

## **Report on Algorithmic Bias**

### **Reference Information**

**Title:** Algorithms help people see and correct their biases, study shows

**Author:** Carey K. Morewedge

**Published by:** The Conversation

**Publication Date:** May 10, 2024

**Link:** [The Conversation Article](#)

### **Summary of the Article**

The article, authored by Carey K. Morewedge and published in The Conversation, discusses a study exploring the dual role of algorithms in both reflecting and potentially correcting human biases. The study was conducted by Morewedge and his colleagues, Begum Celikitan and Romain Cadario, and was published in the Proceedings of the National Academy of Sciences. The research focused on how algorithms, often trained on human decisions, not only inherit these biases but also can make these biases more visible to individuals.

The study involved nine experiments where participants rated Uber drivers or Airbnb listings based on various attributes like driving skill and trustworthiness. Participants were provided with relevant details alongside irrelevant, biasing information such as the age, gender, and attractiveness of the drivers or hosts. After making their ratings, participants were shown summaries of either their own ratings or those generated by an algorithm trained on their ratings. They were then asked to assess the influence of biasing factors on these ratings.

The findings revealed that participants were more likely to recognize biases in the algorithm's decisions than in their own. This "algorithmic mirror effect" persisted even when participants were told that their ratings were produced by an algorithm. The study concluded that algorithms could serve as tools to help people recognize and correct their biases, as participants were more inclined to adjust ratings they believed were made by algorithms.

### **Ethically Relevant Issues**

The primary ethical issue raised by the article is the role of algorithms in perpetuating versus correcting human biases. On one hand, algorithms, when trained on biased human data, can perpetuate and even amplify biases related to race, gender, and socioeconomic status. This phenomenon is well-documented in areas like judicial sentencing, hiring practices, and urban development, where algorithmic decisions have shown biases against marginalized groups.

On the other hand, the article highlights a potentially beneficial use of algorithms: serving as mirrors that reflect human biases back to individuals, thereby making these biases more recognizable and correctable. This duality presents an ethical dilemma. Can and should algorithms be trusted to both identify and correct biases, and under what conditions? Is it

ethical to use biased algorithms as a tool for self-improvement, or does this perpetuate a reliance on inherently flawed systems?

### **Assessment of the Handling of Issues**

In assessing whether the issues discussed in the article are being handled properly, it is essential to apply an ethical framework. One appropriate framework is the principle of beneficence from bioethics, which emphasizes actions that promote well-being and prevent harm.

From this perspective, the study's findings offer a promising approach to mitigating bias through awareness and correction. By leveraging algorithms to expose biases, the research aligns with the principle of beneficence, as it aims to improve decision-making and reduce discriminatory practices. However, the ethical implementation of this approach requires several additional steps:

1. **Transparency and Accountability:** There must be transparency in how algorithms are trained, the data used, and the potential biases embedded in them. Users should be informed about the biases present and how they may affect outcomes.
2. **Continuous Monitoring and Adjustment:** Algorithms should not be static; they need continuous monitoring and adjustment to minimize biases over time. This requires robust mechanisms for feedback and improvement.
3. **Education and Training:** Users must be educated on the limitations and potential biases of algorithms. Training programs should be implemented to help individuals understand how to interpret and act on the bias reflections provided by algorithms.
4. **Diverse and Representative Data:** Ensuring that the data used to train algorithms is diverse and representative of all population segments can help mitigate the initial biases that algorithms might perpetuate.
5. **Ethical Oversight:** Establishing independent ethical oversight committees can help ensure that the use of algorithms adheres to ethical standards and promotes fairness.

### **Conclusion**

The study discussed in the article sheds light on a novel approach to addressing algorithmic bias by using algorithms as reflective tools. While this method shows promise in helping individuals recognize and correct their biases, its ethical implementation requires a multifaceted approach involving transparency, continuous monitoring, education, representative data, and ethical oversight. By adhering to the principle of beneficence and implementing these additional steps, the potential harms of algorithmic bias can be mitigated, promoting fairer and more equitable decision-making processes.