Preprints.org

Article

Not peer-reviewed version

Text Mining Approaches for Exploring Research Trends in the Security Applications of Generative Artificial Intelligence

Jinsick Kim , Byeongsoo Koo , Moonju Nam , Kukjin Jang , Jooyeoun Lee * , Myoungsug Chung * , Yungseo Song *

Posted Date: 13 February 2025

doi: 10.20944/preprints202502.1011.v1

Keywords: Generative Artificial Intelligence(GAI); AI Security; ChatGPT Security; Text Mining Analysis; Research Trends in AI Security

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Text Mining Approaches for Exploring Research Trends in the Security Applications of Generative Artificial Intelligence

Jinsick Kim [1,2], Byeongsoo Koo [1], Moonju Nam [1], Kukjin Jang [1], Jooyeoun Lee [3,*], Myoungsug Chung [3,*] and Yungseo Song [3,*]

[1] Research Center for Science and Technology Policy and Convergence, Ajou University, 206, Suwon 16499, Republic of Korea
[2] Technology Policy Research Center, Institute for Advanced Engineering, Yongin 175-28, Republic of Korea
[3] Department of Industrial Engineering, Ajou University, 206, Suwon 16499, Republic of Korea
* Correspondence: jooyeoun325@ajou.ac.kr (J.L.); andrewchung@ajou.ac.kr (M.C.); slong@ajou.ac.kr(Y.S.)

Abstract: This study examines the security implications of generative artificial intelligence (GAI), focusing on models such as ChatGPT. As GAI technologies are increasingly integrated into industries like healthcare, education, and media, concerns are growing regarding security vulnerabilities, ethical challenges, and potential for misuse. To address these concerns, this research analyzes 1,047 peer-reviewed academic articles from the SCOPUS database using scientometric methods, including term frequency-inverse document frequency (TF-IDF) analysis, keyword centrality analysis, and latent dirichlet allocation (LDA) topic modeling. The results highlight significant contributions from countries such as the United States, China, and India, with leading institutions like the Chinese Academy of Sciences and the National University of Singapore driving research on GAI security. In the keyword centrality analysis, "ChatGPT" emerged as a highly central term, reflecting its prominence in the research discourse. However, despite its frequent mention, "ChatGPT" showed lower proximity centrality than terms like "model" and "AI." This suggests that while ChatGPT is broadly associated with other key themes, it has a less direct connection to specific research subfields. Topic modeling identified six major themes, including AI and security in education, language models, data processing, and risk management. The analysis emphasizes the need for robust security frameworks to address technical vulnerabilities, ensure ethical responsibility, and manage risks in the safe deployment of AI systems. These frameworks must not only incorporate technical solutions but also ethical accountability, regulatory compliance, and continuous risk management. This study underscores the importance of interdisciplinary research that integrates technical, legal, and ethical perspectives to ensure the responsible and secure deployment of GAI technologies.

Keywords: Generative Artificial Intelligence(GAI); AI Security; ChatGPT Security; Text Mining Analysis; Research Trends in AI Security

1. Introduction

1.1. Backgrounds

In recent years, interest in generative artificial intelligence (GAI) has surged, driven largely by its ability to autonomously generate various types of content, including text, images, speech, and videos. Traditionally, these tasks were performed by humans, but the advent of AI has significantly transformed this landscape. The global interest in generative AI has been further fueled by the introduction of innovative, user-friendly services like ChatGPT, which have made these technologies more accessible to the public [1]. ChatGPT, a large-scale AI language model, responds to questions with remarkable speed, accuracy, and quality, surpassing traditional chatbot services [2].

As shown in Figure 1, by August 2024, ChatGPT dominated the GAI chatbot market in the United States, holding a 60% market share, followed by the Microsoft Copilot program, also based on ChatGPT, at 14.6% [3].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

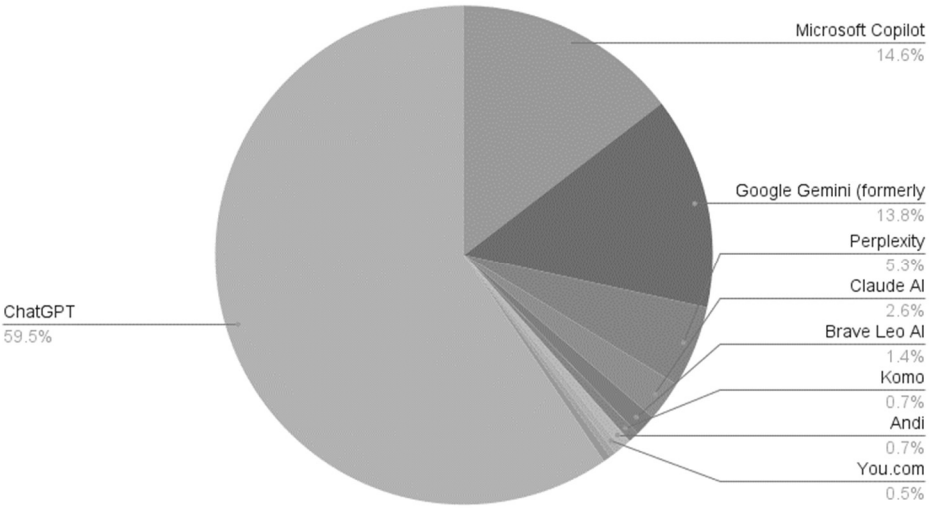

Figure 1. Generative AI chatbots by market share.

Recent studies, including Aljanabi's, have demonstrated that ChatGPT is capable of emulating human-like discourse and providing insightful responses to complex queries [4]. Trained on an extensive corpus of textual data, ChatGPT has gained a deep understanding of numerous subjects, enabling it to engage in coherent and contextually relevant discussions. One of its most remarkable features is its ability to address intricate inquiries. It can analyze nuanced questions, break them down into meaningful components, and generate structured responses that consider the context and subtleties of the conversation [5].

As illustrated in Figure 2, ChatGPT's capabilities have led to its widespread use in various fields, including healthcare [6,7], business [8], education [9,10], and content creation [11]. Remarkably, ChatGPT reached one million users within just five days of its launch, a feat that took Netflix three and a half years and Instagram two and a half months to achieve [12].

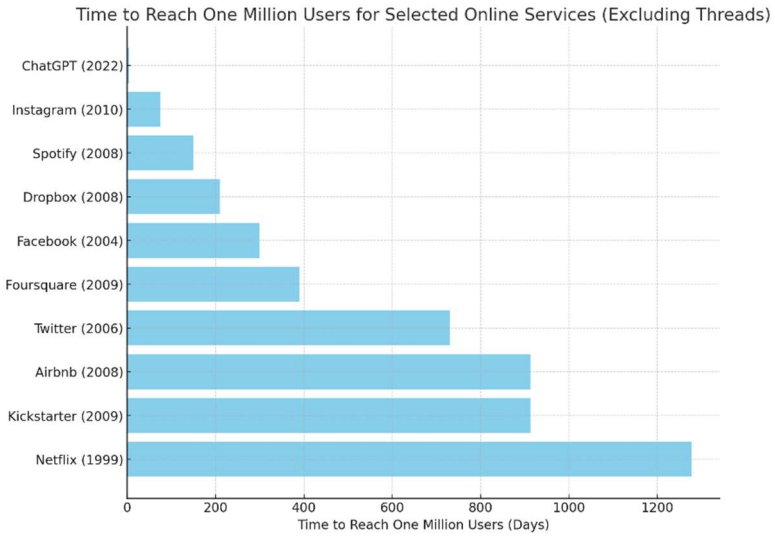

Figure 2. Time to reach one million users for selected online services.

Despite the widespread adoption of ChatGPT, a debate continues regarding the potential benefits and challenges posed by GAI models in various industries [13]. On the positive side, large-scale AI services like ChatGPT have the potential to enhance productivity in business processes and significantly impact industries such as healthcare, law, and media [14]. For instance, in healthcare, ChatGPT can analyze patient data to suggest personalized treatment plans; in law, it can efficiently review vast volumes of legal documents; and in media, it can automatically generate content to

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

provide timely information. The adoption of such AI technologies has the potential to reduce costs, improve efficiency, and enhance service delivery.

However, concerns remain about the risks associated with these technologies, particularly the potential for misinformation, misuse, and exploitation by malicious actors for cyberattacks and criminal activities [15]. For example, ChatGPT could be used to develop malicious software, automate phishing attacks, or spread misinformation, leading to social disruptions. Consequently, it is crucial to uphold ethical standards and establish regulations to mitigate these risks [14]. As GAI technologies continue to evolve rapidly, the discussion surrounding their safe integration into industries and the responses to these changes remains ongoing.

Therefore, further research is essential to explore the use of GAI models from a security perspective, ensuring that they can be safely and appropriately deployed across various industrial applications in the future. The objective of this study is to assess the potential security impact of using GAI models, such as ChatGPT, and to provide practical recommendations for their safer deployment in real-world applications. To achieve this, 1,047 SCI-level academic articles on GAI and security were collected and analyzed from the SCOPUS database using scientometric methods. The goal of this analysis is to identify both the strengths and weaknesses of GAI, offering strategic insights into the opportunities and threats it may pose to industrial security.

1.2. Previous Research on Generative AI

Building upon the discussion in the Background section, it is essential to explore research trends related to GAI, particularly in the context of security. As GAI technologies, such as ChatGPT, are increasingly integrated into various industries, it is crucial to identify and address the potential issues arising from this integration. Despite the relatively recent emergence of GAI, a significant body of research has already been established, exploring not only the technical capabilities of these models but also their applications across various socio-economic sectors, including tourism, marketing, manufacturing, trade, and education [4].

For instance, Rudolph et al. (2023) examined the challenges of value judgment and the application of GAI in higher education, emphasizing the complexities of integrating such advanced technologies into academic settings [16]. Similarly, Bai-doo-Anu et al. (2023) analyzed the potential benefits of utilizing GAI in educational environments, highlighting its ability to enhance student engagement and enable personalized learning experiences [17]. In the financial and business sectors, Wang et al. (2023) and Srivastava (2023) investigated how GAI could improve operational efficiency and support decision-making processes, demonstrating its profound impact on industry practices [18,19]. In healthcare, Kim (2025) utilized insights from ChatGPT to examine the transformative role of mobile health (mHealth) in managing obesity. The study identified trends such as personalized interventions, emerging technologies, remote monitoring, behavioral strategies, and user engagement. Additionally, it highlighted mHealth's potential to improve equity, self-management, and evidence-based practices, while addressing challenges related to privacy and interdisciplinary collaboration [20].

Despite the recognized potential of GAI models like ChatGPT, there is growing discourse surrounding the ethical and legal challenges arising from their technological limitations. Ethical concerns include biases embedded in training datasets, which may result in skewed or unfair outcomes, as well as the risk of plagiarism due to inaccuracies in citations and sources [21]. Furthermore, the use of AI-generated content raises broader questions about its authenticity and originality. On the legal front, discussions have centered around issues such as authorship recognition, intellectual property rights, and copyright infringements related to content generated by ChatGPT [22].

In the educational context, significant concerns have emerged regarding academic dishonesty, particularly the potential for cheating during exams due to an over-reliance on AI-generated content [23]. From an academic integrity standpoint, there are worries about the misuse of generative AI to produce fabricated data or entirely fictitious research papers, which could severely undermine the credibility and trustworthiness of scholarly work [24].

This study aims to distinguish itself from previous research by employing advanced topic modeling techniques to analyze and visualize the key issues surrounding GAI and security. By uncovering deeper semantic structures within the existing body of research, this approach seeks to

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

provide a more nuanced understanding of the potential opportunities and threats posed by the integration of GAI across industries, with a particular focus on security.

1.3. Previous Research on Bibliometrics

Building on the discussion of research trends related to GAI and the importance of analyzing these trends in the context of security, it is essential to explore methods that effectively quantify and evaluate the growing body of literature. While qualitative reviews provide valuable insights, they are limited in their ability to comprehensively assess large volumes of research. To address this, scientometric methods, such as text mining, have emerged as powerful quantitative tools for evaluating the significance of articles and authors in domains like cybersecurity and AI technology [25]. These methods enhance the review process by offering a structured, data-driven approach to understanding the relationships between research papers through graphical representation. However, while text mining can identify patterns in research, it often lacks the granularity needed to capture the nuanced, topic-specific context of these studies [26].

To bridge this gap, latent dirichlet allocation (LDA) topic modeling has emerged as a more advanced technique for discovering and mining latent information within a research domain. LDA topic modeling utilizes statistical and optimization algorithms to extract semantic insights from large text corpora, making it a valuable tool for systematically analyzing textual data [27]. This method has been widely applied across various fields, including scientific research trends, social networks, political discourse, biomedical recommendations, and public health initiatives [25]. In technology and cybersecurity, LDA topic modeling is particularly useful for identifying patterns in cybersecurity incidents, studying public perceptions of privacy, and uncovering emerging trends in artificial intelligence research.

A significant body of research has focused on the application of text mining, including LDA topic modeling, to provide overviews of scientific papers and uncover research topics. These studies can be broadly categorized into two groups [27]. In the first category, text mining is employed to provide broad overviews across a wide range of research areas. For instance, Hall et al. used unsupervised text-mining techniques to analyze historical trends in computational linguistics [28], while Yau et al. applied LDA topic modeling to cluster scientific documents and evaluate their findings [29].

The second category focuses on identifying trends within specific fields through text mining. For example, Jiang et al. conducted a bibliometric analysis using text mining to evaluate articles in hydropower research, revealing research developments, current trends, and intellectual structures [30]. Similarly, Choi et al. used text mining to analyze document abstracts and identify trends in personal data privacy research [31]. More recently, Kim et al. applied structural text network analysis to study articles on maritime autonomous surface ships (MASSs), identifying key trends in this rapidly evolving field [32].

In alignment with these advanced methodological approaches, the current study uses text mining, including LDA topic modeling—a natural language processing technique—to investigate key security topics within the context of GAI, such as ChatGPT. Before conducting the text mining analysis, frequency-reverse frequency analysis and centroid analysis were performed to identify the main keywords. LDA topic modeling was then applied to uncover deeper insights into the research. The findings, derived from both intersubjective and independent perspectives, were subsequently analyzed in relation to existing research. The goal of this study is to offer a comprehensive overview of current research trends, provide valuable insights and implications for ongoing studies, and suggest potential directions for future research on the safety of GAI in industrial applications.

2. Materials and Methods

2.1. Research Subjects

2.1.1 Collecting Research Data

The data for this study were obtained from Elsevier's SCOPUS database, one of the largest and most comprehensive scientific citation indexing sources. SCOPUS offers rich metadata, including abstracts and citations, from a wide array of scientific articles. The database encompasses over 76 million records from more than 39,100 journals, 120,000 conferences, and 206,000 books [33]. Given its extensive coverage, SCOPUS serves as an ideal resource for researching the growing field of GAI

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

and its security implications. To gain a comprehensive overview of the existing literature, a targeted search query was formulated. The search terms included "generative AI," "ChatGPT," "GPT-3," "large language model," and "LLM," combined with security-related terms such as "security," "cybersecurity," "privacy," "threat detection," and "adversarial attack." This query was applied to the SCOPUS database, covering the period from June 2022 to June 2024—an important timeframe in which significant advancements and discussions surrounding ChatGPT and related GAI technologies occurred. The initial search resulted in 1,241 articles, which were then refined through a systematic literature review. Articles lacking abstracts or not directly related to generative AI and its security implications were excluded. After this refinement, 1,047 articles remained, forming the foundation for subsequent text mining and scientometric analyses. These articles provide the necessary basis for uncovering key insights and trends related to the security challenges and opportunities posed by ChatGPT and other generative AI technologies in various industrial contexts.

2.1.2 Refining Research Data

Before analysis, the collected dataset underwent several pre-processing steps to enhance data quality and ensure the accuracy of the analysis. Since keyword data in academic articles is often entered inconsistently—due to variations in terminology, formatting, and use of subscripts—direct use of such data could introduce errors[32]. To mitigate these issues, the keyword data was carefully refined. The SCOPUS "analyse search results" function was then employed to assess the number of publications from key countries and research institutions in the field of GAI and security. This evaluation aimed to identify research trends and highlight prominent national and international institutions focused on this area.

The pre-processing steps included the following critical actions:

- Case unification: All keywords were converted to lowercase to eliminate duplicates caused by case variations. For example, "ChatGPT" and "chatgpt" were standardized to a single keyword.
- Removal of special characters: Special characters such as hyphens (-), slashes (/), and spaces were removed to unify keywords with the same meaning. For example, "AI-based" and "AI based" were combined into a single term.
- Elimination of non-essential terms: Irrelevant terms that did not contribute to the analysis objectives were removed to retain only keywords with significant analytical value.
- Consolidating synonymous terms: Terms with similar meanings were standardized to ensure consistency across the dataset.

These refinements were crucial to ensuring the data's consistency and reliability, which in turn enhanced the credibility of the analysis results. Table 1 provides examples of the refined keywords used in the final analysis. NetMiner 4, a Python-based analysis tool, was utilized for the analysis. The analysis included keyword frequency-reverse frequency analysis, keyword centrality analysis, and LDA topic modeling [34]. Before topic modeling, the coherence score was measured using the Cv value, and the number of topics with the six highest scores was determined. Representative papers with high weighting for each topic were selected, offering a comprehensive explanation of the rationale for topic selection and a detailed discussion of its content and significance.

Table 1. Examples of the refined keywords used in the final analysis.

Preprocessing Type	Purpose	Example
Synonym Consolidation	Standardizes multiple words with the same meaning for consistency	"chatGPT," "chatgpt," “Chatgpt,” "CHATGPT" → "ChatGPT"
Case Conversion	Prevents duplication caused by letter case variations	"Ai" and "ai" → "AI"
Stop Words Removal	Removes words that do not add analytical value	"for," "with," "on" → removed

2.2. Research Method

2.2.1 Analysis Tool

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

This study utilized NetMiner 4.4.3 (Cyram Inc., Seongnam, Republic of Korea), a software tool well-regarded for its capabilities in network analysis and visualization. Alongside other widely used tools such as UCINET and Pajek, NetMiner is considered one of the leading applications in the field of social network analysis [34]. The research model for this study is outlined in Figure 3.

The research process began with a comprehensive review of the concept and characteristics of GAI, forming the foundation of the study. Bibliographic data related to GAI and security were then gathered from the SCOPUS database. The collected data underwent a meticulous refinement process, which involved standardizing terms, removing irrelevant or redundant items, and filtering the dataset through a curated dictionary to ensure consistency and relevance.

Following this, the abstracts, author keywords, and titles of the papers were extracted for in-depth analysis. This enabled a thorough exploration of the emerging trends and themes in the field of GAI security.

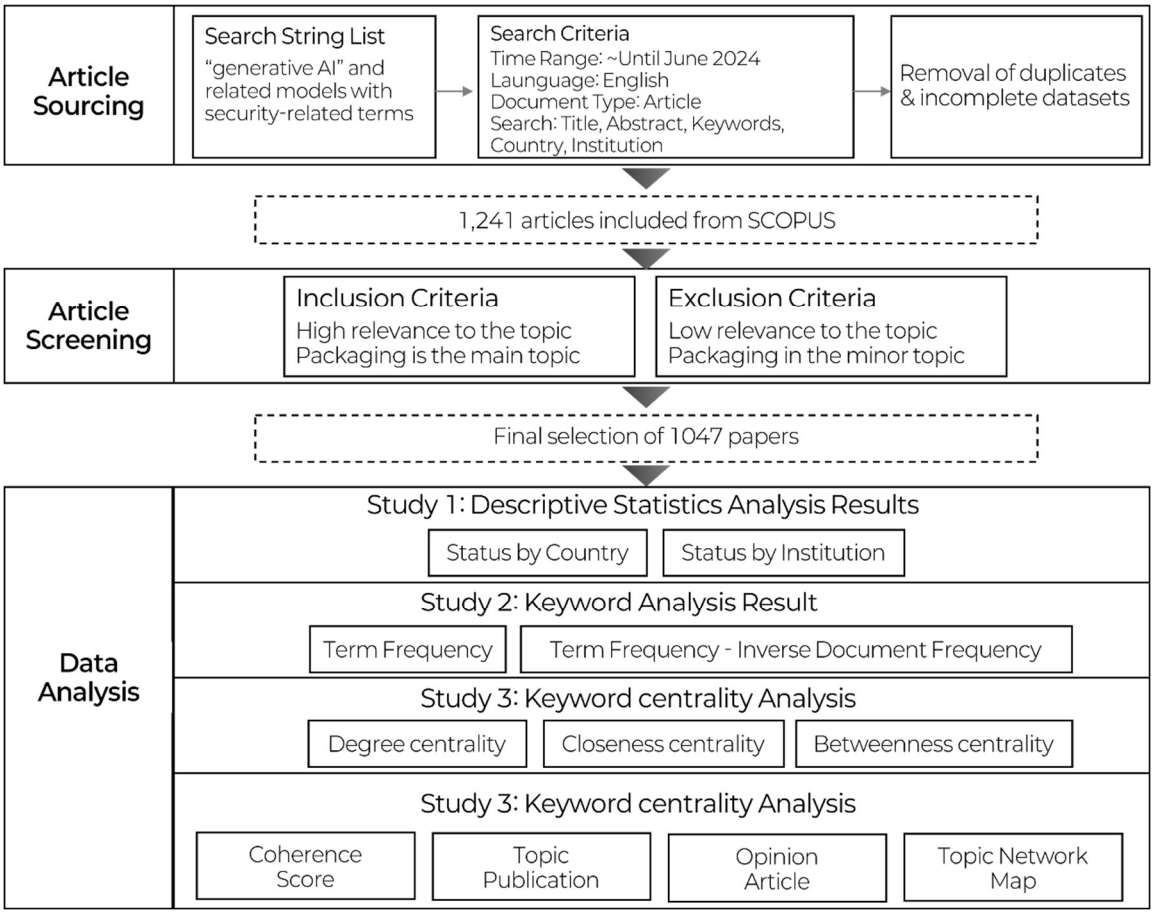

Figure 3. Systematic literature review process.

2.2.2 Term Frequency and Term Frequency-Inverse Document Frequency Analyses

Term frequency-inverse document frequency (TF-IDF) is a widely used statistical technique in text mining, particularly valuable for assessing the importance of specific words within a corpus of documents [35]. TF measures how frequently a particular word appears in a document, helping to identify key terms within that document. In contrast, IDF calculates the frequency of a word across multiple documents by taking the inverse of the number of documents in which the word appears. This metric highlights the uniqueness of a word and its relative significance across the entire document set.

In this study, TF-IDF values were calculated to evaluate both the frequency of words in individual documents and their uniqueness within the broader corpus. A word that appears frequently in a specific document but infrequently in others will have a higher TF-IDF value, indicating its relative importance within that document [36]. This approach provides a nuanced understanding of which terms are central to particular research papers and which are more commonly used across the literature.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

To perform these analyses, term frequencies (TFs) and TF-IDF values were extracted using NetMiner. After cleaning and refining the data, it was transformed into a document-term matrix. This matrix formed the basis for further analyses, including keyword frequency-reverse frequency analysis and LDA topic modeling, which provided deeper insights into the key themes and trends in the field of GAI and security.

2.2.3 Keyword Centrality Analysis

To uncover the underlying knowledge structure within GAI research and its security implications, a keyword centrality analysis was performed using the document-term matrix created earlier. Centrality is a key metric in network analysis that indicates the importance of nodes—represented here by keywords—within a research network. Keywords with high centrality play a central role in shaping discourse and highlight essential areas of focus. This study utilized three commonly used centrality measures in text network analysis: degree centrality, closeness centrality, and betweenness centrality [37].

- **Degree centrality** measures the number of direct connections a keyword has within the network, indicating the frequency with which it co-occurs with other keywords. Keywords with a high degree of centrality often co-occur with other significant terms, underscoring their prominence in the research network [38].
- **Closeness centrality** (also known as proximity centrality) assesses how close a keyword is to all other keywords in the network. It reflects a keyword's ability to connect disparate concepts within the broader research landscape, thereby demonstrating its accessibility and importance [38].
- **Betweenness centrality** evaluates the extent to which a keyword acts as a bridge between different clusters of keywords. Keywords with high betweenness centrality facilitate the flow of information across disconnected groups, maintaining the coherence of the research network [38].

This analysis aimed to identify the most influential keywords in GAI and security, offering valuable insights into the core concepts and their interconnections. Understanding these central keywords provides a better perspective on the evolving themes in the GAI security landscape and highlights areas for future research focus.

2.2.4 LDA Topic Modelling

LDA was employed as the primary technique for topic modeling in this study. LDA is particularly effective when articles tend to converge on certain focal points, making it a preferred method for extracting latent topics from large corpora. It is widely used in research across various domains, including academic papers, news articles, and patents [39].

To derive meaningful insights from LDA, certain parameters must be specified, particularly the alpha and beta values, as well as the number of topics to be generated. The alpha parameter controls the distribution of topics across documents, while the beta parameter governs the distribution of words within topics. These parameters typically range from 0.01 to 0.99, and their values are fine-tuned to optimize the model [40]. Additionally, selecting the appropriate number of topics is a critical challenge in LDA topic modeling, which is addressed by the coherence score. This metric helps determine the optimal number of topics and the relevance of keywords within each topic [41].

The coherence score is used as a proxy for the interpretability and consistency of the topic model. It calculates how frequently words within each topic co-occur in the dataset, assessing the coherence of topics [32]. To determine the optimal number of topics and parameter settings in this study, the alpha and beta values were varied between 0.01 and 0.99, and the number of topics was tested from two to ten, considering the total number of articles in the dataset. The coherence index was calculated for each configuration, enabling comparisons across different topic compositions.

Through this comparison, the study identified core topics—those with a high density of consistent keywords—and variable topics, which showed fluctuations based on model settings. The analysis also involved interpreting the meanings and implications of the subject areas represented by each topic, providing deeper insights into the thematic structure within the field of GAI and security. Figure 4 illustrates the conceptual model of the LDA topic modeling process used in this study.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

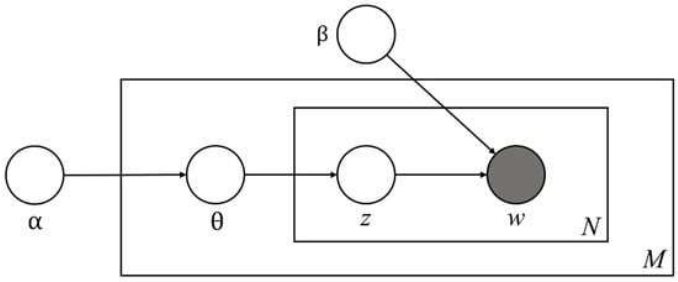

Figure 4. Conceptual model of LDA topic modeling.

One of the most challenging aspects of LDA topic modeling is determining the optimal number of topics [42]. To address this, the coherence score was used as a reference index, specifically employing the Cv metric. The Cv metric has been shown to perform effectively in assessing the interpretability of extracted topics, based on the pairwise similarity between words within a topic [43]. This metric assumes that higher pairwise similarity between words within a topic corresponds to greater coherence.

The Cv metric conceptualizes a keyword as a context vector, representing its co-occurrence frequency with surrounding terms. It calculates the cosine similarity between these vectors and averages the pairwise similarities to produce the coherence score. The formula for calculating the Cv value is provided in equation (1) [44]. This approach yields a more accurate measure of topic coherence by considering the relationships between neighboring words, rather than simply measuring direct word similarity [44].

$$C_v = \frac{1}{N \times K} \sum_{k=1}^{K} \sum_{n=1}^{N} s \cos\left(\overrightarrow{w_{n,k}} \cdot \overrightarrow{w_k^*}\right) \quad (1)$$

3. Results

3.1. Literature Analysis Results

3.1.1. Status by Country

The analysis of GAI research related to security was conducted by examining the distribution of publications across various countries. Figure 5 illustrates the number of documents published by each country or territory, providing a global overview of research activity in this area.

As shown in the figure, the United States leads the research output, with over 250 publications, reflecting its significant focus and investment in GAI and security. China follows closely with nearly 200 publications, demonstrating its substantial contribution to the field. India ranks third, with more than 100 documents, indicating a growing presence in GAI-related research. The United Kingdom, with a notable number of publications, holds the fourth position globally.

Other countries, including Australia, Germany, Canada, and Italy, contribute moderately to the research landscape, with publication numbers ranging between 50 and 100 each. South Korea and Saudi Arabia, though with lower publication volumes, also participate in the global research effort on GAI and security. This distribution underscores the international nature of GAI research, with significant contributions from both Western and Asian countries. It highlights the widespread recognition of GAI's importance in addressing security challenges in various regions.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

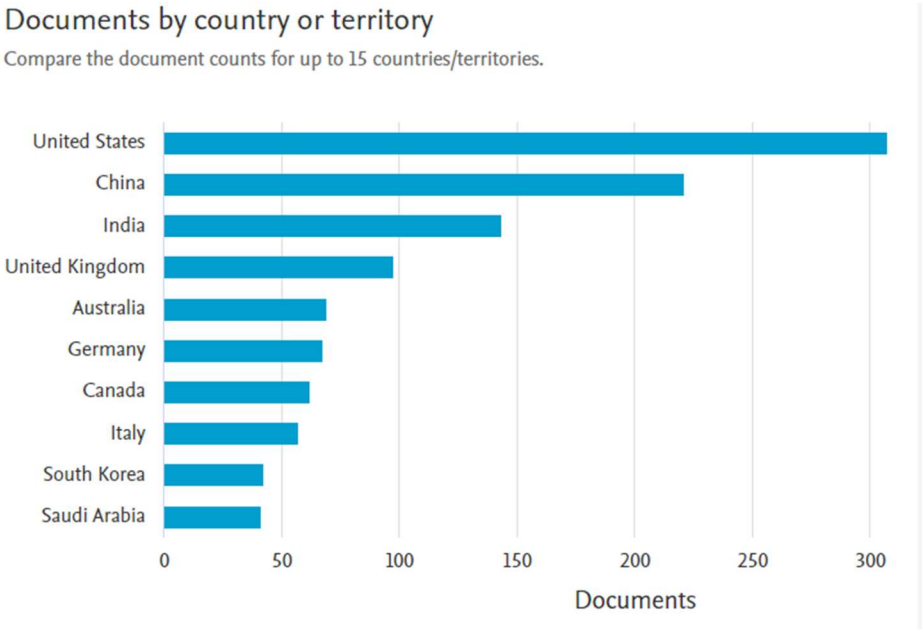

Figure 5. Documents by country or territory.

3.1.2. Status by Institution

Further analysis of GAI research related to security was conducted by examining the contributions of academic and research institutions worldwide. Figure 6 presents the number of documents published by leading institutions, offering insight into the primary contributors in the field.

The Chinese Academy of Sciences emerges as the most prolific institution, with over 20 publications, reflecting China's strong emphasis on advancing GAI research, particularly in the context of security. It is closely followed by the Ministry of Education of the People's Republic of China, underscoring the impact of national education policies and initiatives on driving research output. This demonstrates a coordinated national effort to prioritize GAI research.

The National University of Singapore is also a major contributor, reflecting Singapore's strategic investment in technology and innovation. Tsinghua University and the University of the Chinese Academy of Sciences are other prominent Chinese institutions contributing significantly to the field, further reinforcing China's leadership in GAI research.

Internationally, institutions such as the University of Illinois Urbana-Champaign, University College Dublin, and UNSW Sydney are key players, each contributing between 10 and 15 publications. These institutions reflect the global nature of GAI research, with active participation from leading universities in the United States, Europe, and Australia. Other institutions, including Nanjing University and Universiti Sains Malaysia, also make noteworthy contributions, demonstrating the diversity of actors involved in advancing GAI and security research.

This analysis highlights the dominance of Chinese institutions in GAI security research, while also recognizing the significant contributions of leading institutions from other regions. This global participation is vital for the continued development and application of GAI technologies to address security challenges.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

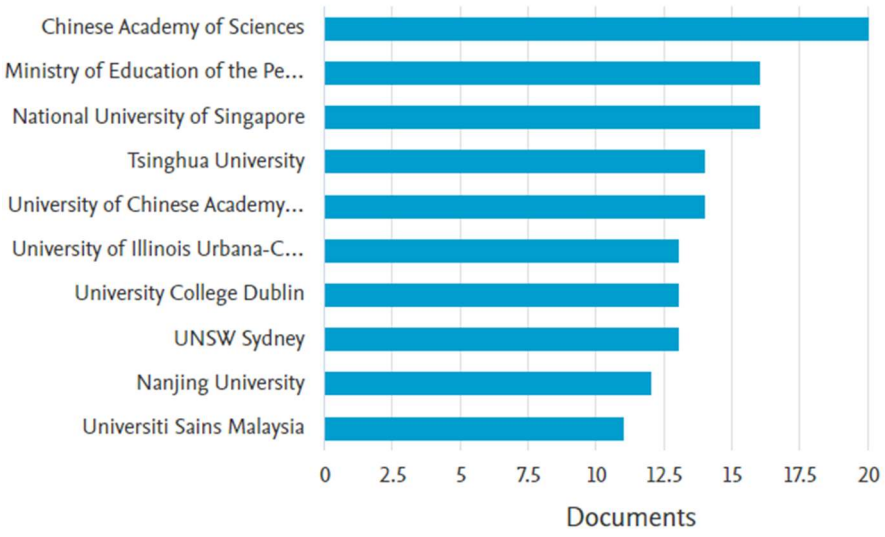

Figure 6. Documents by institution.

3.2. Results of Keyword Frequency (TF-IDF) Analysis

The TF-IDF analysis was conducted to assess the significance of specific words within the corpus of research documents related to GAI and security. This analysis provides insights into the most frequently occurring words, as well as those with unique significance, as indicated by their TF-IDF scores in Table 2.

The term "ChatGPT" emerged as the most frequent word, appearing 2,799 times across the documents. However, its TF-IDF score of 602 is relatively moderate. This suggests that while "ChatGPT" is commonly mentioned, its occurrence is spread across multiple documents, thus diminishing its contribution to the uniqueness of any individual document. Consequently, the term is central to the overall research landscape but may not be the defining focus of specific studies.

In contrast, the term "model" appears 1,503 times and has a high TF-IDF score of 602, highlighting its importance both in terms of frequency and its distinct presence in certain documents. This likely reflects discussions on GAI models and their implications for security. Similarly, "AI" (Artificial Intelligence) is mentioned 1,271 times, with a notable TF-IDF score of 433, emphasizing its pivotal role in research centered on Generative AI.

Other prominent terms include "language," "data," and "intelligence," which rank highly in both frequency and TF-IDF scores. For example, the term "language" appears 958 times, with a TF-IDF score of 530, underscoring its critical importance in discussions of large language models (LLMs), a key component of GAI research. The terms "data" (TF-IDF: 408) and "intelligence" (TF-IDF: 367) are also central to analyses of information systems and intelligence frameworks—crucial aspects in the application of GAI to security challenges.

Terms such as "technology," "tool," and "application" also exhibit moderate to high TF-IDF scores, signaling their relevance in discussions about the practical implementation of GAI technologies across various domains. These words reflect the growing interest in how GAI can be operationalized for specific use cases, particularly in areas like security and technology management.

Additionally, terms such as "challenge" (TF-IDF: 307) and "analysis" (TF-IDF: 297) suggest ongoing discussions about the challenges and methodological considerations in integrating GAI into existing security frameworks. These words highlight the practical obstacles and strategic analysis required for the effective deployment of GAI technologies in sensitive environments.

Finally, terms related to "security" (TF-IDF: 265) and "information" (TF-IDF: 250) are central to the discourse, reflecting the research focus on securing data and systems in the context of GAI applications. Although these terms appear less frequently, their high TF-IDF scores indicate their critical importance in specific documents, especially those addressing core research questions on the security implications of GAI.

Overall, this analysis demonstrates that while certain terms, such as "ChatGPT" and "AI," are frequently mentioned across the document set, the TF-IDF scores provide a deeper understanding of

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

the nuanced importance of other terms in specific contexts. These insights help to clarify the key concepts driving research in GAI and security, revealing both the breadth and depth of ongoing discussions in the field.

Table 2. Results of keyword frequency (TF-IDF) analysis.

Rank	Word	Frequency	Word	TF-IDF	Rank	Word	Frequency	Word	TF-IDF
1	ChatGPT	2,799.00	model	602	11	education	499	analysis	297
2	model	1,503.00	language	530	12	intelligence	472	use	294
3	AI	1,271.00	AI	433	13	system	466	security	265
4	language	958	datum	408	14	result	464	development	256
5	datum	768	intelligence	367	15	use	454	information	250
6	LLM	735	technology	360	16	challenge	453	system	239
7	technology	684	result	348	17	analysis	449	finding	237
8	tool	611	application	334	18	security	432	method	236
9	application	552	tool	331	19	user	420	task	232
10	student	544	challenge	307	20	information	403	education	231

3.3. Results of Keyword Centrality Analysis

A keyword centrality analysis was conducted to identify the most influential terms in research related to GAI and security. This analysis utilized three centrality measures: degree centrality, closeness centrality, and betweenness centrality, as summarized in Table 3.

Degree Centrality indicates the number of direct connections a keyword has within the network, highlighting its overall importance. The term "ChatGPT" emerged with the highest degree centrality score (0.93878), making it the most connected and central keyword in the network. It is followed closely by "AI" (0.87755) and "model" (0.81633), both of which also exhibit strong connectivity, signifying their prominent presence in the research corpus. Other keywords, including "result" (0.81633), "datum" (0.79592), and "LLM" (0.77551), also display high degree centrality, reflecting their frequent co-occurrence with other terms and their integral role in the discourse surrounding GAI and security.

Closeness Centrality measures how easily a keyword can connect to all other keywords in the network, providing insight into its accessibility and influence. The term "model" ranks highest in closeness centrality (0.67606), indicating that it is central to the network and can be quickly connected to other keywords. Other highly ranked terms include "language" (0.5931) and "LLM" (0.28357), which play key roles in bridging various concepts within the GAI and security discourse. Interestingly, despite its high degree centrality, "ChatGPT" has a relatively lower closeness centrality (0.25657), suggesting that although it is frequently mentioned, its connections are less direct or influential in linking other topics within the network. This finding implies that while "ChatGPT" is an essential term, it does not serve as a primary hub for reaching other keywords in the network.

Betweenness Centrality assesses how effectively a keyword acts as a bridge, facilitating connections between different groups of keywords. "ChatGPT" once again leads with the highest betweenness centrality (0.02865), underscoring its role as a key connector in the research landscape. "AI" (0.02539) and "result" (0.02225) also play vital bridging roles, helping to link different research topics and clusters. Other terms, such as "datum" (0.01983), "model" (0.01907), and "technology"

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

(0.01786), exhibit strong betweenness centrality, indicating their importance in connecting different research areas within the broader GAI and security domain.

In summary, this analysis highlights the varying roles of "ChatGPT", "AI", and "model" within the research network. While "ChatGPT" and "AI" are highly connected and serve as important bridges within the research landscape, "model" stands out in terms of accessibility, playing a crucial role in directly linking different keywords across the landscape. The prominence of terms like "technology", "LLM", and "analysis" across all centrality measures emphasizes their significant contributions to discussions of GAI and security, further illustrating the complexity and interconnectedness of research topics in this field.

Table 3. Results of keyword centrality analysis.

Rank	Word	Degree Centrality	Word	Closeness Centrality	Word	Betweenness Centrality
1	ChatGPT	0.93878	model	0.67606	ChatGPT	0.02865
2	AI	0.87755	language	0.5931	AI	0.02539
3	model	0.81633	LLM	0.28357	result	0.02225
4	result	0.81633	ChatGPT	0.25657	datum	0.01983
5	datum	0.79592	AI	0.1293	model	0.01907
6	LLM	0.77551	use	0.07502	technology	0.01786
7	technology	0.77551	tool	0.06478	LLM	0.01683
8	analysis	0.7551	application	0.05542	analysis	0.01675
9	development	0.71429	technology	0.0525	development	0.015
10	application	0.69388	learning	0.03999	challenge	0.01393
11	tool	0.67347	intelligence	0.03711	application	0.01301
12	education	0.67347	capability	0.035	student	0.01177
13	security	0.67347	performance	0.03384	education	0.01171
14	challenge	0.67347	education	0.02849	security	0.01165
15	capability	0.63265	impact	0.02775	capability	0.01117
16	student	0.63265	generation	0.0238	user	0.01016
17	system	0.63265	response	0.02245	tool	0.01007
18	performance	0.61224	student	0.01998	information	0.00916
19	field	0.61224	datum	0.0199	response	0.00914
20	method	0.61224	system	0.01972	performance	0.00883

3.4. LDA Topic Modeling

3.4.1. Coherence Score Measurement Results

To determine the optimal number of topics for LDA topic modeling, the parameters were set following the methodology outlined by W. Zhao et al. [00]. Specifically, α was set to 0.1, β to 0.01, and the number of iterations to 1,000. The number of topics was gradually increased from two to fifteen to evaluate coherence scores, a critical metric for assessing the interpretability of the topics generated by the model. The six topics with the highest coherence score of 0.523 were identified as the most meaningful and interpretable, as shown in Figure 7.

Once the optimal number of topics was established, the top ten most frequent keywords for each of the six topics were identified and summarized in Table 3. To ensure the accuracy and representativeness of the topic labeling, a thorough review of the keywords and relevant abstracts from the research bibliographies was conducted. This review involved collaboration with experts, including researchers with over five years of experience in IT institutions and professors specializing in text mining. Properly naming the topics was crucial for providing clear and concise representations of the central themes emerging from the LDA model. Through a meticulous examination of the keywords and a comprehensive review of the context provided by the abstracts, the most appropriate

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

topic names were assigned, ensuring they accurately reflected the underlying research areas related to GAI and security.

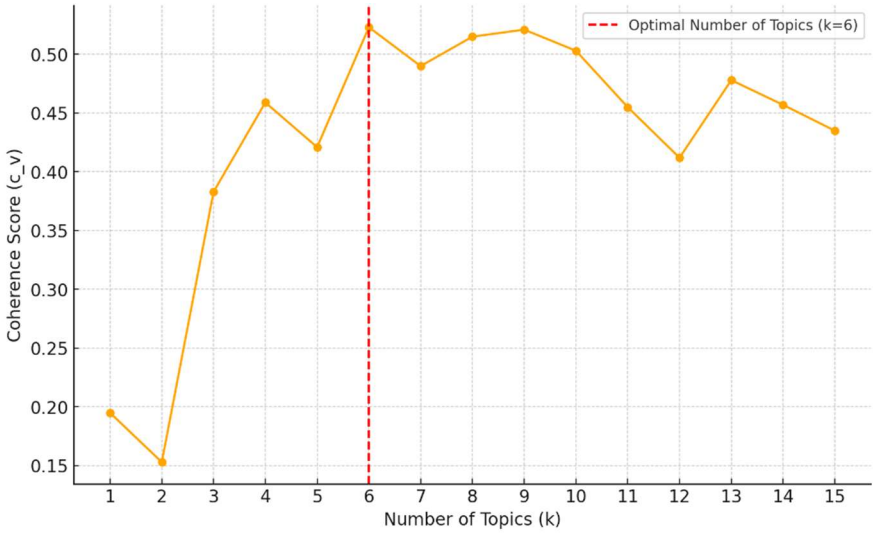

Figure 7. Coherence scores by number of topics.

The analysis revealed six distinct topics within the research on GAI and its security applications, as illustrated in Figure 8. Among these, the topic "AI in Educational Security" emerged as the most prominent, representing 40% of the papers analyzed. This suggests a significant focus on the integration of AI in educational environments, particularly with respect to security concerns. In contrast, the topic "Healthcare Security with AI" was the least represented, comprising only 5% of the topics. This reflects a specialized but growing area of research within the broader field of GAI and security. The distribution of topics highlights the diverse applications of GAI, with a strong emphasis on its role in education, while acknowledging the increasing importance of security in healthcare, albeit on a smaller scale.

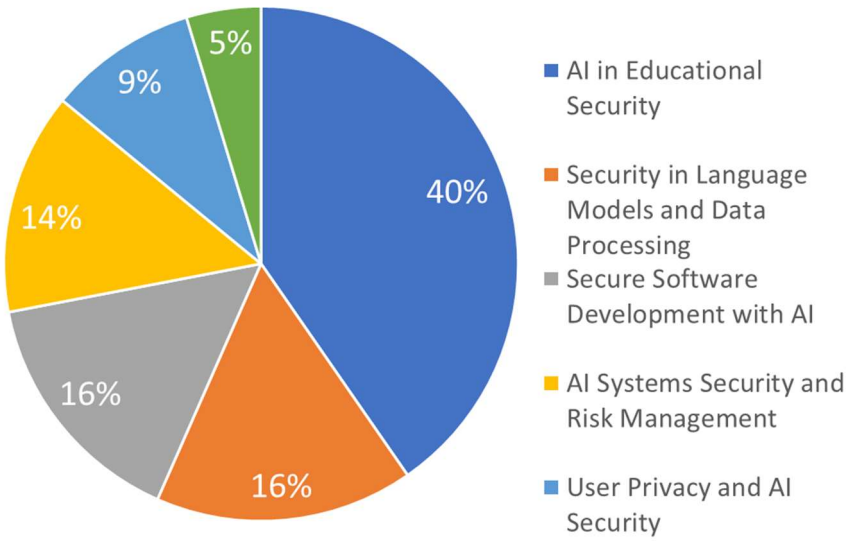

Figure 8. Percentage distribution by topics.

3.4.2. Topic Classification Results

The topic modeling analysis identified six distinct themes within the research on GAI and its applications to security. Each theme was analyzed based on its primary keywords and contextual relevance, providing a comprehensive understanding of the research areas within the domain. These results are presented in Table 4.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

Table 4. Main keywords and topic names per topic.

Topic 1 (422 Papers)			**Topic 2 (170 Papers)**			**Topic 3 (161 Papers)**		
Rank	Keyword	Prob	Rank	Keyword	Prob	Rank	Keyword	Prob
1	ChatGPT	0.062	1	model	0.073	1	ChatGPT	0.044
2	AI	0.042	2	LLM	0.054	2	code	0.034
3	education	0.02	3	language	0.046	3	model	0.021
4	student	0.02	4	datum	0.022	4	result	0.018
5	technology	0.017	5	ChatGPT	0.019	5	question	0.017
6	tool	0.017	6	text	0.016	6	tool	0.016
7	use	0.014	7	task	0.014	7	language	0.015
8	intelligence	0.013	8	dataset	0.012	8	software	0.015
9	challenge	0.012	9	method	0.012	9	task	0.013
10	concern	0.011	10	performance	0.011	10	Detection	0.013
AI and Security in Education (40%)			Security in Language Models and Data Processing (16%)			Secure Software Development with AI (16%)		
Topic 4 (146 Papers)			**Topic 5 (98 Papers)**			**Topic 6 (49 Papers)**		
Rank	Keyword	Prob	Rank	Keyword	Prob	Rank	Keyword	Prob
1	security	0.025	1	ChatGPT	0.051	1	healthcare	0.03
2	system	0.02	2	user	0.028	2	patient	0.02
3	datum	0.018	3	factor	0.022	3	health	0.019
4	AI	0.016	4	analysis	0.012	4	care	0.016
5	technology	0.016	5	attitude	0.012	5	treatment	0.012
6	service	0.015	6	medium	0.012	6	response	0.012
7	development	0.015	7	intention	0.011	7	medicine	0.011
8	application	0.014	8	perception	0.01	8	datum	0.008
9	ChatGPT	0.013	9	result	0.01	9	ChatGPT	0.008
10	model	0.013	10	model	0.01	10	accuracy	0.008
AI Systems Security and Risk Management (14%)			User Privacy and AI Security (9%)			Healthcare Security with AI (5%)		

The initial theme, "AI and Security in Education," which comprises 40% of the total, is defined by keywords such as "ChatGPT," "education," "student," and "technology." This theme addresses security concerns arising from the use of AI in educational settings. Research within this theme focuses on safeguarding student data, enhancing the security of online learning platforms, and mitigating potential security threats in AI-integrated learning environments. A representative study by Torres et al. (2023) examined the influence of generative AI on higher education, with a particular emphasis on ethical considerations and maintaining academic integrity. The study highlighted that, while tools like ChatGPT offer new possibilities for teaching and learning, they also present challenges related to information reliability, privacy, and security. Proposed security measures included encrypting student data, implementing rigorous access controls, and providing security awareness training to maintain academic integrity in AI-driven learning environments [45].

The second theme, "Security in Language Models and Data Processing," focuses on security issues related to language models (LLMs) and data processing techniques. Keywords such as "model," "LLM," "language," and "data" point to research on dataset integrity, vulnerabilities within language models, and privacy concerns. A representative study by Alawida et al. (2023) examined the architecture, training data, and associated security risks, including potential cyberattacks. The study proposed countermeasures such as data anonymization, model hardening, and rigorous data

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

validation to mitigate these risks, while also emphasizing the importance of responsible AI use and the need for legal safeguards to prevent misuse [46].

The third theme, "Secure Software Development with AI," addresses the security challenges inherent in developing software that utilizes AI technologies. Keywords like "ChatGPT," "code," "model," and "software" reflect research on AI's role in coding, the security of AI-generated code, and methods for enhancing security throughout the software development lifecycle. A study by Horne et al. (2023) discussed the potential risks associated with AI code assistants, such as GitHub Copilot, particularly the possibility of introducing unintended vulnerabilities. The paper proposed rigorous human review and secure software development practices as ways to mitigate these risks [47].

The fourth theme, "Security and Risk Management of AI Systems" (14%), focuses on the security and risk management of AI systems. Keywords like "security," "system," "AI," and "technology" indicate research into the inherent vulnerabilities of AI systems and the necessary security considerations during their development. Charfeddine et al. (2024) investigated the dual role of ChatGPT in cybersecurity, analyzing both its benefits and risks. The study recommended comprehensive threat analysis, the implementation of misuse prevention mechanisms, and the establishment of secure development environments as essential measures for managing the risks associated with AI systems [48].

The fifth theme, "User Privacy and AI Security" (9%), explores the security and privacy challenges that arise from user interactions with AI systems, as evidenced by keywords such as "user," "factor," "analysis," and "perception." In a comprehensive examination of public opinion regarding the privacy implications of ChatGPT, Tang and Bashir (2024) identified significant concerns about data privacy and security. The study proposed countermeasures such as minimizing data collection, implementing rigorous data management practices, and incorporating feedback loops to prevent security threats and avoid misinterpreting user intent [49].

The final theme, "Healthcare Security with AI" (5%), addresses the security challenges posed by AI technologies in the healthcare sector, particularly regarding the protection of patient data and the security of AI-powered diagnostic systems. A study by Chen, Walter, and Wei (2024) investigated the deployment of ChatGPT-like AI tools in healthcare, emphasizing the need to safeguard privacy, ensure confidentiality, and comply with regulatory norms. The study recommended encrypting patient data, conducting periodic security assessments, and validating AI models to strengthen healthcare systems against potential security breaches [50].

3.4.3. Weight by topics

Figure 9 presents a network visualization that provides a comprehensive overview of the top 100 keywords identified through topic modeling. This visual representation illustrates the complex interrelationships between various topics and their associated terms in the context of GAI and its security applications. The network highlights central themes within GAI security, revealing common keywords that link multiple topics and identifying broader thematic overlaps across

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

different domains.

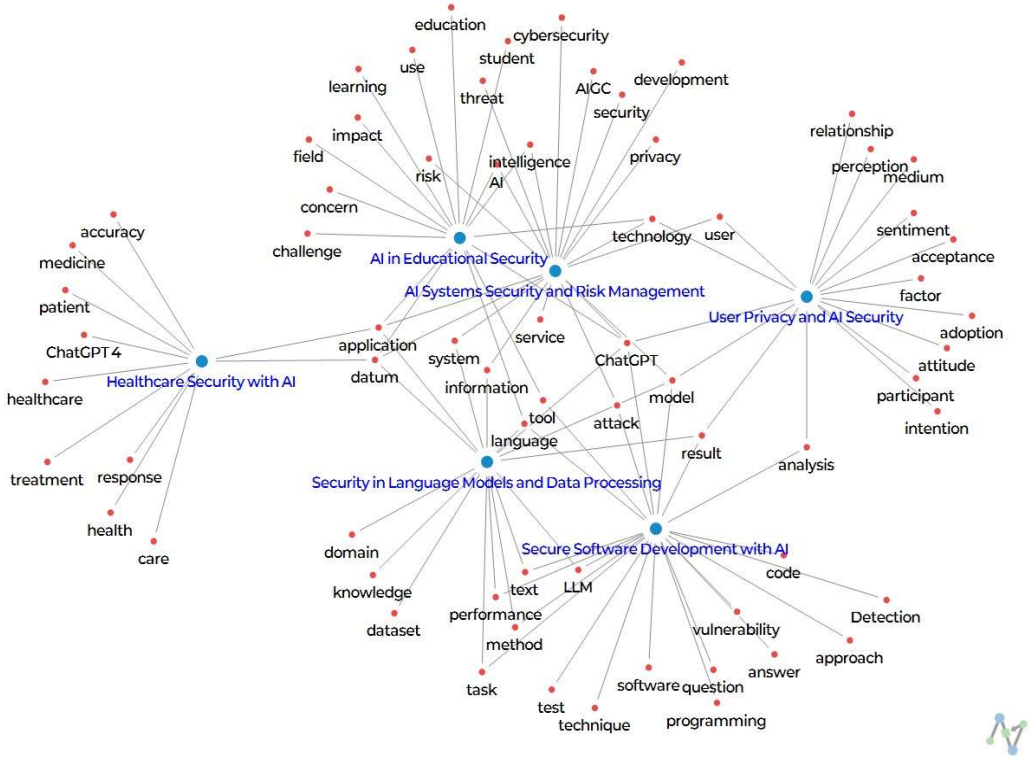

Figure 9. Network visualizations for keywords by topics.

The keyword "ChatGPT" stands out as particularly significant, appearing frequently in contexts related to healthcare security with AI, user privacy and AI security, and AI systems security and risk management. This widespread presence underscores the multifaceted role of ChatGPT, an advanced AI language model, across various sectors. For example, in healthcare, ChatGPT is being explored as a tool to enhance communication between patients and healthcare providers, with the potential to improve diagnostic accuracy and patient engagement. However, its use raises serious concerns about data privacy, security vulnerabilities, and the potential for misuse, especially in sensitive areas like healthcare and cybersecurity. The repeated appearance of "ChatGPT" in these contexts emphasizes the need for robust, tailored security measures to address both its benefits and associated risks.

Similarly, the keyword "security" emerges as a central theme that links multiple topics, such as AI Systems Security and Risk Management, User Privacy and AI Security, and Healthcare Security with AI. This highlights the universal importance of security in the development, deployment, and operation of AI systems. Protecting patient data in healthcare, safeguarding user privacy in AI interactions, and securing AI systems against cyber threats are all critical areas where security is paramount. The frequent mention of "security" underscores the essential role it plays in ensuring the resilience and trustworthiness of AI technologies across various domains.

The term "data" also appears consistently across a number of topics, including Security in Language Models and Data Processing, AI Systems Security and Risk Management, and Healthcare Security with AI. Data is fundamental to AI, serving as the foundation on which AI systems are built and operate. Ensuring the security of data, especially in sensitive sectors like healthcare, is critical. The recurrence of "data" across different topics highlights the urgent need for robust data protection measures, such as encryption, secure storage, and rigorous access controls, to protect sensitive information throughout its lifecycle in AI applications.

The term "privacy" is closely associated with several topics, including User Privacy and AI Security, Healthcare Security with AI, and AI Systems Security and Risk Management. As AI systems increasingly handle personal and sensitive data, privacy concerns have become a major issue. The repeated mention of "privacy" reflects the ongoing challenge of balancing the advantages of AI with the need to protect individuals' privacy and comply with regulatory standards. It underscores the necessity for implementing technologies and practices that safeguard the confidentiality of user data.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

Lastly, the keyword "system" frequently appears in discussions related to AI Systems Security and Risk Management and Secure Software Development with AI, reflecting broad concerns about the security and integrity of AI systems. The term "system" plays a central role, whether it refers to securing the entire AI infrastructure or incorporating security measures early in the software development process. The emphasis on "system" highlights the importance of taking a comprehensive approach to AI security, considering the entire ecosystem and implementing safeguards against potential threats.

In conclusion, the common keywords identified in this network visualization—such as "ChatGPT," "security," "data," "privacy," and "system"—serve as critical connectors between various topics, illustrating broader themes that are relevant across multiple domains. These keywords underscore the interconnectedness of AI security issues and highlight the need for a holistic security strategy that addresses concerns ranging from data integrity to user privacy, while also ensuring the overall security of AI systems. Understanding the significance of these keywords provides valuable insight into the challenges of AI security, guiding both research and practice toward more integrated and effective solutions. The visualization not only maps out specific areas of focus within AI security but also emphasizes critical cross-cutting issues that require attention, contributing to the development of robust and resilient AI technologies capable of safely supporting a wide range of applications.

4. Discussion

This study presents a thorough analysis of trends in GAI research, with a specific focus on security. Descriptive statistics reveal that the United States, China, and India are the leading countries in GAI and security-related research, underscoring the global recognition of the importance of AI technologies in addressing security challenges. Prominent research institutions, such as the Chinese Academy of Sciences and the National University of Singapore, also play a pivotal role, highlighting the concentrated efforts from both Western and Asian regions in advancing GAI security research.

The results from TF-IDF and keyword centrality analyses provide valuable insights into the key terms shaping the discourse around GAI and security. While "ChatGPT," "AI," and "model" emerged as central terms, their varying TF-IDF and centrality scores suggest different roles within the research context. For instance, although "ChatGPT" is frequently mentioned, it does not hold the same centrality in linking core topics as terms like "model" and "AI." This distinction emphasizes the complexity of GAI applications, which span diverse industries—from education to healthcare—with distinct security implications.

LDA topic modeling identified six major themes within the GAI and security landscape: *AI and Security in Education*, *Security in Language Models and Data Processing*, *Secure Software Development with AI*, *Security and Risk Management of AI Systems*, *User Privacy and AI Security*, and *Healthcare Security with AI*. The prominence of AI in education highlights the growing concerns over protecting student data and preserving academic integrity in AI-driven environments. At the same time, the focus on language models and data processing underscores the critical need to secure the foundational datasets and algorithms powering GAI systems. The relatively smaller share of research on healthcare security reflects an emerging but increasingly vital area of concern, as AI is integrated into sensitive sectors such as healthcare.

A recurring theme across all these topics is the emphasis on establishing robust security frameworks that address both technical vulnerabilities and ethical challenges. The frequent appearance of keywords such as "privacy," "data," and "system" across various topics points to the necessity of comprehensive security strategies that cover data protection, system integrity, and user privacy. This study highlights the interconnected nature of AI and security concerns, emphasizing the need for interdisciplinary research and policy development to address the full spectrum of risks associated with GAI technologies.

5. Conclusion

This study examines the security implications of GAI by analyzing 1,047 academic articles sourced from the SCOPUS database. Using scientometric methods, including TF-IDF analysis, keyword centrality analysis, and LDA topic modeling, we identified key trends, themes, and security challenges associated with GAI technologies, particularly in relation to systems like ChatGPT.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

The findings reveal that GAI security research is a global effort, with significant contributions from countries such as the United States, China, and India. Leading institutions, including the Chinese Academy of Sciences and the National University of Singapore, are at the forefront of advancing this field. Our analysis also highlighted the centrality of terms like "ChatGPT," "AI," and "model" in current research, reflecting the prominence of large language models and AI systems in discussions of security.

The six themes identified through LDA topic modeling—ranging from AI in education to healthcare security—illustrate the diverse applications and associated risks of GAI. While research on AI in education and language models dominates the landscape, emerging areas like healthcare security are gaining attention and warrant further investigation to fully understand and mitigate the risks GAI poses in sensitive sectors.

Despite the valuable insights this study provides, it has several limitations. First, the dataset is confined to articles published in SCOPUS, which, although comprehensive, may not encompass all relevant research, particularly non-English literature or studies from less prominent journals. This may introduce a regional or perspective-based bias in the findings. Additionally, the study's data collection period (June 2022 to June 2024) captures recent developments but excludes earlier foundational research on AI security. Given the rapid pace of GAI advancements, the conclusions drawn may become outdated as new technologies and challenges emerge.

Furthermore, while LDA topic modeling offered a structured view of research trends, this method relies on statistical probabilities, which may not fully capture the depth and nuances of each article's content. A more in-depth qualitative analysis could complement these findings by providing richer insights into the contextual details of GAI security challenges.

In conclusion, as GAI technologies continue to evolve, developing comprehensive security frameworks is essential to address both technical vulnerabilities and broader ethical and societal challenges. Future research should focus on interdisciplinary approaches that integrate technical, legal, and ethical perspectives to ensure the safe and responsible deployment of GAI technologies across industries. Expanding the scope of analysis by incorporating diverse data sources and employing hybrid quantitative-qualitative methods would provide a more holistic understanding of GAI security trends.

Author Contributions: Methodology, M.N and Y,S.; Project administration, J.L.; Resources, B.K.; Software, K.J.; Supervision, M.C.; Writing—Original draft, K.J.; Writing—Review and editing, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Policy Expert Development and Support Program through the Ministry of Science and ICT of the Korean government, grant number S2022A066700001. The funder was not involved in the study design, data collection, analysis, or interpretation, nor in the writing of this article or the decision to submit it for publication.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to express their gratitude to all participants who generously contributed their time and effort to this research.

Conflicts of Interest: The authors declare no conflicts of interest related to this research.

References

1. Lund, B.D.; Wang, T. Chatting about ChatGPT: How May AI and GPT Impact Academia and Libraries? Library Hi Tech News 2023, 40, 26–29.
2. Zhang, S.; Liau, Z.Q.G.; Tan, K.L.M.; Chua, W.L. Evaluating the Accuracy and Relevance of ChatGPT Responses to Frequently Asked Questions Regarding Total Knee Replacement. Knee Surg. Relat. Res. 2024, 36, 15.
3. FirstPageSage. Top Generative AI Chatbots by Market Share – December 2024. Available online: https://www.firstpagesage.com (accessed on August 2024).
4. Aljanabi, M. ChatGPT: Future Directions and Open Possibilities. Mesopotamian J. Cybersecur. 2023, 2023, 16–17.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

5. Gill, S.S.; Kaur, R. ChatGPT: Vision and Challenges. Internet Things Cyber-Phys. Syst. 2023, 3, 262–271.
6. Javaid, M.; Haleem, A.; Singh, R.P. ChatGPT for Healthcare Services: An Emerging Stage for an Innovative Perspective. BenchCouncil Trans. Benchmarks Stand. Eval. 2023, 3, 100105.
7. George, A.S.; George, A.H. A Review of ChatGPT AI's Impact on Several Business Sectors. Partn. Univ. Int. Innov. J. 2023, 1, 9–23.
8. Li, L.; Ma, Z.; Fan, L.; Lee, S.; Yu, H.; Hemphill, L. ChatGPT in Education: A Discourse Analysis of Worries and Concerns on Social Media. arXiv 2023, arXiv:2305.02201.
9. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. Educ. Sci. 2023, 13, 410.
10. Rathore, B. Future of AI & Generation Alpha: ChatGPT beyond Boundaries. Eduzone Int. Peer Rev./Ref. Multidiscip. J. 2023, 12, 63–68.
11. Kaddour, J.; Harris, J.; Mozes, M.; Bradley, H.; Raileanu, R.; McHardy, R. Challenges and Applications of Large Language Models. arXiv 2023, arXiv:2307.10169.
12. Buchholz, K. Threads Shoots Past One Million User Mark at Lightning Speed. Statista, January 24, 2023. Available online: https://www.statista.com/chart/29174/time-to-one-million-users (accessed on 5 January 2023).
13. Ahmad, N.; Murugesan, S.; Kshetri, N. Generative Artificial Intelligence and the Education Sector. Computer 2023, 56, 72–76.
14. Javaid, M.; Haleem, A.; Singh, R.P. A Study on ChatGPT for Industry 4.0: Background, Potentials, Challenges, and Eventualities. J. Econ. Technol. 2023, 1, 127–143.
15. Alawida, M.; Abu Shawar, B.; Abiodun, O.I.; Mehmood, A.; Omolara, A.E.; Al Hwaitat, A.K. Unveiling the Dark Side of ChatGPT: Exploring Cyberattacks and Enhancing User Awareness. Information 2024, 15, 27.
16. Rudolph, J.; Tan, S.; Tan, S. ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education? J. Appl. Learn. Teach. 2023, 6, 342–363.
17. Baidoo-Anu, D.; Owusu Ansah, L. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. J. AI 2023, 7, 52–62.
18. Wang, J. et al. Is ChatGPT a Good NLG Evaluator? A Preliminary Study. arXiv 2023, arXiv:2303.04048.
19. Srivastava, M. A Day in the Life of ChatGPT as an Academic Reviewer: Investigating the Potential of Large Language Model for Scientific Literature Review. Preprint 2023.
20. Kim, S.-D. Trends and Perspectives of mHealth in Obesity Control. Appl. Sci. 2025, 15, 74.
21. Wu, T. et al. A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. IEEE/CAA J. Autom. Sin. 2023, 10, 1122–1136.
22. Hosseini, M.; Horbach, S.P.J.M. Fighting Reviewer Fatigue or Amplifying Bias? Considerations and Recommendations for Use of ChatGPT and Other Large Language Models in Scholarly Peer Review. Res. Integr. Peer Rev. 2023, 8, 4.
23. Cotton, D.R.E.; Cotton, P.A.; Shipway, J.R. Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT. Innov. Educ. Teach. Int. 2024, 61, 228–239.
24. Younis, H.A. et al. A Systematic Review and Meta-Analysis of Artificial Intelligence Tools in Medicine and Healthcare: Applications, Considerations, Limitations, Motivation and Challenges. Diagnostics 2024, 14, 109.
25. Zhou, K.; Wang, J.; Ashuri, B.; Chen, J. Discovering the Research Topics on Construction Safety and Health Using Semi-Supervised Topic Modeling. Buildings 2023, 13, 1169.
26. Sun, L.; Yin, Y. Discovering Themes and Trends in Transportation Research Using Topic Modeling. Transp. Res. Part C Emerg. Technol. 2017, 77, 49–66.
27. Nie, B.; Sun, S. Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research. Appl. Sci. 2017, 7, 401.
28. Hall, B. Text Mining and Data Visualization: Exploring Cultural Formations and Structural Changes in Fifty Years of Eighteenth-Century Poetry Criticism (1967–2018). Data Vis. Enlight. Lit. Cult. 2021, 153–195.
29. Yau, S.C. et al. Detection of Topic on Health News in Twitter Data. Emerg. Adv. Integr. Technol. 2021, 2, 23–29.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

30. Ma, B.; Liu, S.; Pei, F.; Su, Z.; Yu, J.; Hao, C.; Li, Q.; Jiang, L.; Zhang, J.; Gan, Z. Development of Hydrogen Energy Storage Industry and Research Progress of Hydrogen Production Technology. In Proceedings of the 2021 IEEE 4th International Electrical and Energy Conference, CIEEC 2021, Wuhan, China, 28–30 May 2021.
31. Choi, C.; Lee, J.; Machado, J.; Kim, G. Big-Data-Based Text Mining and Social Network Analysis of Landscape Response to Future Environmental Change. Land 2022, 11, 2183.
32. Kim, J.; Han, S.; Lee, H.; Koo, B.; Nam, M.; Jang, K.; Lee, J.; Chung, M. Trend Research on Maritime Autonomous Surface Ships (MASSs) Based on Shipboard Electronics: Focusing on Text Mining and Network Analysis. Electronics 2024, 13, 1902.
33. Baas, J.; Schotten, M.; Plume, A.; Côté, G.; Karimi, R. Scopus as a Curated, High-Quality Bibliometric Data Source for Academic Research in Quantitative Science Studies. Quant. Sci. Stud. 2020, 1, 377–386.
34. Park, S.; Park, J. Identifying the Knowledge Structure and Trends of Outreach in Public Health Care: A Text Network Analysis and Topic Modeling. Int. J. Environ. Res. Public Health 2021, 18, 9309.
35. Aizawa, A. An Information-Theoretic Perspective of TF-IDF Measures. Inf. Process. Manag. 2003, 39, 45–65.
36. Xiang, L. Application of an Improved TF-IDF Method in Literary Text Classification. Adv. Multimed. 2022, 2022, 9285324.
37. Park, C.S. Using Text Network Analysis for Analyzing Academic Papers in Nursing. Perspect. Nurs. Sci. 2019, 16, 12–24.
38. Zhang, J.; Luo, Y. Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. In Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017), Bangkok, Thailand, 26–27 March 2017; pp. 300–303.
39. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X. Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. Multimed. Tools Appl. 2019, 78, 15169–15211.
40. Vorontsov, K.; Potapenko, A.; Plavin, A. Additive Regularization of Topic Models for Topic Selection and Sparse Factorization. In Proceedings of the Statistical Learning and Data Sciences: Third International Symposium, SLDS 2015, Egham, UK, 20–23 April 2015; pp. 193–202.
41. O'Callaghan, D.; Greene, D.; Carthy, J.; Cunningham, P. An Analysis of the Coherence of Descriptors in Topic Modeling. Expert Syst. Appl. 2015, 42, 5645–5657.
42. Durbin, J.; Watson, G.S. Testing for Serial Correlation in Least Squares Regression. III. Biometrika 1971, 58, 1–9.
43. Seo, Y.; Kim, K.; Kim, J.-S. Trends of Nursing Research on Accidental Falls: A Topic Modeling Analysis. Int. J. Environ. Res. Public Health 2021, 18, 3963.
44. Röder, M.; Both, A.; Hinneburg, A. Exploring the Space of Topic Coherence Measures. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2–6 February 2015; pp. 399–408.
45. Gallent Torres, C.; Zapata-González, A.; Ortego-Hernando, J.L. The Impact of Generative Artificial Intelligence in Higher Education: A Focus on Ethics and Academic Integrity. RELIEVE 2023, 29, 1–19.
46. Alawida, M. et al. A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity. Information 2023, 14, 462.
47. Horne, D. PwnPilot: Reflections on Trusting Trust in the Age of Large Language Models and AI Code Assistants. In Proceedings of the 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE), IEEE, 2023.
48. Charfeddine, M. et al. ChatGPT's Security Risks and Benefits: Offensive and Defensive Use-Cases, Mitigation Measures, and Future Implications. IEEE Access 2024.
49. Tang, L.; Bashir, M. A Comprehensive Analysis of Public Sentiment Towards ChatGPT's Privacy Implications. In Proceedings of the International Conference on Human-Computer Interaction, Cham, Springer Nature Switzerland, 2024.
50. Chen, C.W.; Walter, P.; Wei, J.C.-C. Using ChatGPT-Like Solutions to Bridge the Communication Gap Between Patients with Rheumatoid Arthritis and Health Care Professionals. JMIR Med. Educ. 2024, 10, e48989.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 February 2025 doi:10.20944/preprints202502.1011.v1

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.