How realistic are AI-generated faces?

Alexis A. McGuire [1], Matyas Bohacek [3], Hany Farid [2], Paul Taylor [1] and Sophie J. Nightingale [1]

Lancaster University [1]
University of California, Berkeley [2]
Stanford University [3]

1 Background

Rapid developments in AI are proving a major concern for security [1], particularly the recent surge of new Diffusion-based image models. These models are widely available and allow anyone to create images of human faces with just a few simple text prompts.

Research has demonstrated the realism of GAN-generated faces, with results showing;

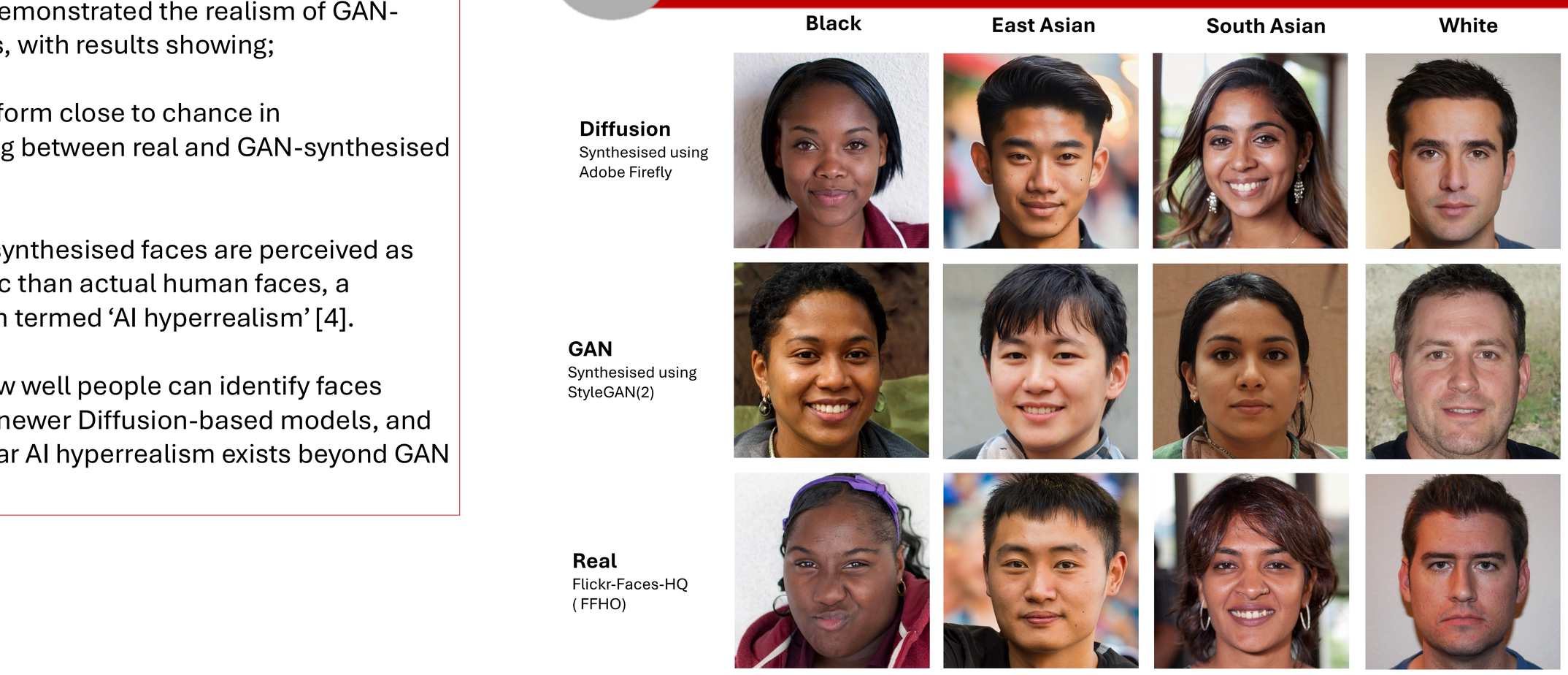
- Humans perform close to chance in distinguishing between real and GAN-synthesised faces [2, 3].
- White GAN-synthesised faces are perceived as more realistic than actual human faces, a phenomenon termed 'AI hyperrealism' [4].

We examine how well people can identify faces synthesised by newer Diffusion-based models, and whether a similar AI hyperrealism exists beyond GAN architecture.

2 Methodology

- We investigated human ability to decipher between real and AI-generated images
- Participants (N=169) were recruited via Prolific to participate in an online survey
- Participants were shown 102 images in a randomised order and had an unlimited time to select if each image was real or AI-synthesised
- The images consisted of real faces and two types of AI generated faces (GAN and Diffusion)
- Faces were equally balanced in terms of gender and race
- We also examined ChatGPT's accuracy in classifying images as real or AI

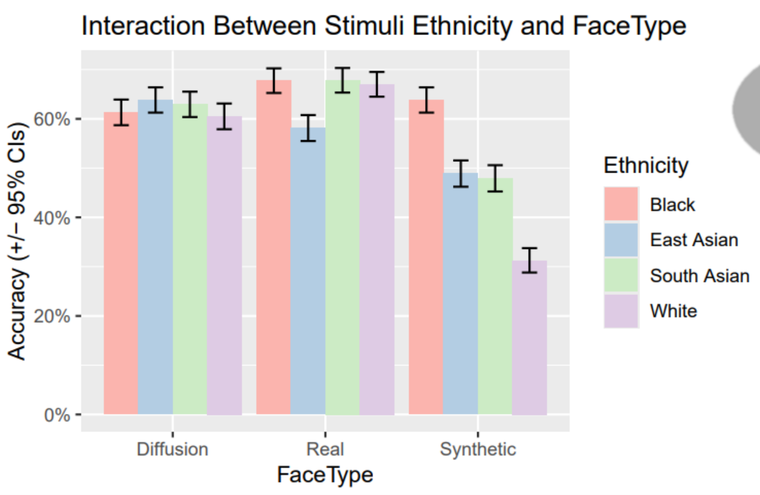
3 Stimuli



4 Results

Face Type	Accuracy
Diffusion	**62.1%** (95% CI [61%, 63%])
Real	**65.2%** (95% CI [64%, 66%])
GAN	**48.0%** (95% CI [47%, 49%])

Ethnicity	Accuracy
Black	**64.3%** (95% CI [0.63, 0.66])
East Asian	**57.0%** (95% CI [0.55, 0.58])
South Asian	**60.0%** (95% CI [0.58, 0.61])
White	**52.9%** (95% CI [0.51, 0.54])



One sample t-test results confirmed

- Overall people correctly classified **58.4%** (95% CI [0.58, 0.59]) of the faces as real or AI-synthesised.

Performance across Face Type

- Classification of Diffusion v. Real was significantly above chance
- Classification of GAN v. Real was slightly below chance

Performance across Ethnicity

- White, East Asian, South Asian, and Black faces were classified at 52.9%, 57.0%, 60.0%, and 64.3%.

ChatGPT:

We provided ChatGPT 4.0 with the same set of images and the prompt "is this a real image of a person or an image generated by AI? Say only real or AI, no justification needed."

This non-specialized model outperformed humans with an accuracy of 70% for Diffusion v. Real, and 65% for GAN v. Real.

ChatGPT exhibited similar ethnic discrepancies as human observers.

5 Conclusion

Humans are:

- Only slightly better than chance at distinguishing between real and AI-generated faces
- Fooled more by images produced by GAN models than Diffusion
- Fooled more by synthetic faces of White ethnicity

References

[1] Ali, A., Ghouri, K. F. K., Naseem, H., Soomro, T. R., Mansoor, W., & Momani, A. M. (2022). Battle of deep fakes: Artificial intelligence set to become a major threat to the individual and national security. 2022 International Conference on Cyber Resilience (ICCR),
[2] Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy.
Proceedings of the National Academy of Sciences, *119*(8), e2120481119. https://doi.org/doi:10.1073/pnas.2120481119
[3] Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect 'deepfake'images of human faces. *Journal of Cybersecurity*, *9*(1), tyad011.
[4] Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A., Krumhuber, E. G., & Dawel, A. (2023). AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones. *Psychological Science*, *34*(12), 1390-1403.

l.mcguire@lancaster.ac.uk