Unveiling public perception of AI ethics: an exploration on Wikipedia data

Mengyi Wei, Yu Feng, Chuan Chen, Peng Luo, Chenyu Zuo, Liqiu Meng

Angaben zur Veröffentlichung / Publication details:

Wei, Mengyi, Yu Feng, Chuan Chen, Peng Luo, Chenyu Zuo, and Liqiu Meng. 2024. "Unveiling public perception of AI ethics: an exploration on Wikipedia data." EPJ Data Science 13 (1): 26. https://doi.org/10.1140/epjds/s13688-024-00462-5.

Wei et al. *EPJ Data Science* (2024) 13:26
https://doi.org/10.1140/epjds/s13688-024-00462-5

EPJ Data Science
a SpringerOpen Journal

RESEARCH Open Access

Unveiling public perception of AI ethics: an exploration on Wikipedia data

Mengyi Wei[1], Yu Feng[1], Chuan Chen[1], Peng Luo[1], Chenyu Zuo[2*] and Liqiu Meng[1]

*Correspondence:
chenyu.zuo@csfm.ethz.ch
[2]Center for Sustainable Future Mobility (CSFM), ETH Zurich, Universitätstrasse 41, UNO D 12, Zürich, 8092, Switzerland
Full list of author information is available at the end of the article

Abstract

Artificial Intelligence (AI) technologies have exposed more and more ethical issues while providing services to people. It is challenging for people to realize the occurrence of AI ethical issues in most cases. The lower the public awareness, the more difficult it is to address AI ethical issues. Many previous studies have explored public reactions and opinions on AI ethical issues through questionnaires and social media platforms like Twitter. However, these approaches primarily focus on categorizing popular topics and sentiments, overlooking the public's potential lack of knowledge underlying these issues. Few studies revealed the holistic knowledge structure of AI ethical topics and the relations among the subtopics. As the world's largest online encyclopedia, Wikipedia encourages people to jointly contribute and share their knowledge by adding new topics and following a well-accepted hierarchical structure. Through public viewing and editing, Wikipedia serves as a proxy for knowledge transmission. This study aims to analyze how the public comprehend the body of knowledge of AI ethics. We adopted the community detection approach to identify the hierarchical community of the AI ethical topics, and further extracted the AI ethics-related entities, which are proper nouns, organizations, and persons. The findings reveal that the primary topics at the top-level community, most pertinent to AI ethics, predominantly revolve around knowledge-based and ethical issues. Examples include transitions from Information Theory to Internet Copyright Infringement. In summary, this study contributes to three points, (1) to present the holistic knowledge structure of AI ethics, (2) to evaluate and improve the existing body of knowledge of AI ethics, (3) to enhance public perception of AI ethics to mitigate the risks associated with AI technologies.

Keywords: AI ethics; Public perception; Wikipedia data; Community detection; Knowledge structure; Visualization

1 Introduction

AI ethics is increasingly valued in the development of artificial intelligence (Siau and Wang [37]; Tomašev et al. [42]; Stahl [40]). However, it is difficult to be aware of the potential risks in most situations because of the complexity of AI technologies. For example, the wide use of online social media has created a dependence on AI technologies. Many users have unconsciously experienced privacy violations that can result in reputational and financial damage (Srivastava and Geethakumari [39]). Companies progressively rely on algorithms

for decision-making to screen job applicants, but have to face risks of hidden discrimination that deprives job seekers of equal employment opportunities (Köchling and Wehner [21]). Social psychology research shows that public perception not only influence people's understanding of the world, but also shape their actual behavior (Ferguson and Bargh [9]). Individuals with a high AI awareness are more likely to recognize its usefulness than those without (Araujo et al. [1]). Conversely, a low level of AI perception amongst the public can lead to skepticism or negativity, potentially impeding the development of AI systems on the one hand, and undermining the need to defend human fairness on the other. Furthermore, the less the public knows about AI ethics, the more challenging it becomes to address ethical issues. According to a Global AI Survey 2023 at the University of Queensland, nearly half (49%) of individuals possess limited knowledge about AI (Gillespie et al. [13]). This underscores the importance of addressing AI ethical issues, where the focus should not just be on the measures or guidelines, but on the need to improve public perception of AI ethics to prevent latent risks.

Most traditional studies on public perception rely on survey or social media like Twitter. For instance, Kelley et al. [19] utilized statistical analysis to categorize the public sentiment about AI by questionnaire (Kelley et al. [19]), while Fu et al. [12] identified the prevalent topics of AI ethics discussed by the public through word frequency counting (Fu et al. [12]). These studies only reflect a part of public perception, which is more like a simple statistic of topics and does not reflect the knowledge involved behind the topics.

Knowledge structures are how individuals or groups understand, categorize, relate, and express information. Revealing the knowledge structure of AI ethics can enhance public perception and help them better understand AI ethical issues (Moy et al. [26]). The knowledge structure in Wikipedia can be represented by a community structure (Lizorkin et al. [22]). The community structure is usually constructed based on common characteristics of complex networks, such as social networks and transportation networks. The network graphs usually contain densely connected groups of nodes called communities, while they in turn are sparsely connected. Community detection is an algorithm designed to extract useful information hidden in the structure of myriad nodes and links (Bohlin et al. [2]), which makes it possible to identify existing communities.

In this study, we employed a community detection method to analyze the hierarchical knowledge structure related to AI ethics in Wikipedia, illustrating the relationships between these communities through visualization. Our findings, distinct from traditional word frequency counting methods for popular topics, contribute to presenting a body of knowledge of AI ethics, including the focus topics, related people, and research institutions. First, our results revealed a hierarchical structure of public perceptions of AI ethics, highlighting critical topics like Corporate Social Responsibility and Information Theory, enabling an evaluation of Wikipedia's current knowledge structure on AI ethics. Last but not least, our approach enhances public awareness of AI ethics by providing a clear overview and links to in-depth definitions and discussions, addressing the public knowledge gap, and mitigating potential ethical issues.

2 Related work

AI ethical issues have drawn much attention from different sectors of society. Governments, companies, and research institutions have proposed a range of measures to regulate AI technologies in ways that better serve humans. These studies and measures have

concentrated on rule-making and public perception research. We summarize the state-of-the-art research in these two parts and investigate the status of knowledge mining from Wikipedia.

2.1 AI ethics and its public perception

AI ethics is a field that has emerged as a response to growing concerns about the impact of AI. It is centered on ethical considerations and values throughout the entire lifecycle of AI systems, encompassing design, development, deployment, and utilization. The overarching objective of this field is to guarantee that the evolution and implementation of AI technologies adhere to ethical standards at both societal and individual levels, aiming to optimize societal well-being. On the one hand, much research takes efforts to propose AI guidelines to reduce ethical risks (Jobin et al. [18]; Microsoft [25]), but these guidelines are often too vague and theoretical to be translated into concrete action programs. On the other hand, gaining an in-depth understanding of the public perception is equally important.

The term "public perception" is challenging to define precisely. Dowler et al. [7] propose that public perception is essentially derived from public opinion surveys, which involve questioning a group of individuals about their thoughts on a specific issue or event (Dowler et al. [7]). Given that the essence of AI technology is to serve humans, understanding the public opinions of AI ethics can help to regulate AI technology. Fast and Horvitz [8] examined 30 years of New York Times coverage to study public perceptions of AI. They noted a significant rise in AI discussions since 2009, generally characterized by optimism. However, ethical concerns peaked around 2017, amid a simultaneous increase in optimism about AI's potential in healthcare and education (Fast and Horvitz [8]). In a global survey by Kelley et al. [19], public sentiment toward AI was characterized by four themes, including exciting, useful, worrying and futuristic (Kelley et al. [19]). Additionally, Ikkatai et al. [17] investigated public views on eight AI guideline themes, such as privacy, accountability, safety, transparency, fairness, human control, professional responsibility, and promotion of human values. Results showed that public opinions on AI usage differed across scenarios (including "Singer", "Service", "Weapon" and "Crime") in an online survey (Ikkatai et al. [17]).

Overall, previous research data on public perceptions of AI ethics is limited to surveys or social media data, such as Twitter. These data can only reflect part of public perceptions, which is more like simple statistics on the topics, and cannot uncover the knowledge behind the topics. Our paper opts for Wikipedia data to delve into the public perception of AI ethics. This approach not only uncovers the topics of public concern but also elucidates the knowledge gaps, contributing to enhancing public awareness of AI ethics.

2.2 Knowledge mining with Wikipedia

Wikipedia, most widely accessed online encyclopedia, has a significant user base and offers a platform for users to freely express their opinions (Fichman and Hara [10]; Vieira Bernat [44]). The diverse range of Wikipedia users spans all ages, ethnic, and socioeconomic groups, reducing bias to some extent (Zickuhr and Rainie [46]). Central to Wikipedia's mission is the pursuing a "neutral point of view" (Majchrzak [23]), wherein multiple viewpoints are presented alongside one another, ensuring the representation of all opinions. For example, a survey examining the tendentiousness in U.S. political articles on

Wikipedia reveals that although many articles possess inherent bias, Wikipedia's extensive history of accommodating diverse perspectives has resulted in a gradual decrease in overall bias and a shift towards neutrality on numerous topics (Greenstein and Zhu [14]). With its collaborative editing approach and extensive user base, Wikipedia transcends the limitations of individual viewpoints and strives to present topics neutrally.

As a valuable knowledge base, Wikipedia provides good research stage for knowledge mining. Nastase and Strube [27] presented an approach to extract knowledge from Wikipedia categories and the category network, with results supporting the notion that Wikipedia category names are a rich source of valuable and reliable knowledge (Nastase and Strube [27]). Furthermore, Nguyen et al. [31] structured knowledge by analyzing the entity relationships in Wikipedia, and eventually proved the method's effectiveness by manually annotating the data (Nguyen [31]). In addition, the vast knowledge base makes Wikipedia a useful resource for question-and-answer tasks (Buscaldi and Rosso [3]) with an essential role in the transmission and influence of knowledge (Medelyan et al. [24]; Di Lauro and Johinke [6]). As a universal tool for information search, Wikipedia has a search pattern that reflects public interest and opinion on a topic, allowing it to influence public perception broadly (Smith and Gustafson [38]). Halatchliyski et al. [16] found that experts tend to contribute to core critical articles in their area of expertise (Halatchliyski et al. [16]), which has a significant impact on non-expert access to knowledge. In addition, Das et al. [5] found a strong tendency of Wikipedia to overly "push" the spread of ideas. They attempted to introduce new behavioral indicators and controversy scores to create a favorable knowledge transfer environment (Das et al. [5]). Given these distinctive attributes, Wikipedia emerges as a primary resource for examining the ethical structure of AI knowledge.

3 Methodology

This study employs the community detection based on information theory and complex graph theory to uncover the knowledge structure in Wikipedia data related to AI ethics (van Steen [43]). The whole research process involves four steps: (1) Acquire Wikipedia data related to AI ethics through web crawling. (2) Obtain the hierarchical communities of AI ethics knowledge based on the Infomap method. (3) Evaluate the reliability of the community detection algorithm. (4) Visualize the network structure for subsequent in-depth analysis. The flow chart of the study is shown in Fig. 1.

3.1 Data acquisition

This study utilized the "ethics of artificial intelligence" as the seed source and took the Infomap method to crawl the "see also" section of each Wikipedia entry. The "see also" section contains terms that are associated with the field of AI ethics, as perceived by the Wikipedia editors. The final dataset consisted of 983 nodes and 1399 links.

3.2 Community detection for AI ethics knowledge from Wikipedia data

3.2.1 Community detection

Wikipedia data presents complex network structures (Lizorkin et al. [22]), so we utilized community detection to uncover multilevel structures and their relationships in networks. In network science (Newman [28]), categorization of nodes into modules with community detection algorithms is mostly limited to two levels (Fortunato [11]), but the real-world

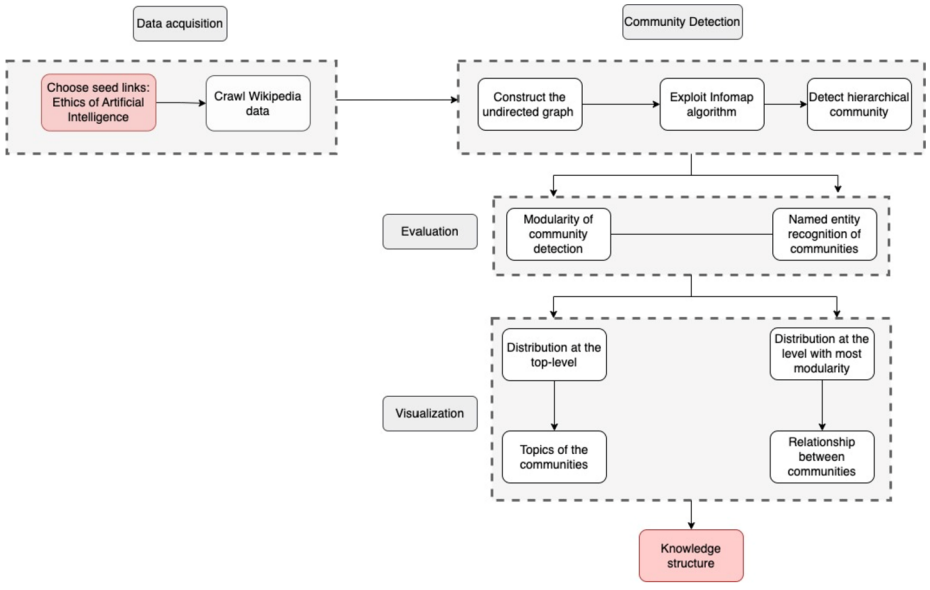

Figure 1 The workflow of detecting Wikipedia knowledge communities concerned with AI ethics

networks have much richer structures (Palla et al. [32]). Among all the hierarchical clustering algorithms (Newman [29]), the Infomap algorithm is regarded as one of the best performing non-overlapping clustering methods. In order to reveal the multiple levels of interdependences in the nodes of Wikipedia network, we choose the Infomap community detection method (Rosvall and Bergstrom [34, 35]; Bohlin et al. [2]).

Based on the complex graph theory and information theory, Infomap leverages random walks as a proxy for real flows, and capitalizes on the duality between message compression and pattern recognition in the underlying structure (Grünwald et al. [15]). Infomap method is described by giving a directed/undirected network, and then finding the optimal "groupings", that is, constructing the minimum information entropy. Huffman coding is applied in the Informap algorithm, which saves space by assigning short codes to common events or objects and long codes to rare ones, similar to how words that occur frequently in spoken languages are short (Zipf [47]). The minimum information entropy corresponds to the shortest average encoding length. As such, the primary objective of Infomap is to identify the optimal coding scheme, which minimizes the total average coding length. This method looks for a module partition $\mathbf{M}$ of n nodes into m modules to minimize the expected description length of a random walk. The average description length of a single step is given by (Rosvall and Bergstrom [34]):

$$L(M) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} p^i_{\circlearrowright} H\left(\mathcal{P}^i\right) \tag{1}$$

The equation includes two terms: the first represents the entropy of movement between modules, and the second the entropy of movement within modules. Exiting the module is also treated as a movement. $q_{\curvearrowright}$ denotes the probability that the random walk transits between modules during a given step. $H(\mathcal{Q})$ is the entropy associated with the module names. $H(\mathcal{P}^i)$ corresponds to the entropy of within-module movements, and the $p^i_{\circlearrowright}$ is

equal to the proportion of within-module movements for module i, including the probability of exiting module i so that $\sum_{i=1}^{m} p_{\circlearrowleft}^{i} = 1 + q_{\curvearrowright}$.

To achieve a balance between effect and speed, the formula (1) can be calculated by a fast stochastic and recursive search algorithm producing a two-level partition (Rosvall et al. [33]). On this foundation, the hierarchical Infomap method removes the constraint of a single index codebook and enables multiple nested index codebooks to define transitions between modules, submodules, and more granular levels of modules. Therefore, we can reveal rich multilevel organization and relationship with such a hierarchical Infomap. Each level consists of different communities, with the top level representing the whole network and the lowest level representing the smallest communities that cannot be subdivided further. Communities at different levels are arranged in a hierarchical relationship, with communities at higher levels encompassing those at lower levels.

3.2.2 Modularity

The algorithm outputs a tree diagram that represents the nested hierarchy of potential community divisions in the network. Therefore, it is important to find out which of these divisions is most suitable for a given network, that is, where to cut the dendrogram to obtain a reasonable network division. To address this question, Newman et al. [30] proposed a metric called modularity, which evaluates the quality of a specific portion of the network (Newman and Girvan [30]).

The network is partitioned into κ communities, and a $\kappa \times \kappa$ symmetric matrix $\rceil$ is defined, whose element e_{ij} is the fraction of all edges in the network that link vertices in community i to vertices in community j, and e_{ii} denotes the fraction of all edges within community i. The trace of the matrix, $\mathrm{Tr}(e) = \sum_i e_{ii}$,indicates the total number of edges within communities. A higher $\mathrm{Tr}(e)$ value means the community has more internal connections, indicating a better community detection outcome. To prevent the entire network from becoming a single community, the row (or column) sum $a_i = \sum_j e_{ij}$ is introduced to represent all edges connected to community i. *Modularity* is expressed as:

$$Q = \sum_i \left(e_{ii} - a_i^2\right) = \sum_i (e_{ii}) - \sum_i \left(a_i^2\right) = \mathrm{Tr}(e) - \left\| e^2 \right\| \tag{2}$$

Modularity Q is an accumulation operation for all communities in the network, and subtracts the degree of nodes in the community from the degree of edges in the community. It is a special form of generalization measure. The larger the Q value, the clearer the structure of the community, and its value range is $[-0.5, 1)$. Moving down the dendrogram, the Q for each split of a network into communities is calculated, and the local peak of its value, which indicates a particularly satisfactory split, is looked for. The height of the peak is a measure of the strength of the community division.

3.3 The evaluation of community detection

Named entity recognition (NER) is a fundamental task in natural language processing with diverse applications. It involves identifying entities with specific meanings or strong references in unstructured text, including person, location, organization, date, time, and proper nouns (Cucerzan [4]). NER systems can extract these entities from text and identify additional types based on specific requirements. In the context of Wikipedia communities,

entities can be classified into person, organization, and proper nouns. To evaluate the results obtained by the community detection, we introduce the concept of mode. In statistics, mode (also known as plurality) refers to the value that appears most frequently in a dataset. Mode is usually used to describe the central tendency of a dataset. The term "Mode proportion" was utilized in this study to describe the percentage of the most frequent occurring categories in a community:

$$Mode\ proportion_i = \frac{Count_i}{Count_j} \tag{3}$$

$Count_i$ is the number of entities with largest proportion in community j, and $Count_j$ is the total number of entities in community j.

4 Results

This results section will present the public concerns about AI ethics according to a hierarchical structure from the top to the sub-level. Meanwhile, we explored the relationship between these community topics. Finally, the community detection method was validated.

4.1 Hierarchical community structure

4.1.1 Top-level analysis of community structure

This study crawled Wikipedia data with depth value as two, three, four, and the quality of the inferred network structure is compared to determine the optimal number of depths. The depth value is the number of link extensions of each entry from the primary node (ethics of artificial intelligence), e.g., the depth value of the entry (information society project) to the primary node (ethics of artificial intelligence) is two. When the depth is three, the network structure presents useful and non-redundant information, while the number of links obtained is 1399 (Table 1). The results of hierarchical community detection are shown in Table 2. Using Infomap method, an undirected network containing 983 nodes and 1338 links is constructed. The final best terminal modularization solution is divided into four levels. Seven top-level communities have a codelength of 0.11, and ninety-five total communities correspond to a total codelength of 4.76.

Table 1 Raw data crawled from Wikipedia data

Source	Target	Depth
Ethics of artificial intelligence	Oxford internet institute	1
Oxford internet institute	Information society project	2
Information society project	Oxford internet institute	3

Table 2 The results of hierarchical community detection based on Wikipedia data

	Modules	Leaf Nodes	Child Degree	Codelength for Modules	Codelength for leaf nodes	Codelength Total
Top level	7	0	7	0.11	0	0.11
Second level	91	0	12	1.07	0	1.07
Third level	95	933	11.15	0.01	3.42	3.42
Fourth level	95	50	12.5	0	0.16	0.16
sum	95	983	11.256 (average)	1.19	3.58	4.76

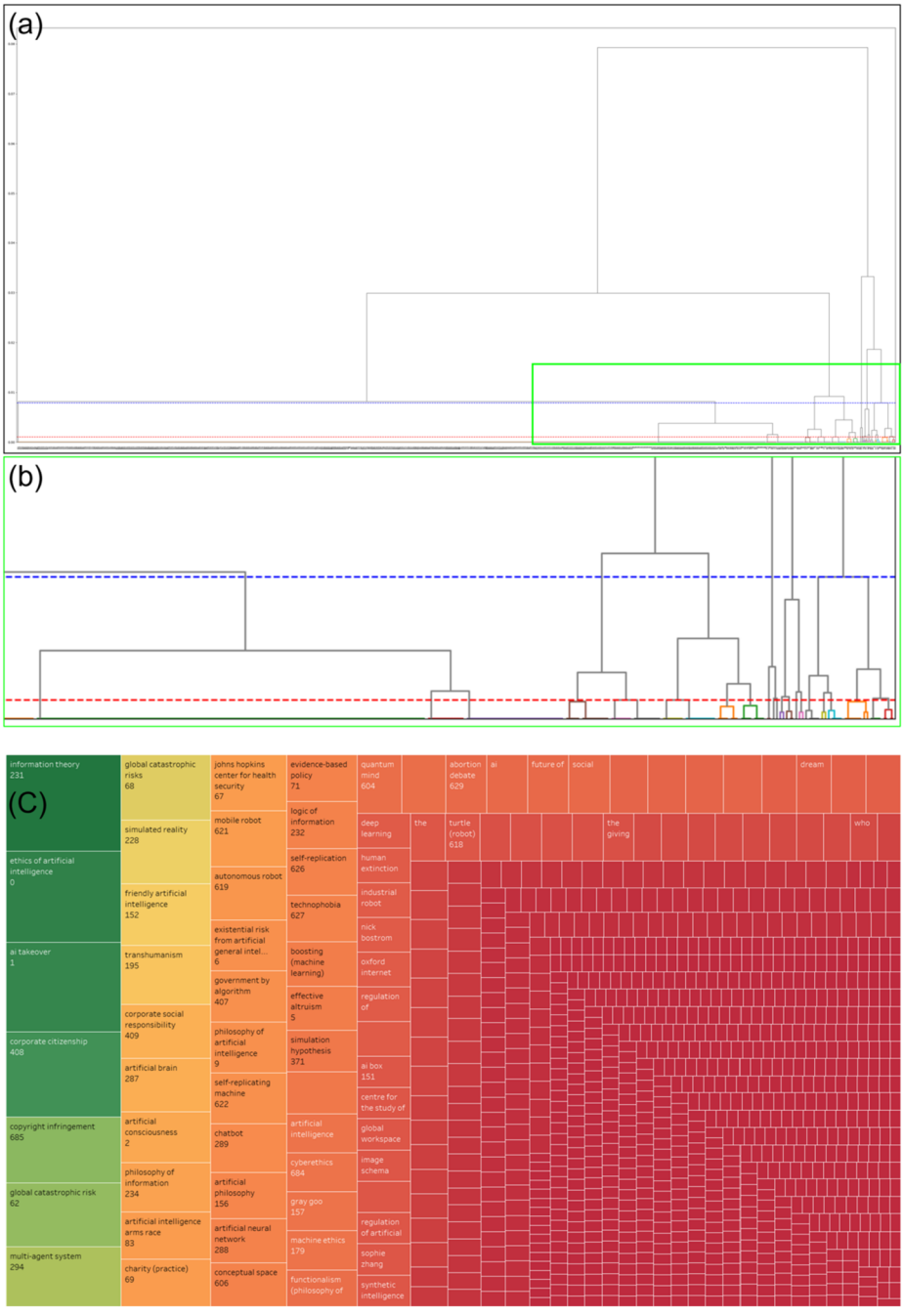

Figure 2 The hierarchical community structure based on Wikipedia data related to AI ethics. (a) The dendrogram figure. (b) Detailed display of Fig. 2(a) (green box). (c) The hierarchical tree

A hierarchical tree or dendrogram can help understand the community structure of Wikipedia data related to AI ethics (Fig. 2). Fig. 2(a) demonstrates Wikipedia data, splitting from top to bottom into smaller and smaller communities, or connecting from bottom to top vertices to form larger and larger communities. A cross-section of the tree at any level, such as that indicated by the dotted line, will give the communities at that level. The vertical height of the split points in the tree simply represents the order in which splits (or

joins) occur. Fig. 2(b) is a detailed presentation of the area marked with green bounding box in Fig. 2(a), where the blue dashed line represents the top-level communities, and the red one corresponds to the perfect identification of the communities. The hierarchical tree diagram in Fig. 2(c) displays the relationship expanded from the backbone nodes, where those nodes most directly connected to AI ethics are "information theory", "ai takeover", "corporate citizenship", "global catastrophic risk", "multi-agent system", reflecting the high level of public interest in these items. The comparison between (a) and (c) in Fig. 2 reveals that the left part of Fig. 2(a) corresponds to the red part in Fig. 2(c) with more nodes, and more modules exist in the second layer (module = 84), which has a compression rate of 44%. The compression rate measures the extent to which information in the network is compressed after applying the Infomap algorithm for network partitioning. A higher compression rate indicates a better partitioning result. The AI ethics-related Wikipedia data present an unbalanced tree structure. When a tree structure is unbalanced, some subtrees can be significantly deeper than others. The left part of Fig. 2(a) shows that the subtree in Community A is deeper and farther from the node "ethics of artificial intelligence," so it does not present an explicit topic. Community B to Community G are the communities that are closely related to the node "ethics of artificial intelligence," and their topics reflect the current public focus.

Figure 3 illustrates seven communities at the top level, demonstrating the division of Wikipedia data related to AI ethics. Word clouds are used to depict the topics associated with each community in Fig. 4. Community A, which has 496 nodes and 41 submodules at the lowest level, lacks a cohesive theme. Conversely, Community B – Community G feature significant topics that are elaborated on in detail below. We always use the center node name of the communities as the community's name.

- Community B (corporate citizenship): it highlights the importance of ethics at the corporate level, with a focus on sustainable business development and social responsibility.
- Community C (information theory): it addresses the theoretical knowledge of AI ethics, drawing on mathematics, physics, and algorithms to explore the technical theories behind ethical decision-making in AI.
- Community D (autonomous robot): it mainly focuses on ethical knowledge related to robots and other automated machines, which shows that robot ethical issues have attracted a widespread attention.
- Community E (semantic space theory): it is related to semantic space and model theory in the field of natural language. From the analysis, this topic is transformed from the field related to artificial intelligence cognition and the exploration of knowledge related to computer language.
- Community F (copyright infringement): it is concerned with copyright infringement issues related to the Internet, including illegal downloading, illegal copying by dongles, and privacy violations by online piracy.
- Community G (philosophy of artificial intelligence): it collects knowledge related to cognitive and consciousness issues, including philosophical cognition, quantum thinking, and more. The brain is studied by interpreting consciousness as an electromagnetic phenomenon, which can be applied to the field of artificial intelligence.

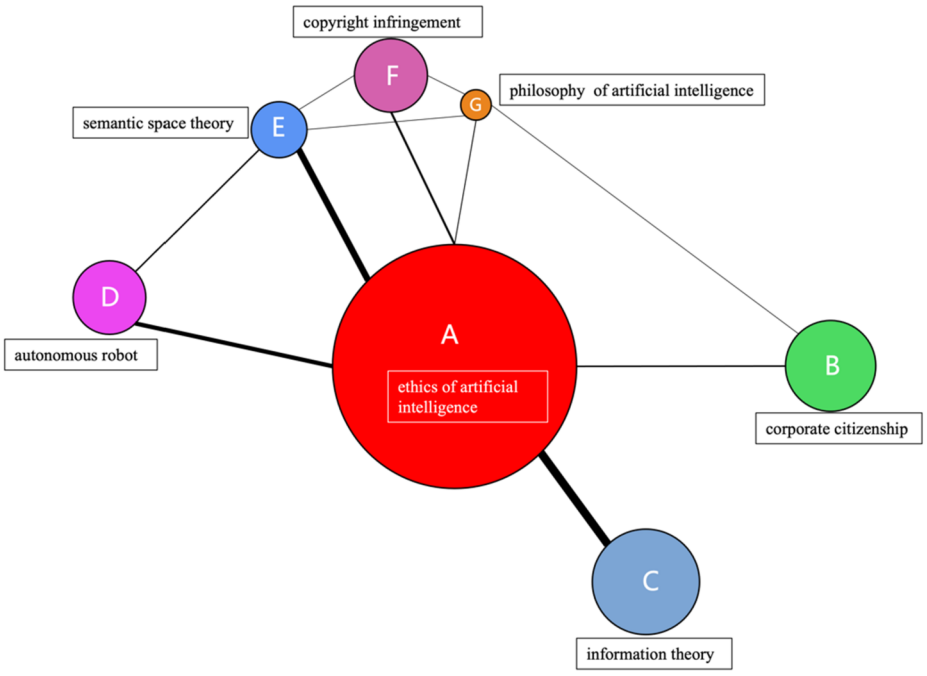

Figure 3 Community Detection results at the top level. The left table shows the correspondence between community numbers and community center nodes

4.1.2 Sub-level community structure based on modularity

The raw Wikipedia data is presented as a disorganized network, making it difficult to derive useful information from the graph. In the hierarchical community division obtained by Infomap, we found that the modularity has a strong peak at 24 communities with a value of $Q = 0.805$. These communities were extracted from the corresponding tree diagram and visualized by VOSviewer, as shown in Fig. 5(a). Each community is represented by the same color. The node size is determined by the degree, which is the number of edges connected to the node within the community. A larger node generally indicates greater importance of the node within that community. In Fig. 5(a), the community's name is based on the name of the central node with the highest degree within the community. Between the communities, starting from the branch node "ethics of artificial intelligence", each branch node is related to each other. The more connections, the stronger relationships between the communities. For example, the community with the most connections to the node "ethics of artificial intelligence" is the community whose branch node is "ai takeover". In Fig. 5(b), the network is reduced to groups, where each community is displayed as a circle, whose size roughly corresponds to the number of individuals in the community. The lines between the communities depict the connections between AI ethical knowledge, with the thickness of the lines proportional to the association.

The 24 communities in this level exhibit a distinct thematic coherence when compared to the seven communities in the top-level. For instance, in Fig. 5(b), the community whose branch node is "global catastrophic risk" (dark purple section) and the community whose branch node is "ai takeover" (red section) both originate from top-level Community A. Figure 6 displays the word cloud of the two communities, showing that the topics of both communities relate to risk. Nevertheless, the Community "global catastrophic risk" fo-

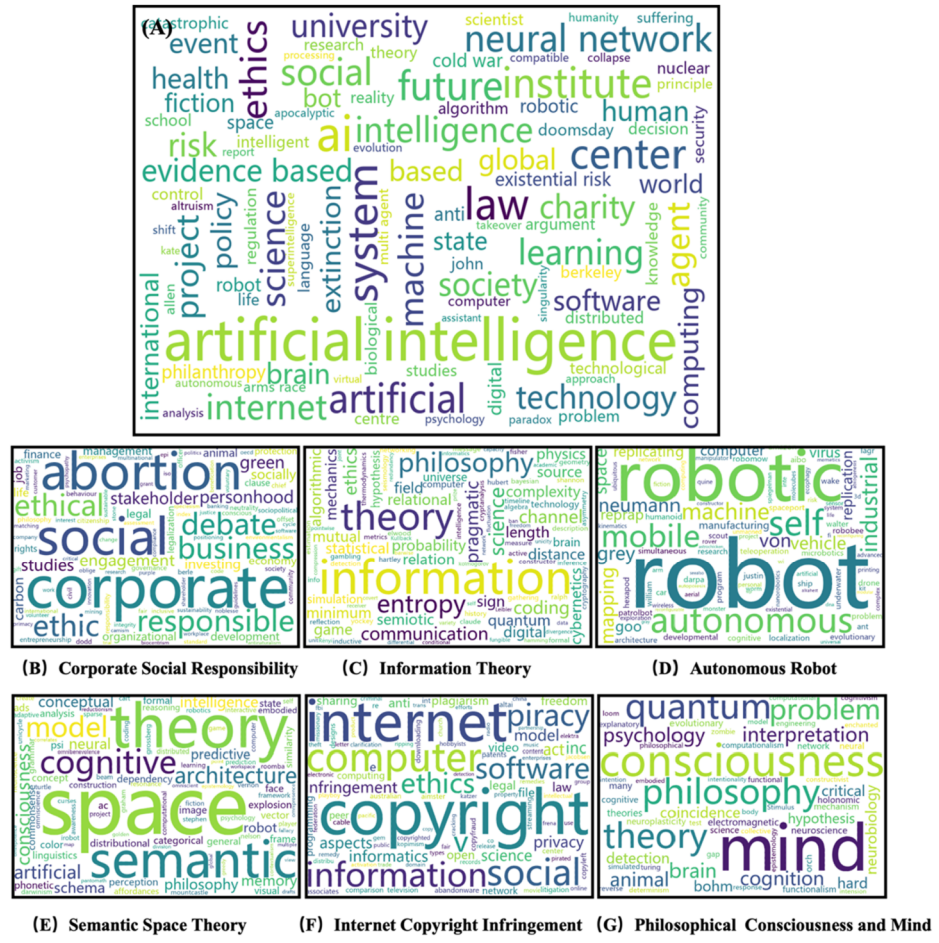

Figure 4 Topics learned from Wikipedia regarding the AI ethics. Community B- Community G give clear themes

cuses more on future risks, such as doomsday scenarios, human extinction, and the future of the earth. On the other hand, the Community "ai takeover" mainly explores the current risks in AI technology to ensure that machines serve humans better.

4.2 The evaluation of community detection

This study evaluates the results of community detection by calculating the mode proportion of 24 communities based on three categories: proper nouns, person, and organization. Communities are comprised of nodes with strong connections and similarities. In general, nodes with the same category are more closely connected, resulting in higher values of mode proportion within a community. Moreover, communities with lower mode proportion may reveal hidden information we can't discover directly. Table 3 presents the statistical results, which are further analyzed as follows.

Table 3 has a total of 971 statistical nodes, including 852 proper nouns, 52 persons, and 67 organizations. Among them, only Community 1 (ethics of artificial intelligence) and Community 4 (johns hopkins center for health security) had the mode proportion of less than 80%. Upon further analysis, community 1 (ethics of artificial intelligence) mainly consists of researchers in the field of AI ethics, such as computer scientists, sociologists,

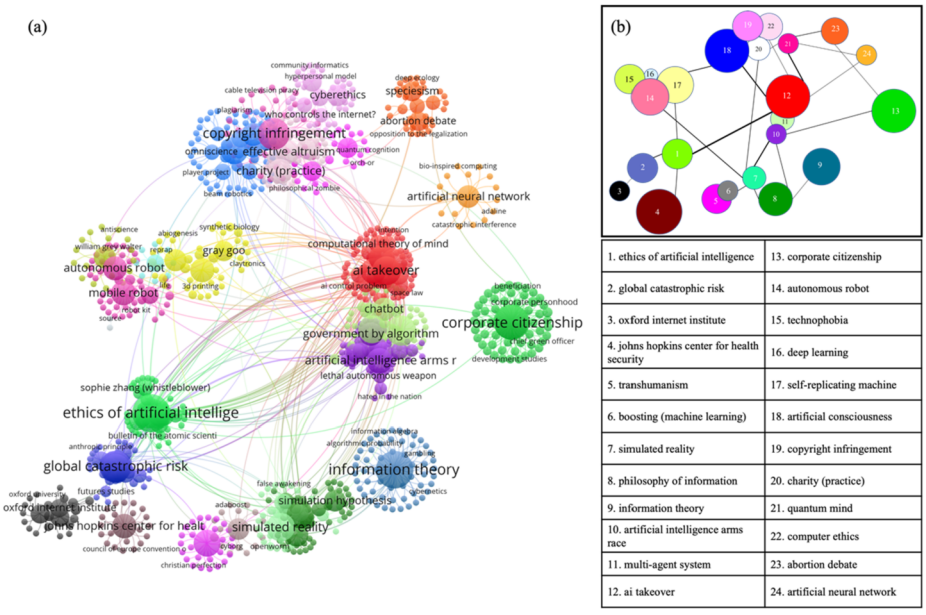

Figure 5 The detected communities in Wikipedia data related to AI ethics. (a) The coarse-grained structure results of community detection based on the maximum modularity, where each community is represented by a single node and the edges indicate the collaboration between communities. The table below shows the corresponding central nodes of the 24 communities. (b) The elaborated demonstration of community structure with node labels

Figure 6 Topics of Community "global catastrophic risk" and Community "ai takeover"

and philosophers, whose work is focused on topics such as AI ethics, machine ethics, algorithmic discrimination, sexism, and fairness. The institutions in this community are primarily research institutes associated with the researchers, and a few proper nouns that are related to their research fields are also included. Community 4 (johns hopkins center for health security) is mainly composed of proper nouns and research institutions, with proper nouns being the dominant type. However, this community brings together almost 33% of research institutions related to AI ethics, whose number of institutions is similar to community 3 (oxford internet institute), a community with 100% mode proportion of institutions. Community 4 (johns hopkins center for health security) is primarily comprised of research institutions that specialize in brain and health, which is significantly

Table 3 The mode proportion statistics for the 24 communities

Name of Community	Proper Nouns	Person	Organization	Mode proportion	Community	Proper Nouns	Person	Organization	Mode proportion
1. ethics of artificial intelligence	6	34	9	69.39%	13. corporate citizenship	65	0	2	97.01%
2. global catastrophic risk	44	1	0	97.78%	14. autonomous robot	44	1	0	97.78%
3. oxford internet institute	0	0	22	100.00%	15. technophobia	18	0	0	100.00%
4. johns hopkins center for health security	60	1	20	74.07%	16. deep learning	9	0	0	100.00%
5. transhumanism	36	0	0	100.00%	17. self-replicating machine	54	1	0	98.18%
6. boosting (machine learning)	16	1	0	94.12%	18. artificial consciousness	67	2	0	97.10%
7. simulated reality	25	3	0	89.29%	19. copyright infringement	38	1	3	90.48%
8. philosophy of information	50	2	1	94.34%	20. charity (practice)	38	0	7	84.44%
9. information theory	56	2	0	96.55%	21. quantum mind	38	0	0	100.00%
10. artificial intelligence arms race	15	0	0	100.00%	22. computer ethics	40	0	2	95.24%
11. multi-agent system	23	0	0	100.00%	23. abortion debate	27	0	0	100.00%
12. ai takeover	67	3	1	94.37%	24. artificial neural network	16	0	0	100.00%

different from the institutions in Community 3 (oxford internet institute), which mostly focus on law and policy. Due to the proper nouns in Community 4 (johns hopkins center for health security) being related to brain and health, the institutions associated with them are also classified in Community 4 (johns hopkins center for health security), resulting in a decrease in the community's mode proportion.

5 Discussion

Wikipedia provides a good platform for recognizing the current state of knowledge about AI ethics. Utilizing the community detection method to organize the knowledge into a hierarchical structure, we can summarize current public concerns into seven communities in a coarse intensity. Additionally, a more detailed knowledge division is presented layer by layer to cater to a more refined understanding. At the top level, Community A (ethics of artificial intelligence) coarsely includes persons, organizations, and proper nouns related to AI ethics. The six other communities, which are distinctly thematic, show that public concerns focus on corporate social responsibility, mathematical-related technical knowledge, automated robots, semantic space theory, internet copyright infringement, and philosophical perception and mind. These topics not only expose public knowledge needs but also reveal the current issues focusing on AI ethics. Community B (corporate citizenship) reveals that major technology companies usually spearhead the promotion and application of AI technologies. For each company, a better balance between the pursuit of profit and social responsibility is needed to avoid the proliferation of AI ethical issues, which affects the sustainable development of AI technologies. Community C (information theory) and Community G (philosophy of artificial intelligence) reveal the fundamental knowledge behind AI technology development, which is supported by natural science such as mathematics, physics, and algorithms on the one hand, and social science such as philosophy and psychology on the other. The topics of these two communities reflect an inquiry into the knowledgeable focus behind technology and the intellectual underpinnings of AI technology development. The topics of Community D (autonomous robot), Community E (semantic space theory), and Community F (copyright infringement) indirectly manifest the current AI ethical issues on Wikipedia, which are consistent with the leading AI ethical issues obtained from news data. Notably, the three major areas of

AI ethical issues obtained through content analysis are Intelligent Service Robots, Language/Vision Models, and Autonomous Driving (Wei and Zhou [45]). The topic within Community D (autonomous robot) indeed includes Intelligent Service Robots and Autonomous Driving, which focuses on the issues raised by automated machines. Community E (semantic space theory) is centered around the semantic space theory, providing theoretical support for technologies like natural language processing, whose topic is compatible with AI ethical issues in the "Language/Vision Model" field. In contrast to news data, the theme of Community F (copyright infringement) reveals a different issue from a macro perspective, namely Internet copyright infringement. The development of AI technology has profoundly impacted individual privacy and monopoly rights. Technology has made it easier to infringe on these rights, as seen with illegal online downloading and copying, which is difficult to counteract. On the other hand, the relevant laws still need to be established. The rapid development of technology has resulted in a certain degree of lag in the law, which in turn has further exacerbated ethical issues.

Clipping the dendrogram when the modularity = 0.805 yields a reasonable community division. An exciting finding is the understanding of current researchers and institutions in AI ethics. Researchers around the central node of "ethics of artificial intelligence" are mainly computer scientists, followed by philosophers and social activists. Most of them hold research positions in universities or technology companies, whose research areas are data, algorithms, and the fairness and sustainability of technology. Besides, one node in this community extends to another small community of persons who share a common characteristic of bringing widespread attention to unethical behavior in large technology companies. For example, a data scientist once uncovered details of Facebook's superficial monitoring of online political manipulation (Susser et al. [41]). From the perspective of institutions, the community detection results show that one group of institutions is the "oxford internet institute" for law and guidelines on AI ethics, and the other group is the "Johns Hopkins center for health security" for brain and health research. The distinct types of institutions exist independently in two different communities. A finer granularity of community detection can help us learn more detailed knowledge about AI ethics.

The development of AI aims to better serve human beings. Public perception of AI is equally important in shaping its development (Kelley et al. [19]; Kieslich et al. [20]; Sartori and Bocca [36]). Users' inadequate knowledge of AI technology can exacerbate fears and prevent them from effectively regulating AI technology for human benefit. Distinguished from examining public concerns about AI ethics or identifying emotional categories (Fast and Horvitz [8]), utilizing Wikipedia to investigate public perceptions not only highlights the trending topics of public interest but also unveils the knowledge underpinning these issues. This approach serves to enhance public awareness of AI ethics from a knowledge-related perspective. We utilize a community detection approach to analyze and present the AI ethical knowledge on Wikipedia in a structured way, which helps to fill the gap in the public knowledge of AI ethics and minimize ethical issues.

6 Conclusion

This paper explores public perception towards AI ethics based on Wikipedia data, which presents the hierarchical structure and the links among the detected nodes of AI ethics. The hierarchical knowledge structure of AI ethics in Wikipedia data can be mapped, using community detection. Specifically, we extracted seven communities at the top level and

discovered six topics with high public attention, that is "corporate social responsibility", "information theory", "autonomous robot", "semantic space theory", "internet copyright infringement", "philosophical consciousness and mind". These topics reflect knowledge related to AI technology and expose major AI ethical issues. We compared these results with the AI ethical issues from news data statistics and found a good match. In addition, we rationalized 24 communities based on modularity and visualized the connections among these communities in depth. Finally, we identified the main research organizations, relevant people, and proper nouns related to AI ethics, and assessed the reliability of the community divisions. Revealing the holistic knowledge structure of AI ethical topics and the relations among the subtopics, contributing to better education on AI ethics. Our results help improve the existing body knowledge of AI ethics and enhance public perception of AI ethics to mitigate the ethical risks.

There are three limitations of this study. First, the only the texts in English were included in our sample data, so we might lose some information from the texts in other languages. Second, most of the Wikipedia content is contributed by experts. To better serve the goal of reflecting the perception of AI ethics, more is needed to disseminate complete knowledge to non-expert users. Third, our analysis of relationships within and between communities is limited to correlation, representing only the initial stage of the relationship (Pearl, 2018). To better regulate AI technologies and build ethical guidelines, we need to explore intervening relationships between communities.

It should be noted that Wikipedia is not merely a dataset, but also serves as a dynamic platform for disseminating knowledge and shaping public perception regarding AI ethics. These changes may take time to become evident. Thus, it would be necessary to revalidate Wikipedia's AI ethics knowledge structure by comparing any shifts in public perception, e.g., five years from now, to regulate the development of AI technology.

Acknowledgements
Not applicable.

Funding
Open access funding provided by Swiss Federal Institute of Technology Zurich.

Abbreviations
AI, Artificial Intelligence; MEPs, Members of European Parliament; NER, Named entity recognition.

Data availability
The datasets generated and analyzed during the current study are available in the https://figshare.com/account/home.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author contributions
All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by MW, PL, and CZ. The first draft of the manuscript was written by MW and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.
Conceptualization: MW, CZ, LM; Methodology: MW, PL, LM; Formal analysis and investigation: MW, CC; Writing-original draft preparation: MW; Writing-review and editing: MW, YF, CC, PL, CZ, LM; Supervision: LM.

Author details
[1]Chair of Cartography and Visual Analytics, Technical University of Munich, Arcisstrasse 21, Munich, 80333, Germany.
[2]Center for Sustainable Future Mobility (CSFM), ETH Zurich, Universitätstrasse 41, UNO D 12, Zürich, 8092, Switzerland.

Received: 11 October 2023 Accepted: 11 March 2024 Published online: 26 March 2024

References

1. Araujo T, Helberger N, Kruikemeier S, de Vreese CH (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc 35:611–623. https://doi.org/10.1007/s00146-019-00931-w
2. Bohlin L, Edler D, Lancichinetti A, Rosvall M (2014) Community detection and visualization of networks with the map equation framework. In: Ding Y, Rousseau R, Wolfram D (eds) Measuring scholarly impact. Springer, Cham, pp 3–34
3. Buscaldi D, Rosso P (2006) Mining knowledge from Wikipedia for the question answering task
4. Cucerzan S (2007) Large-scale named entity disambiguation based on Wikipedia data
5. Das S, Lavoie A, Magdon-Ismail M (2011) Pushing your point of view: behavioral measures of manipulation in Wikipedia
6. Di Lauro F, Johinke R (2017) Employing Wikipedia for good not evil: innovative approaches to collaborative writing assessment. Assess Eval High Educ 42:478–491. https://doi.org/10.1080/02602938.2015.1127322
7. Dowler E, Bauer M, Green J, Gasperoni G (2006). Assessing public perception: issues and methods
8. Fast E, Horvitz E (2016) Long-term trends in the public perception of artificial intelligence
9. Ferguson MJ, Bargh JA (2004) How social perception can automatically influence behavior. Trends Cogn Sci 8:33–39. https://doi.org/10.1016/j.tics.2003.11.004
10. Fichman P, Hara N (2014) Global Wikipedia: international and cross-cultural issues in online collaboration. Rowman & Littlefield, Totowa
11. Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174. https://doi.org/10.1016/j.physrep.2009.11.002
12. Fu Y, Zhuang Z, Zhang L (2022) AI ethics on blockchain: topic analysis on Twitter data for blockchain security
13. Gillespie N, Lockey S, Curtis C et al (2023) Trust in artificial intelligence: a global study. University of Queensland, Brisbane
14. Greenstein S, Zhu F (2012) Is Wikipedia biased? Am Econ Rev 102:343–348. https://doi.org/10.1257/aer.102.3.343
15. Grünwald P, Myung J, Pitt M (2005) Advances in minimum description length
16. Halatchliyski I, Moskaliuk J, Kimmerle J, Cress U (2014) Explaining authors' contribution to pivotal artifacts during mass collaboration in the Wikipedia's knowledge base. Int J Comput-Support Collab Learn 9:97–115. https://doi.org/10.1007/s11412-013-9182-3
17. Ikkatai Y, Hartwig T, Takanashi N, Yokoyama HM (2022) Octagon measurement: public attitudes toward AI ethics. Int J Hum-Comput Interact 38:1589–1606. https://doi.org/10.1080/10447318.2021.2009669
18. Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1:389–399. https://doi.org/10.1038/s42256-019-0088-2
19. Kelley PG, Yang Y, Heldreth C et al (2021) Exciting, useful, worrying, futuristic: public perception of artificial intelligence in 8 countries. In: Proceedings of the 2021 AAAI/ACM conference on AI, ethics, and society. Assoc. Comput. Mach., New York, pp 627–637
20. Kieslich K, Keller B, Starke C (2022) Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. Big Data Soc 9:205395172210929. https://doi.org/10.1177/20539517221092956
21. Köchling A, Wehner MC (2020) Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Bus Res 13:795–848. https://doi.org/10.1007/s40685-020-00134-w
22. Lizorkin D, Medelyan O, Grineva M (2009) Analysis of Community Structure in Wikipedia (Poster)
23. Majchrzak A (2009) Comment: where is the theory in wikis? MIS Q 33:18–20. https://doi.org/10.2307/20650275
24. Medelyan O, Milne D, Legg C, Witten IH (2009) Mining meaning from Wikipedia. Int J Hum-Comput Stud 67:716–754. https://doi.org/10.1016/j.ijhcs.2009.05.004
25. Microsoft (2022) Microsoft responsible AI standard v2 general requirements. Impact assess
26. Moy CL, Locke JR, Coppola BP, McNeil AJ (2010) Improving science education and understanding through editing Wikipedia. J Chem Educ 87:1159–1162. https://doi.org/10.1021/ed100367v
27. Nastase V, Strube M (2008) Decoding Wikipedia Categories for Knowledge Acquisition
28. Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256. https://doi.org/10.1137/S003614450342480
29. Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256. https://doi.org/10.1137/S003614450342480
30. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:026113. https://doi.org/10.1103/PhysRevE.69.026113
31. Nguyen DPT (2007) Relation extraction from Wikipedia using subtree mining
32. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435:814–818. https://doi.org/10.1038/nature03607
33. Rosvall M, Axelsson D, Bergstrom CT (2009) The map equation. Eur Phys J Spec Top 178:13–23. https://doi.org/10.1140/epjst/e2010-01179-1
34. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci 105:1118–1123. https://doi.org/10.1073/pnas.0706851105
35. Rosvall M, Bergstrom CT (2011) Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PLoS ONE 6:e18209. https://doi.org/10.1371/journal.pone.0018209
36. Sartori L, Bocca G (2022) Minding the gap(s): public perceptions of AI and socio-technical imaginaries. AI Soc. https://doi.org/10.1007/s00146-022-01422-1
37. Siau K, Wang W (2020) Artificial Intelligence (AI) ethics: ethics of AI and ethical AI. J Database Manag 31:74–87. https://doi.org/10.4018/JDM.2020040105

38. Smith BK, Gustafson A (2017) Using Wikipedia to predict election outcomes. Public Opin Q 81:714–735. https://doi.org/10.1093/poq/nfx007
39. Srivastava A, Geethakumari G (2013) Measuring privacy leaks in online social networks. In: 2013 international conference on advances in computing, communications and informatics (ICACCI). IEEE, Mysore, pp 2095–2100
40. Stahl BC (2021) Ethical issues of AI. In: Artificial intelligence for a better future. Springer, Cham, pp 35–53
41. Susser D, Roessler B, Nissenbaum H (2019) Online manipulation: hidden influences in a digital world. Georget Law Technol Rev 4:1–46
42. Tomašev N, Cornebise J, Hutter F et al (2020) AI for social good: unlocking the opportunity for positive impact. Nat Commun 11:2468. https://doi.org/10.1038/s41467-020-15871-z
43. van Steen M (2010) Graph theory and complex networks
44. Vieira Bernat M (2023) Topical classification of images in Wikipedia: development of topical classification models followed by a study of the visual content of Wikipedia
45. Wei M, Zhou Z (2022) AI ethics issues in real world: evidence from AI incident database
46. Zickuhr K, Rainie L (2011) Wikipedia. past and present
47. Zipf GK (1949) Human behavior and the principle of least effort: an introduction to human eoclogy

Publisher's Note