INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO CAMPUS GUARULHOS

Av. Salgado Filho, 3501 - Centro, Guarulhos - SP, 07115-000

MARIA PAULA XAVIER DA SILVA

PEDRO SANTOS DE OLIVEIRA

PROF. DRA. GEMA GALGANI RODRIGUES BEZERRA

APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NO COMBATE ÀS *FAKE NEWS* PRODUZIDAS POR MEIO DE *DEEPFAKES*

Período de desenvolvimento: abril de 2023 até setembro de 2023.

SUMÁRIO

1 RESUMO

A pesquisa aborda o uso da Inteligência Artificial (IA) no combate às *fake news*, com foco particular nos *deepfakes*, que são conteúdos manipulados por IA para distorcer a realidade. As *fake news*, amplificadas pelas redes sociais, representam um desafio crescente para a verificação da verdade e a confiança nas informações. Este estudo visa desenvolver ferramentas baseadas em IA para identificar e sinalizar *deepfakes*, contribuindo para a mitigação da disseminação de notícias falsas. O objetivo da pesquisa é criar uma ferramenta eficaz para detectar *deepfakes*, utilizando técnicas de IA para analisar vídeos e imagens manipuladas. A metodologia envolveu testes em plataformas de detecção de deepfakes e o desenvolvimento de um protótipo de reconhecimento facial em Python. Os testes foram realizados em plataformas como *BioID Playground*, *Deepware* e *BrandWell*, utilizando vídeos e imagens gerados por *deepfakes* para avaliar a precisão das ferramentas existentes. Simultaneamente, um protótipo inicial foi desenvolvido para reconhecer características faciais e comparar imagens com vídeos em tempo real. Os resultados mostraram que a *BioID Playground* apresentou a maior precisão na identificação de *deepfakes*, enquanto a *Deepware* teve eficácia parcial, dependendo da análise humana. A *BrandWell* foi menos eficaz na detecção de imagens manipuladas. O protótipo de reconhecimento facial em *Python*, utilizando bibliotecas como *OpenCV, Dlib e face_recognition*, mostrou-se promissor para identificação básica de rostos e comparações entre imagens e vídeos, embora ainda precise de melhorias para lidar com *deepfakes* complexos. Em conclusão, o estudo demonstra que ferramentas de IA podem ser valiosas no combate a *deepfakes*, com o protótipo desenvolvido oferecendo uma base sólida para futuras melhorias. A pesquisa destaca a importância de combinar algoritmos avançados com análises humanas para enfrentar os desafios impostos pela desinformação digital. Futuras investigações podem focar na integração de tecnologias mais sofisticadas e na expansão das capacidades de detecção em diferentes condições.

Palavras-chave: Detecção; Reconhecimento; *Python*.

2 INTRODUÇÃO

A pesquisa visa compreender o que são inteligências artificiais, como funcionam e como podem ser usadas no combate às *fake news*, com enfoque naquelas criadas a partir de ferramentas que distorcem e manipulam conteúdos audiovisuais, como, por exemplo, os *deepfakes*.

As *fake news*, ou notícias falsas em português, podem ser definidas como "um documento deliberadamente falso com o objetivo de manipular os consumidores" (Meneses, 2018, p. 47). Essa prática de produzir e veicular conteúdos falsos ocorre desde que Johannes Gutenberg inventou a imprensa, em 1439, quando, de um modo geral, as notícias começaram a ser amplamente distribuídas (Soll, 2016). Porém, com o avanço das tecnologias informacionais e comunicacionais, a partir da década de 1990, as práticas e as discussões acerca daquele fenômeno se popularizaram. Tanto é que o termo *fake news* foi eleito como a palavra do ano em 2017 pelo dicionário inglês da editora britânica *Collins* (BBC, 2017).

Atualmente, o meio no qual a disseminação de *fake news* tem o seu maior alcance são as redes sociais. Nesse sentido, o relatório *Digital News Report*, do *Reuters Institute* (Levy et al., 2017), aponta que, em uma pesquisa realizada com mais de 70 mil pessoas, de 36 países diferentes, apenas 24% dos entrevistados acreditam que as redes sociais são boas em distinguir o que é fato e o que é falso.

Aliados a isso, dados do laboratório especializado em segurança digital da *PSafe, dfndr lab*, revelam que um a cada dois brasileiros pode ter sido um disseminador de notícias falsas (Pecsen, 2020). Conforme Kaufman e Santaella (2020), as crenças individuais são as causadoras de compartilhamentos. Ao disseminar tais notícias, as pessoas buscam convencer e levar outras a acreditarem nelas, sem diferenciarem o que é opinião e o que é conhecimento. Portanto, é perceptível que existe um problema crescente relacionado às *fake news*. Nesse contexto, a evolução das tecnologias de inteligência artificial (IA) desempenha um papel significativo na criação e disseminação dessas notícias falsas, notadamente os *deepfakes*, que são conteúdos (vídeos, imagens, áudios) manipulados por IA, que distorcem a realidade e a percepção de quem os visualiza (Molina, Berenguel, 2022).

Considerando o crescente desafio da disseminação de *fake news* nas redes sociais, uma abordagem promissora para enfrentar esse problema é o uso de inteligência artificial para identificar e sinalizar conteúdo falso, com foco na detecção de *deepfakes*.

Nesse sentido, esse projeto busca responder à seguinte questão: como a IA pode ser eficaz no combate a *fake news* que fazem uso desses conteúdos audiovisuais manipulados? Isso implica entender quais características dos *deepfakes* podem ser identificadas por algoritmos de IA, além de considerar a aplicação da IA para lidar com a disseminação desses conteúdos enganosos.

Além disso, a pesquisa visa avaliar como a implementação bem-sucedida da IA na identificação de *deepfakes* pode melhorar a confiabilidade nas informações compartilhadas nas redes sociais e, com base nos resultados, desenvolver estratégias práticas para combater a disseminação de notícias falsas *online*, dando enfoque a vídeos, áudios e imagens adulteradas por inteligências artificiais.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 *Fake news*

O dicionário *Cambridge* (2018) define *fake news*, um dos termos mais populares da sociedade contemporânea, como histórias falsas, espalhadas pela Internet ou em outros meios, que aparentam ser notícias reais, normalmente criadas para influenciar visões políticas ou como forma de piada (Cambridge Dictionary, 2018). Sem perder de vista essa definição geral, diversos outros autores escreveram sua própria definição, como, por exemplo, Allcott e Gentzkow (2017), que as caracterizaram como artigos de notícias que são, verificável e intencionalmente, falsos, e que tendem a enganar os leitores.

Ou, também, o *Reuters Institute for the Study of Journalism* (Instituto Reuters para o Estudo do Jornalismo), que definiu o termo *fake news* como informações falsas veiculadas com intenções estratégicas específicas - sejam elas políticas ou comerciais - que, normalmente, disfarçam-se como reportagens legítimas, mas que se apoiam em teorias da conspiração ou outros assuntos carregados de apelos emocionais que confirmam crenças pré-existentes (RISJ, 2017).

A partir das definições dadas por diferentes autores, é possível associar, de um modo geral, o termo *fake news* a notícias intencionalmente falsas, mas que foram feitas de modo a parecerem verdadeiras e que usam apetrechos estratégicos para, deliberadamente, enganar os leitores.

3.2 Inteligência artificial

Em 1950, Alan Turing, matemático, cientista da computação, filósofo e considerado o pai da computação, definiu o termo inteligência artificial pelo teste que leva o seu nome. Essa avaliação possui uma premissa simples: caso um ser humano converse com uma máquina por cinco minutos sem perceber que é uma máquina, o computador passa no teste, sendo considerado uma IA.

O teste de Turing levou muitos engenheiros e cientistas a raciocinarem sobre o que nos faz realmente humanos, visto que durante a aplicação dos testes muitas pessoas que tentavam decifrar se a mensagem era emitida por um humano ou computador eram enganadas pelos sistemas que propositalmente enviavam textos com erros de digitação e não necessariamente apresentavam uma “conversa coerente”.

Segundo Teixeira (2019), a IA pode ser definida como “produção de comportamento inteligente”. Portanto, podemos considerar que a IA é qualquer sistema computacional ensinado a simular comportamentos humanos inteligentes. Hoje em dia, vários programas conversam com milhões de pessoas em todo o mundo.

3.3 *Deepfakes*

O termo *deepfake* surgiu em 2017, após um usuário da plataforma *"Reddit"* postar vídeos com conteúdos pornográficos utilizando as imagens faciais de celebridades (Hall, 2018). De acordo com Molina e Berenguel (2022), os *deepfakes* são conteúdos, incluindo vídeos, imagens e áudios, modificados por meio de ferramentas que utilizam inteligências artificiais, que têm a capacidade de distorcer a realidade e influenciar as pessoas que os visualizam. Representam avanços tecnológicos quando se trata da produção de *fake news* e são amplamente disseminados nas redes sociais. Além disso, o avanço das inteligências artificiais também possibilita a produção em massa de *fake news* em formato de texto, agravando ainda mais a propagação desse tipo de desinformação nas redes sociais.

O uso da inteligência artificial na edição de fotografias, áudios e vídeos pode envolver fraudes e, conforme a tecnologia relacionada aos *deepfakes* avança, os conteúdos ficam mais precisos, tornando-se mais difícil distinguir se são *fake news* ou não (Molina, Berenguel, 2022).

3.4 Combate à desinformação produzida por meio de IA

Com a rápida evolução da tecnologia, o combate a *deepfakes* e outras formas de desinformação se torna um desafio crescente. A legislação em vários países tem se adaptado para lidar com essas novas ameaças, buscando criar um ambiente mais seguro contra conteúdos manipuladores. A crescente frequência de vazamentos de dados, aliada ao uso de Inteligência Artificial (IA) por *hackers*, transformou a forma como ataques cibernéticos são realizados. Como mencionado por DeMello (2022), *hackers* utilizam IA para entender o comportamento dos usuários *online*, potencializando suas ações maliciosas.

Contudo, com a evolução da IA, existem ferramentas por meio das quais se pode identificar e denunciar as notícias falsas, combatendo-as. A busca por novas tecnologias no combate às *fake news* pode trazer diversas oportunidades para empresas produzirem soluções, até mesmo para a detecção de *deepfakes* (Westerlund, 2019).

Para enfrentar esses desafios, surgem também ferramentas de identificação e denúncia de notícias falsas. A inovação nesse campo oferece soluções para o reconhecimento de *deepfakes*, como por exemplo a empresa *"Sensity"*, que desenvolveu uma plataforma que possibilita identificá-los. Dessa forma, o usuário envia um arquivo nos formatos mp4, mov, png, jpeg ou tiff, ou insere uma URL de um vídeo *online*, e a plataforma analisa e verifica se o arquivo enviado é um *deepfake*.

A necessidade de soluções eficazes no combate a *deepfakes* é mais urgente do que nunca. À medida que as tecnologias de criação de conteúdo se tornam mais acessíveis e sofisticadas, o desenvolvimento de ferramentas de verificação e detecção deve acompanhar esse ritmo.

4 MATERIAIS E MÉTODOS

Essa pesquisa partiu de revisão bibliográfica e, em seguida, iniciou testes em algumas plataformas disponíveis gratuitamente na Internet, cuja finalidade é identificar *deepfakes*. Os testes basearam-se na análise da eficiência desses sites em detectar corretamente conteúdos produzidos e modificados por inteligência artificial. Para isso, vídeos e imagens produzidos por meio de *deepfakes* foram submetidos à avaliação desses sites. Com a análise dos resultados dos testes, buscou-se verificar se as plataformas já existentes na Internet, e que são semelhantes àquela objetivada pela pesquisa, são realmente eficazes na identificação de conteúdos fraudados com o uso de inteligência artificial, mais comumente conhecidos como *deepfakes*.

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*. Os três sites fazem a identificação de conteúdos manipulados com o auxílio de IA através de APIs (*Application Programming Interface*), que são como uma espécie de contrato de serviço, que une duas aplicações com funções distintas (AWS, 2023) .

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. O vídeo 1 apresenta um *deepfake* utilizando o rosto e a voz do ator Morgan Freeman (Diep Niep, 2021); o vídeo 2 mostra um conteúdo manipulado com o rosto e voz do empresário Elon Musk (Civic Sentinel, 2022); o vídeo 3 coloca o comediante Jerry Seinfeld dentro do filme *"Pulp Fiction"*, do qual ele nunca fez parte (Desifakes, 2022); o vídeo 4 é um *deepfake* do ator Tom Cruise (Vecanoi, 2021) e o vídeo 5 mostra o ator e ex-governador da Califórnia, Arnold Schwarzenegger, como se ele fizesse parte do filme "Mágico de Oz", do qual ele não participou (Brianmonarch, 2023). Já as imagens foram todas retiradas do site *This Person Does Not Exist* (Essa pessoa não existe), que gera, por meio de uma IA, rostos humanos novos a cada vez que a página é recarregada.

Em relação às plataformas, a *BioID Playground* foi utilizada para identificar tanto os vídeos, quanto as imagens. Ao enviar um arquivo para o site, ele retorna uma mensagem, que pode ser "Esse(a) vídeo/imagem foi feito(a) por uma pessoa real" ou "Esse(a) vídeo/imagem é falso(a)" e, além disso, retorna uma espécie de pontuação entre 0 e 1, baseada na análise da API, sendo 0 um conteúdo indubitavelmente falso e manipulado, e 1, um conteúdo totalmente verídico.

Já a plataforma *Deepware* foi utilizada apenas para a detecção dos vídeos. Esse site utiliza quatro APIs para identificação de *deepfakes*, porém ele também possui a função de denúncia de vídeos falsos. Ao serem

denunciados, os vídeos passam por um analista humano e, caso sejam realmente identificados como tal, toda vez que qualquer usuário enviar o mesmo vídeo, ele já será informado sobre a sua procedência, mesmo que as APIs analisem de uma maneira diferente. Ao enviar um arquivo para o site, ele pode retornar três mensagens diferentes, que são "*Deepfake* detectado", "Nenhum *Deepfake* detectado" ou "Suspeito"; além disso, cada API identifica a probabilidade de o vídeo ser falso. Por fim, a plataforma *BrandWell* foi utilizada apenas para a identificação das imagens. Ao enviar um arquivo para o site, ele retorna apenas a probabilidade de aquela imagem ter sido fraudada por IA.

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*. Após uma pesquisa exploratória, foram identificadas diversas linguagens de programação que poderiam ser utilizadas, como *Python, C++, Java, MATLAB, C#, JavaScript e Ruby*. No entanto, a *Python* foi escolhida por ser uma linguagem de fácil aprendizado e por possuir uma vasta gama de bibliotecas e *frameworks* que facilitam o desenvolvimento de algoritmos de aprendizado de máquina e processamento de imagens. Conforme descrito por Raschka, Patterson e Nolet (2020):

> Com o seu foco principal na legibilidade, Python é uma linguagem de programação interpretada de alto nível, que é amplamente reconhecida por ser fácil de aprender, mas ainda capaz de aproveitar o poder das linguagens de programação de nível de sistema quando necessário. Para além dos benefícios da própria linguagem, a comunidade em torno das ferramentas e bibliotecas disponíveis tornam o Python particularmente atrativo para aplicações nas áreas de ciência de dados, machine learning e computação científica (Raschka, Patterson e Nolet, 2020, p.2).[1]

Na fase inicial da segunda etapa, o intuito principal foi aprender e explorar as funcionalidades da linguagem *Python*. Para essa fase inicial, utilizamos o editor *Visual Studio Code*, instalando o pacote *PIP* para gerenciar as bibliotecas necessárias, e a biblioteca *OpenCV*, que fornece ferramentas para processamento de imagens.

O reconhecimento facial é um tipo de biometria que estuda as características únicas das pessoas. Assim como a digital ou a íris, o rosto possui detalhes únicos, como o formato da testa, o tamanho do nariz e a distância entre os olhos, que são usados para criar uma "assinatura facial". O sistema captura a imagem ou o vídeo do rosto de uma pessoa, transforma essas características em números e os compara com um banco de dados para identificar se há uma correspondência.

[1] Tradução realizada do inglês para o português do Brasil pelos autores.

Para o desenvolvimento do protótipo, foram feitas 3 versões, com diferentes resultados que serão analisados na próxima seção.

5 RESULTADOS

Em relação aos testes de eficiência dos sites analisados, apresentamos a identificação dos vídeos e imagens na Tabela 1 e os resultados dos testes nas plataformas *BioID Playground*, *Deepware* e *BrandWell* nas Tabelas 2, 3 e 4, respectivamente. Além disso, o resultado geral das plataformas e a porcentagem de acertos de cada uma foram apresentados na tabela 5. Para facilitar a análise, a pontuação do site *BioID Playground* foi transformada em porcentagem e a porcentagem da plataforma *Deepware* foi obtida a partir das médias informadas pelas quatro APIs presentes no site.

Tabela 1. Identificação de vídeos e imagens utilizados nos testes

	Descrição
Vídeo 1	*Deepfake* do ator Morgan Freeman (Diep Niep, 2021)
Vídeo 2	*Deepfake* do empresário Elon Musk (Civic Sentinel, 2022)
Vídeo 3	*Deepfake* do ator Jerry Seinfeld, como se ele fizesse parte do filme "*Pulp Fiction*" (Desifakes, 2022)
Vídeo 4	*Deepfake* do ator Tom Cruise (Vecanoi, 2021)
Vídeo 5	*Deepfake* do ator e ex-governador da Califórnia, Arnold Schwarzenegger, como se ele fizesse parte do filme "O mágico de Oz" (Brianmonarch, 2023)
Imagem 1	Pessoa fictícia 1, retirada do site *This Person Does Not Exist*
Imagem 2	Pessoa fictícia 2, retirada do site *This Person Does Not Exist*
Imagem 3	Pessoa fictícia 3, retirada do site *This Person Does Not Exist*
Imagem 4	Pessoa fictícia 4, retirada do site *This Person Does Not Exist*
Imagem 5	Pessoa fictícia 5, retirada do site *This Person Does Not Exist*

Elaborado pelos autores.

Tabela 2. Resultados dos testes na plataforma *BioID Playground*

Conteúdo	Mensagem exibida	Probabilidade informada para conteúdo falso	Analisou corretamente?
Vídeo 1	"Esse vídeo é falso!"	99,9%	SIM
Vídeo 2	"Esse vídeo é falso!"	97,4%	SIM
Vídeo 3	"Esse vídeo foi gravado por uma pessoa real!"	44,6%	NÃO
Vídeo 4	"Esse vídeo é falso!"	86,2%	SIM
Vídeo 5	"Esse vídeo foi gravado por uma pessoa real!"	32%	NÃO
Imagem 1	"Essa imagem é falsa!	99,9%	SIM
Imagem 2	"Essa imagem é falsa!	99,9%	SIM
Imagem 3	"Essa imagem é falsa!	92,9%	SIM
Imagem 4	"Essa imagem é falsa!	99,9%	SIM
Imagem 5	"Essa imagem é falsa!"	99,9%	SIM

Elaborado pelos autores.

Tabela 3. Resultados dos teste na plataforma *Deepware*

Conteúdo	Mensagem exibida	Probabilidade informada para conteúdo falso	Analisou corretamente?
Vídeo 1	"*Deepfake* detectado!"	1%[2]	SIM
Vídeo 2	"*Deepfake* detectado!"	24,5%	SIM
Vídeo 3	"Nenhum *Deepfake* detectado!"	8,5%	NÃO
Vídeo 4	"*Deepfake* detectado!"	1,2%	SIM
Vídeo 5	"Suspeito"	54,5%	NÃO

Elaborado pelos autores.

Tabela 4. Resultados dos testes na plataforma *BrandWell*

Conteúdo	Mensagem exibida	Probabilidade informada para conteúdo falso	Analisou corretamente?
Imagem 1	Não exibe mensagens	15%	NÃO
Imagem 2	Não exibe mensagens	13%	NÃO
Imagem 3	Não exibe mensagens	57%	SIM
Imagem 4	Não exibe mensagens	6%	NÃO
Imagem 5	Não exibe mensagens	22%	NÃO

Elaborado pelos autores.

Tabela 5. 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)

Elaborado pelos autores.

[2] A incongruência entre a probabilidade para conteúdo falso e a mensagem exibida é explicada pelas funcionalidades de denúncia e análise de especialistas humanos, presentes na plataforma *Deepware* e explicadas na seção anterior.

Analisando os resultados dos testes, é possível concluir que a plataforma *BioID Playground* é a que possui a maior eficiência, tendo errado apenas na identificação dos vídeos 3 e 5. Em relação à *Deepware*, o site se mostrou parcialmente eficiente, já que, quando identificou corretamente os vídeos, foi por conta dos analistas humanos presentes na plataforma e não por conta das quatro APIs utilizadas. Já a *BrandWell* se mostrou como a mais ineficiente das três, tendo identificado, erroneamente, quatro imagens de cinco. Para a plataforma em desenvolvimento, a junção de um algoritmo com bons resultados, como o da *BioID Playground*, com as funcionalidades de denúncias de vídeos e análises feitas por seres humanos, como as da *Deepware*, parece ser o caminho a ser seguido.

Na etapa de produção do nosso protótipo, durante a primeira fase o código foi testado com sucesso em 14 tipos de fotos, variando em iluminação, rostos, ambientes e distâncias. Além de utilizar imagens estáticas, o sistema também foi capaz de identificar as regiões dos olhos, nariz e boca em vídeo em tempo real através da *webcam*.

Na segunda fase dessa etapa, o projeto foi expandido para incluir a capacidade de comparar duas imagens e verificar se elas pertencem à mesma pessoa. Para essa fase, além do *OpenCV* já utilizado na primeira versão, foram instaladas as bibliotecas *`face_recognition`* e *`Dlib`* pelo terminal, utilizando os comandos *`pip install face_recognition`* e *`pip install dlib`*, respectivamente, e a ferramenta *`CMake`* com o comando *`pip install cmake`*.

A biblioteca *`Dlib`* é conhecida mundialmente por realizar a detecção de faces e foi desenvolvida em *C++*. Sua principal característica é a capacidade de identificar *"landmarks"*, que são pontos de referência usados para reconhecer padrões faciais. Já a biblioteca *`face_recognition`* é um pacote de manipulação e reconhecimento de rostos que utiliza algoritmos de aprendizado profundo e apresenta uma precisão de 99,38% no *benchmark Labeled Faces in the Wild* (Rostos Etiquetados na Natureza). O *`CMake`* é uma ferramenta de construção multiplataforma que facilita o desenvolvimento e a implementação de projetos complexos.

A terceira fase e, até o presente momento, a versão mais atual do protótipo, é uma melhoria da segunda, em que o sistema tornou-se capaz de comparar uma imagem estática com um vídeo em tempo real capturado pela *webcam*. Para o desenvolvimento dessa versão, foram utilizadas as bibliotecas *`OpenCV`* e a classe *`SimpleFacerec`*, que permite carregar imagens, detectar rostos e codificar essas informações para o reconhecimento facial.

O código dessa versão funciona importando as bibliotecas necessárias e configurando a *webcam* para captura de vídeo. Durante a execução, o sistema lê um *frame* (quadro) da *webcam* e utiliza o método "*detect_known_faces*" (detectar faces conhecidas) para encontrar rostos no *frame* atual. Ele delimita o formato do rosto e exibe o nome do arquivo de imagem, para o qual se recomenda escolher o nome da pessoa na foto, sobre o rosto detectado na *webcam*. Ao final, o comando "*video_capture.release*()" libera o recurso da câmera.

Assim, no estágio atual, o protótipo em desenvolvimento consegue identificar a pessoa cujo rosto está aparecendo na *webcam*, a partir de um banco de imagens. Apesar de ainda não ser possível identificar *deepfakes*, o algoritmo em *Python* já é capaz de analisar e fazer a diferenciação entre um rosto humano e outro, o que se mostra como um importante passo inicial para uma futura ferramenta que consiga, eficazmente, identificar *deepfakes*, combatendo, assim, notícias falsas criadas com o uso dessas tecnologias.

6 CONCLUSÕES

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.

Do ponto de vista econômico, a implementação de um sistema como este é acessível, pois os recursos computacionais necessários, como *webcams* e computadores de médio porte, são comuns. No entanto, para a produção em larga escala, seria necessário considerar custos adicionais relacionados à infraestrutura de armazenamento e processamento de dados em tempo real, especialmente para grandes volumes de vídeo.

As aplicações desse sistema são diversas e abrangem áreas que vão desde segurança pública, autenticação de identidade e monitoramento de acessos até o combate a *deepfakes*. A proposta inicial do projeto "Aplicações da Inteligência Artificial no Combate às *Fake News* produzidas por meio de *Deepfakes*" visa utilizar o reconhecimento facial para comparar rostos em vídeos com fotos reais e verificar se eles pertencem à mesma pessoa, identificando assim manipulações fraudulentas.

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 esse 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 iniciais 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.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ALLCOTT, H.; GENTZKOW, M. Social media and fake news in the 2016 election. **The journal of economic perspectives: a journal of the American Economic Association**, v. 31, n. 2, p. 211–236, 2017.

AWS. O que é uma API (interface de programação de aplicações), 2023. Disponível em: https://aws.amazon.com/pt/what-is/api/#seo-faq-pairs#what-is-an-api. Acesso em: 09 set. 2024.

BBC. 'Fake News' é eleita palavra do ano e ganhará menção em dicionário britânico, 2017. Disponível em: https://www.bbc.com/portuguese/internacional-41843695. Acesso em: 09 set. 2024.

BRIANMONARCH. Arnold Schwarzenegger Sings About Rainbows. **YouTube**, 18 de maio de 2023. Disponível em: <https://youtu.be/K_XwseDwmuQ>. Acesso em: 09 set. 2024.
CIVIC Sentinel. Ultra realistic Deepfake of Elon Musk. **YouTube**, 22 de dezembro de 2022. Disponível em: <https://youtu.be/XuKUkyPegBE>. Acesso em: 08 set. 2024.

DEMELLO, M. **Inteligência Artificial na cibersegurança:** do combate aos ataques ao fim das senhas, 2022. Disponível em: https://www.psafe.com/blog/inteligencia-artificial-na-cibersegurança/.

DESIFAKES. Jerry Seinfeld in Pulp Fiction [DeepFake]. **YouTube**, 6 de fevereiro de 2022. Disponível em: <https://youtu.be/S1MBVXkQbWU>. Acesso em: 09 set. 2024.

DIEP Niep. This is not Morgan Freeman - A Deepfake Singularity. **YouTube**, 7 de julho de 2021. Disponível em: <https://youtu.be/oxXpB9pSETo>. Acesso em: 07 set. 2024.

FAKE news. **Cambridge Dictionary**, 2019. Disponível em: https://dictionary.cambridge.org/dictionary/english/fake-news. Acesso em: 06 out. 2024.

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. Acesso em: 09 set. 2024.

HOW can we combat fake news? – The role of platforms, media literacy, and journalism. **RISJ**, 24 mar. 2017. Disponível em:

https://reutersinstitute.politics.ox.ac.uk/news/how-can-we-combat-fake-news-role-platforms-media-literacy-and-journalism. Acesso em: 06 out. 2024.

KAUFMAN, D.; Santaella, L. O papel dos algoritmos de inteligência artificial nas redes sociais. **Revista FAMECOS**, [S. l.], v. 27, n. 1, p. e34074, 2020. DOI: 10.15448/1980-3729.2020.1.34074. Disponível em: https://revistaseletronicas.pucrs.br/ojs/index.php/revistafamecos/article/view/34074. Acesso em: 08 set. 2024.

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. Acesso: 09 set. 2024.

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/c2ce/e77a45ef6d79bf2c8b941c52a0476051334f.pdf. Acesso em: 09 set. 2024.

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. DOI: 10.33448/rsd-v11i6.29533. Disponível em: https://rsdjournal.org/index.php/rsd/article/view/29533. Acesso em: 07 set. 2024.

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/. Acesso em: 04 set. 2024.

RASCHKA, Sebastian; Patterson, Joshua; Nolet, Corey. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. **Information**, v. 11, n. 4, p. 193, 2020.

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/. Acesso: 09 set. 2024.

TEIXEIRA, J. O que é inteligência artificial. [s.l.] **E-Galáxia**, 2019.

VECANOI. Very realistic Tom Cruise Deepfake | AI Tom Cruise. **YouTube**, 28 de fevereiro de 2021. Disponível em: <https://youtu.be/iyiOVUbsPcM>. Acesso em: 09 set. 2024.

WESTERLUND, M. The Emergence of Deepfake Technology: A Review. **Technology Innovation Management Review**, v. 9, n. 11, 2019. Disponível em: https://timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf. Acesso em: 06 out. 2024.