



INSTITUTO
FEDERAL
São Paulo

APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NO COMBATE ÀS FAKE NEWS

Campus
Guarulhos

Maria Paula Xavier da Silva
Pedro Santos de Oliveira
Prof. Dra. Gema Galgani Rodrigues Bezerra

Introdução

A pesquisa investiga o uso de Inteligência Artificial (IA) na identificação e combate de *fake news* produzidas por *deepfakes*, que são conteúdos audiovisuais manipulados com o intuito de distorcer a realidade. *Fake news*, definidas como “documentos deliberadamente falsos com o objetivo de manipular os consumidores” (Meneses, 2018, p. 47), têm se proliferado com o avanço das tecnologias digitais, particularmente desde a década de 1990 (Soll, 2016). Com a popularização das redes sociais, a disseminação de *fake news* tornou-se mais ampla e impactante. Dados do *Digital News Report* (Levy et al., 2017) e do *dfndr lab* (Pecsen, 2020) indicam uma crescente dificuldade em distinguir informações verdadeiras de falsas. Os *deepfakes*, surgidos em 2017 (Hall, 2018), utilizam IA para criar conteúdos audiovisuais falsificados que desafiam a detecção tradicional (Molina & Berenguel, 2022). Este estudo busca responder como a IA pode identificar eficazmente *deepfakes* e melhorar a confiança nas informações compartilhadas nas redes sociais. A hipótese é que algoritmos de IA podem distinguir *deepfakes* com maior precisão do que métodos convencionais, e o objetivo é desenvolver uma ferramenta baseada em IA para essa finalidade, propondo estratégias para mitigar o impacto das *fake news*.

Metodologia

A pesquisa foi dividida em duas etapas. Na primeira - testes de análise de eficiência de sites já existentes - foram selecionadas três plataformas que identificam *deepfakes*, disponíveis gratuitamente na internet: *BioID Playground*, *Deepware* e *BrandWell*. Para os testes, foram utilizados cinco vídeos e cinco imagens feitas com o uso de *deepfakes*.

A segunda etapa da pesquisa iniciou-se com a produção do protótipo, isto é, o desenvolvimento de um sistema de reconhecimento facial utilizando a linguagem de programação *Python*.

Resultados

Para os testes, foram utilizados cinco vídeos e cinco imagens feitas com o uso de *deepfakes*. Os vídeos, todos feitos com o rosto de celebridades, foram retirados do Youtube. Em relação às plataformas, a *BioID Playground* foi utilizada para identificar tanto os vídeos, quanto as imagens, a plataforma *Deepware* foi utilizada apenas para a detecção dos vídeos e a plataforma *BrandWell* foi utilizada apenas para a identificação das imagens. Os conteúdos foram identificados na tabela 1 e os resultados gerais dos testes e a comparação do percentual de acertos entre as três plataformas foram apresentados na tabela 2.

Tabela 1. Identificação dos conteúdos utilizados nos testes

	Descrição
Vídeo 1	<i>Deepfake</i> do ator Morgan Freeman (Diep Niep, 2021)
Vídeo 2	<i>Deepfake</i> do empresário Elon Musk (Civic Sentinel, 2022)
Vídeo 3	<i>Deepfake</i> do ator Jerry Seinfeld, como se ele fizesse parte do filme “Pulp Fiction” (Desifakes, 2022)
Vídeo 4	<i>Deepfake</i> do ator Tom Cruise (Vecanoi, 2021)
Vídeo 5	<i>Deepfake</i> do ator e ex-governador da Califórnia, Arnold Schwarzenegger, como se ele fizesse parte do filme “O mago de Oz” (Brainmonarch, 2023)
Imagem 1	Retratação do site <i>This Person Does Not Exist</i>
Imagem 2	Retratação do site <i>This Person Does Not Exist</i>
Imagem 3	Retratação do site <i>This Person Does Not Exist</i>
Imagem 4	Retratação do site <i>This Person Does Not Exist</i>
Imagem 5	Retratação do site <i>This Person Does Not Exist</i>

Elaborada pelos autores

Tabela 2. Resultados e comparação do percentual de acertos das três plataformas

	BioID Playground	Deepware	BrandWell
Vídeo 1	Analisou corretamente	Analisou corretamente	não analisa vídeos
Vídeo 2	Analisou corretamente	Analisou corretamente	não analisa vídeos
Vídeo 3	Analisou incorretamente	Analisou incorretamente	não analisa vídeos
Vídeo 4	Analisou corretamente	Analisou corretamente	não analisa vídeos
Vídeo 5	Analisou incorretamente	Analisou incorretamente	não analisa vídeos
Imagem 1	Analisou corretamente	não analisa imagens	Analisou incorretamente
Imagem 2	Analisou corretamente	não analisa imagens	Analisou incorretamente
Imagem 3	Analisou corretamente	não analisa imagens	Analisou corretamente
Imagem 4	Analisou corretamente	não analisa imagens	Analisou incorretamente
Imagem 5	Analisou corretamente	não analisa imagens	Analisou incorretamente
Percentual de acertos	80% (8 de 10)	60% (3 de 5)	20% (1 de 5)

Elaborada pelos autores

Considerações finais

O protótipo de reconhecimento facial utilizando *Python* e suas bibliotecas robustas, como *OpenCV*, *Dlib* e *face_recognition*, tem demonstrado resultados promissores em termos de eficiência e flexibilidade. O sistema foi projetado para reconhecer regiões faciais e realizar comparações entre imagens estáticas e vídeos em tempo real, o que abre portas para inúmeras aplicações, especialmente em segurança digital e combate aos *deepfakes*.

A capacidade operacional da implementação desse sistema de reconhecimento facial é alta, dado o uso de bibliotecas *Python* amplamente adotadas na indústria e nas universidades, como *OpenCV*, *Dlib* e *face_recognition*. Essas bibliotecas são de código aberto, o que anula quaisquer custos para o desenvolvimento. Além disso, a escolha de *Python*, uma linguagem de programação de fácil aprendizado e alta legibilidade, facilita a manutenção e o aprimoramento contínuo do sistema por novos desenvolvedores e pesquisadores.

A expansão do sistema para detectar *deepfakes* teria um impacto significativo no combate à desinformação. *Deepfakes* são uma preocupação crescente, pois podem ser usados para criar conteúdos enganosos que parecem reais, afetando a opinião pública e até mesmo a segurança nacional. Implementar este sistema em larga escala poderia ajudar a mitigar tais ameaças, fornecendo uma ferramenta confiável para identificar e combater a desinformação digital.

Embora o sistema tenha demonstrado um desempenho satisfatório nas fases de desenvolvimento e teste, há várias áreas para futuras melhorias. Uma delas é a integração de algoritmos mais avançados de aprendizado profundo que possam lidar melhor com condições adversas, como pouca luz ou ângulos de visão incomuns. Outro ponto é o desenvolvimento de métodos para reconhecimento facial em uploads de vídeos e comparação em tempo real com fotos, para verificar a autenticidade de um vídeo e identificar *deepfakes* de maneira eficiente.



Referências

Brainmonarch. Arnold Schwarzenegger Sings About Rainbows. YouTube, 18 de maio de 2023. Disponível em: <https://youtu.be/K_XwseDwmuQ>.

Civic Sentinel. Ultra realistic Deepfake of Elon Musk. YouTube, 22 de dezembro de 2022. Disponível em: <<https://youtu.be/XuUKyPegBE>>.

Desifakes. Jerry Seinfeld in Pulp Fiction (DeepFake). YouTube, 6 de fevereiro de 2022. Disponível em: <<https://youtu.be/SIMBVXkQbWU>>.

Diep Niep. This is not Morgan Freeman - A Deepfake Singularity. YouTube, 7 de julho de 2021. Disponível em: <<https://youtu.be/oxXP89p5ETO>>.

Hall, H. K. Deepfake videos: When seeing isn't believing. Cath. UJL & Tech, 27, 51, 2018. Disponível em: <<https://scholarship.law.edu/jlt/vol27/iss1/4>>.

Levy, D. et al. Reuters Institute Digital News Report, 2017. Disponível em: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf>.

Meneses, J. P. Sobre a necessidade de conceptualizar o fenómeno das fake news. Observatorio, v. 2018, p. 37-53, 2018. Disponível em: <<https://pdfs.semanticscholar.org/c2ca/e77a6566d79b7c8b94c32a0476051344f.pdf>>.

Molina, A. C.; Berenguel, O. L. Deepfake: The Evolution of fake news. Research, Society and Development, [S. l.], v. 11, n. 6, p. e56211629533, 2022. Disponível em: <<https://rsdjournal.org/index.php/rsd/article/view/29533>>.

Pecsen, T. 1 a cada 2 brasileiros afirma já ter compartilhado Fake News sem saber, 2020. Disponível em: <<https://www.psafe.com/blog/1-a-cada-2-brasileiros-afirma-ja-ter-compartilhado-fake-news-sem-saber/>>.

Soll, J. The Long and Brutal History of Fake News, 2016. Disponível em: <<https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/>>.

Vecanoi. Very realistic Tom Cruise Deepfake | AI Tom Cruise. YouTube, 28 de fevereiro de 2021. Disponível em: <<https://youtu.be/lyIOVbPcM>>.