

Historical Bias in Artificial Intelligence

To what extent does artificial intelligence's bias in historical representation affect public awareness?

2177

Historical Bias in Artificial Intelligence

History helps people determine their identity and shape their views on the world. So artificial intelligence's ability to present history, impacts people's awareness of historical events. In article "AI in history" Matthew Jones, a Professor at Princeton University, states that artificial intelligence doesn't exactly create bias in history but amplifies biases already present in historical data, resulting in the spread of distorted information (Jones, 2023). If history is told inaccurately, it changes the way people view past events. Johan Norberg, a senior at a public policy research organization, explains how people often view the past better than the present. This view of "false nostalgia" is used by politicians to influence public ideals (Norberg, 2022). AI replicates this by repeating and spreading false, idealized versions of history. Additionally, writing award winner Judith Binney emphasizes how history is usually told by those who are in power, while other minority viewpoints are ignored (Binney, 2024). Similarly, AI follows this pattern with its limited access to certain input information. This way it could continue telling history from a one-sided perspective, leaving out other important ones. When history is changed or presented unequally, people are not given the whole truth about the past. And as AI repeats this mistake to a greater extent, then this misinformation will shape public beliefs, creating misunderstandings in society. This prompts the question: To what extent does artificial intelligence's bias in historical representation affect public awareness? Although Artificial intelligence can present history in a neutral perspective, AI's bias in historical representation affects public awareness to a greater extent by promoting one-sided information and shaping public perception. To combat the bias, it should be required to put on labels or disclaimers on AI-produced historical content to inform viewers about potential biases. This reduces the effect of AI spreading misinformation with its biased and one-sided history.

The spread of historical bias is a significant issue because it reinforces misinformation, which ends up influencing societal behaviors. In the peer-reviewed journal article “Managing Bias in AI”, it states that AI is biased because it presents historical data that already contains historical bias, making it difficult to generate neutral or fully accurate historical narratives (Roselli et al., 2019). The article emphasizes that since AI retrieves information from existing historical records it inherits their limitations, and biases. This is significant because people use AI as a source to retrieve historical information, which means these inaccuracies reach a wide audience and become accepted as the truth. Similarly, executive strategist, Tamella Coval discusses how a Facebook AI chatbot learned offensive language from users which led it to post harmful messages on the platform, leading it to eventually being shut down by Microsoft (Coval, 2018). Coval acknowledges that AI systems learn from biased human inputs, which can spread outdated harmful ideas. This reinforces false and negative beliefs, shaping how people behave and think. Overall, these articles demonstrate how AI can spread historical bias and harmful misinformation. This emphasizes the need to fix these biases, so that public perception isn’t influenced by inaccurate or one-sided information.

AI promotes one-sided information

AI driven historical bias promotes one-sided information due to reliance on the data inputted into its systems. A United Nations report on racism and AI explains how artificial intelligence reflects existing biases in historical data, which result in the buildup of racial and cultural stereotypes, influencing public knowledge of history (*Racism and AI*, 2024). This demonstrates how AI’s biased versions of history influence public memory and how people perceive the past. In addition, recently schools have started to accept the use of artificial intelligence in classrooms. But artificial intelligence’s tendency of producing biased history in

educational environments, result in students receiving one-sided education and an incomplete view of history. For example, Senior Research Manager at Climate Policy Initiative, Leme Lopes discusses how AI-driven historiography can unintentionally favor dominant narratives due to its reliance on biased datasets, affecting how history is taught and learned (Lopes, 2023). From an educational lens, this shows how the limited and biased historical data present in AI systems reinforces dominant narratives and limits students' ability to think critically about history. This is also seen in Judith Binney's emphasis on how history currently is being shaped mainly by dominant narratives. Causing minority perspectives of historical events to be silenced (Binney, 2024). The social and educational views of AI presenting one sided history show that its biased historical data reinforces unfair stereotypes and limits our understanding of the past. Fixing inherited biases in AI is essential for a more balanced view of history.

AI shapes public perception

Historical bias in AI systems shapes public perception by selectively amplifying certain historical narratives while downplaying others. Sangmin-Michelle Lee, a Ph.D. Professor in Educational Technology, discusses how AI bias affects education by shaping how information is presented to students, especially in language learning. He states that AI bias in foreign language education can arise due to demographic and cultural diversity in student populations. For example, AI language tools might be trained mostly on Western English speakers, causing them to misunderstand or unfairly evaluate non-Western accents (Junseong & Lee, 2024). This demonstrates, that just like how AI might favor Western English speakers in language learning; in historical analysis AI could be trained mostly on Western historical sources, leading it to prioritize Western perspectives over others in historical events. Correspondingly, A peer reviewed journal article "A synthesis of AI in higher education: Shaping the future." describes

how AI is changing higher education by affecting research, teaching, and administration. It highlights how AI learns from historical datasets, which reflect inherent gender and racial biases, as well as societal prejudices. AI algorithms then amplify biases in education, affecting student evaluations, admissions decisions, and learning opportunities (Jacques et al., 2024). From an equality lens this reveals that since AI systems inherit biases from historical data, it reinforces existing gender and racial inequalities in education. Consequently, decisions such as student evaluations and admissions can become skewed, limiting fair access and opportunities for marginalized groups. In summary, the educational, social, and equality lens show how the embedded historical bias in artificial intelligence shapes the understanding of history by amplifying dominant narratives and sustaining unfair stereotypes, leading to a skewed incomplete view of the past.

AI is capable of presenting history without human bias

Artificial intelligence can present history in a neutral, unbiased way by relying on factual data rather than human interpretation. The article by the Australian Human Rights Commission, a trusted government organization focused on fairness and equality, states that if AI implements more appropriate data and train it to incorporate fairness into its decisions then AI can eliminate biases present in existing gender pay gap historical data (*Historical Bias in AI Systems*, 2020). Similarly, Matthew Jones discusses the importance of understanding the historical development of artificial intelligence, particularly its data-driven nature and its embedded biases, for historians to effectively utilize and critically engage with AI tools in their research (Jones, 2023). These articles emphasize AI's capability to present historical data better than how it is currently being taught. History is being taught differently around the world due to different perspectives and views of historical text and events. As seen in the magazine article "False Nostalgia", people

tend to have an inaccurate or biased memory of their past, causing them to fantasize about how good the past was compared to their present (Norberg, 2022). This feeling of false nostalgia is a contributing factor to the different views of history present today. However, while these pieces of evidence show how AI can provide a neutral perspective and eliminate existing biases in historical data, these adaptations are yet to be implemented into AI systems, making them currently biased, and an unreliable source of information. For example, the Australian Human Rights Commission also states that there are current privacy regulations that must be cleared before more accurate data can be passed into artificial intelligence systems (*Historical Bias in AI Systems*, 2020). This article demonstrates the limitations AI faces before it can become an appropriate tool used for history. Although some argue that AI's reliance on massive, factual data sets could enable it to potentially correct embedded historical biases, the reality is that current systems still fall short. Ultimately, until these challenges are fully addressed, AI cannot yet be trusted to offer entirely objective historical accounts.

Input accurate data and train AI systems to be more accurate

Issues with current existing solutions for AI bias are that their implementation must go through various regulations and reduce AI accuracy, making them impractical to use. Some solutions to AI's historical bias issue are to input recent and more accurate data into AI systems, and to train AI systems to be fairer in making decisions. For example, the article "Historical Bias in AI systems" states some mitigation action that can be taken to reduce unfairness in AI systems. Some of which are to gather more appropriate data, and to adjust AI systems to incorporate fairness into its decision-making skills. However, these proposed solutions are inadequate due to the possible privacy violations and reduction in AI accuracy. Stating that when gathering more appropriate data for AI systems, this may cause violations of people's data

privacy. And while trying to train AI systems to be more accurate this may cause it to create false positives in AI algorithms, reducing the system's accuracy (*Historical Bias in AI Systems*, 2020). This article demonstrates how current solutions aim to reduce AI bias by updating data and promoting fairness, but end up falling short due to strict regulations, privacy risks, and potential losses in accuracy.

Require disclaimers on AI generated content and train with more diverse data sets

Some solutions that can be implemented to reduce AI's historical bias in affecting public awareness are to require clear labels or disclaimers on AI-produced historical content, and to train AI with a more diverse set of historical data. By requiring clear labels or disclaimers on AI-produced historical content, it informs viewers about potential biases. The article "Managing Bias in AI", states that given the risks, it is essential to create clear guidelines to help AI developers avoid bias. A key step in this process is establishing a framework to identify, measure, and reduce sources of bias (Roselli et al., 2019). This shows how the solution helps reduce the effect of AI's historical bias on public awareness by informing the public of the possible bias present in AI systems. This reduces the effect of AI spreading misinformation with its biased and one-sided versions of history. The second solution is to train AI with a more diverse set of historical data to help reduce its one-sided representation of history. The United Nations report states that AI systems can reinforce biases when trained on past arrest data, as systemic racism in the criminal justice system distorts their predictions (*Racism and AI*, 2024). This demonstrates that training AI with a more diverse set of historical data will help ensure that AI-generated historical narratives are more accurate, balanced, and reflective of different viewpoints. Implementing clear labels on AI history content and training AI with diverse historical data will help reduce bias, minimizing AI's historical bias impact on public awareness.

Require labels and disclaimers on AI generated historical content

Out of the two proposed solutions, requiring labels or disclaimers on AI-produced historical content is the best solution to implement. This solution reduces the effect of AI's historical bias shaping public awareness and perspectives of history by informing the public of the bias present in AI systems. Positive implications of this solution are that it will increase viewers awareness, encourage users to question and verify AI-generated historical narratives, and reduce the spread of misinformation. Some negative implications are that it reduces people's trust in AI and causes viewers to have over dependence on labels and assume labeled content as fully accurate without further scrutiny. The article "Managing Bias in AI" also states the importance of transparency of input data, stating that the only way to verify data is accurate and does not contain protected or incorrect data (Roselli et al, 2019). This demonstrates the importance of having data transparency with AI users to ensure accurate data is being presented and used. Some limitations of this solution are the implementation challenges in standardizing and enforcing labels across AI platforms, and the bias in labeling AI platforms as it could be influenced by subjective or political bases. Despite these challenges, implementing this solution is still plausible, as transparency measures can help reduce the spread of misinformation and promote more balanced historical narratives.

Conclusion

Artificial intelligence's bias in historical representation affects public awareness to a greater extent because it spreads historical bias by amplifying existing misinformation, shaping public perception, and reinforcing one-sided narratives. These bias's impacts education, society, and equality, leading to distorted understandings of history. Requiring labels on AI-generated historical content and training AI with diverse historical data can help reduce bias. While

challenges exist, such as implication regulations and influences on labeling, these solutions promote transparency and decrease historical bias in AI.

References

- Binney, J. (2024). Maori oral narratives, pakeha written texts: Two forms of telling history. *New Zealand Journal of History*, 38(2), 203–214. Project MUSE.
<https://muse.jhu.edu/pub/426/article/878294/pdf>
- Coval, T. (2018). Artificial intelligence, bias & your business: Think you have no place for the “mixed (virtual, augmented and gamification)” realities within your business offerings? Think again. *Journal of Property Management*, 83(2), Gale Power Search.
<https://link.gale.com/apps/doc/A534565611/GPS?u=napl44696&sid=bookmark-GPS&xid=fcc9038a>
- Historical bias in AI systems*. (2020, November 24). Australian Human Rights Commission.
<https://humanrights.gov.au/about/news/media-releases/historical-bias-ai-systems>
- Jacques, P. H., Moss, H. K., & Garger, J. (2024). A synthesis of AI in higher education: Shaping the future. *Journal of Behavioral & Applied Management*, 24(2), 103–111. Ebsco Host.
<https://doi.org/10.21818/001c.122146>
- Jones, M. (2023). AI in history. *American Historical Review*, 128(3), 1360–1367. Ebsco Host.
<https://doi.org/10.1093/ahr/rhad361>
- Junseong, B., & Lee, S.-M. (2024). Addressing bias for equity in ai-driven customized learning systems. *Multimedia-Assisted Language Learning*, 27(4), 70–86. Ebsco Host.
<https://doi.org/10.15702/mall.2024.27.4.70>
- Lopes, L. (2023). Artificial history? Inquiring chatgpt on historiography. *Rethinking History*, 27(4), 709–749. Ebsco Host. <https://doi.org/10.1080/13642529.2023.2234227>
- Norberg, J. (2022). False nostalgia. *Reason.com*. <https://reason.com/2021/12/05/false-nostalgia/>

Racism and AI: “Bias from the past leads to bias in the future.” (2024, July 30). United Nations.

<https://www.ohchr.org/en/stories/2024/07/racism-and-ai-bias-past-leads-bias-future>

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. *ACM Digital Library*, 539

–544. Google Scholar. <https://doi.org/10.1145/3308560.3317590>