

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

UM LEVANTAMENTO DE FERRAMENTAS DE IA PARA PRODUÇÃO DE DEEPFAKES

GUSTAVO M. CARDOSO¹, CLAUDIO H. YAMAMOTO²

¹ Graduando em Ciência da Computação, IFSP, Campus Salto, g.milan@aluno.ifsp.edu.br.

² Doutorado em Ciências no programa Ciências de Computação e Matemática Computacional pela USP, professor efetivo da área de Informática - Programação e Banco de Dados - do IFSP Câmpus Salto, haruo@ifsp.edu.br.

Área de conhecimento (Tabela CNPq): 1.03.00.00-7 Ciência da Computação

RESUMO: A inteligência artificial (IA) está revolucionando diversos setores, trazendo inovações e transformações significativas. Uma das técnicas que utiliza IA é o *deepfake*, que permite a criação de conteúdos falsificados extremamente realistas com o auxílio de aprendizado de máquina e aprendizado profundo. Este trabalho realizou um levantamento das ferramentas existentes no mercado para a produção de *deepfakes*, tanto de imagem e vídeo quanto de clonagem de voz. Foi realizada uma análise comparativa dessas ferramentas, considerando características como gratuidade, capacidade de reutilização sem cobranças e facilidade de uso. É possível afirmar que a variedade e facilidade de produção de *deepfakes* permite que qualquer indivíduo, por mais leigo que seja, possa utilizar essas ferramentas para diversas aplicações sejam com boas ou más intenções.

PALAVRAS-CHAVE: deepfake; inteligência artificial; machine learning; voice clone;

A SURVEY OF AI TOOLS FOR DEEPFAKES PRODUCTION

ABSTRACT: Artificial intelligence (AI) is revolutionizing several sectors, bringing significant innovations and transformations. One of the techniques that uses AI is deepfake, which allows the creation of extremely realistic fake content with the help of machine learning and deep learning. This work carried out a survey of the existing tools on the market for producing deepfakes, both image and video and voice cloning. A comparative analysis of these tools was carried out, considering characteristics such as licensing model, ability to be reused without charge and ease of use. It is possible to say that the variety and ease of producing deepfakes allows any individual, no matter how novice, to use these tools for different applications, whether with good or bad intentions.

KEYWORDS: deepfake; artificial intelligence; machine learning; voice clone

INTRODUÇÃO

A inteligência artificial (IA) tem transformado diversos setores, como a tecnologia e a saúde, ao aumentar a eficiência e permitir inovações. No entanto, técnicas como o *deepfake* — que utilizam IA para manipular vídeos, fotos e áudios — têm gerado preocupações éticas e sociais. O uso indevido de *deepfakes* pode resultar em falsidade ideológica, estelionato e invasão de privacidade, o que ressalta a importância de se estudar essa tecnologia (Citron; Chesney, 2019). Além disso, os avanços em IA podem agravar problemas como desemprego e discriminação, conforme apontado por pesquisas (Acemoglu; Restrepo, 2020).

Este trabalho tem como objetivo mapear e analisar comparativamente as principais ferramentas disponíveis para a criação de *deepfakes* de imagem, vídeo e áudio, discutindo seus riscos e implicações éticas. Como hipótese, propõe-se que a ampla acessibilidade dessas ferramentas, associada à falta de regulamentação adequada, contribui para o aumento de crimes digitais. A relevância deste estudo se dá pela crescente popularização dos *deepfakes* e pela necessidade de maior compreensão sobre as potenciais ameaças dessa tecnologia. Embora existam pesquisas sobre *deepfakes* (Kietzmann et al., 2020; Mirsky; Lee, 2021), este trabalho oferece uma contribuição ao focar na categorização das ferramentas mais usadas atualmente.

MATERIAL E MÉTODOS

Para a realização deste estudo, foi adotado um processo sistemático para identificar e analisar ferramentas de inteligência artificial utilizadas na criação de *deepfakes*. As buscas foram realizadas entre os meses de fevereiro e junho nas plataformas Google, YouTube, TikTok. Foram utilizadas as seguintes palavras-chave e composições de termos: “ferramentas de *deepfake*”, “criação de *deepfake* de vídeo”, “*deepfake* de áudio com IA” e “tecnologias para *deepfake*”.

Os resultados das buscas foram avaliados com base em critérios pré-definidos, que incluíam a facilidade de uso da ferramenta, sua popularidade e o tipo de licenciamento (gratuito ou pago). Além disso, foram analisadas revisões técnicas disponíveis em publicações especializadas, comparando o número de acessos e visualizações em vídeos explicativos nas plataformas YouTube e TikTok.

A popularidade das ferramentas foi medida através da frequência de busca e relevância nas plataformas citadas, considerando tanto o número de downloads quanto as menções em comunidades online. Por fim, o tipo de licenciamento foi categorizado para facilitar a comparação das opções de uso disponíveis aos usuários, observando as permissões e restrições oferecidas por cada ferramenta.

RESULTADOS E DISCUSSÃO

Nas tabelas 1 e 2, são apresentados os resultados comparativos das ferramentas de *deepfake* analisadas, organizadas de acordo com critérios como gratuidade, capacidade de reutilização sem limitações, e facilidade de uso. A Tabela 1 apresenta uma comparação das ferramentas de imagem e vídeo, enquanto a Tabela 2 trata das ferramentas de clonagem de voz.

TABELA 1. Análise comparativa entre as ferramentas de *deepfake* de imagem e vídeo, nas linhas as ferramentas encontradas, nas colunas as características de cada uma delas.

| | Gratuidade | Reutilização sem limitação ou cobrança | Facilidade de uso | Fonte |
|-----------------------------------|------------|--|-------------------|---------|
| Midjourneyai | ✓ | ✓ | X | TikTok |
| Swapface | ✓ | ✓ | ✓ | YouTube |
| This Person Does Not Exist | ✓ | ✓ | ✓ | Google |
| Viggle | ✓ | ✓ | ✓ | TikTok |
| Wefaceswap | ✓ | X | ✓ | Google |

Midjourneyai (2024) é um software de código aberto encontrado em um repositório no Git Hub, sendo possível baixar e fazer todo o processo de *deepfake* no seu hardware pessoal. Porém essa ferramenta foi descartada da pesquisa devido à necessidade de ter um conhecimento mais profundo de repositório e instalações de pacotes via CMD para que seja possível executar a aplicação.

Swapface (2024) é uma aplicação desktop que utiliza redes neurais convolucionais para identificar e mapear rostos. As redes neurais convolucionais (Convolutional neural network ou CNNs) são um subconjunto de aprendizado de máquina para tarefas de classificação e visão computacional, oferecendo uma abordagem mais dimensional (Alves, 2018). No caso do Swapface, estas redes são treinadas com grandes volumes de dados contendo rostos humanos, o que permite ao software reconhecer e substituir com alto grau de realismo as características faciais de uma pessoa pela de outra. A primeira coisa que o software faz é encontrar e definir as áreas faciais nas imagens ou vídeos

fornecidos. Em seguida, coleta características faciais e únicas, como a forma do rosto, os olhos, o nariz e a boca. Para garantir uma transição suave e natural, parâmetros como cor de pele e iluminação são ajustados para mapear e alinhar as características faciais da pessoa de origem com as da pessoa de destino. Além da sua gratuidade e fácil usabilidade, o Swapface possui um limite de produção de 10 *deepfakes*, quantidade que é zerada todos os dias, permitindo assim um uso constante da ferramenta.

This Person Does Not Exist (2024) é um site que gera rostos hiper-realistas de pessoas que, na verdade, não existem. Criado pelo programador Phillip Wang, ele utiliza um algoritmo chamado StyleGAN para combinar características de fotos reais e criar imagens falsas. Criado pela NVIDIA, o StyleGAN é uma arquitetura de rede neural que permite tal formulação de rostos altamente realistas, se destacando pela possibilidade de aplicação de “estilos” para manipular qual tipo de imagens serão geradas. Outra característica desse algoritmo é sua técnica de crescimento progressivo, onde a rede gera imagens de baixa resolução e gradualmente, aumenta a resolução dessas imagens com o intuito de melhorar a capacidade do treinamento e qualidade das imagens produzidas (Sales, 2024). Ao testar a ferramenta é possível notar logo de cara que além de sua gratuidade é possível criar rostos aleatórios repetidamente, sem nenhum tipo de cobrança e de maneira simples.

Viggle (2024), é uma plataforma que permite a criação e animação de vídeos com imagens de personagens ou comandos de texto, possibilita que qualquer pessoa anime personagens com uma compreensão real da física usando o modelo de vídeo-3D JST-1, que é um modelo de IA desenvolvido exclusivamente para a geração de vídeos manipuláveis em 3D (Lemos, 2024). O Viggle se destaca por ter um funcionamento através de um servidor de Discord, isso favorece a sua disseminação, pois não necessita de download nem de processamento de hardware pessoal.

Wefaceswap (2024) é um software para geração de *deepfake* que apesar de ser bastante eficiente na troca de rostos, apresentava um sistema de créditos, que após esgotado por cada utilização da ferramenta, se tornava obrigatório algum custo monetário.

TABELA 2. Análise comparativa das ferramentas de voice clone nas linhas as ferramentas encontradas, nas colunas as características de cada uma delas.

| | Gratuidade | Reutilização sem limitação ou cobrança | Facilidade de uso | Fonte |
|--------------------|------------|--|-------------------|---------|
| ElevenLabs | ✓ | ✓ | ✓ | YouTube |
| Murf AI | ✓ | X | ✓ | Google |
| PlayHT | ✓ | ✓ | ✓ | YouTube |
| Rask AI | ✓ | ✓ | ✓ | Google |
| Resemble AI | ✓ | X | X | Google |

Criado em 2022, ElevenLabs (2024) é um dos principais softwares de síntese de voz e fala e se destaca em identificar emoções como raiva, tristeza, felicidade nos elementos do texto. Isso permite ao sistema entender os sentimentos do usuário, permitindo produzir resultados mais realistas, semelhantes à humana. Já sua ferramenta de clonagem de voz permite que os usuários pagantes carreguem amostras de voz personalizadas para criar estilos vocais.

O Murf AI (2024) é uma plataforma de texto para voz baseado em nuvem, porém apresenta o principal problema da maioria das ferramentas: a limitação na quantidade de uso, sendo necessário pagar por algum plano após utilizar 10 minutos de produção de voz gerada por IA. Logo foi descartado sua utilização na pesquisa.

PlayHT (2024) foi inicialmente desenvolvido como uma extensão do Chrome em 2016 com o objetivo de permitir que pessoas ocupadas ouvissem artigos de internet. A intenção é fornecer uma forma útil de consumir conteúdo escrito sem precisar dedicar tempo exclusivo à leitura. A extensão rapidamente se tornou popular e ao perceberem o potencial escondido, decidiram expandir o software destinado a auxiliar na criação de conteúdo de áudio realista permitindo no plano gratuito a inserção de amostra de vozes reais para voice clone, viabilizando a produção de *deepfakes* com vozes de terceiros.

Rask (2024) é uma ferramenta de IA para vídeo que oferece serviços automatizados de narração, legendas e tradução. Ele pode localizar rapidamente o conteúdo de vídeo, o que permite que empresas criem vídeos multilíngues. Suporta vários idiomas, incluindo o inglês, o chinês, o árabe, o russo, o turco,

o ucraniano, o francês, o alemão, o espanhol etc. Diversas empresas e softwares utilizam Rask de alguma forma, dentre as principais estão a Amazon, Universal Music Group e inclusive o próprio ElevenLabs. Se alinha aos critérios selecionados e surpreende por ser a única de todas as ferramentas estudadas de voice clone que apresenta dublagem instantânea de um vídeo de uma outra língua.

Resemble AI (2024), é semelhante ao ElevenLabs, cria vozes através do texto inserido pelo usuário, porém não foi considerado neste estudo pelo fato de a sua gratuidade ser limitada a alguns segundos de uso para teste, inviabilizando sua usabilidade futuramente.

CONCLUSÕES

Foram encontradas cinco ferramentas de *deepfake* para imagem e vídeo e cinco ferramentas para clonagem de voz, sendo a grande maioria com licenciamento gratuito. As ferramentas de *deepfake* para imagem e vídeo permitem a modificação de rostos, criação de vídeos e imagens a partir de poucas instruções. No entanto, apesar da gratuidade, muitos licenciamentos limitam a utilização da ferramenta, exigindo o consumo de anúncios ou o pagamento para liberar funcionalidades adicionais. As ferramentas de clonagem de voz, mesmo permitindo modulação de vozes reais e a criação de vozes artificialmente, também apresentam questões semelhantes de licenciamento.

Conclui-se, após o desenvolvimento desta pesquisa, que a variedade e a facilidade de produção de *deepfakes* tornam essas tecnologias acessíveis para qualquer indivíduo, independentemente do nível de conhecimento técnico. Tais ferramentas podem ser aplicadas para fins de entretenimento, acadêmicos e profissionais, mas, se utilizadas de forma indevida, têm o potencial de causar grandes prejuízos, como a difusão de informações falsas (*fake news*), estelionato, falsidade ideológica e violação de privacidade.

Para trabalhos futuros, será relevante aprofundar o estudo sobre as regulamentações e legislações que envolvem o uso de *deepfakes*, além de desenvolver e aprimorar métodos de detecção dessas falsificações. Além disso, há uma necessidade crescente de investigar como as novas tecnologias de IA, podem contribuir tanto para a criação destes conteúdos quanto para a mitigação dos seus impactos negativos. A pesquisa poderá evoluir para o desenvolvimento de ferramentas que ajudem na identificação de *deepfakes*, contribuindo para a segurança digital e a ética no uso da IA.

CONTRIBUIÇÕES DOS AUTORES

Gustavo M. Cardoso contribuiu com a redação, curadoria dos dados, desenvolvimento metodológico e análise dos resultados. Todas as etapas do trabalho foram executadas por G.M.C., incluindo a coleta e análise de informações, além da elaboração do manuscrito.

Claudio H. Yamamoto contribuiu com a supervisão, orientação metodológica e revisão crítica do trabalho. Seu apoio foi essencial para o delineamento da estrutura do estudo e a validação dos resultados obtidos.

Todos os autores contribuíram com a redação, revisão e aprovaram a versão final do trabalho submetido.

AGRADECIMENTOS

A realização deste trabalho não seria possível sem o apoio e a colaboração de diversas pessoas, às quais expresso minha profunda gratidão. Em especial, agradeço ao meu orientador, Claudio Haruo Yamamoto, por sua orientação, paciência e dedicação ao longo deste processo. Suas valiosas contribuições foram essenciais para o desenvolvimento deste trabalho.

Por fim, agradeço ao IFSP, por proporcionar um ambiente acadêmico estimulante e recursos indispensáveis para a realização deste trabalho.

REFERÊNCIAS

ACEMOGLU, D.; RESTREPO, P. *Artificial Intelligence, Automation, and Work*. *Journal of Economic Perspectives*. University of Chicago Press, v. 34, n. 2, p. 1-20, 2020.

ALVES, Gisely. Entendendo Redes Convolucionais (CNNs). Medium, 2018. Disponível em: <https://medium.com/neuronio-br/entendendo-redes-convolucionais-cnns-d10359f21184>. Acesso em: 27 out. 2024.

CITRON, D. K.; CHESNEY, R. *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. California Law Review, v. 107, p. 1753, 2019.

ELEVENLABS. ElevenLabs, 2024. *Free Text to Speech & AI Voice Generator*. Disponível em: <https://elevenlabs.io/>. Acesso em: 27 out. 2024.

KIETZMANN, J.; LEE, L.; MCCARTHY, I.; KIETZMANN, T. *Deepfakes: Trick or treat? Business Horizons*, v. 63, n. 2, p. 135-146, 2020.

LE MOS, Diego A. *Viggle AI: Animação 3D Fácil com a IA Viggle. Ia Básico*, 2024. Disponível em: <https://www.iabasico.com/posts/viggle-ai-animacao-3d-facil-com-a-ia-viggle>. Acesso em: 27 out. 2024.

MIDJOURNEY. Midjourney Inc, 2024. *New research lab. Exploring new mediums of thought*. Disponível em: <https://www.midjourney.com/explore?tab=top>. Acesso em: 27 out. 2024.

MIRSKY, Y.; LEE, W. *The Creation and Detection of Deepfakes: A Survey*. ACM Computing Surveys (CSUR), v. 54, n. 1, p. 1-41, 2021.

MURF.AI. Murf AI, 2024. *AI voice generator that transforms your text into realistic speech in minutes*. Disponível em: <https://murf.ai/>. Acesso em: 27 out. 2024.

PLAYHT. PlayHt, 2024. *AI voice generator: Most realistic text to speech AI*. Disponível em: <https://play.ht/>. Acesso em: 27 out. 2024.

RASK AI. Rask AI, 2024. *Ferramenta líder de localização e dublagem de vídeo com IA*. Disponível em: <https://pt-br.rask.ai/>. Acesso em: 27 out. 2024.

RESEMBLE AI. Resemble AI, 2024. *The All-in-One AI Voice Platform*. Disponível em: <https://www.resemble.ai/>. Acesso em: 27 out. 2024.

SALES, Janderson de. *StyleGAN: Revolucionando a arte das redes adversárias generativas*. Guia de Hospedagem, 2024. Disponível em: <https://guiadehospedagem.com.br/stylegan-revolucionando-a-arte/>. Acesso em: 27 out. 2024.

SWAPFACE. Swapface, 2024. *The most light, hyperrealistic, real-time*. Disponível em: <https://www.swapface.org/#/home>. Acesso em: 27 out. 2024.

THIS PERSON DOES NOT EXIST. This Person Does Not Exist, 2023. *Random Face Generator*. Disponível em: <https://this-person-does-not-exist.com/en>. Acesso em: 27 out. 2024.

VIGGLE. Viggle, 2024. *Controllable Video Generation*. Disponível em: <https://viggle.ai/>. Acesso em: 27 out. 2024.

WEFACESWAP. Wefaceswap, 2024. *Faceswap sem esforço na nuvem para fotos, vídeos e GIFs*. Disponível em: <https://www.wefaceswap.com/>. Acesso em: 27 out. 2024.