

LLM-BASED ANALYSIS FOR DISCOVERING CYBERSECURITY PATTERNS IN PUBLIC ADMINISTRATION CYBER EVENTS

Puya PAKSHAD*, Maurice DAWSON*, Enkel HOXHA**, Abdul Hadi KHAN*

*Illinois Institute of Technology, Chicago, USA

**AAB College, Pristina, Kosovo

ppakshad@hawk.illinoistech.edu, maurice.e.dawson@gmail.com, enkelh21@gmail.com,
akhan172@hawk.illinoistech.edu

Abstract: Public Administration cyber event records often contain incomplete, ambiguous, or undetermined information, which makes it difficult to understand incident behavior, affected government functions, and hidden cybersecurity risks through structured data alone. This paper proposes an LLM-assisted semantic analysis approach to interpret public-sector cyber event records by combining available structured attributes with narrative incident descriptions. The framework defines three interpretable metrics: Incident Security Semantics, Governance Function Criticality, and Cyber Intelligence Gap and Inference. These metrics capture the security meaning of each incident, the criticality of the affected public-sector function, and the intelligence value of incomplete records. The goal is to uncover security patterns that are not directly visible in the original records, including service disruption, sensitive data exposure, public safety impact, governance-related risk, and unresolved intelligence gaps. The proposed approach transforms fragmented event information into interpretable cybersecurity insights and supports context-aware analysis of public-sector incidents. The results demonstrate that the proposed LLM-assisted analysis can reveal recurring hidden patterns in cyber events and provide cyber threat intelligence for understanding risks to government services, citizen-facing systems, and sensitive public-sector operations.

Keywords: cyber event analysis, large language models, public administration, hidden cybersecurity patterns

1. Introduction

In recent years, data-centric cybersecurity analysis has become an important research direction for understanding cyber threats, incident behavior, and organizational exposure. Cybersecurity data can be collected from different sources, including network traffic, cyber threat intelligence feeds, vulnerability repositories, incident reports, malware analysis, security logs, open-source intelligence, and social media platforms [1]. These data sources help researchers and analysts study attack trends, threat actor behavior, affected domains, attack motives, and the operational impact of cyber incidents. Accordingly, historical

cyber event datasets provide a useful source of cyber intelligence because they record real-world cyber. Public Administration is one of the most important sectors in cyber incident analysis because it is connected to government operations, public services, law enforcement, taxation, emergency response, public safety, and digital services. Therefore, studying Public Administration cyber events is essential for cyber threat intelligence. Despite the growing use of open-source cyber event repositories, their analytical value is often constrained by incomplete or undetermined fields. Key attributes such as attacker motive, actor attribution, impact scope, and event type are

often missing, inconsistently reported, or only partially described [2,3]. This limitation makes it difficult to track incidents and extract deeper knowledge from the reports. More importantly, many strategic cybersecurity questions cannot be answered directly from structured dataset fields alone, such as which public-sector functions are repeatedly affected when attribution or motive is unknown and what forms of citizen-service disruption, or sensitive data risk can be inferred from incomplete records. Therefore, there is a need for a semantic analysis that can transform incomplete and semi-structured cyber event records into interpretable cybersecurity patterns for Public Administration cyber threat intelligence.

To address this limitation, we propose an large language model (LLM)-assisted semantic analysis framework for uncovering hidden cybersecurity patterns in Public Administration cyber event records. Rather than depending only on structured fields or descriptive summaries, the proposed framework uses LLM-based reasoning to interpret incident narratives, contextual evidence, and public-sector relevance. In general, the main contributions of this paper are as follows. First, we propose an LLM-assisted semantic analysis framework designed to analyze Public Administration cyber event records and uncover hidden cybersecurity patterns that are not directly observable from structured fields alone. Second, we introduce three interpretable analytical metrics—Incident Security Semantics, Governance Function Criticality, and Cyber Intelligence Gap and Inference—that support the semantic interpretation of cyber events from security, governance, and intelligence perspectives. Third, we propose a comprehensive analysis shows the hidden-pattern representation that integrates these three metrics to capture recurring cybersecurity themes in public-sector incidents.

2. Background and Related Work

Cyber threat intelligence studies use

heterogeneous data sources, including logs, vulnerability repositories, incident reports, OSINT, and threat feeds, to extract evidence about attacks, affected sectors, adversary behavior, and operational impact [1,2]. Since many cyber event records are semi-structured, incomplete, or expressed in natural language, previous work emphasizes the need for NLP-based methods to interpret both explicit fields and contextual incident evidence [1,2]. Previous work on cyber threat intelligence illustrates that unstructured cybersecurity texts contain valuable information about incidents, vulnerabilities, threat actors, attack techniques, and security risks [1]. NLP-based methods such as feature extraction, topic modeling, entity recognition, and semantic analysis have therefore been used to transform textual CTI sources into structured and interpretable knowledge [1,6]. Building on this line of research, several automated CTI extraction studies show that narrative reports can be used to identify attack vectors, tools, vulnerabilities, threat entities, and their semantic relationships [2,7,8]. However, these approaches focus mainly on technical CTI extraction, while Public Administration incident analysis also requires interpretation of public functions, governance relevance, service disruption, and organizational exposure. This distinction is important because public-sector cyber incidents are not only technical security events; they may also affect administrative continuity, citizen-facing services, regulatory operations, and sensitive public data.

A related stream of research focuses on cyber threat attribution. Cyber threat attribution studies show that unstructured CTI reports can support the identification of threat actors by extracting contextual and behavioral features [3,11]. Since attribution-related fields such as attacker identity, motive, impact scope, and event type are often uncertain or incomplete, cyber event records require semantic interpretation beyond direct field-based analysis [3,5,14].

This issue is particularly relevant for Public Administration cyber event repositories, where attribution and motive may remain unknown even when the incident has governance implications.

NLP-based studies on emerging cyber threats further show that public and unstructured sources can be used to identify new threats and characterize threat behavior [4,15]. For Public Administration incidents, such sources may lack complete technical details but can still reveal patterns related to affected organizations, service disruption, sensitive data exposure, and public-sector consequences [1,4,6]. Therefore, public cyber event records can provide useful intelligence even when they are incomplete, if they are analyzed through methods capable of interpreting narrative and contextual evidence.

Although these methods help connect fragmented cybersecurity evidence, most focus on technical cyber-entity relationships rather than governance-oriented interpretation of public-sector incidents [7,8,17]. As a result, existing semantic approaches provide basis foundations for structuring cyber knowledge, but they do not properly address how cyber incidents affect public-sector functions and administrative responsibilities.

Recent LLM-based cybersecurity studies show their power in threat intelligence synthesis, incident response, security analytics, CTI extraction, and knowledge graph construction [5,7,9,10]. However, these studies address broad automation or technical CTI tasks, while less attention has been given to interpretable LLM-assisted metrics for analyzing incomplete Public Administration cyber event records from security, governance, and intelligence-gap perspectives [1–10]. To address this gap, this paper proposes an LLM-assisted semantic analysis framework based on three metrics: Incident Security Semantics, Governance Function Criticality, and Cyber Intelligence Gap and Inference. The framework extends prior CTI, NLP, semantic modeling, and

LLM-based research by transforming incomplete and semi-structured Public Administration cyber event records into interpretable hidden-pattern representations for context-aware cyber threat intelligence [11–20].

3. Proposed Design

Figure 1 displays the architecture of the proposed LLM-assisted semantic analysis framework for Public Administration cyber event records. The input of this architecture is a raw cyber event dataset containing structured fields and textual incident descriptions. In the first step, Public Administration-related cyber events are collected and assigned unique event identifiers. In the second step, the dataset is prepared by removing duplicates, standardizing fields, reducing noise, preserving missing and undetermined values, and selecting the features required for semantic analysis. In this study, the selected features include event date, organization, motive, event subtype, magnitude, scope, description, victim country, and actor country. In the third step, the prepared records are fed into LLM-processable batches and analysed via a fixed prompting strategy based on three metrics: Incident Security Semantics, Governance Function Criticality, and Cyber Intelligence Gap and Inference. In the fourth step, these metric outputs are integrated to generate M(TOTAL), which represents the hidden cybersecurity pattern of each incident. Lastly, the verification module checks the consistency, traceability, and evidence support of the LLM-generated outputs. Thus, the architecture consists of six main components: data collection, dataset preparation, LLM prompting strategy, LLM-based semantic analysis, security hidden pattern reasoning, and verification, which are described in more detail in the following subsections.

3.1 Data Collection

The data collection module obtains cyberevent records related to Public

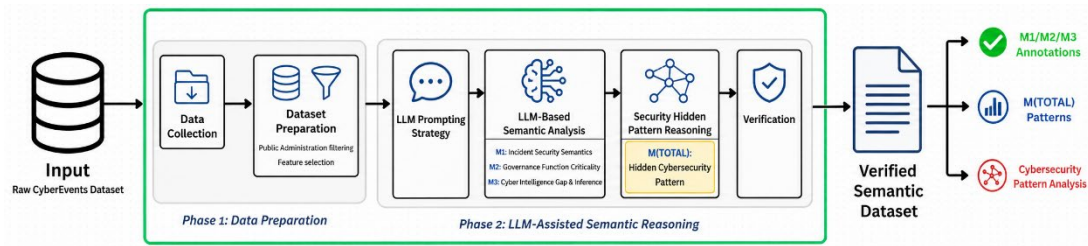


Figure 1- Proposed Design

Administration, including incidents affecting government agencies, public services, law enforcement, taxation, emergency response, public safety, public digital platforms, and other administrative functions. The collected records include structured fields such as event date, victim country, affected organization, motive, event subtype, magnitude, scope, and actor country, as well as semi-structured or textual descriptions that provide context about the cyber incident. Each record is stored with its original textual description and available metadata to preserve the contextual information required for semantic analysis. This is important because many Public Administration cyber events contain missing, null, or undetermined values in fields such as attacker country, motive, impact scope, operational magnitude, or event subtype, while the incident narrative may provide indirect evidence about the affected public function, possible service disruption, sensitive data exposure, or governance-related impact. The output is a raw cyber event dataset in which each record has a unique event identifier for later verification and traceability.

3.2 Dataset Preparation

The dataset preparation module transforms the raw Public Administration cyber event dataset into a structured input format suitable for LLM-based semantic analysis. The main objective is to retain the original meaning of each incident while reducing noise, inconsistency, duplicate records, irrelevant symbols, and non-informative textual elements. Incomplete records are not removed because missing and undetermined values are important for analysing cyber

intelligence gaps. After filtering Public Administration-related records, the module standardizes naming formats, harmonizes overlapping fields such as target name and affected organization, and selects the features required for semantic interpretation. The final selected feature set includes event identifier, event date, organization, motive, event subtype, magnitude, scope, description, victim country, and actor country. These attributes are retained because they capture the temporal context of the incident, the affected public-sector entity, the reported security meaning and operational impact, the incident narrative, and the available attribution context.

The incident description is treated as a key feature because it may contain contextual evidence not available in structured fields, such as data exposure, service disruption, ransomware, unauthorized access, public safety impact, or citizen-service interruption. The scope and magnitude fields are also preserved because they may describe which systems, services, or government functions were affected and how severe the reported impact was. Similarly, motive and actor country are preserved even when they are marked as “Undetermined” because such values are useful for identifying cyber intelligence gaps.

3.3 LLM Prompting Strategy

The LLM prompting strategy guides the model toward consistent, interpretable, and evidence-based semantic analysis. Instead of asking the LLM to summarize each cyber incident, the prompt instructs it to evaluate each record according to the predefined objectives of the framework. Each prompt

includes task instruction, metric definitions, incident records, and the required output format. The task instruction defines the LLM role as a cybersecurity analyst focusing on Public Administration cyber events. The metric definitions explain Incident Security Semantics, Governance Function Criticality, and Cyber Intelligence Gap and Inference. The incident records include the selected structured fields and the original narrative descriptions, while the output format requires structured labels, short explanations, evidence cues, and uncertainty indicators. The same prompt is applied for all batches to support consistency and reproducibility. Each batch contains a couple of records, and the LLM is instructed to produce outputs for every event identifier included in the batch. To improve reliability, the LLM is instructed to use only the information provided in the incident record. If evidence is insufficient, the result must be marked as unknown, unclear, or partially inferable rather than inventing unsupported details. The output of this module is a structured LLM response for each cyber event record, which is later written into an annotation dataset.

3.4 LLM-Based Semantic Analysis

The LLM-based semantic analysis module evaluates each prepared cyber event record using three interpretable metrics: Incident Security Semantics (M1), Governance Function Criticality (M2), and Cyber Intelligence Gap and Inference (M3). These metrics capture different dimensions of Public Administration cyber incidents and extract meanings that may not be directly available from structured fields.

Incident Security Semantics (M1) identifies the cybersecurity meaning of the incident itself. It uses the description, scope, magnitude, event subtype, and motive fields to determine whether the event reflects ransomware, data exposure, denial-of-service, phishing, unauthorized access, website compromise, malware infection, vulnerability exploitation, espionage, sabotage, service disruption, or operational

degradation. This metric focuses on what the incident means from a security perspective. Governance Function Criticality (M2) evaluates the affected public-sector function. It uses the organization, victim country, description, and scope fields to determine whether the incident relates to citizen services, public safety, emergency response, taxation, legal administration, identity services, public records, municipal operations, government communication, public digital platforms, or other governance functions. This metric focuses on the type and sensitivity of the affected public-sector target rather than only the technical form of the attack. Cyber Intelligence Gap and Inference (M3) assess uncertainty and missing information in the record while also determining whether the remaining semantic evidence still supports useful cyber intelligence inference. It uses motive, event subtype, magnitude, scope, description, organization, victim country, actor country, and event date. This metric identifies whether fields such as motive, actor country, event subtype, scope, or magnitude are missing, null, or marked as “Undetermined,” and whether the description or other retained fields still provide meaningful evidence about public-sector impact, service disruption, data exposure, or governance sensitivity. For each metric, the LLM produces a semantic label, short justification, evidence cue, and uncertainty indicator. The output is a metric-level semantic annotation dataset containing M1, M2, and M3 for each incident. This annotation dataset is stored with the event identifier so that each metric-level output remains traceable to the corresponding original record.

3.5 Security Hidden Pattern Reasoning

The security hidden pattern reasoning module integrates M1, M2, and M3 to generate an overall hidden-pattern interpretation for each cyber event, referred to as M(TOTAL). While the individual metrics describe separate dimensions of the incident, M(TOTAL) captures the combined

meaning of the event from security, governance, and intelligence-gap perspectives. The input of this module is the annotation dataset generated in the previous step, including M1, M2, M3 labels, explanations, evidence cues, and uncertainty indicators. This module identifies recurring patterns that are not directly visible from structured fields alone. For example, incidents with different event subtypes may still represent the same hidden pattern if they involve disruption of citizen-facing digital services, repeated targeting of municipal infrastructure, exposure of public record systems, public safety impact, or unresolved attribution despite observable operational consequences. The LLM reasons over the metric outputs rather than over the raw record alone. It then assigns an M(TOTAL) interpretation and explains how the three metrics interact. For example, an event may be interpreted as “public safety disruption with unresolved attribution” when M1 indicates operational disruption, M2 identifies a public safety or law enforcement function, and M3 indicates that motive or actor country is missing or undetermined. Similarly, an event may be interpreted as “citizen-service data exposure” when M1 identifies data exposure, M2 identifies a citizen-facing government service, and M3 shows that incomplete structured fields still contain enough narrative evidence for inference. The output is an integrated semantic representation that supports incident comparison, event grouping, recurring threat-theme identification, and context-aware cyber threat intelligence analysis.

We can formulate the selected feature set and the semantic metrics as follows:

$$X_i = \{ED_i, OR_i, MO_i, ES_i, MA_i, SC_i, DE_i, VC_i, AC_i\} \quad (1)$$

$$M1_i = f1(ES_i, MO_i, MA_i, SC_i, DE_i, AC_i) \quad (2)$$

$$M2_i = f2(ED_i, OR_i, VC_i, MA_i, SC_i, DE_i) \quad (3)$$

$$M3_i = f3(ED_i, MO_i, ES_i, MA_i, SC_i, DE_i, VC_i, AC_i) \quad (4)$$

$$MTOTAL_i = f4(M1_i, M2_i, M3_i) \quad (5)$$

where, X_i = selected features of incident i , ED_i = event date, OR_i = organization, MO_i = motive, ES_i = event type, MA_i = magnitude, SC_i = scope, DE_i = description, VC_i = victim country, AC_i = actor country, $M1_i$ = Incident Security Semantics of incident i , $M2_i$ = Governance Function Criticality of incident i , $M3_i$ = Cyber Intelligence Gap and Inference of incident i , and $MTOTAL_i$ = hidden-pattern representation of incident i .

3.6 Verification

The verification module checks whether the LLM-generated outputs are consistent, traceable, and supported by the original incident records. This module is necessary because LLM-based analysis may produce unsupported, overly confident, or conceptually inconsistent interpretations. Verification is performed at three levels. First, the output format is checked to ensure that each incident includes the event identifier, M1, M2, M3, M(TOTAL), explanations, evidence cues, and uncertainty indicators. Second, the generated explanations and evidence cues are compared with the original selected features and incident description to ensure that the model does not introduce information that is not present or reasonably inferable from the record. Third, the consistency between metric labels, explanations, and M(TOTAL) is reviewed to ensure that the final hidden-pattern interpretation is supported by M1, M2, and M3. The module also supports traceability by linking each generated result to the unique event identifier assigned during data collection. If an output is unsupported, contradictory, too uncertain, or not clearly connected to the selected input fields, the record can be flagged for manual review. The final output is a verified semantic dataset containing the event identifier, selected input features, M1, M2, M3, M(TOTAL), explanations, evidence cues, uncertainty indicators, and verification status.

4. Experimental Analysis

This section presents the experimental analysis of the proposed LLM-assisted semantic framework. The goal is to examine whether Public Administration cyber event records can be transformed into semantic indicators and hidden cybersecurity patterns. The experiment follows the proposed architecture by preparing the dataset, processing records through LLM-based semantic analysis, generating M1, M2, M3, deriving M(TOTAL), and analysing the resulting hidden patterns.

4.1 Setup

The experiment uses cyber event records from the CyberEvents dataset [11]. Since this paper focuses on Public Administration, the dataset was first filtered to retain only records related to government operations, public services, taxation, law enforcement, emergency response, public safety, and citizen-facing digital platforms. After filtering, nine attributes were retained: event date, organization, motive, event subtype, magnitude, scope, description, victim country, and actor country. These features were selected because they capture the time of the event, the affected public-sector entity, the incident narrative, the reported impact, and the available attribution context. A total of 3,000 Public Administration records were selected and divided into 100 batches, with 30 records in each batch. A fixed prompt was used for all batches. In the first stage, the LLM generated three semantic indicators for each record: Incident Security Semantics (M1), Governance Function Criticality (M2), and Cyber Intelligence Gap and Inference (M3). In the second stage, M1, M2, and M3 were provided to the LLM again to generate M(TOTAL), which represents the hidden cybersecurity pattern of each incident.

4.2 Dataset Statistics

Table 1 summarizes the experimental dataset configuration. The dataset contains 3,000 Public Administration cyber event records.

These records include both structured attributes and narrative descriptions. A key characteristic of the dataset is the frequent presence of incomplete, null, ambiguous, or “Undetermined” values in fields such as motive, event subtype, magnitude, scope, and actor country. This incompleteness is important because the proposed design investigates whether meaningful cyber intelligence can still be extracted from incomplete records. In many cases, the description field provides contextual evidence about service disruption, data exposure, unauthorized access, or public safety impact. In addition, The dataset focuses on cyber event records from the United States, the United Kingdom, Australia, and Canada.

Table 1. Dataset Configuration

#	Dataset	CyberEvents [11]
1	Filtered sector	“Public Administration“
2	# of records	3,000
3	Selected features	event_date, organization, motive, event_subtype, magnitude, scope, description, country, actor_country
4	Batch size	30
5	# of batches	100

4.3 Hidden Pattern Generation

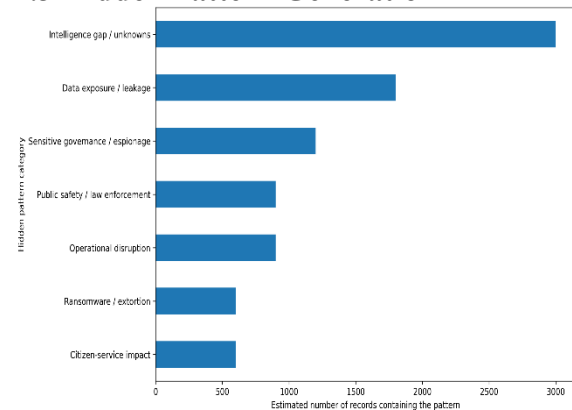


Figure 2- Frequency distribution of hidden cybersecurity patterns identified from 3,000 Public Administration cyber event records using the proposed LLM-assisted semantic analysis framework

In this step, the security records were processed using a Python script¹ connected

¹ <https://github.com/ppakshad/LLM-BASED-ANALYSIS-FOR-DISCOVERING-CYBERSECURITY-PATTERNS-IN-PUBLIC-ADMINISTRATION-CYBER-EVENTS.git>

to the GPT-5.2 engine through an API key. The script iterated over 100 batches, where each batch contained 30 records, and submitted the selected fields with the fixed prompt described in Section 3.3. For each record, the LLM generated three semantic annotations: M1 for the security meaning of the incident, M2 for the affected government function, and M3 for the cyber intelligence gap and inferential value of incomplete fields. The generated M1, M2, and M3 outputs were written into a new annotation dataset using the event identifier for traceability. In the second loop, the annotation dataset was provided again to the LLM, and M1, M2, and M3 were conceptually merged to produce M(TOTAL), which represents the hidden cybersecurity pattern of each incident. After generating M(TOTAL), we obtained about 3,000 enriched records, each describing the hidden security meaning of one Public Administration cyber event. These outputs were then analyzed by the LLM to find repeated patterns across the dataset. This step discovers important hidden patterns, such as service disruption, public safety impact, governance exposure, data leakage, and unresolved attribution.

4.5 Analysis and Key Findings

After generating M(TOTAL) for the selected Public Administration records, the enriched outputs were analyzed to identify recurring hidden cybersecurity patterns. This analysis was reviewed by two cybersecurity experts to examine whether the LLM-generated interpretations provide meaningful security insights beyond the visible structured fields of the original dataset. The purpose of this step is not only to count cyber events, but to understand how incomplete records can still reveal hidden relationships among incident behavior, affected public-sector functions, and intelligence gaps. As shown in Figure 2, the most frequent hidden pattern is Intelligence Gap / Unknowns. This finding is valuable from a cybersecurity perspective because incident response and threat intelligence teams often need to make

decisions even when attribution, motive, scope, or impact details are incomplete. The analysis shows that records with missing fields are not necessarily low-value; their descriptions may still reveal important signs of service disruption, data exposure, public safety impact, or governance-related consequences. Therefore, the framework helps analysts prioritize incomplete but security-relevant incidents for further investigation. The second major pattern is Data Exposure / Leakage, which indicates that many Public Administration incidents involve compromised citizen data, employee information, public records, emails, databases, or sensitive government files. This is important because public-sector data often contains personally identifiable information, legal records, administrative documents, and operational records. Identifying this hidden pattern helps security teams understand where privacy risk, compliance risk, and public trust risk may exist, even when the dataset does not explicitly classify the event as a data breach. The results also show a strong presence of Sensitive Governance / Espionage patterns. These cases may involve ministries, senior officials, public agencies, government communication channels, or politically sensitive functions. From a cybersecurity viewpoint, this pattern is important because the technical severity of an incident may appear limited, while its governance impact may be significant. The Public Safety / Law Enforcement and Operational Disruption patterns also provide important security value. These patterns show that cyber events may affect police departments, sheriff offices, fire departments, emergency communications, municipal systems, or government service continuity. This matters because disruptions in these functions can create real-world consequences beyond IT system failure. For example, an incident with unknown motive may still require high priority if the narrative indicates disruption to emergency communication, law enforcement reporting, or municipal service

delivery. Finally, the Citizen-Service Impact and Ransomware / Extortion patterns show how Public Administration cyber incidents can directly affect citizens' access to government services, payment systems, public portals, records, benefits, or municipal operations. This is useful for public-sector cybersecurity because it connects cyber incidents to service continuity and citizen-facing risk. Overall, this paper presents an AI-based framework that moves raw cyber records to generate more advanced analytical cyber information for CTI Repository.

6. Conclusions

We propose an LLM-assisted semantic analysis framework for discovering cybersecurity patterns in Public Administration cyber event records. The framework moves from raw cyber records to more advanced analytical security information using narrative incident descriptions. To support this, we introduced

three cyber metrics: Incident Security Semantics, Governance Function Criticality, and Cyber Intelligence Gap and Inference, which analyze each event from security, public-sector, and intelligence-gap perspectives. These metrics are used to extract hidden cybersecurity pattern. The results showed that our framework can generate meaningful cyber threat intelligence and discover a couple of cyber patterns.

Acknowledgements: The authors would like to express their sincere gratitude to the Center for Cyber Security and Forensics Education (C²SAFE) at Illinois Institute of Technology for its support, resources, and access to core research facilities, including the Ed Kaplan Family Institute for Innovation and Tech Entrepreneurship. The authors also gratefully acknowledge the financial support provided by the Motorola Solutions Foundation, which helped facilitate the development and completion of this work.

References

- [1] Rahman MR, Mahdavi-Hezaveh R, Williams L. A literature review on mining cyberthreat intelligence from unstructured texts. In: 2020 International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE; 2020. p. 516-525.
- [2] Jo H, Lee Y, Shin S. Vulcan: Automatic extraction and analysis of cyber threat intelligence from unstructured text. *Comput Secur.* 2022;120:102763.
- [3] Irshad E, Siddiqui AB. Cyber threat attribution using unstructured reports in cyber threat intelligence. *Egypt Inform J.* 2023;24(1):43-59.
- [4] Marinho R, Holanda R. Automated emerging cyber threat identification and profiling based on natural language processing. *IEEE Access.* 2023;11:58915-58936.
- [5] Karras A, Theodorakopoulos L, Karras C, Theodoropoulou A, Kalliampakou I, Kalogeratos G. LLMs for cybersecurity in the big data era: a comprehensive review of applications, challenges, and future directions. *Information.* 2025;16(11):957.
- [6] Albarrak M, Salonitis K, Jagtap S. Natural language processing (NLP)-based frameworks for cyber threat intelligence and early prediction of cyberattacks in Industry 4.0: a systematic literature review. *Appl Sci.* 2026;16(2):619.
- [7] Liu J, Zhan J. Constructing knowledge graph from cyber threat intelligence using large language model. In: 2023 IEEE International Conference on Big Data (BigData). Piscataway: IEEE; 2023. p. 516-521.
- [8] Bratsas C, Anastasiadis EK, Angelidis AK, Ioannidis L, Kotsakis R, Ougiaroglou S. Knowledge graphs and semantic web tools in cyber threat intelligence: a systematic literature review. *J Cybersec Priv.* 2024;4(3):518-545.

- [9] Alturkistani H, Chuprat S. Artificial intelligence and large language models in advancing cyber threat intelligence: a systematic literature review [Internet]. Research Square; 2024 [cited 2026 Apr 30].
- [10] Huang L, Zhang M, Chen Z, Ma S, Liu Z, Ye Y, Xiao X. LLM4CTI: Uncovering security entities and their interactions from unstructured cyber threat intelligence. In: International Conference on Learning Representations (ICLR 2026). 2026.
- [11] Harry, C., & Gallagher, N. (2018). Classifying cyber events. *Journal of Information Warfare*, 17(3), 17-31.
- [12] Schlette, D., Caselli, M., & Pernul, G. (2021). A comparative study on cyber threat intelligence: The security incident response perspective. *IEEE Communications Surveys & Tutorials*, 23(4), 2525-2556.
- [13] Sun, N., Ding, M., Jiang, J., Xu, W., Mo, X., Tai, Y., & Zhang, J. (2023). Cyber threat intelligence mining for proactive cybersecurity defense: A survey and new perspectives. *IEEE Communications Surveys & Tutorials*, 25(3), 1748-1774.
- [14] Sarhan, I., & Spruit, M. (2021). Open-cykg: An open cyber threat intelligence knowledge graph. *Knowledge-based systems*, 233, 107524.
- [15] Hu Y, Zou F, Han J, Sun X, Wang Y. LLM-TIKG: Threat intelligence knowledge graph construction utilizing large language model. *Computers & Security*. 2024.
- [16] Arikkat, D. R., Vinod, P., KA, R. R., Nicolazzo, S., Nocera, A., & Conti, M. (2024, June). Relation extraction techniques in cyber threat intelligence. In *International Conference on Applications of Natural Language to Information Systems* (pp. 348-363). Cham: Springer Nature Switzerland.
- [17] Lanz, Z. (2022). Cybersecurity risk in US critical infrastructure: An analysis of publicly available US government alerts and advisories. *International Journal of Cybersecurity Intelligence & Cybercrime*, 5(1), 43-70.
- [18] [20] Lisiak-Felicka D. A comparative analysis of information security incidents in public administration in selected European Union countries. 2023.
- [19] Ghazi, Y., Anwar, Z., Mumtaz, R., Saleem, S., & Tahir, A. (2018, December). A supervised machine learning based approach for automatically extracting high-level threat intelligence from unstructured sources. In *2018 International Conference on Frontiers of Information Technology (FIT)* (pp. 129-134). IEEE.
- [20] Orbinato, V., Barbaraci, M., Natella, R., & Cotroneo, D. (2022, October). Automatic mapping of unstructured cyber threat intelligence: An experimental study:(practical experience report). In *2022 IEEE 33rd International symposium on software reliability engineering (ISSRE)* (pp. 181-192). IEEE.