1. Reconhecimento da questão ética

Escolhemos uma questão ética que se encontra no âmbito deste caso, que pode ser formulada da seguinte forma: *Será eticamente correto uma aplicação tornar fotos modestas em avatares de natureza sexual por falha de aprendizagem da IA?*

Duas opções possíveis: *gerar* ou *não gerar* estes avatares?

2. Obter os factos (só alguns)

- Prisma Lab's CEO and co-founder told us that the behavior we observed in our article can only happen if the AI is intentionally provoked into creating NSFW content and this represents a breach against its terms of use.
- To enhance the work of Lensa, we are in the process of building the NSFW filter. It will effectively blur any images detected as such.
- The selfies are uploaded to Amazon or Google cloud servers. Then, using the latent diffusion model and CLIP, which is a dataset of over 400 million images, the AI studies your selfies and generates photos.
- When it comes to male avatars, the portraits exude both masculinity and strength: generating men as warriors, astronauts, and something resembling a Top-Gun-like themed photoshoot.
- CLIP is a neural network which efficiently learns visual concepts from natural language supervision.
- The app calls its latest self-portrait generation feature "Magic Avatars," and it uses deep-learning to create dreamy selfies in various art styles.

Envolventes: a empresa + investidores, a comunidade tecnológica e o utilizador.

3. Avaliar Alternativas

Abordagem Utilitarista: Os utilizadores beneficiam da utilização desta aplicação, porque podem criar imagens mais estéticas de si próprios, aumentando a sua confiança e fantasiando sobre cenários fantásticos. Contudo, esta aplicação de *deep learning* é baseada numa base de dados de grande escala o que, por vezes, pode fazer com que estas criações não sejam bem sucedidas, em específico, sexualizando o conteúdo. A nível da empresa e dos seus investidores, bem como da comunidade tecnológica, este feito revela o avanço da tecnologia, por isso, estes são beneficiados com a evolução da aplicação. Atualmente, para poder haver esse progresso, a IA precisa da informação de uma vasta rede de fotos, incluindo as fotos inseridas pelos utilizadores. É de notar que os Termos e Condições da aplicação proíbem o anexo de fotografias de crianças e sexuais. Ou seja, não é ético a aplicação sexualizar fotos modestas, mas isto nunca aconteceria se os utilizadores não inserissem fotos sexuais. Além disto, não nos podemos esquecer que uma IA funciona como uma rede neuronal, pretendendo reproduzir aquilo que um humano faz intuitivamente. Posto isto, esta IA aprende com a nossa sociedade e se esta tem standards (sexualizando e discriminando racialmente), a IA vai reproduzir estes padrões. Assim, a opção que melhor otimizará a relação benefício-custo é *gerar os avatares*, isto porque, a IA funciona de maneira a produzir avatares cada vez mais adequados e que vão de encontro com o objetivo da empresa e utilizadores.

Abordagem dos Direitos: Focando-se nos direitos humanos, a opção que melhor representa os direitos de todas as partes envolvidas é *não gerar os avatares.*

Em relação aos direitos dos utilizadores é óbvio que a aplicação não pode criar avatares de caráter sexual, a partir das fotos submetidas (dentro dos parâmetros requeridos pela app), sem o consentimento dos mesmos, muito menos, sabendo que têm o direito de autor sobre a imagem criada. No que toca à empresa, o utilizador não pode enviar fotos para a tecnologia que não estejam dentro do conteúdo apropriado e esse direito foi, de facto, quebrado várias vezes "poluindo" o algoritmo, fazendo com que fotos modestas gerem mais facilmente a imagens de natureza sexual. Dito isto, temos também que considerar que a tecnologia já tinha sim uma base de dados (CLIP) com mais de 400 milhões de imagens da qual não podemos ter a certeza que não está a "poluir" e ser uma causa para essa transformação de fotos. Por fim, a empresa e os investidores, bem como, a comunidade tecnológica têm o direito à evolução do produto/tecnologia, mas este não pode ser feito às custas da violação dos direitos de imagem dos utilizadores, mesmo estando nos Termos e Condições da aplicação.

4. Tomar uma decisão e testá-la

Sendo que apenas foram adotadas 2 abordagens concorrentes, analisando a questão ética de forma mais crítica e profunda, a melhor decisão é *gerar os avatares*, porque sabemos que a é a injeção de imagens de conteúdo sexual que fazem com que a criação de avatares sofra. Logo, se tal não acontecer, nem os direitos dos utilizadores e empresa são quebrados e o objetivo da aplicação é cumprido, tornando os utilizadores mais felizes, bem como os restantes envolvidos.

5. Agir e avaliar o desfecho

Não sendo possível avaliar as consequências imediatas da decisão, esta poderia ter forte impacto na comunidade tecnológica e na sociedade em geral, levando a uma discussão aprofundada sobre as questões éticas e legais que se prendem com o desenvolvimento tecnológico e o seu potencial.