Preprints.org

Article

Using KeyGraph and ChatGPT to Detect and Track Topics Related to AI Ethics in Media Outlets

Wei-Hsuan Li and Hsin-Chun Yu *

Article

Using KeyGraph and ChatGPT to Detect and Track Topics Related to AI Ethics in Media Outlets

Wei-Hsuan Li and Hsin-Chun Yu *

Department of Information Management, Tunghai University, Taichung 40704, Taiwan

* Correspondence: hsyu@thu.edu.tw

Abstract

This study analyzes the semantic dynamics and thematic shifts in artificial intelligence (AI) ethics over time, addressing the scarcity of longitudinal perspectives in the literature. In response to the rapid AI technology evolution and associated ethical risks and societal influences, this research integrates the theory of chance discovery with the KeyGraph algorithm. Guided by the double helix model of human–AI interaction, this work constructs a keyword co-occurrence network via iterative semantic exploration. Building on the co-occurrence structures and original textual data, this work employs ChatGPT for semantic interpretation, enhancing the accuracy and comprehensiveness of topic detection. The study analyzes AI ethics reports released between 2022 and 2024 by reputable, authoritative institutions, revealing that the thematic focus has expanded from technical risks to broader issues of institutional governance and societal trust. Various keywords, including *bias*, *privacy*, and *ethical*, have emerged as core nodes across multiple years, indicating a shift in AI ethics discourse from technical development to regulatory policy. This evolution highlights the formation of an integrated governance framework, encompassing technological robustness, institutional adaptability, and social consensus. This dynamic semantic analysis framework offers empirical contributions to AI ethics governance and knowledge development and valuable insight for researchers and interdisciplinary stakeholders.

Keywords: AI ethics; topic detection; KeyGraph; keyword network; ChatGPT

1. Introduction

The rapid advancement of artificial intelligence (AI) has transformed lifestyles by offering novel solutions to real-world problems and challenges and has introduced revolutionary changes across application domains. Although AI systems enhance efficiency and generate value, the ethical and societal issues they raise have become global concerns. For instance, the rapid development of various technologies (e.g., autonomous driving, drones, smart health care, and generative language models) has led to ethical risks, including algorithmic bias, data privacy infringements, opacity in decision-making, and ambiguous accountability. In response to these challenges, multiple international AI ethics guidelines and governance frameworks have been introduced, including the European Union's AI Act and various national white papers on technology ethics, which aim to address the ethical and social risks posed by technological advancements through policy and institutional design [1,2]. As AI applications penetrate deeper into decision-making processes and social governance, building AI systems with ethical sensitivity and social legitimacy has become inevitable in technological development and is a cornerstone for maintaining social trust and promoting sustainable development.

Against this backdrop, academic and public interest in AI ethics has significantly increased worldwide. Systematic searches of scholarly databases reveal that existing research encompasses diverse topics, including data governance, model fairness, transparency design, and accountability ethics, forming an interdisciplinary and multifaceted body of knowledge [3–6]. Nevertheless, most prior studies have focused on static articles or in-depth analyses of individual issues, without a

systematic examination of whether semantic shifts, thematic evolution, or value-focused transformations occur in AI ethics discourse over time. Given the continuous and rapid evolution of AI technology and its application contexts, discourse on ethical issues may significantly change over time and potentially shift in thematic focus due to event-driven factors, policy interventions, and public opinion.

Motivated by these considerations, this study employs topic detection techniques in text mining to conduct an in-depth analysis of multiple unstructured articles. This work explores whether the themes addressed in AI ethics reports demonstrate stable and consistent focal points or reveal dynamic shifts and contextual changes alongside technological and temporal developments. This research aims to identify the underlying semantic shifts and trends in AI ethics topics by thoroughly examining and comparatively analyzing recent AI ethics articles published by academic institutions, media outlets, and nonprofit organizations [7–9].

This study integrates the KeyGraph text mining algorithm grounded in the theory of chance discovery with the generative AI capabilities of the large language model (LLM)-based ChatGPT tool to address the challenges of semantic analysis in unstructured textual datasets. This integration establishes an innovative research workflow combining structured mining with semantic interpretation. KeyGraph, a graph-based method, constructs keyword networks and knowledge graphs by calculating the frequency and co-occurrence strength of keywords. Notably, KeyGraph identifies "chance" keywords that, despite their low frequency, possess significant bridging value in the network structure, revealing latent topics or emerging concepts in the dataset. This approach surpasses traditional frequency-based methods by uncovering hidden associations, offering the substantial potential for analyzing thematic evolutions and shifts in ethical focus [10].

Unlike standard topic modeling techniques, such as latent Dirichlet allocation (LDA), correlated topic models, and Pachinko allocation models, which are unsupervised learning methods requiring predefined topic numbers and statistical distributions to infer latent topics [11–13], topic detection emphasizes identifying semantic associations and dynamic thematic changes. Topic detection is suitable for analyzing rapidly evolving, cross-temporal, or dynamic problem-oriented datasets. The KeyGraph algorithm is a well-recognized topic detection tool that captures nonlinear co-occurrence relationships between keywords, making it appropriate for exploring complex ethical issues characterized by competing values and shifting contexts, as in this study [14–16].

Furthermore, this study employs ChatGPT as an auxiliary tool for semantic interpretation and summary generation to overcome the subjective limitations in conventional topic analysis methods that rely on expert interpretation for deep semantic understanding and domain knowledge. By applying ChatGPT's advanced language comprehension and summarization capabilities, researchers can achieve more accurate and focused thematic interpretations of each keyword cluster generated by KeyGraph, enhancing the precision of topic detection and the overall efficiency and interpretability of the analytical process [17].

Researchers have employed a human–AI interactive mechanism based on the double helix model to adjust node parameters, reclassify keyword clusters, and evaluate the semantic consistency and logical coherence of generated summaries dynamically. Through iterative HCI, this process constructs a logically coherent thematic structure for AI ethics [18].

This study analyzes AI ethics-related articles published between 2022 and 2024 by international academic institutions, news media, and nonprofit organizations. Rigorous selection criteria were applied during data collection. Sources were limited to reputable, authoritative organizations, and articles focusing exclusively on single disciplines or industry-specific applications were excluded. Comprehensive reports reflecting global trends were prioritized to ensure the dataset represented diverse perspectives and contentious issues in AI ethics worldwide, establishing a neutral and macro-level textual database. Eight representative documents were selected for each year as the basis for the KeyGraph algorithm-based topic detection and chance exploration.

This study conducts comparative analyses of annual thematic clusters using a multistage processing approach, including text preprocessing, keyword network construction, topic detection,

and semantic interpretation. Core themes, logical shifts, and changing focal points were identified, and the chance keywords were explored to detect potentially nascent but promising ethical issues.

In summary, this research combines the structural keyword mining capabilities of KeyGraph with the semantic comprehension of ChatGPT to develop a dynamic, extensible, and semantically rich method for topic detection. The approach aims to provide empirical evidence and strategic insight for trend monitoring, governance planning, and knowledge construction in AI ethics while enabling nonspecialist audiences to understand the evolving dynamics of AI ethics themes quickly.

2. Literature Review

2.1. AI Ethics Literature Review

Recently, ethical issues surrounding AI have become a primary focus in academia and the public sphere, reflecting their increasing importance in the global development of technology and social governance. Systematic searches of scholarly databases have revealed that the existing literature covers diverse themes, with some studies focusing on specific topics (e.g., data transparency and privacy protection), and others attempting to integrate multiple dimensions of ethical concerns for interdisciplinary comprehensive analyses [7,19]. However, structural examinations of the interrelationships between various ethical issues remain limited, especially in the context of rapidly evolving and emerging technology, including generative AI and autonomous driving. Thus, capturing the contextual shifts and logical frameworks of ethical controversies across interdisciplinary articles remains a significant research gap [20]. Against this background, AI ethics-related reports published by news media, academic institutions, and nonprofit organizations (characterized by their timeliness, readability, and public engagement) have become critical data sources for observing thematic evolution and tension regarding value. These reports complement the limited academic articles in clarifying practical contexts and ethical value conflicts.

The ethical issues surrounding AI are highly diverse and complex, encompassing multiple dimensions, including technology, law, society, and philosophy [19,20]. In response to these characteristics, this study employs the KeyGraph algorithm combined with text mining techniques to construct keyword co-occurrence networks based on the frequency and co-occurrence relationships of keywords. This work aims to reveal the ethical topics of concern systematically and visually across articles from various sources while assessing the keyword diffusion structure and intrinsic logical associations [21].

A core subtopic in AI ethics is AI governance, focusing on ensuring the trustworthiness and social acceptance of AI systems in public decision-making and societal applications. The literature indicates that trustworthy AI should encompass multiple elements, including accuracy, robustness, transparency, accountability, fairness, explainability, interpretability, legality, appeal mechanisms, and human oversight. However, in practice, these value indicators often cannot be satisfied simultaneously, resulting in significant priority conflicts and a lack of comparability when using a single value. For example, enhancing the accuracy of AI systems often relies on highly complex models, sacrificing transparency and explainability. Similarly, conflicts may arise in fairness metrics. The statistical impossibility triangle indicates that, under differing group base rates, simultaneously balancing false-positive rates and false-negative rates is not possible [20,22].

Furthermore, the integration of AI technology with the Sustainable Development Goals (SDGs) has become a focal point in AI ethics research. Although AI plays is critical for promoting technological applications, social innovation, and resource efficiency, it must also address the challenges posed by responsible governance and sustainable deployment.

Relevant studies have identified four primary challenges in integrating AI ethics with sustainability initiatives. First, in the ethical and social dimension, issues of transparency, fairness, bias, and accountability have gained increasing attention as automated decision-making and algorithmic deployment become more widespread. Second, the sustainability of AI imposes environmental pressure, particularly due to the high energy consumption and resource use required by large-scale model computations. Third, existing governance and regulatory frameworks often fail

to respond effectively to the challenges and issues introduced by technological evolution, typically remaining at a reactive level. Fourth, technical bottlenecks persist, including insufficient model explainability, data governance challenges, and the design of multidimensional performance metrics, such as accountability, fairness, and accuracy [9,22].

Moreover, the large volume of sensitive data generated by AI applications exacerbates risks related to data privacy and information security. In AI-driven knowledge management systems, ethical risks have emerged regarding various issues (e.g., privacy, bias, and transparency). A systematic review of 102 AI ethics research articles reveals that privacy and algorithmic bias account for 27.9% and 25.6% of the addressed topics, respectively, making them the most frequently addressed concerns. Additionally, transparency, accountability, and fairness remain core concerns [23].

This finding closely aligns with the ethical value frameworks identified in other studies, illustrating that the current AI ethics risks have gradually shifted from abstract principles to practical operational challenges, requiring interdisciplinary integration and technological governance for effective resolution [20,23,24]. To address these challenges, some studies have recommended adopting decentralized data governance models (e.g., federated learning and distributed data architectures) to balance data utility and privacy protection. Moreover, AI system performance evaluation should move beyond traditional single accuracy metrics to multidimensional frameworks encompassing environmental effects, social implications, and ethical compliance to reflect sustainability performance comprehensively. Furthermore, promoting inclusive development processes and human-centered design principles by incorporating diverse stakeholder perspectives can help avoid systemic biases and strengthen the legitimacy of ethical governance.

Although existing AI ethics research and policy documents have proposed core principles, such as transparency, accountability, fairness, and privacy, implementing these principles at the organizational level remains a considerable challenge. A notable gap exists between normative formulation and organizational practice in AI ethics, and relying solely on external guidelines and compliance requirements is insufficient to address the ethical controversies and issues encountered in practice. This gap underscores the importance of organizational AI ethics. Organizations should proactively establish internal governance mechanisms that integrate AI ethics principles throughout the technology life cycle, deliberation processes, and risk assessment frameworks to enhance ethical sensitivity and adaptive capacity [9,19,20,23].

In summary, AI ethics has progressed from the initial stages of principle declaration and conceptual proposals to practice-oriented institutional construction and governance innovation. Future research should further develop integrated analytical frameworks and multilevel governance models encompassing design principles, regulatory mechanisms, ethical conflict identification, and social participation, to ensure that AI development promotes technological innovation and upholds ethical values, social order, and the core objectives of sustainable development [9,19,20].

2.2. Chance Discovery Theory

The chance discovery theory is an interdisciplinary data mining framework designed to identify rare yet highly valuable "chances" in data via HCI and structural data analysis, which may critically influence future decision-making or system development. In this theory, *chance* is broadly defined as important information that can guide decision-makers or automated decision systems to make significant responses. These chances may indicate emerging opportunities not yet explicitly recognized or indicate potential undiscovered risks and crises [10,25–28].

Unlike the discovery of random events, chance discovery stresses conscious awareness rather than random detection [26]. Chance discovery analyzes the frequency of information or keyword occurrences and places greater importance on their associations with other critical information. Keywords and their connections in articles can be visualized via structured keyword analyses and visualization tools (e.g., the Polaris visualization tool implementing the KeyGraph algorithm) [25], identifying bridging words that are highly associated with multiple keyword clusters. These bridging

words may indicate potential chances. This theory has significant application potential in innovation early warning, risk governance, and strategic planning, while providing theoretical support and practical foundations for structured semantic analyses and topic detection methods [10,29,30].

Compared to traditional data analysis methods that assume a stable data structure and known variables, chance discovery focuses on information nodes with low frequency but potentially profound semantic or systemic implications. This process is a gradual unfolding of understanding rather than an immediately identifiable event. The theory of chance discovery underscores that the most critical insight is often hidden in unstructured information, and this theory stresses the need for decision-makers to engage in the dynamic evolution of information. This theory combines the human sensitivity to context and situational awareness with the computational capabilities in data processing and visualization to effectively detect chance information and strategic responses [10,29,30].

Professor Ohsawa proposed the following three criteria for judging chance [10,29,30]:

- **Establishing and uncovering innovative models and variables:** Rather than relying on existing data models and variables, this approach incorporates contextual factors into the analysis to identify noteworthy variables emerging in specific situations, preventing the results from diverging from practical needs and enhancing the accuracy of chance detection.
- **Identifying tail events:** Tail events are rare events with a low frequency but profound influence on the system or domain. Undiscovered chances or phenomena can be identified by observing and analyzing such tail events.
- **Relying on human–AI interaction for interpretation and judgment:** Whether a tail event represents a genuine chance can be discerned by applying extensive human background knowledge and contextual sensitivity. This approach is necessary because the rarity and ambiguity of tail events make it challenging for fully automated data mining methods to assess their true value and significance accurately.

Therefore, the chance discovery theory emphasizes HCI, asserting that the effective awareness, understanding, and realization of chances can be facilitated only by combining computational data processing and visualization capabilities with the rich background knowledge of domain experts, enhancing the accuracy and practical utility of chance identification.

In this exploratory process, information is generated and interpreted in a nonlinear and intertwined manner, giving rise to the double helical model as a cognitive framework for exploring latent meaning via HCI. Computers continuously mine data from the environment and human expression, providing visualized feedback to assist human understanding and judgment. Humans iteratively adjust their comprehension and focus, inspiring new exploration directions. By further integrating the subsumption architecture cognitive model, this approach presents a nonlinear, concurrently operating interactive feedback loop comprising computer mining, feedback results, human understanding, and decision-making processes, requiring continuous and iterative exploration that more closely aligns with human cognitive patterns [10,26,31].

The theory of chance discovery has been widely applied across multiple domains, including health care, business innovation, marketing, disaster prediction, risk management, and decision-making, underscoring its generative rather than merely analytical power. This theory facilitates a deeper understanding of existing data and helps identify rare events and latent information that are difficult to detect via conventional analyses, yet have a significant influence on future decisions. This theory provides a data analysis method and integrated knowledge creation framework that combines human–computer interaction (HCI), semantic construction, and abductive reasoning, fostering the formation of new hypotheses and value creation [10,25,26].

2.3. Double Helix Model: Human–Machine Collaborative Framework for Chance Discovery

This study adopts the theory of chance discovery as its theoretical foundation and applies its core framework (the double helix model) to conduct semantic network analyses and topic detection via an HCI process. The model comprises two interwoven components: the computer- and human-driven processes, forming a spiral cognitive feedback mechanism akin to the structure of DNA. This approach represents the dynamic interplay between data-driven analyses and knowledge interpretation in HCI. The process is designed to assist researchers in identifying low-frequency yet semantically significant nodes (*chances*) in the keyword network [21,25,26,31–33].

Based on this integrative explanation, this study applies a dual-dimensional interactive structure, advancing through a spiral formation that illustrates the iterative nature of human–AI collaboration. As depicted in Figure 1, researchers can follow the spiral pathway to track the feedback and adjustment mechanisms dynamically across each stage. According to this double helix model, the HCI loop is divided into four primary stages, each embodying iterative cycles between the human- and computer-driven processes. This ongoing dynamic feedback mechanism facilitates the progressive refinement and optimization of the semantic structure and topic detection [21,25,26,31–33].

The specific phases of the human–AI interaction cycle are as follows [10,21,33–35]:

- **Human-driven process: New setting for analysis (inputting article data and initializing parameters).** The researcher conducts this phase, which marks the starting point of the overall HCI cycle. Based on the research objectives, the researcher provides the original article dataset and employs the Polaris visualization tool to set the initial parameters for the KeyGraph algorithm, such as the number of default bridging nodes (represented as red nodes) and high-frequency black keyword nodes, establishing the foundation for an automated computer analysis (see Phase 1 in Figure 1).

- **Computer-driven process: Data mining (data mining and keyword network construction).** After the initial parameter setting, the process transitions to a computer-driven process, entering the phases of data mining and keyword network construction. At this stage, the computer autonomously executes the KeyGraph algorithm to conduct in-depth mining of the article dataset. The system constructs a co-occurrence network graph by calculating the co-occurrence frequency and structural relationships between keywords. This phase applies the KeyGraph algorithm to extract latent knowledge structures and keywords automatically from large-scale articles, providing a foundation for semantic interpretation (see Phase 2 in Figure 1).

- **Computer-driven process: Visual results (network graph visualization).** After data mining is completed, the computer transforms the keyword network generated by the KeyGraph algorithm into a visualized graph illustrating the connections (i.e., the co-occurrence relationships) between keywords, including the red nodes. This visual representation serves as a bridge between the computer and human user, converting abstract keyword associations into intuitive, interpretable images that facilitate information integration and semantic judgment. At this stage, the computer completes its intermediate task and awaits human intervention for further examination (see Phase 3 in Figure 1).

- **Human-driven process: Understanding and re-evaluation (interpretation and review).** This phase represents the core of human knowledge interpretation in the double helix model, highlighting the iterative nature of HCI. In this process, researchers do not directly engage in topic detection; instead, based on their domain expertise and semantic comprehension abilities, they systematically evaluate the topic detection and semantic interpretation results generated by ChatGPT from the keyword network visualizations constructed using the KeyGraph algorithm. Researchers examine the thematic and keyword relationships in the visualized graphs and assess whether ChatGPT's initial interpretations are logically coherent, substantively meaningful, and effectively reveal the latent semantics in the articles. For instance, if ChatGPT produces an illogical output (e.g., "I beer"), researchers employ the Polaris visualization tool to adjust the node parameters of the KeyGraph algorithm (e.g., the number of high-frequency black

keywords or red chance nodes) until the output becomes coherent and logically consistent (e.g., "I love to drink beer"). This ongoing process of understanding the output and continuously tuning parameters, which triggers new data mining and visual output, is the critical objective of the iterative cycles in the double helix model. Thus, this approach forms a continuously optimized spiral iteration process (see Phase 4 in Figure 1).

The mentioned HCI process is not a one-time analysis but integrates efficient computer processing capabilities with human critical thinking and domain expertise. The iterative refinement of keyword structures and thematic interpretations forms a continuous and spiraling process via repeated interactive cycles. This iterative cycle is the core driving force of the double helix model, ensuring the analysis process maintains high flexibility and adaptability [21]. Guided by their deep understanding of the data, researchers continually refine the model, progressively transforming implicit keyword structures into explicit knowledge, providing a solid foundation for informed decision making. Researchers can interpret and understand the emerging or evolving themes and issues via this process, enabling more informed and progressive decisions. This example illustrates the advantages of HCI in addressing complex problem-solving [10,32,34,35].
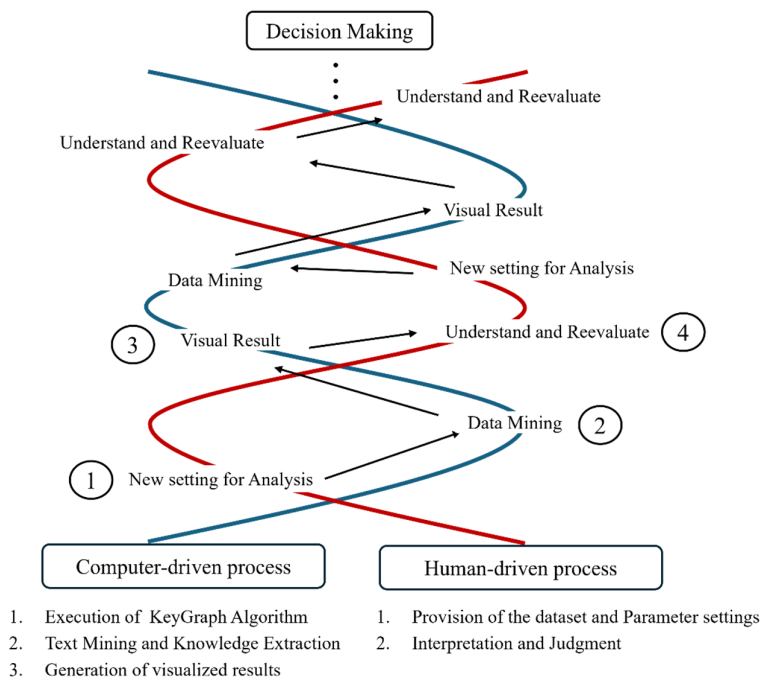

Figure 1. Schematic diagram of the double helix model.

2.4. KeyGraph Algorithm Overview

The KeyGraph algorithm is a core tool for implementing chance discovery. This graph-based text mining technique aims to uncover critical contexts and latent events with significant influence that are hidden in texts via a keyword network model, which is compared to traditional keyword frequency-dependent algorithms (e.g., term frequency-inverse document frequency or LDA). The KeyGraph algorithm is widely applied in topic detection research. The operational mechanism extracts high-frequency keywords from the text as nodes and analyzes the co-occurrence relationships between these keywords to construct a keyword network graph [36–40].

Furthermore, the uniqueness of this algorithm lies in its ability to identify "keywords that have structural bridging value but occur with low frequency," revealing latent nonexplicit topics and interdisciplinary conceptual connections in the text. This approach enables the derivation of core issues and underlying value perspectives in articles. This method does not require manual annotation or prior knowledge and can automatically extract representative keywords or topics from the

collected technical texts or academic literature. This method constructs a keyword network graph to present the associative structure of keywords, enhancing the transparency and interpretability of topic detection [36,37,41,42].

2.4.1. KeyGraph Keyword Network Structure and Visualization

The KeyGraph algorithm is suitable for applications, including knowledge structure exploration, thematic context evolution analysis, and emerging topic detection, owing to its advantages in knowledge structure construction and analysis. Ohsawa first proposed this algorithm in 1998, initially as an automatic indexing technique based on keyword co-occurrence graphs in texts, and it was later developed into a core analytical tool for chance discovery applications. In Ohsawa's original conceptualization, the KeyGraph algorithm is explained using an architectural structure analogy. The literature is viewed as a building, where the foundation represents the fundamental concepts of the literature, constructed by analyzing co-occurrence relationships between high-frequency keywords. The pillars symbolize the associations between the keywords and foundational concepts, forming a structured network connecting concepts in the literature. The roof represents the core viewpoints in the literature, typically constituted by low-frequency bridging keywords strongly connected to multiple conceptual clusters, reflecting the primary perspectives or innovative points in the literature [10,38,41].

In terms of technical implementation, the KeyGraph algorithm produces a graph structure by analyzing co-occurrence relationships between keywords, where nodes represent keywords and edges indicate the strength of these co-occurrences. As depicted in Figure 2, the concepts of black and red nodes are further introduced to describe this structure precisely. Black nodes represent high-frequency keywords with strong connections to fundamental concepts. These nodes are responsible for the interpretability of the knowledge structure, serving as the backbone of the knowledge graph and the foundational structure of KeyGraph. Typically, multiple black nodes form clusters that contain latent topics embedded in the articles. Red nodes are keywords with lower frequencies but strong co-occurrence relationships with multiple clusters. They are often metaphorically associated with chance discovery, reflecting atypical yet potentially valuable keywords in articles, called chance nodes in this work. Black edges represent strong co-occurrence relationships between black nodes, forming stable keyword clusters via their connections. Red edges denote bridging relationships between red nodes and clusters, highlighting the value of rare events [10,37,38].

A distinctive feature of the KeyGraph algorithm is its ability to extract key concepts and their associations automatically from articles, revealing implicit relationships between these concepts. Generating visualized knowledge graphs uncovers latent and significant information in articles, enabling users without a technical background to comprehend the complex structures and interrelations of keywords intuitively [37,41,43] (Figure 2).
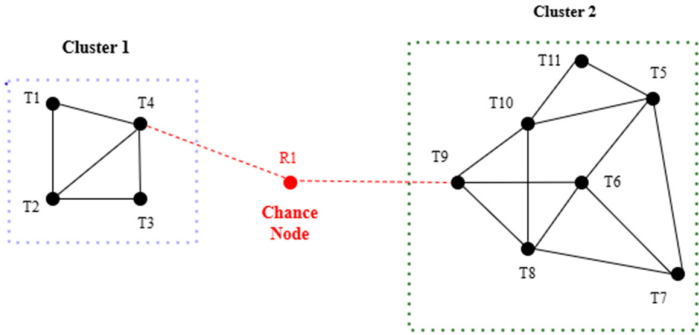

Figure 2. Illustration of the KeyGraph algorithm.

2.4.2. KeyGraph Algorithm

1. **Data preprocessing:** Data preprocessing serves as the foundation for constructing a keyword network in the KeyGraph algorithm. In this study, this phase involves several steps, including tokenization, normalization, stop-word removal, and part-of-speech filtering. Tokenization and normalization establish a stable keyword base, whereas stop-word removal and part-of-speech filtering reduce semantic noise, enhancing the accuracy of co-occurrence analysis and the network structure quality, optimizing the performance of topic detection and chance node identification [38,44,45].

2. **High-frequency keyword extraction:** Based on the preprocessed data, a new dataset D, is generated, comprising a series of sentences, each representing a set of keywords. All keywords are ranked according to their frequency of occurrence in dataset D, and the highest-frequency keywords are selected to form a high-frequency keyword set. These keywords serve as the nodes of the network cluster G [38,44–46].

3. **Calculation formula for keyword network co-occurrence:** In the KeyGraph algorithm adopted in this study, the co-occurrence relationship between keywords serves as the core basis for constructing the keyword network. Each keyword is regarded as a network node. When two keywords co-occur in the same semantic unit (e.g., a sentence or paragraph), a link (edge) is formed between the nodes. The KeyGraph algorithm employs a specific measure called the co-occurrence strength to quantify the co-occurrence relationship between keywords [38,47,48], calculated in Eq. (1):

$$assoc(w_i, w_j) = \sum_{s \in D} min\ (|w_i|_s, |w_j|_s) \tag{1}$$

where $assoc(w_i, w_j)$ represents the co-occurrence strength between keywords w_i and w_j in all semantic units s in dataset D. This measure is calculated by summing the minimum occurrence frequencies of the two keywords in the same semantic unit, reflecting their co-occurrence count. In addition, $|w_i|_s$ and $|w_j|_s$ denote the frequencies of keywords w_i and w_j, respectively, in the semantic unit s, where min $(|w_i|_s, |w_j|_s)$ indicates the minimum occurrence frequency between w_i and w_j in the semantic unit s. This metric aggregates the minimal occurrence counts across semantic units to capture the overall semantic linkage strength between the keyword pair. This approach helps construct a semantic backbone comprising high-frequency terms and reveals latent nodes that may have lower surface frequencies yet critical semantic significance.

In addition to the absolute co-occurrence values described above, studies often employ normalized indicators of co-occurrence strength, such as the Jaccard similarity coefficient, as a relative measure [38], calculated in Eq. (2):

$$Jaccard(w_i, w_j) = \frac{Freq(w_i \cap w_j)}{Freq(w_i \cup w_j)}, \tag{2}$$

where $Freq(w_i \cap w_j)$ indicates the number of times the keywords w_i and w_j co-occur in the same semantic unit in dataset D, and $Freq(w_i \cup w_j)$ refers to the total frequency of either w_i or w_j appearing in dataset D. The value of the *Jaccard* similarity coefficient ranges from 0 to 1, where a higher value represents a stronger similarity between keywords, implying greater semantic similarity. As a normalized similarity measure, the *Jaccard* similarity coefficient can be employed to adjust the edge weights in the network generated by the KeyGraph algorithm, mitigating the linking bias caused by high-frequency terms and enhancing the performance of the keyword network in topic detection and keyword identification. This approach helps reveal the relational paths in the deeper semantic structure more effectively.

4. **Co-occurrence measurement between keywords and keyword clusters:** The KeyGraph algorithm employs the Co-Occurrence Strength Index called $based(w, g)$, which measures the degree of their co-occurrence within articles to calculate the connection strength between a keyword w and a single keyword cluster g [38,47,48], as in Eq. (3):

$$based(w, g) = \sum_{s \in D} |w|_s \, |g - w|_{s''} \tag{3}$$

where w refers to a retained keyword in the preprocessed dataset D, g denotes the cluster to which the keyword belongs, and D represents the dataset obtained after preprocessing. The co-occurrence strength is calculated based on sentences s, the fundamental semantic units, and the sentences are typically treated as sets of keywords that form the basis for defining co-occurrence relationships. In addition, $|w|_s$ indicates the frequency of keyword w appearing in sentence s, whereas $|g - w|_s$ represents the total number of occurrences of all keywords in cluster g, excluding w in the same sentence s. This value is zero if no other keywords from the cluster appear in the sentence.

This formula processes each sentence in dataset D to determine whether the target keyword w appears. If it does, this formula calculates the frequency of keyword w and multiplies it by the total frequency of the other keywords in cluster g (excluding w) in the same sentence. The resulting products are summed across all sentences to quantify the overall co-occurrence strength between keyword w and cluster g.

5. **Calculating the co-occurrence potential of all keywords in cluster g:** In a keyword network analysis, the association between a keyword and the keyword cluster g depends on the degree of co-occurrence and the contextual interactions between the cluster and other keywords. The KeyGraph algorithm provides a standardized metric for evaluating such associations by defining a cluster-level semantic quantification measure called $neighbors(g)$, estimating the potential of a specific keyword cluster g to interact semantically with other keywords throughout the dataset [38,47,48], calculated in Eq. (4):

$$neighbors(g) = \sum_{s \in D} \sum_{w \in S} |w|_s |g - w|_s \tag{4}$$

where s represents each sentence in dataset D, which is regarded as a set of co-occurring keywords and is the basis for defining co-occurrence relationships between keywords. In addition, S denotes the set of all keywords, $w \epsilon S$ indicates any keyword in set S, $|w|_s$ refers to the frequency of keyword w appearing in sentence s, and $|g - w|_s$ denotes the total frequency of all keywords in cluster g (excluding keyword w) appearing in the same sentence s. This value is zero if none of the keywords in the cluster appear in the sentence.

This formula traverses each sentence s in dataset D and, for each keyword in the high-frequency keyword set S, calculates its co-occurrence strength with other keywords in cluster g when w appears in the sentence. These co-occurrence values are summed, and this equation measures the total co-occurrence strength between keyword cluster g and keywords in set S across dataset D. A higher $neighbors(g)$ value indicates that cluster g has a stronger connection with high-frequency keywords in the keyword network, which may suggest its potential significance in the semantic structure.

6. **Evaluation of the importance of the potential of keywords across clusters:** This study adopts the keyness calculation formula proposed in the KeyGraph algorithm to evaluate the connective role played by keyword w in the overall keyword network graph to determine whether a specific keyword possesses the semantic potential to bridge clusters [38,47,48], as computed in Eq. (5):

$$key(w) = 1 - \prod_{g \subset G} \left[1 - \frac{based(w,g)}{neighbors(g)} \right], \tag{5}$$

where $key(w)$ represents the importance score of keyword w, ranging between 0 and 1, and G denotes the set of all keyword clusters in the keyword network graph, where each g refers to an individual keyword cluster. The expression $g \subset G$ indicates that every cluster in set G is evaluated individually to compute the semantic relevance of keyword w to each cluster. Moreover, $based(w, g)$ refers to the co-occurrence strength between keyword w and cluster g, whereas $neighbors(g)$ represents the total co-occurrence strength of cluster g.

This formula represents the product of the co-occurrence complement of keyword w with each keyword cluster $g \subset G$, indicating the overall probability that keyword w has no co-occurrence association with various keyword clusters. The final score $key(w)$ is obtained by subtracting this product from 1, reflecting the importance of keyword w in bridging multiple clusters in the keyword network.

When keyword w exhibits significant co-occurrence relationships with multiple keyword clusters, its $key(w)$ score approaches 1, indicating that keyword w is highly associated with multiple keyword clusters and may represent a potential chance node. Conversely, when $key(w)$ approaches 0, keyword w may have little to no co-occurrence with keyword clusters, reflecting a lower importance in the keyword network structure.

3. Materials and Methods

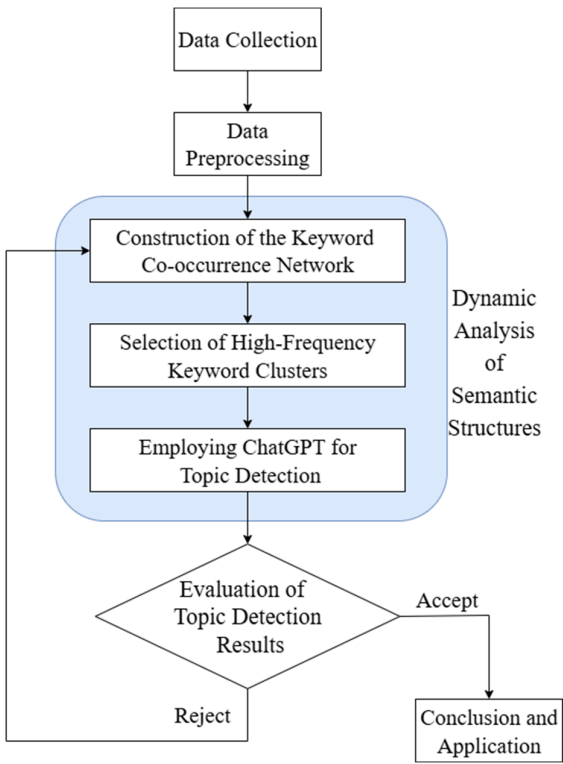The research process is divided into the five stages at the beginning of Figure 3.

Figure 3. Research process flowchart.

3.1. Data Collection

The data collection method in this study primarily employs the search terms *AI ethics, ethical AI, and responsible AI*, focusing on the overall concepts and frameworks related to AI ethics, cross-industry universal guidelines, and globally influential and controversial problems, while avoiding biases toward specific industry applications. Moreover, AI ethics emphasizes moral principles at the technical level, addressing ethical issues in the design and operation of AI systems [49]. Ethical AI concerns the implementation of ethical principles in the development and application of AI technology, ensuring compliance with moral standards [50]. Responsible AI highlights the responsibilities and regulations of developers and users regarding the social influence of AI technology, promoting accountability and governance [51]. These three concepts are complementary and synergistic, facilitating an in-depth exploration of the ethical challenges faced by AI and

advancing discourse on its practical applications and social responsibilities toward greater depth and breadth.

This study prioritizes the selection of reports and articles related to AI ethics published in English between 2022 and 2024 via online searches to emphasize data timeliness and capture recent developments in the field. The aim is to clarify the perspectives and latest trends in AI ethics discourse systematically against the background of rapid AI technological evolution during this period (see Tables 1–3).

Table 1. Data sources for international reports and articles in 2022.

	Original Title	Data Sources	Publication Date
1	The 2022 AI Index: Industrialization of AI and Mounting Ethical Concerns	Stanford HAI	2022/03
2	AI Ethics And AI Law Grappling With Overlapping And Conflicting Ethical Factors Within AI	Forbes	2022/11
3	The 2022 AI Index: AI's Ethical Growing Pains	Stanford HAI	2022/03
4	Prioritising AI & Ethics: A perspective on change	Deloitte	2022/03
5	Top Nine Ethical Issues In Artificial Intelligence	Forbes	2022/10
6	AI Ethics And AI Law Are Moving Toward Standards That Explicitly Identify And Manage AI Biases	Forbes	2022/10
7	Evaluating Ethical Challenges in AI and ML	ISACA Journal	2022/07
8	We're failing at the ethics of AI. Here's how we make real impact	World Economic Forum, WEF	2022/01

Table 2. Data sources for international reports and articles in 2023.

	Original Title	Data Sources	Publication Date
1	The Ethics Of AI: Navigating Bias, Manipulation And Beyond	Forbes	2023/06
2	The Ethics Of AI: Balancing Innovation And	Forbes	2023/12
3	Responsibility	Forbes	2023/07
4	AI Ethics In The Age Of ChatGPT—What Businesses Need To Know	Forbes	2023/05
5	96% Of People Consider Ethical And Responsible AI To Be Important	Forbes	2023/03
6	How Businesses Can Ethically Embrace Artificial Intelligence	CNN	2023/12
7	Experts call for more diversity to combat bias inartificial intelligence	Georgia Tech	2023/08
8	5 AI Ethics Concerns the Experts Are Debating	Bloomberg	2023/06

Table 3. Data sources for international reports and articles in 2024.

	Original Title	Data Sources	Publication Date
1	AI's Trust Problem	Harvard Business Review	2024/05
2	'Uncovered, unknown, and uncertain': Guiding ethics in the age of AI	Yale News	2024/02
3	AI Regulation Is Evolving Globally and Businesses Need to Keep Up	Bloomberg Law	2024/12
4	AI is not ready for primetime	CNN Business	2024/03
5	With AI warning, Nobel winner joins ranks of laureates who've cautioned about the risks of their own work	CNN	2024/12
6	Navigating The Ethics Of AI: Is It Fair And Responsible Enough To Use?	Forbes	2024/11
7	AI And Ethics: A Collective Responsibility For A Safer Future	Forbes	2024/10
8	AI Started as a Dream to Save Humanity. Then, Big Tech Took Over.	Bloomberg	2024/09

A rigorous screening process was conducted during the data collection phase to ensure the comprehensiveness and authority of the research data. Data sources were limited to internationally recognized and credible academic research institutions, news media, and nonprofit organizations. Articles focusing solely on a single domain or industry application were excluded, prioritizing comprehensive reports reflecting international trends. This approach aimed to ensure that the data adequately represent diverse global perspectives and contentious issues regarding AI ethics, constructing a neutral and macro-level article dataset. Eight highly comprehensive and interpretative articles were selected for each year as the analytical corpus for the KeyGraph-based topic detection and chance exploration.

3.2. Data Preprocessing

This study employs the KeyGraph algorithm to extract topics from articles, based on co-occurrence relationships between keywords, and constructs a structured visual graph. However, the original articles often contain numerous stop words, irrelevant information, and punctuation marks. If the articles are analyzed directly using the KeyGraph algorithm without preprocessing, extracting the associations between keywords becomes difficult, resulting in overly cluttered or off-focus keyword networks. This study segments sentences, tokenizes words, and filters out meaningless words and stop words prior to analysis to enhance the accuracy of keyword identification and reduce graph noise, ensuring the accuracy and visualization quality of the keyword co-occurrence network graph.

This study employs Python combined with the Natural Language Toolkit (NLTK) for article preprocessing during the data processing phase to achieve the objectives. First, each collected article was segmented into independent sentences based on periods, and each sentence was converted to lowercase and tokenized into individual words while removing punctuation marks but retaining contractions containing apostrophes (e.g., in the work *don't*). The stopword functionality for NLTK was used to filter out semantically insignificant words (e.g., *the*, *is*, and *and*) to reduce data noise and highlight keywords related to AI ethics. The filtered word list was written line by line into output files with the words separated by spaces. A manual review and filtering process was also conducted

to remove any remaining irrelevant words and tokenization errors to enhance the thematic relevance of the data. This process effectively streamlines the data, improves keyword prominence, and establishes a structured data foundation for the keyword network of the KeyGraph algorithm, ensuring the accuracy and visualization quality of the analyses.

3.3. Construction of the Keyword Co-Occurence Network

3.3.1. Chance Discovery in AI Ethics Using KeyGraph

This study employs the graph-based KeyGraph text mining technique as the core tool for topic detection and chance discovery in AI ethics articles to overcome the limitations of traditional text mining methods in analyzing topic evolution and identifying latent issues. As described in Section 2, KeyGraph is a keyword network analysis method that integrates lexical co-occurrence structures with chance identification. By constructing a co-occurrence graph of keywords, KeyGraph reveals the intrinsic topic structures and semantic linkages in articles, making it suitable for exploring cross-topic or dynamic contexts and emerging keywords. A distinctive feature of KeyGraph is its ability to identify low-frequency but highly connected chance nodes linked to multiple topic clusters, uncovering latent issues that traditional high-frequency analyses often fail to capture. The KeyGraph analytical process can be divided into the following five core strategies, serving as the foundation for semantic mining and topic detection in AI ethics articles:

- **Keyword frequency and co-occurrence calculation:** First, the occurrence frequency of all words in the articles is calculated and sorted. The top consecutive high-frequency words are selected as keywords, representing the core foundational concepts of the articles. Using paragraphs or sentences as the calculation units, the co-occurrence relationships between all keywords are computed and applied to establish connections.
- **Node role classification and keyword clustering:** Based on the frequency of keyword occurrences and their structural positions in the co-occurrence network, nodes are classified into three categories, which lays the foundation for chance discovery.
  - **High-frequency keywords:** Keywords with high occurrence frequency that are concentrated in specific topic clusters represent the primary concepts of the topics. In this study, these are consistently represented by high-frequency black nodes.
  - **Chance keywords:** These keywords (known as bridging words) have lower occurrence frequencies but are associated with multiple topic clusters. They typically indicate emerging concepts or interdisciplinary issues and are valuable for discovering latent topics. In this study, they are represented by red nodes.
  - **General terms:** Keywords lacking structural significance are excluded from the visualization network.
- **Keyword co-occurrence network construction and thematic cluster identification:** A keyword association graph is constructed with keywords as nodes and the co-occurrence strength as weighted edges. This method aggregates high-frequency terms and forms thematic clusters.
- **Keyword network visualization:** The nodes and links are visualized using tools (e.g., Polaris), which map co-occurrence relationships between keywords to construct their association network graphs. By adjusting parameters (e.g., frequency thresholds, co-occurrence strength, and the number of nodes), different levels of keyword structures are explored to enhance the understanding of potential keyword clusters and association pathways.

In the practical implementation, this study preprocesses the articles, including word segmentation and stopword removal, and segments them according to the time series from 2022 to 2024 to observe changes and shifts in potential topics or issues over time. The Polaris visualization tool was employed in combination with statistical methods (e.g., word frequency and co-occurrence analysis) and data mining techniques to execute the KeyGraph algorithm [52], which automatically calculates the keyword co-occurrence frequency and co-occurrence strength. This tool constructs a keyword network graph to explore keyword structures at various levels, promoting semantic

interpretation and chance discovery. In the network, high-frequency black nodes represent keywords with a high frequency and stable semantic cores, whereas red nodes represent potential keywords with lower frequency but strong connections to multiple topic clusters. Black edges between nodes reflect the co-occurrence strength between keywords.

This study applies the parameter adjustment functions provided by Polaris to identify an appropriate keyword network structure and dynamically set conditions (e.g., keyword frequency thresholds, co-occurrence link strength, and maximum number of nodes), controlling the hierarchical levels and cluster partitioning of the keyword network graph. This approach enhances the structural clarity of the visualization, facilitating a comparative analysis and interpretation of the keyword networks across years and enabling further observation of the formation, expansion, and evolution of topic clusters.

This method reveals the explicit topic structures in AI ethics articles and can uncover potential low-frequency keywords and emerging issues that traditional techniques often fail to capture, providing a solid foundation for topic detection and chance discovery. Overall, by integrating keyword co-occurrence structures, latent chance nodes, and visualization tools, KeyGraph effectively extracts explicit and implicit issues in articles, enhancing the exploratory and strategic aspects of topic detection.

However, the interpretive precision of the keyword network graph and the clarity of topic detection often depend on the number of high-frequency black nodes and the ratio between high-frequency black nodes and red chance nodes. An excessive or insufficient number of these nodes may affect the semantic clarity and accuracy of topic detection, influencing the reliability of the analytical results. The following section analyzes the relationship between semantic node density and the effectiveness of topic detection, investigating the effect of node density on the keyword structure and topic differentiation.

3.3.2. Analysis of Keyword Network Node Density and Topic Detection Accuracy

When conducting a KeyGraph keyword network analysis, setting too many high-frequency black nodes (e.g., designating 100 out of 1,000 (10%) distinct terms as black nodes) may lead to an overly complex network structure, adversely affecting the accuracy and focus of topic detection. In KeyGraph, high-frequency black nodes represent the core terms in the keyword network, outlining the principal thematic structure of the articles. However, when the number of high-frequency black nodes is too large, the co-occurrence density of high-frequency keywords increases significantly, resulting in an overly dense network. This density can blur thematic boundaries, intensify the semantic overlap between nodes, and hinder the convergence of co-occurrence paths, weakening the ability to detect latent topics and contextual structures [53–56].

Second, an excessive number of high-frequency terms may include morphologically varied but semantically similar words (e.g., make, makes, and made). Although these high-frequency terms frequently co-occur, they may lack clear thematic referentiality, disrupting the focus of the keyword network. This disruption often leads to topic analysis results that are biased toward overly generalized dominant themes or may even trigger semantic hallucinations, limiting the ability to identify subtle, overlapping, or emerging topics [53–56].

From an operational perspective, setting too many high-frequency black nodes can lead to an overly complex graph structure, reducing the feasibility of cluster partitioning and cross-validation and increasing the difficulty of topic analysis. This outcome may result in the omission of critical information during topic summarization or affect the interpretability of the detection process, weakening the depth and novelty of the conclusions. Conversely, setting too few high-frequency black nodes may cause essential topic-related keywords to be inadequately captured. In particular, some secondary terms (although not highly frequent) may carry significant semantic meaning but be excluded from the analysis, leading to an unbalanced topic distribution. This outcome compromises the formation of structured keyword clusters, dilutes the network core, and undermines the coherence of the keyword network and the overall inference of topic evolution.

In summary, determining the optimal number of key nodes requires iterative parameter tuning and visual inspection to control node density appropriately and identify the most suitable keyword network. This approach helps maintain the structural stability of the keyword network, reducing the risk of overlap and semantic hallucination in the topic analysis and enhancing the overall interpretability and exploratory depth of the analysis.

3.4. Selection of High-Frequency Keyword Clusters

After constructing the keyword network, this study conducts manual classification and clustering based on the network structure formed by the high-frequency black nodes. The initial clustering process employs the chance nodes (i.e., red nodes), identified by the KeyGraph algorithm, as the starting points for keyword diffusion. These red nodes are considered anchors for potential emerging or latent themes due to their role in bridging high-frequency keywords despite having a relatively low frequency themselves. From each red node, the diffusion extends outward to directly connected high-frequency black nodes, with the number of connected black nodes limited to a maximum of six to seven per red node. This setting helps control the cluster size and prevents excessive expansion that may blur or overgeneralize the results of the topic analysis.

After completing the initial clustering, each cluster was input into ChatGPT for topic detection and semantic interpretation. ChatGPT analyzes the complex relationships between the keywords in each cluster and infers the potential ethical issues or discourse associated with the cluster. If the interpretations generated by ChatGPT lack logical coherence or sufficient semantic clarity, researchers can adjust the relevant parameters of the KeyGraph algorithm (e.g., the number of high-frequency black nodes, connection strength, or red nodes) to reconstruct a new keyword structure and cluster distribution. This dynamic human–AI collaborative adjustment mechanism is iteratively repeated until the resulting cluster division demonstrates semantic clarity and structural coherence.

Overall, this procedure embodies a human–AI collaborative mechanism for semantic construction. The KeyGraph algorithm segments clusters based on the proximity of red nodes and high-frequency black nodes, considering their co-occurrence. In contrast, ChatGPT provides complementary support for semantic interpretation, and researchers can manually adjust parameters to ensure coherence. This approach reflects the core aim of HCI in the double helix model, forming a dynamic and iterative process of topic analysis that enhances the semantic focus and ensures the structural integrity of the keyword network.

3.5. Employing ChatGPT for Topic Detection

This section elaborates on how the KeyGraph algorithm is employed to conduct topic detection and chance discovery in AI ethics articles while examining the limitations of traditional keyword network analysis methods. This work employs ChatGPT as an auxiliary tool for semantic interpretation to overcome these constraints during the potential topic detection phase of keyword clusters. The technical advantages and application strategies of this integration are detailed below.

3.5.1. Limitations of Previous Methods

Before the widespread adoption of LLMs, such as ChatGPT, article mining for keyword networks and topic detection primarily relied on interpreting the semantic relationships between node clusters and bridging chance nodes in the keyword network graph. Researchers subjectively assigned thematic meanings to the keyword structures and topics based on these relationships. Researchers needed to trace the data back to the original article and examine the contextual usage for semantic interpretation to clarify the semantic association of a bridging node (e.g., accountability) with multiple clusters [57–60]. In this analytical framework, topic detection and semantic clarification were achieved primarily through the following approaches.

1. **Topic cluster identification and core concept summarization:** KeyGraph identifies high-frequency keywords in articles and designates them as high-frequency nodes (i.e., black nodes)

in the keyword network structure. Based on the co-occurrence relationships between these keywords, tightly connected clusters naturally form, reflecting the primary themes or subdomains in articles. Researchers can summarize representative thematic labels based on the characteristics and co-occurrence patterns of keywords in each cluster, producing an initial thematic summary and classification of the core article content.

2. **Chance keyword identification and pairwise semantic relationship mining:** The uniqueness of KeyGraph lies in its ability to identify chance keywords that, despite their low frequency, connect multiple thematic clusters. Although these keywords appear infrequently, they serve as bridging nodes linking thematic clusters in the keyword network. Researchers conduct in-depth analyses of these chance keywords by tracing their contextual usage back to the original articles, manually interpreting their semantic roles and how they connect with multiple thematic clusters. This process facilitates identifying emerging topics, interdisciplinary integration points, or potential trends.

However, although KeyGraph can provide structural information and identify potential chance keywords, without LLM assistance, theme summarization and semantic clarification still heavily rely on manual interpretation and domain-expert knowledge. Researchers must manually integrate the keyword network graph, centrality metrics, and the contextual usage of terms in the original articles to trace and interpret the semantic roles of potential keywords and their connections to multiple thematic clusters. This process is time-consuming, and the results are often limited by the researchers' professional judgment, reducing the efficiency and scalability of semantic mining. These traditional methods commonly face several significant limitations [59,60].

1. **Topic summarization heavily relies on manual interpretation, resulting in subjectivity and inconsistency:** Although traditional keyword network graphs can visually present co-occurrence relationships between high-frequency keywords, their semantic connections often lack systematic explanatory mechanisms, typically relying on researchers' expertise and experience for semantic interpretation and topic detection. This process is time-consuming, labor-intensive, and prone to inconsistencies due to variations in interpreters' knowledge, affecting the objectivity of topic summarization. These problems become pronounced when analyzing multiple articles or conducting comparative analyses over time.

2. **Limited ability to identify low-frequency, high-value keywords, making latent topic detection difficult:** Traditional text mining methods using statistical frequency focus on topic clusters formed by high-frequency keywords, often overlooking low-frequency keywords and chance nodes that play bridging or transitional roles in the keyword structure. These low-frequency keywords often represent emerging concepts, topic intersections, or contextual shifts, holding significant value for uncovering latent research topics and policy chance information. However, traditional methods struggle to identify and interpret their semantic roles systematically, limiting the efficiency and usefulness of topic exploration.

3. **Difficulty tracking dynamic contexts hinders automating topic-evolution pattern analysis:** When managing cross-temporal texts, such as AI ethics articles from 2022 to 2024, traditional keyword network analysis often requires a manual comparison of keyword structural changes at various time points and cannot effectively or automatically track how topic keywords undergo semantic shifts or experience topic merging and splitting as the context evolves. This limitation hinders researchers' understanding and forecasting of topic evolution trajectories, resulting in analyses without the capacity to present temporal and dynamic characteristics.

4. **Visualization maps are challenging to convert into structured data for inference:** Although keyword network graphs offer a high degree of visual intuitiveness and help reveal thematic contexts and lexical and relational structures in texts, their results are often presented as images. When the number of keyword nodes in topic clusters is high, the clarity and readability of these visuals significantly decrease, leading to blurred outcomes or difficulty in interpretation during advanced analyses (e.g., topic classification, semantic comparison, or cross-validation).

3.5.2. Technical Background: Semantic Comprehension and Topic Extraction in ChatGPT

The ChatGPT LLM is based on the deep learning transformer architecture that Vaswani et al. proposed in 2017. Unlike the widely used LDA, this model undergoes unsupervised pretraining on large-scale textual data to generate high-dimensional semantic embeddings, capturing the syntactic structures and semantic relationships in sentences. Its core self-attention mechanism captures semantic associations and contextual features in the text, transforming textual data into semantic vector representations to infer deep linguistic patterns and latent word relationships. In practical applications, ChatGPT demonstrates capabilities in processing long texts, generating summaries, and performing topic detection.

The experimental results from existing studies indicate that ChatGPT achieves highly accurate content detection and classification tasks, significantly surpassing the current benchmark methods. Notably, ChatGPT displays outstanding performance in zero-shot learning scenarios. The literature has suggested that ChatGPT can directly understand and execute most tasks without any additional training or fine-tuning, with performance typically exceeding that of other mainstream LLMs, demonstrating exceptional generalizability. Moreover, studies have found that, in specific tasks, the performance of ChatGPT surpasses even that of fine-tuned models. This finding highlights the potential of ChatGPT as a foundation model, achieving or exceeding the performance of task-specific trained models without special optimization, highlighting stronger adaptability and broader application prospects [61–64].

This study integrates ChatGPT into the topic detection and summary generation tasks of the keyword networks produced by KeyGraph to apply the semantic analysis potential of KeyGraph fully and overcome the limitations of traditional manual interpretation. When the structure of the keyword network, particularly starting from the red nodes, is expanded layer by layer based on the connection strength and semantic distance with high-frequency black nodes, the corresponding original texts are input into ChatGPT individually. This LLM employs the following steps to conduct semantic interpretation and topic detection processes [64–66]:

1. **Comprehension of keyword network structures and semantic interpretation:** ChatGPT tokenizes the input text, including the original AI ethics articles and translated descriptions of the KeyGraph keyword network structure, and processes it via its multilayer transformer model for deep syntactic and semantic analyses. The built-in attention mechanism in ChatGPT accurately captures complex relationships between tokens and their contextual meaning, constructing a comprehensive, detailed semantic representation. This approach enables the model to understand the meaning of individual tokens and their positions and roles in the keyword network.

2. **Topic identification:** The model identifies frequently recurring keywords and their semantic relationships in the text, grouping them into coherent thematic clusters. Notably, ChatGPT applies its strong contextual reasoning to generate semantically complete and representative thematic descriptions, facilitating the discovery of core concepts in the network structure.

3. **Semantic interpretation and text summarization:** ChatGPT extracts critical insight from text based on semantic logic and generates contextually coherent and concise summaries. Researchers can control the content and length of these summaries using precise prompt engineering (e.g., restricting the summary to the imported text) to meet specific analytical requirements. This control considerably enhances the efficiency of extracting insight from complex network graphs [66,67].

Through these steps, this study expands the keyword network structure constructed by KeyGraph layer by layer, starting from the red nodes and analyzing their associations with the high-frequency black nodes. Then, this study employs ChatGPT to interpret the natural language and summarize the themes of the topic clusters and bridging nodes of the original AI ethics articles. This integrated process significantly enhances the analytical efficiency and accuracy of thematic induction,

realizing the HCI and semantic reciprocal interpretation emphasized by the chance discovery theory and providing a more systematic and operable framework for dynamic topic detection.

3.5.3. Method: Integrating KeyGraph and ChatGPT for Topic Detection

The KeyGraph application emphasizes that mining, understanding, and topic detection should be conducted and interpreted by domain experts to ensure the contextual relevance of the keyword network interpretation. However, given the limited availability of domain experts, this study integrates the KeyGraph keyword network analysis with ChatGPT to enhance the semantic interpretability of detected topics and the ability to determine potential chances by constructing keyword association structures and exploring the topics. The analysis process begins with the KeyGraph algorithm generating a keyword network graph based on the co-occurrence frequency and distance between keywords, identifying potential cross-topic connective chance nodes (red nodes), which serve as the starting points for semantic expansion and topic detection.

Using the red nodes as the starting points for semantic diffusion aligns with the core aim of the chance discovery theory. Although these nodes have a relatively low frequency, they form connections with multiple clusters, displaying cross-topic bridging characteristics and capturing hidden information more effectively. Furthermore, although these nodes may not represent the focus of the texts, they can reveal potential topics at the boundaries of the keyword network, offering high informational value and strategic significance.

For example, starting from the red node R1, a strong co-occurrence relationship exists between R1 and the high-frequency black node T4, which connects to other high-frequency black nodes (e.g., T1 and T3), with T3 linking to T2. Notably, T2 forms a closed-loop connection with both T1 and T4. This hierarchical node expansion enables the construction of a structured semantic diffusion path, delineating the internal relationships and boundaries of a topic (Figure 4).
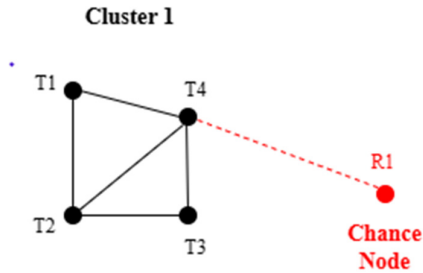

Figure 4. Co-occurrence relationship mapping in the association graph.

This study integrates the ChatGPT model with an HCI mechanism to apply the semantic diffusion structure to semantic inference and topic interpretation tasks, enhancing the semantic accuracy and interpretative consistency of topic understanding. The keyword clusters identified by KeyGraph are reconstructed into logically coherent semantic diffusion paths, with red nodes serving as the starting points for interpretation. These paths are input into ChatGPT, which is guided to perform semantic judgment and topic mining based on the provided semantic diffusion path. Figure 4 presents an input example, the R1 semantic diffusion path, where the red node is R1 and the high-frequency black nodes in Cluster 1 include T1, T2, T3, and T4.

3.6. R1 Semantic Diffusion Path

The co-occurrence structure in Cluster 1 consists of the following edges: {(R1, T4), (T4, T1), (T4, T2), (T4, T3), (T1, T2), (T2, T3)}. Although this method integrates the KeyGraph algorithm with ChatGPT for topic detection and semantic interpretation, practical implementation still faces challenges. As an LLM, ChatGPT's outputs may exhibit risks, including semantic hallucination, topic

ambiguity, or semantic overextension. This study adopts a dual-layer control strategy to mitigate these biases effectively [68].

The first layer employs a refined prompt engineering mechanism to guide ChatGPT to focus on topic inference in a specific context, ensuring that the generated semantics rely solely on the semantic diffusion path and imported textual data. In this process, researchers initially use role-playing prompt strategies to assign ChatGPT a particular identity or perspective. This initial setting helps to converge the model's understanding of the specific domain or context before topic inference, aligning its behavior more closely with expectations and enhancing the efficiency and accuracy of tasks. Clear task assignments are given to direct ChatGPT to perform various functions, including topic classification, semantic interpretation, or summary generation. Throughout the process, researchers control the scope of contextual input, restricting the topic summary content to the imported textual data, compensating for ChatGPT's potential limitations in domain-specific knowledge and reducing the influence of semantic hallucinations on the interpretative results.

The second layer involves a repetitive manual review. The topic interpretation results generated by ChatGPT are subject to manual examination and validation by researchers. The aim is not to perform a word-for-word comparison of the semantic interpretations but to assess and revise the logical coherence and consistency of the topic summaries to ensure semantic logic and interpretative consistency. Researchers can adjust the algorithm parameters based on the review outcomes to conduct the next iteration of analyses, progressively refining and optimizing the semantic structure and topic detection results.

This risk management mechanism balances generative capability with interpretative reliability, illustrating the dynamic optimization loop of data-driven insight combined with knowledge reasoning, as described in the double helix model, facilitating incremental knowledge construction. By integrating keyword diffusion structures with HCI workflows, this study deepens the understanding and identification of latent topics. This approach enhances the sensitivity of topic detection and the accuracy of interpretation, improving the overall explainability of semantic analysis and its capacity to support decision-making.

This study conducts topic detection using keyword network graphs and keyword clusters generated by KeyGraph, supplemented by ChatGPT for topic interpretation and summary generation. Researchers interpret the results and make logical judgments, adjusting the model parameters to guide analysis iterations. This iterative process establishes an HCI mechanism for exploring topics and constructing knowledge.

Although the combined use of KeyGraph and generative AI demonstrates strong potential for topic detection and semantic interpretation, its practical implementation still faces challenges. As an LLM, ChatGPT's outputs may involve risks, including semantic hallucination, topic ambiguity, or excessive semantic expansion. This study adopts a dual-layer control strategy to mitigate these biases. First, the generated semantics are constrained via refined prompt engineering to rely solely on the semantic diffusion paths and imported textual data. Second, a repeated manual review is conducted to enhance semantic consistency and logical thematic accuracy. This risk management mechanism balances generative capability with interpretative reliability, improving the overall explainability of the semantic analysis and decision support.

4. Result Analysis

This section presents a systematic analysis employing the Polaris visualization tool combined with the KeyGraph algorithm to analyze English-language articles related to AI ethics. The dataset comprises representative articles published between 2022 and 2024 by reputable academic institutions, news media, and nonprofit organizations providing comprehensive discourse. Eight representative articles for each year were selected for analysis to deepen the understanding of the evolving context of AI ethics issues and construct a structured keyword network supporting topic detection and interpretation.

Due to the numerous high-frequency black nodes in the initially generated keyword network, the analysis focused on those high-frequency black nodes with explicit connections (i.e., co-occurrence relationships represented by solid black edges). Isolated high-frequency black nodes without connections to other nodes were excluded from the keyword network. This approach improves the readability and relational strength of the keyword network and helps identify potential emerging chance trends.

Next, a structural analysis was conducted on the keyword network formed by these high-frequency black nodes, and keyword clusters were manually delineated. Clustering originated from chance nodes (red nodes) and expanded outward, with the number of linked high-frequency keywords limited to about six to seven per cluster. This process was supported by a keyword review and thematic convergence procedures to ensure semantic coherence and appropriateness within clusters. Following the preliminary clustering, the semantic diffusion paths for each keyword cluster were individually input into ChatGPT for in-depth topic detection and semantic interpretation, revealing the core topics embedded in each cluster.

Cross-cluster thematic integration analyses were conducted to explore latent common topics and interwoven keyword structures of strongly related clusters. All outputs underwent manual inspection to ensure the readability, reliability and logical consistency of the results. This process illustrates the HCI interpretative loop in the double helix model, where dynamic cycles of semantic network visualization, topic clustering, and generative AI-assisted interpretation collectively enhance the depth and explanatory power of topic detection.

4.1. Yearly Analysis of Topic Evolution and Keyword Structures (2022–2024)

The eight articles from 2022 contained 2,595 distinct terms; hence, this study set a parameter of 75 high-frequency keywords as black nodes, connected by 135 solid black edges depicting the co-occurrence network of these keywords to present the core associative structure of AI ethics topics. Additionally, four red nodes were designated as the potential chance discovery nodes. Under these conditions, a keyword network was generated using *automaker*, *dignity*, *behavior*, and *statistical* as the red nodes, highlighting the key themes in the 2022 AI ethics articles (Figure 5). The following bullet points summarize the thematic content.

- **Cluster A-1:** The semantic cluster around the red node automaker focuses on the implementation of autonomous driving technology and the ethical challenges faced by AI in automotive applications. This red node extends through its connection to *self* to include the keywords *based*, *car*, *driver*, *vehicles*, and *autonomous*, outlining application scenarios involving HCI. The keywords *driver*, *task*, and *autonomous* intertwine, reflecting issues of responsibility allocation and control authority. In situations where automated and manual control are combined, the attribution of responsibility for accidents (whether borne by the driver or system) requires further clarification via regulatory frameworks and technical design. Furthermore, task transparency and the interface design are also critical. For example, whether drivers can quickly grasp the operational status and decision rationale of the system directly affects their safety judgments and behavioral responses. Establishing trust and risk perception cannot be overlooked. An insufficient HCI design and information transmission may cause driver overtrust or erroneous reliance, increasing safety risks. Overall, the keyword structure emphasizes several topics, including the behavior prediction of autonomous technologies, system safety, and user responsibility attribution.

- **Cluster A-2:** The red node *behavior* forms a keyword network related to AI risk prediction, system deployment, and ethical practices. Through its connections to *consequences*, the network gradually expands to include the keywords *risk*, *privacy*, *discrimination*, *design*, and *capabilities*, reflecting the multifaceted and uncertain outcomes of AI system behavior. Notably, *discrimination* is intertwined with *risk*, indicating that failure to address data sources and algorithmic bias properly in real-world applications may reinforce existing societal inequalities and trigger ethical crises of systemic discrimination. The association between *design* and

foundational highlights the need to judiciously consider fundamental principles and ethical values during the initial stages of AI development. Overall, this cluster maps the potential externalities that may arise during AI deployment, emphasizing that developers must assume the corresponding responsibility for the potential social and ethical consequences of system behavior.

- **Cluster A-3:** The keyword network extended from the red node *statistical* focuses on the computational logic and algorithmic architecture of AI systems. The strong co-occurrence relationships, with the keywords *computational*, *learning*, *machine*, *critical*, and *implementation*, reveals core problems including statistical biases, risk governance, and explainability in current AI technology. The direct and indirect connections between the keywords *issues*, *concerns*, *ethical*, *implementation*, and *critical* reflect that AI ethics is not merely a conceptual discussion but is involved in the development, design, and deployment stages of AI systems. Furthermore, the connections emphasize that the realization of AI ethics must integrate value judgments and ethical norms as essential foundations for technical practice. This cluster demonstrates the role of ethical issues in institutional frameworks, industrial applications, and technical design, indicating that ethical practice has become a critical factor that cannot be overlooked in the development of responsible technology.

- **Cluster A-4:** The keyword network constructed around the red node *dignity* focuses on human rights protection and ethical principles. This red node displays high co-occurrence frequencies with the keyword *responsibility*, *trust*, *justice*, *transparency*, *principles*, and *ethics*, reflecting that current AI technology developers should assume the corresponding moral responsibilities to avoid problems (e.g., bias, discrimination, and structural inequality). Ensuring the transparency of algorithms and data processing allows users to understand the decision-making logic and behavioral patterns of AI systems, safeguarding human dignity and fundamental rights. The connections between *justice*, *guidelines*, and *harm* highlight the necessity of designing AI ethical frameworks and indicate that the lack of appropriate ethical judgment and operational guidance may harm individuals or society, causing discrimination or unfairness. Overall, this cluster focuses on protecting human rights and strengthening ethical norms and institutional justice as core principles, constructing an AI governance mechanism characterized by social legitimacy and long-term trust.

- **Combined cluster of A-2 and A-3:** The keyword network reveals a significant intersection and complementary structure, highlighting the dual technological and societal dimensions of AI ethics issues. Through red chance nodes, including *bias*, *risk*, *understand*, *issues*, and *implementation*, cross-cluster bridging nodes emerge, uncovering a risk propagation chain that spans from statistical logic to behavioral consequences. Bias often originates from flaws in algorithm design and training data, and further permeates the societal domain after system deployment, leading to concrete and potentially escalating ethical consequences. This analysis indicates that AI ethics challenges must be examined from an integrated, multilayered perspective spanning technical construction and societal influence. Accordingly, ethical practice in AI should focus on identifying and mitigating potential ethical risks during the early stages of technological development (e.g., data preprocessing and model training). A comprehensive ethics governance framework encompassing bias detection, transparency enhancement, and regulatory mechanisms must be promoted to ensure responsible and sustainable AI applications.

- **Combined cluster of A-2 and A-4:** The analysis reveals that AI behavior must be guided and constrained by ethical principles to prevent harm to human dignity and privacy, enabling the deployment of trustworthy and responsible AI. The behavioral logic of AI systems should be grounded in human rights protection and ethical values, with corresponding regulations (e.g., bias detection and privacy protection standards) introduced during the early design stages to ensure legitimacy and credibility during deployment. The consequences of AI behavior (e.g., bias and privacy infringement) must be directed by ethical principles and implemented through

technical practice. This interactive relationship emphasizes that ethics should not be treated as an external constraint to technology but as an internal structure embedded throughout the life cycle of AI design, development, and application, advancing responsible and human-centered AI development. This perspective aligns with the discourse in the 2022 AI ethics articles, including *The 2022 AI Index: AI's Ethical Growing Pains* and *AI Ethics and AI Law: Grappling with Overlapping and Conflicting Ethical Factors Within AI*, and identifies the integration of bias management and privacy protection into a unified ethical framework as an emerging research chance.

- **Combined cluster of A-2, A-3, and A-4:** The semantic co-construction of these three clusters reveals that AI ethics challenges cannot be viewed as problems confined to a single level. The behavioral risks of AI systems (e.g., technical bias, discriminatory outcomes, and privacy infringement) are closely linked to their underlying statistical construction logic, indicating that once deployed, AI may produce irreversible and substantive ethical consequences. If such consequences are not addressed through institutionalized ethical safeguards that ensure prevention and accountability, AI technology risks losing social trust and legitimacy. Moreover, ethical AI practice must adopt a cross-level integration approach to address these challenges, spanning from model training and system deployment to institutional regulation, constructing a full-process ethical governance framework based on the triad of technology, behavior, and values. This structure is critical for preserving human dignity and developing trustworthy and responsible AI.
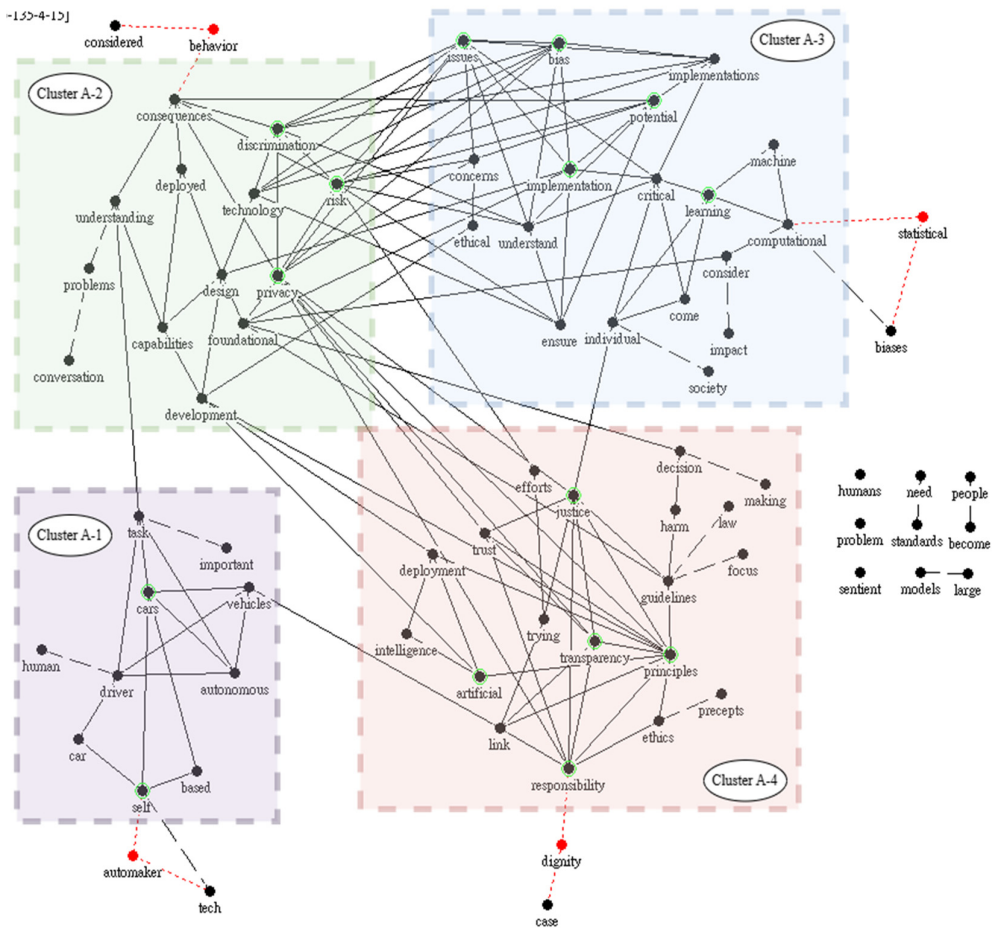

Figure 5. KeyGraph co-occurrence network of AI ethics articles in 2022.

The eight articles from 2023 contain 1,606 distinct terms. Based on customized parameters, 65 high-frequency keywords were selected as black nodes and were connected by 85 solid black lines to construct the co-occurrence keyword network. Two red nodes were designated as potential chance discovery nodes. Under these conditions, a keyword network was generated with *machine* and *misuse* as the red nodes, revealing the structural configuration of keywords related to AI ethics and highlighting the emerging chances (Figure 6). In this network, the red node *machine* forms a keyword structure divided into two subsemantic clusters, connected via the terms *trained* and *develop*. The thematic clusters are summarized as follows:

- **Cluster B-1:** With *trained* as the primary node, the network extends to the keyword *data* and further expands to the keywords *models*, *privacy*, and *customer*, reflecting early-stage concerns in AI development regarding the legitimacy of data sources and the protection of user information. The node *models* branches out to include *intelligence*, *ChatGPT*, *generative*, and *bias*, indicating attention to the algorithmic biases embedded in generative AI models (e.g., ChatGPT). The bidirectional links between *privacy*, *customer*, and *system* highlight ethical considerations regarding user privacy and data security in AI application contexts. The connections between *system* and the keywords *customer*, *create*, and *generative* reveal the interplay between system design and generative technology in practice, raising concerns about technological transparency and ethical accountability. The keyword *artificial* is linked to *intelligence*, *lead*, and *ChatGPT*, forming a semantic structure centered on AI model generation and leadership in application. This cluster reveals deep ethical concerns related to the legitimacy of data usage, model bias, privacy protection, and user participation during the training and deployment phases of AI systems.

- **Cluster B-2:** The primary node *develop* connects with *systems* and *human*, revealing the bidirectional relationship of HCI in technological construction. *Systems* further expands to *make* and *decisions*, reflecting the role AI systems play in decision-making processes. *Decisions* links to *making*, *humans*, and *believe*, forming a cluster centered on how AI decision-making influences human beliefs. *Technology* co-occurs with the terms *ethics* and *concerns*, indicating heightened attention to ethical regulations and institutional policies during AI development. Through the node *concerns*, the keyword *ethics* connects to *potential*, *business*, and *responsibility*, outlining the importance businesses place on ethical risks and responsibilities when applying AI technology. Overall, this semantic group illustrates the institutional and ethical challenges faced during AI development, emphasizing the importance of bias governance, technical regulation, and establishing user trust.

- **Cluster B-3:** With *misuse* as the red node, the initial connection to *government* further extends to *industry* and *society*, forming a semantic cluster focusing on institutional roles. The node *industry* links to *insurance*, which connects to *using*, *policy*, and *responsible*, highlighting an ethical discourse focused on risk transfer mechanisms and institutional responsibility. The keyword *policy* is a central node connecting *responsible*, *insurance*, and *using*, indicating that policy should address AI misuse risks via clear responsibility allocation, technical application guidelines, and industry-level risk management, especially concerning privacy protection and social impact. The keywords *ethical*, *ensure*, *responsible*, and *using* are closely interlinked, underscoring that ethical principles must be embedded in technical usage and institutional regulation. These principles, when supported by accountability structures and protective measures, can mitigate risks of misuse, particularly in areas related to data privacy and societal consequences. The connection between *impact* and *society* further indicates the potential and far-reaching effects of technological misuse on social structures. Overall, this semantic cluster illustrates that AI ethical principles should be integrated into institutional design and technological application processes and that clear accountability and regulatory mechanisms are critical for reducing the potential negative influences of AI misuse on societal systems.

- **Combined cluster of B-1 and B-2:** These two clusters, centered on the red node machine, focus on model training and system development, respectively, revealing, through the lens of practical

application, the crucial ethical challenges spanning the AI life cycle, from data training and system development to deployment. Both clusters emphasize data ethics (e.g., privacy and bias) and the governance of potential negative influences of AI systems on society and humanity, including decision-making influence and responsibility attribution. Together, these semantic clusters reveal that the core of AI ethics lies in the technology itself and, more critically, in the processes of interaction between AI, humans, and society, particularly regarding risk management and the realization of accountability. The clusters collectively emphasize that achieving a vision of AI development that balances innovation and responsibility requires the parallel construction of responsible governance mechanisms throughout the innovation process.

- **Combined cluster of B-2 and B-3:** In the KeyGraph keyword network, Clusters 2 and 3 are centered on the keywords *develop* and *misuse*, respectively, illustrating an ethical link from AI technology development to its potential misuse. The keyword structures revealed by these two clusters reflect that AI ethics challenges originate from individual acts of technical development and extend across broader societal institutions and governance dimensions. The ethical risks posed by AI technology can be effectively addressed only by constructing an integrated accountability framework encompassing development, deployment, and misuse prevention, ensuring that advancement contributes to positive and sustainable social value.

- **Combined cluster of B-1, B-2, and B-3:** These three semantic clusters correspond to three stages of AI ethical risk, model training, system development, and actual misuse along with social impact, respectively, forming a progressive chain from ethical considerations to governance responses. The keyword structures reveal a trajectory that begins with micro-level concerns, including data bias and generative misinformation, and extends to challenges of decision-making and ethical design during the development process. The structures indicate misuse risks and governance responsibility at the societal level. This progression reflects that AI ethics issues are not isolated incidents but constitute a foreseeable and preventable chain of ethical risks. An integrated ethical framework must be established that encompasses data governance, technical design, and misuse prevention, enabling the realization of an AI development vision guided by social values to address multilevel challenges (e.g., bias, manipulation, and misuse).
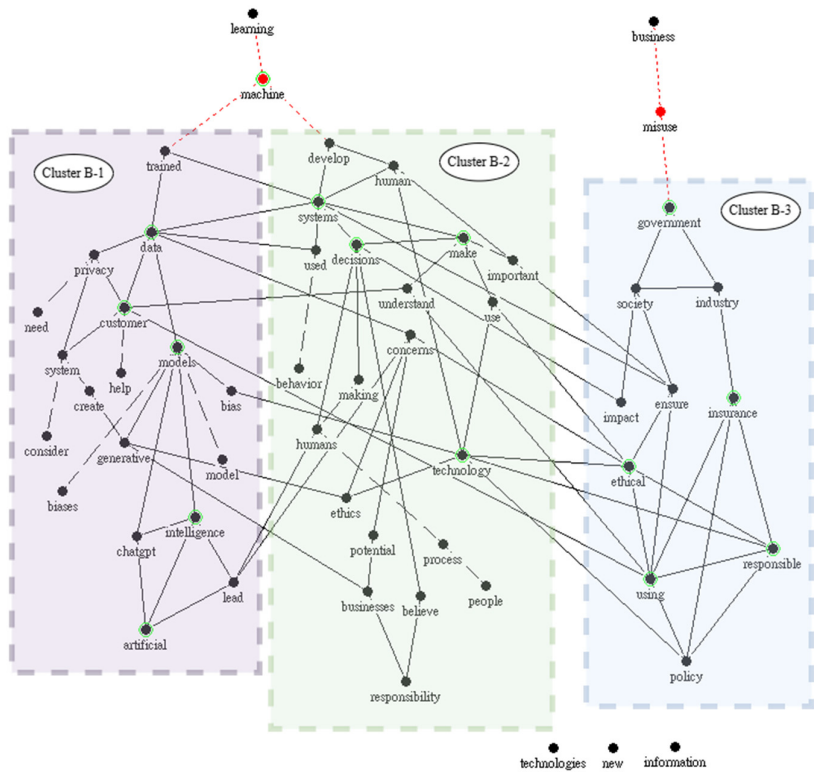
Figure 6. KeyGraph co-occurrence network of AI ethics articles in 2023.

The eight articles from 2024 encompass 2,612 distinct terms. Based on the customized parameters, 60 high-frequency keywords were selected as black nodes and were connected by 95 solid black lines to construct the co-occurrence network of keywords. Two red nodes were designated as potential chance discovery nodes. Under these conditions, keyword networks were generated using *media* and *security* as the red nodes, revealing the relationships between content production, technological transparency, and institutional responsibility in AI ethics issues (Figure 7). The bullet points below summarize the thematic clusters.

- **Cluster C-1:** With *media* as the red node, the network connects to *social*, which links to *content*, *genAI*, and *used*, revealing that generative AI technology has been widely integrated into social platforms and public communication spaces. The connection between *social* and *ethical*, further extending to *risks* and then *deployment*, *challenges*, *technology*, and *responsible*, indicates that societal concerns have shifted beyond technical applications to the ethical risks and responsibility attribution involved in deployment processes, especially regarding misinformation, information manipulation, and bias problems arising in social media environments. The keywords *digital*, *technology*, *innovation*, *industry*, and *development* converge at the nodes *essential*, *become*, and *important*, demonstrating that generative AI has become a core driving force behind contemporary digital innovation and industrial transformation, with its ethical challenges escalating into systemic problems. Overall, this semantic cluster highlights that AI ethics attention has moved toward ethical challenges triggered by the application of generative AI in social and media contexts, emphasizing the importance of responsible technological deployment in these settings.

- **Cluster C-2:** The red node *security* co-occurs with *technologies* and extends to *data*, *models*, and *training*, forming a semantic cluster. The keyword *biases* forms a triangular co-occurrence structure with these three terms, indicating that data sources and processing methods underpin AI system security, and that biases hidden within training data influence model behavior, representing the intersection of ethics and security. *Transparency* and *privacy* connect through *technologies* and further co-occur with *regulatory*, reflecting that AI ethics discourse has reached institutional dimensions and emphasizing the reliance on and necessity of regulatory mechanisms for system transparency and privacy protection. Via *ensure* and *tools*, the keyword *decision* links to *generative* and *ChatGPT* and is associated with *businesses* and *trust*, revealing the critical role of explainability and trust mechanisms in generative AI decision-making processes in corporate and societal applications. Overall, this keyword network reveals the interdisciplinary interconnection of AI ethics issues in 2024, providing a structured analytical perspective for technology development, policy regulation, and industry practice.

- **Combined cluster of C-1 and C-2:** The integration of these two clusters highlights the increasingly multilayered and interdisciplinary complexity of AI ethics issues in 2024. The AI ethical themes are no longer confined to a single domain but require addressing systemic governance challenges while promoting AI development, especially generative AI, in the fields of digital content and media. These challenges include core concerns, including data privacy, algorithmic bias, social trust, lack of transparency, and regulatory compliance. The importance of achieving trustworthy and responsible AI governance via the collaborative operation of social and technical dimensions is emphasized, with collective responsibility shared by developers, businesses, policymakers, and civil society. The current AI ethical frameworks must be established under risk contexts characterized by uncertainty and the undiscovered and unknown, guiding AI development toward a more legitimate and sustainable future.
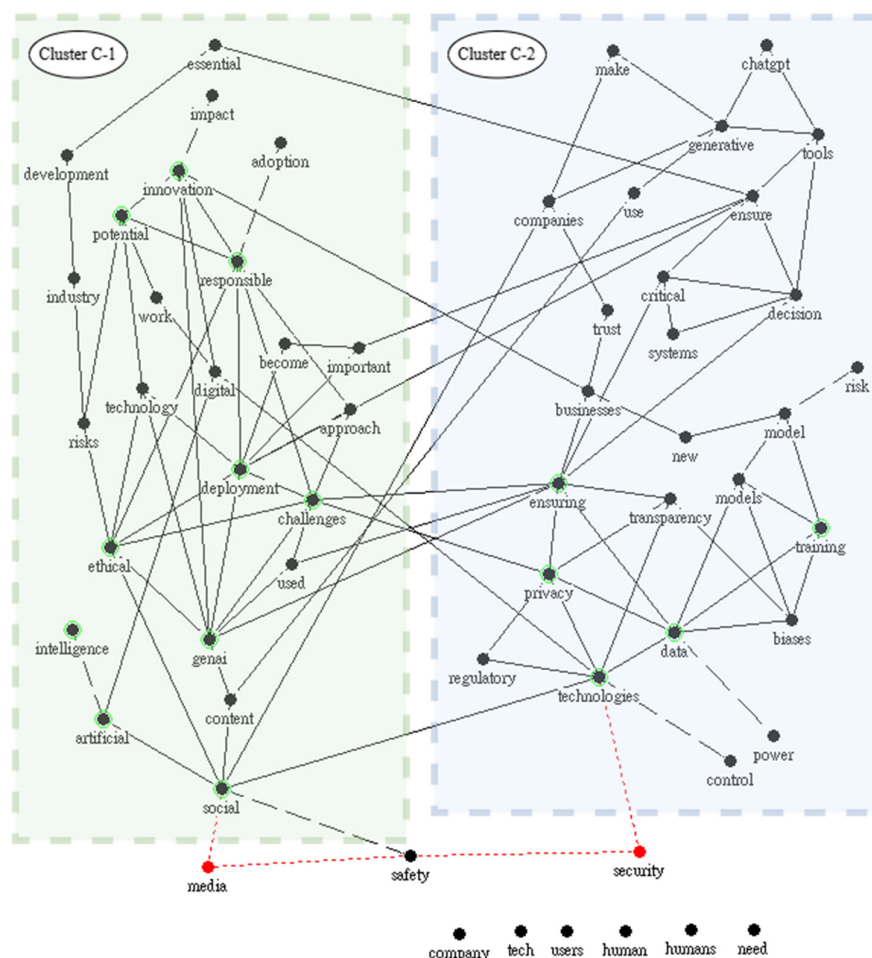
Figure 7. KeyGraph co-occurrence network of AI ethics articles in 2024.

4.2. Integrative Analysis and Trend Summary

The keyword network analysis results from 2022 to 2024 reveal that AI ethics issues exhibit an evolving trend that progressively expands from technical implementation to institutional regulation and social responsibility. Moreover, the keyword structures demonstrate increasingly complex and interdisciplinary governance concerns each year.

The 2022 AI ethics discourse emphasized the inseparable relationship between technical implementation, statistical construction logic, and the resulting social consequences. Without institutionalized ethical safeguards for prevention and accountability, AI technologies risk losing social trust and legitimacy. Therefore, the governance framework, spanning from technology development to institutional regulation, must incorporate ethical practices, addressing core problems (e.g., bias, transparency, and human rights protection) to promote the sustainable and equitable application of AI, achieving trustworthy and responsible AI development.

The 2023 AI ethics keyword structure displays a characteristic chain of ethical risk semantics linking training data, technical design, and governance policies. This structure indicates that AI ethics challenges are not isolated, discrete incidents but span from micro-level issues (e.g., data bias and generative misinformation) to decision-making effects and ethical design in the technical development process, culminating in systemic challenges of misuse risk and governance responsibility at the societal level. This keyword structure reflects the core perspective in 2023, emphasizing the establishment of an integrated ethical framework encompassing data governance, technical design, and misuse prevention. This framework ensures that responsible governance mechanisms are constructed while developing and innovating AI technology, achieving a

development vision guided by social values and ensuring that AI deployment serves human well-being and social value.

The 2024 AI ethics issues present a multilayered and interdisciplinary complexity. The scope of AI ethics has expanded from single technical applications to an integrated framework encompassing technology deployment, data sources, institutional regulation, and social trust. Although the rapid development of generative AI has promoted digital innovation and industrial upgrading, it has also introduced unprecedented ethical risks and governance pressures. As a result of systemic challenges (e.g., algorithmic bias, privacy infringement, lack of system transparency, and regulatory lag), future AI governance must be based on the collaborative operation of technological and social dimensions, emphasizing the collective responsibility of developers, businesses, policymakers, and civil society. Moreover, under the uncertain and rapidly evolving risks posed by generative AI, establishing a forward-looking ethical framework addressing uncertainty and potential risks is a critical direction indicated by the 2024 keyword network, guiding AI development toward a more legitimate and sustainable future.

5. Conclusions

This study adopts the KeyGraph algorithm as its core analytical method to examine the evolving semantic structures in AI ethics discourse. By constructing keyword co-occurrence networks, this work offers an in-depth understanding of the multilayered thematic architecture and conceptual transitions in this domain. The systematic semantic analysis of representative articles from 2022 to 2024 revealed that the KeyGraph algorithm identifies core nodes and semantic linkages over time. The clustering of keywords and identification of chance nodes facilitated the mapping of the developmental trajectory and internal logic of AI ethics over time.

This analysis of time-specific keyword networks demonstrates that the evolution of AI ethics discourse is not static. Instead, the evolution dynamically shifts in response to the interplay between technological advancement, regulatory change, and societal interaction, prompting continuous semantic reconstruction and thematic reorientation. The findings reveal latent interconnections between technology, law, and society, offering a novel perspective for understanding the complexity of AI ethics. These results validate the effectiveness of KeyGraph in processing large-scale unstructured textual data, extracting critical information, and constructing meaningful keyword networks for semantic exploration.

KeyGraph's text-mining results were integrated with the generative AI ChatGPT-based text summarization technique to enhance the analytical rigor and accuracy of the study. This integration allows for guided semantic interpretation and topic detection in co-occurrence-driven keyword clusters. The combined approach strengthens the depth, accuracy, and interpretive consistency of latent topic detection from complex articles. The cross-validation between the keyword co-occurrence network analysis and generative language models improves the scientific robustness and credibility of our understanding of AI ethics discourse and its semantic evolution.

The proposed method has broad potential for cross-domain applications. Beyond the AI ethics domain, integrating the KeyGraph algorithm with temporal analysis strategies can be extended to issue exploration and trend tracking in various fields (e.g., social media, e-commerce, and news media). The dynamic analysis of keyword node structures and co-occurrence intensities over time can reveal shifts in public attention and capture the evolving focus of technology and topics with greater precision. This approach enables the early identification of emerging concepts that may pose potential risks or offer innovative value (e.g., market reactions to new products, the progression of societal hotspots, or changes in user sentiment).

Such dynamic tracking capabilities offer valuable insight for corporate decision-making, public policy formulation, and academic research, supporting more accurate forecasting and timely responses to evolving challenges. Overall, this study demonstrates the applicability of the keyword co-occurrence network analysis in AI ethics and proposes a transferable framework that can be extended to other interdisciplinary issues. With the continued refinement of the KeyGraph algorithm

and the integration of advanced LLMs for summarization, this approach is promising for advancing a deeper understanding and anticipatory reflection of the ethical implications of AI technology.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
HCI	Hyper-converged Infrastructure
LDA	Latent Dirichlet allocation
LLM	Large language model
SDGs	Sustainable Development Goals

References

1. Shetty, D.K.; Arjunan, R.V.; Cenitta, D.; Makkithaya, K.; Hegde, N.V.; Bhatta B, S.R.; Salu, S.; Aishwarya, T.R.; Bhat, P.; Pullela, P.K. Analyzing AI Regulation through Literature and Current Trends. *J Open Innov: Technol Mark Complex* **2025**, *11*, 100508.
2. Tallberg, J.; Lundgren M.; Geith, J. AI Regulation in the European Union: Examining Non-State Actor Preferences. *Bus Politics* **2024**, *26*, 218-239.
3. Ong, J.C.L.; Chang, S.Y.; William, W.; Butte, A.J.; Shah, N.H.; Chew, L.S.T.; Liu, N.; Doshi-Velez, F.; Lu, W.; Savulescu, J.; Ting, D.S.W. Ethical and Regulatory Challenges of Large Language Models in Medicine. *Lancet Digit Health* **2024**, *6*, e428-e432.
4. Huang, C.; Zhang, Z.; Mao, B.; Yao, X. An Overview of Artificial Intelligence Ethics. *IEEE Trans Artif Intell* **2023**, *4*, 799-819.
5. Tabassum A, Elmahjub, E.; Padela, A.I.; Zwitter, A.; Qadir, J. Generative AI and the Metaverse: A Scoping Review of Ethical and Legal Challenges. *IEEE Open J Comput Soc* **2025**, *6*, 348-359.
6. Taeihagh, A. Governance of Generative AI. *Policy Soc* **2025**, *44*, 1–22.
7. Morley, J.; Elhalal, A.; Garcia, F.; Kinsey, L.; Mökander, J.; Floridi, L. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds Mach* **2021**, *31*, 239-256.
8. Mittelstadt, B.D. Principles alone cannot guarantee ethical AI. *Nat Mach Intell* **2019**, *1*, 501-507.
9. Cath, C.; Wachter, S.; Mittelstadt, B.; Taddeo, M.; Floridi, L. Artificial intelligence and the 'Good Society': The US, EU, and UK Approach. *Sci Eng Ethics* **2018**, *24*, 505-528.
10. Ohsawa, Y.; McBurney, P. *Chance Discovery.* Springer-Verlag: Berlin/Heidelberg, Germany, 2003.
11. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J Mach Learn Res* **2003**, *3*, 993-1022.
12. Vayansky, I.; Kumar, S.A.P. A Review of Topic Modeling Methods. *Inf Syst* **2020**, *94*, 101582.
13. Barde, B.; Bainwad, A.M. An Overview of Topic Modeling Methods and Tools. In Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE: Piscataway, NJ, USA, 2017; pp. 745–750.
14. Sayyadi, H.; Raschid, L. A Graph Analytical Approach for Topic Detection. *ACM Trans Internet Technol* **2013**, *13*, 1-23.
15. Hayashi, T.; Ohsawa, Y. Information Retrieval System and Knowledge Base on Diseases Using Variables and Contexts in the Texts. *Procedia Comput Sci* **2019**, *159*, 1662-1669.

16. Wang, J.; Lai, J.Y.; Lin, Y.H. Social Media Analytics for Mining Customer Complaints to Explore Product Opportunities. *Comput Ind Eng* **2023**, *178*, 109104.

17. Guler N.; Kirshner, S.N.; Vidgen, R. A Literature Review of Artificial Intelligence Research in Business and Management Using Machine Learning and ChatGPT. *Data Inf Manag* **2024**, *8*, 100076.

18. Nissen, H.E. Using Double Helix Relationships to Understand and Change Information Systems. *Informing Sci Int J Emerg Transdiscip J* **2007**, *10*, 21-62.

19. Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; Srikumar, M. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *Berkman Klein Center Res Publ* **2020**, *2020*, 1.

20. de Fine Licht, K. Resolving Value Conflicts in Public AI Governance: A Procedural Justice Framework. *Gov Inf Q* **2025**, *42*, 102033.

21. Fruchter, R.; Ohsawa, Y.; Matsumura, N. Knowledge Reuse through Chance Discovery from an Enterprise Design-Build Enterprise Data Store. *New Math Nat Comput* **2005**, *1*, 393-406.

22. Buolamwini, J.; Gebru, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; Friedler, S.A., Wilson, C., Eds.; PMLR: New York, NY, USA, 2018; Volume 81, pp. 77–91.

23. Njiru, D.K.; Mugo, D.M.; Musyoka, F.M. Ethical considerations in AI-based user profiling for knowledge management: A critical review. *Telemat Inform Rep* **2025**, *18*, 100205.

24. Luomala, M.; Naarmala, J.; Tuomi,V. Technology-Assisted Literature Reviews with Technology of Artificial Intelligence: Ethical and Credibility Challenges. *Procedia Comput Sci* **2025**, *256*, 378-387.

25. Ohsawa, Y. Data Crystallization: A Project beyond Chance Discovery for Discovering Unobservable Events. In *Proceedings of the 2005 IEEE International Conference on Granular Computing*, Beijing, China, 25–27 July 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 1, pp. 51–56.

26. Holzinger, A. Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together? In *Availability, Reliability, and Security in Information Systems and HCI*; Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8127.

27. Ohsawa, Y.; Fukuda, H. Chance discovery by stimulated groups of people: Application to understanding consumption of rare food. *J Contingencies Crisis Manag* **2002**, *10*, 129-138.

28. Ko, N.; Jeong, B.; Choi, S.; Yoon, J. Identifying Product Opportunities Using Social Media Mining: Application of Topic Modeling and Chance Discovery Theory. *IEEE Access* **2018**, *6*, 1680-1693.

29. Ohsawa, Y. Chance Discoveries for Making Decisions in Complex Real World. *New Gener Comput* **2002**, *20*, 143–163.

30. Ohsawa, Y.; Nishihara, Y. *Innovators' Marketplace: Using Games to Activate and Train Innovators*; Springer: Berlin/Heidelberg, Germany, 2012.

31. Ho, T.B.; Nguyen, D.D. *Chance Discovery and Learning Minority Classes. New Gener Comput* **2003**, *21*, 149–161.

32. Ohsawa, Y.; Tsumoto, S. *Chance Discoveries in Real World Decision Making: Data-Based Interaction of Human Intelligence and Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 30.

33. Ohsawa, Y.; Nara, Y. Modeling the Process of Chance Discovery by Chance Discovery on Double Helix. In Proceedings of the AAAI Fall Symposium on Chance Discovery; AAAI Press: Arlington, VA, USA, 2002; pp. 33–40.

34. Wang, H.; Ohsawa, Y. Nishihara, Y. Innovation Support System for Creative Product Design Based On Chance Discovery. *Expert Syst Appl* **2012**, *39*, 4890–4897.

35. Wang, H.; Ohsawa, Y. Idea discovery: A Scenario-Based Systematic Approach for Decision Making in Market Innovation. *Expert Syst Appl* **2013**, *40*, 429–438.

36. Yang, S.; Sun, Q.; Zhou, H.; Gong, Z.; Zhou, Y.; Huang, J. A Topic Detection Method Based on KeyGraph and Community Partition. In Proceedings of the 2018 International Conference on Computing and Artificial Intelligence (ICCAI 2018), Chengdu, China, 12–14 May 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 30–34.

37. Ohsawa, Y. KeyGraph as Risk Explorer in Earthquake–Sequence. *J Contingencies Crisis Manag* **2002**, *10*, 119–128.

38. Ohsawa, Y.; Benson, N. E.; Yachida, M. KeyGraph: automatic indexing by co-occurrence graph based on building construction metaphor. In Proceedings IEEE International Forum on Research and Technology Advances in Digital Libraries, 1998; 12-18.

39. Wanchia, K.; Yufei, J.; Hsinchun, Y. Discovering Emerging Financial Technological Chances of Investment Management in China via Patent Data. *Int J Bus Econ Aff* **2020**, *5*, 1–8.

40. Geum, Y.; Kim, M. How to Identify Promising Chances for Technological Innovation: Keygraph-Based Patent Analysis. *Adv Eng Inform* **2020**, *46*, 101155.

41. Sakakibara, T.; Ohsawa, Y. Gradual-Increase Extraction of Target Baskets as Preprocess for Visualizing Simplified Scenario Maps by KeyGraph. *Soft Comput*, **2007**, *11*, 783-790.

42. Kim, K.-J.; Jung, M.-C.; Cho, S.-B. KeyGraph-Based Chance Discovery for Mobile Contents Management System. *Int J Knowl-Based Intell Eng Syst*, **2007**, *11*, 313–320.

43. Perera, K.; Karunarathne, D. KeyGraph and WordNet Hypernyms for Topic Detection. In Proceedings of the 2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chonburi, Thailand, 22–24 July 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 303–308.

44. Manning, C. D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*. Cambridge University Press: Cambridge, UK, 2008.

45. Liu, B. *Sentiment Analysis and Opinion Mining*. *Synth Lect Hum Lang Technol* **2012**, *5*, 1–167.

46. Beliga, S.; Meštrović, A.; Martinčić-Ipšić, S. An overview of graph-based keyword extraction methods and approaches. *J Inf Organ Sci*, **2015**, *39*, 1–20.

47. Pan, R.C.; Hong, C.F.; Huang, N. Hsu, F.C.; Wang, L.H.; Chi, T.H. One-Scan KeyGraph Implementation. The 3rd Conference on Evolutionary Computation Applications & 2005 International Workshop on Chance Discovery, Taiwan, 2005.

48. Nezu Y.; Miura, Y. Extracting Keywords on SNS by Successive KeyGraph. In 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), Baltimore, MD, USA, 2020; 9–11 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 997–1003.

49. Jobin, A.; Ienca, M.; Vayena, E. The Global Landscape of AI Ethics Guidelines. *Nat Mach Intell* **2019**, *19*, 389-399.

50. Mittelstadt B.; Allo P.; Taddeo M.; Wachter S.; Floridi L. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society* **2016**, *3*, 1-21.

51. Dignum, V. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way.* In Artificial Intelligence: Foundations, Theory, and Algorithms; Springer: Cham, Switzerland, 2019.

52. Okazaki, N.; Ohsawa, Y. Polaris: An Integrated Data Miner for Chance Discovery. In Proceedings of The Third International Workshop on Chance Discovery and Its Management, Crete, Greece, 2003.

53. Sayyadi, H.; Hurst, M.; Maykov, A. Event Detection and Tracking in Social Streams. *Proc Int AAAI Conf Web Soc Media* **2009**, *3*, 311–314.

54. Jo, Y.; Lagoze, C.; Giles, C.L. Detecting Research Topics via the Correlation between Graphs and Texts. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '07), San Jose, CA, USA, 12–15 August 2007; ACM: New York, NY, USA, 2007; pp. 370–379.

55. Lozano, S.; Calzada-Infante, L.; Adenso-Díaz, B.; García, S. Complex Network Analysis of Keywords Co-Occurrence in the Recent Efficiency Analysis Literature. *Scientometrics* **2019**, *120*, 609–629.

56. Zhou, Z.; Zou, X.; Lv, X.; Hu, J. Research on Weighted Complex Network Based Keywords Extraction. In Proceedings of the 7th International Conference on Advanced Data Mining and Applications (ADMA 2013), Wuhan, China, 14–16 May 2013; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8229, pp. 442–452.

57. Grimmer, J.; Stewart, B.M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Polit Anal* **2013**, *21*, 267–297.

58. Firoozeh, N.; Nazarenko, A.; Alizon, F.; Daille, B. Keyword extraction: Issues and methods. *Nat Lang Eng* **2020**, *26*, 259-291.

59.   de Graaf, R.; van der Vossen, R. Bits versus brains in content analysis. Comparing the advantages and disadvantages of manual and automated methods for content analysis. *Commun* **2013**, *38*, 433–443.

60.   Lewis, S.C.; Zamith, R.; Hermida, A. Content analysis in an era of big data: A hybrid approach to computational and manual methods. *J Broadcast Electron Media* **2013**, *57*, 34–52.

61.   Feng, Y. Semantic Textual Similarity Analysis of Clinical Text in the Era of LLM. In 2024 IEEE Conference on Artificial Intelligence, (CAI), Singapore, 22–24 May 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1284–1289.

62.   Papageorgiou, E.; Chronis, C.; Varlamis, I.; Himeur, Y. A Survey on the Use of Large Language Models (LLMs) in Fake News. *Future Internet* **2024**, *16*, 298.

63.   Maktabdar Oghaz, M.; Babu Saheer, L.; Dhame, K.; Singaram, G. Detection and classification of ChatGPT-generated content using deep transformer models. *Front Artif Intell* **2025**, *8*, 1458707.

64.   Wu, J.; Yang, S.; Zhan, R.; Yuan, Y.; Chao, L.S.; Wong, D.F. A Survey on LLM-Generated Text Detection: Necessity, Methods, and Future Directions. *Comput Linguist* **2025**, *51*, 275–338.

65.   Domínguez-Diaz, A.; Goyanes, M.; de-Marcos, L. Automating Content Analysis of Scientific Abstracts Using ChatGPT: A Methodological Protocol and Use Case. *MethodsX* **2025**, *15*, 103431.

66.   Ma, X.; Zhang, Y.; Ding, K.; Yang, J.; Wu, J.; Fan, H. On Fake News Detection with LLM Enhanced Semantics Mining. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024); Al-Onaizan, Y.; Bansal, M.; Chen, Y.-N., Eds.; Association for Computational Linguistics: Miami, FL, USA, 2024; pp. 508–521.

67.   Yang, X.; Li, Y.; Zhang, X.; Chen, H.; Cheng, W. Exploring the Limits of ChatGPT for Query or Aspect-Based Text Summarization. *CoRR* **2023**, abs/2302.08081.

68.   Ohsawa, Y. Context design for chance discovery: Words, agents, and interactions. *New Gener Comput* **2006**, *20*, 143-164.

69.   Bang, Y.; Cahyawijaya, S.; Lee, N.; Dai, W.; Su, D.; Wilie, B.; Lovenia, H.; Ji, Z.; Yu, T.; Chung, W.; Do, Q.V.; Xu, Y.; Fung, P. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. In Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP–AACL 2023), Park, J.C.; Arase, Y.; Hu, B.; Lu, W.; Wijaya, D.; Purwarianti, A.; Krisnadhi, A.A., Eds.; Association for Computational Linguistics: Nusa Dua, Bali, Indonesia, 2023; pp. 675–718.