Deepfakes além do algoritmo: uma revisão sobre indícios visuais identificáveis a olho nu na era da inteligência artificial generativa

Camily Cantão[1], **Iago Medeiros**[1]

[1]Universidade Federal do Pará (UFPA)

Abstract. *This article analyzes the visual detection of deepfake videos with the naked eye, focusing on human perception as a tool to aid in the increasingly sophisticated image and video generation techniques. It addresses the limitations of current automatic detectors and explores, based on the literature, recurring visual cues in faked videos, such as the absence of natural blinks, artificial expressions, exaggerated fluidity, homogeneous skin texture, and inconsistencies between the figure and background. A practical example with artificial intelligence (AI)-generated video complements the analysis. The findings reinforce that, despite high aesthetic realism, visual flaws can still be perceived by attentive observers. It concludes that critical and visual observation are viable avenues for combating audiovisual disinformation.*

Resumo. *Este artigo realiza uma análise sobre a detecção visual de vídeos deepfake a olho nu, com foco na percepção humana como ferramenta auxiliar frente à crescente sofisticação das técnicas de geração de imagens e vídeos. Aborda-se a limitação dos detectores automáticos atuais e explora-se, com base na literatura, sinais visuais recorrentes em vídeos falsificados, como ausência de piscadas naturais, expressões artificiais, fluidez exagerada, textura de pele homogênea e inconsistências entre figura e fundo. Um exemplo prático com vídeo gerado por inteligência artificial (IA) complementa a análise. Os achados reforçam que, apesar do alto realismo estético, falhas visuais ainda podem ser percebidas por observadores atentos. Conclui-se que a observação crítica e visual são caminhos viáveis no combate à desinformação audiovisual.*

1. Introdução

Nos últimos anos, a ascensão das tecnologias de inteligência artificial generativa tornou possível a criação de vídeos ultrarrealistas manipulados digitalmente, conhecidos como deepfakes [Tolosana et al. 2020]. Esses conteúdos são produzidos por modelos como as *Generative Adversarial Networks (GANs)*, capazes de sintetizar rostos humanos, imitar expressões, movimentos e até vozes com alto grau de realismo. Com a popularização de ferramentas gratuitas como *DeepFaceLab*[1] e *Zao*[2], a geração desses vídeos se tornou amplamente acessível, intensificando preocupações sobre seu uso em fraudes, desinformação e manipulações sociais [Lee 2019].

Em resposta a essa ameaça, diversas abordagens automatizadas vêm sendo propostas, com destaque para algoritmos de aprendizado de máquina, análise de ruído residual e detecção de padrões espectrais [Guarnera et al. 2020]. No entanto, esses métodos

[1]https://github.com/iperov/DeepFaceLab

[2]ZAO - app desativado/descontinuado

enfrentam limitações críticas: sua eficácia tende a diminuir diante de modelos generativos mais recentes e, além disso, sua complexidade e custo os tornam inacessíveis ao público geral. Segundo [Diel et al. 2024], a acurácia média humana na detecção de deepfakes é de apenas 55%, o que equivale ao acaso, indicando a dificuldade crescente de distinguir vídeos reais de manipulados.

Apesar disso, cresce o interesse por abordagens centradas na percepção humana, que buscam explorar indícios visuais sutis que escapam à manipulação perfeita. Pesquisas baseadas em *eye-tracking*, como o banco de dados *FakeET* [Gupta et al. 2020], revelam que observadores tendem a fixar a visão em regiões-chave do rosto, como olhos e boca, ao tentar julgar a autenticidade de um vídeo. Além disso, estudos recentes demonstram que a amplificação visual de artefatos pode aumentar significativamente a capacidade de percepção humana [Josephs et al. 2023], reforçando o potencial da análise visual como ferramenta complementar de detecção.

Diante deste contexto, este artigo propõe uma revisão crítica da literatura com foco nos principais sinais visuais que podem ser percebidos a olho nu em vídeos deepfake como expressões artificiais, fluidez exagerada, inconsistências no fundo, ausência de piscadas, textura da pele homogênea e sincronização labial imperfeita. Ao reunir essas evidências e aplicá-las a um exemplo gerado por IA, busca-se destacar o papel da observação humana como recurso acessível e relevante na identificação de manipulações visuais, especialmente em contextos onde o uso de ferramentas automáticas é inviável.

2. Trabalhos Relacionados

A crescente sofisticação dos vídeos gerados por inteligência artificial tem ampliado os desafios para a detecção de conteúdos manipulados. Como resposta, a literatura tem se dividido entre abordagens automatizadas baseadas em redes neurais convolucionais e propostas que investigam os limites da percepção humana frente aos deepfakes. Este artigo se insere na segunda linha, buscando compreender quais características visuais ainda permitem que um observador identifique manipulações sem o auxílio de ferramentas computacionais.

Entre os trabalhos mais relevantes está o estudo de [Gupta et al. 2020], que realizou uma metanálise de 56 publicações sobre a capacidade humana de identificar vídeos falsos. Os autores apontam que, mesmo diante de modelos avançados, a percepção humana ainda consegue detectar inconsistências em elementos como expressões faciais, rigidez no movimento, olhar vazio e sorrisos artificiais, confirmando a importância da análise visual como ferramenta complementar aos detectores automáticos. O estudo reforça ainda que o treinamento da atenção visual em regiões como os olhos, a boca e os limites do rosto pode aumentar significativamente a acurácia da detecção por humanos.

Por outro lado, [Josephs et al. 2023] exploraram o uso de amplificação de artefatos (*artifact magnificatio*n) como ferramenta para melhorar a percepção de alterações visuais sutis. Seus experimentos demonstraram que, ao tornar mais evidentes as falhas geradas durante a síntese do vídeo como transições borradas, fluidez artificial e iluminação incoerente os participantes passaram a identificar os deepfakes com maior precisão e confiança. A pesquisa mostra que os erros, embora muitas vezes mascarados em uma primeira visualização, continuam presentes e podem ser potencializados por métodos visuais ou simplesmente percebidos por observadores mais experientes.

Apesar disso, cresce o interesse por abordagens centradas na percepção humana, focando nos indícios visuais que ainda escapam à manipulação perfeita. Estudos como o de [Korshunov and Marcel 2023] mostram que, mesmo em vídeos com alta resolução, ainda é possível identificar falhas visuais como movimentos labiais incoerentes, olhos que não piscam naturalmente, falta de expressividade emocional, iluminação artificial, saturação exagerada e fluidez de movimentos comprometida. Tais sinais, embora sutis, podem ser percebidos por observadores atentos, e tornam-se especialmente relevantes em contextos onde o uso de ferramentas automáticas de detecção é inviável.

Por fim, [Tolosana et al. 2020] detalham as dificuldades das redes neurais em preservar regiões dinâmicas da face, como testa, sobrancelhas e contornos dos olhos, que exigem ajustes contínuos em tempo real para parecerem naturais. Isso faz com que vídeos gerados por IA apresentem inconsistências locais mesmo quando o todo parece coeso. A iluminação, a nitidez e a coerência espacial entre rosto e fundo também são frequentemente comprometidas, oferecendo aos observadores atentos sinais valiosos para inferir a inautenticidade do conteúdo.

Esses trabalhos, em conjunto, apontam para a importância de considerar a percepção humana como ferramenta válida e ainda insubstituível na detecção de deepfakes. Ao mesmo tempo, demonstram que, apesar do alto grau de realismo das produções atuais, ainda existem fragilidades sistemáticas nos rostos e movimentos gerados por IA que escapam à perfeição algorítmica, especialmente quando se leva em conta o comportamento facial natural de seres humanos. Este trabalho propõe uma abordagem alternativa centrada na observação humana. A ideia é investigar como sinais visuais perceptíveis a olho nu podem ser sistematicamente utilizados na identificação de vídeos falsificados, considerando que características como expressões faciais incoerentes, ausência de piscadas naturais ou distorções em elementos do rosto ainda escapam à síntese perfeita promovida por IA. Ao reunir e analisar criticamente esses indícios, busca-se demonstrar que a percepção visual, embora subjetiva, pode representar um recurso complementar importante no combate à desinformação digital.

3. Sinais Visuais Perceptíveis a Olho Nu em Vídeos Deepfake

Mesmo com os avanços das redes generativas e o alto nível de realismo alcançado por vídeos sintéticos, ainda persistem indícios visuais que escapam à manipulação perfeita e podem ser percebidos a olho nu. Neste contexto, torna-se relevante examinar sinais como a incoerência na articulação labial, rigidez ou ausência de piscadas, expressões faciais pouco naturais, distorções no fundo, saturação exagerada, fluidez artificial nos movimentos e ausência de envolvimento emocional. Estes elementos, ainda que sutis, representam uma alternativa viável de análise quando ferramentas automáticas não estão disponíveis ou apresentam limitações.

3.1. Expressões faciais e trejeitos

Um dos principais fatores que comprometem a naturalidade dos vídeos deepfake é a ausência de expressões faciais orgânicas. Microexpressões como contrações rápidas de músculos ao redor da boca, olhos e testa são frequentemente suprimidas ou uniformizadas por modelos geradores, devido à dificuldade de replicar com precisão os detalhes motores sutis do rosto humano [Josephs et al. 2023]. Isso gera sorrisos que não afetam as

regiões orbitais (ao redor dos olhos), e expressões emocionais que parecem "ensaiadas" ou desprovidas de profundidade emocional.

Segundo [Diel et al. 2024], mesmo quando a movimentação geral do rosto é convincente, a ausência de variação emocional ao longo do tempo pode causar desconforto ou desconfiança no espectador. Este fenômeno é frequentemente associado ao "vale da estranheza" (*uncanny valley*), conceito que descreve a rejeição instintiva do ser humano frente a representações quase humanas, mas não plenamente naturais.

3.2. Movimentos dos olhos e piscadas

A movimentação ocular é um dos aspectos mais sensíveis da comunicação não verbal. Deepfakes frequentemente falham em replicar a dinâmica dos olhos de forma crível: os olhos podem parecer fixos em um ponto, movimentar-se em sincronia artificial ou exibir piscadas espaçadas demais ou ausentes. Esses elementos, que normalmente passam despercebidos em interações reais, tornam-se evidentes quando ausentes.

[Gupta et al. 2020], por meio do banco de dados FakeET, demonstraram que observadores humanos tendem a fixar o olhar em regiões como olhos e boca durante a visualização de vídeos. Essa concentração revela a importância desses pontos para a percepção de autenticidade. Quando os olhos em um vídeo não acompanham o ambiente ou expressam pouca variação de foco, o resultado é uma percepção de frieza ou inexpressividade muitas vezes descrita como "olhar morto".

3.3. Sincronização labial

A relação entre fala e movimento labial em deepfakes, embora cada vez mais refinada, ainda carrega inconsistências sutis. Algumas redes generativas produzem movimentos labiais suavizados demais, sem a tensão muscular típica da articulação de fonemas fortes como **/b/**, **/p/** ou **/m/**. Isso resulta em uma leve dissociação entre o som e o movimento: perceptível especialmente em vídeos com boa qualidade de áudio.

[Diel et al. 2024] apontam que essa sincronia artificial contribui para uma detecção intuitiva por parte do espectador. Além disso, os movimentos da mandíbula, bochechas e queixo geralmente acompanham a fala em vídeos reais, o que nem sempre ocorre em vídeos gerados por IA, que tendem a animar apenas a região labial com pouca ou nenhuma influência nas demais partes do rosto.

3.4. Textura da pele e nitidez

As redes generativas, mesmo as mais avançadas, frequentemente produzem uma pele com textura homogênea, poros pouco definidos e iluminação difusa, semelhante a um filtro de suavização. Essa suavização pode ser suficiente para enganar a percepção superficial, mas, ao ser observada em detalhes, revela a ausência de imperfeições naturais como rugas dinâmicas, pequenas manchas, oleosidade ou variação na tonalidade da pele.

[Tolosana et al. 2020] demonstram que a região da testa e ao redor do nariz são áreas particularmente sensíveis para identificar esse tipo de falha, já que são locais de constante movimentação muscular e acúmulo de textura em vídeos reais. Além disso, a transição entre rosto e cabelo também tende a apresentar artefatos ou borrões, especialmente durante movimentos bruscos.

3.5. Fundo e profundidade de campo

Um dos elementos mais reveladores em vídeos falsificados é a relação entre o rosto e o fundo da imagem. Em muitos casos, observa-se um desfoque artificial no fundo, semelhante a efeitos de "modo retrato", mas com bordas excessivamente suaves ou instáveis. Isso ocorre devido à segmentação automática da figura humana feita por alguns modelos de geração, que têm dificuldade em preservar a coerência visual entre o plano principal e o fundo.

Em sua análise de artefatos visuais, [Korshunov and Marcel 2023] destacam que fundos pouco naturais ou mal integrados ao corpo principal indicam manipulação. Além disso, a iluminação frequentemente não se distribui de maneira coerente entre o rosto e o ambiente, gerando sombras ausentes ou contraditórias.

3.6. Saturação, iluminação e fluidez dos movimentos

A saturação excessiva, especialmente em tons de pele, cabelo ou roupas, pode ser um subproduto da tentativa dos algoritmos de tornar a imagem mais "viva". No entanto, isso resulta em uma coloração que, apesar de agradável à primeira vista, parece artificial ao olhar técnico. A iluminação costuma ser distribuída de forma uniforme e sem fonte clara outra pista visual de que a cena não é real.

Além disso, os movimentos do rosto e do corpo em vídeos deepfake tendem a ser suavizados por interpolação digital, o que gera fluidez exagerada. [Josephs et al. 2023] observam que esse tipo de suavização remove os microtremores e hesitações naturais do corpo humano, resultando em gestos contínuos demais, quase mecânicos.

4. Avaliação

Realizamos uma análise exploratória de um vídeo e uma imagem gerados pela nova plataforma de criação audiovisual baseada em inteligência artificial do Google, a Veo 3[3]. O objetivo do teste foi observar, na prática, a ocorrência de sinais visuais artificiais descritos na literatura, a fim de ilustrar como esses elementos ainda podem ser percebidos a olho nu, mesmo em conteúdos de alta qualidade estética[4].

Figura 1. Foto tirada do video gerado pelo Google VEO3.

[3]https://veo3.ai/dashboard
[4]http://bit.ly/4kQLN4W

Durante a análise da do vídeo, observou-se uma composição visual bastante convincente à primeira vista. No entanto, detalhes como o olhar fixo e emocionalmente neutro, o sorriso dissociado dos músculos oculares, a textura da pele suavizada e a ausência de poros ou imperfeições revelaram características típicas de manipulação. A separação entre figura e fundo também se mostrou artificial, com desfoque excessivo e contornos suavizados demais, sugerindo o uso de segmentação automática.

Já no vídeo correspondente, foi possível identificar movimentos faciais excessivamente fluidos, com pouca variação de expressividade, além de piscadas espaçadas e sincronização labial suavizada. A fala, embora tecnicamente bem alinhada ao áudio, carecia de tensão muscular compatível com a articulação fonética real. Esses indícios refletem exatamente os pontos descritos em trabalhos como os de [Gupta et al. 2020][Josephs et al. 2023] [Diel et al. 2024], reforçando a ideia de que mesmo os modelos de última geração ainda apresentam fragilidades detectáveis visualmente por humanos atentos.

5. Conclusão

A análise visual de vídeos deepfake, mesmo diante de tecnologias cada vez mais avançadas, ainda se mostra uma via promissora e necessária para a detecção de manipulações audiovisuais, especialmente em contextos onde ferramentas automáticas são inacessíveis ou ineficazes. Este artigo, ao reunir os principais estudos sobre percepção humana diante de vídeos sintéticos e ilustrar sinais visuais como rigidez facial, olhar inexpressivo, fluidez artificial e inconsistência de fundo, reforça que a observação atenta permanece uma ferramenta poderosa contra a desinformação. Ao contrário do que sugerem os avanços algorítmicos, a imperfeição continua presente e visível nos detalhes. Resta à sociedade, portanto, fomentar a literacia visual e a conscientização crítica como formas de resistência diante da crescente sofisticação da falsificação digital.

Referências

Diel, A., Lalgi, T., Schröter, I. C., MacDorman, K. F., Teufel, M., and Bäuerle, A. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. *Computers in Human Behavior Reports*, 16:100538.

Guarnera, L. et al. (2020). A deepfake video detection method based on convolutional neural networks and frequency analysis. *Sensors*, 20(18):5112.

Gupta, P., Chugh, K., Dhall, A., and Subramanian, R. (2020). The eyes know it: Fakeet – an eye-tracking database to understand deepfake perception.

Josephs, E., Fosco, C., and Oliva, A. (2023). Artifact magnification on deepfake videos increases human detection and subjective confidence.

Korshunov, P. and Marcel, S. (2023). Survey of visual artifacts in deepfake videos. *arXiv preprint arXiv:2311.10824*.

Lee, A. (2019). How puny humans can spot devious deepfakes. https://www.wired.com/story/how-to-spot-deepfake-video/. Acesso em: Julho 2025.

Tolosana, R., Romero-Tapiador, S., Fierrez, J., and Vera-Rodriguez, R. (2020). Deepfakes evolution: Analysis of facial regions and fake detection performance.