Perceptions of AI Ethics on Social Media

Ayse Ocal
Information Science and Technology
Syracuse University
Syracuse, NY
ayocal@syr.edu
ORCID: 0000-0002-1925-4305

Abstract— Since the emergence of Artificial intelligence (AI), despite a common expectation that AI should be 'ethical' [1], there are many different interpretations, assumptions, and expectations about what constitutes "ethical AI" and which ethical problems and requirements are pointed out by the public. Even though many private companies and research institutions have highlighted present and possible future problems, needs, and guidelines associated with AI ethics, relevant public visions regarding how "ethical AI" can be constituted [1] have not been explored sufficiently. For obtaining public opinions, although questionnaires and interviews are commonly used, the questions in these methods are designed based on only the researchers' preferences, and this could be a limitation. Social media data, however, are produced by users freely [2], and many people share their ideas in social media discussions [3]. Social media data usage, therefore, has been growing in various research studies. Researchers intending to utilize social media as a data source predominantly harness Twitter data, yet in recent years Reddit has also gained the attention of scholars with the same research purpose, as in [2], [3]. Reddit is a huge social media platform involving over 50 million daily active users with diverse mentalities shaped by different backgrounds, prior beliefs, personal experiences, and personalities, from various geographical locations, and 100 thousand active communities, thereby it brings different segments of the public together. Moreover, users benefit from a level of anonymity on Reddit not typically accomplished on other social media platforms [4], thereby users may feel more secure and share more honest thoughts on a topic, thus the Reddit data have been used to gather public opinions in prior research as in [3]. Through the lens of technological frames [5], to explore social media users' interpretations, assumptions, and expectations about how ethical AI is built, and which problems hinder building ethical AI, Reddit conversations were analyzed. More specifically, a corpus consisting of 998 unique Reddit post titles and their corresponding 16611 comments extracted from 15 AI-related subreddits were identified by using topic modelling supported by human judgment for frame identification as in [6] based on BERTopic [7]. The findings show that perceptions about AI ethics are clustered around several themes (AI's gender bias; humans' gender bias about perceived gender of bots; regulation and patent laws related to AI use; AI spreading disinformation; AI making fake faces, videos, music; misuse of personal data; and AI impact on crime), with deviations about how these themes are interpreted, what problems or actors they pertain to, and what appropriate measures should be taken to address problems pointed out by the public. While some of these ethical issues were also highlighted in prominent AI ethics literature as in [8], the findings of this study indicated new insights such as humans' gender bias about the perceived gender of bots. The findings offer important implications. First, as a practical implication, the findings can enrich current public voice-centric explorations of AI ethics. Also, they could help designing suitable interfaces that allow proper human-AI task coordination and collaboration and deploying innovative solutions for existing or anticipated ethical problems. Second, expected outcomes can demonstrate areas where misconceptions and unrealistic visions about AI ethics are widespread, which may trigger speculative fears or concerns. Researchers may be encouraged to more focus on the areas where public misconceptions are more common; educational programs may be arranged to reduce speculative fears or concerns or take necessary measures for real ethical risks. Academia, industry and government communities may collaborate for research and policy arrangements in those areas. Third, through employing computer-aided textual analysis, this study reveals frames in social media conversations to showcase the latest perceptions from different viewpoints. This method may be an example method for relevant future research.

Keywords—AI ethics, social media, Reddit, public perceptions, technological frames, bias.

References

[1] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, Art. no. 9, Sep. 2019, doi: 10.1038/s42256-019-0088-2.

[2] A. Öcal, L. Xiao, and J. Park, "Reasoning in social media: insights from Reddit 'Change My View' submissions," *Online Information Review*, vol. 45, no. 7, pp. 1208–1226, Jan. 2021, doi: 10.1108/OIR-08-2020-0330.

[3] K. Chen and D. Tomblin, "Using Data from Reddit, Public Deliberation, and Surveys to Measure Public Opinion about Autonomous Vehicles," *Public Opinion Quarterly*, vol. 85, no. S1, pp. 289–322, Sep. 2021, doi: 10.1093/poq/nfab021.

[4] N. Proferes, N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer, "Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics," *Social Media + Society*, vol. 7, no. 2, p. 20563051211019004, Apr. 2021, doi: 10.1177/20563051211019004.

[5] W. J. Orlikowski and D. C. Gash, "Technological frames: making sense of information technology in organizations," *ACM Trans. Inf. Syst.*, vol. 12, no. 2, pp. 174–207, Apr. 1994, doi: 10.1145/196734.196745.

[6] L. Guo et al., "Proposing an Open-Sourced Tool for Computational Framing Analysis of Multilingual Data," *Digital Journalism*, pp. 1–22, Feb. 2022, doi: 10.1080/21670811.2022.2031241.

[7] G. Hristova and N. Netov, "Media Coverage and Public Perception of Distance Learning During the COVID-19 Pandemic: A Topic Modeling Approach Based on BERTopic," in *2022 IEEE International Conference on Big Data (Big Data)*, Dec. 2022, pp. 2259–2264. doi: 10.1109/BigData55660.2022.10020466.

[8] N. Bostrom and E. Yudkowsky, "The Ethics of Artificial Intelligence," in *Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, 2011.