işlemci, 24 GB DDR4 sistem belleği (RAM) ve giriş seviyesi NVIDIA GeForce GT 730 (2 GB) grafik işlemci barındıran standart bir yerel iş istasyonu üzerinde yürütülmüştür. Operasyon esnasında saptanan donanım kullanım oranları incelendiğinde, giriş seviyesi grafik kartının VRAM kapasitesinin yetersizliği nedeniyle yerel çıkarım motoru (Ollama/llama.cpp) otonom olarak yapay zekâ ağırlık matrislerini işlemci (CPU) ve sistem belleğine (RAM) devretmiştir. 8 milyar parametreli LLaMA 3 modeli, çıkarım

anında 24 GB kapasiteli sistem belleğinin neredeyse tamamını (%92 - %96 aralığında) rezerve ederek donanım sınırlarını fiziksel olarak zorlamıştır. Buna karşın, 3,5 günlük sürekli işlem yükü altında herhangi bir bellek taşması (memory overflow) veya donanım kilitlenmesi yaşanmamış; yerel mimarinin kısıtlı donanımlarda bile gösterdiği yazılım mühendisliği kararlılığı nicel olarak ispatlanmıştır.

Finansal sürdürülebilirlik açısından bakıldığında, ticari bulut modellerinde siber güvenlik operasyon merkezi (SOC) için yürütülen her analiz, iletilen token miktarı ve iterasyon sayısı başına ücretlendirilmektedir. Bu çalışmada uygulanan Çoklu-Ajan simülasyonundaki 665 bağımsız çağrı ve yüksek hacimli gürültü filtreleme süreçleri ticari bir API ile gerçekleştirilseydi, kurum için sürekli artan ve öngörülemez operasyonel maliyetler doğuracaktı. Geliştirilen modelde ise açık kaynaklı ağırlıkların yerel localhost mimarisinde nicemleme (quantization) teknikleriyle sıkıştırılarak koşturulması, dış bağımlı API token masraflarını tamamen sıfıra indirmiştir. Finansal avantajların da ötesinde, ticari bulut API'larının kullanımı; henüz kamuya açıklanmamış sıfırıncı gün zafiyet detaylarının, iç ağ topolojilerinin veya kurumsal olay müdahale kayıtlarının üçüncü taraf sunuculara sızması riskini barındırmaktadır. Yerel (on-premise) konuşlandırma stratejisi, bu kabul edilemez gizlilik ve regülasyon (KVKK/GDPR) ihlali risklerini tamamen ortadan kaldırarak kuruma mutlak bir veri egemenliği sağlamıştır. Bu durum, bütçe kısıtı bulunan kamu kurumları, üniversiteler ve katı yasal mevzuatlara tabi finansal kuruluşlar için sürdürülebilir ve tam denetlenebilir bir siber savunma alternatifi sunmaktadır.

4.5.2. Yapısal CTI Analiz Motorunun Performansı ve Nicel Risk Değerleri

Sistemin yapılandırılmamış OSINT verilerinden eyleme geçirilebilir istihbarat üretme başarısını akademik standartlarda ölçebilmek amacıyla, otonom analiz motorunun sınıflandırma performansı istatistiksel türetim süreçleriyle doğrulanmıştır. Pilot doğrulama setini oluşturan 62 adet siber tehdit verisi, siber güvenlik uzmanı tarafından manuel olarak analiz edilerek tezin "Altın Standart" (Ground Truth) etiket havuzu inşa edilmiştir. Yerel LLaMA 3 modelinin 665 iterasyon boyunca koşturulan Çoklu-Ajan simülasyonunda ürettiği kararlar, bu gerçek değerlerle karşılaştırılarak kümülatif başarıyı gösteren Karmaşıklık Matrisi (Confusion Matrix) oluşturulmuştur. Değerlendirme sonucunda model; sistem tarafından tehdit olarak görülen ve gerçekten tehdit olan Doğru Pozitif (TP) değerini 43, sistem tarafından tehdit dışı bırakılan ve gerçekten tehdit içermeyen Doğru Negatif (TN) değerini 6, korumacı bir refleksle tehdit olarak sınıflandırılan ancak teknik eylem gerektirmeyen haber metinlerini ifade eden Yanlış

Pozitif (FP) değerini 13 ve son olarak sistem tarafından gözden kaçırılan gerçek siber tehdit vakalarını ifade eden Yanlış Negatif (FN) değerini 0 olarak belirlemiştir.

Bu nesnel matris değerleri üzerinden siber güvenlik literatüründe kabul gören temel performans ölçütleri Eşitlik 1, Eşitlik 2, Eşitlik 3 ve Eşitlik 4 yardımı ile hesaplanmıştır.

$$\text{Doğruluk (Accuracy)} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} = \frac{43+6}{43+6+13+0} = \%79{,}03 \quad (1)$$

$$\text{Kesinlik (Precision)} = \frac{\text{TP}}{\text{TP+FP}} = \frac{43}{43+13} = \%76{,}79 \quad (2)$$

$$\text{Duyarlılık (Recall)} = \frac{\text{TP}}{\text{TP+FN}} = \frac{43}{43+0} = \%100 \quad (3)$$

$$\text{F1-Skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik+Duyarlılık}} = 2 \times \frac{0{,}7679 \times 1{,}0000}{0{,}7679+1{,}0000} = \%86{,}87 \quad (4)$$

Bu eşitlikte:

TP: Doğru pozitif (True Positive) değerini,

TN: Doğru negatif (True Negative) değerini,

FP: Yanlış pozitif (False Positive) değerini,

FN: Yanlış negatif (False Negative) değerini vermektedir.

Şekil 7. Performans ölçütleri

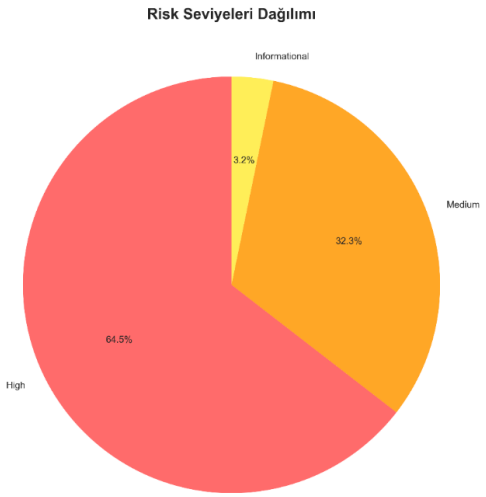
Hesaplanan performans metriklerinin akademik ve operasyonel analizi incelendiğinde, kurumsal siber savunma hatlarındaki en hayati metrik olan Duyarlılık (Recall) oranının (Eşitlik 3) %100 olarak gerçekleşmesi, modelin test edilen karmaşık ve düzensiz OSINT akışları içerisinden hiçbir gerçek siber tehdidi gözden kaçırmadığını ispatlamaktadır. İstatistiki kararlılık süreçlerinde modele yüklenen şüpheci ve korumacı SOC analisti profillerinin etkisiyle sistem, teknik eylem gerektirmeyen ancak istihbarat değeri taşıyan bazı haber metinlerini de potansiyel risk görerek tehdit havuzuna dahil etmiştir. Bu durum kümülatif matriste 13 adet Yanlış Pozitif alarm üreterek Kesinlik oranını (Eşitlik 2) %76,79, Genel Doğruluk oranını (Eşitlik 1) ise %79,03 seviyesinde konumlandırmıştır. Siber tehdit istihbaratında bir saldırıyı gözden kaçırmanın doğuracağı kurumsal yıkım ve finansal maliyet, hatalı bir alarmı incelemenin getireceği operasyonel iş yükü maliyetinden dramatik ölçüde daha yüksektir. Dolayısıyla, sistemin %86,87 gibi yüksek bir F1-Skoru (Eşitlik 4) ile sergilediği bu sıfır risk odaklı korumacı eğilim, gerçek

dünya SOC ve SIEM operasyonları için hedeflenen en güvenli mühendislik marjını temsil etmektedir.

4.5.3. Nicel Risk Sınıflandırması Başarımı

Sistemin ham OSINT metinlerinden elde ettiği siber tehditleri proaktif bir şekilde sınıflandırma kabiliyeti; araştırmacı yanlılığından tamamen arındırılmış, deterministik bir yapay zekâ muhakemesi ve istatistiksel konsensüs algoritması üzerine inşa edilmiştir. Otonom risk derecelendirme sürecinin temelinde, büyük dil modelinin olasılıksal doğasının katı istem şemaları ve rol atama teknikleriyle sınırlandırılması yatmaktadır. Model, metin içerisindeki varlık isimlerini ve tehdit vektörlerini anlamsal olarak analiz ederek teknik bir ciddiyet puanı üretmeye zorlanmıştır. Ancak, tekil bir model çıktısının kurumsal bir Güvenlik Operasyon Merkezi (SOC) için bilimsel güvenilirliği bulunmadığından, analiz edilen her bir veri modele bağımsız iterasyonlarla gönderilerek sanal bir konsensüs odası simüle edilmiştir. Bu iterasyonlar sonucunda üretilen risk puanlarının aritmetik ortalaması, standart sapması ve %95 Güven Aralığı hesaplanarak nihai ve deterministik bir risk skoru elde edilmiştir. İşlem maliyetlerini optimize etmek amacıyla başlangıçta 10 iterasyonla çalışan sistem, modelin ürettiği güven skorunun %40 ile %65 arasındaki kritik kararsızlık aralığına düştüğünü saptadığında insan müdahalesine gerek duymadan iterasyon çapını otonom olarak genişletmektedir. İşletilen bu dinamik ölçekleme mekanizması, istatistiksel hata payını daraltarak sınırda kalan vakaların doğru risk sınıfına atanmasını güvence altına almaktadır.

LLM motoru, analiz edilen her bir ham metin (tweet) için başarılı bir şekilde nicel ve kategorik sınıflandırma yapmıştır:

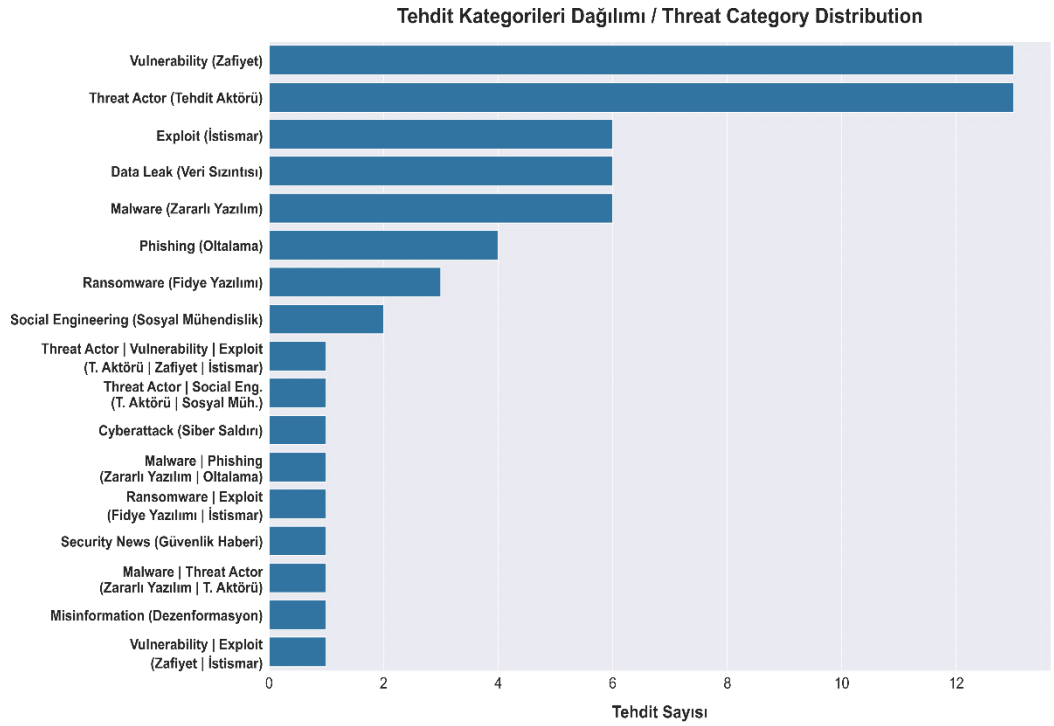


Şekil 8. Risk seviyesi dağılımı

Elde edilen istatistiksel veriler ve LLM motorunun otonom analizleri ışığında, kurumsal triyaj süreçlerini tetikleyecek nesnel risk eşikleri kurgulanmış ve pilot veri seti üzerindeki sınıflandırma bulguları Şekil 5'te görselleştirilmiştir. Buna göre; metin içerisinde aktif olarak sömürülen bir zafiyet kodu (CVE), sıfırıncı gün istismarı, fidye yazılımı yayılımı veya doğrulanmış bir veri sızıntısı barındıran durumlar "Kritik/Yüksek Risk" (High) sınıfına atanmış olup, analiz edilen toplam vakaların %64,5'ini (40 adet) oluşturmaktadır. Potansiyel etki alanı daha sınırlı olan ve aktif sömürüsü henüz kanıtlanmamış sektörel güvenlik açığı duyuruları ise "Orta Risk" (Medium) grubunda derecelendirilmiş ve dağılımın %32,3'lük (20 adet) dilimini temsil etmiştir. Kurum altyapılarını doğrudan tehdit etmeyen küresel siber güvenlik haberleri ve sektörel bilgilendirmeler ise operasyonel alarm yorgunluğu yaratmasının önüne geçmek amacıyla bilgilendirici sınıfa dahil edilmiş ve veri setinin yalnızca %3,2'sini (2 adet) oluşturmuştur. Bu yapısal dağılım, modelin karmaşık tehditleri kurumsal aciliyet düzeylerine göre yüksek bir kararlılıkla önceliklendirebildiğini kanıtlamaktadır.

4.5.4. Tehdit Vektörlerinin Sınıflandırılması ve Olay Analizi

Sistemin risk analiz katmanında sergilediği otonom ayrıştırma yeteneği, tehdit vektörlerinin kategorik olarak sınıflandırılması aşamasında da insan müdahalesinden bağımsız şekilde yürütülmüştür. Gerçekleştirilen kümülatif test bulgularında "Vulnerability" (Güvenlik Açığı - 13 adet) ve "Threat_Actor" (Tehdit Aktörü - 13 adet) kategorilerinin en yüksek yoğunlukta saptanması tesadüfi bir mühendislik çıktısı olmayıp, açık kaynak istihbaratı (OSINT) alanındaki siber güvenlik literatürünün yapısal dinamikleriyle doğrudan örtüşmektedir. X platformunun küresel güvenlik araştırmacıları ve beyaz şapkalı hackerlar tarafından yeni keşfedilen zafiyetlerin kamuya anlık olarak ifşa edildiği birincil erken uyarı ağı olarak işlev görmesi, sistem analiz motorunun "Vulnerability" kategorisinde yoğunlaşmasıyla tam bir bilimsel uyum içerisindedir. Güvenlik operasyon merkezlerinin ve analistlerin gelişmiş tehdit gruplarının veya fidye yazılımı çetelerinin aktif saldırı kampanyalarını sosyal medyada anlık analiz ederek paylaşmaları, "Threat_Actor" bağlamının da doğal olarak yüksek çıkmasını tetiklemektedir. Geliştirilen yerel LLaMA 3 motorunun metin içindeki bu aktör desenlerini yüksek hassasiyetle ayrıştırabilmesi, modelin siber güvenlik taksonomilerine MITRE ATTveCK / STIX) dair anlamsal ilişkilendirme derinliğini doğrulamaktadır.



Şekil 9. Kategori dağılımı

Entegrasyon döngüsünde "Exploit" (6 adet), "Data_Leak" (6 adet) ve "Malware" (6 adet) gibi kategorilerin de net bir şekilde ayırt edilmesi, tasarlanan yerel CTI analiz hattının tek boyutlu bir anahtar kelime filtresi olmadığını; aksine insan bilişini taklit ederek bütüncül bir durumsal farkındalık üretebildiğini göstermektedir. Özellikle tek bir olayda hem aktörü, hem zafiyeti hem de kullanılan zararlı yazılım ailesini çapraz etiketleyebilen karmaşık çoklu şemaların tespiti, Transformer tabanlı Dikkat Mekanizmasının uzun mesafe bağlamsal ilişkileri kurmadaki başarısını akademik olarak tescillemektedir.

4.5.4.1. Örnek Vaka İncelemesi: Otonom Zafiyet ve Aktör Tespiti

Geliştirilen otonom CTI analiz motorunun anlama ve yapısal veri çıkarma yeteneklerini göstermek amacıyla, veri toplama aşamasında tespit edilen gerçek bir siber güvenlik olayı detaylandırılmıştır. Süreç, TheHackersNews veya benzeri bir otorite kaynaktan otomasyon aracılığıyla çekilen yapılandırılmamış ve gürültülü bir ham metin ("URGENT: Active exploitation of CVE-2025-12345 in Microsoft Exchange Server detected in the wild. Threat actor APT29 is dropping web shells. Patch immediately! #cybersecurity #APT29") ile başlamaktadır. Bu ham metin, arka planda çalışan yerel LLaMA 3 motoruna iletilmiş ve model insan müdahalesine gerek kalmadan yapılandırılmış bir istihbarat JSON raporu üretmiştir.

4.5.4.2. Çıktı (LLM Tarafından Üretilen Yapısal JSON Formatı)

Büyük Dil Modeli (LLM) tarafından üretilen JSON formatındaki bu çıktı, karmaşık bir metnin bilgisayarlar tarafından kolayca okunup işlenebilecek "yapısal ve düzenli" bir rapora dönüştürülmüş halidir. Otonom olarak üretilen JSON çıktısının içerdiği veriler; modelin "is_threat" alanıyla metnin gerçek bir siber saldırı uyarısı olduğunu teyit etmesini ve "confidence_score" ile bu analize %95 oranında güvendiğini göstermesini kapsamaktadır. Sistem, olayın saldırı türünü "Zafiyet" (Vulnerability) olarak sınıflandırırken, "risk_level" ve "severity_score" alanlarıyla tehdidin aciliyetini "Kritik" (9/10) olarak otonom bir biçimde puanlamıştır. Bu yapı, kurumsal sistemlerin hangi siber olaya daha önce müdahale etmesi gerektiğini insan yardımı olmadan belirlemesini sağlamaktadır.

Dil bariyerinin aşılması bağlamında sistem, İngilizce olan ham metni sadece kelime kelime çevirmekle kalmayıp, "summary_tr" alanında olayın Türkçe özetini sunmakta ve "recommended_action_tr" alanı aracılığıyla sistem yöneticilerine "Sistemleri derhal yamayın..." şeklinde eylem planı önermektedir. Son olarak çıktı; metnin içine gizlenmiş olan "CVE-2025-12345" kodunu ve saldırgan grup ismini "APT29" olarak Standart Saldırı Göstergeleri (IoC) dizilerinde başarıyla ayrıştırmaktadır.

4.5.4.3. Vaka Değerlendirmesi

Bu örnek vaka incelendiğinde, sistemin ham İngilizce metindeki aciliyet vurgusunu doğru bir şekilde yorumlayarak risk seviyesini "Critical" (Kritik) ve “tehdit şiddetini” (severity_score) 10 üzerinden 9 olarak otonom bir şekilde belirlediği görülmektedir. Model, metin içerisine gizlenmiş olan "CVE-2025-12345" zafiyet kimliğini Standart Saldırı Göstergesi (IOC) formatında (cve_ids) başarıyla ayrıştırmış; eş zamanlı olarak "APT29" grubunu bir Tehdit Aktörü olarak etiketlemiştir. En kritik mühendislik başarılarından biri ise, sistemin yabancı dildeki (İngilizce) bir tehdit verisini eşzamanlı olarak Türkçe özetleyebilmesi (summary_tr) ve güvenlik analistleri için eyleme geçirilebilir Türkçe önleme stratejileri (recommended_action_tr) üretebilmesidir. Başarıyla yapılandırılan bu veri bloğu, doğrudan yerel SQLite veritabanına kaydedilmiş ve yüksek risk kategorisinde olduğu için SMTP uyarı mekanizmasını tetikleyerek ilgili ekiplere saniyeler içinde iletilmiştir.

5. SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde, gerçekleştirilen tez çalışması neticesinde elde edilen temel bulgular sentezlenmekte; geliştirilen otonom siber tehdit istihbaratı modelinin operasyonel çıktıları, literatüre sunduğu katkılar ve gelecekteki araştırma potansiyelleri kapsamlı bir şekilde değerlendirilmektedir.

Siber tehdit ortamının giderek karmaşıklaştığı günümüzde, geleneksel savunma mekanizmaları reaktif kalmakta ve saldırıları gerçekleşmeden önce öngörme konusunda yetersiz düşmektedir. Sosyal medya platformları gibi açık kaynak istihbaratı (OSINT) kanalları, yaklaşan tehditler hakkında zengin ve erken uyarı niteliğinde veriler sunmasına rağmen; bu verilerin devasa hacmi, gürültülü ve yapılandırılmamış doğası, manuel analiz süreçlerini imkânsız kılmaktadır. Bu tez çalışması, söz konusu yapısal darboğazı aşmak amacıyla; açık kaynak verilerini otonom olarak toplayan, anlamsal bütünlüğünü kavrayan ve karar vericiler için eyleme geçirilebilir net istihbarat paketlerine dönüştüren bütüncül bir Siber Tehdit İstihbaratı (CTI) modeli ortaya koymuştur.

5.1. Araştırma Boşluklarının Kapatılması ve Özgün Katkılar

Tez çalışmasının temel motivasyonu, güncel siber güvenlik operasyonlarında karşılaşılan üç temel problemi (veri egemenliği, bütüncül otomasyon eksikliği ve yapısal hata kırılganlığı) çözmek üzerine kurgulanmıştır. Elde edilen bulgular, geliştirilen modelin bu üç araştırma boşluğunu kurumsal güvenlik ihtiyaçlarına yanıt verecek düzeyde kapattığını göstermektedir.

Gerçekleştirilen bu araştırmanın literatüre ve sektörel uygulamalara sunduğu en temel katkı, "veri egemenliği" prensibinin siber güvenlik operasyonlarında pratik olarak uygulanabilirliğini kanıtlamış olmasıdır. Literatürde son dönemde önerilen yapay zekâ destekli istihbarat sistemlerinin büyük çoğunluğu, metin analizi için verileri üçüncü taraf bulut sunucularına ileterek kurumların hassas güvenlik mahremiyetini riske atmaktadır. Bu çalışmada kurgulanan yerel (on-premise) altyapı mimarisi (LLaMA 3), Büyük Dil Modellerinin üstün analitik yeteneklerini tamamen kapalı ve izole bir ağ ortamında kullanıma sunarak veri sızıntısı riskini tamamen ortadan kaldırmıştır. Bu durum, bütçe kısıtlaması olan veya regülasyonlara (KVKK, GDPR) tabi olan kurumsal siber güvenlik ekipleri, kamu kurumları ve kritik altyapılar için operasyonel maliyetleri sıfırlayan stratejik bir alternatiftir.

Çalışmanın literatüre sunduğu en radikal mühendislik katkısı ise "Yapısal Hata Toleransı" bağlamında geliştirilen Otonom JSON Sağlamlaştırma Protokolü'dür. Büyük

dil modelleri olasılıksal doğaları gereği, SIEM ve SOAR gibi deterministik güvenlik yazılımlarının beklediği katı veri şemalarını (JSON) bozma eğilimindedir. Sistemin 62 olay üzerinden gerçekleştirdiği 665 bağımsız LLaMA 3 iterasyonunda, 601 iterasyonda kusursuz üretim yapılmış, ancak 64 iterasyonda (%9.6 hata oranı) yapısal kırılmalar yaşanarak "JSONDecodeError" üretildiği saptanmıştır. Normal şartlarda tüm analiz hattını çökertecek olan bu durum, mimariye entegre edilen Regex tabanlı otonom protokolün devreye girerek "Satır Sonu Temizliği" uygulaması sayesinde anında yakalanmış, bozulan 64 çıktının tamamı onarılarak sisteme geri kazandırılmıştır. Bu müdahale, yapay zekânın üretkenliği ile kurumsal yazılımların katı kuralları arasında bir köprü kurmuş ve insan müdahalesine gerek kalmadan %100 kurtarma başarısı ile sıfır sistem çökme oranına ulaşarak "bütüncül otomasyon" hedefini gerçekleştirmiştir.

5.2. Performans Başarımı ve Literatürle Karşılaştırmalı Analiz

Modelin sınıflandırma performansına yönelik gerçekleştirilen genel sistem doğrulama testlerinde (TP: 43, TN: 6, FP: 13, FN: 0 metrikleri üzerinden) %79.03 Doğruluk (Accuracy), %76,79 Kesinlik (Precision), %86,87 F1-Skoru ve %100 Duyarlılık (Recall) başarısı elde edilmiştir. Kurumsal siber savunma stratejilerinde en hayati metrik olan Duyarlılık (Recall) oranının %100 seviyesinde gerçekleşmesi; sistemin, test edilen gürültülü OSINT verileri içerisindeki hiçbir gerçek siber tehdidi gözden kaçırmadığını (Sıfır Yanlış Negatif) ve kurumu savunmasız bırakmadığını ispatlamaktadır. Sistem, proaktif savunma refleksi gereği, nadiren de olsa teknik detayı zayıf olan bazı şüpheli haber metinlerini potansiyel risk kabul ederek "Yanlış Pozitif" (False Positive) olarak işaretleyebilmekte, bu durum Doğruluk (Accuracy) oranını %79,03 bandında konumlandırmaktadır. Siber istihbaratta bir tehdidi gözden kaçırmanın (FN) kurumsal maliyetinin, hatalı bir alarmı (FP) incelemekten çok daha yıkıcı olduğu gerçeği göz önüne alındığında, modelin sergilediği bu sıfır risk odaklı korumacı eğilim, operasyonel SOC ekipleri için oldukça başarılı ve güvenilir kabul edilmektedir.

Geliştirilen bu modelin literatürdeki güncel çalışmalarla karşılaştırılması, çalışmanın akademik değerini daha da belirginleştirmektedir. Literatürde tehdit istihbaratı için sıklıkla kullanılan SecureBERT gibi mimariler, siber güvenlik terminolojisine hakim olmalarına rağmen ince ayar ve karmaşık etiketleme süreçlerine bağımlı olan "Transformer-Encoder" modelleridir. Oysa bu çalışmada kullanılan LLaMA 3 modeli, "sıfır atışlı" (zero-shot) üretim yeteneği sayesinde hiçbir ince ayara ihtiyaç duymadan, geleneksel kural tabanlı sistemlerin aksine metnin niyetini okuyabilmektedir. Benzer şekilde, literatürdeki CTIKG ve LLM-TIKG gibi güncel çalışmalar, büyük dil

modellerini Tehdit İstihbaratı Bilgi Grafikleri (Knowledge Graph) oluşturmak için çoklu-ajanlı yapılarla kullanmaktadır. Ancak bu çalışmaların temel zafiyeti, LLM ajanlarının ürettiği grafiksel çıktıların, güvenlik duvarı veya SIEM sistemleri tarafından anlık olarak işlenemeyecek kadar karmaşık olması ve format bozulmalarına karşı tolerans mekanizmalarının bulunmamasıdır. Bu tez kapsamında önerilen sistem ise, CTIKG ve LLM-TIKG'nin aksine doğrudan olay müdahalesine odaklanarak, makine okunabilir (JSON) IOC'leri hata toleranslı ve deterministik bir şekilde üretmekte; böylece teorik bilgi çıkarımını aşarak doğrudan SOC ekiplerine entegre edilebilir bir yapay zekâ hattı sunmaktadır.

5.3. Kurumsal Siber Güvenlik Operasyonlarına Pratik Katkılar

Geliştirilen modelin en önemli operasyonel katkısı, Güvenlik Operasyon Merkezlerindeki (SOC) Birinci Seviye (Tier-1) analistlerin yerine geçerek triyaj sürecini otonomlaştırmasıdır. Sistem; karmaşık, yabancı dildeki gürültülü OSINT verilerini okumakta, zafiyetin kritiklik derecesine karar vermekte ve hedef kurumun dilinde (Türkçe) uygulanabilir bir eylem planı üreterek e-posta yoluyla iletmektedir. Analistlerin üzerindeki yüksek veri işleme yükünü ve "alarm yorgunluğunu" azaltan bu yapı, tespit ile müdahale arasındaki süreyi (MTTR) minimize etmektedir.

Kamu kurumları, bankalar ve kritik altyapı işletmecileri açısından bu mimarinin pratik değeri çok büyüktür. Tespit edilen IOC'lerin (IP, Domain, CVE) otonom olarak JSON şemaları halinde SIEM platformlarına beslenmesi, insan analistlerin güncel tehditleri okuyup manuel engelleme kuralı yazma zorunluluğunu ortadan kaldırmakta, potansiyel siber saldırıların kurum ağına ulaşmadan önce sınır bilişim noktasında durdurulabilmesine zemin hazırlamaktadır.

5.4. Çalışmanın Sınırlılıkları

Araştırma bulgularının operasyonel geçerliliğine rağmen, çalışmanın birtakım metodolojik sınırlılıkları bulunmaktadır. İlk olarak, sistemin genel karar başarısı 62 adet tehdit verisi üzerinden hesaplanmıştır. Bu göreceli dar örneklem boyutu; modelin otonom kararlarını istatistiksel olarak doğrulamak için her bir verinin siber güvenlik uzmanı tarafından manuel olarak etiketlenmesi (Altın Standart - Ground Truth) zorunluluğundan ve LLaMA 3 motorunun yerel donanımda (24 GB RAM, GT 730 GPU) koşturulmasının yarattığı işlem süresi maliyetinden (62 vaka için 665 iterasyonluk döngünün 35,6 dakikada tamamlanması) kaynaklanmıştır. İlerleyen süreçte daha güçlü GPU kümeleri ile

binlerce verilik setler test edildiğinde, %79,03'lük mevcut doğruluk (Accuracy) oranında dalgalanmalar yaşanması muhtemeldir.

İkinci olarak, sistemin veri sağlama (OSINT) kaynağı yalnızca X (Twitter) platformundaki otorite hesaplarla sınırlandırılmıştır. Farklı sosyal ağların kendilerine has dil yapıları (argo, teknik jargon yoğunluğu) bu çalışmada test edilmemiştir. Son olarak, LLaMA 3 modeli standart ağırlıklarıyla (zero-shot) kullanılmış olup, siber tehdit istihbaratına özgü geniş çaplı bir veri setiyle özel bir ince ayar (fine-tuning) işlemine tabi tutulmamıştır.

5.5. Gelecek Çalışmalar ve Öneriler

Bu tez çalışması, Büyük Dil Modellerinin açık kaynak istihbaratı süreçlerindeki devasa potansiyeline yönelik kavramsal ve mimari bir temel atmıştır. Geliştirilen sistem mimarisinin operasyonel kapasitesinin artırılması ve literatürdeki diğer boşlukların doldurulması adına gelecek araştırmalara yönelik çeşitli öneriler getirilmektedir. Öncelikle veri entegrasyonunun genişletilmesi bağlamında; mevcut sisteme X platformunun yanı sıra siber suçluların aktif olduğu karanlık ağ (Dark Web) sızıntı forumları, Telegram kanalları, Reddit siber güvenlik alt dizinleri, RSS beslemeleri ve Ulusal Zafiyet Veritabanları (NVD/CVE) eşzamanlı veri kaynakları olarak entegre edilmelidir.

Açık kaynak modellerin performansının artırılması amacıyla ince ayar (fine-tuning) işlemleri gündeme alınmalıdır. Genel geçer LLaMA modelinin, siber güvenlik uzmanlarınca doğrulanmış geniş çaplı CTI raporları ve MITRE ATTveCK teknikleriyle eğitilerek alana özgü (Domain-Specific) bir modele dönüştürülmesi, saldırı tekniklerini öngörme ve tehdit aktörlerini tanıma başarımını en üst seviyeye taşıyacaktır. Ayrıca, SOAR (Güvenlik Orkestrasyonu, Otomasyonu ve Yanıt) platformları ile tam otonom reaksiyon yeteneğinin geliştirilmesi gerekmektedir. Çalışmanın bir sonraki evresinde, sistemin yalnızca SIEM platformlarına log göndermekle kalmayıp, SOAR sistemleriyle otonom REST API bağlantı noktaları üzerinden entegre edilmesi hedeflenmelidir. Böylece tespit edilen yüksek riskli saldırgan IP'lerinin kurumsal Firewall üzerinden insan onayı olmadan bloklanabildiği tam otonom aktif savunma mekanizmaları geliştirilebilecektir. Son olarak, yapay zekâ tabanlı sistemlerin analiz süreçlerindeki kararlılığını maksimize etmek amacıyla gelecekteki mimarilerde "Çoklu-Ajan (Multi-Agent)" yapıları kurgulanmalıdır. Veriyi toplayan, risk skorunu hesaplayan ve elde edilen bulguları bağımsız olarak doğrulayan farklı yapay zekâ ajanlarının birbirini denetlediği

sistem tasarımları, otonom siber güvenlik çözümleri için oldukça değerli bir araştırma alanı olarak öne çıkmaktadır.

KAYNAKÇA

Afane, K., Wei, W., Mao, Y., Farooq, J., ve Chen, J. (2024, Aralık). Next-generation phishing: How LLM agents empower cyber attackers. *2024 IEEE International Conference on Big Data (BigData)* içinde (ss. 2558-2567). IEEE.

Alturkistani, H., ve Chuprat, S. (2024). Artificial intelligence and large language models in advancing cyber threat intelligence: A systematic literature review. *Artificial Intelligence Review*.

Ayoade, G., Chandra, S., Khan, L., Hamlen, K., ve Thuraisingham, B. (2018, Ekim). Automated threat report classification over multi-source data. *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)* İçinde (ss. 236-245). IEEE.

Balasubramanian, P., Liyana, S., Sankaran, H., Sivaramakrishnan, S., Pusuluri, S., Pirttikangas, S., ve Peltonen, E. (2025). Generative AI for cyber threat intelligence: Applications, challenges, and analysis of real-world case studies. *Artificial Intelligence Review*, *58*(11), 1-134.

Brown, R., ve Lee, R. M. (2019). *The evolution of cyber threat intelligence (CTI): 2019 SANS CTI survey*. SANS Institute.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... ve Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877-1901.

Clement, M. (2025). *Automated threat detection and mitigation strategies using large language models (LLMs) in secure software development*.

Dale, D. S., Mcclanahan, K., Elder, W. S., ve Li, Q. (2025). Twitter-based OSINT for cyber event analytics. *IEEE Access*.

De Bellis, A. (2023, Kasım). Structuring the unstructured: An LLM-guided transition. *International Semantic Web Conference (ISWC)* içinde.

Ferrag, M. A., Ndhlovu, M., Tihanyi, N., Cordeiro, L. C., Debbah, M., Lestable, T., ve Thandi, N. S. (2024). Revolutionizing cyber threat detection with large language models: A privacy-preserving BERT-based lightweight model for IoT/IIoT devices. *IEEE Access*, *12*, 23733-23750.

Fieblinger, R., Alam, M. T., ve Rastogi, N. (2024, Temmuz). Actionable cyber threat intelligence using knowledge graphs and large language models. *2024 IEEE*

European Symposium on Security and Privacy Workshops (EuroSvePW) içinde (ss. 100-111). IEEE.

Guru, K., Moss, R. J., ve Kochenderfer, M. J. (2025). On technique identification and threat-actor attribution using LLMs and embedding models. *arXiv preprint arXiv:2505.11547*.

Hu, Y., Zou, F., Han, J., Sun, X., ve Wang, Y. (2024). LLM-TIKG: Threat intelligence knowledge graph construction utilizing large language model. *Computers ve Security*, *145*, 103999.

Huang, L., ve Xiao, X. (2024, Ekim). CTIKG: LLM-powered knowledge graph construction from cyber threat intelligence. *First Conference on Language Modeling*.

Huff, P., ve Li, Q. (2021, Eylül). Towards automated assessment of vulnerability exposures in security operations. *International Conference on Security and Privacy in Communication Systems* içinde (ss. 62-81). Springer International Publishing.

Jakstaite, D., ve Czekster, R. M. (2023, Kasım). Extracting cyber threat intelligence from social media with case studies in Twitter/X and Reddit. *International Symposium: From Data to Models and Back* içinde (ss. 46-65). Springer Nature Switzerland.

Jelodar, H., Bai, S., Hamedi, P., Mohammadian, H., Razavi-Far, R., ve Ghorbani, A. (2025). Large language model (LLM) for software security: Code analysis, malware analysis, reverse engineering. *arXiv preprint arXiv:2504.07137*.

Ji, H., Yang, J., Chai, L., Wei, C., Yang, L., Duan, Y., ... ve Li, Z. (2024). SevenLLM: Benchmarking, eliciting, and enhancing abilities of large language models in cyber threat intelligence. *arXiv preprint arXiv:2405.03446*.

Jurafsky, D., ve Martin, J. H. (2020). *Speech and language processing* (3. baskı). https://web.stanford.edu/jurafsky/slp3/

Khatami, S., ve Frantz, C. (2024). Prompt engineering guidance for conceptual agent-based model extraction using large language models. *arXiv preprint arXiv:2412.04056*.

Krishnamurthy, O. (2023). Enhancing cyber security enhancement through generative AI. *International Journal of Universal Science and Engineering*, *9*(1), 35-50.

Lee, R. M. (2020). *2020 SANS Cyber Threat Intelligence (CTI) survey* (Tech. Rep.). SANS Institute.

Mahboubi, A., Luong, K., Aboutorab, H., Bui, H. T., Jarrad, G., Bahutair, M., ... ve Gately, H. (2024). Evolving techniques in cyber threat hunting: A systematic review. *Journal of Network and Computer Applications*, *232*, 104004.

Martin, J. H., ve Jurafsky, D. (2019). *Vector semantics and embeddings*. Speech Lang. Process.

Mior, M. J. (2024). Large language models for JSON schema discovery. *arXiv preprint arXiv:2407.03286*.

Mo, Y., Qin, H., Dong, Y., Zhu, Z., ve Li, Z. (2024). Large language model (LLM) AI text generation detection based on transformer deep learning algorithm. *arXiv preprint arXiv:2405.06652*.

Nath, H. V., ve Mehtre, B. M. (2014, Mart). Static malware analysis using machine learning methods. *International Conference on Security in Computer Networks and Distributed Systems* içinde (ss. 440-450). Springer.

Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C. P., ... ve Wu, Q. J. (2022). A review of generalized zero-shot learning methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(4), 4051-4070.

Purba, M. D., ve Chu, B. (2023, Ekim). Extracting actionable cyber threat intelligence from Twitter stream. *2023 IEEE International Conference on Intelligence and Security Informatics (ISI)* içinde (ss. 1-6). IEEE.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., ve Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.

Tellache, A., Korba, A. A., Mokhtari, A., Moldovan, H., ve Ghamri-Doudane, Y. (2025). Advancing autonomous incident response: Leveraging LLMs and cyber threat intelligence. *arXiv preprint arXiv:2508.10677*.

Tounsi, W. (2019). What is cyber threat intelligence and how is it evolving? *Cyber-Vigilance and Digital Trust: Cyber Security in the Era of Cloud Computing and IoT* içinde (ss. 1-49).

Urgan, S., ve Erdoğan, P. (2024). *Dijital organizasyonlarda siber saldırılar ve siber güvenlik*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... ve Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... ve Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.

Xu, H., Wang, S., Li, N., Wang, K., Zhao, Y., Chen, K., ... ve Wang, H. (2024). Large language models for cyber security: A systematic literature review. *ACM Transactions on Software Engineering and Methodology*.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	: Sait Emre ORAL

Eğitim Durumu

Lisans Öğrenimi	: Endüstri Mühendisliği
Yüksek Lisans Öğrenimi	: Yönetim Bilişim Sistemleri Anabilim Dalı
Bildiği Yabancı Diller	:
Bilimsel Faaliyetler ve Yayınlar	:

İş Deneyimi

Stajlar	:
Projeler	:
Çalıştığı Kurumlar	: Gümüşhane Üniversitesi Bilgi İşlem Daire Başkanlığı

Tarih **:** 05.08.2026