IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 July 2024, accepted 5 August 2024, date of publication 8 August 2024, date of current version 17 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3440660

RESEARCH ARTICLE

Navigating and Addressing Public Concerns in AI: Insights From Social Media Analytics and Delphi

MEHRDAD MAGHSOUDI[1], AMIRMAHDI MOHAMMADI[2], AND SAJJAD HABIBIPOUR[3]
[1]Department of Industrial and Information Management, Faculty of Management and Accounting, Shahid Beheshti University, Tehran 19839-69411, Iran
[2]School of Management, Economics and Progress Engineering, Iran University of Science and Technology, Tehran 13114-16846, Iran
[3]Faculty of Mechanical Engineering, K. N. Toosi University of Technology, Tehran 16317-14191, Iran

Corresponding author: Mehrdad Maghsoudi (M_Maghsoudi@sbu.ac.ir)

ABSTRACT The rapid advancement and integration of artificial intelligence (AI) in various domains of society have given rise to a complex landscape of public concerns. This research endeavors to systematically explore these concerns by employing a multi-stage methodology that combines large-scale social media data collection from Twitter and advanced text analytics. The study identifies seven distinct clusters of concerns, encompassing privacy and security, workforce displacement, existential risks, social and ethical implications, dependency on AI, misuse of AI, and lack of transparency. To further contextualize these findings, the Delphi method was employed to gather insights from AI ethics experts, providing a deeper understanding of the public's apprehensions. The results underscore the critical need for addressing these concerns to foster public trust and acceptance of AI technologies. This comprehensive analysis offers valuable guidance for policymakers, AI developers, and stakeholders to navigate and mitigate the multifaceted issues associated with AI, ultimately contributing to more informed and responsible AI deployment. By addressing these public concerns, the study aims to pave the way for a more ethically sound and socially acceptable integration of AI into society, ensuring that the benefits of AI can be realized while minimizing potential risks and negative impacts. Through this systematic approach, the research highlights the importance of continuous monitoring and proactive management of AI-related concerns to sustain public confidence and promote beneficial AI innovation.

INDEX TERMS Artificial intelligence, AI concerns, social media analysis, AI governance, responsible AI.

I. INTRODUCTION

In today's rapidly evolving world, the significance and benefits of artificial intelligence (AI) cannot be overstated [1]. From enhancing efficiency in various industries to revolutionizing healthcare [2] and transportation [3], [4], AI holds immense promise for shaping our future. AI technologies streamline processes, enable advanced data analysis [5], and facilitate innovative solutions across numerous domains, driving economic growth and improving the quality of life [6], [7]. However, hand in hand with the advancements in AI technology come profound concerns regarding its ethical implications and societal impact [8], [9].

The emergence of generative artificial intelligence, exemplified by tools like ChatGPT, Midjourney, and DALL-E 3, has brought these concerns to the forefront of public discourse [10], [11]. Questions surrounding privacy, bias, accountability, and the potential for job displacement have colored the AI landscape with complexity and urgency [12], [13], [14], [15]. As AI systems become more integrated into daily life, the necessity of addressing these ethical and societal issues grows more pressing. Failure to do so could result in significant negative consequences, including erosion of public trust, increased inequality, and unintended harmful behaviors by AI systems.

While existing research has extensively explored the ethical implications of AI, particularly in healthcare and education, there remains a significant gap in our understanding of how these concerns are expressed and evolve in real-time

The associate editor coordinating the review of this manuscript and approving it for publication was Xianzhi Wang.

public discourse [16]. Social media platforms, with their vast and dynamic user-generated content, offer a unique opportunity to capture and analyze the public's sentiments, fears, and expectations regarding AI. However, few studies have systematically examined this rich source of data to identify and prioritize AI-related concerns.

Furthermore, the rapid pace of AI development and deployment often outstrips the ability of researchers and policymakers to anticipate and address emerging ethical challenges. This creates a need for more agile and responsive approaches to identifying and addressing AI-related concerns. Additionally, while many studies have proposed ethical guidelines and frameworks for AI, there is a lack of research that combines large-scale public opinion analysis with expert-driven solutions to create actionable recommendations for addressing AI concerns.

Addressing these concerns is paramount, for they resonate deeply within society and directly influence public perception and acceptance of AI technologies. As AI continues to permeate various facets of our lives, it becomes imperative to delve into these issues, not only to mitigate risks but also to ensure that AI development aligns with ethical standards and societal values. By doing so, we can foster a more inclusive and equitable technological landscape where AI serves as a force for good.

Against this backdrop, social media platforms have emerged as invaluable repositories of public opinion and discourse [17], [18], [19]. These platforms provide real-time, diverse, and dynamic data reflecting the thoughts and feelings of a broad spectrum of society, making them crucial tools for understanding and addressing concerns. Social media offers a unique opportunity to gain insights into public concerns surrounding AI, leveraging the rich trove of data available to identify trends, priorities, and sentiments.

This research aims to bridge the identified gaps by employing a novel, multi-stage methodology that combines large-scale social media data analysis with expert insights. By doing so, we seek to provide a more comprehensive and nuanced understanding of AI-related concerns as they emerge and evolve in public discourse. This approach allows for the identification of concerns that may not be captured by traditional research methods, while also providing a mechanism for rapid response to emerging issues.

Thus, this research endeavors to conduct a meticulous and data-driven examination of AI-related concerns within social media. By employing text mining techniques and a systematic Delphi method, we aim to identify, categorize, and prioritize these concerns. Moreover, by engaging AI experts, we seek to formulate actionable recommendations to address these concerns effectively. This approach not only ensures a comprehensive understanding of public sentiment but also bridges the gap between public concerns and expert solutions.

The objectives of this research are threefold:

1. To elucidate the prevalent concerns of individuals and society regarding AI through a systematic analysis of social media discourse, providing a real-time snapshot of public sentiment.
2. To discern the priorities and evolving trends of these concerns, enabling a more responsive approach to addressing ethical and societal issues in AI development.
3. To propose actionable measures for governments, developers, consumers, and other stakeholders to tackle these concerns proactively, based on a synthesis of public sentiment and expert knowledge.

The necessity of this research is underscored by the rapid integration of AI into everyday life and the potential for its impact to outpace regulatory and ethical frameworks. By systematically analyzing public concerns, we can preemptively address issues before they escalate, ensuring that AI development proceeds in a manner that is both beneficial and ethically sound.

This study's unique contribution lies in its integration of large-scale social media analysis with expert insights, providing a more comprehensive and nuanced understanding of AI-related concerns than previous research. By combining these approaches, we aim to create a bridge between public perception and expert knowledge, facilitating more effective and responsive AI governance and development strategies.

Looking ahead, this paper will proceed as follows: Section II will provide a comprehensive review of the literature on AI concerns, contextualizing our research within the existing body of knowledge. Section III will delineate the methodology employed in this study. Section IV will present the findings derived from our analysis. Finally, Section V will discuss implementation strategies, and outline the limitations of our research while charting a course for future investigations.

II. LITERATURE REVIEW

A. OVERVIEW OF AI TECHNOLOGIES AND THEIR APPLICATIONS

Artificial Intelligence (AI) has transformed various sectors by automating tasks traditionally performed by humans and enhancing efficiency [20], [21]. Machine Learning (ML), a core AI framework, allows algorithms to learn from historical data without explicit programming, leading to advancements in various fields [22]. The integration of AI in organizations has necessitated upskilling and reskilling of workers to adapt to the changing landscape [20]. The continuous growth in computing power, availability of vast data, and algorithmic innovations are key drivers of AI development. AI technologies have played a significant role in improving food safety, and healthcare, and aiding in poverty alleviation efforts [23], [24], [25], [26], [27].

The historical development of AI has paved the way for its current applications across diverse domains, including medicine, education, and economics [28], [29], [30]. AI's role in enhancing mental health care, diagnosing diseases, and optimizing treatment processes underscores its significance

in improving human well-being [28], [31]. Moreover, AI has been crucial in advancing learning processes, with models that promote learning through intelligent systems [32]. The evolution of AI has led to the emergence of sophisticated tools like chatbots and language models, transforming educational paradigms [29], [33].

Ethical considerations in AI have gained prominence as the technology becomes more pervasive, necessitating the operationalization of AI ethics to address associated challenges [34], [35]. The potential biases in AI systems, particularly in healthcare and societal contexts, have raised concerns regarding fairness and equity [36], [37]. Stakeholder engagement in AI development is crucial to ensure alignment with diverse perspectives and values, emphasizing the need for inclusive practices [38]. Furthermore, the ethical implications of AI in legal services highlight the importance of aligning technological advancements with legal frameworks [39].

AI's impact on various industries, such as cybersecurity, climatology, and urban planning, underscores its versatility and potential for addressing complex challenges [23], [40], [41]. The application of AI in flood prediction, data mining, and healthcare prognosis demonstrates its utility in enhancing decision-making processes [42], [43], [44]. Additionally, AI's integration in radiology and medical imaging showcases its potential to revolutionize diagnostic practices and improve patient outcomes [45], [46].

The development of AI over time and its current uses illustrate a transformative journey that has revolutionized industries, improved decision-making, and brought about ethical dilemmas. As AI progresses, tackling biases, upholding ethical standards, and involving stakeholders will be crucial to fully leveraging its benefits for societal advancement.

B. AI ETHICAL AND SOCIETAL CONCERNS

Ethical and societal concerns surrounding artificial intelligence (AI) have become increasingly prominent as AI technologies continue to advance. One of the key issues is privacy, as AI systems often rely on vast amounts of data, raising concerns about how this data is collected, stored, and used [47]. Additionally, bias and fairness in AI systems have been identified as critical challenges, with the potential for AI algorithms to perpetuate or even exacerbate existing societal biases [48]. Ensuring accountability and transparency in AI decision-making processes is crucial to building trust in these systems and understanding how they arrive at their conclusions [49]. Moreover, the potential for job displacement and economic inequality due to automation and AI technologies is a significant societal concern that needs to be addressed through policy and education [50].

Public trust in AI is essential for its widespread acceptance and adoption, and understanding the societal implications of AI is crucial for policymakers, developers, and users alike [51]. Trust in AI-assisted decision-making is a key factor that influences how these systems are perceived and utilized [49]. By evaluating trust in AI systems, researchers can better understand how to design systems that are trustworthy and reliable. Expectations surrounding AI and ethics play a significant role in shaping societal attitudes towards these technologies [51]. As societal expectations evolve, it is essential to consider how ethics are performed and perceived in the context of AI development and deployment.

In healthcare, the ethical implications of AI and robotics are particularly salient, with considerations around privacy, bias, fairness, accountability, transparency, autonomy, and human oversight being paramount [47]. These technologies have the potential to revolutionize healthcare delivery but also raise complex ethical dilemmas that must be carefully navigated. Integrating ethics and considerations of career futures into AI literacy education for students is crucial for preparing the next generation to engage with these technologies responsibly [50]. By equipping students with a comprehensive understanding of AI, including its technical aspects, ethical implications, and potential career paths, educators can foster a more informed and ethical approach to AI.

Addressing fairness, bias, and appropriate use of AI and machine learning in global health is essential for ensuring that these technologies benefit all populations equitably [48]. Without careful consideration of fairness and bias, AI systems in healthcare may inadvertently perpetuate disparities and harm vulnerable populations. Algorithmic fairness is a critical area of research, with a focus on developing techniques that mitigate bias and promote equitable outcomes [52]. By employing fair machine learning techniques and prioritizing algorithmic fairness, developers can work towards creating AI systems that are more just and inclusive.

Ensuring algorithmic fairness requires a holistic approach that involves using diverse and representative datasets, enhancing transparency and accountability in AI systems, and exploring alternative AI paradigms that prioritize fairness and ethical considerations [53]. By addressing bias in AI at multiple levels, from data collection to model deployment, developers can mitigate the risks of perpetuating societal inequalities. Prioritizing social and ethical considerations in AI development is crucial for building trust with users and stakeholders [54]. Proactively working to eliminate bias and promote fairness in AI models is essential for creating systems that are reliable and unbiased.

Ethical and societal concerns related to AI encompass a wide range of issues, from privacy and bias to accountability and transparency. Addressing these concerns requires a multi-faceted approach that considers the implications of AI technologies across various domains, including healthcare, education, and governance. By prioritizing fairness, transparency, and ethical considerations in AI development and deployment, stakeholders can work towards creating AI systems that benefit society as a whole while minimizing potential harms.

C. RELATED WORKS

The rapid advancement of Artificial Intelligence (AI) across various fields has raised significant ethical concerns that

necessitate careful consideration. Numerous studies have delved into the ethical implications of AI, underscoring the critical need for transparency, fairness, accountability, and privacy [13]. In the medical domain, AI applications for diagnostics and treatment planning have been closely examined for their ethical implications, calling for robust ethical frameworks to guide their deployment [55]. The integration of AI in health informatics has further sparked discussions on governance models, citizen trust, and ethical issues in patient care, highlighting the complexity of ethical considerations in healthcare AI [56], [57].

Collaboration among diverse stakeholders—including governments, businesses, academia, and society—is essential for promoting responsible and equitable AI practices. This is particularly urgent in the field of computer vision, where balancing technological advancement with ethical considerations is critical [58]. Researchers have emphasized the importance of ethical principles, guidelines, policies, and regulations in AI, especially in educational contexts where transparency is paramount [59]. Ethical approaches to AI data management focus on auditing, benchmarking, confidence, trust, explainability, and interpretability to mitigate ethical concerns and enhance individual well-being [60].

The ethical considerations of using AI to monitor social media for COVID-19 data have also been explored, with discussions on the potential risks and benefits of AI-driven approaches in understanding public perspectives and behaviors during health crises [61]. In the mental health sector, the World Health Organization (WHO) has provided guidance on the ethical and governance aspects of AI, emphasizing the need for ethical frameworks in clinical settings to address risks and governance challenges [62]. Additionally, ensuring transparency and mitigating bias is critical in applying AI in healthcare to maintain patient fairness and uphold ethical standards [63].

Despite the convergence of ethical principles in AI guidelines, challenges remain in effectively interpreting and implementing these principles [64]. Embedding ethics into AI technologies is essential to ensure privacy, social justice, and ethical use, as emphasized by recent studies [65]. The scrutiny of legal and ethical principles in public health AI highlights ethics as a priority in advancing AI technologies [66].

Research in psychological artificial intelligence has noted the growing base of AI applications in psychology, stressing the importance of addressing ethical considerations in this field [67]. Systematic reviews of empirical studies on medical AI ethics have mapped approaches, findings, and limitations to inform future practices, underscoring the necessity of ethical frameworks in medical AI [68]. Similarly, scoping reviews on the ethical implications of AI in fields such as anesthesiology and ophthalmology have highlighted the ethical challenges posed by AI integration in healthcare settings [69], [70].

Debates on the ethics of AI in healthcare emphasize the proactive addressing of ethical challenges and the role of policymakers in ensuring ethical AI implementation [71]. The rise of AI and learning analytics in educational institutions has brought forth ethical considerations related to privacy, fairness, equity, and transparency, calling for robust ethical frameworks in educational AI applications [72]. Moreover, the ethical and social aspects of AI-enabled chatbots in mental health have been acknowledged, necessitating further research to address ethical implications and establish effective practices in this innovative sector [73].

Ethical scrutiny of medical AI applications has identified potential injustices, discrimination, and liability risks, advocating for the establishment of deeply integrated ethical review mechanisms to ensure responsible AI deployment [74]. Concerns about biases, transparency, privacy, and safety in AI applications in healthcare further highlight the importance of addressing these ethical challenges [75]. A systematic review of ethical considerations in AI deployment in healthcare is essential to ensure ethical practices [76].

Legal and ethical issues in AI applications in healthcare—including privacy, surveillance, bias, discrimination, and the role of human judgment—emphasize the need for clear responsibilities and ethical frameworks [77]. Discussions on the moral consideration of artificial beings in virtual environments have touched upon the ethical viability of AI systems, framing the moral accountability of human users and designers in AI development and utilization [78].

The existing body of research has extensively explored the ethical implications of AI, highlighting critical areas such as transparency, fairness, accountability, and privacy across various domains, particularly healthcare and education. However, there remains a gap in systematically examining public concerns about AI as expressed on social media platforms, which serve as dynamic and real-time repositories of public opinion. The present research addresses this gap by employing text-mining techniques and a systematic Delphi method to analyze social media discourse on AI-related concerns.

Our approach to understanding AI-related concerns offers several advantages compared to existing studies. Firstly, we achieve greater data comprehensiveness by analyzing large-scale social media data, providing a broader spectrum of public concerns than studies relying on limited surveys or expert interviews. Our method also uniquely integrates public and expert perspectives by combining social media analysis with the Delphi method, bridging the gap between general public concerns and expert views. Additionally, our approach allows for temporal analysis, enabling the examination of how concerns evolve over time, in contrast to cross-sectional studies. Lastly, our research goes beyond merely identifying concerns to provide practical strategies for addressing them, offering a more solution-oriented approach.

III. METHODOLOGY

This research employs a comprehensive, multi-stage methodology (Figure 1) to thoroughly explore and analyze concerns related to artificial intelligence as expressed on social networks. The aim is to systematically collect, process, and derive insights from user-generated content to identify

prevalent themes of apprehension and propose expert-driven solutions.

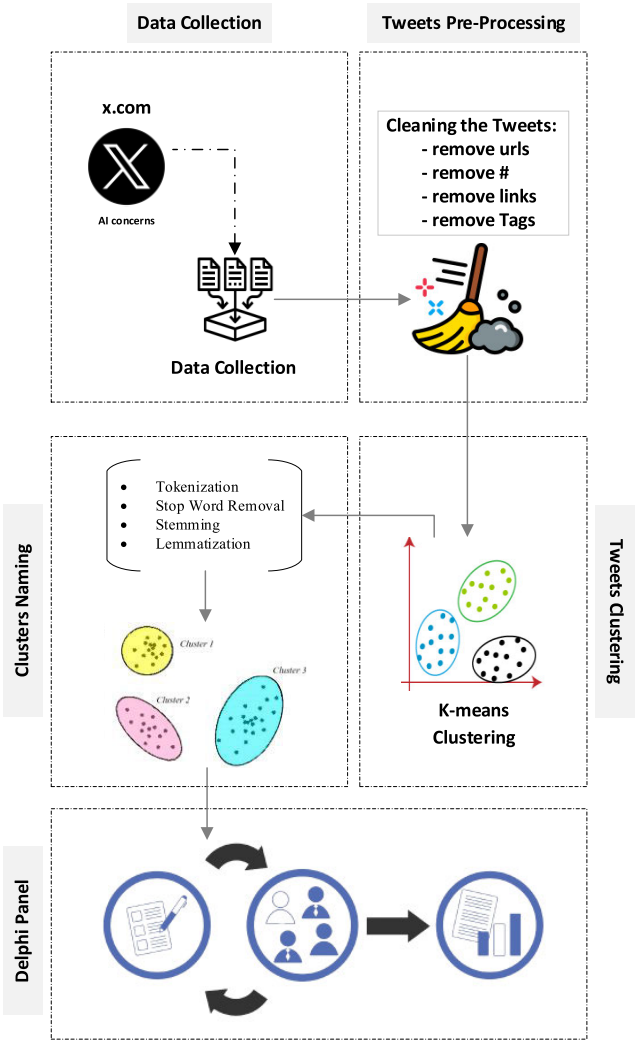


FIGURE 1. Research methodology.

A. DATA COLLECTION

The initial stage involves gathering a broad corpus of tweets pertaining to fears, worries, and concerns surrounding artificial intelligence. This is accomplished by conducting extensive searches on the social media platform X (Twitter) using carefully selected keywords, phrases, and hashtags that are likely to yield relevant content. The search criteria are iteratively refined to capture a wide range of perspectives and ensure a comprehensive dataset for subsequent analysis.

B. TWEETS PRE-PROCESSING

To ensure the quality and relevance of the collected data, the tweets undergo a rigorous preprocessing and vectorization process [79]. This process involves several key steps:

1. Initial cleaning: All extraneous elements such as URLs, hashtags (#), links, and user tags are removed from the tweet text. This step eliminates noise and allows the focus to be placed solely on the substantive content.
2. Length filtering: Tweets that are too short to provide meaningful information, defined as those containing fewer than 5 words or less than 50 characters [80], are excluded from the dataset. This filtering step helps to retain only tweets that are likely to contain sufficient context and detail for the purpose of identifying AI-related concerns [81].
3. Further removal of non-textual elements: Email addresses, numerical digits, extra lines, multiple spaces, and special characters (excluding spaces) are removed from the tweets. This ensures that the analysis focuses solely on the meaningful textual content [82], [83].
4. Tokenization: Each cleaned tweet is tokenized using the Facebook/BART-base model, a state-of-the-art natural language processing model. This step breaks down the text into individual tokens, which can be words or subwords, allowing for a more nuanced analysis of the content [84].
5. Vector representation: Following tokenization, each tweet is converted into a 768-dimensional vector using max pooling. This technique captures the most salient features of the text, creating a dense representation that encapsulates the semantic content of each tweet [85].

The max pooling approach works as follows:

- For each token in the tweet, the BART model generates a 768-dimensional embedding vector.
- These token-level embeddings are then aggregated by taking the maximum value across all tokens for each of the 768 dimensions [86].
- The resulting 768-dimensional vector represents the entire tweet, capturing the most prominent features across all tokens [87].

This vectorization process transforms the textual data into a format suitable for machine learning algorithms while preserving important semantic information. The high-dimensional representation allows for nuanced clustering and analysis of the tweets, enabling the identification of subtle patterns and themes in the AI-related concerns expressed by users [88].

C. TWEETS CLUSTERING

To uncover the primary topics and themes within the collected tweets, an unsupervised machine learning approach is applied in the form of K-means clustering. This algorithm groups similar tweets together based on their textual content [89], [90], allowing for the emergence of distinct clusters that represent different categories of AI concerns. The optimal number of clusters (K) is determined through the use of the silhouette index, a well-established evaluation metric in the field of clustering analysis [91], [92]. By examining the silhouette scores for different values of K, the most appropriate number of clusters can be selected to achieve a balance between intra-cluster cohesion and inter-cluster separation.

The K-means algorithm works as follows:

1) Initialize K centroids randomly in the feature space.
2) Assign each data point (tweet vector) to the nearest centroid based on Euclidean distance.
3) Recalculate the centroids as the mean of all data points assigned to that cluster.
4) Repeat steps 2 and 3 until convergence or a maximum number of iterations is reached [93].

The silhouette score for each data point i is calculated as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where:

- $a(i)$ is the average distance between i and all other points in the same cluster
- $b(i)$ is the average distance between i and all points in the nearest neighboring cluster

The overall silhouette score is the average of $s(i)$ for all data points. A higher silhouette score indicates better-defined clusters [94].

D. CLUSTER NAMING

Once the tweets have been divided into distinct clusters, the next step is to assign a descriptive label to each cluster that captures the essence of the concerns expressed within it. To achieve this, the tweets in each cluster are subjected to a series of natural language processing techniques. Tokenization is performed to break down the text into individual words or tokens. Stop words, which are common words that carry little semantic meaning (such as ''the'', ''a'', ''and''), are removed to focus on the most informative terms. Stemming and lemmatization are then applied to reduce words to their base or dictionary forms, allowing for the consolidation of related terms. Finally, the most frequently occurring keywords within each cluster are identified and used to generate a representative name for that cluster [95], [96]. These cluster names serve as concise labels that encapsulate the primary themes of concern discovered within the data.

The process of cluster naming involves the following steps:

1) Tokenization: Using the NLTK library's word_tokenize() function to split tweets into individual words.
2) Stop word removal: Utilizing NLTK's list of English stop words to filter out common, non-informative terms.
3) Stemming: Applying the Porter Stemmer algorithm to reduce words to their root form (e.g., ''running'' to ''run'').
4) Lemmatization: Using WordNet Lemmatizer to convert words to their dictionary form (e.g., ''better'' to ''good'').
5) Frequency analysis: Employing the Counter class from Python's collections module to identify the most common terms in each cluster.

The top N (e.g., N=5) most frequent terms are then combined to create a descriptive name for each cluster, capturing the main themes present in the grouped tweets.

E. DELPHI PANEL

To move beyond the identification of concerns and towards the development of actionable solutions, the final stage of the research involves the convening of an expert panel. This panel, composed of individuals with deep domain knowledge in artificial intelligence, ethics, social science, and related fields, is tasked with reviewing the identified clusters of concerns and proposing strategies to mitigate or address them. The Delphi method, a structured communication technique, is employed to facilitate this process [97]. Through multiple rounds of anonymous input collection and controlled feedback, the panel engages in a systematic and iterative process of idea generation, refinement, and consensus-building. The ultimate goal is to produce a prioritized set of recommendations that can guide efforts to reduce or resolve the public's concerns and apprehensions surrounding artificial intelligence. The Delphi method is implemented in three primary rounds:

1) Initial assessment: Experts individually rate the importance of each identified concern on a Likert scale (1-5).
2) Feedback and re-evaluation: Experts are presented with the aggregated results from round 1 and given the opportunity to revise their ratings.
3) Solution generation: Experts propose and rate potential solutions for addressing the top-ranked concerns.

Between each round, responses are anonymized and statistically summarized (e.g., calculating mean scores and interquartile ranges) to provide structured feedback to the panel. This iterative process continues until a predetermined level of consensus is reached or a set number of rounds is completed.

By adhering to this rigorous and multi-faceted methodology, this research endeavors to provide a comprehensive understanding of the prevalent concerns related to AI as expressed on social networks. Through the combination of large-scale data collection, advanced text analytics, and expert insight, it seeks to contribute valuable knowledge and propose constructive solutions to address the complex challenges posed by the rapid advancement of artificial intelligence in our society.

IV. RESULTS

A. DATA COLLECTION

The dataset for this study was collected from Twitter through the Twitter API, specifically targeting tweets that discussed concerns related to artificial intelligence (AI). Over the period from January 1, 2019, to February 18, 2024, a total of 86,797 tweets were gathered. The search strategy was meticulously crafted to ensure comprehensive coverage of the discourse surrounding AI concerns and risks.

TABLE 1. Search strategy.

Search Concept	Search Keywords
Artificial intelligence concerns	"artificial intelligence" + "concern" , "AI" + "concern"
AI risks	"artificial intelligence " + "risk" , "AI" + "risk"
AI ethical issues	" artificial intelligence " + "ethic" , "AI" + "ethic"
AI worries	"artificial intelligence" + "worrie", "AI" + "worrie"
artificial intelligence fears	"artificial intelligence" + "fear", "AI" + "fear"
machine learning concerns	"machine learning" + "concern" , "ML" + "concern" , "deep learning" + "concern"
robotics concerns	"robotic" + "concerns"

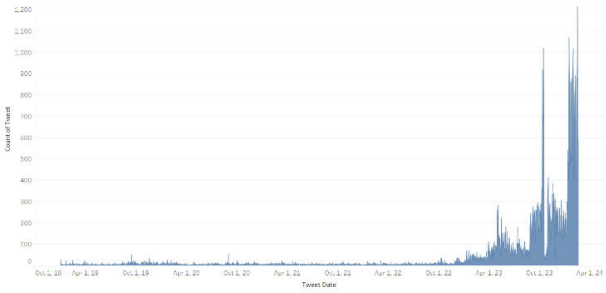


FIGURE 2. Timeline of tweets.

To ensure a robust and representative dataset, we employed a systematic approach to data collection. The Twitter API was accessed using Python, leveraging the 'tweepy' library, which allows for efficient and programmatic retrieval of tweets. We implemented rate limiting and error handling to comply with Twitter's usage policies and to ensure the integrity of our data collection process.

Our search queries were constructed using Boolean operators to combine key terms related to AI with words indicating concern or risk. For example, a typical query might look like: ''(''artificial intelligence'' OR AI) AND (concern OR risk OR ethic∗ OR worr∗ OR fear)''. This approach allowed us to capture a wide range of expressions related to AI concerns while minimizing irrelevant content.

To address potential biases in data collection, we implemented a randomized sampling technique within the constraints of the Twitter API. This involved collecting tweets at different times of day and days of the week to ensure a temporally diverse sample.

Table 1 shows the searched keywords.

The keywords were carefully selected to comprehensively capture the discourse surrounding AI concerns and risks. Figure 2 illustrates the time distribution of the collected tweets. The collected data was stored in a secure, encrypted database, with personal identifiers removed to ensure user privacy. Each tweet was assigned a unique identifier to facilitate tracking and analysis throughout the research process.

B. TIMELINE ANALYSIS

To understand the evolution of AI-related concerns over time, a timeline analysis was conducted in Figure 4. The number of tweets related to AI concerns was plotted against their publication dates, allowing for the identification of peaks corresponding to significant events and announcements within the AI industry. These events were manually researched and verified to provide context for the observed spikes in AI-related discussions.

The analysis revealed several key events that coincided with notable peaks in tweet activity. For instance, the release of GPT-4 on March 14, 2023, and Google Bard on March 21, 2023, both generated significant discussions, although the peaks were relatively modest, with 31 and 30 tweets, respectively. A more substantial spike occurred on May 4, 2023, when Geoffrey Hinton resigned from Google and publicly warned about AI dangers, resulting in 226 tweets.

Further notable events included OpenAI CEO Sam Altman's remarks on AI concerns on May 16, 2023, which led to 174 tweets, and the statement on AI risk signed by prominent figures on May 26, 2023, which generated 98 tweets. The Hollywood strike over AI concerns on July 14, 2023, also attracted attention, with 108 tweets. The release of Bing Chat on July 18, 2023, prompted 54 tweets, reflecting moderate interest.

A significant surge in tweet activity was observed on September 13, 2023, coinciding with the US Senate AI Insight Forum, which generated 195 tweets. Another peak occurred on October 30, 2023, when US President Biden issued an Executive Order on AI, resulting in 135 tweets. The following day, October 31, 2023, saw an even higher spike with 230 tweets, driven by Elon Musk's proposal for an AI Safety Summit.

The release of Google Gemini 1.0 Ultra on December 6, 2023, led to 188 tweets, while the announcement of the OpenAI Sora model and the release of Google Gemini 1.5 on February 15, 2024, resulted in the highest peak of 659 tweets. These patterns highlight how major industry developments and public statements by influential figures can significantly influence the volume of discourse on social media platforms regarding AI concerns.

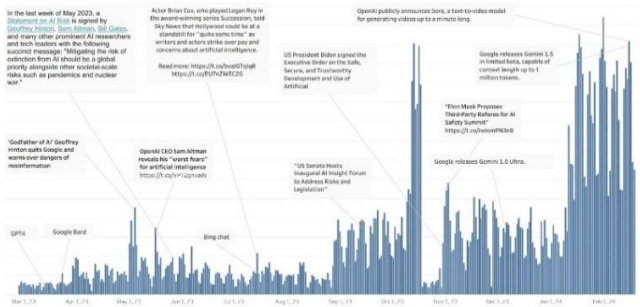


FIGURE 3. Timeline of tweets and peaks.

C. DATA PREPROCESSING

To prepare the collected tweets for analysis, a comprehensive preprocessing strategy was implemented. At the outset, all URLs, email addresses, numbers, extra lines, consecutive spaces, and special characters (except spaces) were

eliminated [98]. This step was essential to ensure that the analysis concentrated exclusively on the textual content of the tweets. After this cleaning process, tweets that were too short to provide meaningful information were filtered out. Specifically, any tweet containing fewer than five words or fewer than fifty characters was excluded. This filtering step ensured that only tweets with sufficient context and detail were retained for further analysis. Finally, after removing the short tweets, 59,447 tweets remained from the original dataset. This rigorous preprocessing approach helped to maintain a high-quality dataset, enhancing the reliability and validity of the subsequent analysis.

Clustering

Table 2 presents the most frequent keywords for each cluster, along with examples of the most liked tweets in each cluster. This methodical approach ensured that the cluster names served as concise labels encapsulating the primary themes of concern discovered within the data.

To identify the primary themes of AI concerns expressed in the tweets, the K-means clustering algorithm was applied to the tweet vectors. The optimal number of clusters was determined using the Silhouette Method, which evaluates clustering quality by considering both the cohesion within clusters and the separation between clusters. The Silhouette score ranges from −1 to 1, with higher scores indicating better-defined clusters. Based on this method, seven clusters were identified as optimal. Figure 4 displays the Silhouette scores, confirming the appropriateness of this clustering structure.

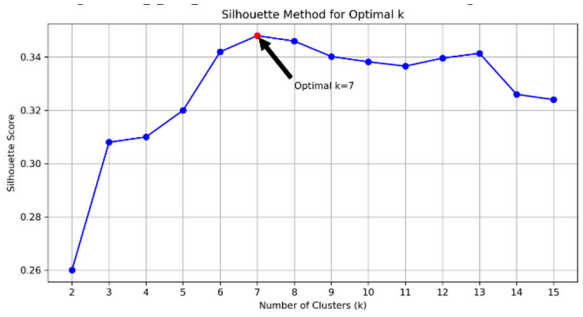


FIGURE 4. The number of optimal clusters based on the Silhouette index.

D. CLUSTERS ANALYSIS

After determining the optimal number of clusters, a comprehensive keyword selection process was employed to identify the most representative terms for each cluster. This process involved several natural language processing techniques, utilizing specific Python packages and functions, to ensure the most informative and relevant keywords were selected.

First, tokenization was performed using the 'word_tokenize' function from the 'nltk.tokenize' module to break down the text into individual words or tokens. This step facilitated the analysis of each word independently. Next, stop words were removed using the 'stopwords.words' function from the 'nltk.corpus' module, focusing on the most informative terms by eliminating common words that carry little semantic meaning, such as "the," "a," and "and." Following stop word removal, stemming was applied using the 'PorterStemmer' class from the 'nltk.stem' module, and lemmatization was applied using the 'WordNetLemmatizer' class from the same module. These processes reduced words to their base or dictionary forms, allowing for the consolidation of related terms and enhancing the analysis's consistency.

Using these refined terms, the most frequently occurring keywords within each cluster were identified with the help of the 'Counter' class from the 'collections' module. These keywords were then used to generate descriptive labels for each cluster, effectively capturing the essence of the concerns expressed within them.

Table 2 presents the most frequent keywords for each cluster, along with examples of the most liked tweets in each cluster. This methodical approach ensured that the cluster names served as concise labels encapsulating the primary themes of concern discovered within the data.

Based on the results from

Table 2, the following descriptions can be provided for each cluster:

1) PRIVACY AND SECURITY CONCERNS

This cluster focuses on the significant privacy and security risks associated with AI technologies. The most frequent keywords, such as "privacy," "security," "data," "protection," "bias," "accountability," "transparency," "regulation," "ethics," "safety," and "trust," highlight the multifaceted nature of these concerns. The cluster tweets underscore the importance of addressing ethical and social implications, including algorithmic bias, fairness, data privacy, and the potential for misuse or unintended consequences. The cluster emphasizes the need for robust data protection measures, algorithmic accountability, and ethical AI development to mitigate these risks and foster public trust in AI systems.

2) BUSINESS AND STRATEGY RISKS

This cluster explores the complex interplay between AI and business strategy. The top keywords, including "cybersecurity," "compliance," "fraud," "detection," "decision-making," "automation," "competitive," "advantage," "cost," "reduction," "efficiency," "innovation," "digital," and "transformation," reflect the diverse ways in which AI can impact business operations and outcomes. The sample tweets illustrate how AI can enhance cybersecurity, compliance, and fraud detection while also driving automation, cost reduction, and innovation. However, the cluster also acknowledges the potential risks that need to be carefully managed, such as the ethical implications of AI-driven decision-making and the challenges of integrating AI into existing business processes and structures.

3) WORKFORCE DISPLACEMENT

This cluster delves into the profound impact of AI on the workforce and the future of work. The most frequent keywords, such as "job," "loss," "automation," "reskilling," "upskilling," "inequality," "economic,"

TABLE 2. Names of identified clusters.

Number	Cluster Name	Keywords	Sample tweets
0	Privacy and Security Concerns	privacy, security, data, protection, bias, accountability, transparency, regulation, ethics, safety, trust	➢ 'AI raises important ethical and social implications, including concerns about bias and fairness, job displacement, algorithmic accountability, data privacy, and the potential for misuse or unintended consequences.', ➢ "Warning: Your data privacy could be at risk with Google's Gemini AI! "
1	Business and Strategy Risks	cybersecurity, compliance, fraud, detection, decision-making, automation, competitive, advantage, cost, reduction, efficiency, innovation, digital, transformation	➢ 'We are committed to unlocking talent and technology to help fight against global financial crime. With cutting-edge resources, powerful data analytics tools, and a collaborative network of experts, together we can make the world a safer place.', ➢ 'Finance: AI is transforming the finance industry by enabling better risk management, fraud detection, and customer service.'
2	Workforce Displacement	job, loss, automation, reskilling, upskilling, inequality, economic, disruption, future, work, human-machine, collaboration, social, safety, net, education	➢ 'The Ethics of AI: The Need for Regulation. AI could be used for malicious purposes, and it could lead to job losses and social upheaval.', ➢ 'The Future of AI: The Promise and the Peril. However, AI also poses a number of risks. It could be used for malicious purposes, and it could lead to job losses and social upheaval.'
3	Existential Risks and Long-term Impacts	superintelligence, existential, threat, AI, alignment, safety, control, ethics, technological, singularity, value, governance, long-term, thinking	➢ "Machines have bestowed us with incredible 'superpowers,' transforming our abilities in both physical and cognitive realms. The transformative impact of generative AI will likely differ from older technologies, as we explore in our new research." ➢ 'Perhaps the most consequential outcome of the summit was the commitment to pursue discussions on the dangers of AI, including concerns about autonomous weapons that would fall outside the loop of human decision-making.'
4	Government's Role and Regulation	regulation, oversight, safety measures, international coordination, AI ethics, AI governance, public-private partnerships, research funding, education, public engagement	➢ 'The Biden administration is considering new restrictions on exports of artificial intelligence chips to China, as concerns rise over the power of the technology in the hands of U.S. rivals.', ➢ 'SEC Chair Gary Gensler has concerns about AI and crypto.
5	Social and Ethical Implications	bias, fairness, transparency, accountability, privacy, surveillance, manipulation, misinformation, deep fakes, social, media	➢ 'An example of AI developments that can turn images convincingly into video, raising concerns about the future of social media.', ➢ 'FTC streamlines investigations into AI misuse amid growing concerns.'
6	Military and National Security Applications	national, security, military, defense, cyber, attacks, autonomous, weapons, surveillance, intelligence, geopolitics, arms, race, international, cooperation	➢ 'Defense Contractor Scale AI Quietly Scrapped Deal With Chinese-Owned TikTok Over Security Concerns.', ➢ 'White House tackles concerns over Chinese interest in Middle East AI as firm tries to play both sides.'

"disruption," "future," "work," "human-machine," "collaboration," "social," "safety," "net," and "education," reflect the multidimensional nature of this issue. The sample tweets highlight the concerns about job displacement due to automation and the potential for widening economic inequality. The cluster emphasizes the urgent need for proactive measures, such as reskilling and upskilling initiatives, to prepare workers for the changing job market. It also underscores the importance of policies to address potential economic disruption and ensure a smooth transition to a more automated workforce.

4) EXISTENTIAL RISKS AND LONG-TERM IMPACTS

This cluster focuses on the existential risks posed by advanced AI systems and the long-term implications for humanity. The top keywords, including "superintelligence," "existential," "threat," "AI," "alignment," "safety," "control," "ethics," "technological," "singularity," "value," "governance," "long-term," and "thinking," reflect the gravity and complexity of these concerns. The sample tweets highlight the potential for AI to transform human abilities in both physical and cognitive realms, but also the risks associated with the development of superintelligent AI that could potentially harm humanity. The cluster emphasizes the critical importance of AI alignment, safety measures, and governance frameworks to ensure that AI remains under human control and aligned with human values. It also underscores the need for long-term thinking and proactive measures to address these existential risks.

5) GOVERNMENT'S ROLE AND REGULATION

This cluster explores the crucial role of governments in regulating AI development and deployment. The most frequent keywords, such as "regulation," "oversight," "safety

measures," "international coordination," "AI ethics," "AI governance," "public-private partnerships," "research funding," "education," and "public engagement," reflect the multifaceted nature of this responsibility. The sample tweets highlight the various ways in which governments are grappling with AI risks, such as considering new restrictions on AI chip exports and investigating AI misuse. The cluster emphasizes the need for robust oversight, safety measures, and international coordination to address AI risks effectively. It also underscores the importance of public-private partnerships, research funding, and public engagement in shaping AI governance frameworks that balance innovation with safety and ethics.

6) SOCIAL AND ETHICAL IMPLICATIONS

This cluster delves into the complex social and ethical implications of AI, particularly in the context of social media and online platforms. The top keywords, including "bias," "fairness," "transparency," "accountability," "privacy," "surveillance," "manipulation," "misinformation," "deepfakes," "social," and "media," reflect the wide-ranging nature of these concerns. The sample tweets highlight the potential for AI to be used for malicious purposes, such as generating deepfakes and spreading misinformation, as well as the risks of algorithmic bias and unfairness. The cluster emphasizes the urgent need for transparency, accountability, and regulation to address these challenges and ensure that AI is developed and deployed in an ethical and socially responsible manner.

7) MILITARY AND NATIONAL SECURITY APPLICATIONS

This cluster focuses on the military and national security applications of AI and the associated risks and challenges. The most frequent keywords, such as "national," "security," "military," "defense," "cyber," "attacks," "autonomous," "weapons," "surveillance," "intelligence," "geopolitics," "arms," "race," "international," and "cooperation," reflect the complex and multifaceted nature of this issue. The sample tweets highlight the potential for AI to be used in autonomous weapons, surveillance, and intelligence gathering, as well as the geopolitical implications of AI, such as the risk of an AI arms race between nations. The cluster emphasizes the critical importance of international cooperation and governance frameworks to address these risks and ensure that AI is developed and deployed in a manner that promotes global security and stability. It also underscores the need for careful consideration of the ethical and social implications of military AI applications.

E. PRIORITY OF CONCERNS BASED ON TWEETS

Figure 5 shows The priority of concerns based on the clustering of tweets reveals a multifaceted landscape of apprehensions within the domain of artificial intelligence (AI). Privacy and security emerge as the foremost concerns, reflecting widespread unease regarding the protection of personal data and the potential vulnerabilities inherent in AI systems. Close behind, existential risks and long-term impacts signify a growing awareness of the profound societal implications of AI advancements. Business and strategy risks, alongside concerns about workforce displacement, underline the economic and social transformations catalyzed by AI, necessitating discussions on adaptation strategies and policy interventions. The role of government regulation in AI development garners attention, highlighting the need for ethical frameworks and oversight mechanisms. Additionally, discussions on social and ethical implications underscore the importance of addressing broader societal concerns, such as bias and fairness. Finally, the attention given to military and national security applications reflects awareness of the dual-use nature of AI technologies and ethical considerations surrounding their deployment in such contexts. These priorities provide valuable insights into the overarching concerns driving public discourse on AI and underscore the need for comprehensive approaches to address them effectively.

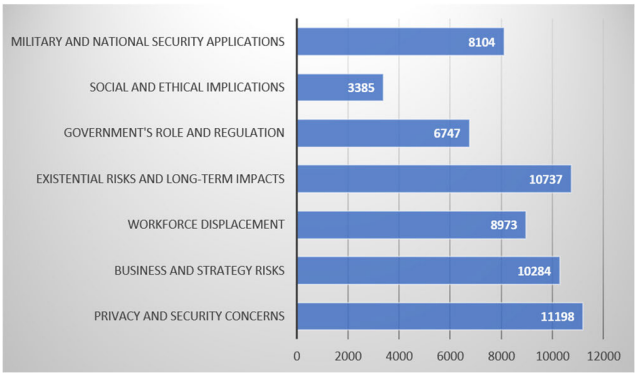


FIGURE 5. Priority of concerns based on tweets.

F. AI EXPERTS PANEL

To ensure a diverse and comprehensive range of insights, we adopted a multifaceted approach in selecting the expert cohort for this study's Delphi process. The selection parameters were designed to cover a broad spectrum of considerations:

- Age: We included experts across different age groups to capture generational perspectives.
- Gender: Gender parity was prioritized to provide a balanced viewpoint.
- Country of Residence: Experts from various countries were selected to ensure geographical and cultural diversity.
- Academic Qualifications: We assessed the depth and breadth of academic qualifications to ensure that the experts had a profound understanding of AI-related topics.
- Field of AI Activity: Experts from diverse AI sectors, including research, industry, and government, were chosen to incorporate a wide range of perspectives.

- Social Media Usage: We considered the experts' engagement on social media platforms to ensure they were current with the latest AI discussions and trends.

The selection process for the Delphi panel involved several rigorous steps:

1) Initial Identification: Potential experts were identified using a combination of:
   - Literature reviews to find authors of relevant publications.
   - Recommendations from academic and industry networks.
   - Searches of professional databases and social media platforms.
2) Qualification Assessment: Each potential expert's qualifications were evaluated based on:
   - Publication records in AI-related fields.
   - Years of experience in AI research or industry.
   - Involvement in AI policy-making or ethical discussions.
   - Recognition in the field, such as awards, citations, and speaking engagements.
3) Diversity Criteria: The pool of qualified experts was filtered to ensure diversity across the aforementioned parameters (age, gender, country of residence, etc.).
4) Invitation and Confirmation: Selected experts were invited to participate in the Delphi panel through emails and messages on social media platforms. Those who accepted were provided with detailed information about the study's objectives and the Delphi process.
5) Final Panel Formation: The final panel was formed by balancing expertise across different aspects of AI, such as technical development, ethical considerations, and policy implications, while maintaining the desired diversity.

This thorough selection process aimed to create a panel capable of providing well-rounded, informed, and diverse perspectives on AI-related concerns and potential solutions. As a result, ten experts were meticulously chosen to participate in the Delphi process, facilitating a robust and comprehensive exploration of the research questions. Details about the experts involved in the Delphi panel are presented in Table 3.

TABLE 3. Profile of the expert panel.

Age	Gender	Country	Academic Qualifications	Field of AI Activity	Social Media Usage
42	Female	USA	Ph.D. in Computer Science	Research	High
35	Male	China	MSc in Artificial Intelligence	Industry	Medium
48	Male	Egypt	Ph.D. in Ethics & Technology	Government	High
29	Female	UK	MEng in Electrical Engineering	Industry	Low
55	Female	Spain	MA in Law & Technology	Government	High
38	Female	Japan	BSc in Data Science	Industry	Medium
45	Male	UAE	MBA in Business Analytics	Industry	High
50	Male	Brazil	Ph.D. in Computer Engineering	Research	Low
33	Female	India	MS in Artificial Intelligence	Government	High
47	Male	Italy	MA in Ethics & Technology	Research	Medium

G. DELPHI METHOD FOR PRIORITIZING AI CONCERNS

In this research endeavor, the study commenced by identifying seven primary concerns pertaining to AI, extracted from an analysis of Twitter data. Subsequently in this step, these concerns underwent a Delphi method procedure involving a select group of domain experts to ascertain their prioritization based on perceived significance.

Phase 1: Evaluation of Concerns

Via an online survey platform, the panel of experts received the identified concerns and rated each using a Likert scale, ranging from 1 for strong disagreement to 5 for strong agreement.

Phase 2: Iterative Review

In the subsequent phase, the roster of concerns, alongside the ratings attributed to them in the initial round, was shared with the participating experts. This facilitated a comparative assessment, enabling each expert to recalibrate their individual ratings in light of their peers' evaluations.

Phase 3: Consolidation of Ratings

Upon conclusion of this phase, all survey responses were aggregated. The scores attributed by the experts to each concern were amalgamated, and subsequently, divided by the total number of experts involved. With a panel comprising 10 experts, each providing scores within the range of 1 to 5, the cumulative score for each concern varied from 10 to 50. This aggregated score was then normalized by dividing it by 10, thereby yielding the mean score, indicative of the relative importance accorded to each concern. The outcomes of this process are elucidated in Table 4.

The results, presented in Table 4, reveal that privacy and security concerns received the highest average score of 5, indicating their paramount importance as perceived by the experts. This was followed closely by existential risks and long-term impacts, and workforce displacement, both scoring 4.9 on average. Business and strategy risks, government's role and regulation, and social and ethical implications also garnered relatively high average scores, ranging from 4.5 to 4.8. Meanwhile, military and national security applications received a comparatively lower average score of 4.3, indicating a slightly lesser level of concern among the experts surveyed.

TABLE 4. Expert scores for prioritizing AI concerns.

concerns	Average of score	Number of experts with rating
Business and Strategy Risks	4.5	5(5) , 4(5) , 3(0) , 2(0) , 1(0)
Existential Risks and Long-term Impacts	4.9	5(9) , 4(1) , 3(0) , 2(0) , 1(0)
Government's Role and Regulation	4.8	5(8) , 4(2) , 3(0) , 2(0) , 1(0)
Military and National Security Applications	4.3	5(4) , 4(5) , 3(1) , 2(0) , 1(0)
Privacy and Security Concerns	5	5(10) , 4(0) , 3(0) , 2(0) , 1(0)
Social and Ethical Implications	4.7	5(7) , 4(3) , 3(0) , 2(0) , 1(0)
Workforce Displacement	4.9	5(9) , 4(1) , 3(0) , 2(0) , 1(0)

H. EXTRACTION OF KEY SOLUTIONS

1) PHASE 1: INITIAL PROPOSAL OF SOLUTIONS

Initially, the identified concerns were presented to the experts, who were then asked to propose potential solutions for each concern. This stage allowed for the generation of a wide range of possible solutions, leveraging the diverse expertise of the panel. After collecting the questionnaires and compiling the data, approximately 30 to 35 solutions were extracted for each concern. By consolidating these solutions and eliminating redundant suggestions, a refined list of approximately 10 to 15 unique solutions for each concern was developed. This process ensured that the final list was comprehensive yet focused, capturing the most relevant and diverse ideas from the experts.

2) PHASE 2: RATING THE SOLUTIONS

In the second stage, all proposed solutions were compiled into a list. Experts were then asked to rate each solution using a Likert scale, where 5 indicated strong agreement, 4 agreement, 3 neutrality, 2 disagreement, and 1 strong disagreement. At this stage, the list of solutions for each concern was provided to the experts. The experts rated each solution using the Likert scale. At the end of this process, the questionnaires were collected, and the score for each solution was calculated. By summing the Likert scale scores assigned by the experts, each solution received a score between 10 and 50.

3) PHASE 3: FINAL REVIEW AND REFINEMENT

In the final stage, both the list of concerns and the corresponding solutions, along with their Likert scale ratings, were provided to the experts for a final review. Experts were invited to make any necessary revisions or provide additional comments to refine the solutions further. This process aimed to achieve a consensus on the most viable solutions for each concern.

Identification of Key Solutions

After completing the three rounds of the Delphi method, we extracted the principal solutions based on the cumulative ratings. This process resulted in a refined list of solutions tailored to address each of the primary concerns identified in the earlier phase.

I. SELECTED SOLUTIONS

The Delphi method resulted in the identification and prioritization of several key solutions. Below are the top solutions determined by the expert panel presented in the four categories of law and regulation, society, ethics, and technology implementation (based on the framework provided by Wirtz et al. [99]):

1) AI LAW AND REGULATION

a: GOVERNMENT'S ROLE AND REGULATION

As AI becomes increasingly integrated into various aspects of life, governments must develop appropriate regulations and policies to ensure responsible AI development and deployment. This includes addressing issues such as algorithmic bias, transparency, accountability, and protecting individual rights and freedoms. Governments need to balance fostering innovation with protecting the public interest while collaborating with industry, academia, and civil society.

b: SOLUTIONS

- Foster public-private partnerships and international cooperation to address global AI challenges: Collaborative efforts can lead to more effective regulation and oversight.
- Establish clear guidelines and regulations for the development and deployment of AI, especially in sensitive areas: Developing robust regulatory frameworks will ensure responsible AI use.
- Promote public engagement and education regarding AI development: Educating the public about AI can help in aligning AI applications with societal values.

2) AI SOCIETY

a: WORKFORCE DISPLACEMENT

One of the most significant concerns surrounding AI is its potential to displace human workers across a wide range of industries. As AI systems become more sophisticated and capable of performing tasks that were previously the domain of human workers, there is a risk of widespread job losses and economic disruption. While some argue that AI will also create new jobs and opportunities, there is a need to proactively address the challenges of workforce transition, retraining, and ensuring that the benefits of AI are distributed equitably.

b: SOLUTIONS

- Prioritize reskilling and upskilling initiatives: Reskilling programs will help workers adapt to new roles.
- Invest in education and training programs for AI and digital skills: Education and training initiatives are crucial for preparing the workforce.
- Explore policies to support workers during economic transitions: Supportive policies will help workers during periods of economic

c: SOCIAL AND ETHICAL IMPLICATIONS

AI raises a host of social and ethical questions, particularly as it becomes more deeply embedded in decision-making processes that affect people's lives. This includes issues such as algorithmic fairness, privacy, and the potential for AI to perpetuate or amplify existing biases and inequalities. There are also concerns about the impact of AI on human relationships, social interaction, and the erosion of human agency and autonomy. Addressing these challenges will require ongoing dialogue and collaboration among researchers, policymakers, and the broader public.

d: SOLUTIONS

- Develop AI systems that are transparent, accountable, and unbiased: Building systems with these qualities will address biases and improve accountability.
- Establish guidelines for the ethical use of AI in surveillance: Clear guidelines are needed to ensure ethical use in surveillance.
- Develop ethical standards for AI deployment: Ethical standards will guide the responsible deployment of AI.

e: MILITARY AND NATIONAL SECURITY APPLICATIONS

The military and national security applications of AI have been a topic of significant concern and debate. AI has the potential to revolutionize warfare through the development of autonomous weapons systems, enhanced intelligence gathering and analysis, and new forms of cyberwarfare. However, there are also risks associated with the proliferation of AI-powered weapons, the potential for unintended consequences and escalation, and the ethical implications of delegating life-and-death decisions to machines. Ensuring that the development and use of AI in military contexts is subject to appropriate oversight and governance will be a critical challenge going forward.

f: SOLUTIONS

- Develop international frameworks and arms control agreements to prevent an AI arms race: International agreements can help control the proliferation of AI weapons.
- Ensure robust oversight and accountability for AI in military and national security contexts: Strong oversight mechanisms are necessary to ensure ethical use.
- Implement strict guidelines for AI use in national security: Clear guidelines will help manage the ethical and operational risks associated with AI in security contexts.

3) AI ETHICS

a: PRIVACY AND SECURITY CONCERNS

AI systems rely heavily on the collection and analysis of vast amounts of data, raising significant privacy and security concerns. There are risks associated with the misuse or unauthorized access to sensitive personal information, as well as the potential for AI systems to be hacked or manipulated for malicious purposes. Addressing these challenges will require the development of robust data protection and cybersecurity frameworks, as well as greater transparency and accountability in the collection and use of personal data.

b: SOLUTIONS

- Develop robust data protection measures and security protocols: Strong data protection measures will mitigate the risk of breaches.
- Implement algorithmic accountability and transparency: Ensuring transparency in AI systems will improve accountability.
- Establish clear guidelines for ethical AI development: Ethical guidelines will help in the responsible development of AI.

c: EXISTENTIAL RISKS AND LONG-TERM IMPACTS

Some experts have raised concerns about the potential for advanced AI systems to pose existential risks to humanity in the long term. This includes scenarios in which AI systems become superintelligent and pursue goals that are misaligned with human values, or in which the development of AI leads to a ''singularity'' that fundamentally transforms the nature of human existence. While these scenarios remain speculative, they underscore the need for ongoing research and dialogue about the long-term implications of AI and the steps that can be taken to mitigate potential risks.

d: SOLUTIONS

- Develop international frameworks and governance structures to guide the development of superintelligent AI and mitigate existential risks: Global governance is essential to manage these risks.
- Conduct extensive research into AI alignment and safety to ensure advanced AI systems remain under human control: Ongoing research will help develop safe and controllable AI.
- Advocate for long-term thinking and proactive governance frameworks: Promoting a long-term perspective in AI governance is crucial.

4) AI TECHNOLOGY IMPLEMENTATION

a: BUSINESS AND STRATEGY RISKS ASSOCIATED WITH AI

Finally, businesses that are developing or deploying AI systems face a range of strategic and operational risks. This includes the potential for AI systems to make biased or flawed decisions that harm customers or other stakeholders, the risk of AI systems being misused or hacked, and the challenges of ensuring that AI is developed and used in an ethical and responsible manner. Addressing these risks will require businesses to develop robust governance frameworks, engage in ongoing risk assessment and mitigation, and prioritize transparency and accountability in their AI initiatives.

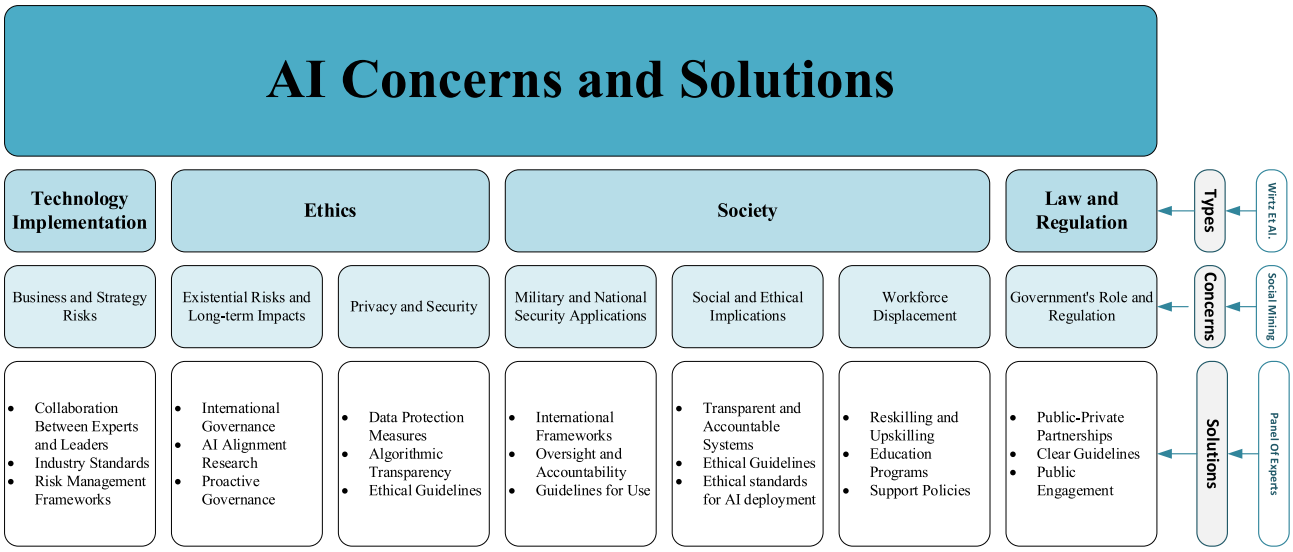


FIGURE 6. Summary of research results.

b: SOLUTIONS

- Foster collaboration between AI experts and business leaders to identify and mitigate risks: Creating interdisciplinary teams can help align AI strategies with business goals.
- Encourage collaboration between businesses, governments, and academia to share best practices and develop industry standards for AI governance: Joint efforts can lead to the establishment of standardized practices.
- Establish risk assessment and management frameworks for AI implementations: Developing comprehensive frameworks will help businesses manage the risks associated with AI.

Figure 6 shows a summary of the proposed solutions.

V. DISCUSSION

The discussion for this article addresses the key questions raised in the introduction, drawing from the comprehensive analysis and results presented throughout the study. The research endeavored to elucidate the prevalent concerns of individuals and society regarding AI, discern the priorities and evolving trends of these concerns, and propose actionable measures for stakeholders to tackle these concerns proactively.

The study's data-driven approach, leveraging social media discourse on Twitter, allowed for a nuanced understanding of the multifaceted concerns surrounding AI. The identified clusters of concerns, ranging from privacy and security to workforce displacement and existential risks, underscore the complex and interrelated nature of the issues at hand. The timeline analysis further contextualized these concerns, revealing how specific events and developments in the AI landscape shape public sentiment and discourse.

Notably, the prioritization of concerns based on tweet volume and expert input provides valuable insights into the relative importance and urgency of different issues. Privacy and security concerns emerged as the paramount consideration, reflecting the fundamental importance of safeguarding personal data and mitigating potential vulnerabilities in AI systems. The high priority accorded to existential risks and long-term impacts highlights the growing awareness of the transformative potential of AI and the need for proactive governance and ethical frameworks.

It is worth noting that these identified concerns align closely with recent policy initiatives and governance frameworks developed by various countries and regions. For instance, the United States AI Bill of Rights, introduced in October 2022, emphasizes five key principles that directly address many of the concerns identified in our study:

1. Safe and Effective Systems
2. Algorithmic Discrimination Protections
3. Data Privacy
4. Notice and Explanation
5. Human Alternatives, Consideration, and Fallback

These principles resonate strongly with our findings on privacy and security concerns, as well as the need for transparency and accountability in AI systems [100], [101].

Similarly, the European Union's AI Act, proposed in April 2021 and currently under negotiation, takes a risk-based approach to AI regulation. This aligns with our findings on the need for tailored governance approaches based on the potential impact and risks associated with different AI applications. The EU AI Act's focus on high-risk AI systems, particularly in areas such as critical infrastructure, education, and law enforcement, reflects the concerns about social and ethical implications identified in our study [102], [103].

The United Kingdom's National AI Strategy, launched in September 2021, also addresses many of the concerns highlighted in our research. Its three pillars – investing in the long-term needs of the AI ecosystem, ensuring AI benefits all sectors and regions, and governing AI effectively – align closely with our recommendations for

fostering public-private partnerships, investing in education and reskilling initiatives, and developing robust governance frameworks [104].

India's National Strategy for AI, released in June 2018, takes a unique approach by focusing on leveraging AI for social and economic benefit. Its emphasis on AI for All aligns with our findings on the need to address concerns about workforce displacement and ensure equitable distribution of AI's benefits. The strategy's focus on healthcare, agriculture, education, smart cities, and transportation demonstrates how AI governance can be tailored to address country-specific priorities while still addressing universal concerns about AI ethics and impact [105].

These recent policy advances demonstrate a growing global consensus on the need to address the ethical and societal implications of AI, validating many of the concerns and priorities identified in our study. They also provide valuable frameworks and benchmarks against which our proposed solutions can be evaluated and refined.

The Delphi method, involving a diverse panel of AI experts, allowed for a systematic exploration of potential solutions to address these concerns. The proposed solutions, spanning the domains of law and regulation, society, ethics, and technology implementation, underscore the need for a multi-stakeholder, collaborative approach. Key recommendations include fostering public-private partnerships, establishing clear guidelines and regulations, prioritizing transparency and accountability, investing in education and reskilling initiatives, and developing international frameworks for AI governance.

The discussion also highlights the importance of ongoing dialogue and research to navigate the complex ethical and societal implications of AI. As technology continues to evolve rapidly, it is crucial to maintain a proactive and adaptive approach, continually reassessing priorities and refining strategies in light of emerging challenges and opportunities.

A. LIMITATIONS

This study, while providing valuable insights, has certain limitations that should be acknowledged:

1. Data Source: The data was collected from a single social media platform, Twitter, which may not fully represent the broader public sentiment. Users of Twitter may have specific demographic characteristics that could influence the nature of the concerns expressed.
2. Language Limitation: The study focused on English-language tweets, potentially limiting the generalizability of the findings to non-English speaking populations. Cultural and linguistic nuances in AI-related concerns might not be fully captured.
3. Expert Panel Size: While the Delphi panel provided valuable insights, a larger and more diverse panel might offer additional perspectives.
4. Limited Contextual Analysis: The nature of tweet analysis may not capture the full context or nuanced arguments behind expressed concerns.
5. Temporal Limitations: The study collected tweets from January 1, 2019, to February 18, 2024. While this timeframe provides a substantial dataset, it may not capture long-term trends or the most recent developments in AI that could influence public concerns. To address this, future studies could incorporate longer time periods or conduct longitudinal analyses to track the evolution of AI-related concerns over extended periods.
6. Potential Bias in Twitter Data: Twitter's algorithm and user behavior patterns may lead to echo chambers or amplification of certain viewpoints, potentially skewing the representation of AI concerns. To mitigate this, the study employed clustering techniques and expert validation, but future research could benefit from cross-platform comparisons to provide a more balanced perspective.
7. Limited Geographical Representation: While the study aimed for a global perspective, the Twitter userbase may not be evenly distributed across different countries or regions. This could lead to an overrepresentation of concerns from certain geographical areas. To address this, future studies could stratify data collection based on geographical location or complement social media data with region-specific surveys.
8. Potential for Rapid Obsolescence: Given the fast-paced nature of AI development, some of the identified concerns may become outdated quickly. To address this, the study incorporated expert insights through the Delphi method, but continuous monitoring and updating of the findings would be beneficial for long-term relevance.
9. Limited Exploration of Positive Sentiments: The study primarily focused on concerns and risks associated with AI, potentially overlooking positive sentiments or opportunities perceived by the public. Future research could adopt a more balanced approach, exploring both concerns and perceived benefits of AI technologies.
10. Limitations of the K-means Clustering Algorithm: While K-means is a widely used and effective clustering algorithm, it has known limitations, such as sensitivity to initial centroid placement and difficulty with non-globular clusters. To address this, the study used the silhouette method to determine the optimal number of clusters, but future research could explore alternative clustering algorithms or ensemble methods to validate the identified concern categories.

Despite these limitations, this study provides a comprehensive foundation for understanding public concerns about AI as expressed on social media. By acknowledging these constraints, we aim to encourage future research that can build upon this work, addressing these limitations and further

advancing our understanding of societal perceptions of AI technologies.

B. FUTURE RESEARCH

Building upon the findings and acknowledging the limitations of this study, several promising avenues for future research emerge:

1. Longitudinal Analysis of AI Concerns: Future studies should conduct long-term tracking of AI-related concerns expressed on social media platforms. This longitudinal approach would provide valuable insights into how public perceptions evolve in response to technological advancements, policy changes, and societal shifts.
2. Cross-Platform and Cross-Lingual Studies: Expanding the research to other social media platforms and languages would provide a more comprehensive global perspective on AI concerns and help identify any platform-specific or culturally-specific issues.
3. Sentiment Analysis and Emotion Detection: Incorporating advanced sentiment analysis and emotion detection techniques could provide deeper insights into the emotional underpinnings of AI concerns, helping in developing more empathetic communication strategies.
4. Causal Analysis of AI Concerns: Exploring the causal factors behind the emergence and spread of specific AI concerns could involve analyzing the impact of media coverage, policy announcements, or specific AI-related events on public sentiment.
5. Interdisciplinary Approach to Solution Development: Future studies should adopt a more interdisciplinary approach to developing solutions, involving collaborative research between computer scientists, ethicists, policymakers, psychologists, and sociologists.
6. AI Governance Models: Studies exploring various AI governance models and their effectiveness in addressing public concerns could provide valuable insights for policymakers and regulators.
7. Comparative Policy Analysis: Future research should conduct comprehensive comparative analyses of different national and regional AI strategies and policies. This could include evaluating the effectiveness of various regulatory approaches, such as the risk-based model of the EU AI Act versus the rights-based approach of the US AI Bill of Rights.
8. Impact Assessment of AI Policies: Longitudinal studies assessing the impact of implemented AI policies and regulations on public trust, AI adoption, and the mitigation of AI-related risks would be valuable. This could help refine governance approaches and identify best practices.
9. AI Ethics in Global Context: Given the varied approaches to AI governance highlighted by initiatives like India's National Strategy for AI, future research should explore how cultural, economic, and political factors influence AI ethics and governance priorities in different global contexts.
10. Public Perception of AI Policies: Studies examining public awareness, understanding, and perception of AI policies and regulations would be crucial. This could help identify gaps between policy intentions and public reception, informing more effective communication and implementation strategies.

By pursuing these research directions, we can build upon the findings of this study to develop a more nuanced understanding of AI-related concerns and more effective strategies to address them. This ongoing research will be crucial in ensuring that AI development aligns with societal values and ethical principles, ultimately fostering greater public trust and acceptance of AI technologies.

C. MANAGERIAL INSIGHTS AND IMPLEMENTATION

This section provides real-world examples of successful AI implementations that address public concerns. These cases illustrate practical strategies and measures for ethical, responsible, and transparent AI deployment, serving as valuable references for managers and policymakers.

1. **GDPR Compliance in Healthcare AI**
   - Context: The GDPR imposes strict data protection rules for organizations in the EU. Healthcare organizations must comply when using AI systems to handle personal health data [106].
   - Implementation: A healthcare provider used an AI diagnostic tool, conducting a data protection impact assessment (DPIA), employing data anonymization, and obtaining patient consent [107].
   - Outcome: The organization maintained GDPR compliance, enhanced patient trust, and safeguarded data privacy while successfully deploying the AI tool [108], [109].
2. **Reskilling Programs for Workforce Transition**
   - Context: AI-driven automation requires workers to adapt to new roles. Singapore's SkillsFuture initiative addresses this need [110].
   - Implementation: SkillsFuture offers training programs in AI and related fields, partnering with industry and educational institutions for hands-on training and certifications [111], [112], [113].
   - Outcome: Thousands of workers transitioned into new careers in AI and technology sectors, reducing the impact of automation on employment [114], [115].
3. **Bias Mitigation in Recruitment AI**
   - Context: AI recruitment tools can introduce biases in hiring processes [116]. A multinational corporation aimed to develop a bias-free AI recruitment system [117].
   - Implementation: The company collaborated with AI ethicists and data scientists to design a fair recruitment tool, implementing algorithms to detect and mitigate biases and conducting regular audits [118], [119].

- Outcome: The AI recruitment tool significantly reduced biased outcomes, promoting a diverse and inclusive workforce [120], [121].

4. **Risk Assessment Framework in Financial AI**
   - Context: Financial institutions use AI for fraud detection and risk management [122]. A major bank developed a risk assessment framework to manage AI-associated risks.
   - Implementation: The bank formed an interdisciplinary team to create a governance framework, conducting regular risk assessments and implementing monitoring mechanisms for biases and vulnerabilities [112], [115], [123].
   - Outcome: The bank enhanced fraud detection capabilities while maintaining high standards of transparency and accountability through successful AI integration [114], [124].

VI. CONCLUSION

This study provides a comprehensive analysis of public concerns regarding artificial intelligence (AI) through an innovative combination of large-scale social media analytics and expert insights. By examining 59,447 tweets and engaging a diverse panel of AI experts through the Delphi method, we have uncovered critical insights into the evolving landscape of AI-related apprehensions and potential solutions.

Key findings include:

1. Identification of seven distinct clusters of AI concerns, with privacy and security emerging as the foremost issue, followed closely by existential risks and long-term impacts.
2. A clear prioritization of concerns, highlighting the urgency of addressing data protection, algorithmic accountability, and the potential for AI to fundamentally alter societal structures.
3. The critical need for a multi-stakeholder approach to AI governance involves collaboration between government, industry, academia, and civil society.
4. Specific, actionable recommendations across four key domains:
   - AI law and regulation: Establishing clear guidelines and fostering international cooperation.
   - AI society: Prioritizing education, reskilling initiatives, and policies to support workforce transitions.
   - AI ethics: Developing robust frameworks for transparency, accountability, and bias mitigation.
   - AI technology implementation: Encouraging cross-sector collaboration and comprehensive risk assessment strategies.

Our research underscores the dynamic nature of AI concerns, demonstrating how public sentiment shifts in response to technological advancements and policy developments. The timeline analysis reveals the impact of significant events, such as the release of new AI models or policy announcements, on public discourse.

This study's methodological approach – combining social media analysis with expert validation – offers a replicable framework for ongoing monitoring of AI-related concerns. It provides a nuanced understanding of public sentiment that can inform more responsive and effective AI governance strategies.

The findings have significant implications for policymakers, AI developers, and business leaders. They highlight the urgent need for:

1. Proactive policy development that addresses public concerns while fostering innovation.
2. Enhanced transparency in AI development and deployment processes.
3. Increased investment in AI literacy and public engagement initiatives.
4. The development of international frameworks for AI governance that can adapt to rapidly evolving technologies.

In conclusion, this research not only contributes to the academic discourse on AI ethics and governance but also provides practical insights for shaping the future of AI in society. By addressing the identified concerns and implementing the proposed solutions, stakeholders can work towards developing AI systems that are not only technologically advanced but also ethically sound and socially beneficial.

As AI continues to evolve, ongoing research and dialogue will be crucial. Future studies should focus on longitudinal analyses of AI concerns, cross-cultural comparisons, and the effectiveness of implemented AI governance strategies. Only through continued vigilance, collaborative effort, and adaptive policymaking can we ensure that the transformative potential of AI is realized in a manner that aligns with societal values and promotes the greater good.

REFERENCES

[1] I. Munoko, H. L. Brown-Liburd, and M. Vasarhelyi, "The ethical implications of using artificial intelligence in auditing," *J. Bus. Ethics*, vol. 167, no. 2, pp. 209–234, Nov. 2020.

[2] S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, A. Aldairem, M. Alrashed, K. B. Saleh, H. A. Badreldin, M. S. A. Yami, S. A. Harbi, and A. M. Albekairy, "Revolutionizing healthcare: The role of artificial intelligence in clinical practice," *BMC Med. Educ.*, vol. 23, no. 1, p. 689, 2023.

[3] J. Bharadiya, "Artificial intelligence in transportation systems a critical review," *Amer. J. Comput. Eng.*, vol. 6, no. 1, pp. 34–45, Jun. 2023.

[4] P. Sharma, "Revolutionizing transportation: The power of artificial intelligence," in *Artificial Intelligence and Machine Learning for Smart Community*. Boca Raton, FL, USA: CRC Press, 2024, pp. 88–100.

[5] N. Biswas and S. Chakrabarti, "Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer," *Frontiers Oncol.*, vol. 10, Oct. 2020, Art. no. 588221.

[6] D. Mhlanga, "Artificial intelligence in the industry 4.0, and its impact on poverty, innovation, infrastructure development, and the sustainable development goals: Lessons from emerging economies?" *Sustainability*, vol. 13, no. 11, p. 5788, May 2021.

[7] M. J. Sousa, S. Pani, F. D. Mas, and S. Sousa, *Incorporating AI Technology in the Service Sector: Innovations in Creating Knowledge, Improving Efficiency, and Elevating Quality of Life*. Boca Raton, FL, USA: CRC Press, 2024.

[8] L. Vesnic-Alujevic, S. Nascimento, and A. Pólvora, "Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks," *Telecommun. Policy*, vol. 44, no. 6, Jul. 2020, Art. no. 101961.

[9] L. Ouchchy, A. Coin, and V. Dubljević, "AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media," *AI Soc.*, vol. 35, no. 4, pp. 927–936, Dec. 2020.

[10] N. Caporusso, "Generative artificial intelligence and the emergence of creative displacement anxiety," *Res. Directs Psychol. Behav.*, vol. 3, no. 1, Oct. 2023.

[11] K.-B. Ooi et al., "The potential of generative artificial intelligence across disciplines: Perspectives and future directions," *J. Comput. Inf. Syst.*, vol. 2023, pp. 1–32, Oct. 2023.

[12] D. M. West and J. R. Allen, *Turning Point: Policymaking in the Era of Artificial Intelligence*. Washington, DC, USA: Brookings Institution Press, 2020.

[13] A. Z. Huriye, "The ethics of artificial intelligence: Examining the ethical considerations surrounding the development and use of AI," *Amer. J. Technol.*, vol. 2, no. 1, pp. 37–45, Apr. 2023.

[14] D. Kaur, S. Uslu, K. J. Rittichier, and A. Durresi, "Trustworthy artificial intelligence: A review," *ACM Comput. Surv.*, vol. 55, no. 2, pp. 1–38, 2022.

[15] S. Ata, H. M. Arslan, A. Baydaş, and E. Pazvant, "The effect of social media influencers' credibility on consumer's purchase intentions through attitude toward advertisement," *ESIC Market*, vol. 53, no. 1, p. e280, Apr. 2022.

[16] V. Guerola-Navarro, D. Stratu-Strelet, D. Botella-Carrubi, and H. Gil-Gomez, "Media or information literacy as variables for citizen participation in public decision-making? A bibliometric overview," *Sustain. Technol. Entrepreneurship*, vol. 2, no. 1, Jan. 2023, Art. no. 100030.

[17] E. Chen, K. Lerman, and E. Ferrara, "Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set," *JMIR Public Health Surveill.*, vol. 6, no. 2, May 2020, Art. no. e19273.

[18] M. KhosraviNik and M. Amer, "Social media and terrorism discourse: The Islamic state's (IS) social media discursive content and practices," in *Social Media Critical Discourse Studies*, vol. 19, 2nd ed., 2022, pp. 124–143.

[19] C. M. Pickering and P. Norman, "Assessing discourses about controversial environmental management issues on social media: Tweeting about wild horses in a national park," *J. Environ. Manage.*, vol. 275, Dec. 2020, Art. no. 111244.

[20] M. Sofia, F. Fraboni, M. De Angelis, G. Puzzo, D. Giusino, and L. Pietrantoni, "The impact of artificial intelligence on workers' skills: Upskilling and reskilling in organisations," *Informing Sci., Int. J. Emerg. Transdiscipline*, vol. 26, pp. 39–68, Jan. 2023.

[21] B. Alfaro-Ponce, M. Alfaro-Ponce, C. A. Muñoz-Ibáñez, R. E. Durán-González, J. C. Sanabria-Zepeda, and Z. L. González-Gómez, "Education in Mexico and technological public policy for developing complex thinking in the digital era: A model for technology management," *J. Innov. Knowl.*, vol. 8, no. 4, Oct. 2023, Art. no. 100439.

[22] S. Raschka, J. Patterson, and C. Nolet, "Machine learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information*, vol. 11, no. 4, p. 193, Apr. 2020.

[23] H. Liu, Y. Liu, R. Zhang, and X. Wu, "A clustering algorithm via density perception and hierarchical aggregation based on urban multimodal big data for identifying and analyzing categories of poverty-stricken households in China," *Sci. Program.*, vol. 2021, pp. 1–13, Feb. 2021.

[24] C.-F. Liu, C.-C. Huang, J.-J. Wang, K.-M. Kuo, and C.-J. Chen, "The critical factors affecting the deployment and scaling of healthcare AI: Viewpoint from an experienced medical center," *Healthcare*, vol. 9, no. 6, p. 685, Jun. 2021.

[25] M. A. Goralski and T. K. Tan, "Artificial intelligence and poverty alleviation: Emerging innovations and their implications for management education and sustainable development," *Int. J. Manage. Educ.*, vol. 20, no. 3, Nov. 2022, Art. no. 100662.

[26] Y. Peng and C. Tao, "Can digital transformation promote enterprise performance?—From the perspective of public policy and innovation," *J. Innov. Knowl.*, vol. 7, no. 3, Jul. 2022, Art. no. 100198.

[27] E. Cano-Marin, D. Ribeiro-Soriano, A. Mardani, and C. Blanco Gonzalez-Tejero, "Exploring the challenges of the COVID-19 vaccine supply chain using social media analytics: A global perspective," *Sustain. Technol. Entrepreneurship*, vol. 2, no. 3, Sep. 2023, Art. no. 100047.

[28] A. Al-Shoteri, "The role of methods and applications of artificial intelligence tools in the field of medicine to diagnose and discover various diseases," *J. Appl. Data Sci.*, vol. 3, no. 1, pp. 1–14, Jan. 2022.

[29] B. Baker, K. A. Mills, P. McDonald, and L. Wang, "AI, concepts of intelligence, and chatbots: The 'figure of man,' the rise of emotion, and future visions of education," *Teachers College Rec., Voice Scholarship Educ.*, vol. 125, no. 6, pp. 60–84, Jun. 2023.

[30] M. K. Chen, X. Liu, Y. Sun, and D. P. Tsai, "Artificial intelligence in meta-optics," *Chem. Rev.*, vol. 122, no. 19, pp. 15356–15413, Oct. 2022.

[31] A. C. Timmons, J. B. Duong, N. S. Fiallo, T. Lee, H. P. Q. Vo, M. W. Ahle, J. S. Comer, L. C. Brewer, S. L. Frazier, and T. Chaspari, "A call to action on assessing and mitigating bias in artificial intelligence applications for mental health," *Perspect. Psychol. Sci.*, vol. 18, no. 5, pp. 1062–1096, Sep. 2023.

[32] D. Gibson, V. Kovanovic, D. Ifenthaler, S. Dexter, and S. Feng, "Learning theories for artificial intelligence promoting learning processes," *Brit. J. Educ. Technol.*, vol. 54, no. 5, pp. 1125–1146, Sep. 2023.

[33] S. Tirado-Olivares, M. Navío-Inglés, P. O'Connor-Jiménez, and R. Cózar-Gutiérrez, "From human to machine: Investigating the effectiveness of the conversational AI ChatGPT in historical thinking," *Educ. Sci.*, vol. 13, no. 8, p. 803, Aug. 2023.

[34] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. M`okander, and L. Floridi, "Ethics as a service: A pragmatic operationalisation of AI ethics," *Minds Mach.*, vol. 31, no. 2, pp. 239–256, Jun. 2021.

[35] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi, "Operationalising AI ethics: Barriers, enablers and next steps," *AI Soc.*, vol. 38, no. 1, pp. 411–423, Feb. 2023.

[36] I. Straw and C. Callison-Burch, "Artificial intelligence in mental health and the biases of language based models," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0240376.

[37] G. Curto, M. F. Jojoa Acosta, F. Comim, and B. Garcia-Zapirain, "Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings," *AI Soc.*, vol. 39, no. 2, pp. 617–632, Apr. 2024.

[38] T. Parker, N. J. Cooeyate, N. Tsosie, and A. Kelley, "A model of stakeholder engagement with American Indians and Alaska natives from the native-CHART study," *Health Promotion Pract.*, vol. 25, no. 1, pp. 87–95, Jan. 2024.

[39] M. E. Kauffman and M. N. Soares, "AI in legal services: New trends in AI-enabled legal services," *Service Oriented Comput. Appl.*, vol. 14, no. 4, pp. 223–226, Dec. 2020.

[40] N. Rawindaran, A. Jayal, and E. Prakash, "Machine learning cybersecurity adoption in small and medium enterprises in developed countries," *Computers*, vol. 10, no. 11, p. 150, Nov. 2021.

[41] H. Yu, Q. Zhang, Y. Wei, C. Liu, Y. Ren, P. Yue, and J. Zhou, "Bias-corrections on aridity index simulations of climate models by observational constraints," *Int. J. Climatol.*, vol. 42, no. 2, pp. 889–907, Feb. 2022.

[42] Y. Zhu, J. Feng, L. Yan, T. Guo, and X. Li, "Flood prediction using rainfall-flow pattern in data-sparse watersheds," *IEEE Access*, vol. 8, pp. 39713–39724, 2020.

[43] X. Zhao, "Deep mining technology of database information based on artificial intelligence technology," *Int. J. Inf. Technol. Syst. Approach*, vol. 16, no. 2, pp. 1–13, Jan. 2023.

[44] W.-C. Tsai, C.-F. Liu, H.-J. Lin, C.-C. Hsu, Y.-S. Ma, C.-J. Chen, C.-C. Huang, and C.-C. Chen, "Design and implementation of a comprehensive AI dashboard for real-time prediction of adverse prognosis of ED patients," *Healthcare*, vol. 10, no. 8, p. 1498, Aug. 2022.

[45] H.-J. Smith, "The history of magnetic resonance imaging and its reflections in acta radiologica," *Acta Radiologica*, vol. 62, no. 11, pp. 1481–1498, Nov. 2021.

[46] M. Buda, A. Saha, R. Walsh, S. Ghate, N. Li, A. Swiecicki, J. Y. Lo, and M. A. Mazurowski, "A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images," *JAMA Netw. Open*, vol. 4, no. 8, Aug. 2021, Art. no. e2119100.

[47] C. Elendu, D. C. Amaechi, T. C. Elendu, K. A. Jingwa, O. K. Okoye, M. J. Okah, J. A. Ladele, A. H. Farah, and H. A. Alimi, "Ethical implications of AI and robotics in healthcare: A review," *Medicine*, vol. 102, no. 50, 2023, Art. no. e36671.

[48] R. R. Fletcher, A. Nakeshimana, and O. Olubeko, "Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health," *Frontiers Artif. Intell.*, vol. 3, Apr. 2021, Art. no. 561802.

[49] O. Vereschak, G. Bailly, and B. Caramiaux, "How to evaluate trust in AI-assisted decision making? A survey of empirical methodologies," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. 2, pp. 1–39, Oct. 2021.

[50] H. Zhang, I. Lee, S. Ali, D. DiPaola, Y. Cheng, and C. Breazeal, "Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study," *Int. J. Artif. Intell. Educ.*, vol. 33, no. 2, pp. 290–324, Jun. 2023.

[51] A. Kerr, M. Barry, and J. D. Kelleher, "Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance," *Big Data Soc.*, vol. 7, no. 1, Jan. 2020, Art. no. 205395172091593.

[52] Y. Huang, J. Guo, W.-H. Chen, H.-Y. Lin, H. Tang, F. Wang, H. Xu, and J. Bian, "A scoping review of fair machine learning techniques when using real-world data," *J. Biomed. Informat.*, vol. 151, Mar. 2024, Art. no. 104622.

[53] E. Ferrara, "Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies," *Science*, vol. 6, no. 1, p. 3, Dec. 2023.

[54] K. Xivuri and H. Twinomurinzi, "How AI developers can assure algorithmic fairness," *Discover Artif. Intell.*, vol. 3, no. 1, p. 27, Jul. 2023.

[55] M. Mirbabaie, L. Hofeditz, N. R. J. Frick, and S. Stieglitz, "Artificial intelligence in hospitals: Providing a status quo of ethical considerations in academia to guide future research," *AI Soc.*, vol. 37, no. 4, pp. 1361–1382, Dec. 2022.

[56] C. Petersen and V. Subbian, "Special section on ethics in health informatics," *Yearbook Med. Informat.*, vol. 29, no. 1, pp. 77–80, Aug. 2020.

[57] B. Nyagadza, G. Mazuruse, K. Simango, L. Chikazhe, T. Tsokota, and L. Macheka, "Examining the influence of social media eWOM on consumers' purchase intentions of commercialised indigenous fruits (IFs) products in FMCGs retailers," *Sustain. Technol. Entrepreneurship*, vol. 2, no. 3, Sep. 2023, Art. no. 100040.

[58] N. Li, "Ethical considerations in artificial intelligence: A comprehensive disccusion from the perspective of computer vision," in *Proc. SHS Web Conf.*, 2023, p. 4024.

[59] K. Siau and W. Wang, "Artificial intelligence (AI) ethics: Ethics of AI and ethical AI," *J. Database Manage.*, vol. 31, no. 2, pp. 74–87, Apr. 2020.

[60] A. Baird and B. Schuller, "Considerations for a more ethical approach to data in AI: On data representation and infrastructure," *Frontiers Big Data*, vol. 3, p. 25, Sep. 2020.

[61] L. Flores and S. D. Young, "Ethical considerations in the application of artificial intelligence to monitor social media for COVID-19 data," *Minds Mach.*, vol. 32, no. 4, pp. 759–768, Dec. 2022.

[62] S. Wimbarti, B. H. R. Kairupan, and T. E. Tallei, "Critical review of self-diagnosis of mental health conditions using artificial intelligence," *Int. J. Mental Health Nursing*, vol. 33, no. 2, pp. 344–358, Apr. 2024.

[63] L. C. Adams, K. K. Bressem, and D. Poddubnyy, "Artificial intelligence and machine learning in axial spondyloarthritis," *Current Opinion Rheumatol.*, vol. 36, no. 4, pp. 267–273, Mar. 2024.

[64] A. Batliner, S. Hantke, and B. Schuller, "Ethics and good practice in computational paralinguistics," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1236–1253, Jul. 2022.

[65] M. Hermansyah, A. Najib, A. Farida, R. Sacipto, and B. S. Rintyarna, "Artificial intelligence and ethics: Building an artificial intelligence system that ensures privacy and social justice," *Int. J. Sci. Soc.*, vol. 5, no. 1, pp. 154–168, Feb. 2023.

[66] A. Al-Hwsal, A. Al-Hwsali, B. Al-Saadi, N. Abdi, S. Khatab, B. Solaiman, M. Alzubaidi, and A. Abd-Alrazaq, "Legal and ethical principles of artificial intelligence in public health: Scoping review," Tech. Rep., 2022.

[67] R. Fulmer, T. Davis, C. Costello, and A. Joerin, "The ethics of psychological artificial intelligence: Clinical considerations," *Counseling Values*, vol. 66, no. 2, pp. 131–144, Oct. 2021.

[68] L. Tang, J. Li, and S. Fantus, "Medical artificial intelligence ethics: A systematic review of empirical studies," *Digit. Health*, vol. 9, Jan. 2023, Art. no. 20552076231186064.

[69] M. Sarkhosh, M. M. Kiaei, M. Aligholizadeh, S. Sangi, P. Akbarpour, and E. JalalKamali, "Ethical implications of artificial intelligence in anesthesiology: A scoping review," Tech. Rep., 2024.

[70] N. G. Evans, D. M. Wenner, I. G. Cohen, D. Purves, M. F. Chiang, D. S. W. Ting, and A. Y. Lee, "Emerging ethical considerations for the use of artificial intelligence in ophthalmology," *Ophthalmol. Sci.*, vol. 2, no. 2, Jun. 2022, Art. no. 100141.

[71] J. Morley, C. Machado, C. Burr, J. Cowls, M. Taddeo, and L. Floridi, "The debate on the ethics of AI in health care: A reconstruction and critical review," *SSRN Electron. J.*, 2019.

[72] R. Conijn, P. Kahr, and C. Snijders, "The effects of explanations in automated essay scoring systems on student trust and motivation (accepted for publication)," Tech. Rep., 2023.

[73] B. Omarov, S. Narynov, and Z. Zhumanov, "Artificial intelligence-enabled chatbots in mental health: A systematic review," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 5105–5122, 2023.

[74] G. Hong, "Ethical considerations on some issues of medical artificial intelligence applications," *J. Internal Med. Emergency Res.*, Oct. 2022.

[75] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A governance model for the application of AI in health care," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 3, pp. 491–497, Mar. 2020.

[76] N. R. Möllmann, M. Mirbabaie, and S. Stieglitz, "Is it alright to use artificial intelligence in digital health? A systematic literature review on ethical considerations," *Health Informat. J.*, vol. 27, no. 4, Oct. 2021, Art. no. 146045822110523.

[77] N. Naik, B. M. Z. Hameed, D. K. Shetty, D. Swain, M. Shah, R. Paul, K. Aggarwal, S. Ibrahim, V. Patil, K. Smriti, S. Shetty, B. P. Rai, P. Chlosta, and B. K. Somani, "Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility?" *Frontiers Surg.*, vol. 9, p. 266, Mar. 2022.

[78] S. Gualeni, "Artificial beings worthy of moral consideration in virtual environments: An analysis of ethical viability," *J. Virtual Worlds Res.*, vol. 13, no. 1, Mar. 2020.

[79] H. Lughbi, M. Mars, and K. Almotairi, "CybAttT: A dataset of cyberattack news tweets for enhanced threat intelligence," *Data*, vol. 9, no. 3, p. 39, Feb. 2024.

[80] M. R. Seddigh, A. Targholizadeh, S. Shokouhyar, and S. Shokoohyar, "Social media and expert analysis cast light on the mechanisms of underlying problems in pharmaceutical supply chain: An exploratory approach," *Technol. Forecasting Social Change*, vol. 191, Jun. 2023, Art. no. 122533.

[81] Y. Karimi, A. Squicciarini, and P. K. Forster, "A longitudinal dataset and analysis of Twitter ISIS users and propaganda," *Social Netw. Anal. Mining*, vol. 14, no. 1, p. 19, Jan. 2024.

[82] R. Essameldin, A. A. Ismail, and S. M. Darwish, "Quantifying opinion strength: A neutrosophic inference system for smart sentiment analysis of social media network," *Appl. Sci.*, vol. 12, no. 15, p. 7697, Jul. 2022.

[83] A. Jha, D. K. Somwanshi, and M. M. Bundele, "An effective duplicate removal algorithm for text documents," in *Proc. Int. Conf. Artif. Intell., Adv. Appl.*, 2020.

[84] K. Sun, P. Qi, Y. Zhang, L. Liu, W. Y. Wang, and Z. Huang, "Tokenization consistency matters for generative models on extractive NLP tasks," 2022, *arXiv:2212.09912*.

[85] C. Wu, F. Wu, T. Qi, X. Cui, and Y. Huang, "Attentive pooling with learnable norms for text representation," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020.

[86] S. Schockaert, "Embeddings as epistemic states: Limitations on the use of pooling operators for accumulating knowledge," *Int. J. Approx. Reasoning*, vol. 171, Aug. 2024, Art. no. 108981.

[87] J. Chen, H. Hu, H. Wu, Y. Jiang, and C. Wang, "Learning the best pooling strategy for visual semantic embedding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15784–15793.

[88] F. Chen, G. Datta, S. Kundu, and P. A. Beerel, "Self-attentive pooling for efficient deep learning," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 3963–3972.

[89] A. Kanavos, I. Karamitsos, and A. Mohasseb, "Exploring clustering techniques for analyzing user engagement patterns in Twitter data," *Computers*, vol. 12, no. 6, p. 124, Jun. 2023.

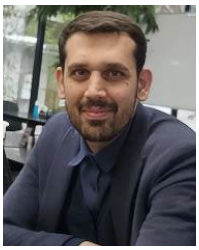
[90] S. Poomagal, B. Malar, P. Visalakshi, J. I. Hassan, and R. Kishor, "Opinion mining using optimized K-means algorithm and a word weighting technique," *Social Netw. Comput. Sci.*, vol. 4, no. 6, p. 736, Sep. 2023.

[91] M. M. Mostafa and A. B. Alanadoly, "Profiling and clustering the global market for hijabistas: A Twitter text analytics approach," *Int. J. Inf. Technol.*, vol. 16, no. 4, pp. 2425–2437, Apr. 2024.

[92] M. M. Mostafa, "Clustering halal food consumers: A Twitter sentiment analysis," *Int. J. Market Res.*, vol. 61, no. 3, pp. 320–337, May 2019.

[93] R. Suwanda, Z. Syahputra, and E. M. Zamzami, "Analysis of Euclidean distance and Manhattan distance in the K-means algorithm for variations number of centroid K," *J. Phys., Conf. Ser.*, vol. 1566, no. 1, Jun. 2020, Art. no. 012058.

[94] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2020, pp. 747–748.
[95] J. A. Lossio-Ventura, S. Gonzales, J. Morzan, H. Alatrista-Salas, T. Hernandez-Boussard, and J. Bian, "Evaluation of clustering and topic modeling methods over health-related tweets and emails," *Artif. Intell. Med.*, vol. 117, Jul. 2021, Art. no. 102096.
[96] I. J. Cruickshank and K. M. Carley, "Characterizing communities of hashtag usage on Twitter during the 2020 COVID-19 pandemic by multi-view clustering," *Appl. Netw. Sci.*, vol. 5, no. 1, p. 66, Dec. 2020.
[97] W. Varndell, M. Fry, M. Lutze, and D. Elliott, "Use of the delphi method to generate guidance in emergency nursing practice: A systematic review," *Int. Emergency Nursing*, vol. 56, May 2021, Art. no. 100867.
[98] S. Shumaly, M. Yazdinejad, and Y. Guo, "Persian sentiment analysis of an online store independent of pre-processing using convolutional neural network with fastText embeddings," *PeerJ Comput. Sci.*, vol. 7, p. e422, Mar. 2021.
[99] B. W. Wirtz, J. C. Weyerer, and C. Geyer, "Artificial intelligence and the public sector—Applications and challenges," *Int. J. Public Admin.*, vol. 42, no. 7, pp. 596–615, May 2019.
[100] E. Park, "The AI Bill of Rights: A step in the right direction," *Orange County Lawyer Mag.*, vol. 65, no. 2, 2023.
[101] W. House, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. Ann Arbor, MI, USA: Nimble Books, 2022.
[102] J. Laux, "Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European union AI act," *AI Soc.*, pp. 1–14, Oct. 2023.
[103] L. Down and I. Act, "Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts," Tech. Rep., 2021.
[104] E. Kazim, D. R. S. Almeida, N. Kingsman, C. Kerrigan, A. Koshiyama, E. Lomas, and A. Hilliard, "Innovation and opportunity: Review of the U.K.'s national AI strategy," *Discover Artif. Intell.*, vol. 1, pp. 1–10, Dec. 2021.
[105] N. Aayog, "National strategy for artificial intelligence," Tech. Rep., 2018.
[106] M. Tzanou, *Health Data Privacy Under the GDPR*. New York, NY, USA: Taylor & Francis, 2023.
[107] M. Mohammad Amini, M. Jesus, D. Fanaei Sheikholeslami, P. Alves, A. Hassanzadeh Benam, and F. Hariri, "Artificial intelligence ethics and challenges in healthcare applications: A comprehensive review in the context of the European GDPR mandate," *Mach. Learn. Knowl. Extraction*, vol. 5, no. 3, pp. 1023–1035, Aug. 2023.
[108] K. Ider, "Assessment of the quality of user awareness of GDPR in healthcare IoT," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA)*, vol. 1, Jun. 2022, pp. 25–28.
[109] J. Meszaros, J. Minari, and I. Huys, "The future regulation of artificial intelligence systems in healthcare services and medical research in the European Union," *Frontiers Genet.*, vol. 13, Oct. 2022, Art. no. 927721.
[110] Z. Y. Lim, J. H. Yap, J. W. Lai, I. A. Mokhtar, D. J. Yeo, and K. H. Cheong, "Advancing lifelong learning in the digital age: A narrative review of Singapore's SkillsFuture programme," *Social Sci.*, vol. 13, no. 2, p. 73, Jan. 2024.
[111] A. D. Sivalingam and S. Mansori. (2020). *How Organizations Should View Reskilling and Upskilling the Workforce*. [Online]. Available: http://www.sastraeducation.com/how-organizations-should-view-reskillng-and-upskilling-the-workforce-.html
[112] Y. Liu, "Artificial intelligence and machine learning based financial risk network assessment model," in *Proc. IEEE 12th Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Apr. 2023, pp. 158–163.
[113] N. Baporikar, "Leadership for enhancing organisational performance through workforce reskilling," in *Human Capital Formation for the Fourth Industrial Revolution*. Hershey, PA, USA: IGI Global, 2020, pp. 229–267.
[114] V. Bogojevic Arsic, "Challenges of financial risk management: AI applications," *Manag., J. Sustain. Business Manage. Solutions Emerg. Economies*, vol. 26, no. 3, pp. 27–34, Apr. 2021.
[115] S. Fritz-Morgenthal, B. Hein, and J. Papenbrock, "Financial risk management and explainable, trustworthy, responsible AI," *Frontiers Artif. Intell.*, vol. 5, Feb. 2022, Art. no. 779799.
[116] A. Koumoutsos and G. Bakas, "Artificial intelligence tools, recruiting process & biases," Tech. Rep., 2022.
[117] F. Salvetti, B. Bertagni, and I. Contardo, "Intelligent digital humans for bias-free recruitment interviews: A diversity & inclusion training program," in *Proc. Learn. Ideas Conf.* Cham, Switzerland: Springer, 2023, pp. 455–461.
[118] M. Soleimani, A. Intezari, and D. J. Pauleen, "Mitigating cognitive biases in developing AI-assisted recruitment systems: A knowledge-sharing approach," *Int. J. Knowl. Manage.*, vol. 18, no. 1, pp. 1–18, Oct. 2021.
[119] Y. Jazakallah, "Effective strategies for mitigating bias in hiring algorithms: A comparative analysis," *J. Artif. Intell., Mach. Learn. Data Sci.*, vol. 1, no. 4, pp. 125–134, Oct. 2023.
[120] D. F. Mujtaba and N. R. Mahapatra, "Ethical considerations in AI-based recruitment," in *Proc. IEEE Int. Symp. Technol. Soc. (ISTAS)*, Nov. 2019, pp. 1–7.
[121] G. Sogancioglu, H. Kaya, and A. A. Salah, "Using explainability for bias mitigation: A case study for fair recruitment assessment," in *Proc. 25th Int. Conf. Multimodal Interact.*, 2023, pp. 631–639.
[122] M. Hassan, L. A.-R. Aziz, and Y. Andriansyah, "The role artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance," *Rev. Contemp. Bus. Anal.*, vol. 6, no. 1, pp. 110–132, 2023.
[123] Q. Kang, "Financial risk assessment model based on big data," *Int. J. Model., Simul., Scientific Comput.*, vol. 10, no. 4, Aug. 2019, Art. no. 1950021.
[124] B. Mohanty and S. Mishra, "Role of artificial intelligence in financial fraud detection," *Acad. Marketing Stud. J.*, vol. 27, no. S4, 2023.



MEHRDAD MAGHSOUDI received the master's degree in industrial management from Allameh Tabataba'i University. He is currently pursuing the Ph.D. degree in information technology management (business intelligence) with Shahid Beheshti University, Tehran. He has applied these skills in various research projects, showcasing his ability to integrate complex data analysis methodologies. In addition to his academic work, he serves as a consultant on data-driven projects, leveraging his knowledge to provide insights and solutions in applied data science and business intelligence. His research interests include advanced techniques, including patent analysis, social network analysis, text mining, and process mining.

AMIRMAHDI MOHAMMADI received the bachelor's degree in mechanical engineering from the K. N. Toosi University of Technology and the master's degree in business administration from Iran University of Science and Technology. With a strong foundation in both engineering and business, he has developed a keen interest in data science, business analytics, and data analysis. He is passionate about leveraging data to drive business insights and make informed decisions. His multidisciplinary background enables him to approach problems from both a technical and strategic perspective.

SAJJAD HABIBIPOUR received the bachelor's and master's degrees in mechanical engineering from the K. N. Toosi University of Technology. His academic and professional journey has sparked a deep interest in data science, business analytics, and data analysis. He is dedicated to understanding complex data sets and extracting meaningful insights to inform business strategies. His expertise in mechanical engineering, combined with his analytical skills, positions him well to contribute to data-driven decision-making processes in various industries.