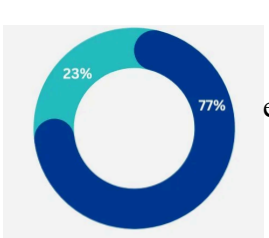
A new era: the intersection of AI and psychology - How do biases in AI form, how can they be mitigated, and what are the implications on users?

By Rishabh Modi

Why does this matter?



Working Professionals



1. 77 % of companies are either using or exploring the use of AI[1].
2. 60% of business owners think AI will increase their productivity[1].







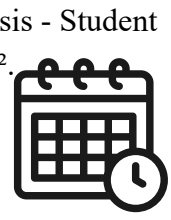
Students



86% of students globally are regularly using AI in their studies, with 54% of them using AI on a weekly basis - Student Survey[2].







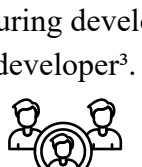
Developing AI Models

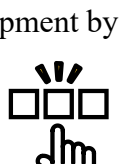


Systems Bias

Data is selected, filtered, or weighted during development by developer[3].

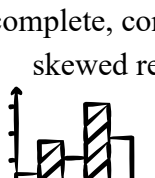


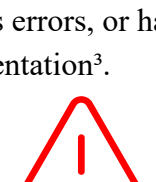




Input Bias

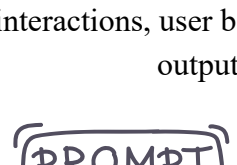
Data used to train the AI model is incomplete, contains errors, or has skewed representation[3].

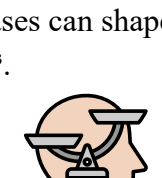




Application Bias

When models learn from interactions, user biases can shape outputs[3].



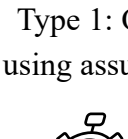


Making Decisions[4]

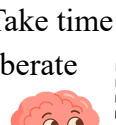


Type 1: Quickly using assumptions

Type 2: Take time to deliberate



Enforce Bias



How do we form bias?

Psychological Perspectives[4]



1. Behaviouralists → systematic errors in rational reasoning

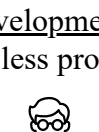


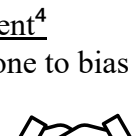
2. Evolutionary theorists → decisions based on context

Bias Mitigation and AI Development[4]

Experts less prone to bias

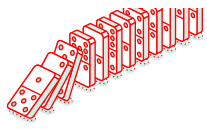






AI developers need experts input to build more accurate models.

Impact of AI Models on People

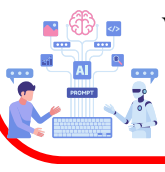


AI Impact on Users Bias

People who repeatedly interacted with biased AI systems learned to be more biased themselves[5].



AI models exploit minor biases to learn how to better respond and leads to interactions between humans and AI exponentially increasing biased views across our society[5].





Perceived Fairness and Trust in AI Models

Another study found that when some people interact with a biased model, they perceive it as less fair and are less likely to use it[6].



Our Ability to Recognize Bias and the Implications.

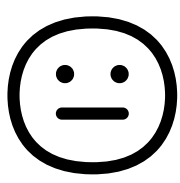
Participants often claimed they were less biased than the "average American[7]."

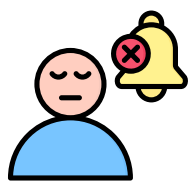




This study also found that knowledge of particular biases in human judgment and the ability to recognize the impacts of that bias on other people neither prevents one from succumbing nor makes one aware of having done so[7].







This information indicates that most people will continue to interact with biased AI systems since they can't accurately determine if it is reliable[7].

Concluding Throughts



As social media becomes more prominent, people are increasingly exposed to content tailored by AI algorithms, making it easier to form biases. Greater reliance on AI for studying and work has also accelerated the spread of biased views. Be mindful of what you consume.







"Up to 95% of youth ages 13–17 report using a social media platform[8]."



Another issue to highlight is WEIRD research in psychology. Many groups are underrepresented, emphasizing the need for more diverse and inclusive research samples when training AI models.

<u>Answer to research question:</u> AI biases emerge from cognitive biases embedded in data, design, and human–AI interactions, can be mitigated through diverse sampling and expert input, and ultimately influence user perception, trust, and decision-making.

Works Cited

1. National University. (2024, March 1). 131 AI Statistics and Trends (2024). National University; National University. https://www.nu.edu/blog/ai-statistics-trends/
2. Digital Education Council. (2024, August 7). What Students Want: Key Results from DEC Global AI Student Survey 2024. Digitaleducationcouncil.com. https://www.digitaleducationcouncil.com/post/what-students-want-key-results-from-dec-global-ai-student-survey-2024
3. Hofmann, B. (2025). Biases in AI: acknowledging and addressing the inevitable ethical issues. Frontiers in Digital Health, 7. https://doi.org/10.3389/fdgth.2025.1614105
4. Toet, A., Brouwer, A.-M., van den Bosch, K., & Korteling, J. E. (2016). Effects of personal characteristics on susceptibility to decision bias: A literature study. International Journal of Humanities and Social Sciences, 8(5), 1–17.
5. Glickman, M., & Tali Sharot. (2024). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, *9*. https://doi.org/10.1038/s41562-024-02077-2
6. Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221115189
7. Pronin, E., Lin, D. Y., & Ross, L. (2002). The Bias Blind Spot: Perceptions of Bias in Self Versus Others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008
8. National Library of Medicine. (2023). Social Media and Youth Mental Health. Www.ncbi.nlm.nih.gov; US Department of Health and Human Services. https://www.ncbi.nlm.nih.gov/books/NBK594759/