International Cartographic Association
Association Cartographique Internationale

Visual analysis of AI ethics cases

Chuan Chen [a*], Mengyi Wei [a], Liqiu Meng [a*],

[a] *Chair of Cartography and Visual Analytics, Technical University of Munich, Munich, Germany, Chuan Chen – chuan.chen@tum.de, Mengyi Wei – mengyi.wei@tum.de, Liqiu Meng - liqiu.meng@tum.de*

* Corresponding author

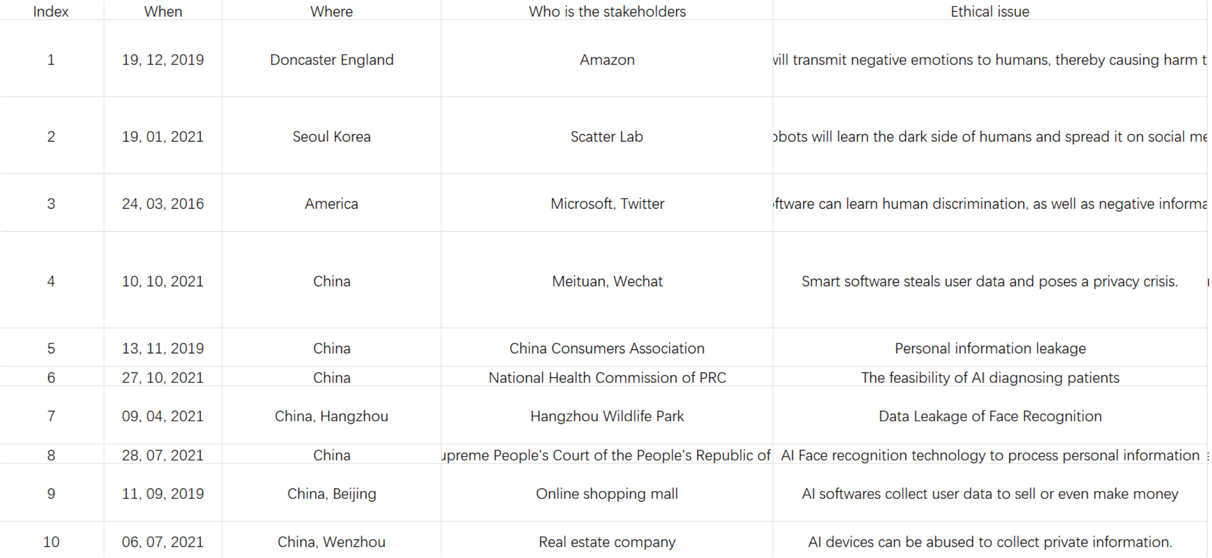
Keywords: AI ethics, Graph convolutional network, Graph visualization, Cross-theme analysis

Abstract:

Artificial intelligence (AI) and humans have become more intertwined than ever before, thus bringing up diverse AI ethics issues. Social media, autonomous driving, web search are just some of the areas where AI and humans are working side-by-side. An in-depth understanding of AI ethics is indispensable to improve the ethical norms of AI and promote better coexistence between AI and humans.

Previous research on AI ethics reveals a general trend of researchers focusing on specific themes such as fairness and privacy. For example, Aïvodji et al. (2019) and Arnold et al. (2022) investigated the fairness of an algorithm in terms of whether potential discrimination exists. Studies on bias resulting from the unfairness are more diverse. Sun et al. (2019) elaborated on gender bias in natural language processing. Hutchinson et al. (2020) examined bias in machine learning for people with disabilities. Baron and Musolesi (2020) have focused on privacy leakage by providing an explainable frameworks to allow people to understand how privacy is stolen. Yet the existing studies have been conducted in isolation and could not answer the question of whether there is homogeneity and heterogeneity between AI ethics cases related to personal privacy leakage and those involving algorithmic fairness. This paper takes a macro perspective that considers different themes as components of a system and examines their potential correlations.

The AI ethics cases are collected from public AI incidents database, which are stored in the structure of attribute matrixes as shown in Figure 1. Figure2 depicts the workflow in detail. The authors visually analyze the correlations of AI ethics cases across themes with the graph convolutional network (GCN) (Kipf and Welling, 2016) and the graph database Neo4j. AI ethics cases are first modeled through attributional organization. Their attributes are encoded into vectors by term frequency–inverse document frequency (TF-IDF) algorithm (Martineau and Finin, 2009). Then a link prediction task is implemented to identify potential correlations between cases through a two-layer GCN. Neo4j is used to visualize the results of link prediction.



Index	When	Where	Who is the stakeholders	Ethical issue
1	19, 12, 2019	Doncaster England	Amazon	vill transmit negative emotions to humans, thereby causing harm t
2	19, 01, 2021	Seoul Korea	Scatter Lab	obots will learn the dark side of humans and spread it on social me
3	24, 03, 2016	America	Microsoft, Twitter	ftware can learn human discrimination, as well as negative informa
4	10, 10, 2021	China	Meituan, Wechat	Smart software steals user data and poses a privacy crisis.
5	13, 11, 2019	China	China Consumers Association	Personal information leakage
6	27, 10, 2021	China	National Health Commission of PRC	The feasibility of AI diagnosing patients
7	09, 04, 2021	China, Hangzhou	Hangzhou Wildlife Park	Data Leakage of Face Recognition
8	28, 07, 2021	China	upreme People's Court of the People's Republic of	AI Face recognition technology to process personal information
9	11, 09, 2019	China, Beijing	Online shopping mall	AI softwares collect user data to sell or even make money
10	06, 07, 2021	China, Wenzhou	Real estate company	AI devices can be abused to collect private information.

Figure 1. Part of the AI ethics attribute matrix

Abstracts of the International Cartographic Association, 6, 35, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-35-2023 | © Author(s) 2023. CC BY 4.0 License.

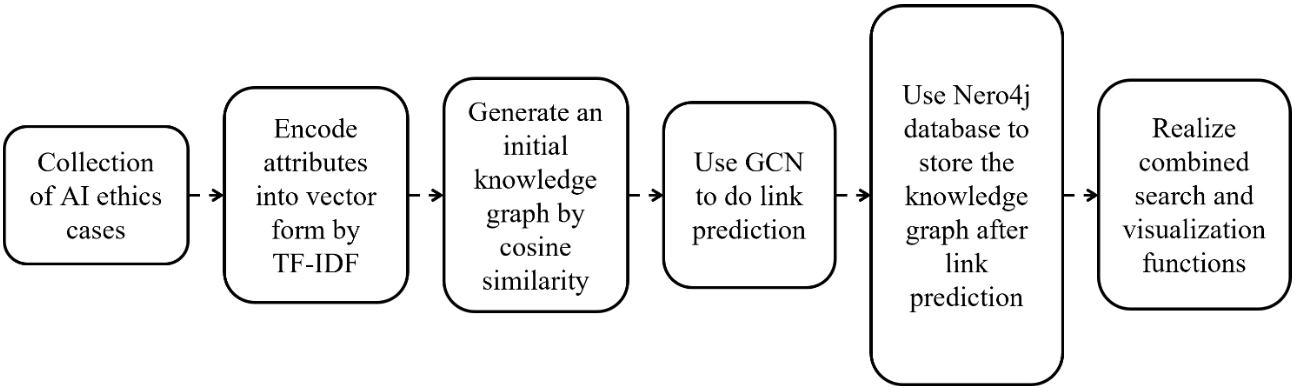

Collection of AI ethics cases
Encode attributes into vector form by TF-IDF
Generate an initial knowledge graph by cosine similarity
Use GCN to do link prediction
Use Nero4j database to store the knowledge graph after link prediction
Realize combined search and visualization functions

Figure 2. Technology route map

In addition to data storage, deletion and modification, the authors have designed rich query functions for this database of AI ethics cases. This enables users to find more useful information about AI ethics issues. Figure 3 visualize 12 cases resulting from a combinational query for intelligent recommendation and negative public opinions. With these preliminary results, the authors aim to appeal to the cartographic community for a systemic understanding of AI ethics issues supported by visual analytics.

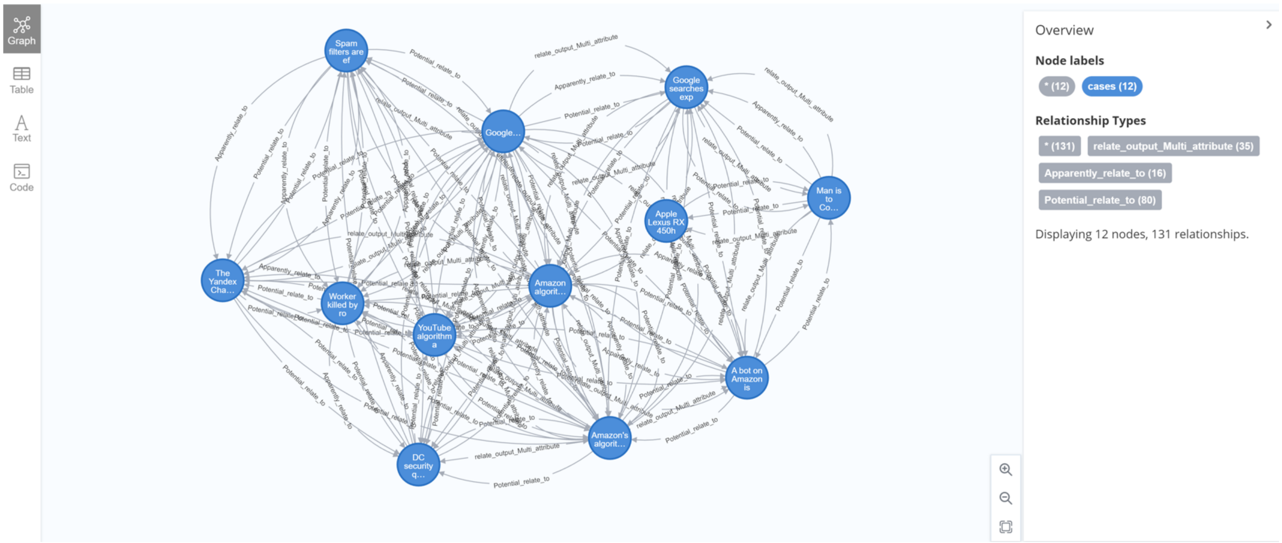


Figure 3. Visualization of 12 cases resulting from a combined search.

References

Arnold, D., Dobbie, W., and Hull, P, 2022. Measuring racial discrimination in bail decisions. American Economic Review, 112(9), pp. 2992-3038.

Aïvodji, U., Arai, H., Fortineau, O., Gambs, S., Hara, S., and Tapp, A, 2019. Fairwashing: the risk of rationalization. In International Conference on Machine Learning, pp. 161-170.

Baron, B., and Musolesi, M, 2020. Interpretable machine learning for privacy-preserving pervasive systems. IEEE Pervasive Computing, 19(1), pp. 73-82.

Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., and Denuyl, S, 2020. Unintended machine learning biases as social barriers for persons with disabilitiess. ACM SIGACCESS Accessibility and Computing, Vol. 1, No. 1, pp. 125.

Kipf, T. N., and Welling, M, 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Martineau, J., and Finin, T, 2009. Delta tfidf: An improved feature space for sentiment analysis. In Proceedings of the International AAAI Conference on Web and Social Media Vol. 3, No. 1, pp. 258-261.

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., ... and Wang, W. Y, 2019. Mitigating gender bias in natural language processing: Literature review. arXiv preprint arXiv:1906.08976.

https://doi.org/10.5194/ica-abs-6-35-2023 | © Author(s) 2023. CC BY 4.0 License.