Ethical artificial intelligence (AI): confronting bias and discrimination in the library and information industry

Hamid Reza Saeidnia

Introduction

Bias and discrimination in artificial intelligence (AI) refers to the unfair or unjust treatment of individuals or groups as a result of biases embedded within AI algorithms or systems (Ferrer *et al.*, 2021). AI algorithms are designed to make decisions or predictions based on patterns and data (Malek, 2022). However, these algorithms can inadvertently reflect the biases and prejudices present in the data they are trained on or the way they are programmed (Ferrer *et al.*, 2021; Malek, 2022).

Bias in AI can occur in various forms. One common type is algorithmic bias, where AI systems produce discriminatory outcomes or decisions that disproportionately affect certain individuals or groups (Akter *et al.*, 2021). This can happen when the training data used to develop the AI algorithm is biased or lacks diversity, leading to biased predictions or recommendations. For example, if a facial recognition system is trained on predominantly white faces, it may have difficulty accurately recognizing or identifying individuals with darker skin tones.

Discrimination in AI can also occur when algorithms are used to make decisions in areas such as hiring, lending or criminal justice (Zuiderveen Borgesius, 2018). If these algorithms are trained on biased historical data or reflect societal biases, they can perpetuate or amplify discriminatory practices. For instance, an AI hiring tool might inadvertently discriminate against certain demographics if it is trained on data that reflects biased hiring decisions from the past (Ferrer *et al.*, 2021; Akter *et al.*, 2021; Zuiderveen Borgesius, 2018).

Addressing bias and discrimination in AI is crucial for ensuring fairness and equality (Ferrer *et al.*, 2021; Akter *et al.*, 2021). This involves examining and mitigating biases in training data, improving algorithmic transparency and explainability, involving diverse stakeholders in the development and evaluation process and implementing ethical guidelines and regulations (Malek, 2022; Zuiderveen Borgesius, 2018). It is important to note that bias and discrimination in AI are not intentional, but rather a result of the underlying data and algorithms used (Chen, 2023). Therefore, it is essential to continuously evaluate and improve AI systems to minimize the impact of biases and promote fairness and equality.

The use of AI systems in the library and information industry has the potential to perpetuate biases and discrimination, hindering the goal of providing equitable access to information. Biases can be introduced through flawed training data, biased decision-making processes and lack of transparency in AI algorithms. Without addressing these issues, AI systems can reinforce existing inequalities and limit the inclusivity of library services. Thus, there is a need to promote ethical AI practices and prioritize education and awareness to mitigate biases and ensure fairness, transparency and accountability in AI technologies within the library and information industry. In this viewpoint article, we delve into the importance of addressing these issues and advocate for the adoption of ethical AI practices within the library and information industry.

Understanding the impact

Bias and discrimination in AI systems within the library and information industry can have profound implications for users and communities. Libraries have long been champions of equal access to information and knowledge, but the introduction of biased AI algorithms can undermine these principles (Chen, 2023; Raso *et al.*, 2018).

One significant impact of biased AI is the potential exclusion of marginalized communities (Kim and Bodie, 2021). Libraries serve diverse populations, and it is crucial that AI systems do not perpetuate existing societal biases. For example, if a recommendation algorithm only suggests materials from a narrow range of perspectives or fails to recognize and cater to the unique needs of certain user groups, it can inadvertently reinforce existing disparities and limit access to information for those who need it most.

Moreover, biased AI systems can contribute to the perpetuation of stereotypes and discrimination (Ferrara, 2023). If algorithms are trained on biased data sets that reflect societal prejudices, they can amplify and perpetuate these biases in their recommendations, search results or decision-making processes (Akter *et al.*, 2021; Chen, 2023; Ferrara, 2023). This can lead to harmful outcomes, such as reinforcing gender stereotypes or perpetuating racial and ethnic biases.

In addition, the trust and credibility of libraries can be eroded if users perceive bias or discrimination in AI systems (Miao, 2019). Users rely on libraries to provide unbiased and objective information, and any indication of bias in AI algorithms can undermine this trust. It is crucial for libraries to ensure that their AI systems are transparent, fair and accountable to maintain the integrity of their services (Bubinger and Dinneen, 2021).

Addressing the impact of bias and discrimination in AI requires a multifaceted approach (Raso *et al.*, 2018; Bubinger and

  0741-9058, DOI 10.1108/LHTN-10-2023-0182

Dinneen, 2021). Libraries must actively engage in ongoing evaluation and monitoring of AI systems to identify and rectify biases (Bubinger and Dinneen, 2021). Collaboration with diverse user communities, including marginalized groups, can provide valuable insights and feedback to improve AI technologies and ensure inclusivity.

Furthermore, libraries have a responsibility to advocate for ethical AI practices and engage in critical discussions about the impact of AI on society (Chen, 2023; Ferrara, 2023; Bubinger and Dinneen, 2021). By fostering awareness and understanding among library staff, users and stakeholders, libraries can promote a culture of accountability and actively work toward mitigating bias and discrimination in AI systems.

Unveiling the root causes

To effectively combat bias and discrimination in AI systems within the library and information industry, it is crucial to identify and address the root causes. By understanding these underlying factors, libraries can implement strategies to mitigate bias and ensure fairness in their AI technologies.

One primary cause of biased AI algorithms is flawed or biased training data (Akter *et al.*, 2021; Zuiderveen Borgesius, 2018). AI systems learn from historical data, and if that data reflects societal biases, the algorithms can unintentionally perpetuate those biases (Akter *et al.*, 2021; Zuiderveen Borgesius, 2018; Ferrara, 2023; Bubinger and Dinneen, 2021). For example, if a library's recommendation algorithm is trained on a data set that predominantly includes works from certain authors or perspectives, it may lead to biased recommendations that overlook diverse voices and perspectives. Therefore, it is essential to critically examine and diversify the training data to ensure it accurately represents the diverse user base libraries serve.

Another root cause of bias in AI is the decision-making process during the development of AI systems (Varona and Suárez, 2022). Biases can be inadvertently introduced through the choices made by developers, such as the selection of features or the formulation of decision rules (Ferrer *et al.*, 2021; Ferrara, 2023; Varona and Suárez, 2022). It is crucial for libraries to promote diversity within AI development teams to challenge and address these biases. A diverse team can bring different perspectives and experiences to the table, helping to uncover and rectify implicit biases that might otherwise go unnoticed.

Moreover, the lack of transparency in AI algorithms can contribute to biased outcomes (Daneshjou *et al.*, 2021). Many AI systems operate as black boxes, making it challenging to understand how they arrive at their decisions (Daneshjou *et al.*, 2021; Lopez and Garza, 2023). This opacity makes it difficult to identify and rectify biases. Libraries should prioritize transparency and explainability in their AI systems, ensuring that users have insight into how decisions are made. By providing transparency, libraries can foster trust and enable users to hold AI systems accountable for their actions.

Regular audits and evaluations of AI algorithms are essential to identify and rectify biases. Libraries should establish mechanisms to continuously monitor and assess the performance of their AI systems. This includes analyzing outcomes, reviewing user feedback and conducting regular audits to ensure fairness and accountability. By regularly assessing AI systems, libraries can proactively identify and address biases, making necessary adjustments to improve the accuracy and inclusivity of their services.

Promoting ethical artificial intelligence practices

To effectively address bias and discrimination in AI systems within the library and information industry, it is crucial to promote and adopt ethical AI practices (Bubinger and Dinneen, 2021; Roberts *et al.*, 2023). By integrating ethical considerations into the development and deployment of AI technologies, libraries can ensure fairness, transparency and accountability (Bubinger and Dinneen, 2021). Here are key strategies for promoting ethical AI practices:

- *Transparency and explainability*: Libraries should prioritize transparency in their AI systems. Users should have a clear understanding of how AI algorithms make decisions, what data is used and how biases are mitigated. Providing explanations for AI-generated recommendations or search results can empower users and foster trust in the technology.
- *Regular audits and evaluations*: Libraries must establish mechanisms to regularly audit and evaluate their AI systems. This involves continuously monitoring the performance and outcomes of AI algorithms, analyzing user feedback and conducting external assessments. By proactively identifying biases and addressing them, libraries can ensure that their AI systems are fair and provide equitable access to information.
- *Inclusive data collection and representation*: Libraries should critically examine the training data used for AI systems to ensure it is diverse, representative and free from biases. This includes considering factors such as race, gender, age and cultural background to avoid perpetuating discriminatory patterns. Collaborating with user communities and stakeholders can provide valuable insights to ensure that AI systems cater to the needs of diverse populations.
- *Collaboration and user engagement*: Libraries should actively engage with user communities and stakeholders to understand their needs, concerns and expectations regarding AI technologies. Involving users in the development and evaluation of AI systems can provide valuable perspectives and help identify potential biases or areas for improvement.
- *Ethical guidelines and policies*: Libraries should develop and adhere to clear ethical guidelines and policies for the development, deployment and use of AI technologies. These guidelines should address issues such as bias mitigation, privacy protection and the responsible use of AI. By establishing clear ethical standards, libraries can ensure that

AI systems align with their mission and values.

- *Continuous learning and training*: Libraries should invest in ongoing education and training for staff involved in AI development and implementation. This includes raising awareness about the potential biases in AI systems, providing training on ethical considerations and fostering a culture of critical engagement with AI technologies.
- *Collaboration with the AI community*: Libraries should actively collaborate with the wider AI community to share best practices, research findings and ethical frameworks. Engaging in discussions and collaborations with AI researchers, developers and policymakers can help libraries stay informed about emerging ethical issues and contribute to the development of industry-wide standards.

Education and awareness

Education and awareness play a crucial role in addressing bias and discrimination in AI systems within the library and information industry (Soleymani *et al*., 2023; Saeidnia, 2023). By promoting knowledge and understanding among library staff, users and stakeholders, libraries can foster a culture of critical engagement with AI technologies and empower individuals to advocate for ethical and inclusive practices (Lebovitz *et al*., 2022). Here are key aspects of education and awareness:

- *Staff training*: Libraries should provide comprehensive training programs for staff members to enhance their understanding of AI technologies, their potential biases and the ethical considerations involved. Staff should be equipped with the knowledge and skills to critically evaluate AI systems, identify biases and take appropriate actions to mitigate them. Training should also focus on promoting awareness of the impact of biased AI on diverse user populations.
- *User education*: Libraries should actively educate users about the presence of AI technologies in their services and the potential biases associated with them. This can be done through workshops, informational materials and online resources that explain how AI algorithms work, their limitations and steps taken to mitigate biases. Users should be empowered to critically evaluate AI-generated recommendations and search results, and to provide feedback on their experiences.
- *Community engagement*: Libraries should engage with their user communities and stakeholders to raise awareness about the impact of biased AI systems. This can involve hosting community discussions, inviting guest speakers or organizing events that highlight the ethical implications of AI technologies. By involving diverse perspectives, libraries can better understand the concerns and needs of their communities and work toward more inclusive AI systems.
- *Collaboration with educational institutions*: Libraries can collaborate with educational institutions, such as universities or schools, to incorporate AI ethics and bias awareness into curricula. By working together, libraries and educational institutions can ensure that future professionals are well-prepared to address bias and discrimination in AI systems.
- *Sharing best practices*: Libraries should actively share best practices and case studies related to addressing bias in AI systems. This can be done through conferences, workshops or online platforms where libraries can exchange ideas, experiences and strategies for promoting ethical AI practices. By sharing knowledge and lessons learned, libraries can collectively work toward creating more inclusive and fair AI systems.
- *Advocacy and policy engagement*: Libraries should actively engage in advocacy efforts to promote ethical AI practices at a broader level. This can involve participating in policy discussions, collaborating with other organizations and contributing to the development of regulations and guidelines that ensure fairness and transparency in AI technologies.

Conclusion

Addressing bias and discrimination in AI systems within the library and information industry requires a multifaceted approach. By unveiling the root causes, promoting ethical AI practices and prioritizing education and awareness, libraries can work toward creating more inclusive and fair AI systems. Understanding the root causes of bias, such as flawed training data, biased decision-making processes and lack of transparency, is essential. Libraries should critically examine and diversify training data, promote diversity within AI development teams, prioritize transparency and explainability and conduct regular audits and evaluations to identify and rectify biases. Promoting ethical AI practices involves integrating fairness, transparency and accountability into the development and deployment of AI technologies. Libraries should prioritize transparency and explainability, regularly audit and evaluate AI systems, collect and represent inclusive data, collaborate and engage with users and stakeholders, establish ethical guidelines and policies and invest in continuous learning and training for staff. Education and awareness are vital in fostering a culture of critical engagement with AI technologies. Libraries should provide staff training to enhance understanding of AI technologies and biases, educate users about the presence and potential biases of AI systems, engage with communities to raise awareness, collaborate with educational institutions to incorporate AI ethics into curricula, share best practices and engage in advocacy and policy discussions. By implementing these strategies, libraries can lead the way in ensuring that AI systems within the library and information industry are fair, inclusive and accountable. By prioritizing fairness, transparency and ethical considerations, libraries can harness the power of AI to enhance services, improve access and promote equitable and unbiased access to information for all users.

REFERENCES

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y.K., D'Ambra, J. and Ning Shen, K. (2021), "Algorithmic bias in data-driven innovation in the age of AI", *International Journal of Information Management*, Vol. 60, p. 102387.

Bubinger, H. and Dinneen, J.D. (2021), "Actionable approaches to promote ethical AI in libraries", *Proceedings of the Association for Information Science and Technology*, Vol. 58 No. 1, pp. 682-684.

Chen, Z. (2023), "Ethics and discrimination in artificial intelligence-enabled recruitment practices", *Humanities and Social Sciences Communications*, Vol. 10 No. 1, p. 567, doi: 10.1057/s41599-023-02079-x.

Daneshjou, R., Smith, M.P., Sun, M.D., Rotemberg, V. and Zou, J. (2021), "Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review", *JAMA Dermatology*, Vol. 157 No. 11, pp. 1362-1369.

Ferrara, E. (2023), "Fairness and bias in artificial intelligence: a brief survey of sources, impacts, and mitigation strategies", arXiv preprint arXiv:230407683.

Ferrer, X., van Nuenen, T., Such, J.M., Coté, M. and Criado, N. (2021), "Bias and discrimination in AI: a cross-disciplinary perspective", *IEEE Technology and Society Magazine*, Vol. 40 No. 2, pp. 72-80.

Kim, P.T. and Bodie, M.T. (2021), "Artificial intelligence and the challenges of workplace discrimination and privacy", *Journal of Labor and Employment Law*, Vol. 35 No. 2, pp. 289-315.

Lebovitz, S., Lifshitz-Assaf, H. and Levina, N. (2022), "To engage or not to engage with AI for critical judgments: how professionals deal with opacity when using AI for medical diagnosis", *Organization Science*, Vol. 33 No. 1, pp. 126-148.

Lopez, A. and Garza, R. (2023), "Consumer bias against evaluations received by artificial intelligence: the mediation effect of lack of transparency anxiety", *Journal of Research in Interactive Marketing*.

Malek, M.A. (2022), "Criminal courts' artificial intelligence: the way it reinforces bias and discrimination", *AI and Ethics*, Vol. 2 No. 1, pp. 233-245.

Miao, Z. (2019), "Investigation on human rights ethics in artificial intelligence researches with library literature analysis method", *The Electronic Library*, Vol. 37 No. 5, pp. 914-926.

Raso, F.A., Hilligoss, H., Krishnamurthy, V., Bavitz, C. and Kim, L. (2018), "Artificial intelligence & human rights: opportunities & risks", Berkman Klein Center Research Publication, (2018-6).

Roberts, H., Cowls, J., Hine, E., Morley, J., Wang, V., Taddeo, M., et al. (2023), "Governing artificial intelligence in China and the European Union: comparing aims and promoting ethical outcomes", *The Information Society*, Vol. 39 No. 2, pp. 79-97.

Saeidnia, H. (2023), "Using ChatGPT as a digital/smart reference robot: how may ChatGPT impact digital reference services?", *Information Matters*, Vol. 2 No. 5.

Soleymani, H., Saeidnia, H.R., Ausloos, M. and Hassanzadeh, M. (2023), "Selective dissemination of information (SDI) in the age of artificial intelligence (AI)", *Library Hi Tech News*, Vol. ahead-of-print No. ahead-of-print.

Varona, D. and Suárez, J.L. (2022), "Discrimination, bias, fairness, and trustworthy AI", *Applied Sciences*, Vol. 12 No. 12, p. 5826.

Zuiderveen Borgesius, F. (2018), "Discrimination, artificial intelligence, and algorithmic decision-making. Línea]", Council of Europe.

Corresponding author
Hamid Reza Saeidnia can be contacted at: hamidrezasaeidnia@gmail.com

Hamid Reza Saeidnia (*hamidrezasaednia@gmail.com*) is based at the Department of Knowledge and Information Science, Tarbiat Modares University, Tehran, Iran.