Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 46 (2025), No. 7, pp. 2329-2338
DOI : 10.47974/JIOS-2129

Integrating natural language processing techniques into cybersecurity workflows : Enhancing cyber protection against emerging threats

Vandana Kalra [†]
Supreet Kaur Sahi *
Sri Guru Gobind Singh College of Commerce
University of Delhi
Delhi 110034
India

Amanpreet Kaur [§]
Bow Valley College
Calgary
Alberta, T2G4V1
Canada

Abstract

Cyberattacks are a constantly evolving danger, requiring cybersecurity to become even more critical. The ever-changing dynamics of attacking threats have finally started to change the terrain of the cyberspace, thus, requiring strong cyber security. The domain of NLP and especially its use in cyber defense has potential to make a significant contribution in addressing these concerns. This research attempts to find any scholarly works that focus on the use of NLP in cyber threat intelligence so that corresponding strategies can be devised to counter novel threats with sophisticated systems that can address the constantly changing dynamics of complex challenges. The paper also tackles the problems of classifying and structuring of many unorganized data sets numbering in terabytes, concentrating the focus on the use of NLP for threat detection, incident response and overall risk assessment, evaluation and compliance control. NLP based cyber systems defence on secondary multi-tiered application architecture NLP frameworks provides stronger positions in countering cyber warfare. This research aims to demonstrate that through NLP, an organization and its

[†] *E-mail:* **Vandana.kalra@sggscc.du.ac.in**
* *E-mail:* **supreetkaursahi@sggscc.du.ac.in (Corresponding Author)**
[§] *E-mail:* **akaur@bowvalleycollege.ca**

critical systems can be digitally fortified, thereby, proving once again the growing necessity of cyber linguistic defense from NLP based frameworks.

Subject Classification: *Primary 93A30, Secondary 49K15.*

Keywords: *Cybersecurity, Natural language processing, Threat intelligence, Incident reports, Sentiment analysis, Named entity recognition.*

1. Introduction

Unfortunately, digital advancements have impacted increasing opportunities for cybercriminals. Cyber terrorists have advanced and utilize military-grade tools, countermeasures and methods for defeating high level defensive systems on networks, applications and other devices. A major threat in digital security is the vicious mass data attack on the organizational cyber security systems. Old tactics, including logs, text communications and other user-generated data, are not able to withstand the attack and result in the organization being unprotected. One alternative is the recognition and understanding sophisticated language and contextual tools to help overcome the challenges in unstructured data and comprehensive text analysis.

Context and language are vital tools in cyber security in toward accomplishing, resolving and thinking counter threats. The NLP system augments the exposure and analytic grammar and contextual cyber security strategy toward a culminating formulated defense system. The contextualized cyber-attack system strategy is to ease and promote organizational understanding toward defensive cyber-attacks. NLP system in cyber security is applied to ease organizational interpretive challenges on understanding and defense mechanisms toward acquisition threat systems. The implementation of NLP to the cyber security focus provides defense against active threats emerging due to verbal communications and simultaneously raises new important issues pertaining to the sensitivity of data, linguistic bias. As the exploration of integrating NLP within the domain of cybersecurity unfolds, this paper is structured around the following pivotal research questions aimed at discovering how NLP can strengthen cybersecurity measures.

- What gaps and challenges are identified in the current cybersecurity literature?
- How is the traditional cybersecurity workflow structured and how can integrating NLP into workflow improve threat detection and response?

- How can a layered defence strategy be implemented using NLP techniques at various stages of the cybersecurity workflow?
- What essential components and considerations should be included in a framework for developing NLP-driven cybersecurity applications?

2. Related Work

In the rapidly evolving landscape of digital security, the integration of NLP into cybersecurity measures has emerged as a groundbreaking approach to fortify digital defences. This study seeks to investigate the use of 'linguistic analysis' and 'computational linguistics' techniques that function as a potent defense against cyber threats in detail.

The use of NLP in cybersecurity is a significant breakthrough in the identification and mitigation of digital dangers. In this field, research seems to be gradually changing from the detection of phishing attacks focusing on various algorithms [1][2] to the adoption of more sophisticated deep learning techniques for classification [3]. Specifically, incorporating Text and Voice enabled communication and Genetic Decision Trees with NLP [4] has shown new capabilities for process automation. For example, automating the process of extracting and analyzing cyber threat intelligence significantly expands the NLP's capabilities to provide speedy reactions to incendiary security threats [5] and the Vulcan system [6].

Further, NLP enhances compliance with regulatory requirements such as automating the translation of complex legal documentation [7] and automating the analysis of large volumes of research [8], underscoring the critical need to alleviate the compliance burden of manual administrative tasks. Research now involves developing strategies to determine software vulnerabilities using the Mixeway system [9] alongside enhancing systems for intrusion monitoring [10][11]. The evolution of cyber threats increasingly incorporates elements of social engineering which underscores the need for NLP to develop counter measures [12]. Online discourse sentiment analysis provides essential intelligence to monitor and prioritize reactive measures to mitigate cyber security threats [13][14]. While integrating NLP to enhance measures in cybersecurity, certain intricate challenges have to be navigated. These challenges of integrating NLP in cybersecurity have been summarized in Table 1 with their descriptions according to existing literature.

3. Cybersecurity workflow

There are five stages in the cybersecurity workflow [15], which include identification, protection, detection, response, and recovery, all supporting improvement of the organization's overall security posture. An enterprise begins the process by identifying assets, assessing risks, and creating security policies. In the protection stage, security measures and training of employees are performed as shown in Figure 1. Continuous monitoring and audits reveal incidents, allowing immediate response actions [16][17]. The focus during the recovery phase pivots towards the restoration of systems and data alongside increased security measures [18]. Post-incident analysis aids in fostering trust within the organization and its stakeholders.

Table 1

Challenges in integrating cybersecurity with NLP

Challenge	Description
Data Quality and Volume	Essential high-quality, labelled data is scarce, complicating NLP model training with data volume.
Contextual Understanding Challenges Multilingual and Code-Switching Issues	Struggle with technical jargon and domain-specific knowledge impacts NLP accuracy. Code-switching in multiple languages add complexity.
Scalability and Performance Hurdles	Real-time processing of large datasets is resource-intensive, challenging scalability.
Accuracy in Threat Detection	Balancing high accuracy in detecting threats while minimizing false positives and negatives is difficult.
Privacy and Ethical Concerns &Integration with Existing Systems	Analyzing communications raises issues, requiring compliance with laws and ethical guidelines. Integrating NLP tools into cybersecurity frameworks.
Evaluation Complexities	The dynamic nature of cyber threats and language makes evaluating NLP's effectiveness challenging.

NLP techniques are reshaping the information security field by automating the processes related to identifying, protecting, detecting, responding, and recovering from incidents [19]. NLP surveils and analyzes structured and unstructured data for risks [20].

4. NLP-Enabled Multi-Layered Defence System

The multi-layered defence strategy, also known as NETWAR, employs the use of several defence mechanisms including preventive measures to the cyber world. With the use of NLP at every layered, these mitigation measures can be more effective for the organizations. Each action in Figure 2 is marked with a description of NLP application in the multi-layered defence strategy. When adopting NLP in every single part of their infrastructure, companies obtain richer data on threat environments and improve their agility to the real-time cybersecurity landscape [18].

5. Framework for Designing NLP-based Cybersecurity Solutions

From step to step, the systematic designing NLP oriented cybersecurity tools for each element of cybersecurity system is followed as described below.

- Cybersecurity Objectives: What do you aim to achieve with the application?
- Data Acquisition and Preparation: What data do you need? Get, filter for relevant data, and train with other unneeded data.
- Approach: Determine applicable NLP methods based on the set cybersecurity goals and create new or revise current methods that meet the needs of the application.
- Model Training and Testing: The models are prepared using the provided data and are then evaluated with a testing set to gauge performance, after which necessary changes are applied.
- Integration and Deployment: Merge and distribute the NLP models into the cybersecurity framework, with provisions of real time monitoring and control of the systems, as well with the periodic refreshing of the models based on new data and subsequent feedback.

The application of cutting-edge technologies in Natural Language Processing in the domain of Cybersecurity is the result of synergy between the domain specialists in cyber security, NLP, and software engineering [10]. Systems without implaced frameworks can face attacks from new vectors and continuously strive to outsmart security systems. With the use of the framework, the potential damage can be mitigated and the system robustness enhanced.

6. Emerging Cybersecurity Threats

Today, the evolvement of cyber security threats poses considerable risks and challenges to an individual, an organization and even the state and is achieving which is quite difficult to do [14]. These challenges mostly emerge from the new and refined state of the art tools and technologies that are continuously growing along with the attack surfaces [21] or new images.

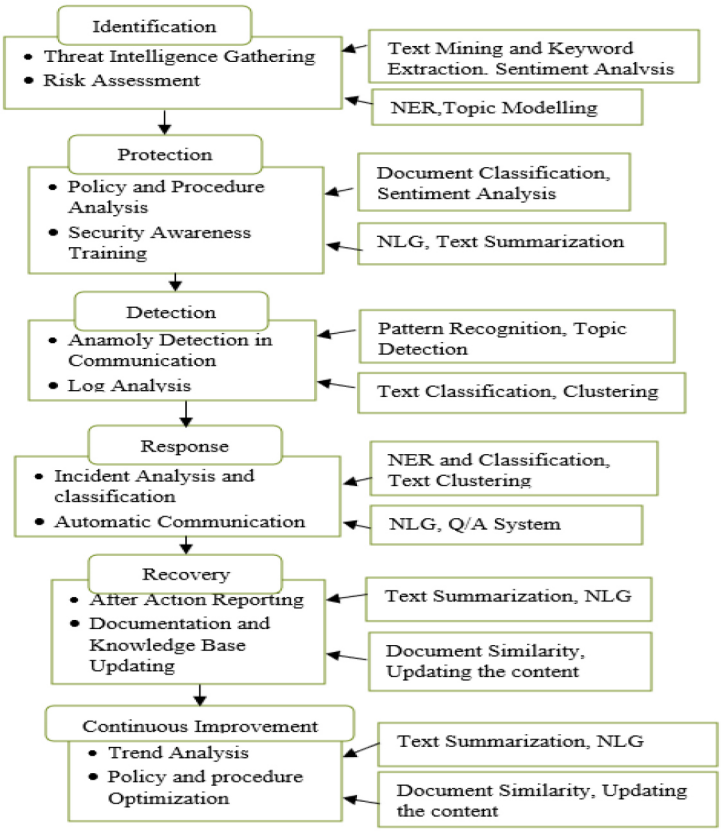

Figure 1

NLP techniques in workflow for cybersecurity

Use of sophisticated actionable cyber threat intelligence in conjunction with real-time monitoring of cyber incidents, followed by swift incident containment and response strategies, greatly mitigates the impact of a cyber-attack. Address the bladed challenges of a multifaceted IT hybrid

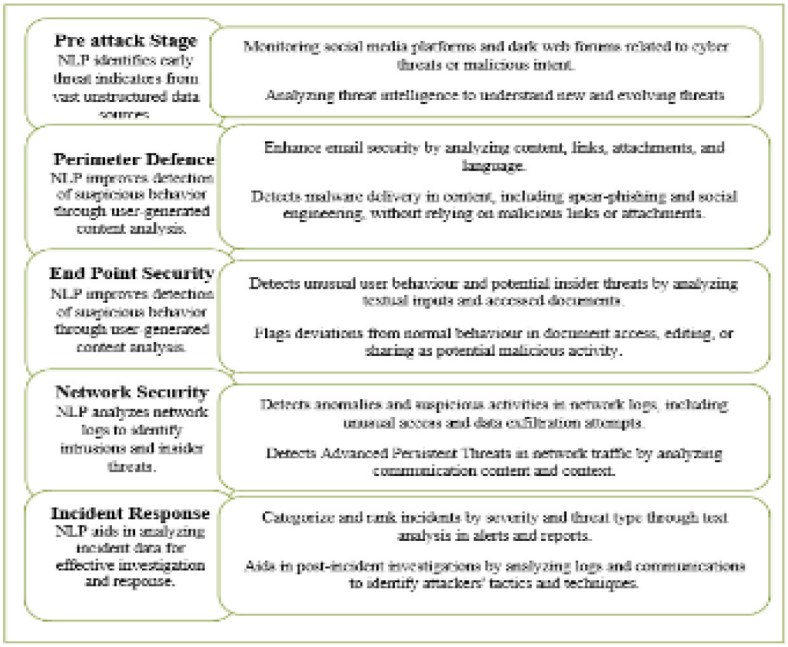

Figure 2

NLP techniques at diverse levels of cyberattacks in the Layer Defence Approach

cloud and network environment through consolidating that integrates all components of the security solution. Whist maintaining the balance of counter cyber security measures, do not lower the requirements of primary obligations of data protection, the protection of other industry standards and the evolving threats to cyberspace of cyber security. The analysis of content, such as deep fakes and AI-created fake phishing that is fake content, poses new challenges in cybersecurity for which the countering is drawn from NLP. NLP enhances cybersecurity by processing vast amounts of unstructured data from diverse sources to identify threats [22].

7. Future Scope

The existing literature often discusses the limitations related to the adaptability of NLP models to the dynamically evolving cybersecurity landscape, the accuracy of threat detection, and the challenges in processing technical and domain-specific language. Enhancing NLP systems to process cybersecurity information across languages, widening

threat intelligence scope. Utilize sophisticated deep learning models, like transformers, to increase accuracy and context awareness in NLP for cybersecurity, aiming for precise threat detection and reduced false alerts. Combine NLP with other AI technologies for more robust cybersecurity solutions, leveraging the strengths of multiple approaches for adaptive protection.

References

[1] S. Salloum, T. Gaber, S. Vadera, and K. Shaalan, "Phishing email detection using natural language processing techniques: A literature survey," *Procedia Computer Science*, vol. 189, pp. 19–28 (2021).

[2] S. Salloum, T. Gaber, S. Vadera, and K. Shaalan, "A systematic literature review on phishing email detection using natural language processing techniques," *IEEE Access*, vol. 10, pp. 65703–65727 (2022).

[3] A. Alhogail and A. Alsabih, "Applying machine learning and natural language processing to detect phishing email," *Computers & Security*, vol. 110, pp. 102414 (2021).

[4] S. S. Ismail, R. F. Mansour, A. El-Aziz, R. M. Rasha, and A. I. Taloba, "Efficient email spam detection strategy using genetic decision tree processing with NLP features," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2022 (2022).

[5] M. R. Rahman, R. Mahdavi-Hezaveh, and L. Williams, "A literature review on mining cyberthreat intelligence from unstructured texts," in *Proc. 2020 Int. Conf. Data Mining Workshops (ICDMW)*, pp. 516–525 (Nov. 2020).

[6] H. Jo, Y. Lee, and S. Shin, "Vulcan: Automatic extraction and analysis of cyber threat intelligence from unstructured text," *Computers & Security*, vol. 120, pp. 102763 (2022).

[7] C. Y. Chaw and H. N. Chua, "A framework system using Word Mover's distance text similarity algorithm for assessing privacy policy compliance," in *IT Convergence and Security: Proc. ICITCS 2021*, pp. 79–89, Singapore: Springer (2021).

[8] O. A. Cejas, M. I. Azeem, S. Abualhaija, and L. Briand, "NLP-based automated compliance checking of data processing agreements (DPAs) against GDPR," *IEEE Transactions on Software Engineering* (2023).

[9] G. Siewruk and W. Mazurczyk, "Context-aware software vulnerability classification using machine learning," *IEEE Access*, vol. 9, pp. 88852–88867 (2021).

[10] S. Sharma and T. Arjunan, "Natural language processing for detecting anomalies and intrusions in unstructured cybersecurity data," *International Journal of Information and Cybersecurity*, vol. 7, no. 12, pp. 1–24 (2023).

[11] Z. T. Sworna, Z. Mousavi, and M. A. Babar, "NLP methods in host-based intrusion detection systems: A systematic review and future directions," *Journal of Network and Computer Applications*, Article ID 103761 (2023).

[12] N. Tsinganos and I. Mavridis, "Building and evaluating an annotated corpus for automated recognition of chat-based social engineering attacks," *Applied Sciences*, vol. 11, no. 22, pp. 10871 (2021).

[13] S. Sharma and A. Jain, "Role of sentiment analysis in social media security and analytics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 5, pp. e1366 (2020).

[14] R. Marinho and R. Holanda, "Automated emerging cyber threat identification and profiling based on natural language processing," *IEEE Access* (2023).

[15] F. Spegni, A. Sabatelli, A. Merlo, L. Pepa, L. Spalazzi, and L. Verderame, "A precision cybersecurity workflow for cyber-physical systems: The IoT healthcare use case," in *Proc. European Symp. Research in Computer Security*, pp. 409–426, Springer (2022).

[16] M. M. Yamin and B. Katt, "Modeling and executing cyber security exercise scenarios in cyber ranges," *Computers & Security*, vol. 116, pp. 102635 (2022).

[17] M. A. Haq, M. A. R. Khan, and M. Alshehri, "Insider threat detection based on NLP word embedding and machine learning," *Intelligent Automation & Soft Computing*, vol. 33, no. 1, pp. 619–635 (2022).

[18] S. Silvestri, S. Islam, D. Amelin, G. Weiler, S. Papastergiou, and M. Ciampi, "Cyber threat assessment and management for securing healthcare ecosystems using natural language processing," *International Journal of Information Security*, pp. 1–20 (2023).

[19] R. Vinaya Kumar, K. P. Soman, P. Poornachandran, and V. K. Menon, "A deep dive on machine learning for cyber security use cases," in

Machine Learning for Computer and Cyber Security, pp. 122–158, CRC Press (2019).

[20] T. M. Georgescu, "Natural language processing model for automatic analysis of cybersecurity-related documents," *Symmetry*, vol. 12, no. 3, pp. 354 (2020).

[21] S. Arvind, B. Unhelkar, S. S. Shankar, P. Chakrabarti, and S. V. Devika, "Advancing cyber threat detection through deep learning in management information systems," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 8, pp. 2409–2417 (2024), doi: 10.47974/JDMSC-2014.

[22] S. K. Mandal, R. Singh, N. Chandu, D. Mehta, S. K. Henge, D. Kothapeta, C. K. Hinge, A. Sharma, and T. Hussain, "Evolutionary multi-authentication cryptographic key implications for secure data access control," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 28, no. 5-B, pp. 1865–1874 (2025), doi: 10.47974/JDMSC-2363.

Received March, 2025