Mapping AI ethics narratives: evidence from Twitter discourse between 2015 and 2022

Mengyi Wei, Puzhen Zhang, Chuan Chen, Dongsheng Chen, Chenyu Zuo, Liqiu Meng

ARTICLE

OPEN

Check for updates

Mapping AI ethics narratives: evidence from Twitter discourse between 2015 and 2022

Mengyi Wei [1], Puzhen Zhang[1], Chuan Chen [1✉], Dongsheng Chen[1], Chenyu Zuo[2] & Liqiu Meng[1]

The ethical issues that arise in the development of AI technologies are closely linked to public engagement. Although Twitter, as an online public sphere, provides a platform for exploring AI ethics discourse, it is difficult for current research to effectively extract fine-grained but meaningful information from the vast amount of social media data. To address this challenge, this paper proposes a research framework for the fine-grained exploration of AI ethics discourse on Twitter. The framework consists of two main parts: (1) combining neural networks with large-scale language models to construct a hierarchically structured topic framework that not only extracts popular topics of public interest, but also highlights smaller, yet significant voices; (2) using narrative metaphors to achieve the integration of fragmented information across levels and topics, ultimately presenting a complete story to help the public better understand the evolution of topics within AI ethics discourse. Our research has revealed that the most significant concern in the current AI ethics discourse is the lag in AI-related laws and ethical guidelines. It also shows that the integration of AI technology with the humanities is essential to promote a good public society. Through cross-level fine-grained mining, this study uncovers information hidden beneath the noise interference, which helps policymakers make targeted adjustments or improvements to policies. In addition, this research framework provides a reference for fine-grained mining of other specific issues in social media data.

[1] Chair of Cartography and Visual Analytics, Technical University of Munich, Munich, Germany. [2] Center for Sustainable Future Mobility (CSFM), ETH Zurich, Zurich, Switzerland. ✉email: chuan.chen@tum.de

Introduction

The essence of AI (Artificial Intelligence) technology is to serve human beings. However, the current development of AI is in the hands of technical and economic elites and lacks active public engagement (Corbett et al., 2023; Dollbo, 2023; Gilman, 2023). The public is reduced to consumers simply waiting for a new generation of AI technology. If public discourse and sentiment are ignored, the decision-making power on how to better deploy AI will be fall into special interests (Rainie, 2018). With the advent of the intelligent era, social media data are increasingly rich in content and information dimensions, allowing the possibility to study AI ethics from the public perspective in an egalitarian manner (Manovich, 2012). Although social media represents only the voices of the users, its powerful influence makes it an important choice for studying public discourse (Chen et al., 2023; Fu et al., 2023; Hua et al., 2022; Liang et al., 2017).

However, information extraction from Twitter data in various fields often focuses on simple topic mining and sentiment analysis, which lacks in-depth exploration of the information beneath the surface. For example, Twitter data has been used to study AI ethics issues in blockchain, identifying only three main topics: security, fairness, and emotional sentiments (Fu et al., 2022). The analysis of tweets during the first month after the launch of ChatGPT simply extracted ten topics that the public was interested in (Taecharungroj, 2023). Social media data is characterized by its massive, unstructured, and diverse nature. Much meaningful information is often drowned out by interference and noise, limiting the ability of researchers to extract valuable insights. A wealth of unpredictable, deep information is hidden beneath the surface.

Considering the characteristics of social media data, this study focuses on achieving cross-level, fine-grained exploration of AI ethics discourse on Twitter. The hierarchical structure of data refers to nested relationships organized according to certain logical or hierarchical connections. Data often has a hierarchical structure or can be computed as such. Data at different structural levels have varying degrees of semantic granularity, and hierarchical data is more conducive to expressing fine-grained exploration (Kontaxaki et al., 2010). Cross-level information exploration investigates the relationships between multiple levels, enabling flexible analysis of complex, large-scale information at different granularities (Y Feng et al., 2023). Fine-grained data refers to detailed information expressed through specific scales and levels, often representing lower-level structures or smaller data units. Fine-grained exploration of Twitter data can uncover hidden information buried in massive amounts of data across multiple dimensions. This unanticipated data mining is an exploratory approach to data analysis that provides new insights for problem solving and decision optimization by revealing overlooked patterns or unforeseen anomalies.

Visual analytics is an alternative data analysis tool in unanticipated data mining (Keim et al., 2008). Especially in the fine-grained exploration of social media data, even if adequate information is obtained through fine-grained exploration, the information is often too fragmented to be easily interpreted (Zuo et al., 2022). A readable, efficient, and easy-to-understand narrative can be constructed only by using visualization analysis methods to assemble fragmented information. Based on the characteristics of AI ethics data on Twitter, this study combines appropriate visualization analysis methods from the perspectives of data spatiality and topic evolution to achieve a cross-level, fine-grained exploration of AI ethics discourse.

In this study, we aim to address the following two research questions:

1. What is the hierarchically structured topic related to AI ethics discourse on Twitter?
2. How can cross-level fine-grained information in AI ethics discourse on Twitter be effectively understood?

The second section provides an overview of the relevant research, followed by an introduction to the research methods in the third section. The results of the two research questions are presented in the fourth section. The fifth section contains a discussion of the results. The limitations and implications of this study are discussed in the sixth section, followed by a conclusion in the final section.

Related work

The literature review of this paper is structured into three aspects. Firstly, it delves into examining research about AI ethics within the context of Twitter data. Secondly, it outlines the methodologies employed for topic extraction in tweets, analyzing their merits and drawbacks. Lastly, it provides a comprehensive overview of visualization studies concerning social media data.

AI Ethics in Twitter Data. While Twitter may not fully represent the entirety of the social media landscape (Dijck, 2013), it has emerged as a prominent platform for individuals to express opinions and engage in online discourse (Anger and Kittl, 2011). Additionally, it serves as a valuable research tool for scholars investigating various scientific topics (Chen et al., 2023; Fu et al., 2022; Hua et al., 2022). Research on AI ethics utilizing Twitter data has been focused either on specific topics in a domain, such as security, equity, and emotional sentiments in the blockchain domain, with the aim to address fairness concerns in blockchain design related to transaction ordering (Fu et al., 2022), or on a single category of ethical issues, such as using Twitter data to examine the relationship between individual privacy settings and self-disclosure on Twitter, considering cultural values across different contexts (Liang et al., 2017). Recent studies have also investigated public discussions and reactions to ChatGPT (Haque et al., 2022). Specifically, by analyzing tweets within a month of ChatGPT's launch, the study identified the most relevant topics and sentiments of the public, and ethical challenges to be addressed as ChatGPT develops (Taecharungroj, 2023). While such studies shed light on prevailing topics of public discourse and sentiment analysis on specific AI domains or ethical concerns, few provide a comprehensive analysis of AI ethics-related topics on Twitter.

Furthermore, research related to Twitter data often focuses on two types of analysis: trending public discussion topics and sentiment changes. For instance, in a blockchain study related to AI ethics, the top 30 keywords associated with #Flashbots and #MEV were extracted (Fu et al., 2022). The tweets about ChatGPT were categorized into nine topics, ranging from "Future Career & Opportunities" to "Disruptions for Software", and the corresponding sentiment distribution for each topic was presented in (Haque et al., 2022). Taecharungroj, (2023) employed the LDA method to extract public topics on ChatGPT, mainly focusing on news, technology, and reactions, and identified five functional domains: creative writing, essay writing, prompt writing, code writing, and answering questions. Topic extraction and sentiment analysis played an important role in using Twitter data to reveal public viewpoints and expressions on online social platforms (Boon-Itt and Skunkan, 2020; Hua et al., 2022). However, this information represents only a small part of Twitter data and is often too fragmented to form coherent, readable information. This study focuses on extracting fine-grained and rich information from Twitter and attempts to present it in a

more readable and coherent narrative, thereby providing a comprehensive perspective for the public to understand AI ethics.

Topic mining. Topic extraction methods in Twitter data are mainly divided into network-based methods and text-based methods. Network-based topic classification methods include social network analysis, graph mining, and topic propagation models. Lee et al., (2011) identified the top five similar topics among 18 popular categories on Twitter based on the number of influential users in common and validated that network-based classification modeling methods can achieve up to 70% classification accuracy. Azam et al. (2015) proposed a social graph generation method, treating tweets as nodes and using the Markov clustering technique to decompose the social graph into various clusters, each corresponding to specific events, thereby achieving event classification. Huang and Mu, (2014) employed a combination of clustering algorithms with hashtag propagation algorithms to detect topics on Twitter, to classify the tweets into different clusters and then using a label propagation mechanism to label tweets that overlap in different clusters. Finally, this method was compared with other clustering algorithms, validating the accuracy of the label-propagation-based algorithm.

Text-based topic extraction from Twitter data includes keyword extraction and topic modeling (Karami et al., 2020). A representative method for keyword extraction is the TF-IDF (Term Frequency–Inverse Document Frequency). Alsaedi et al. (2016) proposed a novel temporal term TF-IDF method, which can overcome the drawbacks of traditional methods that require prior knowledge of the entire dataset by assuming that "words with higher frequencies in documents within specific periods are more likely to be selected for human-created document summaries," and validated the superiority of this method. With regard to topic modeling for the extraction of topics from Twitter data is often, LDA (Latent Dirichlet Allocation) is a commonly used method. Chen et al. (2023) used the LDA method to obtain topics related to "climate strikes" in Twitter data from 2018 to 2021, providing a reference for using social media to construct political issues and collective actions. However, these topic mining methods often suffer from a large amount of noise and a single structured nature, failing to effectively mine valuable information from social media data. Therefore, we proposed to combine neural networks with large language models, presenting hierarchical topic structure, and achieving fine-grained classification of social media data.

Visual analytics for AI ethical data analysis. Data visualization is a data analysis technique that combines data analysis methods (such as machine learning and data mining) with information visualization. It is not only a tool for presenting data, but also a research method for exploring data, discovering patterns, and generating insights. For researchers, visual analytics is particularly useful for unanticipated data mining, as it can reveal unforeseen or easily overlooked results. Studies have shown that humans process visual information much more effectively than textual information (Yang and Jin, 2020). When applied to social media data mining, visual analytics allows researchers to uncover hidden patterns and potential relationships in the data without pre-determining the results. Visual analytics presents the results of data analysis to the public in a graphical format, making complex data relationships intuitive and easy to understand. This significantly lowers the barrier to understanding, especially for non-technical users, and enables information to be communicated quickly and clearly. In the field of AI ethics, the application of visual analytics breaks down knowledge barriers and enables people from different backgrounds to understand AI ethics-

related information. This serves as a basis for improving public engagement with AI development.

As an effective tool for information discovery, visual analytics emphasizes different aspects of information depending on its developmental focus. Story maps typically refer to visual explanations similar to the construction of semantic maps, webs, or networks that present geographically relevant information, events, or themes in a narrative format (Davis and McPherson, 1989; Freedman and Reynolds, 1980; Roth, 2021). Based on the spatial characteristics of social media data, story maps use maps to display the geographic distribution characteristics of the data. For example, Kwon et al. (2023) conducted a comprehensive analysis of how emotions evolve with topics and geographic locations using time series analysis and geographic visualization to capture the dynamic nature of emotions. This approach helped to identify and visualize changes in emotional trends across different geographical regions and topics on Twitter. In AI ethics research, story maps are used to represent the geographic distribution characteristics of Twitter data and to place the data in a specific context through narrative techniques. For example, maps can illustrate how AI ethics discourse spreads globally and how cultural or political factors influence it.

Developing fine-grained information from massive social media data is crucial for data mining (Kontaxaki et al., 2010). Information granularity refers to the level of detail or abstraction at which information is described or expressed. It reflects the degree of precision in analyzing, processing, or presenting data or information. The finer the granularity, the more detailed the information (Kontaxaki et al., 2010). Exploring data at a granular level can reveal more detailed information, but such fine-grained exploration can lead to overly fragmented information, making it difficult for people to understand effectively. This is one of the main reasons why fine-grained information is often overlooked in social media data mining. However, integrating fragmented information and uncovering the abstract narratives hidden beneath the surface is key to enabling people to recognize and understand the importance of fine-grained information(J Feng et al., 2023). Chen (2018) proposed a story synthesis theory, where complex information is integrated into a coherent story using story synthesis support functions. Based on information visualization methods, this study explores fine-grained information related to AI ethics discourse in Twitter data by creating topic evolution view and constructing narrative metaphors to help the public intuitively understand the evolution of AI ethics discourse.

Data and methodology

This section presents the research data and methods, discussing the establishment of the dataset, construction of hierarchical topic structures, presentation of the story map, and topic evolution analysis. Figure 1 provides the entire framework of the study from twitter data to visual analytics.

Dataset construction. To collect tweets related to AI ethics, we used topic tags to identify discourse communities revolving around specific topics (Chen et al., 2023; Jost et al., 2018). We selected seven synonymous expressions (#AI ethics, #Artificial Intelligence ethics, #Ethics of AI, #ai ethics, #Ethics In AI, #Ethical AI) (These search terms have spaces added for clarity, but there are no spaces when searching on Twitter.) to focus on AI ethics in general rather than a single ethical theme, and collected tweets from 1 January 2015 to 31 December 2022, using Python and the Twitter API ($N = 539{,}743$). Only English-language tweets were considered, including text, user location, posting time, user occupation, user verification, and hashtag. We
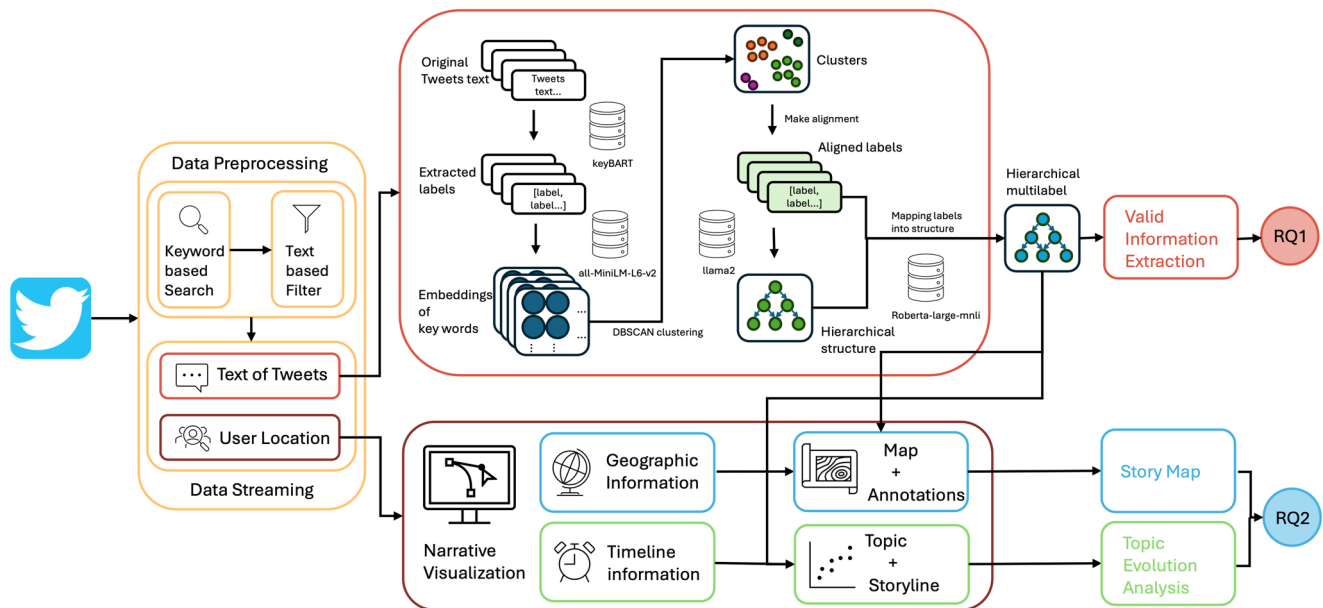
Fig. 1 The Research Framework from Twitter Data to Narrative Visualization.

geocoded the entire tweet dataset, converted textual location descriptions into geographical coordinates, and constructed a structured geographic spatial dataset.

Methodology. To address the two research questions, we conducted tasks of hierarchical topic extraction, sentiment analysis, and narrative visualization of the data.

Hierarchical topic extraction. Traditional topic extraction methods often require extensive data preprocessing and do not allow for creating hierarchically structured topics. We adopt an approach based on the combination of neural networks and large language models to build hierarchically structured topics.

First, the raw text data is fed into the KeyBART model (Kulkarni et al., 2022), and highly summarized vital phrases are generated by setting the num_beams parameter to 10. The value of num_beams determines the number of candidate sequences considered at each step. Typically, the value of num_beams is set between 3 and 10. Generally, a larger num_beams can improve the quality of the generated text. The words within each phrase were then lexically reduced using NLTK's grammar reduction library to align tags with similar grammatical morphology. Next, the reduced labels were converted to vectors using the all-MiniLM-L6-v2 model (Reimers and Gurevych, 2019), and density-based spatial clustering was applied to the application using a noise clustering approach to generate initial clusters (Schubert et al., 2017). Next, the aligned labels are fed into the LLAMA2 model to generate a hierarchical labeling structure (Touvron et al., 2023), and the labels are mapped to the bottom layer in the structure using the RoBERTa-large-mnli model (Liu et al., 2019). Finally, 64 labels at the bottom layer detected in the given dataset are extracted as fine-grained topics. With the hierarchical structure, the percentage of the number of original tweets corresponding to each topic can be calculated. The method realizes the conversion from raw text data to a hierarchical tag structure. In the process, multiple models and algorithms are utilized to align semantically similar tags, thus achieving efficient processing and analysis of the dataset.

Sentiment analysis. To better illustrate the story map related to AI ethics, we employ the TweetNLP integrated platform for

sentiment analysis of Twitter data (Camacho-collados et al., 2022). The core of TweetNLP is based on Transformer language models, which no longer rely on generic models or train language models from scratch but continue the training from RoBERTa and XLM-R checkpoints on Twitter-specific corpora (Conneau et al., 2020; Liu et al., 2019), providing more reliable analysis results (Devlin et al., 2019; Nguyen et al., 2020).

The sentiment analysis of TweetNLP aims to predict the sentiment of a tweet, including three labels: positive, neutral, or negative. Through TweetNLP, we predict the sentiment class of each tweet and provide a score to explain the confidence of the prediction. Moreover, tweets categorized with sentiment are further used to extract keywords of positive or negative sentiments, which helps to analyze the reasons for different sentiments.

Data visualization methods
Story mapping: To extract richer and more meaningful information from Twitter data, we visualize the data on maps to reveal hidden spatial narratives. Story maps intertwine geographic locations with relevant information to narrate spatial distribution stories. This paper primarily maps tweets related to AI ethics, including their topics and sentiments, onto corresponding spatial locations, supplemented with relevant text and image information to present a story map of AI ethics.

Topic evolution view: The data processing of Twitter data typically yields fragmented information, making it challenging to form a coherent and understandable narrative. Therefore, we construct a topic evolution view to integrate scattered information into a comprehensive story that allows the investigation of the development and significance of the AI ethics discourse. Since fragmented social media information may overlook deep semantic understanding when analyzed merely by computing relevance or highest frequency, we employ semantic similarity calculation to analyze the evolution of AI ethics discourse. We select specific keywords as starting words, such as the keywords appearing most frequently in January 2015, and calculate the top five keywords with the highest semantic similarity to the starting word annually as its evolutionary words. This paper utilizes the all-MiniLM-L6-V2 model to convert text labels into vectors
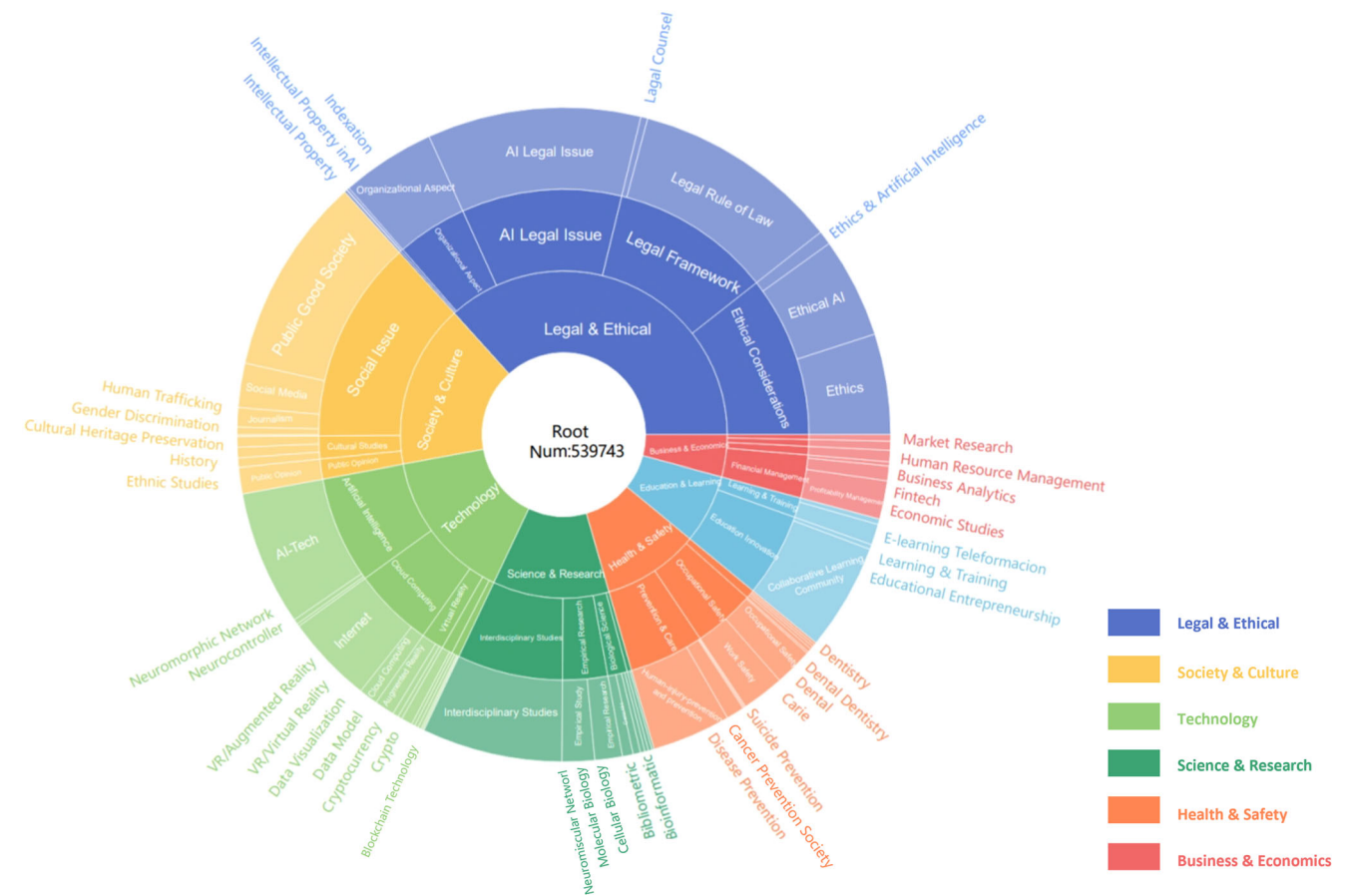
Fig. 2 A Hierarchical Thematic Structure of AI Ethics Twitter Discourse.

(Reimers and Gurevych, 2019), and then use vector representations to compute the similarity between two text fragments, with the cosine similarity formula as follows:

$$Cosine\ Similarity(A, B) = \frac{A \cdot B}{||A||||B||} \quad (1)$$

where A and B are the vector representations of the two text labels, · denotes the dot product of the vectors, and ||A|| and ||B|| are the norms (i.e., lengths) of the vectors A and B, respectively.

Results

The findings of this paper consist of an overview of the hierarchical topic structure, a story map with geographic locations, and a topic evolution view integrating fragmented information.
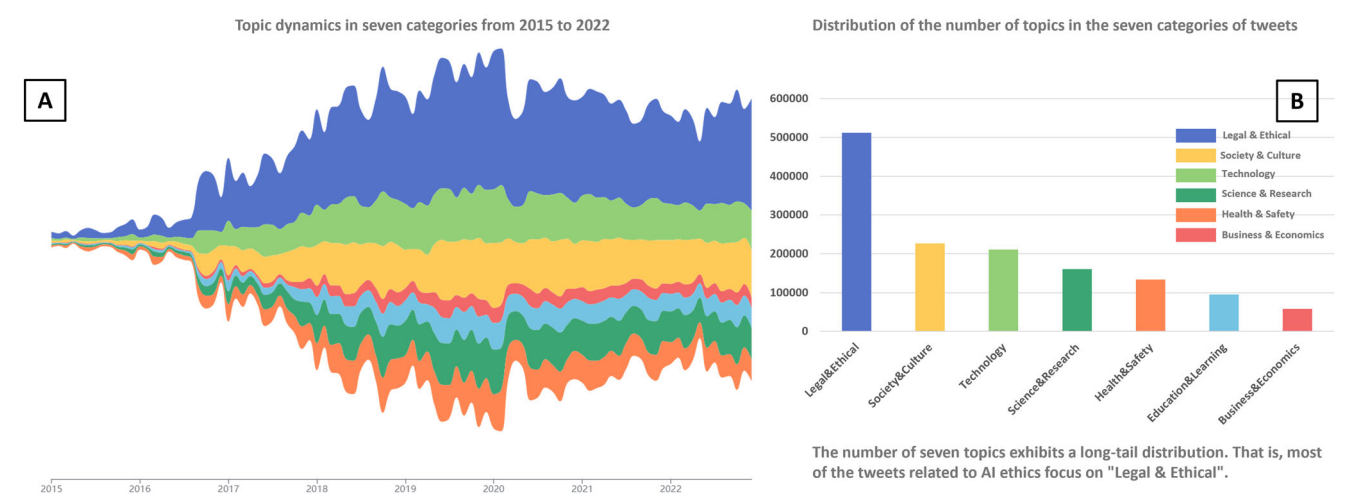
Overview of hierarchical topic structure. To address our first research question (RQ1), about how AI ethics discourse is framed within Twitter discourse overall, we extracted all relevant tweets and clustered them into a hierarchical topic structure, as shown in Fig. 2. This structure consists of three layers. The first layer comprises seven main topics: Legal & Ethical, Society & Culture, Technology, Science & Research, Health & Safety, Education & Learning, and Business & Economics. Figure 2 presents a clear and intuitive visualization of the hierarchical structure of AI ethics discourse through the visualization of the sunburst chart. This shows the categories of each topic and the containment relationships between the hierarchies. Furthermore, the underlying 64 fine-grained topics (lowest level) have not been overlooked, covering mainstream public discourse and issues of small

but critical scope, such as *Intellectual Property in AI* and *Gender Discrimination*.

Table 1 is a codebook for the topic of AI ethics discourse. It describes each category of topics in the top-level structure and provides examples of topics discussed within each category. An in-depth analysis of the topics of the hierarchical structure can be found in the Discussion section.

In addition to the fine-grained hierarchical structure analysis, the temporal trend of topic discussions over time also conveys rich meaning. we chose stream chart and bar chart to illustrate the changes of the seven main topics related to AI ethics from 1 January 2015 to 31 December 2022 as Fig. 3A shows. The overall volume of discussions on AI ethics was relatively low during 2015–2016, gradually increasing from 2017 and peaking in early 2020. There was a slight decline in 2020, possibly influenced by the pandemic outbreak, which may have diverted substantial discourse resources. After 2020, the discourse on AI ethics remained relatively stable. Figure 3B shows that among the seven topics, a significant portion of the discussions revolve around "Legal & Ethical," surpassing 90% of the total tweet volume. The remaining six topics have relatively similar amount of discussion, with "Business & Economics" being the least discussed. This skewed distribution, where a few categories (also called heads) contain a large number of samples, while most categories (also called tails) have very few samples, conforms to a long-tail distribution (Anderson, 2012). The long-tail distribution of topics related to AI ethics reveals that although public concern in this field is focused on Legal & Ethical discussions, niche topics such as Education & Learning and Business & Economics are also significant and should not be overlooked (Agarwal et al., 2012; Mustafaraj et al., 2011).

Table 1 The codebook of AI ethics discourse on the top level.

| Top Level	Percentage	Definition	Detail Example
Legal & Ethical	36.7%	This section refers to the part of the AI ethics discourse that relates to legal and ethical guidelines. This section focuses on developing and applying AI technologies in compliance with laws, regulations, and ethical guidelines.	Intellectual Property in AI, AI Legal Issue, Leagal Framework
Society & Culture	16.2%	This section refers to the part of the AI ethics discourse that is relevant to social and cultural studies. It is concerned with the impact of AI on social structures, cultural values and human interactions.	Gender Discrimination, Public Good Society, Ethnic Studies.
Technology	15.1%	This section relates to AI technologies that have an ethical component. It focuses on whether the design, development and use of AI systems meet ethical standards to ensure safety, transparency, and interpretability.	Neurocontroller, VR/Augmented Reality, Cryptocurrency
Science & Research	11.5%	This section deals with the part of the AI ethics discourse that relates to science and research. The section focuses on the categories of science and research areas involved in AI ethics.	Interdisciplinary Studies, Empirical Research, Bioinformatic
Health & Safety	9.6%	This section addresses the health and safety content addressed in AI ethical discourse. It focuses on using AI in healthcare, public health, and safety.	Work Safety, Suicide Prevention, Disease Prevention
Education & Learning	6.8%	This section is the category of education and learning in the AI ethics discourse. It focuses on AI technologies' new changes and impacts on education and learning.	E-learning Teleformacion, Collaborative Learning Community, Educational Entrepreneurship
Business & Economics	4.1%	This section is the business and economic part of the AI ethics discourse. It looks at the use of AI in business applications and the impact on employee rights and the market economy as it creates value and improves efficiency.	Fintech, Human Resource Management, Market Research

Topic dynamics in seven categories from 2015 to 2022

Distribution of the number of topics in the seven categories of tweets

The number of seven topics exhibits a long-tail distribution. That is, most of the tweets related to AI ethics focus on "Legal & Ethical".

Fig. 3 Temporal Trend and Distribution of AI Ethics Topics on Twitter. A Trend of the number of the seven topics over time; **B** graph of the distribution of the seven topics, with their numbers ranked from highest to lowest.

Visual analytics of AI ethics discourse in Twitter. To answer RQ2, this study presents coherent and readable visual analytics from two parts, including story map and topic evolution diagram.

Story Map: The World and The United States as Examples. We constructed a global story map and a more detailed story map of the United States, which has the highest tweet volume, as an example. Among them, the global story map combines mainstream AI ethics discourse with geospatial information, presenting the distribution of seven topics worldwide. The American Story Map integrates mainstream AI ethics discourse, sentiment information, and spatial locations, built upon the changes in the number of tweets related to AI ethics in the United States from 2015 to 2022.

Figure 4 illustrates the distribution of AI ethics-related tweets worldwide. Most AI ethics discourse is concentrated in the United States and Europe. Some countries, such as China and Cuba, have limited use of Twitter, so their distribution data may not provide accurate references. The surrounding maps in Fig. 4 display the distribution of the seven topics worldwide, they are generally similar but with some subtle differences. For instance, discussions on "Health & Safety" in African countries are more prevalent compared to "Technology" and "Business & Economics," while India has more discussions on "Technology" and "Education & Learning" compared to other topics.

To analyze the information conveyed by the story map in Twitter data further, we present an example using the United States. Figure 5A displays the distribution of AI ethics-related tweets in the United States. We observe that the discussion of AI
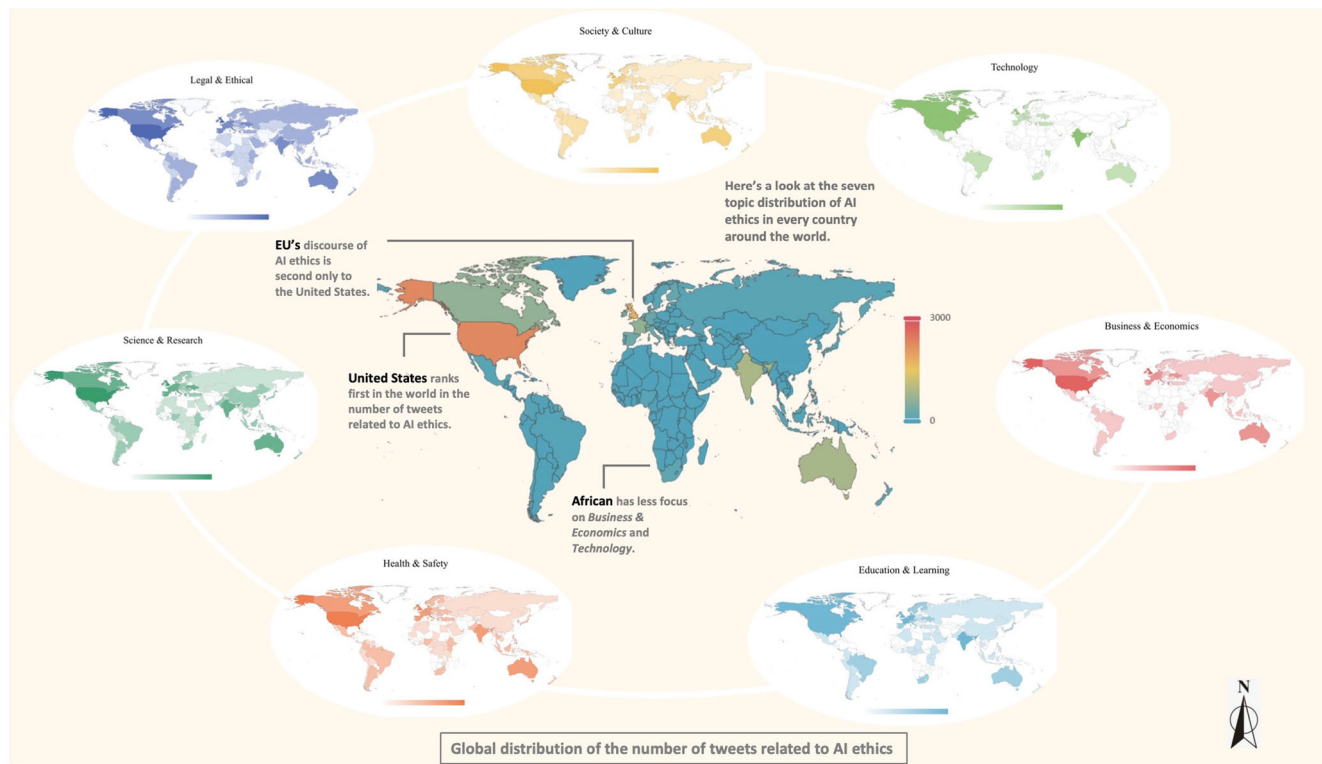
Fig. 4 Global Distribution of AI Ethics Twitter Data and Distribution of Seven Topics Worldwide.

ethics topics is most concentrated in California, New York, and Massachusetts. This concentration might be attributed to the large population size, numerous high-tech companies, and the abundance of universities in these three states. Considering the "long-tail distribution" characteristic of the seven topics related to AI ethics, although the tail-end topics constitute a relatively small proportion, they still reveal significant information. After excluding the most discussed topic "Legal & Ethical," we illustrate the distribution of the remaining topics in Fig. 5B. More than 50% of states focus on "Science & Culture," with "Technology" as the next prominent topic. Interestingly, New Mexico is most interested in the "Health & Safety" topic. This may be due to the diverse social and cultural backgrounds of the different federal states. For example, California and Washington State are home to numerous large tech companies. Still, California's industries include globally renowned tourism and film industries, while Washington State is known for aerospace and agriculture, resulting in differing AI ethics discourse between the two states.

Figure 5C, D present the distribution of positive and negative sentiments related to AI ethics discourse in various states. Interestingly, the top five states with the strongest positive sentiment are the same as the top five states with the strongest negative sentiment: California, New York, Massachusetts, Washington, and Texas. This result reflects the consistent intensity of public sentiment; regions expressing positive sentiments do not necessarily have reduced negative sentiments. Furthermore, we explored the content discussed behind these positive and negative sentiments and displayed them using word clouds. Figures 5E, F show the word clouds corresponding to positive sentiment and negative sentiment respectively. This indicates that people discuss similar topics with different sentiments, focusing on data, humans, and artificial intelligence. However, those expressing positive sentiments are more likely to see the positive impacts of data and AI on humanity, while those expressing negative sentiments demonstrate more ethical concerns and are also more concerned about the potential problems

with data. Figure 5G provides statistical information on the trend of AI ethics discourse in the United States over time, serving as background information for the story map.

Topic evolution view. We integrated fragmented AI ethics discourse information into readable and coherent narratives. Figure 6A depicts the evolution of AI ethics discourse worldwide over time, with some critical events related to AI as background information. Figure 6B, C illustrate the evolution of AI ethics discourse from 2015 to 2022. The horizontal axis represents the timeline, while the vertical axis indicates the magnitude of each topic. Since this topic evolution diagram aims to display the evolution of topics, time and quantity serve as reference information for the distribution of topic bubbles. We first extracted the main topics from the discourse in January 2015, identifying "Internet" as the most discussed one in that month. Then, by calculating semantic similarity, we selected five topics semantically most related to "Internet" in 2015, such as AI-tech, Crypto, etc. The topic with the latest timestamp in the previous year then evolved into the following year's five topics, and so on.

Figure 6 takes 'Internet', the most frequent keyword in AI ethics discourse in January 2015, as an example of how the story metaphor framework can be used to build fragmented information into easy-to-understand stories in cross-level fine-grained social media data mining. For other topics, we can also select an event that happened at a particular time and explore the topic discussions or other events triggered by the event. This paper starts with the evolution of the 'Internet'. Through the consolidation of fragmented information related to AI ethics from 2015 to 2022, combined with background information, we present a topic evolution view. In 2015, the formal launch of Ethereum and incidents such as hacking of Bitcoin exchanges sparked discussions on "Crypto" and "Cryptocurrency" related to AI technology. In 2016, following the theft of Ether from the DAO, an Ethereum-based intelligent contract organization, attention towards financial technology continued to rise.
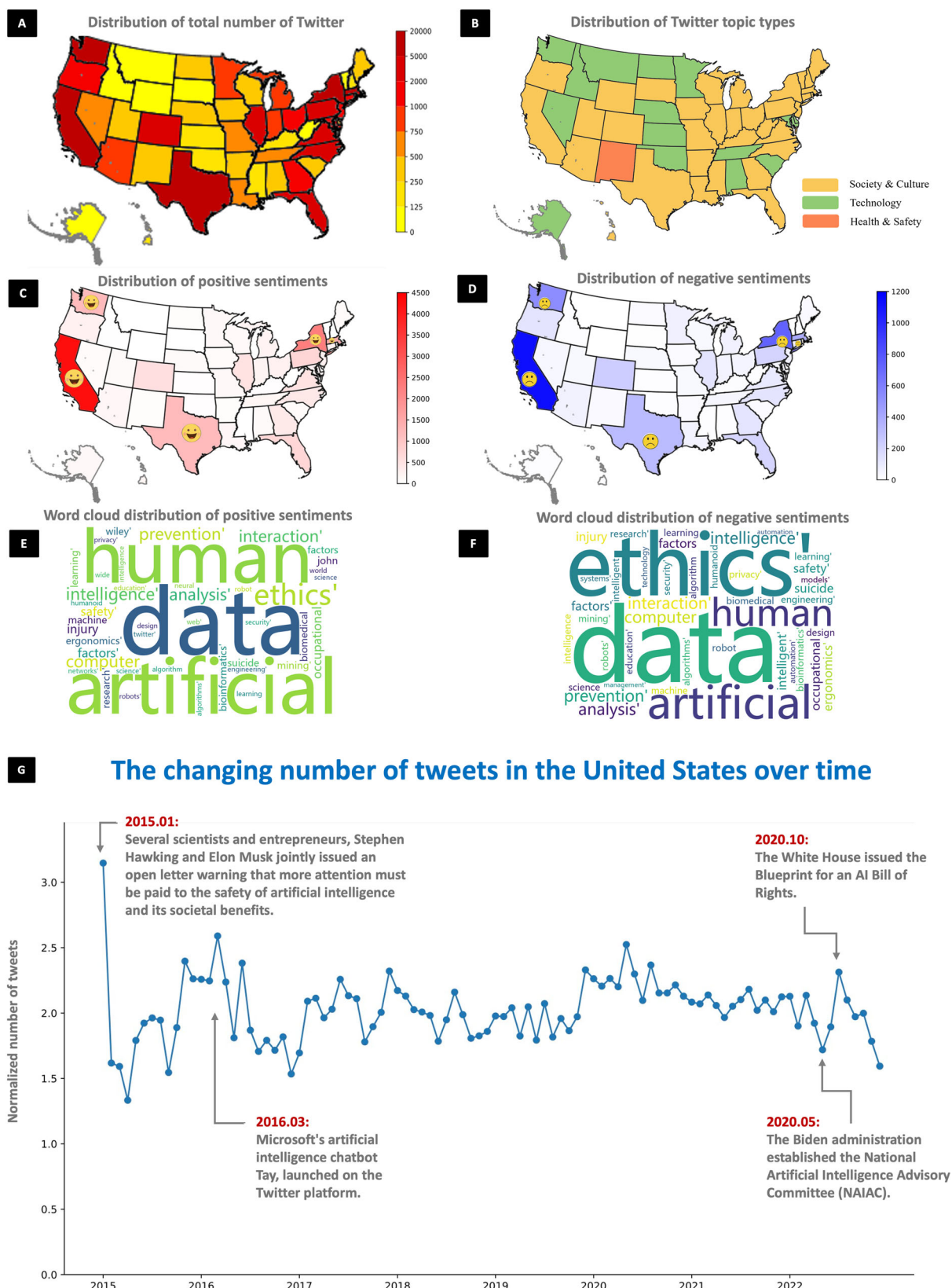
Fig. 5 Story maps of AI ethics discourse in the United States. A The overall distribution of AI ethics discourse in the United States. **B** The distribution of the seven topics of AI ethics discourse across US states. **C** The distribution of positive sentiments in AI ethics discourse across US states. **D** The distribution of negative sentiments in AI ethics discourse across US states. **E** The word cloud distribution of positive sentiments in AI ethics discourse. **F** The word cloud distribution of negative sentiments in AI ethics discourse. **G** The normalized curve showing the number of AI ethics-related tweets over time in the USA, with significant AI-related events annotated for context.
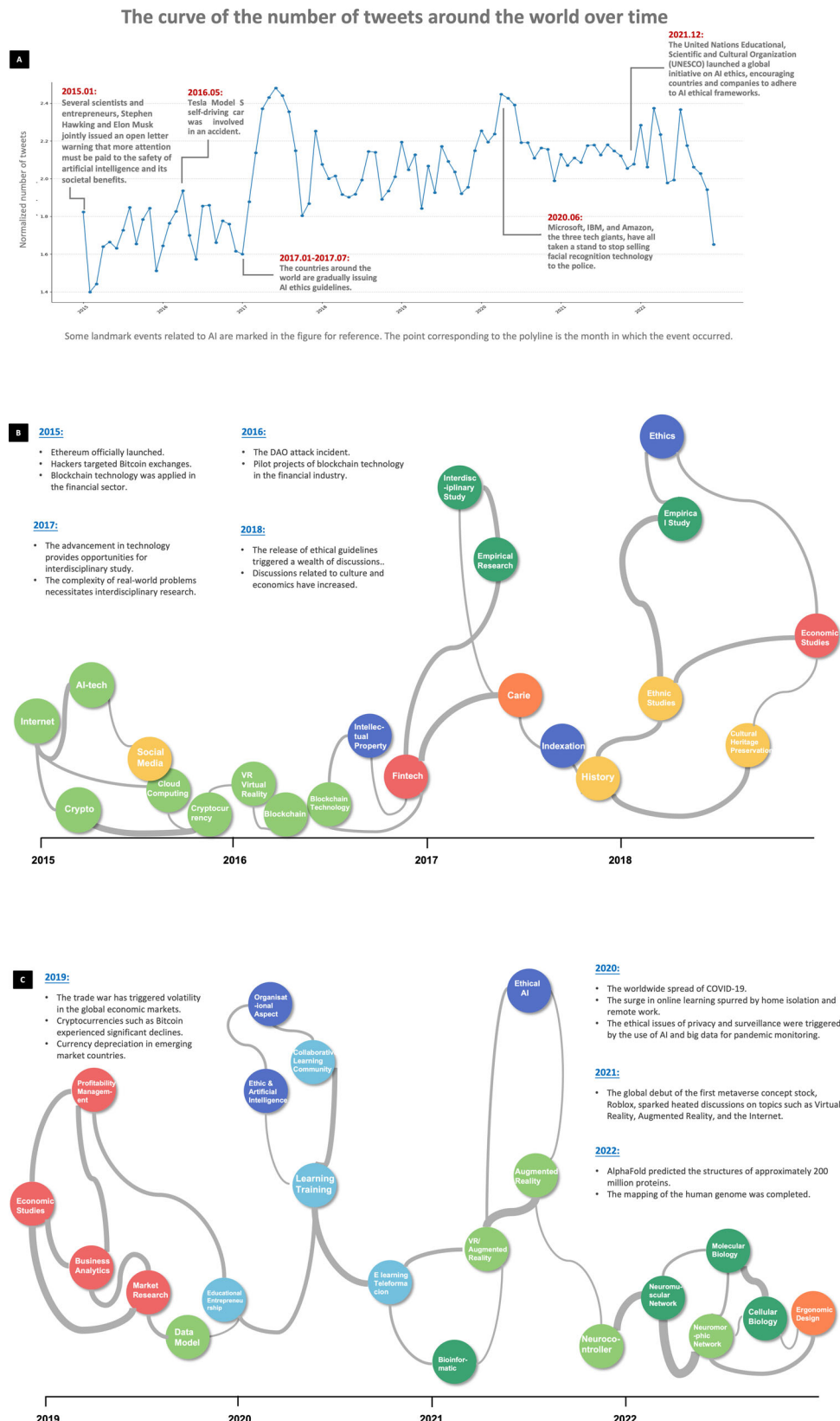
Fig. 6 Topic evolution view of AI ethics discourse on Twitter. A The normalized curve showing the number of AI ethics-related tweets over time in the world, with significant AI-related events annotated for context. **B**, **C** Topic evolution graphs on Twitter from 2015 to 2022, using the topic "Internet" as an example. **B** shows period from 2015 to 2018, while **C** shows the period from 2019 to 2022. The same color represents topics belonging to the same category within the seven topics of AI ethics discourse. The *x*-axis in the figure represents the timeline, indicating the appearance of topics over time, while the *y*-axis represents the number of topics. The higher the bubble representing a topic is distributed on the *Y*-axis in the graph, the larger its relative quantity. Note that this graph is a schematic, and the time and quantity do not represent precise values.

Blockchain technology gradually began to be used in the financial sector to ensure security and compliance. In 2017, the continued development of AI technology drew multidisciplinary attention. Single-discipline advancements were no longer sufficient to address complex real-world issues, leading to an emphasis on interdisciplinary research. In 2018, public discussions on ethics reached unprecedented levels. In May of the same year, the European Union formally implemented the General Data Protection Regulation (GDPR), setting a benchmark for AI ethics regulations. In addition to "Ethics," the public showed increased interest in cultural and economic-related topics. In 2019, the trade war triggered fluctuations in the global economic market. Bitcoin plummeted, leading to widespread closures of mining operations. The strengthening of the US dollar and continuous interest rate hikes by the Federal Reserve caused emerging market currencies to collapse. Topics such as "Economic Studies," "Business Analytics," and "Market Research" became focal points of discussion. In 2020, the outbreak of COVID-19 consumed a significant portion of social media resources, resulting in fewer discussions related to AI ethics. As the pandemic spread, many countries implemented surveillance and tracking measures to control its spread, leading to ethical debates on privacy and surveillance. The pandemic-induced home isolation and remote work shifted learning and work patterns, making topics like "Ethics & Artificial Intelligence," "Collaborative Learning Community," and "E-Learning Teleformacion" new focal points of discussion. In 2021, the global gaming platform Roblox became the first metaverse concept stock listed on the New York Stock Exchange, sparking discussions on "Augmented Reality" and "Virtual Reality." In April of the same year, the European Union proposed the Artificial Intelligence Act (EU AI Act), the world's first comprehensive legislative attempt to address the phenomenon and risks of artificial intelligence. The establishment of this AI regulatory framework led to an increase in discussions on "Ethical AI." In 2022, DeepMind successfully predicted the structures of approximately 200 million proteins from 1 million species using AlphaFold, covering almost all known proteins on Earth, ushering humanity into a new era of digital biology. In April, the international academic journal Science revealed the mystery of human genes, announcing the completion of the first complete map of the human genome. Topics such as "Neuromuscular Network," "Molecular Biology," "Cellular Biology," and "Ergonomic Design" became focal points of discussion.

Discussion

The implications of AI ethics discourse in Twitter

The absence and urgency of AI ethics laws and regulations. Analysis of AI ethics discourse in Twitter reveals that the topic of **Legal & Ethical** (36.7%) has the highest public attention among AI ethics-related discussions. Public discourse highlights concern about the lagging development of AI-related laws. In particular, the rapid advancement of AI technology has outpaced public understanding and acceptance, leading to discomfort regarding issues such as opacity in decision-making processes (black-box effect) and potential privacy risks. The high level of public concern about laws and ethical regulations indicates that trust mechanisms still need to be established, and the public hopes that explicit external constraints (such as laws and ethical guidelines) can ensure the safe and controllable development of AI technology. Through fine-grained data mining, it was found that the public pays significant attention to intellectual property rights within AI ethics laws (see Table 1). These findings suggest that as AI technologies develop, the widespread availability of creative tools allows more people to quickly produce content. However, the issue of ownership over AI-generated content has

complicated intellectual property protection (Abdikhakimov, 2023; Ballardini et al., 2019). For example, there is no consensus on the creator identity and copyright ownership of AI-generated texts, images, or music, and the boundaries of technology usage still need to be clarified (Epstein et al., 2023). The existing legal system is no longer sufficient to meet the demands of the current era and requires innovation and adjustment in legislation and regulation. Furthermore, AI laws and ethical regulations allow the public to understand and participate in technology governance (Atkinson et al., 2020; Surden, 2019). The high level of attention indicates that the public hopes to express their concerns through these mechanisms and influence the direction and limits of AI technology use. This also suggests that the public expects to play a more active role in technology governance rather than simply being passive recipients of technology applications.

Emphasizing the integration of technology and the humanities. **Society & Culture** (16.2%) ranks second among the AI ethics discourse topics in Twitter data, surpassing **Technology** (15.1%). This indicates that the public is equally concerned about the impact of AI on social values, cultural dissemination, and human interaction patterns. Through fine-grained information mining, we found that public discussions include Gender Discrimination, Public Good Society, and Ethnic Studies. These results suggest that the public may be concerned about issues such as cultural homogenization or social fragmentation caused by AI. This means academics and policymakers must prioritize discussions on these issues, studying AI's adaptability and inclusiveness in diverse societies. The integration of technology and humanities plays an indispensable role in the development of AI. The humanities provide in-depth social insights and ethical frameworks, helping technology developers and decision-makers avoid potential social risks and ensure that AI technologies positively impact all groups. Additionally, this highlights future research directions, such as addressing gender bias and racial discrimination through integrating technology and humanities, promoting fairness in AI technology, and improving technology transparency and accountability to foster a sense of social responsibility. Combining the **Science & Research** topic with fine-grained information exploration, we discovered the Interdisciplinary Studies topic, which shows that interdisciplinary collaboration is key to better advancing AI technology for the public good, achieving a dual enhancement of technological innovation and social value.

Promoting multilateral cooperation and public discussion. The complexity of AI ethics issues requires the participation of governments, businesses, academia, and the public. The research findings suggest the establishment of cross-sector collaboration platforms to jointly address the ethical challenges of AI technology through open dialogue, knowledge sharing and joint action. Public opinion expressed through social media provides real-time feedback on policy and technology development. This 'bottom-up' flow of information can help fill the blind spots in top-down designs, making AI ethics more responsive to practical needs. Such multilateral collaboration contributes to building a more comprehensive ethical framework, enhancing public trust in AI technology, and increasing societal acceptance of technological innovation.

Visual analytics in social media data mining. Our work demonstrates the importance of using visual analysis methods for AI ethics discourse data mining on Twitter and provides a reference for mining social media data on other topics.

Visual analytics helps to mine social media data in unanticipated ways. Social media data is vast and diverse, and extracting information directly from text may miss hidden patterns or subtle trends. Visual analytics helps to explore information from social media data where the results cannot be predicted in advance. Graphically presenting the data helps researchers eliminate the interference of irrelevant information and effectively extract the hidden key insights.

Visual analytics helps the public better understand the deeper layers of information about AI ethics. Using the topic evolution view in Fig. 6C as an example, this study demonstrates how to explore topics beneath the surface of vast social media data through a layered data structure and integrate the findings into a cohesive narrative. Taking the topic "Internet," which saw significant discussion in January 2015, as an example, we illustrate how to help the public understand the evolution of topics, explore connections between different themes, and dynamically achieve cross-level topic evolution, effectively linking and integrating fragmented data. Research in AI ethics relies on active public engagement. Visual analysis facilitates easier communication and knowledge sharing among experts from various fields and enables the public to understand and directly participate. Thus, visual analysis is an essential method in AI ethics research.

Transferability of visual analytics to other social media topics. This study applies visualization analysis methods to present results based on the characteristics of AI ethics discourse data on Twitter, focusing on spatial distribution and topic evolution. These methods are not limited to AI ethics topics but have high transferability to studies of other social media topics, such as political orientation analysis and pandemic spread analysis. Our visualization analysis approach provides a cross-level, fine-grained presentation of information, enabling researchers to deeply analyze topics, gain insights into public opinion, understand topic evolution trends, and accordingly develop or adjust policies.

Limitation
Although our results have effectively answered the two research questions, they still have some limitations with regard to data and methods. First, our approach of crawling data using keywords related to AI ethics may not fully capture all relevant tweets. For instance, content without AI ethics-related hashtags added by the tweet publishers would not be retrieved. Moreover, the public's use of social media platforms is unbalanced. For example, it may be more difficult for economically underdeveloped regions to use social media platforms, resulting in limited accessible data. Second, Public opinions expressed on social media still require vigilance regarding the issue of insufficient sample representativeness. Social media users are only partially representative of the general public, and may be predominantly composed of younger individuals, urban dwellers, or specific interest groups. Older adults, low-income populations, or users from remote areas may have lower levels of social media participation, leading to biases in the data sample. However, although younger users on Twitter are more active, curious, and expressive about emerging technology issues, its user base also has a significant proportion of older people (Sloan et al., 2013). Furthermore, Twitter users are diverse in age, occupation, and geographical distribution (Sloan et al., 2015). As such, Twitter is an effective platform for understanding public discourse on AI technology development (Anger and Kittl, 2011). Third, the representation of the AI ethics topic evolution view is limited. The evolution of AI ethics events should ideally include

information such as the time and volume of discussion for each topic. However, our results involve exploring topics at a granular level, leading to significant differences in discussion volume for each topic. Additionally, restrictions in visualization effects and methods make it difficult to accurately represent the temporal distribution of topics within a constrained space. What we have created is only an illustrative view of the evolution of AI ethics discourse by combining time and discussion volume for each topic.

Conclusion and outlook
This study fills a gap in the literature by investigating the cross-level, fine-grained exploration of AI ethics discourse on Twitter. As ethical issues arising from the development of AI technologies receive increasing attention, our research provides insights into public concerns about AI ethics and the evolution of related topics. By integrating neural networks with large-scale language models, this study extracts hierarchical thematic structures of AI ethics discourse on Twitter, enabling fine-grained analysis across levels. Furthermore, the study employs visualization techniques to address the challenge of fragmented fine-grained information in social media. The construction of narrative metaphors - topic evolution - helps the public to better understand the deeper aspects of the AI ethics discourse. The findings offer valuable insights for AI policymaking and demonstrate the feasibility of fine-grained information mining in social media data.

The exploration of AI ethics and social media data is just at the beginning. The essence of AI technology is to serve humanity, and the public ethical concerns and discussions sparked by AI are worthy of attention. For example, future research could combine news and Twitter data to analyze AI ethics issues. Investigating public discussions and feedback on Twitter regarding a specific AI ethics news case can help policymakers formulate more specific guidance to address and mitigate ethical concerns from the public perspective.

Data availability
Data will be made available on request.

References
Abdikhakimov I Unraveling the Copyright Conundrum: Exploring AI-Generated Content and its Implications for Intellectual Property Rights | International Conference on Legal Sciences (2023) Available at: https://www.science-zone.org/index.php/conference/article/view/42 (accessed 19 December 2024)

Agarwal P, Vaithiyanathan R, Sharma S et al. (2012) Catching the long-tail: extracting local news events from Twitter. Proc Int AAAI Conf Web Soc Media 6(1):379–382

Alsaedi N, Burnap P, Rana O (2016) Temporal TF-IDF: a high performance approach for event summarization in Twitter. In: *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, October 2016, pp. 515–521. Available at: https://ieeexplore.ieee.org/abstract/document/7817104?casa_token=LwmLtepqa0wAAAAA:wUiLe7UNFU8XIauBLTDjnPt6kaVILw6wtwD8VGjKSdvsyVZ8bhh7LTPVB4uSkxKP6UuLBR4g (accessed 7 April 2024)

Anderson C (2012) The Long Tail. In: 10. The Long Tail. New York University Press, pp 137–152. Available at: https://www.degruyter.com/document/doi/10.18574/nyu/9780814763025.003.0014/pdf?licenseType=restricted (accessed 30 April 2024)

Anderson J, Rainie L (2018) Artificial intelligence and the future of humans. In: Pew Research Center. Available at: https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/ (accessed 1 May 2024)

Anger I, Kittl C (2011) Measuring influence on Twitter. In: *Proceedings of the 11th International Conference on Knowledge Management and Knowledge*

Technologies, Association for Computing Machinery, New York, pp 1–4. Available at: https://dl.acm.org/doi/10.1145/2024288.2024326 (accessed 22 April 2024)

Atkinson K (2020) Trevor Bench-Capon and Danushka Bollegala Explanation in AI and law: Past, present and future - ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/S0004370220301375?casa_token=Fdyyn7zHv1YAAAAA:stKae04uxjmLjpyF8vMYZ539QypECA8RFsRFoTb-YvjO0v3wtnR2YesBpr1fl_4E_wVYKxxUhGo (accessed 19 December 2024)

Azam N, Jahiruddin, Abulaish, M et al. (2015) Twitter Data mining for events classification and analysis. In: *2015 Second International Conference on Soft Computing and Machine Intelligence (ISCMI)*, November 2015, pp. 79–83. Available at: https://ieeexplore.ieee.org/abstract/document/7414678?casa_token=4HARCwdd1sIAAAAA:h_gTIWDc4lg59QNuIT3Ww5j_vCRrkgSNZRa_yRemj3Res6aEpxEsR22C7SsVFXgSo3km4Khx (accessed 7 April 2024)

Ballardini RM, He K, Roos T (2019) AI-generated content: authorship and inventorship in the age of artificial intelligence. Available at: https://www.elgaronline.com/edcollchap/edcoll/9781788119894/9781788119894.00015.xml (accessed 19 December 2024)

Boon-Itt S, Skunkan Y (2020) Public Perception of the COVID-19 Pandemic on Twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveillance 6(4):e21978

Camacho-collados J, Rezaee K, Riahi T, et al. (2022) TweetNLP: Cutting-Edge natural language processing for social media. In: (W Che and E Shutova (eds) Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Abu Dhabi, UAE, pp 38–49. Available at: https://aclanthology.org/2022.emnlp-demos.5 (accessed 8 March 2024)

Chen K, Molder AL, Duan Z (2023) How climate movement actors and news media frame climate change and strike: evidence from analyzing Twitter and News Media Discourse from 2018 to 2021. Int J Press/Politics 28(2):384–413

Chen S (2018) Big Data in computational social science and humanities | SpringerLink. Available at: https://link.springer.com/book/10.1007/978-3-319-95465-3 (accessed 13 December 2024)

Conneau A, Khandelwal K, Goyal N, et al. (2020) Unsupervised cross-lingual representation learning at scale. arXiv:1911.02116. arXiv. Available at: http://arxiv.org/abs/1911.02116 (accessed 25 April 2024)

Corbett E, Denton E, Erete S (2023) Power and public participation in AI. In: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, New York, pp 1–13. EAAMO '23. Association for Computing Machinery. Available at: https://dl.acm.org/doi/10.1145/3617694.3623228 (accessed 1 May 2024)

Davis ZT, McPherson MD (1989) Story map instruction: a road map for reading comprehension. The Reading Teacher 43(3):232–240

Devlin J, Chang M-W, Lee K, et al. (2019) BERT: pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805. arXiv. Available at: http://arxiv.org/abs/1810.04805 (accessed 25 April 2024)

Dollbo A (2023) Deliberating AI: Why AI in the Public Sector Requires Citizen Participation. Master Thesis. Available at: https://studenttheses.uu.nl/handle/20.500.12932/44530 (accessed 1 May 2024)

Epstein Z, Hertzmann A (2023) the Investigators of Human Creativity, et al. (2023) Art and the science of generative AI. Science 380(6650):1110–1111

Feng J, Wu K, Chen S (2023) TopicBubbler: an interactive visual analytics system for cross-level fine-grained exploration of social media data. Visual Inform 7(4):41–56

Feng Y, Meng Z, Clemmer C, et al. (2023) Data and resources paper: a multi-granularity decade-long geo-tagged twitter dataset for spatial computing. Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems. 1–4

Freedman G, Reynolds EG (1980) Enriching basal reader lessons with semantic webbing. The Reading Teacher 33(6):677–684

Fu Y, Zhuang Z, Zhang L (2023) Ai ethics on blockchain: Topic analysis on twitter data for blockchain security. Science and Information Conference. Cham: Springer Nature Switzerland, 82–100

Gilman ME (2023) Democratizing AI: principles for meaningful public participation. 4628755, SSRN Scholarly Paper. Rochester, NY. Available at: https://papers.ssrn.com/abstract=4628755 (accessed 1 May 2024)

Haque MU, Dharmadasa I, Sworna ZT, et al. (2022) 'I think this is the most disruptive technology': exploring Sentiments of ChatGPT Early Adopters using Twitter Data. arXiv:2212.05856. arXiv. Available at: http://arxiv.org/abs/2212.05856 (accessed 15 March 2023)

Hua Y, Jiang H, Lin S et al. (2022) Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. J Am Med Inform Assoc 29(10):1668–1678

Huang D, Mu D (2014) Topic detection in Twitter Based on Label Propagation Model. In: 2014 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, pp 97–101 Available at: https://ieeexplore.ieee.org/abstract/document/6999065?casa_token=EMx1pkS1XUcAAAAA:sXz37qF5pa4sYDYWreOK1tgocDSfwrQGcJiWpjXjJ4AMGrG9KUk_N0zC5428WCMoSjM-4qHv (accessed 7 April 2024)

J van Dijck (2013) The culture of connectivity: a critical history of social media. Oxford University Press

Jost JT, Barberá P, Bonneau R et al. (2018) How social media facilitates political protest: information, motivation, and social networks. Polit Psychol 39(S1):85–118

Karami A, Lundy M, Webb F et al. (2020) Twitter and research: a systematic literature review through text mining. IEEE Access 8:67698–67717

Keim DA, Mansmann F, Schneidewind J, et al. (2008) Visual analytics: scope and challenges. In: Simoff SJ, Böhlen MH, and Mazeika A (eds) Visual Data Mining: Theory, Techniques and Tools for Visual Analytics. Berlin, Heidelberg: Springer, pp 76–90. Available at: https://doi.org/10.1007/978-3-540-71080-6_6 (accessed 19 December 2024)

Kontaxaki S, Tomai E, Kokla M, et al. (2010) Visualizing multidimensional data through granularity-dependent spatialization. In: Visualization and Data Analysis 2010, pp 232–243. SPIE. Available at: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7530/75300M/Visualizing-multidimensional-data-through-granularity-dependent-spatialization/10.1117/12.838430.full (accessed 13 December 2024)

Kulkarni M, Mahata D, Arora R, et al. (2022) Learning rich representation of keyphrases from Text. arXiv:2112.08547. arXiv. Available at: http://arxiv.org/abs/2112.08547 (accessed 12 February 2024)

Kwon H, Koylu C, Dietrich B (2023) Natural language processing meets spatial time series analysis and geovisualization: identifying and visualizing spatio-topical sentiment trends on Twitter. Cartogr Geogr Inform Sci 50:1–15

Lee K, Palsetia D, Narayanan R, et al. (2011) Twitter Trending Topic Classification. In: 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, p. 251–258. IEEE. Available at: http://ieeexplore.ieee.org/document/6137387/ (accessed 5 April 2024)

Liang H, Shen F, Fu K (2017) Privacy protection and self-disclosure across societies: a study of global Twitter users. New Media Soc 19(9):1476–1497

Liu Y, Ott M, Goyal N, et al. (2019) RoBERTa: A robustly Optimized BERT Pretraining Approach. arXiv:1907.11692. arXiv. Available at: http://arxiv.org/abs/1907.11692 (accessed 25 April 2024)

Manovich L (2012) Trending: the promises and the challenges of big social data. In: Gold MK (ed.) Debates in the Digital Humanities. University of Minnesota Press, pp 460–475. Available at: https://academic.oup.com/minnesota-scholarship-online/book/29340/chapter/243999993 (accessed 1 May 2024)

Mustafaraj E, Finn S, Whitlock C, et al. (2011) Vocal minority versus silent majority: discovering the opinions of the long tail. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, pp 103–110. Available at: https://ieeexplore.ieee.org/abstract/document/6113101 (accessed 30 April 2024)

Nguyen DQ, Vu T, and Nguyen AT (2020) BERTweet: a pre-trained language model for English Tweets. arXiv:2005.10200. arXiv. Available at: http://arxiv.org/abs/2005.10200 (accessed 25 April 2024)

Reimers N, Gurevych I (2019) Sentence-BERT: sentence embeddings using Siamese BERT-Networks. arXiv:1908.10084. arXiv. Available at: http://arxiv.org/abs/1908.10084 (accessed 12 February 2024)

Roth RE (2021) Cartographic design as visual storytelling: synthesis and review of Map-Based Narratives, Genres, and Tropes. Cartogr J58(1):83–114

Schubert E, Sander J, Ester M et al. (2017) DBSCAN revisited, revisited: why and how you should (Still) Use DBSCAN. ACM Transac Database Syst 42(3):21

Sloan L, Morgan J, Burnap P, et al. (2015) Who Tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter User Meta-Data. PLOS ONE 10(3):e0115545

Sloan L, Morgan J, Housley W, et al. (2013) Knowing the Tweeters: deriving sociologically relevant demographics from Twitter. Sociol Res Online 18(3):74–84

Surden H (2019) 'Artificial intelligence and law: an overview' by. Available at: https://scholar.law.colorado.edu/faculty-articles/1234/ (accessed 19 December 2024)

Taecharungroj V (2023) "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. Big Data and Cognitive Computing 7(1):35

Touvron H, Martin L, Stone K, et al. (2023) Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288. arXiv. Available at: http://arxiv.org/abs/2307.09288 (accessed 1 May 2024)

Yang J, Jin H (2020) Application of big data analysis and visualization technology in news communication. Comput Aided Design Appl 17(S2):134–144

Zuo C, Ding L, Yang Z, et al. (2022) Multiscale geovisual analysis of knowledge innovation patterns using big scholarly data. Annals of GIS 28(2):197–212

Author contributions

Mengyi Wei: Conceptualization, Project administration, Methodology, Formal analysis, Writing– original draft, Writing- review & editing. Puzhen Zhang: Data processing, Methodology. Chuan Chen: Data processing, Writing- review & editing. Dongsheng Chen: Methodology, Writing- review & editing. Chenyu Zuo: Writing- review & editing. Liqiu Meng: Formal analysis, supervision, Writing- review & editing.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as the study did not involve human participants.

Informed consent

Informed consent was not required as the study did not involve human participants.

Additional information

Correspondence and requests for materials should be addressed to Chuan Chen.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.