A SURVEY OF AI-POWERED CYBER THREAT DETECTION AND PROFILING USING NATURAL LANGUAGE PROCESSING TECHNIQUES

[#1]Aluvala Anusha, *M.Tech Student,*
[#2]Dr. B. Sateesh Kumar, *Professor & Head of Department,*
Department of Computer Science and Engineering,
JNTUH College of Engineering, Jagtial, Telangana, India.

ABSTRACT: The rapid increase in cyberthreats has brought attention to the shortcomings of identification methods based on rules and signatures. Sophisticated, context-aware detection algorithms are required due to the increasing prevalence of hacks that use social media, complex communication channels, and obfuscation tactics. Natural language processing (NLP) and artificial intelligence (AI) will be discussed in this lecture along with their historical uses in cyber threat detection. We also study hybrid systems, such as deep neural architectures (LSTM, CNN), transformer-based models (BERT, RoBERTa, GPT), and ontologies and knowledge graphs. Using examples from social media conversations, phishing emails, and unstructured cyber threat intelligence (CTI) reports, the article shows how natural language processing (NLP) can be used to identify TTPs. It also tackles problems with clarity, language support, real-time operation, and adequate data. Lastly, it talks about recent advancements that point to a move toward automated, flexible, and user-friendly security systems. Large language models (LLMs), explainable artificial intelligence, and bidirectional learning are a few examples of this.

Keywords: Artificial Intelligence; Natural Language Processing (NLP); Cyber Threat Intelligence (CTI); Transformer Models; Deep Learning; Threat Profiling; Explainable AI; Phishing Detection; Named Entity Recognition; Cybersecurity Automation.

1. INTRODUCTION

The frequency and intensity of cyberattacks against people, businesses, and vital infrastructure have increased within the past ten years. The methods used to identify these attacks become less effective as their language becomes more complicated. Phishing emails, underground chats, and malware how-to tutorials are a few examples. When confronted with new or context-sensitive attacks that use common language to trick and influence people, traditional algorithms that depend on matching signatures or rules become ineffective.

These factors have contributed to the rise in popularity of AI and NLP applications in the defense industry. Natural language processing (NLP) makes it possible for computers to understand text messages, which makes it easier to find, assess, and arrange information about online threats. However, computers may eventually be able to learn from records and identify patterns of inappropriate behavior with the aid of artificial intelligence. They provide an advanced technique for identifying cyberthreats and creating profiles by gleaning valuable information from vast amounts of unstructured textual data, including event reports, vulnerability statements, and online debates.

From simple machine learning algorithms to more complex ones like deep learning and transformer-based systems, research has advanced significantly. Early research have found a correlation between language used on social media and the possibility of exploitative behavior (Sabottke et al., 2015). Tools such as TTPDrill (Husari et al., 2017) and Exaction (Zhang et al., 2021) have made it easier to automate strategies and threat actions utilizing CTI data. With

contextual embeddings, which combined the BERT and RoBERTa models to uncover semantic linkages, cybersecurity research made a huge leap forward. These claims are made by Evangelatos et al. (2021) as well as Rahali and Akhloufi (2021).

New projects that extensively rely on natural language analysis, such LogShield, MalBERT, and TIEF, are demonstrating the future of threat intelligence systems (Afnan et al., 2023; Joy et al., 2025). They enhance the correctness, scalability, and understandability of the frameworks through the use of transformer structures, knowledge graphs, and ontology-driven reasoning. Automated cyber analytics are now more widely available and of greater quality due to changes in transformer design (Avci et al., 2024) and explainable artificial intelligence (Bhardwaj, 2023).

Some problems still exist in spite of these developments. Insufficient data worsens the performance of supervised learning, especially in named cybersecurity datasets. People who don't speak the same language have a harder time communicating on international cybercrime platforms that use multilingual threat communication. We must figure out how to make huge language models more comprehensible and accessible in order to comply with the regulations and win over analysts. One of the most practical challenges is being able to change in real-time without being susceptible to manipulation by an adversary.

This study gathers academic papers over the last ten years that address the use of artificial intelligence to interpret natural language in cybersecurity in order to classify and rate cyberthreats. The main goal of this essay is to give a thorough review of the development of textual intelligence analysis and how it is used in contemporary cybersecurity systems. It highlights open problems and suggests directions for further study into improving cybersecurity ecosystems using federated, explicable, and flexible AI-NLP frameworks.

2. LITERATURE SURVEY

C. Sabottke, O. Suciu, and T. Dumitras (2015): This study investigates the feasibility of using social media data, particularly Twitter data, to detect potentially useful software vulnerabilities in their early stages. The authors use correlations between security-related tweets and public vulnerability reports to predict when an attack will occur. Statistical modeling and machine learning are employed in the study to demonstrate that online discussions can be utilized to reliably forecast potential dangers. Their findings highlight the growing significance of social media intelligence in proactive cybersecurity defense.

G. Husari, E. Al-Shaer, M. Ahmed, B. Chu, and X. Niu (2017): A method for automatically extracting TTPs from unstructured cyber threat intelligence (CTI) material is outlined by the authors. The method extracts threat information from retrieved textual patterns using natural language processing. The study automates CTI processing by using machine learning for syntactic parsing and entity recognition. Gathering information is made more valuable by the test findings showing that identifying threats and behaviors is quite accurate.

M. Ebrahimi et al. (2019): In order to extract cyber danger indicators from unstructured text, the research developed a deep learning method. Malware, vulnerabilities, and exploits are made more cognizant of their environment by the creators using BiLSTM and attention approaches. The approach outperforms more conventional machine learning baselines and exhibits strong cross-dataset generalizability. One of the earliest deep NLP frameworks for autonomously collecting threat information is demonstrated in this paper.

M. Zhang, H. Shen et al. (2021): Exaction, described in this research, is a multimodal deep learning system capable of autonomously extracting malicious behaviors from CTI data. This approach becomes even more adept at identifying attack sequences when contextual information is added to textual properties.

Compared to other natural language processing (NLP) extraction systems, Exaction performs better in experiments. More comprehensive danger information may be provided by merging diverse forms of data, according to the study.

K. Satvat, R. Gjomemo, and V. Venkatakrishnan (2021): Extractor, described in this research, is a system that can transform disordered threat data into attack graphs based on behavioral patterns. Attackers' actions and goals are described using dependency parsing and rule-based language templates. Its great recall and precision are demonstrated by its testing on all available CTI datasets. The approach re-creates the behavior of threats by merging cyber ontology models with NLP.

P. Evangelatos et al. (2021): Cyber threat intelligence's Named Entity Recognition (NER) is investigated using deep transformer-based models. Using a domain-specific dataset (DNRTI), they evaluate BERT, RoBERTa, and XLNet. Contextual embeddings significantly improve threat entity detection, according to the data. The findings highlight the significance of domain-specific cybersecurity NLP job tuning.

C. Gao et al. (2021): A knowledge-and data-based Named Entity Recognition tool tailored to cybersecurity is proposed in this article. The approach use domain ontologies in conjunction with BERT-based embeddings to extract named entities and links from CTI documents. A combination of structured threat knowledge graphs and unstructured data sources is effectively achieved by the hybrid approach. Finding domain-specific entities has advanced greatly, leading to better automated cyber situational awareness, according to the results.

A. Rahali and M. A. Akhloufi (2021): The specialists present MalBERT, a transformer-based deep learning model that can detect malicious software. In order to distinguish between malicious and safe samples, MalBERT uses textual malware descriptions to fine-tune BERT. The approach achieves respectable accuracy and recall on benchmark datasets. The research proves that transformer topologies can successfully mimic contextual malware's characteristics.

M. Liberato (2022): This technical study introduces SecBERT, an alternative to the BERT paradigm for analyzing cybersecurity event logs. Classifying objects and extracting events from disparate text sources seem to be the author's primary interests. Using security corpora for fine-tuning, SecBERT makes cyber event detection and correlation much more accurate. The results of the study demonstrate that natural language processing (NLP) is useful in military settings when domain-specific pretraining is used.

D. Joshi et al. (2022): The study proposes a transformer-based method for describing threat actors and leveraging cyber threat text data collected from internet forums. The system use clustering algorithms and contextual embeddings to categorize talks according to the participants' actions and intentions. The outcomes demonstrate that profiling outperforms conventional topic models in terms of accuracy. Applications of natural language processing in behavioral cyber intelligence are demonstrated in this project.

C. Catal et al. (2022): This comprehensive research examines the application of deep learning to detect frauds in textual and web-based datasets. Important preprocessing techniques and feature representations are emphasized when comparing CNN, LSTM, and hybrid architectures. The outcomes demonstrate that deep models outperform more conventional classifiers. But they struggle to adapt to new domains and explain their outcomes. Using natural language processing (NLP) driven contextual analysis, the study concludes with recommendations for phishing security systems.

Z. Alshingiti et al. (2023): The research recommends training a spam detection model using data collected from the internet as well as email using a combination of LSTM and CNN. The model effectively encapsulates the structure and content of phishing assaults by combining the capture of sequential and geographical data. The experiment's findings demonstrate that the model

outperforms the CNN and LSTM models separately in terms of accuracy. This research establishes the practicality of deep hybrid systems for threat detection via NLP.

S. Afnan, M. Sadia, S. Iqbal, and A. Iqbal (2023): For the purpose of locating APTs in system logs, the authors discuss LogShield, a transformer-based framework. The model identifies patterns that differ from others through self-attention processes; these differences may indicate the presence of covert attacks. The identification rates and false positive rates of this approach are significantly greater than those of LSTM baselines, according to numerous experiments. According to the research, transformers are an effective architecture for log-based cyber threat analytics.

Y. Wang et al. (2023): In order to discover novel cyberthreats on Twitter, this article examines topic models using NLP. Using clustering and Latent Dirichlet Allocation (LDA), the authors examine cybersecurity tweets in real-time. The system is effective in spotting patterns, such as the emergence of new viruses or data breaches. The research demonstrates the feasibility of using social media collecting to detect potential dangers in advance.

S. Bhardwaj (2023): This research proposes a method for Explainable AI (XAI) that describes cyberthreats using understandable machine learning models and natural language processing (NLP). The author demonstrated the relationship between textual hints and threat categories through attention-based visualization. Experiments demonstrate that analysts are more trustworthy and forthcoming when using automatic profiling tools. The study recommends including explainability into AI-powered defense systems.

H. Wang (2019–2024): This paper provides a comprehensive overview of Named Entity Recognition (NER) methodologies, which are utilized in cyber threat intelligence. Extracting entities from disorganized text is examined using deep learning, machine learning, and transformer-based approaches. Learning how environmental embeddings and transfer learning models are replacing rule-based tagging is an intriguing development.

M. Alshomrani et al. (2024): The authors provide a comprehensive analysis of malware detection methods based on transformers. Applications for cybersecurity, such as GPT, BERT, and RoBERTa, are examined. Based on evaluation criteria, techniques for improvement, and dataset types, the study categorizes current methodologies into groups. Results show that transformers outperform previous neural models, despite ongoing issues with data tagging and processing costs.

C. Avci et al. (2024): Various approaches to developing transformer models for usage in cybersecurity contexts are investigated in this study. In order to improve the performance of log data and CTI text, the authors investigate the advantages of attention levels, positional encoding, and tokenization. When it comes to classification, testing has revealed that domain-specific adaptation works wonders. The research provides helpful guidance on modifying deep architectures to integrate with AI security applications.

A. Joy et al. (2025): In order to discover TTP, the Threat Intelligence Extraction Framework (TIEF) consults knowledge graphs and vocabularies in conjunction with NLP, as discussed in this article. The program use relation mapping, object identification, and semantic reasoning to construct AI. The results demonstrate that risk data is both more scalable and easier to comprehend after ontology integration. The approach suggests a future where CTI enrichment is automated with the use of integrated AI and NLP systems.

A. Almadhor et al. (2025): The effectiveness of large transformer models in detecting anomalies in cyber and IoT networks is examined in this study. The work enhances models for detecting anomalous communication patterns by using big language models that have already undergone training. In terms of detecting meaningful

changes, the results demonstrate that transformers outperform conventional anomaly detectors. Full cybersecurity monitoring should include linguistic models, according to the authors.

K. Barik et al. (2025): Using a combination of metaheuristic feature selection and improved convolutional layers, the authors construct EGSO-CNN, a deep learning model for phishing URL detection. In a variety of phishing datasets, their mixed-method approach improves generalization. According to the study, it outperforms typical models in terms of accuracy while producing fewer false positives.

L. Chen, H. Deng, J. Zhang, B. Zheng, and R. Jiang (2025): Using semantic similarity mapping, this study demonstrates a segment-level NER model that can obtain cyber threat information. This concept establishes a platform for the cross-term linking of threat elements. The findings demonstrate an improvement in both the accuracy of extraction and the consistency of the context. The research reveals a novel approach to segment-level entity recognition in defense using natural language processing.

3. COMPARATIVE ANALYSIS OF EXISTING AI–NLP MODELS FOR CYBER THREAT DETECTION AND PROFILING

3.1 Table 1 Early Machine Learning and NLP Approaches

S. No.	Author & Year	Approach / Model	Data Source / Domain	Core NLP / AI Techniques	Application / Use Case	Key Findings / Contributions
1	C. Sabottke et al., 2015	Predictive modeling from social media	Twitter vulnerability discussions	Machine learning, Text mining	Predicting exploit trends	shown how Twitter data could serve as a vulnerability early warning system.
2	G. Husari et al., 2017	TTPDrill framework	CTI reports	NLP parsing, Entity extraction	Automatic extraction of TTPs	automated the process of identifying tactics, methods, and procedures (TTPs) and analyzing threat information.
3	M. Ebrahimi et al., 2019	Deep Learning for CTI extraction	Threat reports & dark web	BiLSTM + Attention	Threat indicator extraction	one of the first deep natural language processing algorithms to extract threat indicators.
4	H. Wang, 2019	Comprehensive survey on NER	Research datasets	ML, DL, Transformer comparison	Named Entity Recognition	investigated the development of contextual NLP models for CTI as opposed to rule-based models.

3.2 Table 2 Deep Learning and Transformer Integration

S. No.	Author & Year	Approach / Model	Data Source / Domain	Core NLP / AI Techniques	Application / Use Case	Key Findings / Contributions
5	M. Zhang et al., 2021	Exaction Multimodal Model	CTI reports + metadata	Multimodal deep learning	Threat action extraction	enhanced accuracy via the utilization of contextual metadata and text.
6	K. Satvat et al., 2021	Extractor Framework	Threat reports	Dependency parsing, linguistic rules	Attack behavior reconstruction	Developed organized behavioral graphs with textual threat data.
7	C. Gao et al., 2021	Knowledge-driven NER	CTI & ontology data	BERT + Knowledge Graph	Threat entity & relation extraction	For situational awareness, domain ontology and NLP were merged.
8	A. Rahali & M. A. Akhloufi, 2021	MalBERT	Malware text descriptions	Transformer (BERT)	Malware detection	Proven transformers exhibit a high accuracy rate in detecting fake text.
9	P. Evangelatos et al., 2021	Transformer-based NER	DNRTI dataset	BERT, RoBERTa, XLNet	Entity extraction	Significant enhancements in performance by transformer fine-tuning.
10	M. Liberato, 2022	SecBERT	Incident reports	BERT adaptation	Event & entity extraction	Domain-specific pretraining in natural language processing has demonstrated advantages.
11	D. Joshi et al., 2022	Transformer-based actor profiling	Hacker forums	Contextual embeddings, clustering	Threat actor profiling	Augmented behavioral categorization and performer identification.
12	C. Catal et al., 2022	Systematic review of DL	Phishing datasets	CNN, LSTM, Hybrid	Phishing detection	Deficiencies in performance, explainability, and adaptability were identified.

3.3 Table 3 Hybrid and Explainable AI Models

S. No.	Author & Year	Approach / Model	Data Source / Domain	Core NLP / AI Techniques	Application / Use Case	Key Findings / Contributions
13	Z. Alshingiti et al., 2023	Hybrid DL model	Email & website datasets	LSTM + CNN	Phishing classification	Higher accuracy was attained by the use of hybrid sequential and spatial characteristics.
14	S. Afnan et al., 2023	LogShield	System logs	Transformer, Self-Attention	APT detection	Detected stealthy APTs better than RNNs.
15	Y. Wang et al., 2023	Topic Modeling	Twitter & social media	LDA, Clustering	Threat trend detection	found online conversations in real time for preventative protection.
16	S. Bhardwaj, 2023	Explainable AI framework	CTI text	XAI, Attention visualization	Threat profiling	increased interpretability and confidence among analysts in AI systems.
17	M. Alshomrani et al., 2024	Survey of transformers	Malware & phishing datasets	BERT, GPT, RoBERTa	Malware detection	highlighted the accuracy and versatility of transformer dominance.
18	C. Avci et al., 2024	Transformer optimization	Log & CTI data	Model tuning, attention design	Cyber text classification	offered recommendations for modifying transformer architectures.
19	A. Joy et al., 2025	TIEF Framework	CTI text + knowledge graphs	NLP + Ontology + Reasoning	TTP extraction	For structured intelligence, NLP and semantic reasoning were combined.
20	A. Almadhor et al., 2025	LLM-based anomaly detection	IoT & cyber systems	Pre-trained LLMs	Anomaly recognition	Cyber data contextual irregularities were successfully detected by LLMs.
21	K. Barik et al., 2025	EGSO-CNN model	Phishing URL datasets	CNN + metaheuristics	Phishing detection	CNN was tuned through evolution to improve accuracy.
22	L. Chen et al., 2025	Segment-level NER	CTI datasets	Semantic similarity mapping	Entity extraction	For CTI situations, cross-sentence semantic NER was introduced.

4. CONCLUSION

According to this study, natural language processing and artificial intelligence have a significant influence on cyber threat identification and profiling. Over the past ten years, rule-based and classical machine learning techniques have evolved into deep learning and transformer-based architectures, which have improved the ability to understand textual threat information. These techniques have significantly improved threat identification from unstructured sources in terms of automation, accuracy, and context. Explainable AI, massive language models, and hybrid AI-NLP frameworks have made cybersecurity solutions more flexible and transparent. Data accessibility, model interpretability, and multilingual analysis continue to present formidable obstacles. In conclusion, proactive cybersecurity is crucial in today's ever-changing digital environment, and AI-powered natural language processing (NLP) technologies are the key. They are helpful in creating scalable and understandable security frameworks.

REFERENCES

1. C. Sabottke, O. Suciu, and T. Dumitras, “Vulnerability disclosure in the age of social media: exploiting Twitter for predicting real-world exploits,” in Proc. 24th USENIX Security Symp., 2015, pp. 1041–1056.
2. G. Husari, E. Al-Shaer, M. Ahmed, B. Chu and X. Niu, “TTPDrill: Automatic and accurate extraction of threat actions from unstructured text of CTI sources,” in Proc. 33rd Annu. Computer Security Applications Conf. (ACSAC), 2017, pp. 103–115.
3. H. Wang, “Threat intelligence named entity recognition techniques: a survey,” (survey article), 2019–2024 (see review articles and datasets).
4. M. Ebrahimi, et al., “Deep learning for threat intelligence extraction from unstructured text,” Computers & Security, vol. 87, 2019.
5. Mohammad Sirajuddin, Dr. B. Sateesh Kumar, Efficient and Secured Route Management Scheme Against Security Attacks in Wireless Sensor Networks, International Conference on Electronics and Sustainable Communication Sys, ISBN No.978-1-6654-2866-8, pp.1052-1058, IEEE, Sept, 2021
6. C. Gao, et al., “Data and knowledge-driven named entity recognition for cyber security,” Cybersecurity, SpringerOpen, 2021.
7. A. Rahali and M. A. Akhloufi, “MalBERT: Using Transformers for Cybersecurity and Malicious Software Detection,” arXiv, Mar. 2021.
8. M. Liberato, “SecBERT: Analyzing reports with BERT-like models” (master’s thesis / technical report), 2022.
9. C. Catal, et al., “Applications of deep learning for phishing detection: a systematic review,” International Journal / PubMed Central (review), 2022.
10. Z. Alshingiti, et al., “A deep learning-based phishing detection system using LSTM, CNN and hybrid LSTM–CNN models,” Electronics, MDPI, vol. 12, no. 1, 2023.
11. S. Afnan, M. Sadia, S. Iqbal and A. Iqbal, “LogShield: A Transformer-based APT detection system leveraging self-attention,” arXiv, Nov. 2023.
12. Y. Wang, et al., “Exploring topic models to discern cyber threats on Twitter,” Comput. & Security / Journal, 2023.
13. S. Bhardwaj, “Explainable AI for cyber threat profiling,” Expert Systems with Applications, 2023 (explainability in CTI contexts).
14. M. Zhang, H. Shen, et al., “Exaction: Automatically extracting threat actions from cyber threat intelligence reports based on multimodal learning,” Security and Communication Networks, 2021.
15. K. Satvat, R. Gjomemo, and V. Venkatakrishnan, “Extractor: Extracting attack behavior from threat reports,” in Proc. IEEE European Symposium on Security and Privacy (EuroS&P), 2021, pp. 598–615.
16. M. Alshomrani, et al., “Survey of transformer-based malicious software detection,” Electronics, MDPI, 2024.
17. C. Avci, et al., “Design tactics for tailoring transformer architectures to cybersecurity

applications," Cluster Computing / Springer, 2024.
18. D. Joshi, et al., "Transformer-based threat actor profiling," ACM Digital Threats: Research and Practice, 2022.
19. A.Joy, et al., "Threat Intelligence Extraction Framework (TIEF) for TTP extraction," MDPI / 2025 (framework combining NLP + ontology/knowledge graph).
20. P. Evangelatos, et al., "Named Entity Recognition in Cyber Threat Intelligence," IEEE/ACTI Workshop / technical report, 2021 (DNRTI experiments and transformer evaluations).
21. A. Almadhor, et al., "Evaluating large transformer models for anomaly detection in IoT / cyber contexts," Scientific Reports / Nature, 2025.
22. K. Barik, et al., "Web-based phishing URL detection model using deep learning (EGSO-CNN)," Jnl, 2025.
23. L. Chen, H. Deng, J. Zhang, B. Zheng, and R. Jiang, "Threat Intelligence Named Entity Recognition based on segment-level information extraction and similar semantic space construction," Symmetry (MDPI), 2025.