AI & SOCIETY
https://doi.org/10.1007/s00146-025-02258-1

CURMUDGEON CORNER

Why AI image generators cannot afford to be blind to racial bias

Mustafa Arif[1] · **Yoshiyasu Takefuji**[1]

Received: 21 October 2024 / Accepted: 17 February 2025
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2025

AI image generators are shaping a new visual reality—one where words turn into pictures with an ease that would have seemed like magic just a few years ago. But beneath the surface of these technological marvels, there is a serious problem we cannot ignore: racial bias. And the issue is not just about bad training data or flawed algorithms. It runs much deeper, reflecting the biases embedded in the very structures of society.

For all the talk of AI as a neutral tool, we know better. As Michel Foucault might put it, technology is never just technology—it is an instrument of power, shaping what we see, how we categorize people, and who gets represented in certain ways. AI-generated images do not just reflect reality; they shape it.

1 Bias isn't just a bug—it is the system at work

Some might argue that the problem of racial bias in AI is just a glitch—something that can be fixed with better data or more careful programming. But that assumption ignores a fundamental truth: AI is trained on our world and our world is not fair. Consider the findings of Lin and colleagues (6), who tested an AI image generator to see if it mirrored racial and gender demographics in medical residency programs. Surprisingly, it did—but that does not necessarily mean the system is bias-free. Did the AI reflect real-world equity Bail (2024), or was it simply aligning with historical trends that already favor certain groups? If it is the latter, that is not fairness—it is just reinforcing the status quo.

And what about applications beyond medicine? Some researchers, like Tortora (10) and others Chen and , Joshi et al. (2024), Patel et al. (2024), suggest that AI-generated images could actually reduce bias in police lineups. That sounds promising, but it comes with a major caveat: if the AI is trained on biased datasets, won't it just recreate the same old racial profiling issues in a more sophisticated disguise? The potential for unintended consequences is huge.

2 AI as a tool of control

Gilles Deleuze described modern society as one of continuous control—a world where power is not just about institutions but about constant surveillance, categorization, and subtle influence. AI fits neatly into this framework. It does not just reinforce biases; it operationalizes them. Park's (7) research on AI's role in spreading misinformation is a perfect example. When generative models help create misleading images Templin et al. (2024)—whether of political figures, crime suspects, or "typical" professionals—they do not just depict reality; they shape our collective perception of it Joshi et al. (2024). And when biases creep into that process, the consequences go far beyond a few flawed pictures.

Even when AI is supposed to be making things "fairer," it often does the opposite. Bell and colleagues (2), for instance, found that AI-generated images of suspects were fairer than traditional police lineup photos. But how do we measure fairness? What happens when subtle aesthetic choices in AI-generated faces—lighting, expressions, angles—affect how we perceive people? In criminal justice, even a slight shift in perception can be the difference between guilt and innocence.

3 So what do we do?

Here is the reality: bias in AI image generation is not going to vanish overnight. But that does not mean we should accept it as an inevitable byproduct of technology. We have options.

Demand transparency: AI developers need to open the black box. Who decides what data gets used to train these models? How are outputs audited for bias? If companies

✉ Mustafa Arif
arifmustafa75@gmail.com

[1] Musashino University, Tokyo, Japan

Published online: 13 March 2025

want to build world-altering tools, they should be willing to show us how they work.

Move beyond "Fixing" bias: It is not enough to tweak algorithms and hope for the best. As Fang's (4) work on large language models shows, even when AI companies try to mitigate bias, it still seeps through in unexpected ways. Instead of just fixing bias when it appears, we need to ask: should AI be generating these images in the first place?

Rethink AI's role in society: Foucault reminds us that systems of power are often disguised as systems of efficiency. AI in policing, hiring, and media is not just about making things faster—it is about who gets to define reality. If AI is amplifying racial biases rather than dismantling them, maybe the real question is not "How do we fix it?" but "Should we be using it at all?"

The bottom line is this: AI-generated images are already shaping our world. If we do not confront their biases now, we will be living in a future where machines reinforce and amplify social inequalities instead of challenging them. This is not just about coding better AI. It is about deciding what kind of world we want to live in.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Author contributions M.A. (Mustafa Arif) conducted all aspects of the research, including data collection, AI model development, data analysis, and manuscript preparation. Y.T. (Yoshiyasu Takefuji) provided guidance, oversight, and feedback throughout the research process. All authors reviewed and approved the final manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interests The authors declare no competing interests.

References

Bail CA (2024) Can generative AI improve social science? Proc Natl Acad Sci U S A 121(21):e2314021121. https://doi.org/10.1073/pnas.2314021121

Bell R, Menne NM, Mayer C, Buchner A (2024) On the advantages of using AI-generated images of filler faces for creating fair lineups. Sci Rep 14(1):12304. https://doi.org/10.1038/s41598-024-63004-z.PMID:38811714;PMCID:PMC11137153

Chen Y, Esmaeilzadeh P (2024) Generative AI in medical practice: in-depth exploration of privacy and security challenges. J Med Internet Res 26:e53008. https://doi.org/10.2196/53008

Fang X, Che S, Mao M, Zhang H, Zhao M, Zhao X (2024) Bias of AI-generated content: an examination of news produced by large language models. Sci Rep 14(1):5224. https://doi.org/10.1038/s41598-024-55686-2

Joshi S, Forjaz A, Han KS, Shen Y, Queiroga V, Xenes D, Matelsk J, Wester B, Barrutia AM, Kiemen AL, Wu PH, Wirtz D (2024) Generative interpolation and restoration of images using deep learning for improved 3D tissue mapping. BioRxiv: the preprint server for biology, 2024.03.07.583909. https://doi.org/10.1101/2024.03.07.583909

Lin S, Pandit S, Tritsch T, Levy A, Shoja MM (2024) What goes in, must come out: generative artificial intelligence does not present algorithmic bias across race and gender in medical residency specialties. Cureus 16(2):e54448. https://doi.org/10.7759/cureus.54448

Park HJ (2024) The rise of generative artificial intelligence and the threat of fake news and disinformation online: perspectives from sexual medicine. Investig Clin Urol 65(3):199–201. https://doi.org/10.4111/icu.20240015

Patel T, Othman AA, Sümer Ö, Hellman F, Krawitz P, André E, Ripper ME, Fortney C, Persky S, Hu P, Tekendo-Ngongang C, Hanchard SL, Flaharty KA, Waikel RL, Duong D, Solomon BD (2024) Approximating facial expression effects on diagnostic accuracy via generative AI in medical genetics. Bioinformatics 40(Supplement_1):110–118. https://doi.org/10.1093/bioinformatics/btae239

Templin T, Perez MW, Sylvia S, Leek J, Sinnott-Armstrong N (2024) Addressing 6 challenges in generative AI for digital health: a scoping review. PLOS Digital Health 3(5):e0000503. https://doi.org/10.1371/journal.pdig.0000503

Tortora L (2024) Beyond discrimination: generative AI applications and ethical challenges in forensic psychiatry. Front Psych 15:1346059. https://doi.org/10.3389/fpsyt.2024.1346059

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.