# Lensa AI – Sexual Avatars

1. The technique and how it is applied

**Lensa AI** is a technology that re-imagines users' photos as anime characters and celestial beings. To be specific, after users provide several selfies and choose the categories for the type of image they want, Lensa will show what they look like as an anime character.

To create these portraits, Lensa uses images as the prompts and builds the product on Stable Diffusion, a deep learning model released to the public. In other words, Lensa acts as a middleman between the users and Stable Diffusion. Stable Diffusion is based on a Latent Diffusion Model: the data flows among three parts, Pixel Space, Latent Space, and Conditioning; the image encoder compresses the image from Pixel Space to Latent Space, adds noise, and performs a diffusion process; then, the CLIP text encoder converts the input descriptors to the condition of the denoising process to obtain the potential representation of the generated image. Finally, the model transforms the image from latent space back to pixel space to generate the final image (See example in Appendix 1).

2. Incident background and details

While popular, Lensa AI is not without its problems. The app has been criticized for perpetuating biases related to nudity and racist stereotypes, particularly against women of Asian and Black descent. One example came from Melissa Heikkilä ,an Asian journalist writing for MIT Technology Review. According to Melissa, Lensa produced several avatars of her that were "nude" or "showed a lot of skin," while her white female colleagues "got significantly fewer sexualized images."[1] Importantly, Melissa didn't submit anything that disobeyed Lensa's terms of use, and it was the app that produced nudity and perpetuated stereotypes of Asian women. This suggests that Lensa AI's algorithms may refer to Asian women as sexual objects. Another user reported that despite having features that span the East and South Asian spectrum, the avatars Lensa generated for her were completely skewed towards East Asian features, indicating a problem with the app's ability to accurately identify and generate Asian faces.

3. Ethical implications

These kinds of outputs from generative AI algorithms can have harmful ethical implications, both at the individual and societal levels. At an individual level, these outputs can perpetuate harmful stereotypes and unrealistic beauty standards, leading to low self-esteem, body

1 https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/

dissatisfaction, and mental health issues. At a societal level, they can contribute to a culture that objectifies and sexualizes individuals based on their physical appearance. When Lensa generates sexualized stereotypes **based on user-uploaded photos**, it can have an even more damaging impact, reinforcing harmful beauty standards and prenoting discrimination and bias. In this example, the sexualization of Asian women based on their physical appearance preserves the stereotype that they are hypersexualized and objectified. It is thus crucial to ensure that AI systems are designed and developed with care and consideration for their potential impacts on individuals and society.

4. Potential solutions and mitigation

Lensa's AI technique produces biased racial results, because of the data it is trained on, and the designing decisions made by developers. As researchers already state, enacting malignant stereotypes and bias is not a new thing to AI techniques.[2]

Our first potential solution to this problem is that Lensa builds a professional team to write rigorous ethical guidelines and actively monitor and correct biases in the data. This can include regularly reviewing outputs and adjusting the training data as needed. Secondly, we suggest strengthening the effectiveness of content filters when generating images on Lensa, which means training content filters on bigger, more diverse corpus data. This could reduce the likelihood of generating images that spread negative stereotypes or biases.

We believe the mitigation of these changes should come from two directions: stakeholder diversification, and user awareness. First, diversifying the company's stakeholders can prevent these problems from happening in the initial stage of development and encourage designers to build a positive brand impression of Lensa AI in the market. Second, a campaign to raise user awareness to these problems can encourage Lensa to create and improve sophisticated content filters.

Overall, bias is a huge industry-wide problem.[3] As a society, we must quickly find solutions for it, to catch-up with the rapid growth and impact of generative AI in our lives.

2 https://dl.acm.org/doi/pdf/10.1145/3531146.3533138
3 https://www.vox.com/future-perfect/23023538/ai-dalle-2-openai-bias-gpt-3-incentives

BA820 Mini Write Up – Generative AI: Dangers and Misuses
Group B4

**Appendix 1**: Lensa AI Avatar Example

Actor Timothee Chalamet and his Lensa avatar. Photo: Getty Images and Instagram
https://www.scmp.com/lifestyle/fashion-beauty/article/3203128/ai-art-app-lensa-and-its-magic-avatars-are-taking-over-internet-why-matters-and-why-some-are-worried

**Appendix 2**: MIT Technology Review Reporter, Melissa Heikkilä Lensa Avatars