INTERFACE TECNOLÓGICA

TECNOLOGIA EM INFORMÁTICA
Interface Tecnológica · v. 23, n. 1, 2026 · p. 52–58
ISSN on-line 2447-0864 · ISSN impresso 1807-3980
https://doi.org/10.31510/infa.v23i1.2439

SIMULAÇÃO DE IDENTIDADE E RISCOS ÉTICOS NA ERA DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

Identity simulation and ethical risks in the age of generative artificial intelligence

Carlos Eduardo Torrichelli de Souza Pinto[1], Carlos Pereira de Castro Filho[2]

AUTORIA

[1] **Carlos Eduardo Torrichelli de Souza Pinto**
Fatec Taquaritinga – SP – Brasil
carloseduardotdes@gmail.com
ORCID: https://orcid.org/0009-0000-2654-0572

[2] **Carlos Pereira de Castro Filho**
Fatec Taquaritinga – SP – Brasil
cpcftq@gmail.com
ORCID: https://orcid.org/0009-0000-3811-2882

HISTÓRICO

Recebido	25/04/2026
Aceito	04/06/2026
Publicado	30/06/2026

COMO CITAR

PINTO, C. E. T. S.; CASTRO FILHO, C. P. Simulação de identidade e riscos éticos na era da inteligência artificial generativa. **Interface Tecnológica**, Taquaritinga, v. 23, n. 1, p. 52-58, 2026. DOI: 10.31510/infa.v23i1.2439.

LICENÇA

Artigo de acesso aberto distribuído sob a licença Creative Commons Atribuição 4.0 Internacional.

cc BY 4.0

RESUMO

O avanço recente da inteligência artificial generativa ampliou significativamente a capacidade de criação de conteúdos sintéticos altamente realistas, incluindo imagens, vídeos e áudios. Entre as aplicações mais sensíveis dessas tecnologias estão os sistemas capazes de simular identidades humanas, como deepfakes e ferramentas de clonagem de voz. Embora essas tecnologias apresentem potencial para aplicações legítimas em áreas como entretenimento, educação e produção audiovisual, também introduzem riscos éticos relevantes relacionados à privacidade, à segurança digital e à confiança social nas interações mediadas por tecnologia. Este artigo analisa os principais riscos associados ao uso de inteligência artificial generativa na simulação de identidades digitais. A pesquisa adota abordagem qualitativa e exploratória baseada em revisão bibliográfica de literatura acadêmica recente, relatórios institucionais e estudos de caso documentados sobre o uso de deepfakes e clonagem de voz. Os resultados indicam que a combinação entre o avanço técnico dos modelos generativos e a ampla disponibilidade de dados pessoais em ambientes digitais amplia significativamente o potencial de fraudes, manipulação informacional e apropriação indevida de identidade. Observa-se também que a crescente dificuldade em distinguir conteúdos autênticos de conteúdos sintéticos pode comprometer a confiança digital em ambientes online. Conclui-se que a mitigação desses riscos exige estratégias multidisciplinares envolvendo desenvolvimento de tecnologias de detecção, marcos regulatórios adequados e iniciativas de educação digital voltadas à conscientização pública.

Palavras-chave: Inteligência artificial generativa. Deepfakes. Clonagem de voz. Identidade digital. Ética em IA.

ABSTRACT

Recent advances in generative artificial intelligence have significantly expanded the ability to create highly realistic synthetic content, including images, videos, and audio. Among the most sensitive applications of these technologies are systems capable of simulating human identities, such as deepfakes and voice cloning tools. Although these technologies present legitimate applications in areas such as entertainment, education, and audiovisual production, they also introduce relevant ethical risks related to privacy, digital security, and social trust in technology-mediated interactions. This paper analyzes the main risks associated with the use of generative artificial intelligence in the simulation of digital identities. The research adopts a qualitative and exploratory approach based on a literature review of recent academic studies, institutional reports, and documented cases involving deepfakes and voice cloning technologies. The results indicate that the combination of technical advances in generative models and the widespread availability of personal data in digital environments significantly increases the potential for fraud, misinformation, and identity appropriati-

on. Furthermore, the growing difficulty in distinguishing authentic from synthetic content may undermine trust in digital environments. The study concludes that mitigating these risks requires multidisciplinary strategies involving detection technologies, regulatory frameworks, and digital literacy initiatives aimed at public awareness.

Keywords: Generative artificial intelligence. Deepfake. Voice cloning. Digital identity. AI ethics.

1 INTRODUÇÃO

A evolução recente da inteligência artificial tem produzido transformações profundas na forma como conteúdos digitais são criados e distribuídos (FLORIDI et al., 2018; WESTERLUND, 2019). Entre os avanços mais significativos se destaca o desenvolvimento da chamada inteligência artificial generativa, capaz de produzir textos, imagens, vídeos e áudios de forma automatizada a partir de grandes volumes de dados utilizados no treinamento de modelos computacionais. Esse tipo de tecnologia tornou-se amplamente acessível ao público, permitindo que indivíduos sem conhecimento técnico avançado possam gerar conteúdos altamente realistas.

Entre as aplicações mais controversas dessa tecnologia estão os sistemas capazes de simular identidades humanas por meio de mídia sintética. Ferramentas conhecidas como *deepfakes* permitem substituir ou manipular rostos em vídeos com elevado grau de realismo, enquanto sistemas de clonagem de voz conseguem reproduzir características vocais específicas de indivíduos a partir de pequenas amostras de áudio (KIETZMANN et al., 2020; MIRSKY; LEE, 2022). Essas tecnologias têm potencial para aplicações legítimas em setores como cinema, entretenimento, educação e acessibilidade digital.

No entanto, a capacidade de reproduzir digitalmente a aparência e a voz de pessoas reais levanta preocupações éticas relevantes (CHESNEY; CITRON, 2018). A disseminação de conteúdos sintéticos pode facilitar a criação de fraudes, manipulação informacional, difamação e outras formas de abuso digital. Além disso, a presença massiva de dados pessoais em redes sociais e plataformas digitais fornece material abundante para o treinamento ou adaptação desses sistemas, ampliando o risco de uso indevido da identidade de indivíduos (ZUBOFF, 2019). Nesse contexto, surge a seguinte questão de pesquisa: de que forma o avanço da inteligência artificial generativa contribui para ampliar os riscos associados à simulação e à apropriação indevida de identidades digitais? A investigação desse problema envolve não apenas aspectos técnicos relacionados ao funcionamento dessas tecnologias, mas também suas implicações éticas, sociais e regulatórias.

Diante desse cenário, o objetivo deste artigo é analisar os principais riscos éticos e sociais associados ao uso de inteligência artificial generativa na simulação de identidades humanas, com foco em tecnologias de *deepfake* e clonagem de voz. Busca-se compreender como essas ferramentas podem impactar a privacidade individual, a segurança digital e a confiança nas interações mediadas por tecnologia.

A relevância deste estudo está relacionada ao rápido crescimento dessas tecnologias e à necessidade de compreender seus impactos sociais antes que seu uso se torne ainda mais difundido. A análise crítica desses riscos contribui para o desenvolvimento de estratégias de mitigação que envolvam regulação adequada, desenvolvimento tecnológico responsável e conscientização pública sobre os desafios da era da mídia sintética.

2 FUNDAMENTAÇÃO TEÓRICA

A literatura recente aponta que o avanço da inteligência artificial generativa deslocou o foco dos riscos tecnológicos tradicionais (como automação e desemprego) para riscos sociais, éticos e identitários. Tecnologias capazes de gerar rostos, vozes e vídeos realistas ampliaram significativa-

mente as possibilidades de impersonificação, fraude e manipulação social, afetando diretamente a confiança digital e a noção de autenticidade.

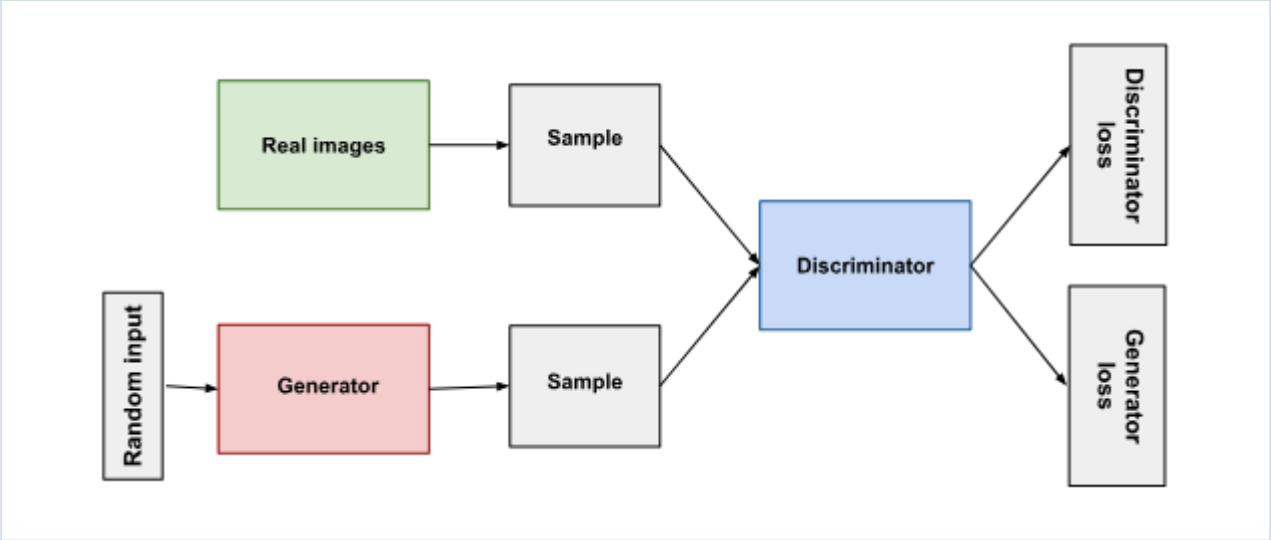
A facilidade de acesso a ferramentas de geração de mídia sintética, aliada à ampla disponibilidade de dados pessoais, contribui para a ampliação de riscos de roubo de identidade digital (CHESNEY; CITRON, 2018; MIRSKY; LEE, 2022; TOLOSANA et al., 2020).

2.1 IA generativa e mídia sintética

A inteligência artificial generativa refere-se a sistemas computacionais capazes de produzir novos conteúdos a partir de padrões aprendidos durante o processo de treinamento em grandes bases de dados. Esses sistemas utilizam técnicas avançadas de aprendizado de máquina, particularmente redes neurais profundas, para identificar relações complexas entre diferentes tipos de informação.

Um dos marcos importantes nesse campo foi o desenvolvimento das *Generative Adversarial Networks* (GANs), apresentadas por Goodfellow et al. (2014). Nesse modelo, duas redes neurais competem entre si: uma responsável por gerar conteúdo sintético e outra responsável por avaliar sua autenticidade. Esse processo competitivo permite a produção de imagens cada vez mais realistas (KIETZMANN et al., 2020).

Figura 1 – Estrutura de uma Generative Adversarial Network (GAN)



Fonte: Google (2019).

Mais recentemente, modelos baseados em arquiteturas *Transformer* também ampliaram significativamente as capacidades da IA generativa. Esses sistemas são capazes de gerar conteúdos multimodais, combinando texto, imagem e áudio em uma mesma estrutura de aprendizado. A expansão dessas tecnologias possibilitou novas aplicações em diversas áreas, incluindo design, produção audiovisual, criação artística e geração automatizada de conteúdo digital. Entretanto, a mesma capacidade tecnológica que permite inovação criativa também pode ser utilizada para fins maliciosos.

2.2 Deepfakes e clonagem de voz na simulação de identidade

Deepfakes são definidos como conteúdos audiovisuais gerados ou manipulados por inteligência artificial que simulam a aparência ou a voz de indivíduos reais. O termo surgiu inicialmente em comunidades online que utilizavam redes neurais para substituir rostos em vídeos, mas rapi-

damente passou a representar um fenômeno mais amplo relacionado à mídia sintética. Os *deepfakes* representam uma nova categoria de ameaça digital, pois combinam técnicas avançadas de visão computacional e aprendizado profundo para produzir conteúdos extremamente convincentes.

Nos últimos anos, observa-se uma evolução significativa nas técnicas de geração de *deepfakes*, especialmente com o surgimento de sistemas capazes de operar em tempo real (MIRSKY; LEE, 2022). Diferentemente dos métodos tradicionais, que exigem processamento prévio e edição posterior de vídeos, essas abordagens permitem a manipulação instantânea de expressões faciais durante transmissões ao vivo, videoconferências ou interações digitais síncronas.

Essa capacidade amplia consideravelmente os riscos associados à simulação de identidade, uma vez que elimina a necessidade de pós-processamento e possibilita a utilização dessas tecnologias em contextos dinâmicos, como chamadas de vídeo em tempo real (WESTERLUND, 2019). Nesse cenário, torna-se mais difícil para os indivíduos identificar sinais de manipulação, o que potencializa o uso dessas ferramentas em fraudes, engenharia social e outras formas de engano digital (TOLOSANA et al., 2020).

Figura 2 – Exemplo de sistema de deepfake em tempo real aplicado à troca facial em vídeo

Fonte: GVCLab (2025).

Avanços recentes têm possibilitado a geração de *deepfakes* em tempo real, ampliando riscos em interações síncronas (MIRSKY; LEE, 2022; TOLOSANA et al., 2020). Projetos públicos disponibilizados em plataformas de desenvolvimento colaborativo demonstram que sistemas capazes de realizar troca facial em tempo real já podem ser implementados com recursos computacionais relativamente acessíveis.

Esse avanço indica uma mudança relevante no panorama de risco, pois a simulação de identidade deixa de ser limitada a conteúdos editados e passa a ocorrer também em interações ao vivo, aumentando seu potencial de impacto em ambientes pessoais, profissionais e institucionais. A clonagem de voz constitui outro avanço relevante nesse campo. Sistemas modernos de clonagem de voz conseguem reproduzir características vocais de indivíduos a partir de pequenas amostras

de áudio, muitas vezes com poucos segundos de gravação (ARIK et al., 2018; JIA et al., 2018; MIRSKY; LEE, 2022). Isso cria possibilidades para aplicações legítimas, como síntese de voz personalizada, mas também abre espaço para fraudes e manipulação de identidade.

2.3 Riscos éticos, sociais e de privacidade

O debate sobre ética em inteligência artificial ganhou destaque nos últimos anos devido ao impacto crescente dessas tecnologias na sociedade. Floridi et al. (2018) destacam que sistemas de IA devem ser desenvolvidos e utilizados considerando princípios éticos fundamentais, como transparência, responsabilidade e respeito à autonomia humana.

No caso específico da mídia sintética, os riscos vão além do dano individual. A disseminação de conteúdos falsificados pode comprometer a confiança social em evidências audiovisuais, que historicamente foram consideradas formas confiáveis de registro da realidade. Esse fenômeno é conhecido como *liar's dividend*, no qual a existência de *deepfakes* permite que indivíduos desacreditem conteúdos autênticos alegando que se tratam de falsificações digitais (CHESNEY; CITRON, 2018).

Além disso, a coleta massiva de dados pessoais na economia digital cria um ambiente no qual imagens, vídeos e gravações de voz estão amplamente disponíveis online. Esse cenário aumenta significativamente o risco de apropriação indevida de identidade por sistemas automatizados.

3 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa se caracteriza como um estudo qualitativo de natureza exploratória, baseado em revisão bibliográfica narrativa da literatura.

A coleta de dados foi realizada por meio de levantamento bibliográfico em bases acadêmicas, incluindo Google Scholar, Scopus e IEEE Xplore, utilizando palavras-chave relacionadas ao tema, como "*deepfake*", "*generative* AI", "*voice cloning*", "AI *ethics*" e "*synthetic media*".

Foram selecionados artigos científicos, relatórios institucionais e publicações acadêmicas relevantes publicados principalmente entre 2018 e 2024. A seleção considerou critérios de relevância temática, impacto acadêmico e reconhecimento institucional das fontes.

Após a seleção das fontes, realizou-se uma análise qualitativa do conteúdo, buscando identificar padrões de risco associados à utilização dessas tecnologias, bem como possíveis estratégias de mitigação discutidas na literatura científica.

4 RESULTADOS E DISCUSSÃO

A partir da revisão bibliográfica e da análise de casos recentes envolvendo o uso de inteligência artificial generativa, foi possível identificar padrões recorrentes de risco associados à simulação de identidade humana. Os resultados evidenciam que o avanço técnico das ferramentas de geração de imagens, vídeos e áudios, aliado à ampla exposição de dados pessoais em ambientes digitais, cria um cenário de vulnerabilidade social ainda pouco compreendido pela maioria da população.

Os achados reforçam que os riscos não se limitam a usos isolados ou experimentais dessas tecnologias, mas se manifestam de forma sistêmica, impactando a confiança digital, a privacidade e a segurança individual (TOLOSANA et al., 2020).

4.1 A exposição massiva de dados pessoais

A análise da literatura indica que um dos principais fatores que possibilitam a simulação de identidade por IA é a ampla disponibilidade de dados pessoais na internet. Fotografias, vídeos e gravações de voz compartilhadas em redes sociais constituem matéria-prima para o treinamento ou adaptação de modelos generativos.

Esse fenômeno está relacionado ao modelo econômico dominante na internet, baseado na coleta e análise massiva de dados pessoais. Segundo Zuboff (2019), a economia digital contemporânea depende da captura contínua de informações sobre o comportamento humano.

Quando combinados com sistemas de IA generativa, esses dados podem ser utilizados para recriar digitalmente a aparência e a voz de indivíduos reais.

4.2 Erosão da confiança digital e impactos sociais

Outro resultado relevante se refere ao impacto dessas tecnologias na confiança digital. A crescente dificuldade em distinguir conteúdos autênticos de conteúdos sintéticos cria um ambiente de incerteza informacional.

De acordo com Chesney e Citron (2018), *deepfakes* podem ser utilizados para manipulação política, difamação e fraudes financeiras. A simples possibilidade de falsificação reduz a confiabilidade de registros audiovisuais, anteriormente considerados evidências relativamente seguras.

Esse fenômeno pode ter consequências relevantes para processos jurídicos, comunicação institucional e interações sociais mediadas por tecnologia.

4.3 Limitações legais e desafios regulatórios

A análise também indica que as estruturas regulatórias atuais ainda estão em processo de adaptação para lidar com os desafios apresentados pela IA generativa. Um exemplo relevante é o *Artificial Intelligence Act* da União Europeia, que estabelece diretrizes para o desenvolvimento e uso responsável de sistemas de inteligência artificial.

Apesar dessas iniciativas, ainda existem lacunas relacionadas à responsabilidade legal pelo uso indevido de identidades digitais geradas por IA. Além disso, sistemas técnicos de detecção de *deepfakes* ainda enfrentam dificuldades diante da rápida evolução dos modelos generativos.

5 CONCLUSÃO

Este estudo analisou os riscos éticos e sociais associados ao uso da inteligência artificial generativa na simulação de identidades digitais. A partir da revisão de literatura acadêmica e de estudos recentes sobre *deepfakes* e clonagem de voz, observou-se que essas tecnologias apresentam potencial significativo tanto para aplicações legítimas quanto para usos maliciosos.

Os resultados indicam que a combinação entre avanços técnicos na geração de mídia sintética e a ampla disponibilidade de dados pessoais na internet cria um ambiente favorável à apropriação indevida de identidades. Fotografias, vídeos e gravações de voz compartilhados online podem ser utilizados para reproduzir digitalmente indivíduos reais sem seu consentimento.

Outro aspecto relevante identificado é o impacto dessas tecnologias na confiança digital. A dificuldade crescente em distinguir conteúdos autênticos de conteúdos gerados por inteligência artificial pode comprometer a credibilidade das interações online e favorecer a disseminação de desinformação.

Além disso, a análise demonstra que os marcos regulatórios ainda estão em processo de adaptação para lidar com os desafios apresentados pela IA generativa. Embora iniciativas regulatórias já estejam em desenvolvimento em diferentes regiões do mundo, persistem lacunas relacionadas à responsabilização e à prevenção do uso indevido dessas tecnologias.

Diante desse cenário, torna-se evidente que a mitigação dos riscos associados à simulação de identidade por inteligência artificial exige uma abordagem multidisciplinar. Estratégias eficazes devem envolver avanços em tecnologias de detecção de mídia sintética, desenvolvimento de políticas públicas adequadas e programas de educação digital voltados à conscientização da população.

Assim, a inteligência artificial generativa representa simultaneamente uma inovação tecnológica significativa e um desafio ético relevante para a sociedade contemporânea, exigindo equilíbrio entre inovação, responsabilidade e proteção da integridade das identidades digitais.

REFERÊNCIAS

ARIK, Sercan O. *et al.* **Neural voice cloning with a few samples**. [*S. l.*]: arXiv, 2018. Disponível em: http://arxiv.org/abs/1802.06006.

CHESNEY, Robert M.; CITRON, D. Deepfakes and the new disinformation war. **Foreign Affairs**, 2018.

EUROPEAN COMMISSION. **AI Act**. [*S. l.*], [*s. d.*]. Disponível em: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. Acesso em: 6 fev. 2026.

EUROPEAN COMMISSION. **AI Act enters into force**. [*S. l.*], [*s. d.*]. Disponível em: https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_en. Acesso em: 17 jan. 2026.

FLORIDI, Luciano *et al.* AI4People: an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. **Minds and Machines**, v. 28, n. 4, p. 689-707, 2018.

GOODFELLOW, Ian J. *et al.* **Generative adversarial networks**. [*S. l.*]: arXiv, 2014. Disponível em: http://arxiv.org/abs/1406.2661.

GOOGLE. **Overview of GAN structure**: generative adversarial networks. [*S. l.*]: Google, [*s. d.*]. Disponível em: https://developers.google.com/machine-learning/gan/gan_structure. Acesso em: 21 fev. 2026.

GVCLAB. **PersonaLive!**: expressive portrait image animation for live streaming. [*S. l.*]: GitHub, [*s. d.*]. Disponível em: https://github.com/GVCLab/PersonaLive. Acesso em: 11 abr. 2026.

JIA, Ye *et al.* **Transfer learning from speaker verification to multispeaker text-to-speech synthesis**. [*S. l.*]: arXiv, 2018. Disponível em: http://arxiv.org/abs/1806.04558.

KALIZHAN, Alikhan. **Deepfake detection methods, a review**. [*S. l.*: *s. n.*], 2025. DOI: https://doi.org/10.13140/RG.2.2.13178.76481.

KIETZMANN, Jan *et al.* Deepfakes: trick or treat? **Business Horizons**, v. 63, n. 2, p. 135-146, 2020.

MIRSKY, Yisroel; LEE, Wenke. The creation and detection of deepfakes: a survey. **ACM Computing Surveys**, v. 54, n. 1, p. 1-41, 2022.

TOLOSANA, Ruben *et al.* Deepfakes and beyond: a survey of face manipulation and fake detection. **Information Fusion**, v. 64, p. 131-148, 2020.

WESTERLUND, Mika. The emergence of deepfake technology: a review. **Technology Innovation Management Review**, v. 9, n. 11, p. 39-52, 2019.

ZUBOFF, Shoshana. **The age of surveillance capitalism**: the fight for a human future at the new frontier of power. New York: PublicAffairs, 2019.

Os conceitos e as opiniões expressos neste artigo são de responsabilidade exclusiva de seus autores.