engineering proceedings

MDPI

Proceeding Paper

A Bibliometric Analysis of Phishing Detection Using NLP in Business Enterprises †

Yadana Myint Hein *, Kumuduni Ranasinghe, Noushad Sahad, Shuang Chiao Wan, Naoki Sekizawa and Yoshitaka Kuroiwa

Graduate Department of Computer and Information Systems, The University of Aizu, Aizuwakamatsu 965-0006, Japan; krsranasinghe@gmail.com (K.R.); sahad.noushad1@gmail.com (N.S.); charliewan705@gmail.com (S.C.W.); aelc1358@gmail.com (N.S.); m5291045@u-aizu.ac.jp (Y.K.)

* Correspondence: yadanamhein@gmail.com

† Presented at the 8th International Global Conference Series on ICT Integration in Technical Education & Smart Society, Aizuwakamatsu City, Japan, 20–26 January 2026.

Abstract

The advancement of natural language processing (NLP), transformer architectures, and large language models (LLMs) has reshaped phishing detection research within business and enterprise environments. However, the structural evolution, thematic transitions, and collaboration patterns of this domain remain insufficiently mapped. This study conducts a bibliometric analysis of Scopus-indexed publications from 2020 to 2025. Using VOSviewer and Bibliometrix (RStudio), we perform performance analysis and science mapping, including co-authorship, co-citation, bibliographic coupling, and keyword co-occurrence analyses. The findings reveal a clear methodological shift from traditional machine learning toward deep learning and transformer-based architectures, particularly after 2023. Two dominant research clusters emerge: conventional feature-based phishing detection and NLP-driven AI security approaches. While large language models and multi-channel phishing detection are gaining prominence, enterprise-level implementation and interdisciplinary integration remain limited. This study identifies emerging trends, collaboration gaps, and underexplored themes, providing directions for future research and practical cybersecurity development.

Keywords: bibliometric analysis; phishing detection; natural language processing; large language models; transformer architectures; business enterprise; science mapping

1. Introduction

Phishing is a cyberattack in which criminals pretend to be trusted organizations to steal sensitive information such as passwords or financial details [1]. Email phishing is a common form of this attack, where fake emails are used to trick recipients into clicking malicious links, downloading harmful attachments, or sharing private data. According to the Verizon 2025 Data Breach Investigations Report [2], phishing is strongly linked to broader social engineering activity and credential compromise. The report indicates that social engineering attacks represent a significant portion of security incidents in both small and large organizations. Phishing accounts for nearly all incidents within this category. Phishing continues to be a serious and growing cybersecurity threat, with 3.8 million attacks reported in 2025 alone. Attackers are no longer relying only on email, but are increasingly targeting different platforms like social media as target victims. Recent findings from the APWG Phishing Activity Trends Report [3] also highlight that social media is a highly

Academic Editors: Debopriyo Roy, George Fragulis and Peter Ilic

Published: 20 July 2026

targeted industry sector influenced by phishing by the end of 2025. Furthermore, those threats can be of two types, Scams and Impersonation.

With the growing integration of natural language processing (NLP) and advanced artificial intelligence (AI) techniques, phishing detection research has shifted from rule-based filtering to traditional machine learning, as well as to deep learning architectures and large language models capable of analyzing contextual, semantic, and behavioral signals across diverse communication channels [4,5]. As phishing detection expands beyond traditional email to collaborative platforms, messaging applications, and social media ecosystems, understanding how this research domain evolves becomes critical for strengthening organizational cybersecurity resilience [6]. Analyzing research trends and gaps is therefore essential to identify dominant technological trajectories, emerging application contexts, underexplored themes, and patterns of global collaboration that shape the intellectual development of NLP-driven phishing detection. By utilizing bibliometric analysis to reveal citation structures, thematic clusters, and collaboration networks often overlooked by traditional reviews, this study maps the intellectual landscape of NLP-based phishing detection in business contexts from 2020 to 2025 to identify dominant themes, shifting trends, and critical research gaps. This study addresses seven key research questions:

1. How does the evolutionary growth of NLP architectures compare to broad-spectrum AI techniques in phishing detection literature between 2020 and 2025?
2. Which NLP architectures and language models occupy central positions in the citation and keyword networks of phishing detection research?
3. What emerging application domains and digital communication channels beyond email are reflected in recent phishing detection research using NLP?
4. Who are the most influential authors, institutions, and countries in this research domain, and how do they collaborate internationally?
5. Which research themes exhibit declining relevance, and which themes demonstrate strong growth or citation bursts over time?
6. What under-researched themes and weakly connected clusters suggest potential gaps in current NLP-based phishing detection strategies for business enterprises?
7. To what extent does the existing literature emphasize organizational and industry-oriented applications compared to methodological or theoretical contributions?

The goal of this study is to quantitatively map the intellectual structure, thematic evolution, and collaboration patterns, technological dominance of the research on phishing detection using NLP in business and enterprise contexts from 2020 onward. NLP techniques are widely used to detect email-based phishing by analyzing the text in the email body, subject, and headers. Under this category, techniques such as basic text preprocessing and feature extraction, applying machine learning, and deep learning to identify suspicious patterns were also used. Recently, LLMs have improved detection by better understanding context and meaning [7]. LLMs have shown strong potential as cost-effective tools for detecting phishing emails, with lightweight open-source models to achieve high accuracy [8]. However, they can still be fooled by highly sophisticated and context-driven social engineering attacks, so they should support human judgment rather than fully replace it in cybersecurity decisions. The remainder of this paper is organized into Section 2, which outlines the literature review, and Section 3 presents the methodology used for this bibliometric analysis, with the statistical representations. Section 4 visualizes the results by addressing the above-mentioned research questions, while Section 5 mentions the limitations of this study. Finally, the paper concludes with Section 5 with the limitations of this analysis and conclusion under Section 6, and future work, followed by the list of references.

https://doi.org/10.3390/engproc2026143044

2. Literature Review

2.1. Bibliometric Analysis Theory

Bibliometric analysis is a quantitative approach to assess and map the intellectual structure of a research area through statistical analysis of publication literature [9]. It allows researchers to find trends of publication, important contributors, evolving themes, and collaborations for a given field. Unlike traditional narrative reviews, bibliometric reviews can derive objective and repeatable conclusions through citation relations, keyword networks, and co-authorship connections [9,10].

In fast-moving technology areas like cybersecurity and NLP, bibliometric studies are of special value. These areas are undergoing rapid methodological change, and their integration takes place at the interface of scientific disciplines with a more global network of collaboration. Researchers can thus systematically investigate the emergence, evolution, and transition of research topics over time using bibliometric approaches. This method will facilitate an evidence-based identification of research gaps and new opportunities [9].

Recent methodological guidelines emphasize the importance of combining performance analysis and science mapping techniques to achieve a comprehensive understanding of a research landscape [9]. Therefore, bibliometric analysis provides a suitable methodological foundation for examining the evolution of NLP-driven phishing detection research within enterprise contexts.

2.2. Types of Bibliometric Analysis

Bibliometric analysis generally consists of two primary approaches: performance analysis and science mapping [9].

Performance Analysis: It analyzes performance by considering research productivity and impact. It considers indexes such as growth in the number of yearly publications, total citations, average number of citations per paper, and h-index [9]. These measurements enable scholars to measure the evolutionary status of a research field as well as identify significant authors, institutions, and nations. In the space of phishing detection literature, performance analysis also highlights how research has focused more since deep learning and transformer-based models have come into prominence [8,11].

Science mapping, in contrast, focuses on the structural relations between concepts within a field. It uses the network analysis method to reveal relationships among authors, articles, keywords, and institutions [9]. Typical science mapping techniques involve co-author analysis, co-citation analysis, bibliographic coupling, and keyword occurrence. These methods can be used to help uncover the main lines of research, classic works, emerging topics, and collaborative patterns.

By integrating performance analysis with science mapping, researchers can obtain both quantitative impact metrics and structural insights into thematic evolution. This combined approach is widely recommended in contemporary bibliometric research methodologies [9,10].

2.3. Bibliometric Studies in Cybersecurity and Phishing Research

Several bibliometric analyses have looked at research trends in cybersecurity, AI, and machine learning applications [10]. Such studies usually examine the growth of publications, collaboration networks, and thematic clusters of large cybersecurity domains. Yet, the majority of previous work addresses phishing detection as a component of the larger cybersecurity environment and fails to focus exclusively on the use of natural language processing methods in this context.

Furthermore, previous bibliometric studies typically concentrate on the general use of machine learning rather than the development of lingual detection approaches. The advent

of transformer models and LLMs has drastically changed the methodology of phishing detection research [8,11,12]. Nonetheless, there is little systematic mapping of the extent to which NLP architectures have impacted enterprise-oriented phishing detection research.

Additionally, while performance-based studies frequently evaluate model accuracy and detection efficiency [11,13], fewer investigations analyze international collaboration patterns, thematic evolution across time periods, and interdisciplinary integration within this specialized domain.

3. Methodology

3.1. Research Design

This study examines the global research landscape of NLP-based phishing detection within business and organizational contexts. The study followed a structured three-phase workflow to ensure transparency and reproducibility and adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to enhance reporting quality. This methodological approach is also inspired by the bibliometric–systematic Literature Review (B-SLR) framework [14]. Specifically, our workflow aligns with the screening, cleaning, and cluster identification protocols outlined in the B-SLR logic, although the systematic literature review part is out of the scope of this study. The study covers publications from 2020 to 2025, a period marked by the rapid advancement of Transformer-based architectures and LLMs [15], which have significantly reshaped both phishing attack strategies and detection methodologies.

3.2. Data Source and Search Strategy

3.2.1. Database Selection

All data were retrieved exclusively from Scopus. The database was selected due to its extensive coverage of peer-reviewed journals and conference proceedings, ensuring a high-quality and comprehensive dataset suitable for bibliometric analysis [16]. Selecting Scopus as the sole data source aligns with [17,18], who emphasize Scopus for its broad reach in security and NLP research. In addition, Scopus provides structured metadata exports compatible with bibliometric software tools, making it widely recognized as a reliable source for large-scale quantitative literature analysis.

3.2.2. Search Strategy and Phased Filtering

The search was conducted within Title, Abstract, and Keywords fields. A three-phase search strategy was implemented to progressively refine the dataset from a broad conceptual scope to a targeted intersection of phishing, AI-driven language technologies, and business applications. This phased approach is to ensure that the final dataset captures studies situated at the intersection of phishing, advanced NLP methodologies, and organizational applications.

- Phase 1-Broad Identification: The term “phishing” was used to capture the overall research landscape.
- Phase 2-Technical Refinement: The results from Phase 1 were refined by incorporating AI- and NLP-related terms. This phase ensured methodological relevance to computational linguistics and modern AI-based detection techniques. The terms included: “natural language processing,” “NLP,” “text classification,” “text mining,” “transformer*” “language model*” “LLM” “generative AI,” “GPT” and “BERT”
- Phase 3-Contextual Focus: The dataset was further narrowed by adding business-related terms such as “business,” “enterprise*” “finance” “corporate” “organization*""company” and “industry” to enhance the specificity of selection within the field of business or organizational specialization.

The asterisk (*) was used as a wildcard symbol to capture variations of a root term (e.g., "transformers" or "language models" or "enterprise") in database searches.

3.2.3. Inclusion and Exclusion Criteria

The comprehensive search string incorporated inclusion, exclusion, and filtering parameters related to publication year, language, document type, and subject area. Publications were limited to the period 2020–2025, English language, and relevant multidisciplinary subject areas including Computer Science, Engineering, Mathematics, Business, Decision Sciences, Social Sciences, Economics, Arts and Humanities, and Multidisciplinary studies. Books and book chapters were excluded to prioritize peer-reviewed research outputs. The exclusion and inclusion criteria were consistent in all three phases that are mentioned above. By restricting the dataset to these parameters, the study minimized disciplinary noise while still allowing sufficient scope to explore interdisciplinary relationships that contribute meaningfully to this bibliometric analysis.

3.3. Screening Process

The process was conducted in sequential stages aligned with the three-phase search strategy.

In Phase 1, a total of 6596 records were identified using the keyword "phishing." Phase 2 refined the dataset by applying AI- and NLP-related terms, which excluded 5652 records and reduced the total to 944. Phase 3 incorporated business and organizational terms, resulting in the exclusion of 684 records and leaving 260 studies for further evaluation.

To ensure data integrity, duplicate records were identified and removed prior to analysis. Duplicate entries in Scopus exports may occur due to multiple export batches or minor metadata inconsistencies. A total of 37 duplicate records were removed using a structured multi-key validation procedure with the following priority hierarchy: (1) DOI (Digital Object Identifier), (2) Title and Year, and (3) Title and First Author. Duplicate detection and removal were conducted programmatically in RStudio using the bibliometrix [19] package rather than relying on manual spreadsheet-based filtering.

The cleaning procedure comprised the following steps:

1. Initial inspection of exported metadata;
2. Removal of duplicate records based on DOI;
3. Secondary validation using combinations of Title and Year.

After completing all screening and cleaning procedures, the final dataset consisted of 223 publications, which were retained for bibliometric and thematic analysis.

Figure 1 illustrates a structured, three-phase bibliometric retrieval process using the Scopus database to isolate relevant literature. It employs a funneling approach, narrowing an initial pool of 6596 papers down to a refined dataset of 223 unique articles by applying specific filters for NLP techniques, business contexts, and a 2020–2025 timeframe.

3.4. Bibliometric Analysis Tools

The final analysis was conducted using the main bibliometric analysis techniques, such as performance analysis and science mapping [9]. Two complementary software tools were employed to examine intellectual structure, thematic development, collaboration patterns, and cold spots within the field. This 'dual-tool' approach (VOSviewer and Bibliometrix) is recognized as an innovative methodological pillar for modern bibliometric studies, as it allows for simultaneous network visualization and rigorous data normalization [20].

VOSviewer (Version 1.6.20) [21] was used to construct and visualize bibliometric networks. The software generated network, density, and overlay visualizations to identify dominant research themes and keyword relationships. These visualizations

highlighted phishing as the central research focus and illustrated both dominant and non-dominant keywords.

R (version 4.5.1) and RStudio (version 2025.09.2+418), together with the Bibliometrix package, were employed to perform statistical analysis and thematic mapping. This tool enabled thematic classification (Basic, Motor, and Niche themes), temporal keyword evolution analysis from 2020 to 2025, and collaboration mapping among authors, institutions, and countries.

The bibliometric techniques applied in this study included co-authorship analysis, co-occurrence analysis, co-citation analysis, and bibliographic coupling. Collectively, these methods revealed collaboration networks, emerging research clusters, and foundational studies, thereby contributing to an understanding of trends in both traditional machine learning and modern transformer-based large language model (LLM) approaches to phishing detection research.

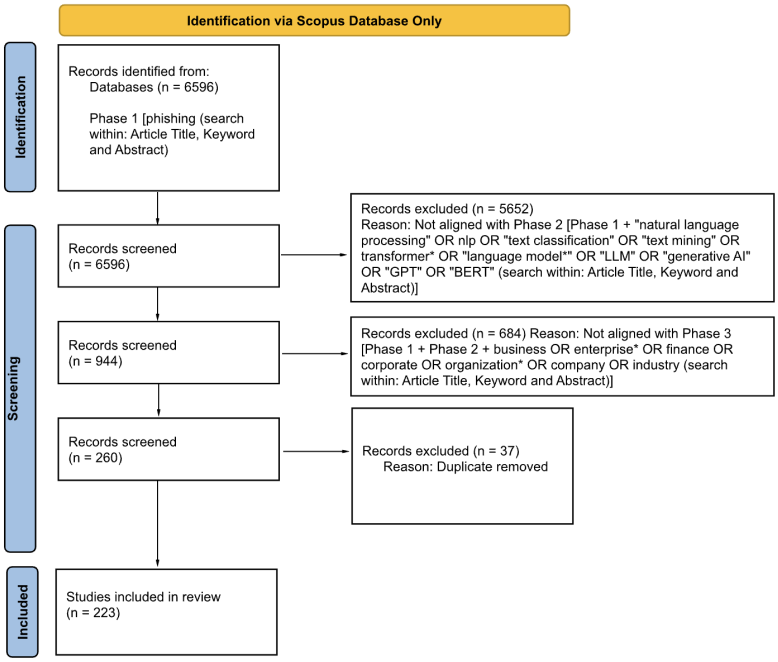

Figure 1. PRISMA diagram for the selection and filtering of research papers on NLP-based phishing detection in business enterprises (2020–2025) for bibliometric analysis.

4. Results and Discussion

As mentioned, we initially gathered 260 papers for this bibliometric analysis and subsequently selected 223 by applying our targeted filters, such as the specified time period and publication type. Finally, a total of 37 papers were retained after removing duplicates. This study analyses literature across several categories, including the evaluation of phishing-related research, technological dominance, the expansion of phishing detection beyond email, the most influential contributions, research trend analysis, existing gaps in current detection strategies, and the overall impact on industry. RStudio and VOSViewer were used to conduct the bibliometric analysis in the research.

4.1. Methodological Evolution

The longitudinal analysis of keyword frequencies between 2020 and 2025 shows a clear shift in the technical focus of the field. The Figure 2 illustrates how the general methodologies moved away and shows the strong emphasis on NLP-related techniques.

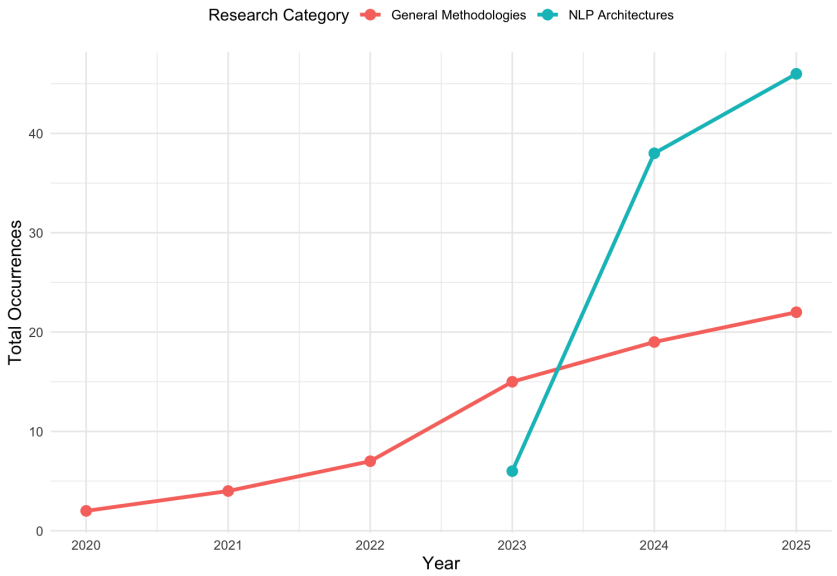

Figure 2. Comparative evolution of research(2020–2025). Aggregated frequency analysis based on keywords, illustrating the growth of NLP techniques relative to existing machine learning methodologies.

While the traditional machine learning methods remain relevant, the current research is largely driven by the NLP architectural background. From 2020 to 2022, research in this field mainly relied on the general machine learning techniques, and a noticeable change began in 2023. Although the traditional methods grew gradually, NLP architecture-based research, including transformers, BERT, and LLMs, increased rapidly from 2023. By 2024, NLP-based approaches surpassed general machine learning methodologies in terms of research publication frequency. The number of studies using NLP increased sharply from around six occurrences in 2023 to nearly 50 by 2025. This rapid growth indicates that the researchers are prioritizing LLMs and generative AI to address the more complex threats in phishing attacks.

4.2. Technological Dominance

The keyword co-occurrence network reveals a clear structural division within the research landscape, indicating a significant technological shift in phishing detection studies. This analysis was conducted to examine technological dominance before and after 2023 because the majority of NLP-based research began to emerge prominently from 2023 onward. By analyzing keyword frequency, network centrality measures, and citation burst patterns, this study identifies which language models and technologies hold central and influential positions.

The Figure 3 illustrates the keyword co-occurrences network, which exhibits a clear bifurcation into two primary clusters based on the year. The red cluster, which denotes the NLP frontier, indicates that the advanced NLP architectures occupy central positions. In this cluster, the dominant research areas like transformers, BERT, and LLMs occupy central positions with high betweenness centrality. Conversely, the blue cluster encompasses foundational machine learning methodologies.

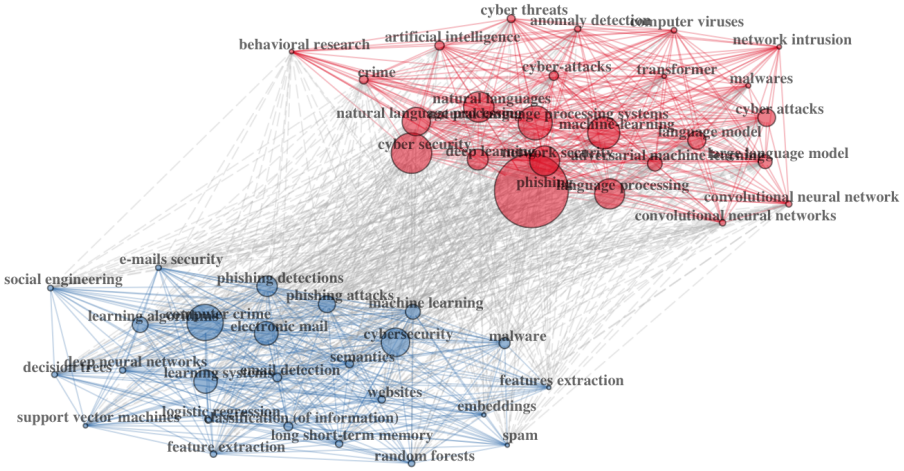

Figure 3. Keyword co-occurrence network revealing a binary structural division that undergoes a significant technological shift in phishing detection research.

4.3. Expansion of the Filed

To examine whether phishing detection research has expanded beyond traditional email-based systems, a thematic cluster analysis was conducted. This provides evidence of the field's gradual expansion towards broader and more diverse digital environments.

The thematic map visualized in Figure 4 categorized the research topics into four quadrants based on centrality and density to examine structural organization and find the under-discovered areas. The resulting quadrant framework visualizes the Motor themes, Niche themes, Basic themes, and Emerging or declining themes [22]. Centrality indicates the relevance of the theme in the overall field, while density defines the internal development and maturity of a specific subtopic.

The current research momentum, identified under the motor themes quadrant, is primarily driven by the integration of network security, cyber attacks, and language models. These themes are both highly relevant and have reached a high level of internal development in the field.

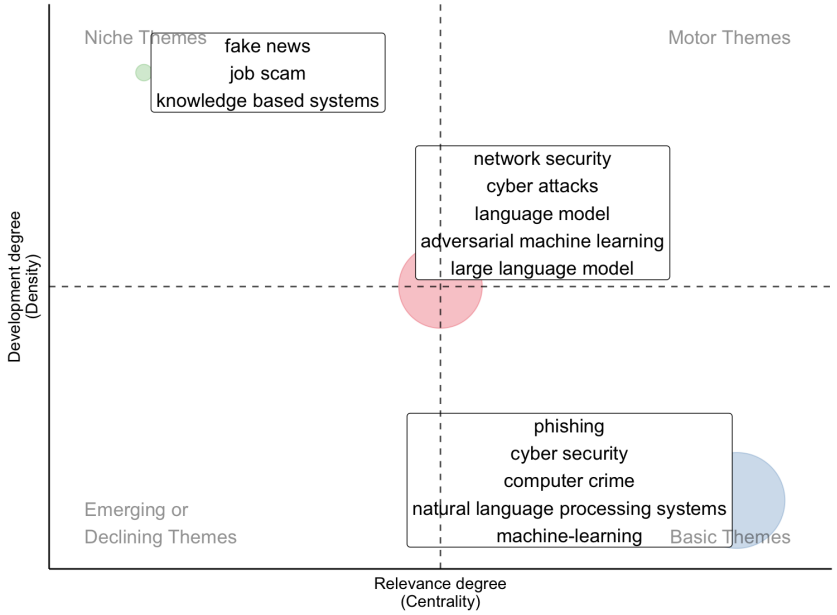

Figure 4. Thematic map illustrating emerging application domains and digital communication channels in NLP-based phishing detection research.

The basic themes that appear in the bottom right quadrant represent the foundational topics within the field. They are critically important and relevant to almost every paper in the dataset, and they include "phishing", "cyber security", "computer crime", and "natural language processing systems". Although they are highly central and relevant to most studies in the data set, they are considered established and general topics rather than specialized areas. Niche themes, located in the top left quadrant, denote the highly specialized and technically developed areas. It remains relatively isolated from the broader research landscape. The areas such as "fake news detection" and "job scams" are not strongly integrated into the cybersecurity and phishing detection frameworks.

4.4. Influential Actors and Collaboration

A bibliography coupling analysis at the country level was conducted to identify the leading countries in this research domain. This reveals the intellectual structure and collaborative intensity of phishing detection across different countries.

The Figure 5 illustrates the coupling network, and the node sizes represent the volume of the publication, while link strength reflects the extent to which the countries share similar reference patterns. The results show that China, the United States, India, the United Kingdom, and Germany emerge as major research hubs. China appears as one of the most prominent modes, by indicating high publication volume and strong bibliographic linkages with multiple nations. The United States acts as a central connector between Europe and Asia while demonstrating a substantial coupling strength. India and Saudi Arabia also form a strong regional cluster. These cross-regional linkages suggest active international collaboration, and these findings confirm that phishing detection research is supported globally.

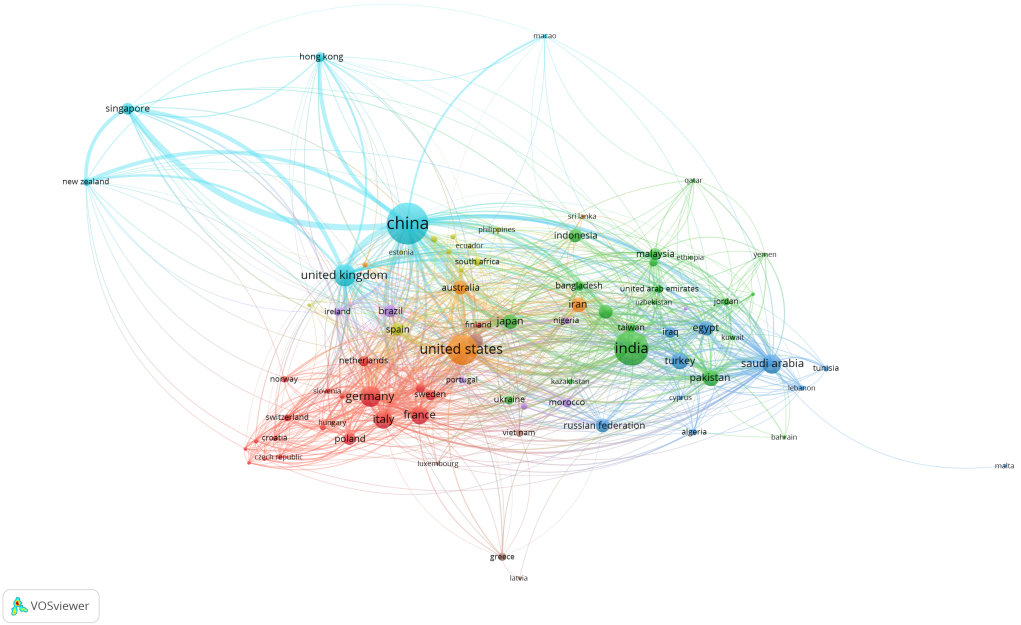

Figure 5. Bibliographic coupling network of countries, illustrating intellectual linkages and collaborative clusters. (2020–2025).

4.5. Keyword Co-Occurrence Analysis

A keyword co-occurrence analysis was conducted to examine the conceptual structure of phishing detection research. This approach identifies how frequently keywords appear together in the literature by revealing major themes and relationships between different topics.

The Figure 6 shows a keyword co-occurrence network generated using VOSviewer. Each node represents a keyword, where the size indicates how frequently a term appears in the literature, and the nodes that are placed together are frequently studied together. Large central nodes like “phishing”, “Computer crime”, and “cyber attacks “ obtained high density and numerous connections, confirming that they are the core concepts linking all sub-domains of the field.

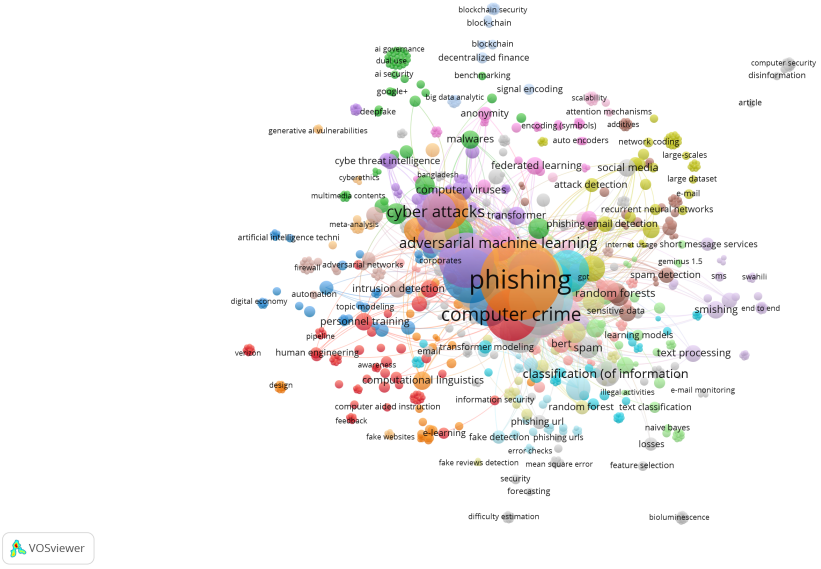

Figure 6. Keyword co-occurrence network illustrating the structural organization of phishing research in business environments.

4.6. Industry Orientation

The orientation of the existing literature was examined to determine whether research in phishing detection is primarily focused on organizational and industry applications or on methodological and theoretical development. A structured keyword-based classification was applied, using predefined industry and theory-related terms identified within paper abstracts. Industry-oriented terms such as “enterprise”, “corporate”, and “deployment” were represented, while “algorithm”, “model”, and “architecture” were used to indicate methodological focus.

Table 1 displays the classification results of the phishing detection studies. Out of 223 publications, 195 studies (87.4%) reference organizational or enterprise context, while 28 studies (12.6%) focus primarily on theoretical aspects like algorithmic or technical modeling. These results indicate that a large portion of the literature represents phishing detection within an organizational or enterprise context. The findings point to a substantial gap between methodological development and deployment level evaluation in the phishing detection field.

Table 1. Distribution of phishing detection studies by research orientation, showing the proportion of industry-oriented and methodological/theoretical publications.

Industry-Oriented	Theory-Oriented
195	28

5. Limitation

While this study provides a structured and quantitative overview of phishing detection research using NLP in enterprise contexts, several limitations must be acknowledged.

5.1. Database Limitation

First, this analysis relied exclusively on the Scopus database. Although Scopus is one of the largest and most comprehensive academic indexing platforms, it does not include all relevant publications. Important studies indexed in other databases (e.g., Web of Science, IEEE Xplore, ACM Digital Library) may have been excluded. Therefore, the dataset may not fully represent the entire body of global research on phishing detection using NLP.

5.2. Language Restriction

For this study, only English-language publications were included. This language restriction may have excluded significant contributions published in other languages. As cybersecurity is a global issue, regional studies conducted in non-English contexts might provide valuable insights that are not reflected in this analysis.

5.3. Differences in Research Quality

At last, bibliometric analysis primarily relies on quantitative indicators such as publication counts and citation metrics. However, citation frequency does not always directly correspond to research quality. Some highly cited works may benefit from early publication timing or broad methodological relevance, while highly rigorous studies published in niche venues may receive fewer citations.

Additionally, certain influential researchers publish high-quality papers infrequently (e.g., once every few years), which may reduce their visibility in productivity-based metrics despite their strong academic impact. Therefore, quantitative centrality in citation networks should not be interpreted as a definitive measure of scholarly excellence.

5.4. Limited Sample Size After Filtering

Finally, after applying inclusion and exclusion criteria, only 223 highly relevant papers were selected for in-depth analysis. While this ensures thematic precision, it also limits statistical generalizability. The relatively small dataset may influence network density and clustering outcomes in science mapping analyses.

6. Conclusions

This study presented a bibliometric analysis of research on phishing detection using NLP techniques in business and enterprise contexts between 2020 and 2025. By applying performance analysis and science mapping techniques to data retrieved from Scopus, the study quantitatively mapped the intellectual structure, thematic evolution, and collaboration patterns of this rapidly developing research domain.

The results reveal a clear methodological transition from traditional machine learning approaches to deep learning and transformer-based NLP architectures. Until approximately 2023, research was largely dominated by feature-based models and classical machine learning algorithms. However, after 2023, transformer-related terms and LLMs became central within citation and keyword networks, indicating a paradigm shift toward generative and contextual language modeling approaches.

Keyword co-occurrence and thematic mapping analyses further demonstrate the emergence of new research directions beyond conventional email-based phishing detection. Recent studies increasingly address multi-channel phishing environments, including social media, messaging platforms, and AI-generated phishing content. Additionally, adversarial machine learning and explainability have begun to form bridging themes between cybersecurity and advanced NLP research.

Despite the rapid academic growth in LLM-based phishing detection, the findings suggest that real-world enterprise deployment remains limited. Organizational integration,

scalability, explainability, and cost-efficiency are still underexplored areas. Therefore, future research should focus not only on improving detection performance but also on bridging the gap between academic innovation and practical enterprise cybersecurity implementation.

7. Future Directions

Furthermore, the under-researched areas and weakly connected clusters in the thematic map suggest the potential gaps for future research. Although specialized areas such as fake news detection and job scams are identified as technically developed (high density), they remain relatively disconnected from the broader phishing and cybersecurity domain. So it is important to connect these specialized subtopics in niche themes to the main security frameworks. Furthermore, for the foundational topics, it needs newer, more dense research. Addressing these gaps in future research could connect these specialized topics with broader cybersecurity models, enabling more integrated and multi-domain threat detection strategies.

Author Contributions: Y.M.H.: Methodology, Supervision, Writing—original draft, Writing—review & editing; K.R.: Data curation, Investigation, Visualization, Writing—original draft, Writing—review & editing; N.S. (Noushad Sahad): Resources, Writing—original draft, Writing—review & editing; S.C.W.: Validation, Writing—original draft, Writing—review & editing; N.S. (Naoki Sekizawa): Writing—original draft, Writing—review & editing; Y.K.: Writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available through the Scopus database and can be accessed subject to Scopus's terms and access policies.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bountakas, P.; Koutroumpouchos, K.; Xenakis, C. A Comparison of Natural Language Processing and Machine Learning Methods for Phishing Email Detection. In Proceedings of the 16th International Conference on Availability, Reliability and Security, New York, NY, USA, 17 August 2021; p. 12. [CrossRef]
2. Hylender, C.D.; Langlois, P.; Pinto, A.; Widup, S. *Verizon 2025 Data Breach Investigations Report*; Technical Report; Verizon Business: New York, NY, USA, 2025.
3. Anti-Phishing Working Group. Phishing Activity Trends Report. In *Technical Report 4th Quarter 2025*; APWG: Lexington, KY, USA, 2026.
4. Popescul, D.; Radu, L.D. AI in Phishing Detection: A Bibliometric Review. *Front. Artif. Intell.* **2025**, *8*, 1496580. [CrossRef] [PubMed]
5. Thapa, J.; Chahal, G.; Gabreanu, S.V.; Otoum, Y. Phishing Detection in the Gen-AI Era: Quantized LLMs vs. Classical Models. In Proceedings of the 2025 IEEE/ACIS 29th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Busan, Republic of Korea, 25–27 June 2025; pp. 856–863. [CrossRef]
6. Wangchuk, T.; Gonsalves, T. Multimodal Phishing Detection on Social Networking Sites: A Systematic Review. *IEEE Access* **2025**, *13*, 1. [CrossRef]
7. Salloum, S.; Gaber, T.; Vadera, S.; Shaalan, K. A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. *IEEE Access* **2022**, *10*, 65703–65727. [CrossRef]
8. Zhang, J.; Wu, P.; London, J.; Tenney, D. Benchmarking and Evaluating Large Language Models in Phishing Detection for Small and Midsize Enterprises: A Comprehensive Analysis. *IEEE Access* **2025**, *13*, 28335–28352. [CrossRef]
9. Donthu, N.; Kumar, S.; Mukherjee, D.; Pandey, N.; Lim, W.M. How to Conduct A Bibliometric Analysis: An Overview and Guidelines. *J. Bus. Res.* **2021**, *133*, 285–296. [CrossRef]
10. Kumar L., M.; George, R.J.; P.S., A. Bibliometric Analysis for Medical Research. *Indian J. Psychol. Med.* **2022**, *45*, 277–282. [CrossRef] [PubMed]

11. Gupta, B.B.; Gaurav, A.; Arya, V.; Attar, R.W.; Bansal, S.; Alhomoud, A.; Chui, K.T. Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems. *Comput. Model. Eng. Sci.* **2024**, *141*, 2165–2183. [CrossRef]
12. Bethany, M.; Galiopoulos, A.; Bethany, E.; Karkevandi, M.B.; Beebe, N.; Vishwamitra, N.; Najafirad, P. Lateral Phishing with Large Language Models: A Large Organization Comparative Study. *IEEE Access* **2025**, *13*, 60684–60701. [CrossRef]
13. Alguliyev, R.; Aliguliyev, R.; Sukhostat, L. An Approach for Business Email Compromise Detection using NLP and Deep Learning. In Proceedings of the IEEE 18th International Conference Application of Information and Communication Technologies (AICT), Turin, Italy, 25–27 September 2024; pp. 1–6. [CrossRef]
14. Marzi, G.; Balzano, M.; Caputo, A.; Pellegrini, M.M. Guidelines for Bibliometric-Systematic Literature Reviews: 10 Steps to Combine Analysis, Synthesis and Theory Development. *Int. J. Manag. Rev.* **2025**, *27*, 81–103. [CrossRef]
15. 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*, 957. [CrossRef]
16. Pranckutė, R. Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications* **2021**, *9*, 12. [CrossRef]
17. Rahim, N. Bibliometric Analysis of Cyber Threat and Cyber Attack Literature: Exploring the Higher Education Context. In *Cybersecurity Threats with New Perspectives*; Sarfraz, M., Ed.; IntechOpen: London, UK, 2021; pp. 1–18. [CrossRef]
18. Hassooni, M.N. Information Technology and Natural Language Processing in Education: A Systematic and Bibliometric Review (2020–2025). *Am. J. Appl. Sci.* **2026**, *8*, 46–71. [CrossRef]
19. Aria, M.; Cuccurullo, C. Bibliometrix: An R-tool for Comprehensive Science Mapping Analysis. *J. Inf.* **2017**, *11*, 959–975. [CrossRef]
20. Stefanis, C.; Stavropoulou, E.; Stavropoulos, A.; Gyriki, D.; Nikolaidis, C.G.; Vassos, V.; Nena, E.; Tsigalou, C.; Bezirtzoglou, E. Innovative methodological pillars for bibliometric studies: AI screening, data normalization and dual-tool analysis. *Discov. Appl. Sci.* **2025**, *7*, 973. [CrossRef]
21. Eck, N.v.; Waltman, L. Software Survey: VOSviewer, A Computer Program for Bibliometric Mapping. *Scientometrics* **2010**, *84*, 523–538. [CrossRef] [PubMed]
22. Cobo, M.J.; López-Herrera, A.G.; Herrera-Viedma, E.; Herrera, F. Science Mapping Software Tools: Review, Analysis, and Cooperative Study among Tools. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 1382–1402. [CrossRef]