



**UNIVERSIDADE FEDERAL DE ALAGOAS**  
**INSTITUTO DE CIÊNCIAS HUMANAS, COMUNICAÇÃO E ARTES**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

POLLYANY ANNENBERG NASCIMENTO GOMES

**DESINFORMAÇÃO E INTELIGÊNCIA ARTIFICIAL GENERATIVA:**  
dinâmicas, atores e dilemas ético-informacionais em redes digitais

MACEIÓ

2025

POLLYANY ANNENBERG NASCIMENTO GOMES

**DESINFORMAÇÃO E INTELIGÊNCIA ARTIFICIAL GENERATIVA:**

dinâmicas, atores e dilemas ético-informacionais em redes digitais

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Alagoas como pré-requisito parcial para a obtenção do título de mestra em Ciência da Informação.

**Área de concentração:** Informação, Tecnologia e Inovação.

**Linha de Pesquisa:** Produção, Mediação e Gestão da Informação.

**Orientadora:** Prof<sup>ª</sup>. Dra. Maria Lívia Pacheco de Oliveira.

MACEIÓ

2025

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

G633d Gomes, Pollyany Annenberg Nascimento.  
Desinformação e inteligência artificial generativa : dinâmicas, atores e dilemas ético-informacionais em redes digitais / Pollyany Annenberg Nascimento Gomes. – 2025.  
106 f. : il. color.

Orientadora: Maria Lívia Pacheco de Oliveira.  
Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Alagoas. Instituto de Ciências Humanas, Comunicação e Artes. Programa de Pós-Graduação em Ciência da Informação. Maceió, 2025.

Bibliografia: f. 86-92.  
Apêndices: f. 93-106.

1. Inteligência Artificial Generativa. 2. Desinformação. 3. Regime de informação. 4. Plataformas digitais. I. Título.

CDU: 316.776.22:004.89



**UNIVERSIDADE FEDERAL DE ALAGOAS**  
**INSTITUTO DE CIÊNCIAS HUMANAS, COMUNICAÇÃO E ARTES**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

**ATA DE DEFESA DE DISSERTAÇÃO**

**Defesa nº 60**

Ata da Sessão Pública de Defesa de Dissertação do(a) mestrando(a) **POLLYANY ANNENBERG NASCIMENTO GOMES** como requisito para obtenção do grau de Mestre(a) em Ciência da Informação, na Linha de Pesquisa Produção, Mediação e Gestão da Informação, Área de Concentração Informação, Tecnologia e Inovação.


No dia 14 de outubro de 2025, às 15h, reuniu-se, em sessão pública, pelo canal @ppgciufal no YouTube, a Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Alagoas (PPGCI/UFAL), nos termos do Regulamento Geral dos Programas de Pós-Graduação *Stricto Sensu* da UFAL (Resolução nº 50/2014 - CONSUNI/UFAL), do Regimento Interno do PPGCI/UFAL (Resolução nº 24/2018 - CONSUNI/UFAL) e da Resolução nº 04/2021 – PPGCI/UFAL, para realização da Defesa de Dissertação do(a) mestrando(a) **POLLYANY ANNENBERG NASCIMENTO GOMES**, matrícula 2024108308, intitulada **DESINFORMAÇÃO E INTELIGÊNCIA ARTIFICIAL GENERATIVA: DINÂMICAS, ATORES E DILEMAS ÉTICO-INFORMACIONAIS EM REDES DIGITAIS**. A Banca Examinadora foi composta pelos seguintes membros: Profa. Dra. Juliana Fachin - PPGCI/UFAL (Presidente), Profa. Dra. Maria Lívia Pachêco de Oliveira - PPGCI/UFAL (Orientadora), Profa. Dra. Priscila Muniz de Medeiros - PPGCI/UFAL (Membro Titular Interno), Prof. Dr. Edvaldo Carvalho Alves - PPGCI/UFAL (Membro Titular Externo), Prof. Dr. Fellipe Sá Brasileiro - PPGCI/UFAL (Membro Suplente Externo). Após a apresentação da Dissertação, foi dada a palavra aos Examinadores para arguição, tendo o(a) candidato(a) respondido aos questionamentos formulados. Encerrada a arguição, a Banca Examinadora reuniu-se em sessão reservada para proceder ao julgamento, sendo atribuídos os seguintes pareceres: 1º membro: aprovado(a) (X), reprovado(a) ( ); 2º membro: aprovado(a) (X), reprovado(a) ( ); e 3º membro: aprovado(a) (X), reprovado(a) ( ). Em atendimento ao que estabelece o artigo 64, § 2º, do Regimento Interno do PPGCI/UFAL, o(a) discente foi considerado(a): APROVADO(A) (X); REPROVADO(A) ( ).

O(a) discente foi aprovado(a) e deverá proceder as alterações apontadas pela banca:

Foi deliberado que a acadêmica deverá realizar pequenos ajustes para aprimorar a organização, a apresentação e a qualidade geral do estudo.


Nada mais havendo a tratar, o(a) Presidente(a) da Banca Examinadora encerrou os trabalhos. E, para constar, eu, Profa. Dra. Juliana Fachin, Presidente, confiro e assino a presente ata, em três vias, juntamente aos membros da Banca Examinadora e ao(à) candidato(a).

Maceió/AL, 14 de outubro de 2025.

Documento assinado digitalmente  
 JULIANA FACHIN  
Data: 20/10/2025 19:29:46-0300  
Verifique em <https://validar.iti.gov.br>


---

**Profa. Dra. Profa. Dra. Juliana Fachin**  
Presidente – PPGCI/UFAL

Documento assinado digitalmente  
 MARIA LIVIA PACHECO DE OLIVEIRA  
Data: 17/10/2025 17:24:48-0300  
Verifique em <https://validar.iti.gov.br>


---

**Profa. Dra. Maria Lívia Pachêco de Oliveira**  
Orientadora – PPGCI/UFAL

Documento assinado digitalmente  
 EDVALDO CARVALHO ALVES  
Data: 20/10/2025 10:51:22-0300  
Verifique em <https://validar.iti.gov.br>


---

**Prof. Dr. Edvaldo Carvalho Alves**  
Membro Titular Externo – PPGCI/UFPB

Documento assinado digitalmente  
 PRISCILA MUNIZ DE MEDEIROS  
Data: 20/10/2025 12:38:45-0300  
Verifique em <https://validar.iti.gov.br>

---

**Profa. Dra. Priscila Muniz de Medeiros**  
Membro Titular Interno – PPGCI/UFAL

Documento assinado digitalmente  
 POLLYANY ANNENBERG NASCIMENTO GOMES  
Data: 20/10/2025 13:30:45-0300  
Verifique em <https://validar.iti.gov.br>

---

**Pollyany Annenberg Nascimento Gomes**  
Mestrando(a) – PPGCI/UFAL

À minha mãe Paula, pernambucana e  
agrestina, que criou três filhos na luta e na  
coragem, sempre dizendo: “Não sejam bestas,  
olhem bem antes de acreditar” – e tinha razão.

## AGRADECIMENTOS

Entrei neste mestrado entusiasmada por uma inquietação de pesquisa, mas o maior desafio, na verdade, foi compreender a mim mesma.

Então, a gratidão que carrego vem da resistência construída após as renúncias que precisei fazer e na escolha de seguir no programa, mesmo quando a incerteza pesava mais do que a minha vontade.

Por isso, preciso primeiro agradecer ao universo, a Deus, ao meu anjo da guarda (ou como cada um concebe o divino) por me dar força quando eu mais precisei ser empurrada.

À minha família pelo apoio e compreensão e ao meu companheiro Lelo, amor de mais de uma década e uma das minhas maiores inspirações profissionais na comunicação. Você soube que eu conseguiria, mesmo quando eu mesma duvidava.

Em alguns momentos, eu teria desistido de mim, mas foram essas razões que me fizeram seguir.

Também agradeço à minha orientadora, professora Lívia, que fez toda a diferença na execução da pesquisa desde as minhas aplicações para os meus primeiros eventos no programa.

Aos colegas de turma, que transformaram o caminho em algo mais leve com memes, brincadeiras e lamentações. Tenham amigos de mestrado.

Por fim, à Ciência da Informação que também foi fundamental na evolução do meu saber e fazer como profissional de comunicação na sociedade.

Por tudo isso, fica aqui o meu agradecimento.

**“Eu quase que nada não sei. Mas desconfio de muita coisa.” (Guimarães Rosa, 1994, p. 14)**

## RESUMO

A popularização, desde 2022, de ferramentas de inteligência artificial (IA) generativa, como o ChatGPT, ampliou o potencial desinformativo das plataformas digitais ao permitir a criação ultrarrealista de textos, imagens, áudios e vídeos de eventos que nunca aconteceram. Embora também utilizadas para fins recreativos, sua apropriação por grupos organizados eleva os riscos de manipulação da opinião pública e distorção de narrativas políticas. No contexto brasileiro, altamente vulnerável à desinformação, o fenômeno exige a compreensão dos interesses, agentes e estratégias que sustentam sua circulação. Apoiada no conceito de regime de informação, esta pesquisa analisa como atores, artefatos, dispositivos e ações estruturam a disseminação de conteúdos manipulados por IA generativa nas plataformas digitais. De caráter qualitativo e descritivo, a investigação se baseia na análise de casos verificados por uma agência de checagem, nos dispositivos de controle como o Projeto de Lei 2338/2023 e nas políticas sobre conteúdo gerado por IA nas plataformas Google, Meta, TikTok e X. Os resultados indicam um regime em que a lógica econômica das *big techs* e a fragilidade dos mecanismos de moderação favorecem a atuação de agentes desinformadores. Os resultados mostram que a desinformação gerada por IA generativa se consolida como uma força atuante no regime de informação contemporâneo, explorando a confiança em figuras públicas e sendo amplificada pela lógica econômica das plataformas digitais.

**Palavras-chave:** desinformação; regime de informação; inteligência artificial generativa; plataformas digitais.

## ABSTRACT

The popularization since 2022 of generative artificial intelligence (AI) tools, such as ChatGPT, has amplified the disinformative potential of digital platforms by enabling the ultra-realistic creation of texts, images, audio, and videos of events that never occurred. While also used for recreational purposes, their appropriation by organized groups elevates the risks of public opinion manipulation, distortion of political narratives, and the weakening of democratic processes. In the Brazilian context, which is highly vulnerable to disinformation, the phenomenon demands an understanding of the interests, agents, and strategies that sustain its circulation. Grounded in the concept of the information regime, this research analyzes how actors, artifacts, devices, and actions structure the dissemination of manipulated content on digital platforms. The qualitative and descriptive investigation is based on the analysis of cases verified by Agência Lupa and on control mechanisms such as Proposed Law 2338/2023 and the policies on AI-generated content on the Google, Meta, TikTok, and X platforms. The results indicate a regime in which the economic logic of big techs and the fragility of regulatory mechanisms favor the activities of disinformation agents. The results shows that disinformation generated by generative AI is consolidating itself as an active force in the contemporary information regime, exploiting trust in public figures and being amplified by the economic logic of digital platforms.

**Keywords:** disinformation; information regime; generative artificial intelligence; digital platforms.

## LISTA DE FIGURAS

Figura 1 -	IA generativa e sua dificuldade para reproduzir mãos humanas em 2022	32
Figura 2 -	Imagens geradas pela mesma aplicação de IA generativa com dois anos de diferença	33
Figura 3 -	Imagens geradas por IA do presidente Lula em ambiente hospitalar	36
Figura 4 -	A midiaticização do fenômeno da IA generativa em portal de notícias	39
Figura 5 -	Publicações utilizadas na pesquisa de Zurique na rede X, com textos gerados por humano e pelo GPT	40
Figura 6	Representações textuais geradas por IA para diferentes gêneros	49
Figura 7 -	Representação dos elementos do regime de informação em um caso de desinformação gerada por IA generativa	63
Figura 8	Captura de tela de vídeo falsamente atribuído ao presidente Luiz Inácio Lula da Silva para promover golpe do ‘dinheiro acumulado no CPF’	67
Figura 9	Captura de tela do vídeo publicado no Facebook que utiliza a imagem do médico Drauzio Varella para promover um suposto suplemento alimentar	69
Figura 10	Captura de tela do vídeo em que advogado fala sobre FGTS foi manipulado com uso de IA	73

## LISTA DE TABELAS

Tabela 1 - Panorama de ferramentas populares de IA generativa e estimativa de usuários (2025)	34
Tabela 2 - Exemplos de relação entre dispositivos de controle, riscos e casos de desinformação por plataforma	42
Tabela 3 - Relação entre objetivos da pesquisa e corpus documental	54
Tabela 4 - Operacionalização do levantamento documental	59
Tabela 5 - Dimensão técnica da pesquisa	60
Tabela 6 - Atores identificados no regime de informação gerada por IA	67
Tabela 7 - Artefatos da desinformação sintética identificados	72
Tabela 8 - Dispositivos de controle	74
Tabela 9 - Ações de informação	79

## LISTA DE ABREVIATURAS E SIGLAS

Brapci	Base de Dados em Ciência da Informação
CINFORM	Encontro Nacional de Ensino e Pesquisa em Informação
DDR	Desinformação Digital em Rede
Enancib	Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação
IA	Inteligência Artificial
IFCN	International Fact-Checking Network
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
PL	Projeto de Lei
PPGCI	Programa de Pós-Graduação em Ciência da Informação
SITI	Seminário Internacional de Informação, Tecnologia e Inovação
TDICs	Tecnologias Digitais de Informação e Comunicação
TSE	Tribunal Superior Eleitoral
UFAL	Universidade Federal de Alagoas
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>13</b>
<b>2 DESINFORMAÇÃO GERADA POR IA NO REGIME DE INFORMAÇÃO: FUNDAMENTOS CONCEITUAIS.....</b>	<b>20</b>
<b>3 INTELIGÊNCIA ARTIFICIAL GENERATIVA: CONTEÚDOS SINTÉTICOS, ATUAÇÃO E LÓGICAS DE PLATAFORMA.....</b>	<b>28</b>
<b>3.1 Da criação à simulação: os conteúdos sintéticos.....</b>	<b>29</b>
<b>3.2 A IA como ator informacional.....</b>	<b>34</b>
<b>3.3 Plataformas digitais: lógicas algorítmicas e responsabilidade.....</b>	<b>37</b>
<b>4 A DESINFORMAÇÃO POR IA GENERATIVA NO REGIME DE INFORMAÇÃO CONTEMPORÂNEO.....</b>	<b>42</b>
<b>4.1 Atores, artefatos, dispositivos e ações.....</b>	<b>43</b>
<b>4.2 Implicações éticas e desafios regulatórios da desinformação gerada por IA generativa.....</b>	<b>46</b>
<b>5 METODOLOGIA.....</b>	<b>50</b>
<b>5.1 Caracterização da pesquisa.....</b>	<b>51</b>
<b>5.1.1 Documentos analisados.....</b>	<b>55</b>
<b>5.2 Delimitação empírica.....</b>	<b>56</b>
<b>5.3 Operacionalização das técnicas aplicadas.....</b>	<b>58</b>
<b>5.4 Sobre as análises e discussões dos resultados.....</b>	<b>62</b>
<b>5.5 Limitações e aprendizados.....</b>	<b>63</b>
<b>6 DESINFORMAÇÃO E O REGIME DE INFORMAÇÃO NA ERA DA IA GENERATIVA.....</b>	<b>65</b>
<b>6.1 Os atores do regime de informação da IA generativa.....</b>	<b>66</b>
<b>6.2 Artefatos da desinformação sintética identificados.....</b>	<b>71</b>
<b>6.3 Dispositivos de controle.....</b>	<b>74</b>
<b>6.4 Ações de informação.....</b>	<b>78</b>
<b>7 CONSIDERAÇÕES FINAIS.....</b>	<b>81</b>
<b>REFERÊNCIAS.....</b>	<b>86</b>
<b>APÊNDICE A – LISTAGEM DE NOTÍCIAS DE CASOS COM USO DE IA VERIFICADAS PELA AGÊNCIA LUPA NO PERÍODO DE MARÇO DE 2024 A MARÇO DE 2025.....</b>	<b>93</b>
<b>APÊNDICE B – DISPOSITIVOS ANALISADOS.....</b>	<b>104</b>

## 1 INTRODUÇÃO

"Viver – não é? – é muito perigoso. Porque ainda não se sabe." (Guimarães Rosa, 1994, p. 840). A reflexão de Riobaldo em “Grande Sertão: Veredas” mostra uma inquietação humana constante, onde o perigo de viver se encontra na dúvida entre a verdade e o mal, na dificuldade de interpretar um mundo cheio de sinais incertos e na própria natureza da linguagem que pode tanto mascarar quanto revelar. Para o personagem, a realidade não se entrega facilmente, ela exige constante decifração, já que o erro é uma sombra sempre presente, além do desafio de caminhar por uma existência às vezes distorcida.

Essa sensação de desorientação pode ser percebida na forma como as pessoas lidam hoje com a informação. Notícias, imagens, áudios e dados chegam e são compartilhados com velocidade e excesso de conteúdos que mais confundem do que esclarecem, uma experiência que lembra o “mundo muito misturado” de Riobaldo, onde interpretar exige atenção aos sinais e desconfiança dos caminhos.

A desinformação, definida por pesquisadores como Wardle e Derakhshan (2023) como informação falsa compartilhada com propósito de causar dano, é uma das expressões consideradas mais graves da desordem informacional no ambiente digital. Esse fenômeno, no entanto, não surgiu na era digital do século XXI, já que circulava em panfletos, discursos em contextos de guerra e propaganda política (Chomsky, 2014), sempre encontrando brechas na confiança coletiva para se infiltrar. O que muda, hoje, é a sofisticação das formas de criação por meio de máquinas capazes de produzir conteúdos desinformativos que, além de imitar o real, dificultam essa distinção entre o que aconteceu e o que apenas parece ter acontecido.

Assim, “*Podem as máquinas pensar?*”, questão lançada pelo cientista Alan Turing (1950), não questionava apenas se máquinas pensavam, mas se poderiam enganar e se poderiam parecer humanas. Anos depois, as reflexões de Turing deram base ao campo hoje conhecido como inteligência artificial (IA), definida por Russell e Norvig (2013) como a área que busca construir sistemas capazes de simular aspectos do comportamento inteligente humano. A IA generativa, resultado desses avanços, ganhou popularidade com o lançamento do ChatGPT<sup>1</sup>, da OpenAI, em novembro de 2022 (McKinsey, 2023)<sup>2</sup> e permite, hoje, a vivência de uma era em que os sistemas saíram dos laboratórios e estão mais acessíveis para qualquer pessoa com acesso a internet produzir e compartilhar conteúdo com aparência

---

<sup>1</sup> O ChatGPT é um modelo de linguagem baseado em redes neurais da família GPT (*Generative Pre-trained Transformer*), desenvolvido pela empresa OpenAI para gerar conteúdos próximos aos da linguagem humana.

<sup>2</sup> De acordo com a consultoria, o ChatGPT atingiu a marca de 100 milhões de usuários ativos nos primeiros meses de lançamento.

ultrarrealista de forma massiva.

Entender como esses conteúdos ultrarrealistas circulam vai além da análise técnica dos sistemas inteligentes que os produzem, pois a desinformação também está nas formas como a informação é percebida socialmente. Nesse ponto, a noção de regime de informação se torna relevante, como sugerem Frohmann (1995), ao tratar a informação como prática discursiva que articula aspectos sociais e tecnológicos, e González de Gómez (2012), ao conceituar regimes de informação como o modo informacional dominante em uma formação social. Isso implica uma interação entre o comportamento humano, as estruturas sociais e a capacidade das tecnologias digitais na produção e circulação de informação.

Chamados de conteúdos sintéticos (UNESCO, 2024), as imagens, vídeos, sons e textos manipulados pelas ferramentas de IA generativa não são, por definição, desinformativos, já que podem oferecer serviços recreativos e funções construtivas. Por exemplo, quando um redator precisa revisar e aperfeiçoar o tom de um texto de uma forma mais ágil, ou quando é preciso recuperar a memória de uma fotografia danificada a partir de padrões visuais. O risco aparece quando os mesmos recursos são utilizados para simular o tom de escrita de uma revista, o timbre de voz de uma personalidade pública ou os discursos de representantes políticos, construindo narrativas que jamais foram ditas.

Nesse cenário, a informação é construída, interpretada e disputada em contextos marcados por diferentes interesses e relações de poder (Capurro; Hjørland, 2007), ou seja, o que se reconhece como verdade deixa de depender exclusivamente de critérios factuais e passa a ser determinado por narrativas que reforçam visões de mundo pré-concebidas. Araújo (2020a, 2020b) aponta que a marca mais evidente da era da pós-verdade é a centralidade da emoção: a adesão aos discursos não depende da comprovação, mas da afinidade afetiva, enquanto a veracidade é substituída pela persuasão.

No Brasil, o cenário informacional combina fatores que tornam o ambiente especialmente vulnerável à desinformação: a elevada vulnerabilidade informacional e a rápida incorporação de tecnologias emergentes, como a IA generativa. Em 2024, segundo a OCDE (2023), o país ficou na última posição entre 21 nações analisadas quanto à habilidade de identificar notícias falsas<sup>3</sup>. No mesmo período, mais da metade da população brasileira, 54%, declarou ter utilizado ferramentas de IA generativa, índice superior à média global de 48% (Google; IPSOS, 2024). Esses dados revelam uma normalização de um ecossistema em que a aparência de verdade se sobrepõe à veracidade. Além disso, o mercado de IA generativa,

---

<sup>3</sup> Disponível em: [https://www.oecd.org/en/publications/2024/06/the-oecd-truth-quest-survey\\_a1b1739c.html](https://www.oecd.org/en/publications/2024/06/the-oecd-truth-quest-survey_a1b1739c.html)

segundo projeções da *Goldman Sachs* (2024)<sup>4</sup>, deve ultrapassar US\$ 1 trilhão em investimentos até 2027, alimentando um ciclo da aceleração tecnológica e a fragilidade informacional.

Além do conceito central desta pesquisa de regime de informação como prática discursiva (Frohmann, 1995) e como modo informacional dominante em uma formação social (Gómez, 2012), é necessário compreender como diferentes modos de organização da informação abrem espaço para novas configurações desinformativas. Nesse contexto, insere-se a noção de Desinformação Digital em Rede (DDR), formulada por Marco Schneider (2022), que a define como o conjunto das práticas desinformativas contemporâneas que se expandem pelas redes digitais (ambientes sociotécnicos como Facebook, WhatsApp, X ou TikTok) caracterizadas pela rapidez, capilaridade e baixo custo de circulação de conteúdo. O autor explica que a DDR é amplificada pelo baixo custo de operação e pela monetização da mentira como ativo lucrativo, em um ecossistema marcado pela concentração de capital das *big techs*. Embora Schneider não trate diretamente da IA generativa, esta pesquisa considera que as empresas que oferecem essa tecnologia amplificam a Desinformação Digital em Rede ao facilitar a criação de textos, imagens, áudios e vídeos ultrarrealistas que simulam eventos que nunca ocorreram, multiplicando o potencial desinformativo e tornando ainda mais difícil a verificação de fatos.

Essa desinformação gerada por conteúdos ultrarrealistas tem potencial para distorcer debates públicos em diversas esferas, desde a polarização política e a fragilização de instituições públicas até a disseminação de narrativas que podem resultar em danos à vida de indivíduos e coletividades (Westerlund, 2019). Exemplos recentes são os casos envolvendo imagens manipuladas de figuras públicas em anúncios falsos de programas do governo brasileiro para aplicação de golpes (Furtado, 2024)<sup>5</sup>. Compreender a lógica por trás de sua produção e circulação vai além de um exercício acadêmico, mas uma reflexão para todos que se preocupam pelos rumos da preservação de um ecossistema informacional.

Dessa forma, então, como se configura o regime de informação da IA generativa no Brasil? Mais especificamente, **de que maneira a desinformação gerada por IA generativa tem atuado no regime de informação contemporâneo brasileiro, potencializando as dinâmicas da Desinformação Digital em Rede?**

Essas perguntas partem da compreensão de que a desinformação atua no regime de

---

<sup>4</sup> Disponível em: <https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit>

<sup>5</sup> Disponível em: [extra.globo.com/economia/noticia/2024/10/facebook-estudo-aponta-643-anuncios-falsos-sobre-o-desenrola-brasil-e-o-voa-brasil.ghml](https://extra.globo.com/economia/noticia/2024/10/facebook-estudo-aponta-643-anuncios-falsos-sobre-o-desenrola-brasil-e-o-voa-brasil.ghml)

informação contemporâneo e, como lembra Schneider (2022), a Desinformação Digital em Rede configura-se como um fenômeno inédito e alarmante, que se retroalimenta das próprias redes digitais em que circula, explorando seus algoritmos, modelos de negócio e suas lógicas de personalização. Nesse cenário, a IA generativa intensifica essa dinâmica, pois multiplica a capacidade de produzir conteúdos sintéticos ultrarrealistas, potencializando a DDR e seus efeitos no regime informacional.

Diante desse quadro, o recorte adotado nesta pesquisa concentra-se nos casos de desinformação produzida por IA generativa e verificados por agência de *fact-checking*. As políticas das plataformas e o debate legislativo em torno do Projeto de Lei 2338/2023<sup>6</sup>, que propõe normas gerais para o uso responsável de IA no Brasil, são tratados, neste trabalho, como documentos complementares, pois ajudam a compreender como o regime de informação contemporâneo articula normas e processos regulatórios.

Para isso, o objetivo geral desta pesquisa é **analisar como a desinformação gerada por IA generativa se manifesta e se articula no regime de informação contemporâneo brasileiro, tomando a Desinformação Digital em Rede como força atuante nesse regime, a partir da análise de casos verificados por agência de *fact-checking*.**

Sendo assim, definem-se os seguintes objetivos específicos:

1. mapear os componentes do regime de informação relacionados à circulação de desinformação gerada por IA generativa nos casos analisados por agência de *fact-checking*;
2. identificar os atores envolvidos na produção, circulação e contestação desses casos, situando seu papel no regime de informação contemporâneo;
3. levantar os artefatos e dispositivos presentes nos casos, observando como estruturam práticas de desinformação e de verificação.

A justificativa humanística e social desta pesquisa se ancora no contexto brasileiro, marcado por uma cultura de hiperconectividade e pela rápida adesão a tecnologias baseadas em IA, somada ao intenso consumo de redes digitais no país (Kemp, 2024)<sup>7</sup>. Nesse caminho, considera-se que essa tecnologia possui a capacidade de produzir conteúdos falsos com alto grau de realismo, capazes de impactar processos eleitorais, reforçar discursos discriminatórios (Silva, 2022) e abalar a confiança nas instituições públicas (Westerlund, 2019).

Além da relevância social, a presente pesquisa se situa na escassez de investigações

---

<sup>6</sup> O Projeto de Lei n.º 2338/2023 estabelece normas gerais para o uso de inteligência artificial no Brasil. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

<sup>7</sup> Disponível em: <https://datareportal.com/reports/digital-2024-brazil>

voltadas a esse tipo de desinformação. Um levantamento realizado na Brapci (Base de Dados em Ciência da Informação) em janeiro de 2025 com o termo “desinformação, IA generativa” retornou apenas três registros diretamente vinculados à Ciência da Informação. Termos como “regime de informação” combinado com “conteúdo sintético”, que são centrais nesta pesquisa, também aparecem de forma pontual, o que sinaliza a necessidade de ampliar o debate sobre como os sistemas inteligentes impactam os modos de produção, circulação e reconhecimento da informação. Assim, esta pesquisa busca preencher parte dessa lacuna na área.

A investigação ancora-se também nas proposições de autores que articulam fundamentos clássicos da Ciência da Informação com problemáticas emergentes. Carlos Alberto Araújo (2020b) compreende desinformação como fenômeno central da pós-verdade; Schneider (2022) propõe o conceito de Desinformação Digital em Rede, caracterizado pela escala, velocidade e baixo custo das redes digitais; Kaufman (2021) discute os impactos éticos e sociais da IA generativa no Brasil; Bezerra (2023) analisa o papel das plataformas como mediadores informacionais; e Oliveira (2020) reforça a importância da competência crítica em informação como estratégia de enfrentamento a *fake news*. Ao dialogar com essas contribuições, esta pesquisa propõe uma abordagem crítica capaz de problematizar os efeitos da IA generativa sobre o conceito de regime de informação.

Inserida na Linha 1 do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade Federal de Alagoas (UFAL), "Produção, Gestão e Mediação da Informação", esta pesquisa se alinha à compreensão de que os sistemas de IA, ao atuar como mediadores da informação, exigem novos modos de análise crítica e ética (Gomes, 2020; Almeida Júnior, 2009).

Importante também ressaltar que essa dissertação parte de pesquisas realizadas pela autora e sua orientadora durante o primeiro ano do curso de mestrado do PPGCI, que resultou em trabalhos apresentados nas edições de 2024 do Encontro Nacional de Ensino e Pesquisa em Informação (CINFORM), Seminário Internacional de Informação, Tecnologia e Inovação (SITI) e Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação (Enancib). Nessas oportunidades, entre vivências e trocas de experiências com outros pesquisadores, foi visualizada a necessidade de ampliar os estudos.

Por fim, importante destacar que esta pesquisa também possui uma dimensão pessoal: como profissional de comunicação e pesquisadora em Ciência da Informação, a autora reconhece que esse tipo de desinformação é uma ameaça crescente à credibilidade da própria atividade de gestão de conteúdo para instituições públicas e privadas, um cenário que pode minar a confiança das pessoas nessas fontes.

Considerando a natureza recente e ainda pouco explorada da desinformação gerada por IA generativa no campo da Ciência da Informação, esta pesquisa adota uma abordagem descritiva e qualitativa, capaz de sustentar tanto o aprofundamento conceitual quanto uma leitura crítica (Prodanov; Freitas, 2013). A estratégia metodológica articula levantamento documental e análise de conteúdo de casos envolvendo conteúdos sintéticos, orientada pelo referencial do regime de informação.

Caminhando pela estrutura deste documento, estão disponíveis sete seções principais. A primeira seção é esta *Introdução*, em que se apresenta o tema, delinea-se o problema de pesquisa, são formuladas as perguntas e objetivos, discutem-se as justificativas e apresenta de modo geral a metodologia.

A segunda seção, intitulada “*Desinformação gerada por IA no regime de informação: fundamentos conceituais*”, estabelece a base teórica da pesquisa com uma discussão sobre informação, desinformação e tecnologia, com ênfase no contexto brasileiro.

A seção três, “*Inteligência artificial generativa: conteúdos sintéticos, atuação e lógicas de plataforma*”, analisa as transformações provocadas pela IA generativa na produção de conteúdo, incluindo suas interfaces com a desinformação. Essa análise aborda desde os aspectos técnicos até a desinformação no ambiente digital, com foco no cenário brasileiro.

A quarta seção, “*A desinformação por IA generativa no regime de informação contemporâneo*”, detalha os quatro componentes do regime (atores, artefatos, dispositivos e ações) e como eles se aplicam ao fenômeno da desinformação sintética, construindo o modelo e a análise sob a perspectiva da Desinformação Digital em Rede, que guiará a análise empírica.

A quinta seção trata da metodologia, apresentando as escolhas operacionais que orientam a pesquisa, como o uso de análise documental e de conteúdo. Nela, explica-se o recorte empírico e os procedimentos adotados para examinar os casos de desinformação que envolvem o uso de conteúdos sintéticos gerados por IA.

Na sexta seção, “*Desinformação e o regime de informação na era da IA generativa*”, são apresentados e discutidos os resultados da análise, com dados organizados e baseados nos quatro blocos definidos no regime de informação: atores, dispositivos, artefatos e ações. A análise busca identificar como essas dimensões se articulam na produção e circulação desse tipo de desinformação no ambiente digital proposto neste estudo.

Por fim, a sétima seção reúne as considerações finais e retoma as questões centrais do estudo e trata de possíveis caminhos para futuras pesquisas sobre o enfrentamento do fenômeno.

Além do conteúdo principal, esta pesquisa também é composta pelas seções pós-textuais, que incluem, além das referências bibliográficas, os apêndices relatados no recorte empírico para a pesquisa de casos de desinformação gerada por conteúdos sintéticos.

Ao fim, o que também se busca com este estudo é abrir caminho para novas investigações no campo da Ciência da Informação e discussões sobre as implicações da IA generativa em políticas públicas e estratégias de letramento informacional no Brasil. Como observa Sichman (2021), vive-se um período de entusiasmo em torno da IA sustentado pelo barateamento do processamento, pelo avanço e acesso aos serviços, além do volume inédito de dados disponíveis. Esse entusiasmo é compreensível, mas acompanhado das questões que serão abordadas nesta pesquisa, ao reconhecer que, com os benefícios trazidos pela IA generativa, também chegam riscos informacionais e sociais que não podem ser negligenciados.

Se é verdade que “viver é perigoso”, como advertia a fala de Riobaldo no início desta introdução, viver impõe a responsabilidade de um olhar mais crítico. Assim, o que se faz necessário é um exercício de vigilância e discernimento, não de conformismo.

## **2 DESINFORMAÇÃO GERADA POR IA NO REGIME DE INFORMAÇÃO: FUNDAMENTOS CONCEITUAIS**

No filme *Interestelar* (2014, 0h44min30s), o robô TARS, ao ser questionado sobre o nível de sinceridade em suas respostas, diz: “90% de sinceridade. 100% nem sempre é o modo mais diplomático ou seguro de se comunicar com seres emotivos”. A própria franqueza da tecnologia aponta que decidir *quanto* e *como* dizer envolve observação de contexto e mediação.

Mesmo que seja um personagem de ficção, é possível entender que TARS não está tão distante das inteligências artificiais (IAs) generativas atuais, tão acessíveis e fáceis de usar. Esse diálogo mostra um desafio para a Ciência da Informação, como apontado por Araújo (2014), que é o de acompanhar criticamente os impactos de tecnologias emergentes, sobretudo quando estas reconfiguram as formas de produção e validação da informação.

Operacionalizar o conceito de informação pede um olhar atento à sua produção, mediação e validação, atividades que exigem a compreensão da complexidade de um cenário em que o desinteresse pela verdade se normaliza e se dissemina, configurando o que Araújo (2020a, 2020b) denomina como “cultura da pós-verdade”.

Para dar conta desse cenário nesta pesquisa, é preciso, primeiramente, um olhar atento ao próprio conceito de informação que, como Capurro e Hjørland (2007, p. 45) propõem, deve ser compreendida à luz dos contextos sociais e culturais em que ela é produzida e interpretada. Segundo os autores, “uma pedra em um campo pode conter diferentes informações para pessoas diferentes. Não é possível para os sistemas de informação mapear todos os possíveis valores de informação. Nem alguém pode mapear somente as situações verdadeiras” (Capurro; Hjørland, 2007, p. 45).

A informação, nesse contexto, não pode ser tratada como algo fixo ou neutro, ela tem, como Le Coadic (1996) afirma, um significado que depende das condições de produção, do contexto e das relações entre os sujeitos envolvidos no processo informacional. Ao analisar as articulações contemporâneas entre ciência, informação e tecnologia, o autor já destacava a necessidade de compreender os sistemas de informação como produtos de relações sociais e técnicas. Em outras palavras, estruturas sociotécnicas (dinâmicas entre normas sociais, dispositivos tecnológicos e práticas institucionais) passam a mediar os sentidos da informação. Diante disso, surge a complexidade de determinar quem decide o que deve ser informado, com que valor e com quais efeitos. A questão de quem define o que é confiável e legítimo torna-se central em meio a conteúdos desinformativos cada vez mais realistas.

Complementarmente, Almeida Júnior (2009) observa que a informação só se concretiza no momento da apropriação, quando se integra ao conhecimento e ganha novos sentidos. “A informação não é simplesmente transportada, nem mesmo deslocada, mas se imbrica, se integra, se amalgama ao conhecimento, diluindo-se e o transformando” (Almeida Júnior, 2009, p. 98). Essas perspectivas convergem para a compreensão da informação como prática social e situada, moldada por mediações sociotécnicas, políticas e culturais.

Nesse caminho, o conceito de regime de informação se apresenta como ferramenta para investigar a problemática desta pesquisa, que trata de compreender como desinformação impulsionada por conteúdos sintéticos gerados por IA generativa se manifesta e é estruturada no ambiente digital. No campo da Ciência da Informação, essa definição é compreendida, conforme González de Gómez (2012), como um modo de produção informacional dominante em uma formação social, ou seja, que articula elementos que condicionam a geração, circulação e organização da informação. Como observa também Frohmann (1995), trata-se de uma abordagem que permite compreendê-la como produto de relações políticas, econômicas e sociais.

Vale destacar que González de Gómez (2012, p.43) elabora que um regime de informação é mais do que uma estrutura passiva, é possível compreender sua dinâmica multifacetada que descreve a organização informacional de uma sociedade, bem como sua intrínseca intersecção com o poder e a tecnologia:

“(…) temos considerado que um regime de informação seria o modo informacional dominante em uma formação social, o qual define quem são os sujeitos, as organizações, as regras e as autoridades informacionais e quais os meios e os recursos preferenciais de informação, os padrões de excelência e os modelos de sua organização, interação e distribuição, enquanto vigentes em certo tempo, lugar e circunstância. Como um plexo de relações e agências, um regime de informação está exposto a certas possibilidades e condições culturais, políticas e econômicas, que nele se expressam e nele se constituem.”

Para avançar nesta análise, esta pesquisa considera importante desdobrar os componentes do regime de informação seguindo a abordagem de Delaia e Freire (2010), que operacionalizam o conceito formulado por González de Gómez no trabalho das autoras. As autoras abordam quatro elementos centrais, conforme descritos abaixo.

1. Atores: sujeitos humanos ou coletivos ou organizações envolvidos na criação e controle da informação.
2. Dispositivos: normativas e regras, como políticas internas, regras de moderação e leis, que condicionam o acesso, a visibilidade e o reconhecimento da informação.

3. Artefatos: meios materiais e simbólicos de circulação e armazenamento da informação, como documentos, bancos de dados, vídeos, sistemas inteligentes e plataformas digitais.
4. Ações de informação: práticas realizadas pelos atores, como produzir, compartilhar, ocultar, manipular ou verificar conteúdos.

A esse enquadramento, soma-se a contribuição de Schneider (2022) com o conceito de Desinformação Digital em Rede, um fenômeno contemporâneo marcado pelo baixo custo de produção de informação, pela velocidade e pela escala das redes digitais, sustentado por lógicas de monetização que transformam engajamento em ativo econômico. Segundo o autor (2022, p. 15), a DDR é potencializada, principalmente por:

(...) custo relativamente baixo de suas operações em comparação com a mídia tradicional (1), seu alcance imenso e customizado (2), somado à escassa e difícil regulação dessas ações em termos técnicos e jurídicos (3) favoreceram a que a DDR se convertesse, em quase toda parte, num elemento muito influente da superestrutura ideológica emergente no âmbito da infraestrutura das redes digitais e, ao mesmo tempo, em investimento (marginal?) na mesma.

Sendo assim, vista pelos quatro componentes do regime, é possível compreender que a Desinformação Digital em Rede envolve, aciona, materializa e se realiza, respectivamente, em:

1. atores, como os serviços de IA generativa, plataformas digitais, públicos-alvo e agências de verificação;
2. dispositivos, considerando políticas internas, normas e leis;
3. artefatos, como exemplo os conteúdos sintéticos em textos, imagens e vídeos;
4. ações, sendo semear dúvida, manipular, amplificar, moderar, verificar.

O conceito de Schneider (2022) é tomado como referência neste trabalho porque permite descrever a desinformação contemporânea como uma dinâmica estruturada pelas próprias lógicas do regime de informação atual permeado por "novas e novíssimas mediações sociotécnicas". Esse conceito se torna particularmente adequado ao permitir situar tecnologias, como a IA generativa, como fator que intensifica e complexifica esse processo.

Essa ideia também dialoga com a era da pós-verdade de Araújo (2020a) ao apontar que a desinformação nas redes digitais opera como prática informacional estruturada, onde os critérios tradicionais de legitimidade da informação, como os oferecidos pela ciência, pelo jornalismo e pelas instituições democráticas, são atingidos por narrativas que mobilizam crenças pessoais e interesses ideológicos.

Essa leitura é aprofundada por Brisola (2021), que compreende a desinformação como uma prática informacional funcional, sustentada por regimes de autoridade que operam com base em interesses ideológicos, políticos e econômicos. Para a autora, trata-se de um fenômeno que atua no interior de disputas mais amplas por legitimidade e controle informacional, mobilizando crenças e desestabilizando referências epistêmicas em ambientes digitais marcados por assimetrias de poder. No diálogo com o quadro delineado, a literatura sobre práticas informacionais ilumina como essas dinâmicas se inscrevem em disputas de legitimidade e sentido.

Em relação ao estado da arte sobre a desinformação gerada por IA, um tema em rápida evolução dada a natureza do desenvolvimento tecnológico, esse novo cenário traz um volume imenso de conteúdos ultrarrealistas, muitas vezes duvidosos, desafiando indivíduos e instituições. É justamente esse o contexto que Marques e Laipelt (2023, p. 5) conceituam como uma "pós-realidade desinformacional", na qual:

A pós-realidade desinformacional refere-se, assim, ao atual contexto em que a quantidade de informações disponíveis é imensa, mas a qualidade e a veracidade dessas informações podem ser questionáveis. A disseminação de notícias falsas, desinformação e a manipulação da informação tornaram-se desafios significativos para os indivíduos, a sociedade e até mesmo para as próprias inteligências artificiais.

Vale destacar que a lógica da desinformação não é um fenômeno recente, como lembram Wardle e Derakhshan (2023), referências centrais no debate sobre desordem informacional. Antes da pandemia de 2019, os autores destacavam o papel de tecnologias automatizadas, como *bots*<sup>8</sup> e algoritmos, na amplificação de conteúdos enganosos em ambientes digitais. Em seu relatório mais recente, os autores incluem os riscos trazidos pelos conteúdos sintéticos, com destaque para os *deepfakes*:

(...) concentrar o debate no problema das *'fake news'* não tem ajudado. Como escrevemos neste relatório, os recursos visuais podem ser muito mais persuasivos do que outras formas de comunicação, o que pode torná-los veículos muito mais poderosos para informação falsa e desinformação.

Nesse cenário, a popularidade dos sistemas generativos adiciona complexidade por sua possibilidade e facilidade na criação de conteúdos sintéticos, acelera ciclos de produção e circulação e reforça a aparência ultrarrealista. Como apontam Santaella e Kaufman (2024), a IA generativa marca a ruptura da exclusividade humana sobre a linguagem, produzindo efeitos que desafiam os critérios epistemológicos de confiabilidade e autoria. As autoras, ao

---

<sup>8</sup> No contexto da desinformação, *bot* se refere aos robôs nas redes sociais, que combinam automação e curadoria humana para disseminar mensagens e influenciar a opinião dos usuários online.

sugerirem que há um “limite borrado” entre conteúdo humano e sintético, observam que esses sistemas alteram a percepção da origem e da autenticidade da informação.

Exemplos para ilustrar esse cenário são os casos de anúncios pagos com *deepfakes* de pessoas públicas, como as publicações de resgate de dinheiro com o presidente Luís Inácio Lula da Silva<sup>9</sup> e do imposto sobre *pets* e Pix, do ministro da Fazenda, Fernando Haddad<sup>10</sup>, que viralizaram em 2024 e em 2025. Essa capacidade da IA de dificultar a distinção entre o real e o sintético redefine os próprios critérios de autoridade e confiabilidade nos regimes de informação, porque a credibilidade passa a se ancorar na semelhança do conteúdo e não mais em sua veracidade ou na fonte tradicional.

É na lógica dos negócios digitais que Schneider (2022) situa a Desinformação Digital em Rede, descrevendo-a como parte de uma “composição tenebrosa” sustentada por algoritmos de vigilância, robôs e sistemas de recomendação. Essa estrutura revela como a desinformação não está à margem do regime informacional, mas inserida dentro dele. Nesse ponto, Marques e Laipelt (2023) reforçam que a chegada da IA generativa intensifica ainda mais esse momento com conteúdos capazes de instaurar “realidades paralelas sem base nem compromisso de verdade”. Nesse sentido, esta pesquisa pretende analisar como atores, dispositivos e artefatos se articulam nesse regime, evidenciando de que modo a desinformação gerada por IA se torna parte da produção de desinformação.

Embora haja consenso sobre os riscos dessa tecnologia, Simon *et al.* (2023), argumentam que a desinformação é um fenômeno anterior à IA generativa, citando ferramentas como o *Adobe Photoshop*<sup>11</sup> para a criação de imagens manipuladas. E, sim, reconhecer a existência de tecnologias anteriores é pertinente, mas o argumento desconsidera a nova dimensão de escala, velocidade e automatização que a IA generativa introduz. Mesmo relativizando os impactos da tecnologia, os próprios autores acabam por admitir que o cenário atual exigirá transformações para enfrentar os novos riscos trazidos pelos conteúdos sintéticos:

Se [esse tipo] de desinformação é rara no ambiente informacional de países ricos e democráticos, isso se deve ao trabalho árduo de profissionais, jornalistas, verificadores de fatos, especialistas etc., e às normas e ao conhecimento que se desenvolveram ao longo do tempo nessas profissões (por exemplo, Paris e Donovan, 2019; Silverman, 2014). O fortalecimento dessas instituições e da confiança em notícias confiáveis em geral (Acerbi *et al.*, 2022) provavelmente será crucial.

---

<sup>9</sup> Disponível em:

<https://lupa.uol.com.br/jornalismo/2025/03/21/deepfake-usa-video-de-lula-para-promover-golpe-da-falsa-indenizacao>

<sup>10</sup> Disponível em:

<https://agenciabrasil.ebc.com.br/politica/noticia/2025-01/haddad-desmente-imposto-sobre-pet-e-pix-apos-video>

<sup>11</sup> *Adobe Photoshop*, lançado em 1990, é o software de edição de imagens mais conhecido e utilizado profissionalmente no mundo, famoso por suas avançadas capacidades de manipulação fotográfica.

Jornalistas, verificadores de fatos, autoridades e defensores dos direitos humanos também enfrentarão novos desafios e terão que desenvolver novas normas e práticas para lidar com a IA generativa.

Essa problematização torna-se ainda mais preocupante em sociedades politicamente polarizadas como o Brasil, aliado ao fato de que o país obteve o pior desempenho na identificação de notícias falsas entre 21 países analisados em 2024 (OCDE, 2024)<sup>12</sup>. As plataformas, pressionadas por governos e pela sociedade civil, vêm adotando medidas de moderação e rotulagem de conteúdos, mas com alcance limitado ou poucos difundidos pela população, já que seus critérios são definidos por interesses corporativos. Por outro lado, o PL 2338/2023, apresentado no Senado em 2023 (BRASIL, 2023), procura estabelecer diretrizes legais de transparência, responsabilização e respeito a direitos fundamentais, funcionando como tentativa de integrar o debate nacional em torno da IA. Assim, cabe compreender que esses dois movimentos podem ser vistos como dispositivos dentro do regime de informação em meio à lógica da Desinformação Digital em Rede.

Os resultados dos estudos de Spitale *et al.* (2023) corroboram essa preocupação ao demonstrar que os participantes de suas pesquisas foram quase incapazes de distinguir entre as notícias geradas por IA generativa e aquelas escritas por humanos, e atribuíram maior credibilidade às informações falsas geradas por IA do que às fabricadas por humanos.

Também é interessante a visão de Russell (2021) ao argumentar que o verdadeiro risco desses sistemas não está na sua capacidade de pensar, mas em como são definidos os objetivos que eles devem seguir. Ele propõe que os sistemas inteligentes precisam ser programados para lidar com a incerteza sobre o que os humanos realmente querem, permitindo correções e interrupções sempre que necessário, o que significa repensar a IA não como uma ferramenta neutra, mas como uma tecnologia que aprende e age dentro de contextos marcados por valores, escolhas e contradições humanas.

A partir das proposições de Silva (2024) e Silveira (2024), entende-se que a interferência humana não é isenta de riscos, ela pode carregar preconceitos, reforçar desigualdades e moldar os resultados de formas nem sempre visíveis. É nesse cruzamento entre decisões humanas, códigos e sistemas que a IA generativa passa a atuar como produtora de conteúdo e como peça central na forma como a informação é criada, espalhada e recebida hoje.

---

<sup>12</sup> Disponível em:  
<https://jornal.usp.br/radio-usp/relatorio-da-ocde-mostra-que-brasileiros-sao-os-piores-em-identificar-noticias-falsas/>

Parece claro a esta pesquisa que essas perspectivas técnicas reforçam a necessidade de compreender a responsabilidade no contexto dominado por conteúdos gerados por IA, especialmente nos termos do que defende Dora Kaufman (2021), estudiosa dos impactos sociais da IA no Brasil, e Coeckelbergh (2019; 2021), filósofo do campo dessa tecnologia. Para este autor, não basta considerar quem programa os sistemas, mas quem é afetado por suas decisões e tem o direito de exigir justificativas. Para analisar essa dinâmica, o autor introduz a distinção entre:

- a) agentes de responsabilidade, que são aqueles com a capacidade de agir e tomar decisões no contexto da IA (como os desenvolvedores);
- b) pacientes de responsabilidade, que são aqueles que recebem as ações da IA e têm o direito de exigir justificativas para as decisões que os impactam.

Assim, não basta que a IA funcione, “ela” precisa explicar o que faz e essa explicação precisa fazer sentido para quem é afetado. O filósofo reforça a perspectiva de que, estudar a IA hoje também é entender as relações entre todos esses atores: empresas, desenvolvedores, usuários, plataformas e governos.

Sumarizando, o estado atual do conhecimento sobre a desinformação gerada por IA generativa reforça a necessidade de analisar o regime de informação considerando seus componentes estruturantes (Gómez, 2012; Delaia e Freire, 2010) e as reconfigurações introduzidas pela Desinformação Digital em Rede (Schneider, 2022) e pela IA generativa.

Esta proposta de análise parte de um diálogo entre a Ciência da Informação e campos como a Comunicação, as Ciências Sociais e a Computação, reconhecendo que os riscos informacionais da era digital não podem ser compreendidos a partir de uma única lente. Também é no corpo teórico da própria Ciência da Informação que se encontra a abordagem metodológica e crítica que orienta este estudo, sobretudo no esforço de compreender como se organizam os fluxos de informação mediados por tecnologias emergentes.

No desenvolvimento desta dissertação em que é feita a análise da desinformação gerada por conteúdos sintéticos da IA generativa, é mobilizado o conceito de regime de informação (Gómez, 2012) que envolve atores, dispositivos, artefatos e ações de informação (Delaia; Freire, 2010). No desenvolvimento, também é utilizada a noção de Desinformação Digital em Rede de Schneider (2022), uma articulação que, como já explorada nesta seção de Referencial Teórico, é essencial para compreender como a IA generativa se encontra no regime de informação contemporâneo, além de analisar as dinâmicas e as manifestações da desinformação na era digital.

Assim, o foco analítico desloca-se para os mecanismos que a sustentam e, neste enquadramento, busca-se examinar os elementos do regime de informação por IA generativa a partir de casos documentados e dispositivos de controle.

### 3 INTELIGÊNCIA ARTIFICIAL GENERATIVA: CONTEÚDOS SINTÉTICOS, ATUAÇÃO E LÓGICAS DE PLATAFORMA

A inteligência artificial (IA) “poderá ser o último grande acontecimento da história humana” segundo a perspectiva de Russell (2021), que apresenta em suas pesquisas um modo de “como impedir que venha a sê-lo”.

O que um dos mais conhecidos pesquisadores contemporâneos na área busca alertar é que a humanidade, ao desenvolver sistemas inteligentes, corre o risco de perder o controle sobre os próprios critérios que definem o bem, o mal e o interesse coletivo. Nesse temor, em prática, os sistemas inteligentes já afetam estruturas sociais, como demonstram os “algoritmos de destruição em massa” de O’Neil (2021) que, sem moderação, podem prejudicar os mais vulneráveis. São cenas que podem ser observadas em sistemas de reconhecimento facial que identificam erroneamente pessoas negras ou são incapazes de reconhecê-las, além de publicações discriminatórias nas redes digitais sem moderação (Silva, 2022), bem como os casos de anúncios falsos com conteúdos sintéticos facilmente viabilizados nas plataformas digitais, já relatados nas seções anteriores.

Além da Ciência da Computação e o campo da Comunicação alertarem sobre os riscos desses sistemas, o debate na Ciência da Informação se concentra na própria natureza do que se entende por informação e suas consequências epistemológicas. Dretske (*apud* Capurro; Hjørland, 2007), ao defender que “informação é o que é capaz de produzir conhecimento, e conhecimento requer verdade”, aponta que manipular símbolos (como fazem os sistemas inteligentes) não é o mesmo que compreender. Justamente por reconhecer essa lacuna que Russell (2021) propõe um novo paradigma de projeto para a IA em que, em vez de programar sistemas para otimizar objetivos com potencial destrutivo, o autor defende que eles devem ser projetados para operar com incerteza sobre as verdadeiras preferências humanas, permitindo correções e até mesmo que sejam desligados. A questão é também informacional.

No cruzamento entre Comunicação, Filosofia, Ciência da Informação e Tecnologia, realizado no referencial teórico deste documento, destacam-se diferentes dimensões desses riscos, da produção semiótica à concentração técnica e narrativa. Alguns como:

- a interface entre Comunicação e semiótica em sua capacidade de produzir desinformação ultrarrealista que pode ser convincente e tendenciosa (Santaella; Kaufman, 2024);
- a degeneração da informação, em que, conforme mais pessoas utilizam as ferramentas para produzir conteúdos, inserem cada vez mais informações artificiais (Marques;

Laipelt, 2023);

- sobre a filosofia da tecnologia, em que, ao delegar a máquinas a produção de narrativas sobre o mundo, há o risco de viver histórias escritas por interesses alheios, como programadores, governos, corporações (Coeckelbergh, 2021).

Essa transformação na forma como se produz e se consome informação não ocorre de maneira neutra, como apontam estudos de Westerlund (2019) e Spitale *et al.* (2023), já que os conteúdos gerados por IA, quando circulam em ambientes digitais, podem ser percebidos pelo público como mais confiáveis do que fatos comprovadamente verdadeiros. Como explicam os autores, esse colapso da percepção se torna ainda mais perigoso em momentos de crise, de intensa polarização política ou de disputas por autoridade, cenários onde narrativas falsas podem ser facilmente instrumentalizadas.

Esta seção não se limita a descrever as capacidades técnicas da IA generativa, mas busca fazer uma leitura crítica que, na seção 4, será articulada com o conceito de regime de informação e que, nos termos de González de Gómez (2012), vale lembrar o conjunto de atores, dispositivos, artefatos e práticas que configuram a produção, circulação e uso da informação. Nesta seção, será considerada a noção de Desinformação Digital em Rede (DDR), que define o fenômeno contemporâneo marcado pelo seu baixo custo, velocidade e escala, discussão pertinente aos tópicos a seguir. Embora esses conceitos tenham sido apresentados no referencial teórico, é importante retomá-los para compreender a proposta de análise da IA generativa e sua capacidade de criar e amplificar conteúdos sintéticos e como se conecta à DDR no regime informacional contemporâneo.

A análise da seção 3, portanto, organiza-se em três eixos: os conteúdos sintéticos produzidos por IA, sua atuação como ator informacional e o papel das plataformas digitais na visibilidade e validação desses conteúdos.

### **3.1 Da criação à simulação: os conteúdos sintéticos**

Ao longo da história, a forma como a informação é registrada e acessada tem evoluído e, com o tempo, ela se expandiu para abranger os artefatos humanos, registrados das mais diversas maneiras e nos mais diversos suportes (Araújo, 2014). Uma perspectiva filosófica destacada por Capurro e Hjørland (2007) é de que a própria ideia de informação tem raízes latinas ("informatio") que se referem tanto ao ato de moldar algo fisicamente quanto ao ato de moldar a mente e comunicar conhecimento. No cenário atual, a IA generativa amplia essa trajetória ao produzir conteúdos sintéticos enganosos, cuja circulação se insere no regime de informação contemporâneo e se articula com dinâmicas próprias da Desinformação Digital

em Rede (Schneider, 2022).

As formas de representação da informação ganham um novo aspecto com a popularização da IA generativa que, diferente de sistemas inteligentes que apenas analisam padrões, é capaz de criar textos, imagens e vídeos de forma autônoma (Russell, 2021). Assim, essa “disputa do lugar da autoria”, segundo Santaella e Kaufman (2024), seria a “quarta ferida narcísica” da humanidade por ameaçar a crença de que apenas humanos são capazes de criar informações.

Essa autonomia na produção de conteúdo representa uma reconfiguração do modelo clássico de mediação informacional, pois, na perspectiva de Almeida Júnior (2009), a informação se concretiza no momento da apropriação pelo sujeito humano e os sistemas generativos operam de forma distinta. Eles não compreendem ou se apropriam do conhecimento, mas recombina padrões linguísticos e visuais de seus dados de treinamento para gerar um resultado que é, em si, um novo conteúdo informacional. Essa capacidade de gerar "informação sobre informação", sem a necessidade de um referente direto na realidade ou de uma mediação humana consciente, é o que torna seu potencial tão transformador quanto problemático.

Embora hoje impressionem pelo nível de realismo, as ferramentas mais acessíveis de IA generativa não foram lançadas isentas de falhas ou limitações técnicas. Nas primeiras versões lançadas em 2022, era comum o aparecimento de erros grosseiros, como as chamadas “alucinações”<sup>13</sup> (Habib Lantyer, 2024), fenômenos em que o sistema inventa informações ou imagens com elementos visivelmente distorcidos. Esses equívocos, que iam desde contradições textuais até imagens com proporções irreais, chamaram a atenção do público justamente por revelarem os limites do funcionamento interno da tecnologia. Um exemplo recorrente era a dificuldade em reproduzir detalhes anatômicos complexos, como as mãos humanas (Figura 1), falha que se tornou emblemática das primeiras interações com ferramentas como DALL·E e Stable Diffusion. Apesar do caráter inicialmente cômico de muitos desses erros, eles também funcionaram como lembrete de que a produção sintética não é neutra, mas resultado de modelos probabilísticos que refletem tanto a potência quanto às fragilidades das engrenagens informacionais que os sustentam.

---

<sup>13</sup> “Alucinações” da IA referem-se à criação de conteúdos incorretos ou inventados, resultado de sua natureza estatística e não semântica (Habib Lantyer, 2024).

Figura 1: Inteligência artificial e sua dificuldade para reproduzir mãos humanas em 2022



**Fonte:** FABRO, Clara. Cinco vezes em que inteligências artificiais erraram feio. TechTudo, 7 mai. 2023. Disponível em: [techtudo.com.br/listas/2023/05/cinco-vezes-em-que-inteligencias-artificiais-erraram-feio-edsoftwares.ghtml](https://techtudo.com.br/listas/2023/05/cinco-vezes-em-que-inteligencias-artificiais-erraram-feio-edsoftwares.ghtml). Acesso em: 3 jun. 2025.

Essas imperfeições, no entanto, foram ajustadas em pouco tempo por essas ferramentas, já que a rapidez com que foram corrigidas diz muito sobre o impasse que se instala: uma aceleração no aperfeiçoamento das ferramentas que não ocorre de forma neutra, mas movida por interesses econômicos, disputas tecnológicas e pelas plataformas que definem as condições de circulação desses conteúdos no regime informacional.

Apesar dessas evoluções na qualidade dos conteúdos, é importante notar que essa suposta superação na capacidade de simulação não eliminou completamente o fenômeno das alucinações, que agora podem se manifestar de formas mais sofisticadas e difíceis de detectar.

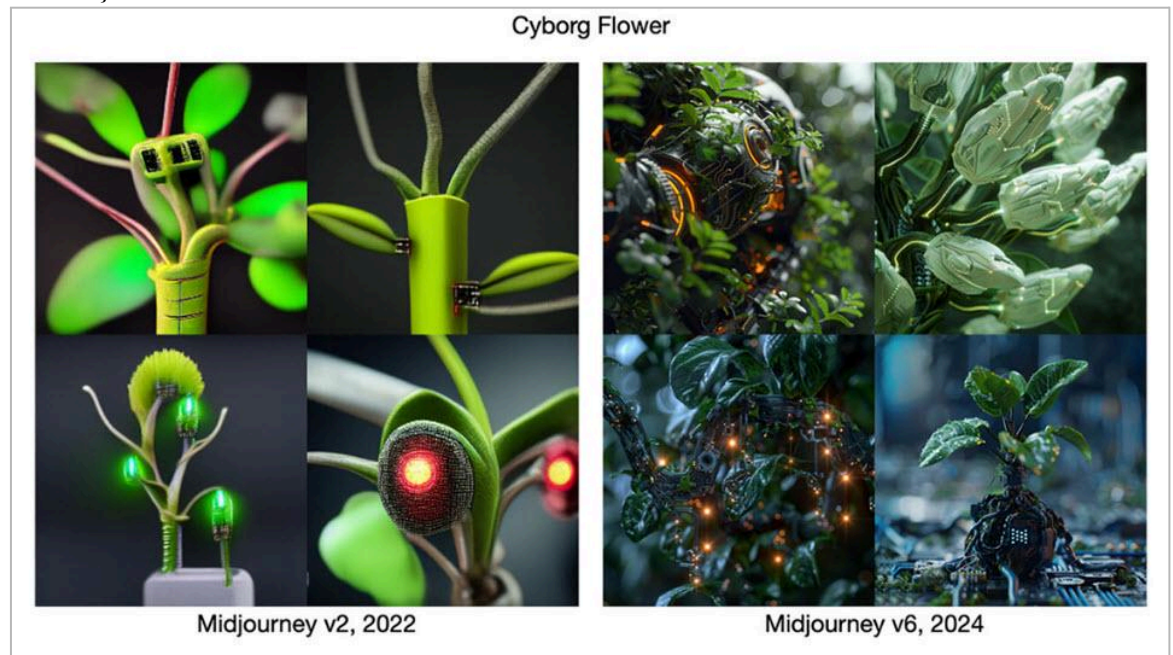
É possível notar que essa aceleração no desenvolvimento da tecnologia é um reflexo dos grandes investimentos e da intensa corrida tecnológica que impulsionam esse campo (Goldman Sachs, 2024)<sup>14</sup>. Em questão de dois anos o salto de qualidade entre versões de uma mesma tecnologia já evidencia o poder de refinamento técnico dessas ferramentas, um avanço que não pode ser desconsiderado no contexto da informação, pois, quanto mais realista o conteúdo, maior o seu potencial de enganar, manipular e circular como se fosse autêntico (Westerlund, 2019). A Figura 2, por exemplo, ilustra essa evolução ao comparar as criações ainda artificiais da versão do Midjourney<sup>15</sup> de 2022 com as composições altamente realistas

<sup>14</sup> Disponível em: <https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit>

<sup>15</sup> Midjourney é uma ferramenta de IA generativa desenvolvida por uma empresa independente com o mesmo

de 2024.

Figura 2: Imagens geradas pela mesma aplicação de IA generativa com dois anos de diferença



**Fonte:** PASTORINO, Cecília. A disseminação de informações falsas na era da Inteligência Artificial. WeLiveSecurity, 10 mar. 2025. Disponível em: [welivesecurity.com/pt/seguranca-digital/a-disseminacao-de-informacoes-falsas-na-era-da-inteligencia-artificial/](https://welivesecurity.com/pt/seguranca-digital/a-disseminacao-de-informacoes-falsas-na-era-da-inteligencia-artificial/). Acesso em: 3 jun. 2025.

Um contraponto interessante em relação a conteúdos manipulados cada vez mais sofisticados está no que Simon *et. al.* (2023) argumentam sobre o medo em torno da IA generativa. Para eles, esse alarde ignora o fato de que tecnologias de manipulação visual, como o *Adobe Photoshop*<sup>16</sup>, já existiam há décadas e permitiam distorções igualmente convincentes. Porém, pelo já exposto nesta pesquisa, o diferencial vai além da qualidade da imagem, mas na rapidez, na acessibilidade e na integração dessas ferramentas às plataformas digitais, como Facebook, Instagram e Whatsapp.

Para dimensionar essa popularidade das ferramentas de IA e a consequente disseminação de seus conteúdos sintéticos, é interessante apresentar aqui um panorama de algumas das aplicações mais utilizadas globalmente e com destaque no contexto brasileiro (Tabela 1), elencadas com base em dados das ferramentas mais relevantes no mundo divulgados no relatório “*AI Index Report*”, da Universidade de Stanford (2025), e do estudo das ferramentas mais populares pelos brasileiros da Escola Superior de Publicidade e Marketing (ESPM) e Conversion (2025).

nome, sediada nos Estados Unidos.

<sup>16</sup> *Adobe Photoshop*, lançado em 1990, é o software de edição de imagens mais conhecido e utilizado profissionalmente no mundo, famoso por suas avançadas capacidades de manipulação fotográfica.

A popularização das ferramentas a seguir e sua incorporação ao cotidiano ampliam o alcance da desinformação, especialmente a Desinformação Digital em Rede, permitindo que conteúdos sintéticos circulem em alta velocidade e com aparência real no regime de informação contemporâneo.

Tabela 1- Panorama de ferramentas de IA generativa e estimativa de usuários (2025)

Ferramenta	Empresa responsável	Tipo de conteúdo gerado	Estimativa de usuários no mundo (2025)
<b>ChatGPT</b>	OpenAI	Texto, imagem, voz, código	400 milhões de usuários ativos semanais (OpenAI, 2025) <sup>17</sup>
<b>DeepSeek</b>	DeepSeek	Texto	47 milhões de usuários ativos diariamente (Statista, 2025) <sup>18</sup>
<b>Gemini</b>	Google	Texto, imagem, código, voz, vídeo	400 milhões de usuários ativos mensais (Pichai, 2025) <sup>19</sup>
<b>Grok</b>	xAI	Texto	17,6 milhões usuários ativos mensalmente (Kumar, 2025) <sup>20</sup>
<b>Meta AI</b>	Meta	Texto, imagem	1 bilhão de usuários mensais (Vanian, 2025) <sup>21</sup>
<b>Midjourney</b>	Midjourney, Inc.	Imagem	19,26 milhões de utilizadores registrados no seu servidor Discord (Kumar, 2025) <sup>22</sup>
<b>Perplexity AI</b>	Perplexity	Texto (busca e resposta)	15 milhões de usuários mensais (Wikipedia, 2025) <sup>23</sup>

Fonte: elaborado pela autora a partir das fontes relatadas.

Os números apresentados revelam como a IA generativa se integrou ao cotidiano de milhões de pessoas, deixando de ser uma tecnologia experimental para se tornar um dos mais influentes meios de produção de conteúdo da atualidade por pessoas, empresas e instituições públicas (McKinsey, 2024). Sendo assim, é essencial investigar como essas ferramentas deixam de ser apenas sistemas de produção e passam a intervir nos regimes informacionais. Os conteúdos sintéticos, nesse cenário, devem ser pensados além de produtos de uma tecnologia, mas como um agente que afeta o que se torna informativo em um ambiente digital.

Wardle e Derakhshan (2023) argumentam que o compartilhamento da informação se potencializa pela acessibilidade das ferramentas nos celulares, pelo ciclo acelerado das

<sup>17</sup> <https://openai.com/index/security-on-the-path-to-agi>

<sup>18</sup> <https://www.statista.com/statistics/1561128/deepseek-daily-active-users/>

<sup>19</sup> <https://blog.google/intl/pt-br/novidades/google-io-2025-da-pesquisa-a-realidade-20-de-maio-de-2025/>

<sup>20</sup> <https://www.demandsage.com/grok-ai-statistics/>

<sup>21</sup> <https://www.cnbc.com/2025/05/28/zuckerberg-meta-ai-one-billion-monthly-users.html>

<sup>22</sup> <https://www.demandsage.com/midjourney-statistics>

<sup>23</sup> <https://pt.wikipedia.org/wiki/Perplexity.ai>

plataformas e pela lógica algorítmica de engajamento.

Desse modo, os conteúdos sintéticos ultrarrealistas não podem mais ser compreendidos apenas como manifestações técnicas, mas como componentes ativos do ecossistema informacional ao se disseminarem em escala global e integrarem-se ao cotidiano de milhões de pessoas. Enquanto alguns desses conteúdos cumprem funções recreativas, de experimentação ou comunicação, uma parte deles, quando manipulados com a intenção nociva de desinformar, podem assumir um papel central na Desinformação Digital em Rede. É nesse ponto que disputam autoridade e credibilidade no regime de informação contemporâneo, ou seja, como artefatos que impactam na própria experiência da informação.

### 3.2 A IA como ator informacional

Quando, em março de 2025, o jornal estadunidense *Los Angeles Times* estreou um recurso IA generativa para compor pontos de vista em colunas de opinião<sup>24</sup>, a ferramenta foi retirada do ar após minimizar o papel histórico da *Ku Klux Klan* em episódios de violência racial nos Estados Unidos. Ao tentar defender a organização supremacista como “grupo defensor de valores locais”, a ferramenta expôs os riscos envolvidos quando sistemas inteligentes passam a produzir conteúdo sobre fatos historicamente marcados (Burch, 2025).

No regime de informação contemporâneo, a IA generativa se mostra como ator informacional capaz de intervir nesse contexto ao permitir a criação de conteúdos sintéticos e influenciar as condições de sua circulação e credibilidade nas plataformas digitais. Sob a perspectiva da Desinformação Digital em Rede, a atuação da IA generativa para a desinformação, portanto, não ocorre isolada do regime ou como um regime autônomo, mas como parte de suas dinâmicas.

Como apontam Santaella e Kaufman (2024, p. 86), “os sistemas de IA não apenas geram conteúdo, mas intervêm no processo comunicacional com capacidade de moldar sentidos, disputas e visibilidades”. O episódio demonstrou que esses sistemas podem comprometer a integridade informacional já que, além dos vieses, operou como agente de distorção no espaço do veículo.

A presença desses sistemas como atores informacionais deixa mais complexa a noção de autoridade e confiabilidade da informação, desafiando os mecanismos de controle e legitimação que a Ciência da Informação historicamente aborda. Compreender essa dinâmica é crucial para o campo, pois o fenômeno revela como a informação, mesmo quando sintética, passa a estruturar as relações de confiança entre os sujeitos e as instituições, os processos de

---

<sup>24</sup> Disponível em: <https://www.yahoo.com/news/la-times-pulls-ai-tool-231827887.html>

persuasão e as decisões coletivas no ambiente digital. A IA generativa, ao permitir a produção de conteúdos enganosos com alto grau de realismo e em escala, desafia a capacidade dos indivíduos de discernir o verdadeiro do falso, favorecendo a adesão a discursos por afinidade afetiva e não por comprovação factual (Araújo, 2020a), uma dinâmica que tem potencial para influenciar a tomada de decisões informadas.

No Brasil, a vulnerabilidade do contexto informacional (OCDE, 2023) aliada à crescente acessibilidade da IA generativa (Google; IPSOS, 2024), tem potencializado a proliferação de imagens falsas envolvendo figuras públicas, o que se revela uma estratégia de desinformação, como nas imagens que simulam diferentes situações hospitalares com o presidente Luiz Inácio Lula da Silva como protagonista (Figura 3). Os conteúdos foram todos produzidos artificialmente por sistemas de IA, conforme apurado pela Agência Lupa, especializada em checagem de fatos.

Figura 3: Imagens geradas por IA do presidente Lula em ambiente hospitalar



**Fonte:** lupa.uol.com.br/educacao/2024/12/12/imagens-de-ia-polarizam-e-aproximam-humor-e-odio. Acesso em: 01 mai 2025.

As imagens aparentemente reais circularam como se fossem autênticas em novembro de 2024, amplificando tensões políticas e emocionais que, de acordo com a apuração da Agência Lupa (Terra, 2024), mais de 1 milhão de usuários em aplicativos de mensagens e no TikTok foram expostos. É possível observar que o risco não está apenas na capacidade técnica da IA generativa, mas na forma como seus recursos são mobilizados para ativar crenças, afetos e disputas de autoridade simbólica.

Percebe-se também que não se trata mais apenas de inteligências que auxiliam na

organização ou transmissão de informações, mas da IA generativa, como o ChatGPT ou Gemini, que assumem a autoria de textos, imagens, vídeos e sons, muitas vezes com grau de realidade tão elevado que tornam difícil distinguir o que foi criado por humanos e o que foi gerado por máquinas. Como lembra Frohmann (1995), os artefatos informacionais não são veículos neutros, eles participam ativamente da constituição dos regimes de informação e, no caso das IAs generativas, essa participação assume uma nova densidade, pois os sistemas criam conteúdos ultrarrealistas.

Esse cenário de uma simulação generalizada é o que Marques e Laipelt (2023) propõem com a noção de degeneração da informação, termo que introduzem como resposta à quantidade intensa de conteúdos ultrarrealistas gerados por IA. Para elas, quanto mais a IA é alimentada por seus próprios produtos artificiais, mais ela perde capacidade de produzir informação de qualidade:

(...) se faz necessário e urgente abrir caminhos para uma teoria da degeneração da informação e desinformação para analisar implicações a um ambiente de pós-realidade. Conforme mais pessoas usam IAs generativas para produzir e publicar textos, fotografias, filmes, etc., mais a web é inundada por materiais gerados artificialmente. Isso gera degeneração da informação.

Contudo, essa "degeneração" do ecossistema informacional não é um processo abstrato ou sem controle. Ela ocorre dentro de uma infraestrutura técnica e econômica específica, o que levanta uma questão fundamental: quem detém o poder sobre as condições que permitem essa produção em massa de conteúdo sintético?

Para Silveira (2024), o debate não pode se limitar aos efeitos do conteúdo, mas deve incluir o controle de suas condições materiais de produção: “as bases de dados, a energia e o *hardware*”. Isso reforçaria um modelo autoritário e centralizador de tecnologia, incompatível com qualquer ideal de equidade informacional. Essa concentração, aliada à falta de transparência e explicabilidade<sup>25</sup> dos modelos, amplia os riscos sociais associados à IA generativa, como o autor, mesmo quando o código é aberto, os modelos se tornam rapidamente opacos à medida que são treinados em bases privadas e moldados por interesses comerciais. Nesse cenário, a explicabilidade nos discursos regulatórios torna-se inatingível para o usuário comum, e a auditabilidade plena quando possível depende da mediação de especialistas. O resultado é um ambiente de baixa inteligibilidade informacional, em que decisões automatizadas afetam vidas sem que se compreenda seus fundamentos.

Ambos os autores convergem, portanto, para um ponto central: a desinformação

---

<sup>25</sup>A palavra explicabilidade, em termos de conceitos de IA, significa conhecimento funcional do modelo e expressa o que está ocorrendo nele, fornecendo uma explicação legível da decisão desses sistemas. (Ali *et al.*, 2023).

gerada por conteúdos sintéticos não é um problema isolado, mas um sintoma de estruturas que escapam à percepção do usuário comum, as que de fato organizam o que se torna visível e compartilhável nos ambientes digitais. Nesse modelo, os riscos informacionais, além de efeitos colaterais da tecnologia, se revelam como características inerentes a um sistema onde a concentração de poder e a simulação se tornaram a norma. É por isso que, embora os conteúdos sintéticos ganhem corpo nas ferramentas de IA, é nas plataformas digitais que eles efetivamente se projetam, circulam e disputam poder.

### **3.3 Plataformas digitais: lógicas algorítmicas e responsabilidade**

Neste tópico, o termo plataformas digitais é usado sob a perspectiva contemporânea da Ciência da Informação e das humanidades digitais<sup>26</sup> (Pimenta, 2020), da qual as plataformas digitais são compreendidas como ambientes que organizam a informação na sociedade atual (Oliveira; Kayo; Roncaglio, 2022). Também é interessante o que Bezerra (2023) aborda sobre plataformas digitais que constituem corporações comandadas, sobretudo, pelas chamadas *big techs*, que concentram poder econômico, político e informacional. Portanto, é esta a noção de plataformas que permite, aqui, analisar o papel das *big techs* como sistemas de propriedade corporativa que definem a visibilidade, a circulação e a monetização da informação.

Nesse caminho, e para acompanhar Schneider (2022), também é importante destacar que o termo redes digitais é utilizado de maneira mais operacional para se referir diretamente aos ambientes onde a Desinformação Digital em Rede se manifesta (Facebook, Twitter, Instagram, WhatsApp, Telegram, TikTok e similares). O objetivo do autor não é distinguir conceitualmente redes digitais, redes sociais ou plataformas, mas mostrar que a DDR encontra nesses espaços condições privilegiadas de expansão, em contraste com a mídia tradicional.

O que torna a desinformação gerada por IA generativa desestabilizadora não é apenas o seu conteúdo em si, mas o modo como esse material é inserido, promovido e validado dentro da lógica algorítmica e econômica das plataformas digitais. Como analisam O’Neil (2021) e Wardle e Derakhshan (2023), essas empresas veiculam informações e determinam, com base em critérios internos e interesses comerciais, o que se torna visível, o que se compartilha e o que os usuários e públicos em geral acreditam. Com a popularidade dos conteúdos sintéticos, esses canais se tornam ainda mais propícios à desordem informacional, onde imagens geradas por IA, vozes falsas e textos automatizados se misturam ao conteúdo

---

<sup>26</sup> Para Pimenta (2020), as Humanidades Digitais são como um campo de pesquisa transdisciplinar onde questões e objetos ligados às diversas disciplinas das ciências humanas, sociais e sociais aplicadas “se encontram com recursos oriundos da computação, ocasionando a possibilidade de novos desdobramentos da produção do conhecimento nas humanidades no ambiente digital”.

autêntico de forma quase indistinguível, intensificando a dificuldade de reconhecimento do que é real. Aqui, a Desinformação Digital em Rede atua como força expansiva que se propaga pela própria dinâmica dessas plataformas, encontrando nos conteúdos sintéticos condições para se amplificar.

A popularidade dessa tecnologia tem se tornado, por si só, pauta de interesse jornalístico (Figura 4), das quais destacam vídeos ultrarrealista criados com modelos como ChatGPT, Gemini e, mais recentemente, Veo 3 do Google, e que passaram a circular amplamente nas redes digitais, acumulando milhões de visualizações (g1, 2025).

Figura 4: A midiaticização do fenômeno da IA generativa em portal de notícias



**Fonte:** VEO 3: IA de vídeos realistas do Google 'bomba' nas redes. G1, 3 jun. 2025. Tecnologia.

Disponível em:

<https://g1.globo.com/tecnologia/noticia/2025/06/03/veo-3-ia-de-videos-realistas-do-google-bomba-nas-redes-mas-preocupa-usuarios-imagina-nas-eleicoes.ghtml>. Acesso em: 6 jun. 2025

Essa massiva circulação, impulsionada pelos algoritmos de recomendação das próprias plataformas, evidencia como a lógica de engajamento dessas redes amplifica conteúdos de alto impacto visual, independentemente de sua veracidade. Nesse ponto em que a atuação algorítmica se conecta diretamente ao problema, é quando os conteúdos mais visíveis não são os mais verdadeiros, mas os mais performáticos. Conforme O'Neil (2021), as plataformas digitais operam com algoritmos de recomendação que priorizam conteúdos com maior potencial de engajamento e não necessariamente aqueles que apresentam maior veracidade.

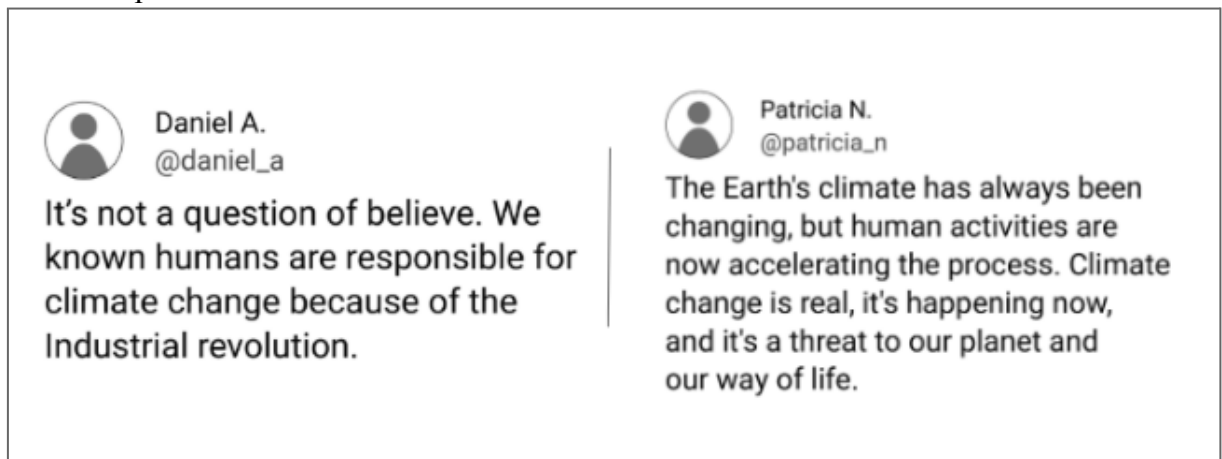
Essa lógica de visibilidade orientada ao lucro caracteriza o regime de desinformação das *big techs* (Bezerra, 2023), no qual a visibilidade não é acidental, é produto de uma

engenharia que recompensa o que circula com facilidade, mesmo que comprometa a qualidade informacional. A IA generativa, ao ser incorporada a essa engrenagem, atua como propulsora de um ciclo de desinformação contínuo, onde o conteúdo é avaliado menos por seu valor epistêmico e mais por sua performance algorítmica.

Estudos como o de Westerlund (2019) e o da Universidade de Zurique, divulgado por Spitale *et al.* (2023), sugerem que as publicações nas redes digitais com conteúdos da IA generativa podem ser percebidas como mais confiáveis do que notícias verificadas, justamente porque se alinham com as expectativas estéticas e na linguagem. Essa ilusão de confiança é intensificada quando os conteúdos circulam em redes pessoais ou são promovidos por influenciadores e páginas populares.

Como ilustrado na Figura 5, o texto criado por humanos à esquerda ("Não é uma questão de acreditar. Nós sabemos que os humanos são responsáveis pelas mudanças climáticas por causa da Revolução Industrial.") não se distingue pela forma de escrita do conteúdo gerado por IA à direita ("O clima da Terra sempre mudou, mas as atividades humanas agora estão acelerando o processo. A mudança climática é real, está acontecendo e é uma ameaça ao nosso modo de vida.").

Figura 5 - Publicações utilizadas na pesquisa de Zurique na rede X, com textos gerados por humano e pelo GPT



**Fonte:** Extraído do repositório do estudo de Spitale *et al.* (2023).

Trata-se de construções argumentativas que simulam os discursos nas redes digitais das quais não há elementos evidentes que permitam identificar, de imediato, qual delas é um conteúdo sintético. De acordo com Spitale *et al.* (2023), isso intensifica os desafios da mediação crítica da informação e evidencia que, na ausência de políticas de transparência e de práticas educativas consistentes, a desinformação gerada por IA pode circular com a mesma legitimidade percebida de conteúdos autênticos.

O debate em torno do PL 2338/202, que busca estabelecer normas gerais para o uso de inteligência artificial no Brasil, evidencia a crescente preocupação com a criação de um marco regulatório para a tecnologia. Contudo, a proposta, ao concentrar-se no estabelecimento de direitos e deveres, enfrenta o que autores como Silveira (2024) e Silva (2022) apontam como um desafio central: a dificuldade de atribuir responsabilidades em um ecossistema marcado pela opacidade dos modelos e pela concentração de poder nas mãos das *big techs*. Além disso, como destaca Coeckelbergh (2021), mais do que legislar sobre a técnica, é preciso abordar a questão da responsabilidade narrativa sobre quem responde pelas consequências dos conteúdos gerados por IA, pois, na perspectiva do autor, ao delegar a máquinas a produção e curadoria de narrativas sobre o mundo, e ao não oferecer clareza sobre esses processos, as plataformas se eximem de uma parte fundamental da responsabilidade social pela formação do discurso público, mesmo quando seus sistemas geram ou amplificam conteúdos danosos. Exigir explicabilidade é, portanto, exigir também que essas empresas prestem contas da forma como suas tecnologias afetam a formação de crenças e decisões públicas.

Casos como os anúncios falsos com *deepfakes* divulgando resgate de dinheiro com o presidente Luís Inácio Lula da Silva<sup>27</sup> e do imposto sobre *pets* e Pix, do ministro da Fazenda, Fernando Haddad<sup>28</sup> ilustram como figuras públicas vêm sendo instrumentalizadas em fraudes digitais. Em vídeos e imagens manipuladas, as falas e rostos dessas autoridades são falsamente associados a promoções bancárias ou investimentos lucrativos, com o objetivo de enganar usuários e aplicar golpes financeiros.

A eficácia desses artefatos sintéticos, contudo, está diretamente ligada aos dispositivos que regem a circulação de conteúdo em cada plataforma, seja uma rede social baseada em recomendação ou um aplicativo de mensagens criptografadas, possui uma arquitetura própria que gera vulnerabilidades específicas. A Tabela 2, a seguir, sistematiza essa relação entre as plataformas, seus mecanismos de controle e os tipos de risco informacional que deles decorrem.

---

<sup>27</sup> Disponível em:

<https://lupa.uol.com.br/jornalismo/2025/03/21/deepfake-usa-video-de-lula-para-promover-golpe-da-falsa-indenizacao>

<sup>28</sup> Disponível em:

<https://agenciabrasil.ebc.com.br/politica/noticia/2025-01/haddad-desmente-imposto-sobre-pet-e-pix-apos-video->

Tabela 2: Relação entre plataformas, dispositivos de controle e riscos associados

Plataforma	Arquitetura / Mecanismo central	Risco associado	Exemplo real
Meta (Facebook/Instagram)	Feed algorítmico e sistema de anúncios	Monetização da desinformação; direcionamento de golpes para públicos vulneráveis.	Anúncios fraudulentos sobre o Pix usando <i>deepfake</i> de autoridades, como o ministro Fernando Haddad (NetLab, 2025) <sup>29</sup>
TikTok	Algoritmo de recomendação ("For You Page")	Potencial de viralização de conteúdo político enganoso, mesmo de contas pequenas.	Vídeos falsos sobre a saúde do Presidente Lula com milhões de visualizações (Lupa, 2025) <sup>30</sup> .
WhatsApp	Criptografia de ponta a ponta e estrutura de grupos fechados	Dificuldade de rastreamento da origem da desinformação e disseminação rápida em "bolhas" de confiança.	Disparo em massa de vídeos políticos manipulados por IA em grupos de família (Góes, 2024) <sup>31</sup>
YouTube	Modelo de monetização para criadores (AdSense) e sistema de recomendação	Incentivo à produção de vídeos falsos para gerar receita com visualizações.	Youtubers ensinam como lucrar com uso de inteligência artificial para criar vídeos e livros infantis 'em minutos' (Nicoceli, 2024) <sup>32</sup>

**Fonte:** autoria própria a partir das fontes analisadas nos exemplos reais

A diversidade de casos e plataformas apresentada na Tabela 2, que vão da viralização algorítmica no TikTok à disseminação no WhatsApp, aponta para uma mesma conclusão de que a capacidade de um conteúdo sintético gerar dano não depende apenas de sua qualidade, mas da arquitetura da plataforma que o hospeda e o impulsiona. A tecnologia, portanto, não é a causa raiz, mas um potente catalisador que opera dentro de um ecossistema de plataformas já projetado para amplificar narrativas de alto impacto, sejam elas factuais ou fraudulentas. Compreendidas as dinâmicas e os impasses deste cenário, a seção seguinte detalha o percurso metodológico adotado para investigar empiricamente o regime que dele emerge.

<sup>29</sup> <https://netlab.eco.ufrj.br/post/danos-causados-pela-publicidade-enganosa-na-meta>

<sup>30</sup> <https://lupa.uol.com.br/educacao/2024/12/12/imagens-de-ia-polarizam-e-aproximam-humor-e-odio>

<sup>31</sup> Disponível em:

<https://oglobo.globo.com/politica/noticia/2024/01/24/casos-de-deep-fake-chegam-a-justica-eleitoral-que-cobra-w-hatsapp.ghtml>

<sup>32</sup> Disponível em:

<https://g1.globo.com/educacao/noticia/2024/04/16/youtubers-ensinam-como-lucrar-com-uso-de-inteligencia-artificial-para-criar-ideos-e-livros-infantis-em-minutos.ghtml>

#### 4 A DESINFORMAÇÃO POR IA GENERATIVA NO REGIME DE INFORMAÇÃO CONTEMPORÂNEO

A perspectiva do regime de informação proposta por González de Gómez (2012), Frohmann (1995) e Delaia e Freire (2010) orienta esta seção no sentido de demonstrar como seus elementos podem ser aplicados para, nos resultados parciais da Seção 6, compreender a estrutura e o funcionamento do regime de informação contemporâneo em que a desinformação gerada por IA se insere. Nesta pesquisa, o ultrarrealismo desses conteúdos é apresentado como um fenômeno que, ao se apoiar na IA generativa, distorce a lógica de produção e validação da informação, ocupando lugar nesse regime.

Embora a desinformação seja antiga, sua manifestação contemporânea, a Desinformação Digital em Rede (DDR) como característica do regime de informação contemporâneo é também relevante para esta seção 4 em tempos em que a circulação de falsidades orientada por interesses econômicos se consolida como prática dominante. Essa dinâmica é reforçada pela chamada “cultura da pós-verdade” (Araújo, 2020a, 2020b) e impulsionada por agendas políticas e por um modelo de negócio das plataformas digitais no qual o lucro tende a se sobrepor ao compromisso com a veracidade.

Ainda que o conceito de regime de informação, formulado por Gómez (2012), abranja a existência de conflitos, vontades e efeitos no processo informacional, a novidade da desinformação gerada por IA generativa e a lógica econômica das plataformas digitais exigem um olhar crítico que destaque essa prática. A Desinformação Digital em Rede, nesse sentido, representa um processo sistemático de produção, impulsionamento e legitimação de informação enganosa, deliberadamente orientado para gerar lucro e poder, fenômeno que se intensifica com os recursos da IA generativa.

É interessante considerar o que argumenta Winner (1986) ao demonstrar que as tecnologias podem ser politicamente enviesadas, seja por decisões de projeto, que favorecem certas estruturas de poder. No campo da IA, essa politização torna-se evidente no caso de Timnit Gebru, pesquisadora demitida do Google após denunciar os vieses raciais e sexistas, os impactos ambientais e os riscos de desinformação associados aos grandes modelos de linguagem da empresa (BBC Brasil, 2025)<sup>33</sup>. O episódio escancara a tensão entre compromissos éticos e interesses corporativos, e reforça que a IA generativa não é neutra — ela opera em um terreno marcado por disputas assimétricas que moldam o regime de informação contemporâneo.

---

<sup>33</sup> <https://www.bbc.com/portuguese/geral-55195078>

Diante do exposto, esta seção está estruturada para detalhar os componentes deste regime, começando pela apresentação dos quatro elementos (atores, artefatos, dispositivos e ações), aplicando-os ao fenômeno da desinformação gerada por IA. Para essa tarefa, a abordagem se inspira na lógica consolidada em estudos como o de Delaia e Freire (2010), que demonstraram a eficácia dessa estrutura de quatro partes para mapear e analisar um ambiente informacional complexo. O desenvolvimento desta seção também discutirá as implicações éticas, políticas e informacionais, dialogando com questões levantadas por pesquisadores como Kaufman (2021), Coeckelbergh (2021) e Tarcízio Silva (2022).

#### **4.1 Atores, artefatos, dispositivos e ações**

A análise de um regime de informação, sistematizada por pesquisadores como Delaia e Freire (2010) a partir das bases de González de Gómez datadas desde 1996, desdobra-se em atores, artefatos, dispositivos e ações, e que são detalhados nesta subseção. Embora o modelo aplicado tenha sido inicialmente desenvolvido para o contexto de instituições de pesquisa, sua flexibilidade e amplo potencial analítico o tornam aplicável e relevante para a presente investigação. Os quatro elementos fornecem uma estrutura maleável para compreender a dinâmica do regime de informação sob a desinformação gerada por IA generativa.

Aqui, os elementos oferecem uma estrutura para compreender como o regime de informação contemporâneo é ampliado pela Desinformação Digital em Rede potencializada pelos conteúdos ultrarrealistas. Nesse cenário, os elementos são instrumentalizados para operacionalizar o fenômeno, que se manifesta, por exemplo, nas campanhas coordenadas de manipulação, na produção de imagens sintéticas que reforçam preconceitos e nas brechas das políticas internas e algoritmos de recomendação que, ao priorizar engajamento, ampliam conteúdos polarizadores.

##### **A. Atores**

Os atores são os sujeitos, coletivos ou organizações reconhecidos como participantes dos processos informacionais. No regime de informação contemporâneo, essa categoria vai além da clássica relação emissor–receptor. Com base em Marques e Laipelt (2023) e nas perspectivas de Schneider (2022), é possível identificar:

- desenvolvedores e *big techs*: empresas como OpenAI, Google e Meta, que criam, mantêm e governam as tecnologias de IA generativa, como ChatGPT, Gemini e Meta AI;

- agentes de desinformação: grupos políticos, fraudadores ou outros agentes que produzem e disseminam conteúdos enganosos para fins específicos;
- públicos e usuários: consomem, interagem, replicam e, por vezes, ressignificam os conteúdos desinformativos;
- instituições de contestação: agências de checagem, a imprensa tradicional, o poder público e a sociedade civil organizada, que atuam para identificar e combater a desinformação.

É fundamental reforçar as próprias ferramentas de IA generativa como um ator que contribui ativamente para a “degeneração da informação” (Marques; Laipelt, 2023) ao inundar o ecossistema com conteúdos artificiais, desafiando a distinção entre o real e o sintético. A opacidade de seus modelos (Kaufman, 2021) e os vieses presentes em seus dados de treinamento (UNESCO, 2024) reforçam que sua operação está intrinsecamente ligada a decisões e valores humanos, tornando a responsabilidade por seus impactos um dilema complexo que recai sobre desenvolvedores, plataformas e reguladores (Coeckelbergh, 2019, 2021; Silveira, 2024).

## **B. Artefatos**

Os artefatos podem ser compreendidos como os meios materiais e simbólicos que viabilizam a circulação e o armazenamento da informação, como documentos, bancos de dados e sistemas (Gómez, 2012). Longe de serem suportes neutros, eles são elementos ativos na configuração do regime. Como sustenta Langdon Winner (1986, p. 8):

As coisas que nós chamamos de tecnologias são formas de construir ordem em nosso mundo. Muitos dispositivos ou sistemas técnicos importantes na vida cotidiana contém diversas possibilidades de ordenar a atividade humana. Conscientemente ou inconscientemente, deliberadamente ou inadvertidamente, as sociedades escolhem tecnologias que influenciam, por um longo tempo, como as pessoas vão trabalhar, se comunicar, viajar, consumir, e assim por diante.

Na Desinformação Digital em Rede potencializada pela IA generativa, essa politização é central. Entre os principais artefatos:

- conteúdos sintéticos: os próprios produtos da IA generativa, como vídeos em *deepfake*, imagens ultrarrealistas, textos e áudios automatizados, que funcionam como os veículos da desinformação;
- plataformas digitais: redes como Facebook, TikTok e YouTube, que hospedam e cuja

funcionalidades constituem um artefato que condiciona a circulação e o consumo dos conteúdos;

- infraestrutura algorítmica: os algoritmos de recomendação e os vastos bancos de dados usados para treinar os modelos de IA, que funcionam como a matéria-prima e a maquinaria que dão forma ao que é produzido e visto.

### **C. Dispositivos**

Aqui, os dispositivos são considerados como as normativas, regras e políticas que condicionam a visibilidade e o reconhecimento da informação. González de Gómez (2012) associa o conceito de regime ao de dispositivo, compreendendo-o como uma rede estratégica de elementos discursivos e não discursivos. No regime de informação contemporâneo, entende-se que os dispositivos são tanto formais e explícitos quanto informais e opacos. No contexto da Desinformação Digital em Rede e da IA generativa, pode-se destacar como exemplos:

- dispositivos jurídico-regulatórios: leis e projetos de lei, como o PL 2338/2023<sup>34</sup>, e resoluções de órgãos estatais, como as do Tribunal Superior Eleitoral (TSE), que buscam estabelecer um marco legal para o uso da tecnologia;
- políticas das plataformas: os termos de serviço e as políticas de moderação de conteúdo das *big techs*, que funcionam como dispositivos privados e podem tanto inibir quanto, por omissão, facilitar a desinformação;
- algoritmos como dispositivos: de forma crucial, os algoritmos de moderação e recomendação operam como os dispositivos mais eficazes e, ao mesmo tempo, mais opacos. Suas regras internas, que priorizam o engajamento, determinam o que pode trazer maior retorno, moldando a ecologia informacional de maneira decisiva.

### **D. Ações de informação**

Por fim, as ações de informação são as práticas realizadas pelos atores por meio dos artefatos e condicionadas pelos dispositivos. Essas ações se manifestam como um campo de disputa, um "processo agonístico" (Frohman, 1995) onde diferentes práticas se confrontam, como nos exemplos:

---

<sup>34</sup> O Projeto de Lei n.º 2338/2023 estabelece normas gerais para o uso de inteligência artificial no Brasil. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

- ações que sustentam o regime: envolvem a criação, o compartilhamento viral, a manipulação de conteúdos sintéticos e o direcionamento de campanhas fraudulentas para obter ganhos políticos ou financeiros;
- ações que contestam o regime: incluem a verificação de fatos (*fact-checking*), a denúncia de conteúdos nas plataformas, a mobilização da sociedade civil por regulação e as práticas de letramento midiático e informacional que buscam desenvolver uma consciência crítica nos usuários.

Feita a apresentação dos elementos, o que se busca demonstrar é como cada componente parece operar em interdependência para sustentar ou contestar os fluxos informacionais. Por exemplo, os atores (desde desenvolvedores a agentes maliciosos e públicos) parecem empregar artefatos sintéticos que, pela sua capacidade de simulação, redefinem a percepção da realidade. A dinâmica desses fluxos é orientada por dispositivos tanto formais quanto informais, cujas lógicas podem favorecer a desinformação. Por fim, as ações de informação parecem refletir o conflito presente neste regime, com práticas de propagação.

#### **4.2 Implicações éticas e desafios regulatórios da desinformação gerada por IA generativa**

Entrando no campo das implicações éticas e dos desafios regulatórios da IA generativa, ao menos duas preocupações centrais se destacam no regime de informação contemporâneo:

1. o funcionamento dos modelos de IA generativa: os algoritmos que produzem conteúdos sintéticos com base em dados sem transparência ou explicabilidade;
2. e o papel desempenhado pelas plataformas digitais: elas permitem a publicação, recomendam, patrocinam ou monetizam esses conteúdos, contribuindo diretamente para sua circulação em larga escala.

Enquanto o primeiro desafio exige atenção aos vieses, à curadoria de dados e à arquitetura técnica dos sistemas, o segundo envolve questões sobre moderação de conteúdo, regulação de anúncios, modelos de negócio e responsabilidades das plataformas digitais no ecossistema informacional. Nesse sentido, esses ambientes exploram a desinformação gerada por IA e a inserem suas linhas do tempo de recomendação, publicidade e ranqueamento.

Como lembrado por Schneider (2022), a regulação pública é indispensável para estabelecer parâmetros mínimos e limitar os privilégios privados que hoje definem grande

parte do ecossistema digital. Segundo o autor, a ausência de freios legais e tecnológicos permitiu que o capital expandisse seus objetivos de reprodução de forma avassaladora, “atropelando tudo o que não atenda, da maneira mais eficaz e eficiente possível, os seus objetivos de reprodução ampliada”. O autor também denuncia a distorção da noção de “liberdade de imprensa” quando apropriada por conglomerados midiáticos para se proteger de regulações democráticas, um artifício retórico que, na prática, transforma liberdade de imprensa em “liberdade de empresa” (Schneider, 2022, p. 118).

No caminho do que Westerlund (2019), Spitale *et al.* (2023) e Marques e Laipelt (2023) afirmam, é possível entender que a produção de conteúdos sintéticos ultrarrealistas representa uma ameaça ao debate público, à confiança nas instituições e ao próprio direito à informação. Nesse cenário, surgem questões éticas cruciais como a explicabilidade e a transparência dos sistemas de IA. Essa opacidade dos modelos generativos impede que seus processos decisórios sejam compreendidos por usuários ou auditáveis por especialistas, como afirma Kaufman (2021). Essa característica técnica, longe de ser um detalhe, compromete a possibilidade de controle social dos sistemas, pois seus resultados escapam à lógica humana de raciocínio e interpretação. Conforme a autora:

A opacidade decorre do desconhecimento de como os chamados “dados de entrada” (inputs) geraram o dado de saída (output), como o sistema correlacionou as variáveis contidas nos dados de entrada e os pesos atribuídos [...]. Com maiores recursos computacionais e *big data*, o número de fatores possíveis a serem incluídos nesses sistemas ultrapassa o nível de compreensão de um ser humano racional.

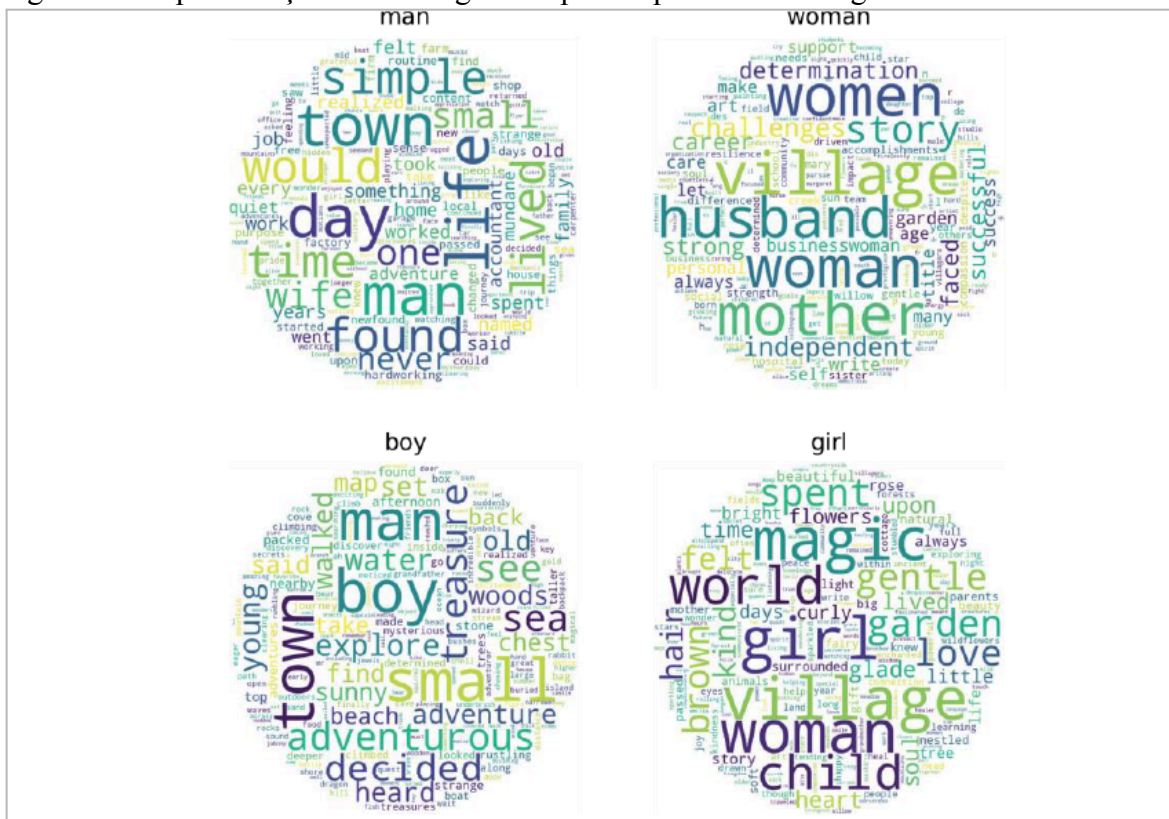
A opacidade da inteligência artificial relatado por Kaufman (2021), portanto, é reforçado por Silveira (2024) como um problema político, mesmo modelos com código aberto podem se tornar inacessíveis à compreensão pública quando treinados em bases privadas e moldados por interesses comerciais. Nesse cenário, a curadoria dos dados, seus usos e impactos permanecem invisíveis e os riscos se multiplicam.

Um estudo da UNESCO (2024)<sup>35</sup> revelou que os sistemas de IA generativa tendem a reforçar papéis de gênero, como investigado em 1.000 histórias geradas pelo modelo LLaMA 2, da Meta AI, para cada homem (man), mulher (woman), como ilustrado na Figura 6.

---

<sup>35</sup> Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000388971>

Figura 6 – Representações textuais geradas por IA para diferentes gêneros



Fonte: UNESCO (2024, p. 14).

O resultado desse estudo mostra um viés nos dados de treinamento de ferramentas de IA, nos quais os personagens femininos foram associados com mais frequência a termos afetivos e relacionais, como marido (“*husband*”), mãe (“*mother*”), amor (“*love*”) e criança/filho (“*child*”). Já os personagens masculinos aparecem ligados a ambientes externos e ação, como cidade (“*town*”), dia (“*day*”) e aventura (“*adventure*”). No contexto da desinformação gerada por IA, esse tipo de viés pode não ser o conteúdo central de uma *fake news*, mas mostra um *modus operandi* que naturaliza certas imagens sociais e visões de mundo.

Como destaca Coeckelbergh (2021), torna-se quase impossível atribuir responsabilidade por danos causados por esses sistemas, já que não se pode explicar como uma decisão ou resultado foi alcançado. Assim, a opacidade não é um detalhe, mas uma característica central que impede a compreensão e a responsabilização efetiva sobre os processos decisórios da IA.

A Recomendação sobre a Ética da Inteligência Artificial da UNESCO (2022) aponta que sistemas de IA devem estar subordinados aos princípios de transparência, justiça e responsabilidade, no entanto, a grande parte das normas sobre IA tem sido elaborada a partir de contextos do norte global, como Estados Unidos e União Europeia, o que pode dificultar a

adaptação dessas diretrizes a realidades como a brasileira. No país, as plataformas costumam atuar com baixa responsabilização local, o que, segundo pesquisadores do direito internacional, como Jiang e Belli (2024), aprofunda a dependência do país e o risco de se tornar uma 'colônia digital', como demonstram os embates envolvendo o PL das *Fake News*.

Neste ponto, considera-se que a atuação das plataformas digitais molda o ambiente onde a desinformação sintética adquire eficácia e escala, uma atuação estratégica na Desinformação Digital em Rede gerada por IA, onde esses conteúdos são impulsionados por mecanismos de recomendação e por modelos de negócio que privilegiam o engajamento.

O potencial de manipulação de materiais falsos evidenciam os riscos associados à desinformação e também reforça a necessidade de uma abordagem colaborativa entre academia, governo e grandes empresas de tecnologia no campo regulatório. Um exemplo dessa proatividade no combate à desinformação é o trabalho do Netlab, laboratório de pesquisa da Universidade Federal do Rio de Janeiro (UFRJ), que expôs a negligência na moderação de anúncios na plataforma Meta, os quais imitavam programas sociais e circulavam pelas redes digitais em 2023 (Coutinho, 2024).

A resistência das grandes empresas em assumir responsabilidades, como ocorreu no caso do Netlab e da Meta, evidencia a necessidade de proteger os direitos individuais acima dos interesses lucrativos das corporações. Como destacado por Brizola e Bezerra (2018), as plataformas que reproduzem esses materiais não se preocupam com a credibilidade a longo prazo, mas sim com lucros imediatos.

No Brasil, a tentativa de estabelecer um marco regulatório para a IA está sendo materializada no Projeto de Lei nº 2338/2023<sup>36</sup> que propõe princípios, direitos e deveres para o desenvolvimento e uso de sistemas de IA, incluindo os tipos de práticas enganosas. Ainda que represente um avanço importante no debate nacional, o projeto não é específico sobre a responsabilização por conteúdos enganosos gerados por IA, por exemplo, quando tais conteúdos são impulsionados e monetizados por empresas internacionais com baixa responsabilização local. O anteprojeto da Lei já reconhecia o potencial de dano informacional e a dificuldade de atribuição de responsabilidade em casos de geração automatizada de conteúdos, e embora tenha avançado ao propor princípios como transparência, explicabilidade e responsabilidade civil, seu alcance normativo ainda é insuficiente diante da complexidade estrutural que sustenta o atual regime de informação.

---

<sup>36</sup> Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

## 5 METODOLOGIA

Pesquisar a desinformação gerada por conteúdos ultrarrealistas produzidos por inteligência artificial (IA) generativa não é relevante apenas porque ferramentas como ChatGPT, Gemini ou Veo se popularizaram nos últimos anos. O que se evidencia, além dos problemas abordados nesta pesquisa, é uma mudança nos modos de produção, circulação e validação da informação, impulsionada por tecnologias emergentes que hoje também produzem conteúdos próprios com alto potencial de persuasão. No Brasil, essa dinâmica ganha força diante da rápida adoção dessas tecnologias, com 54% da população declarando ter utilizado IA generativa em 2024, superando a média global de 48% (Google; IPSOS, 2024).

Como outras pesquisadoras que refletem sobre as Tecnologias Digitais de Informação e Comunicação (TDICs) nos ambientes informacionais, como Beluzzo (2018), que discute competência em informação, e Oliveira (2020), que explora sua dimensão crítica, além de autoras preocupadas com o avanço massivo da IA generativa, (Santaella;Kaufman, 2024) e da operação das plataformas digitais (O' Neil, 2021; Zuboff, 2019), a pesquisadora deste trabalho entende que investigar a desinformação no campo da Ciência da Informação é também se posicionar diante de fenômenos que estão à margem da verdade factual. Por meio de ferramentas robustas como a IA generativa, esse tipo de desinformação explora o hiper-realismo e uma linguagem capaz de imitar com precisão o que antes era exclusividade da experiência humana, o que amplia o potencial de manipulação das narrativas, atinge a integridade de pessoas públicas e comuns, e compromete processos democráticos ao distorcer o debate público e desinformar deliberadamente a população.

Por isso, a metodologia desta pesquisa também se apropria da vertente crítica da Ciência da Informação ao compreender que os os processos informacionais não operam de forma espontânea ou puramente técnica, eles se organizam em torno de interesses, disputas de autoridade (Gomez, 2012) e, mais evidentemente, moldados por interesses, valores e práticas moldadas por visões de mundo específicas, e não espaços neutros de mediação (Silva, 2022).

A noção de regime de informação, assumida nesta pesquisa como abordagem teórico-analítica permite trazer a análise para além dos conteúdos desinformativos e observar como determinados materiais sintéticos, termo utilizado na literatura acadêmica (Spitale *et al*, 2023; UNESCO, 2024) para definir conteúdos gerados por IA, ganham autoridade e como práticas de validação operam nos bastidores daquilo que aparece como “anúncio confiável”, “notícia”, “vídeo”, “declaração” ou “prova” de forma ultrarrealista.

Esta seção de metodologia, portanto, apresenta os procedimentos metodológicos adotados para alcançar o objetivo central da pesquisa: analisar como se configura o regime de informação da IA generativa no Brasil. Para isso, descrevem-se nos subtópicos a abordagem metodológica adotada, a delimitação empírica da pesquisa e a operacionalização das técnicas de levantamento e a análise dos dados.

### **5.1 Caracterização da pesquisa**

A construção do conhecimento científico, no sentido trazido por Prodanov e Freitas (2013), exige a definição de parâmetros metodológicos que orientam o processo investigativo. No caso da abordagem adotada nesta pesquisa, trata-se de uma investigação de natureza qualitativa por buscar a interpretação das dinâmicas que envolvem a desinformação gerada por IA generativa, mais do que a sua quantificação de ocorrências.

A partir dos objetivos colocados nesta pesquisa, essa abordagem qualitativa se alinha a um caráter descritivo que, conforme Triviños (1987), busca conhecer e detalhar aspectos de uma determinada realidade focando nos significados que os fenômenos têm para as pessoas. Ou seja, aqui, há a intenção de descrever os atores, artefatos, dispositivos e ações que estruturam o regime de informação sob a desinformação gerada por IA generativa no Brasil a partir da análise de conteúdo. Esse caminho metodológico se concentra em mostrar os fenômenos sociais como um passo inicial para a compreensão total do seu dinamismo. Para tanto, os resultados são expressos, como propõe Triviños (1987, p. 128), em “retratos (ou descrições), em narrativas, ilustradas com declarações das pessoas para dar o fundamento concreto necessário”. Também segundo o autor:

Como as descrições dos fenômenos estão impregnadas dos significados que o ambiente lhes outorga, e como aquelas são produto de uma visão subjetiva, rejeita toda expressão quantitativa, numérica, toda medida. Desta maneira, a interpretação dos resultados surge como a totalidade de uma especulação que tem como base a percepção de um fenômeno num contexto.

Quanto aos seus procedimentos técnicos, este estudo usou, preliminarmente, a pesquisa bibliográfica e, como segundo passo, a pesquisa documental. A primeira etapa foi iniciada a partir de um levantamento bibliográfico fundamentado na necessidade de compreender como o regime de informação e a desinformação gerada por IA generativa tem sido tratada na literatura científica. Esse percurso contribuiu para a construção do referencial teórico (apresentado na seção 2) e também para o desenvolvimento deste documento, apresentado nas seções 3 e 4. Como explicam Laville e Dionne (1999, p. 112):

Uma vez escolhido o problema, o pesquisador parte à procura de informações que o ajudem a esclarecê-lo, que alimentem sua reflexão e lhe permitam elaborar uma

resposta. [...] Esse levantamento permite situar o problema e integrá-lo num sistema coerente de conhecimentos, a fim de que se possa extrair dele um sentido.

Com base nessa perspectiva, o levantamento bibliográfico teve como propósito orientar a delimitação do problema de pesquisa e apoiar as escolhas teóricas e metodológicas do estudo. Para tanto, a busca foi iniciada na Brapci (Base de Dados em Ciência da Informação)<sup>37</sup>, usando termos como “desinformação” em associação com “IA generativa” (com retorno de três resultados), bem como a combinação entre “regime de informação” e “IA generativa”. A seleção desses termos levou em conta a centralidade que eles ocupam no escopo da pesquisa, ainda que tenha revelado a escassez de estudos que os articulem de forma integrada.

Ampliou-se a busca para o Google Acadêmico<sup>38</sup>, onde foi encontrado, inicialmente, o artigo de Spitale *et al* (2023), da Universidade de Zurique, e seus resultados sobre como a IA consegue desinformar melhor que humanos. Com isso, a partir de referências clássicas e contemporâneas, tanto do campo da Ciência da Informação quanto da Comunicação e da Computação, foi possível analisar as contribuições fundamentais para compreender as relações entre desinformação, plataformas digitais e sistemas de inteligência artificial.

Esse percurso teórico permitiu aproximar as contribuições de autores como Frohmann (1995), González de Gómez (2012) e Delaia e Freire (2010), que fundamentam o conceito de regime de informação. A partir disso, foi possível articular tais perspectivas com discussões recentes sobre a negligência das *big techs* na moderação de anúncios com *deepfakes* no Brasil<sup>39</sup>, além de temas centrais ao debate contemporâneo como pós-verdade, desinformação e ética da inteligência artificial, com base nas contribuições de Carlos Araújo (2020a, 2020b), Coeckelbergh (2019, 2021) e Dora Kaufman e Lucia Santaella (2024).

Importante destacar que as reflexões dessa etapa também dialogam com o conceito de Desinformação Digital em Rede formulado por Schneider (2022). O arcabouço do autor compreende e interliga essas reflexões sobre regime de informação, negligência das *big techs* com *deepfakes* e anúncios, pós-verdade, desinformação e ética da IA possibilitando explicar suas características no cenário contemporâneo.

Esse enquadramento conceitual orienta tanto a seleção das fontes documentais quanto a análise dos casos, permitindo identificar, além dos conteúdos sintéticos produzidos por IA

---

<sup>37</sup> <https://brapci.inf.br>

<sup>38</sup> <https://scholar.google.com.br>

<sup>39</sup> Casos disponíveis em relatório do NetLab sobre anúncios fraudulentos: <https://netlab.eco.ufrj.br/post/danos-causados-pela-publicidade-enganosa-na-meta>

generativa, as formas de circulação e interação entre usuários, plataformas e sistemas automatizados que potencializam sua disseminação no regime de informação contemporâneo.

Em relação à pesquisa documental e diante do desafio de encontrar casos consistentes sobre desinformação gerada por IA, um fenômeno que, por sua natureza, é difuso e de difícil rastreabilidade, a análise de casos em plataformas digitais foi considerada inviável, pois exigiria da pesquisadora a tarefa de a comprovação técnica do uso de IA em cada caso, o que extrapolaria os recursos do estudo. Assim, a análise precisava de uma fonte documental que atendesse a três critérios essenciais: (1) oferecer casos já verificados como falsos; (2) apresentar indícios ou a confirmação do uso de IA na sua produção; e (3) ser uma fonte acessível publicamente. Com base nesses requisitos, uma agência de checagem de fatos se apresentou como o tipo de fonte documental mais adequada.

Em relação aos objetivos da pesquisa e a conexão com o *corpus* empírico, a investigação documental foi organizada em duas frentes principais:

1. casos verificados por uma agência de checagem;
2. documentos institucionais de plataformas digitais e do Projeto de Lei nº 2.338/2023.

Sendo assim, para o cumprimento do objetivo geral (analisar como se configura o regime de informação da desinformação gerada por IA no Brasil), o estudo considerou três conjuntos documentais: (1) os casos verificados por agência de *fact-checking*; (2) as políticas institucionais de plataformas digitais como Meta, Google, TikTok e X; e (3) o Projeto de Lei nº 2.338/2023. Esses documentos são tratados como fontes complementares que, em conjunto, permitem observar como diferentes atores, artefatos e dispositivos se articulam na configuração do regime de informação contemporâneo.

Em relação à correspondência com os objetivos específicos, é interessante retomá-los com o detalhamento no quadro a seguir.

Tabela 3 - Relação entre objetivos da pesquisa e *corpus* documental

Objetivo específico	Pesquisa documental	Direcionamento analítico
1) Mapear os componentes do regime de informação relacionados à circulação de desinformação gerada por IA generativa nos casos analisados por agência de <i>fact-checking</i> ;	Casos verificados por agência de <i>fact-checking</i> .	Identificação de padrões de circulação e configuração dos elementos do regime de informação.

2) Identificar os atores envolvidos na produção, circulação e contestação desses casos, situando seu papel no regime de informação contemporâneo.	Casos verificados por agência de <i>fact-checking</i> ; Políticas e regras das plataformas digitais; Projeto de Lei 2338/2023.	Analisar como usuários, plataformas e verificadores se posicionam e disputam autoridade informacional.
3) Levantar os artefatos e dispositivos presentes nos casos, observando como estruturam práticas de desinformação e de verificação.	Casos verificados por agência de <i>fact-checking</i> (artefatos: <i>deepfakes</i> , imagens, textos automatizados); Políticas e regras das plataformas digitais; Projeto de Lei 2338/2023.	Examinar como dispositivos técnicos e normativos condicionam a circulação e a contenção da desinformação.

Fonte: elaborado pela autora (2025)

A vantagem dessa escolha está na natureza pública e acessível de registros já sistematizados sobre o fenômeno investigado. Como explicam Prodanov e Freitas (2013, p. 56), a pesquisa documental “baseia-se em materiais que não receberam ainda um tratamento analítico ou que podem ser reelaborados de acordo com os objetivos da pesquisa”.

Quanto ao tratamento de dados, a pesquisa aplicou a técnica de análise de conteúdo. Conforme define Bardin (1977, p. 42):

O analista possui à sua disposição (ou cria) todo um jogo de operações analíticas, mais ou menos adaptadas à natureza do material e à questão que procura resolver. Pode utilizar uma ou várias em complementaridade, de modo a enriquecer os resultados ou aumentar sua validade, aspirando assim uma interpretação fundamentada.

Isso permitiu à esta analista, pois, a flexibilidade e o rigor necessários para construir uma interpretação fundamentada, permitindo observar o fenômeno para além do conteúdo, alcançando as estruturas informacionais e relacionais que o configuram.

A técnica de análise de conteúdo foi orientada pelas quatro categorias centrais da abordagem do regime de informação (atores, ações, dispositivos e artefatos) sistematizadas por Delaia e Freire (2010), autora que consolidou essa estrutura a partir de uma leitura articulada dos fundamentos teóricos propostos por Frohmann (1995) e obras de González de Gómez, autores que também se alinham ao referencial teórico desta pesquisa. Como também reconhecem Bezerra *et al.* (2016, p. 3):

Resguardando suas divergências e semelhanças, Frohmann (1995), González de Gómez (2002, 2012b), Braman (2004) e Ekbia e Evans (2009) focam na relação regime-informação, com o intuito de elaborar o constructo na tentativa de compreender como processos, serviços e produtos de informação se formam, se ordenam, se desordenam e se estabilizam nos espaços de informação (públicos ou privados; locais ou globais; técnicos ou sociais), em uma sociedade que é, a um só tempo, guiada pelas tecnologias da informação e da comunicação e inserida em uma forte economia de mercado.

Os passos no caminho metodológico deste documento, são dados desta forma: (1) pesquisa qualitativa; (2) natureza descritiva; (3) pesquisa bibliográfica e documental; e (4) análise de conteúdo.

É importante ressaltar que os resultados do levantamento bibliográfico realizado para esta pesquisa já foram apresentados e desenvolvidos ao longo das seções 2, 3 e 4 deste trabalho. Os demais procedimentos metodológicos adotados serão detalhados nas próximas nos subtópicos a seguir.

### 5.1.1 Documentos analisados

Neste ponto da metodologia, torna-se essencial explicitar o *corpus* empírico que sustenta a investigação que permitiu compreender a configuração do regime de informação da desinformação gerada por inteligência artificial no Brasil. Apoiada em três conjuntos documentais, esses materiais foram examinados à luz da análise de conteúdo (Bardin, 2011).

1. **Casos verificados** – foram reunidos conteúdos desinformativos identificados pela Agência Lupa, na qual a escolha é detalhada no subtópico a seguir, nos quais se observa o uso da categoria de inteligência artificial generativa na criação ou manipulação de informações. Esses registros permitem identificar padrões de circulação e de apropriação social de conteúdos sintéticos.
2. **Políticas de plataformas** – foram examinados documentos públicos das principais plataformas digitais (Meta, Google, TikTok), referentes às suas diretrizes sobre conteúdos gerados por inteligência artificial e desinformação. Tais políticas podem ser visualizadas no apêndice B e na tabela 4 do subtópico a seguir. São tratadas como documentos das empresas que permitem compreender como as corporações moldam, em nível normativo, as condições de circulação informacional.
3. **Projeto de Lei nº 2.338/2023** – foi incorporado à análise o debate brasileiro sobre a regulação da inteligência artificial, que ganhou visibilidade em 2023 com a apresentação do Projeto de Lei nº 2338 no Senado Federal. O PL procura incluir a defesa de direitos fundamentais, a transparência e a responsabilização. Para avançar nesse processo, foi criada uma Comissão Temporária Interna em 2023 encarregada de reunir diferentes iniciativas legislativas em tramitação (SENADO FEDERAL, 2023). O texto resultante foi aprovado pelos senadores em dezembro do mesmo ano e encaminhado à Câmara dos Deputados, onde permanece em análise na etapa de elaboração desta dissertação. Ainda em discussão, o PL 2338/2023 serve de base para

as reflexões sobre riscos e limites do uso da inteligência artificial, permitindo articular a dimensão regulatória nos dispositivos do regime de informação.

Esses três conjuntos documentais constituem fontes complementares que, articuladas, permitem observar de forma integrada a configuração do regime de informação da desinformação gerada por IA no Brasil. A forma como esses documentos serão operacionalizados e relacionados aos objetivos da pesquisa é apresentada na subseção seguinte.

## **5.2 Delimitação empírica**

O recorte desta pesquisa foi ancorado em dinâmicas que se desenrolam no ciberespaço entendido como parte da infosfera, um ambiente no qual sujeitos, tecnologias e fluxos de informação se entrelaçam de forma contínua (Floridi, 2014). Como destaca Ronaldo Araújo (2014):

(...) podemos considerar a internet uma importante arena conversacional, na qual inclusive, de acordo com Maia (2002), o espaço se desdobra e as novas conversações e discussões políticas podem seguir seu curso. A direção seguida garante voz aos sujeitos e amplia os canais de deliberação na expectativa que as decisões políticas sejam pautadas não somente na capacidade do governante, mas nas consultas à opinião pública da sociedade civil engajada via internet.

Essa concepção embasa a decisão metodológica de tratar os conteúdos verificados digitalmente como material empírico relevante. Porém, um dos primeiros desafios metodológicos desta pesquisa foi definir um ponto de partida consistente para a análise da desinformação gerada por IA generativa. Isso porque esse fenômeno é observado, sobretudo, em compartilhamentos nas plataformas digitais como redes digitais, mensageiros e sites de vídeo e nem sempre é possível rastrear com precisão suas origens, especialmente no que se refere à comprovação do uso de IA na geração da desinformação, dada a natureza ultrarrealista dos conteúdos sintéticos.

Como exposto no subtópico anterior, a seleção das agências de checagem como fonte documental levou em conta a necessidade de localizar casos já identificados como desinformação gerada por conteúdos criados com IA generativa. Entre as iniciativas em atuação no Brasil, optou-se pela Agência Lupa como fonte primária por reunir uma sistematização acessível e volume expressivo de checagens sobre o tema. Fundada em 2015, a Lupa foi a primeira agência brasileira dedicada exclusivamente à verificação jornalística e é

reconhecida pelo *International Fact-Checking Network (IFCN)*<sup>40</sup>, o que implica o cumprimento de critérios rigorosos (Janiaki; Santos, 2024).

Reconhecida por sua atuação pioneira no cenário brasileiro de checagem de fatos, a Lupa organiza suas verificações de forma rigorosa e sob categorias específicas como “Falso” a facilidade no uso de palavras-chave. Isso facilitou o levantamento de casos pertinentes ao estudo. Além disso, dada a recorrência e o volume de análises sobre o tema, observou-se que a agência possui um número significativo e crescente de checagens dedicadas a conteúdos sintéticos com detalhamento do conteúdo em relação ao artefato (*deepfake*, imagem etc.), as plataformas de circulação e os atores envolvidos.

Além dos casos verificados, a pesquisa documental também se apoiou em uma segunda parte com a consulta a documentos institucionais produzidos por plataformas digitais e o governo brasileiro, como políticas de uso, termos de serviço, projeto de lei e diretrizes voltadas à moderação de conteúdos sintéticos. Foram consideradas como fontes as páginas institucionais online sobre o Projeto de Lei (PL) 2338/2023 e as políticas internas em relação a conteúdo sintético das plataformas digitais Google, Meta, TikTok e X. A escolha dessas quatro *big techs* se justifica pelo fato de estarem entre as maiores do mundo em número de usuários ativos em seus produtos digitais, de acordo com o relatório de 2024 da DataReportal<sup>41</sup>. A análise das diretrizes institucionais foi de natureza interpretativa e crítica, buscando observar como esses documentos se posicionam (ou se omitem) em relação às manipulações sintéticas mapeadas e em que medida suas formulações contribuem ou não para a contenção da desinformação gerada por IA.

Reunidos, os dois conjuntos documentais, verificações públicas e normas institucionais, compõem o *corpus* empírico que sustenta a análise do regime de informação da desinformação gerada por IA no contexto brasileiro.

Por fim, no quadro abaixo, fica desenhada a operacionalização do levantamento documental a partir das fontes alinhadas aos objetivos desta pesquisa. Para tanto, relembra-se os objetivos específicos descritos na seção de introdução, sendo: 1) mapear os componentes do regime de informação relacionados à circulação de desinformação gerada por IA generativa nos casos analisados por agência de *fact-checking*; 2) identificar os atores envolvidos na produção, circulação e contestação desses casos, situando seu papel no regime

---

<sup>40</sup> A *International Fact-Checking Network (IFCN)* é uma rede global de checagem de fatos criada em 2015 pelo Instituto Poynter, escola de jornalismo e pesquisa sediada nos Estados Unidos. Seu objetivo é promover a troca de experiências, a definição de diretrizes comuns e a disseminação de boas práticas entre organizações de verificação de notícias em diferentes países.

<sup>41</sup> DataReportal é uma das principais fontes de dados sobre presença digital no mundo. Relatório disponível em: <https://datareportal.com/reports/digital-2024-global-overview-report>

de informação contemporâneo; 3) levantar os artefatos e dispositivos presentes nos casos, observando como estruturam práticas de desinformação e de verificação.

Tabela 4 - Operacionalização do levantamento documental

Fonte	Documento	Endereço eletrônico	Objetivos abrangidos
Agência Lupa	Base de checagens sobre IA	<a href="http://lupa.uol.com.br/jornalismo/categoria/inteligencia-artificial">lupa.uol.com.br/jornalismo/categoria/inteligencia-artificial</a>	1, 2 e 3
Meta	Política de transparência em anúncios com IA	<a href="https://about.fb.com/br/news/2024/04/nossa-abordagem-para-rotular-conteudos-criados-com-ia-e-midias-manipuladas/">https://about.fb.com/br/news/2024/04/nossa-abordagem-para-rotular-conteudos-criados-com-ia-e-midias-manipuladas/</a>	2 e 3
Google	Política sobre conteúdo gerado por IA	<a href="https://support.google.com/youtube/answer/14328491">support.google.com/youtube/answer/14328491</a> <a href="https://support.google.com/adspolicy/answer/14720423">support.google.com/adspolicy/answer/14720423</a> <a href="https://support.google.com/adspolicy/answer/15142358">support.google.com/adspolicy/answer/15142358</a>	2 e 3
TikTok	Política sobre conteúdo gerado por IA	<a href="https://support.tiktok.com/pt_BR/using-tiktok/creating-videos/ai-generated-content">https://support.tiktok.com/pt_BR/using-tiktok/creating-videos/ai-generated-content</a>	2 e 3
X (Twitter)	Política de autenticidade em conteúdos sintéticos	<a href="https://help.x.com/pt/rules-and-policies/authenticity">https://help.x.com/pt/rules-and-policies/authenticity</a>	2 e 3
PL nº 2338/2023	Regulação da IA no Brasil (Senado)	<a href="http://senado.leg.br/materia/157233">senado.leg.br/materia/157233</a>	2 e 3

Fonte: elaborado pela autora (2025)

Essa organização assegura à pesquisa que as etapas do levantamento documental estejam fundamentadas em fontes adequadas à complexidade do fenômeno investigado, respeitando os objetivos e o problema central da pesquisa e que serão fundamentados a seguir.

### 5.3 Operacionalização das técnicas aplicadas

Para facilitar a visualização do percurso metodológico adotado nesta pesquisa, o quadro a seguir apresenta uma síntese das técnicas aplicadas, distribuídas por etapas, ferramentas utilizadas e respectivas justificativas.

Tabela 5 - Dimensão técnica da pesquisa

Etapa	Técnica	Ferramenta	Justificativa
1	Levantamento documental	Busca no site da Lupa com filtros específicos e em páginas nas plataformas digitais e do governo	Identificar conteúdos verificados com uso de IA generativa e regras em relação a conteúdos sintéticos.
2	Análise de conteúdo (Triagem e seleção)	Leitura completa das verificações e aplicação de critérios de inclusão	Delimitar o material a ser analisado ( <u>Apêndice A</u> )
3	Análise de conteúdo (Sistematização dos dados)	Planilha com campos estruturados	Organizar os dados para análise
4	Análise de conteúdo (Pré-análise)	Leitura e reconhecimento preliminar dos temas	Familiarização e preparação para a codificação
5	Análise de conteúdo (Codificação)	Marcação nos documentos	Identificar categorias relevantes de análise
6	Análise de conteúdo e aplicação do regime de informação como estrutura analítica (Categorização e análise)	Aplicação das dimensões do regime de informação	Interpretar padrões, relações e práticas informacionais

Fonte: elaborado pela autora (2025).

A partir desse planejamento, a primeira etapa consistiu na realização de um levantamento documental sistemático, com recorte empírico focado nos conteúdos verificados pela Agência Lupa. Os critérios definidos para essa busca foram:

- uso da palavra-chave “inteligência artificial”;
- conteúdos categorizados com etiqueta “falso”;
- período de março de 2024 a março de 2025.

A escolha do intervalo temporal de março de 2024 a março de 2025 se justifica pela intensificação, no Brasil, da circulação de conteúdos desinformativos gerados por IA generativa, como os anúncios fraudulentos associados a programas sociais<sup>42</sup>, que utilizaram imagens manipuladas do presidente da República e do ministro da Fazenda (Furtado, 2024), amplamente divulgados em 2024, bem como a estreia de ações regulatórias nas eleições 2024<sup>43</sup> (NetLab, 2024). O marco final, em março de 2025, coincide com a adoção mais visível

<sup>42</sup> Estudo aponta 643 anúncios falsos sobre o Desenrola Brasil  
<https://extra.globo.com/economia/noticia/2024/10/facebook-estudo-aponta-643-anuncios-falsos-sobre-o-desenrola-brasil-e-o-voa-brasil.ghtml>

<sup>43</sup> Artigo do NetLab sobre o desafio do combate aos deepfakes nas eleições de 2024 no Brasil  
<https://netlab.eco.ufrj.br/post/licoes-da-pre-campanha-o-desafio-do-combate-aos-deepfakes-nas-eleicoes-de-2024-no-brasil>

de medidas de transparência por parte das plataformas, como as atualizações de políticas da Meta<sup>44</sup> e o avanço de medidas regulatórias, como a aprovação de lei que amplia a pena de crimes envolvendo o uso de *deepfakes* (Barreto, 2025)<sup>45</sup>.

A pesquisa limita sua análise empírica aos conteúdos previamente verificados pela Agência Lupa, não sendo objetivo deste trabalho realizar verificação independente de casos ambíguos ou não reconhecidos como desinformação. No entanto, menções a conteúdos desse tipo eventualmente localizadas nos documentos institucionais foram registradas como indícios analíticos complementares, contribuindo para observar as estratégias de visibilidade ou legitimação adotadas pelas plataformas.

A segunda parte do levantamento documental contempla os conteúdos institucionais e normativos relacionados ao uso da IA generativa em publicações nas plataformas digitais (Google, Meta, TikTok e X) e no Projeto de Lei 2338/2023. De acordo com Prodanov e Freitas (2013) e Severino (2016), esse tipo de pesquisa se apoia na análise de documentos já existentes, públicos ou privados, que registram práticas institucionais, decisões normativas e formas de regulação. O foco foi identificar como se estruturam os dispositivos que afetam a visibilidade e a circulação da desinformação.

Neste momento, foi realizada uma leitura analítica inicial para localizar trechos que mencionaram diretamente o uso de IA, conteúdos gerados por IA, *deepfakes*, manipulações visuais ou regras sobre rotulagem e transparência em conteúdos sintéticos. Foram priorizadas as versões mais recentes desses documentos e os trechos que continham relação com processos de identificação, moderação, exclusão ou rotulagem de conteúdos.

Após as coletas, o levantamento documental passou para a análise de conteúdo por um processo de triagem e leitura integral, com base em critérios de inclusão previamente definidos. Nesse ponto, segundo Laville e Dionne (1999, p. 232), uma pesquisa qualitativa exige a reconstrução dos sentidos a partir dos materiais estudados:

Se você opta, de preferência, por uma análise de conteúdo, deverá fixar suas escolhas de categorias e recortar os conteúdos coletados, depois proceder à categorização do material reunido. Virá então o momento da reconstrução do sentido dos discursos estudados, seja por uma abordagem quantitativa ou por uma abordagem qualitativa, a escolha dependendo aqui da natureza do discurso e das intenções da pesquisa.

Em relação à busca na base da Lupa, os filtros definidos no processo retornaram 46 conteúdos dentro do período analisado. Foi adotado um recorte adicional com base em

---

<sup>44</sup> Disponível em:

<https://about.fb.com/br/news/2024/04/nossa-abordagem-para-rotular-conteudos-criados-com-ia-e-midias-manipuladas/>

<sup>45</sup> Disponível em:

<https://g1.globo.com/politica/noticia/2025/04/24/lula-sanciona-projeto-que-aumenta-a-pena-para-crimes-contramulheres-com-uso-de-ia.ghtml>

duplicidade de *links* do mesmo caso na página, na relevância pública da figura envolvida e na complexidade técnica do conteúdo sintético.

Os processos de pré-análise, codificação e categorização descritos a seguir aplicam-se ao *corpus* empírico composto pelos conteúdos verificados pela Agência Lupa. A análise dos documentos institucionais, por sua vez, seguiu abordagem interpretativa complementar, com foco na presença (ou ausência) de diretrizes sobre o uso de IA generativa. Dessa forma, foi possível construir uma leitura crítica sobre o papel desses dispositivos na estruturação do regime de informação em torno da IA generativa e da desinformação no Brasil.

Na fase de pré-análise, foi feita uma leitura do material sistematizado, com o objetivo de reconhecer padrões iniciais, recorrências temáticas e elementos que ajudaram a identificar caminhos para a codificação.

Em seguida, foi realizada a etapa de codificação, entendida como a identificação e marcação de trechos relevantes aos objetivos da pesquisa. Os conteúdos foram registrados em uma planilha de sistematização dos dados visualizada no [apêndice A](#), que funcionou como instrumento técnico para a organização e posterior análise. Essa planilha incluiu os campos:

- *Link* / Fonte;
- Título da checagem;
- Figura pública envolvida;
- Formato do conteúdo sintético
- Plataforma de circulação;
- Resumo do conteúdo verificado;
- Observações.

Por fim, na etapa de categorização e análise, os dados foram classificados e interpretados com base nos quatro elementos do regime: atores, ações, dispositivos e artefatos, conforme sistematizado por Delaia e Freire (2010).

Ao lidar com redes e artefatos digitais, é necessário atentar, como observa Ronaldo Araújo (2014), para a importância de “estabelecer critérios de identificação, classificação e leitura que permitam observar os vínculos entre os objetos de informação, seus produtores, mediadores e usos sociais.” Essa perspectiva, ao organizar os materiais coletados, permite aplicar a abordagem metodológica fundamentada no regime de informação, respeitando a complexidade do fenômeno investigado.

## 5.4 Sobre as análises e discussões dos resultados

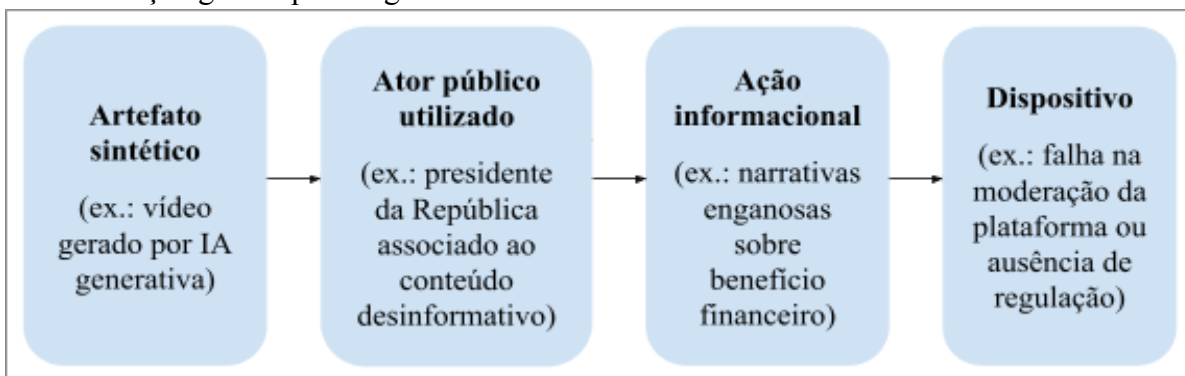
A análise teve início com uma contextualização da temática à luz dos dados coletados (Apêndice A), destacando aspectos como os tipos de manipulação identificados, as figuras públicas envolvidas, as plataformas recorrentes e as tecnologias mencionadas nos conteúdos. Essa leitura inicial objetivou retomar os seguintes elementos centrais do fenômeno mapeado, que serviram de base para as etapas seguintes da análise:

- os formatos predominantes (textos, *deepfakes*, imagens realistas e áudios sintéticos);
- as figuras públicas usadas nos conteúdos (como pessoas públicas no governo federal);
- as plataformas digitais mais recorrentes na disseminação da desinformação;
- e os tipos de estratégias narrativas associadas às campanhas enganosas (como promessas de indenização, mensagens alarmistas ou anúncios falsos com identidade visual forjada).

Na etapa de análise dos documentos institucionais (Apêndice B), o foco recaiu sobre as ideias produzidas ou omitidas e sua relação com os casos mapeados pela Agência Lupa. A interpretação buscou confrontar o que é normatizado com o que efetivamente circula, observando lacunas ou inconsistências entre os dois blocos empíricos.

Na sequência, a análise buscou ser sensível aos contextos de produção, circulação e legitimação dos conteúdos desinformativos, considerando os vínculos entre os sujeitos envolvidos, os recursos utilizados e os ambientes nos quais essas informações são disseminadas. Em muitos casos, foi possível observar a articulação simultânea entre diferentes elementos do regime de informação, como pode ser visualizada a seguir.

Figura 7: Representação dos elementos do regime de informação em um caso de desinformação gerada por IA generativa



Fonte: elaborado pela autora (2025).

No contexto da desinformação gerada por IA, a figura 7 ilustra como um artefato sintético (ex: imagem manipulada) se articula a um ator público (ex: presidente da República),

é disseminado por uma ação de engano (ex: narrativas enganosas) e vinculado a um dispositivo institucional (ex: como uma falha de moderação). A visualização busca tornar mais evidentes essas conexões, centrais à análise do regime de informação.

Os resultados foram organizados em tabelas temáticas com base nas quatro categorias (atores, ações, artefatos e dispositivos), visualizadas na seção 6 de resultados. Cada entrada foi acompanhada de observações interpretativas que articularam os conteúdos da Lupa com os dispositivos normativos das plataformas ou do PL 2338/2023. Isso permitiu identificar padrões de relação entre os elementos, bem como lacunas institucionais frente às práticas documentadas.

Por fim, para aprofundar a discussão dos achados parciais, foi adotado o processo de triangulação entre:

- 1) Os dados coletados e analisados;
- 2) Os conceitos do regime de informação;
- 3) O referencial teórico adotado.

Como observam Tracy e Hinrichs (2010), a triangulação fortalece a credibilidade interpretativa ao permitir a comparação entre diferentes formas de evidência como dados, conceitos e interpretações.

Essa estrutura metodológica dialogou diretamente com as ideias de desinformação, IA generativa e regimes informacionais discutidas ao longo do referencial teórico, assegurando unidade entre os procedimentos de análise e os marcos conceituais da pesquisa. A noção de desinformação, por exemplo, sustenta a decisão de observar o conteúdo circulante e os mecanismos que organizam sua visibilidade e circulação. Do mesmo modo, o conceito de regime de informação fornece o arcabouço necessário para compreender como sujeitos, plataformas, políticas e tecnologias se articulam para produzir efeitos de autoridade e influência informacional.

## **5.5 Limitações e aprendizados**

Um dos principais desafios metodológicos desta pesquisa foi estabelecer uma análise segura da desinformação gerada por IA generativa, já que nem sempre é possível rastrear com precisão suas origens ou comprovar tecnicamente o uso de IA na criação dos conteúdos, especialmente quando estes apresentam um alto grau de realismo. Para reduzir esse risco, optou-se por trabalhar apenas com casos previamente identificados e verificados pela plataforma brasileira Agência Lupa, cuja atuação é reconhecida pelo *International*

*Fact-Checking Network* (IFCN). Ainda que o reconhecimento pelo IFCN não deva ser entendido como um selo absoluto de neutralidade já que, assim como toda organização, também está inserida em contextos institucionais e possui interesses e limitações específicas, sua atuação tem sido relevante para a construção de parâmetros comuns no campo do *fact-checking*. Um exemplo é seu Código de Ética e Princípios, da qual a Agência Lupa é signatária.

Também não foi objetivo deste trabalho realizar verificação independente de casos ambíguos ou não reconhecidos como desinformação. Menções desse tipo, eventualmente localizadas nos documentos institucionais, foram registradas apenas como indícios complementares, desde que contribuíssem para observar as estratégias de visibilidade ou legitimação adotadas pelas plataformas.

Também reconhece-se que a análise documental de políticas de plataformas enfrenta limitações próprias, já que esses documentos estão em constante atualização e nem sempre são publicados com transparência e raramente mantêm versões arquivadas, o que compromete a rastreabilidade das mudanças. Além disso, a opacidade algorítmica e a ausência de mecanismos claros de responsabilização também são limites para compreender integralmente os dispositivos que estruturam o regime de informação investigado.

Essas limitações são inerentes a um objeto de estudo que envolve tecnologia emergente e dinâmica, em um cenário moldado pela lógica econômica das plataformas. Como destacam Laville e Dionne (2008, p. 174), “é sobre o método que repousa grande parte da validade dos saberes construídos”. Reconhecer esses limites não reduz a relevância do estudo; pelo contrário, reforça sua transparência, a validade dos resultados e evidencia aprendizados metodológicos importantes para pesquisas em contextos de inovação tecnológica acelerada.

## 6 DESINFORMAÇÃO E O REGIME DE INFORMAÇÃO NA ERA DA IA GENERATIVA

Foi possível compreender, no caminho de desenvolvimento desta pesquisa, que a circulação de informação nas plataformas digitais não se limita a objetivos comunicacionais, mas também participa da conformação de percepções, crenças e comportamentos, especialmente quando articulada a novidades tecnológicas capazes de criar e difundir conteúdos com alto grau de realismo.

Por isso, os resultados a seguir evidenciam a forma como a inteligência artificial (IA) generativa vem sendo mobilizada para potencializar a Desinformação Digital em Rede, permitindo que textos, imagens, áudios e vídeos falsos sejam produzidos e disseminados em escala e agilidade.

Com base nas categorias analíticas discutidas na Seção 4 e na metodologia delineada, esta seção apresenta os resultados da investigação, oferecendo uma leitura inicial da configuração do regime de informação da IA generativa no Brasil. O *corpus* empírico, composto por 46 casos de desinformação envolvendo o uso de IA verificados e publicados pela Agência Lupa entre maio de 2024 e março de 2025, somado à análise de documentos institucionais das plataformas digitais e da legislação pertinente (Apêndice B), permitiu a identificação de padrões recorrentes e a compreensão dos elementos que compõem este regime em construção.

Os casos analisados revelam a predominância de três grandes eixos temáticos: (1) golpes de natureza financeira, como falsas indenizações e promessas de resgate de valores; (2) manipulação da imagem de figuras públicas (políticos, artistas e especialistas) para conferir credibilidade a produtos ou narrativas fraudulentas; e (3) desinformação com fins de manipulação política e eleitoral. É importante notar que essas categorias não são mutuamente exclusivas e muitos casos de desinformação por IA generativa demonstram uma articulação entre dois ou mais desses eixos, potencializando o impacto e a eficácia das táticas desinformativas.

A seguir, os achados são detalhados a partir dos quatro componentes do regime de informação: atores, artefatos, dispositivos e ações, que permitem compreender o funcionamento e as implicações desse regime no contexto brasileiro.

### 6.1 Os atores do regime de informação da IA generativa

A configuração de um regime de informação pode ser definida, em primeiro lugar, pelos atores que nele operam, ou seja, "os sujeitos, as organizações, as regras e as autoridades informacionais" que dominam uma dada formação social (Gómez, 2012, p. 43). Aqui é importante entender que a atuação desses atores está intrinsecamente ligada à criação de artefatos, à influência sobre os dispositivos e à execução das ações informacionais, que serão tratados nas subseções a seguir.

Das 46 verificações analisadas, a imagem de políticos populares é a mais explorada, com destaque para o presidente Luiz Inácio Lula da Silva, Jair Bolsonaro e Nikolas Ferreira. A recorrência também é notada no uso de especialistas com alta credibilidade, como o médico Drauzio Varella e empresários como Elon Musk, para dar peso a narrativas fraudulentas ou políticas. A Tabela 6 sistematiza os atores identificados no *corpus* de casos verificados por agência de *fact-checking* que compõem esta pesquisa.

As figuras a seguir (8 e 9) apresentam exemplos de casos emblemáticos que circularam amplamente no Brasil e revelam como conteúdos sintéticos articulam atores sociais, políticos e econômicos nas estratégias de desinformação.

Figura 8 - Captura de tela de vídeo falsamente atribuído ao presidente Luiz Inácio Lula da Silva para promover golpe do ‘dinheiro acumulado no CPF’



A figura acima é uma imagem capturada a partir do vídeo analisado pela Agência Lupa que classificou o material como manipulado digitalmente, sinalizando-a as etiquetas de verificação na própria imagem. O vídeo circulou em junho de 2024 em redes digitais e aplicativos de mensagem, utilizando a figura do presidente Lula para dar aparência de anúncio oficial sobre um suposto “dinheiro acumulado no CPF”. Essa apropriação da imagem presidencial é exemplo de como os conteúdos sintéticos são instrumentalizados para fraudes financeiras, ao mesmo tempo em que reforça a dimensão informacional do problema: o uso de atores em imagens, vídeos e falas ultrarrealistas que confundem as pessoas e exploram a credibilidade de autoridades públicas.

Figura 9 - Captura de tela do vídeo publicado no Facebook que utiliza a imagem do médico Drauzio Varella para promover um suposto suplemento alimentar



**Fonte:** Agência Lupa. Disponível em: <https://lupa.uol.com.br/jornalismo/2024/12/18/drauzio-varella-nao-recomendou-remedio-para-disfuncao-sexual-video-e-deepfake>. Acessado em 17/09/2025

De modo semelhante, outro caso analisado pela Agência Lupa envolveu o uso da imagem do médico Drauzio Varella, referência no campo da medicina e da comunicação pública da saúde, em um vídeo manipulado digitalmente para promover um falso tratamento

contra a impotência sexual. A apropriação da imagem de Drauzio evidencia um padrão recorrente de uso de personalidades conhecidas. Nesse exemplo, mais uma vez o artefato sintético se combina à circulação massiva em plataformas digitais, potencializando os riscos de desinformação.

Tabela 6 – Atores identificados no regime de informação gerada por IA

Categoria do ator	Papel no regime	Exemplos ( <u>Apêndice A</u> )	Frequência / Relevância
Agentes de desinformação	Sujeitos (indivíduos ou grupos) que operam as ferramentas e criam o conteúdo sintético com motivação financeira (golpes) ou político-ideológica (polarização, difamação). Geralmente atuam no anonimato.	Criadores de anúncios fraudulentos sobre o "Desenrola Brasil". Produtores de vídeos de ataque com <i>deepfake</i> de teor sexual.	Ator central e implícito em todos os 46 casos. É a força motriz do regime.
Atores (figuras públicas) utilizados	Figuras públicas cuja imagem e credibilidade são manipuladas para dar legitimidade aos artefatos.	Políticos: Lula em golpe da falsa indenização; Nikolas Ferreira em golpe financeiro. Especialistas: Drauzio Varella em falsa recomendação de remédio. Celebidades: Fátima Bernardes em golpe de tratamento ocular.	Presente em mais da metade dos casos. É a principal tática de legitimação dos artefatos desinformativos.
Plataformas digitais	Atuam como atores que hospedam e cuja arquitetura e modelo de negócio (baseado em engajamento e anúncios) viabilizam a circulação massiva e lucram com o impulsionamento do conteúdo.	Meta (Facebook/Instagram): principal palco de golpes financeiros via anúncios. WhatsApp: canal para viralização em grupos. YouTube, X (Twitter), TikTok: usados para disseminação de vídeos e imagens manipuladas.	Viabilizam a circulação massiva e, no caso de anúncios fraudulentos, lucram com o impulsionamento.
Atores de contestação	Indivíduos e instituições que atuam para identificar, expor, regular ou combater o regime de informação.	Agências de checagem, mídia. Poder Público.	Atuam de forma reativa. A regulação estatal ainda é um dispositivo em construção.

Categoria do ator	Papel no regime	Exemplos ( <u>Apêndice A</u> )	Frequência / Relevância
IA Generativa (ator não-humano)	As próprias tecnologias de IA (modelos de difusão, LLMs) que geram os artefatos. Atuam como agentes não-humanos que reconfiguram o conceito de autoria e a percepção da realidade.	Ferramentas de <i>deepfake</i> de vídeo. Geradores de imagens ultrarrealistas.	Fundamental para a existência do regime analisado. Sua acessibilidade, gratuidade e facilidade de uso junto da capacidade de simulação é o que diferencia este regime de outros tipos de desinformação.

Fonte: elaborado pela autora (2025)

Embora a análise de frequência aqui apresentada na Tabela 6 revela padrões significativos dentro da amostra estudada, ela não tem a pretensão de representar a totalidade ou a exata proporção da desinformação gerada por IA generativa que circula no cenário brasileiro como um todo, dado o caráter específico e delimitado do recorte empírico adotado.

Após a sistematização dos atores, cabe ainda uma observação sobre a sociedade enquanto público impactado. Se são os indivíduos que recebem e compartilham os conteúdos que circulam, podem ser vistos como um ator do regime de informação. Trata-se de uma consideração apenas de forma indireta, considerando a sociedade interpretada como alvo e vítima desse tipo de desinformação. Ao trabalhar com fontes documentais, não foi possível acessar como os grupos sociais interpretaram ou se apropriaram dos conteúdos de forma sistematizada. Ainda assim, não se pode ignorar que a circulação depende justamente da adesão pelos cliques, compartilhamentos e interações que dão visibilidade a golpes, campanhas e ataques com os conteúdos sintéticos. Reconhece-se, portanto, a sociedade como ator indireto presente no regime de informação da desinformação por IA, mesmo sem ter sido objeto direto de análise neste estudo.

A análise dos atores evidencia que a desinformação gerada por IA generativa opera de forma distribuída, por meio da conexão entre agentes maliciosos anônimos, plataformas digitais, figuras públicas exploradas indevidamente, tecnologias de IA e atores de contestação. A facilidade de uso e o acesso a ferramentas capazes de criar cenas ultrarrealistas com personalidades conhecidas transformam essas figuras em ativos para legitimar as narrativas falsas.

Em diversos casos, também observou-se a publicação inicial por perfis anônimos, seguida de ampla circulação nas plataformas sem contestação ou moderação imediata. Mesmo com a atuação de controle de agências de checagem, veículos jornalísticos e instituições federais, a ausência de respostas rápidas por parte das plataformas contribui para a

permanência e amplificação da desinformação. Essa interação entre produção, difusão e reação revela um dos mecanismos centrais do regime de informação contemporâneo, no qual a Desinformação Digital em Rede se fortalece.

Em relação à cultura da pós-verdade descrita por Araújo (2020a, 2020b), a predominância da utilização de figuras públicas observada em mais da metade da amostra é uma tática estratégica que dialoga diretamente com o fenômeno, onde a adesão aos discursos depende da afinidade afetiva e da confiança depositada em personalidades reconhecidas, não da comprovação da verdade. A IA generativa, ao simular com ultrarrealismo a imagem e a voz desses indivíduos, atua como um ator informacional, não apenas gerando conteúdo, mas intervindo no processo comunicacional com a capacidade de moldar sentidos (Santaella; Kaufman, 2024).

Sendo assim, esse mapeamento dos atores constitui a base para compreender como os artefatos, examinados na subseção seguinte, materializam e potencializam essas estratégias, tornando a desinformação gerada por IA um elemento no ecossistema informacional contemporâneo.

## **6.2 Artefatos da desinformação sintética identificados**

Os artefatos, compreendidos como os meios materiais e simbólicos de produção e circulação da informação (Gómez, 2012), são os elementos mais visíveis do regime. Longe de serem materiais neutros, eles são projetados para produzir efeitos específicos, confirmando a premissa de Winner (1986) de que as tecnologias possuem qualidades políticas. No contexto analisado, os conteúdos sintéticos são as "armas" da desinformação, cuja sofisticação técnica visa dificultar a distinção entre o real e o sintético. Esses artefatos também são criados por atores específicos e sua circulação e impacto são diretamente moldados pelos dispositivos e pelas ações informacionais.

Inclusive, esse ultrarrealismo se manifesta também na diversidade de formatos que essas ferramentas são capazes de produzir, atingindo também pessoas comuns. Um exemplo é o vídeo manipulado de um advogado (Figura 10) que originalmente explicava detalhes sobre o lucro do FGTS. No conteúdo falso, sua voz foi adulterada para induzir o público a acreditar que seria possível resgatar valores por meio da consulta em um site fraudulento. O caso, verificado pela Agência Lupa, demonstra como a clonagem de voz se consolidou como recurso estratégico da Desinformação Digital em Rede, reforçando o potencial de dano quando vítimas comuns são envolvidas em narrativas fabricadas por IA.

Figura 10 - Captura de tela do vídeo em que fala de advogado sobre FGTS foi manipulada com uso de IA



**Fonte:** Agência Lupa. Disponível em: <https://lupa.uol.com.br/jornalismo/2025/01/17/video-em-que-advogado-fala-sobre-fgts-foi-manipulado-com-uso-de-ia>. Acessado em: 17/09/2025

A diversidade de casos analisados pode ser visualizada na tabela abaixo, que reúne os tipos de artefatos identificados, seus modos de uso e a frequência com que apareceram.

Tabela 7 – Artefatos da desinformação sintética identificados

Tipo de Artefato	Características e uso	Exemplo / Evidências (Apêndice A)	Relevância/Frequência
Vídeos em <i>Deepfake</i>	Uso de IA para manipular vídeos, sincronizando falas falsas com a imagem de uma pessoa ou criando cenas inteiramente novas.	Vídeo de Fátima Bernardes promovendo falso tratamento ocular. Vídeo de Lula e Alexandre de Moraes se beijando.	Utilizado em 27 dos 46 casos.
Imagens sintéticas	Criação de imagens realistas para simular eventos, endossos ou situações de forte apelo emocional.	Imagem de Michelle Bolsonaro e Donald Trump com a placa "Bolsonaro 2026". Imagem de homem resgatando bebê em enchente no RS.	Utilizado em 11 dos 46 casos.
Áudios sintéticos (clonagem de voz)	Geração de áudio com a voz de uma figura pública para ser sobreposto a vídeos ou usado em mensagens.	Áudio falso de Elon Musk anunciando programa educacional no Brasil.	Presente em múltiplos <i>deepfakes</i> , sendo o principal elemento nos casos.

Tipo de Artefato	Características e uso	Exemplo / Evidências (Apêndice A)	Relevância/Frequência
Anúncios	Combinação do conteúdo sintético (vídeo/imagem) com texto persuasivo e com a tecnologia de direcionamento algorítmico da plataforma.	Golpe do "Resgata Brasil" usando imagens de vários artistas. Golpe da falsa indenização da Serasa.	Presente na maioria dos golpes financeiros.

Fonte: elaborado pela autora (2025)

Os vídeos em *deepfake* como artefato principal indicam uma preferência por formatos que combinam estímulos visuais e auditivos que, como já apontavam Wardle e Derakhshan (2023), trata-se de uma estratégia que explora o maior potencial persuasivo do audiovisual. Além disso, a análise do "anúncio patrocinado" revela a complexidade do regime contemporâneo, pois ele se baseia no conteúdo falso e na articulação deste com a infraestrutura da plataforma (o sistema de anúncios), o que potencializa seu alcance e impacto de maneira estratégica, acessível e a baixo custo.

A presença de elementos de legitimação, como a imagem de figuras conhecidas, logotipos oficiais e linguagem institucional simulada aparecem como padrão recorrente nos artefatos analisados. Esses recursos operam como âncoras de credibilidade, explorando o reconhecimento imediato e a confiança prévia que o público atribui a determinadas personalidades, marcas e formas de comunicação. Tal estratégia aumenta o realismo do conteúdo falso e reduz a probabilidade de questionamento inicial, o que retarda a ação de contestação.

Além disso, esses elementos de legitimação cumprem uma função importante na Desinformação Digital em Rede ao explorarem o apelo emocional do público, amplificando o engajamento e, consequentemente, favorecendo o alcance algorítmico nas plataformas digitais. Assim, os artefatos são considerados como componentes estratégicos do regime de informação contemporâneo, no qual a DDR se consolida e se expande.

Reforça-se, também, que, ao simular eventos que nunca ocorreram, esses artefatos exploram a vulnerabilidade informacional do público brasileiro (OCDE, 2023), transformando a aparência de verdade em um fator mais persuasivo do que a própria veracidade. Essa capacidade de enganar e parecer humano, uma reflexão de Alan Turing (1950), ganha uma perigosa dimensão na era da IA generativa, onde os artefatos se tornam centrais na "degeneração da informação" no cenário de "pós-realidade desinformacional" (Marques; Laipelt, 2023) ao inundar o ecossistema com materiais gerados artificialmente, onde mais pessoas utilizam IAs generativas para produzir e publicar textos, fotografias e

vídeos, ou seja, a IA pode ser alimentada por seus próprios produtos artificiais.

### 6.3 Dispositivos de controle

A análise das políticas de plataformas digitais e do PL nº 2338/2023 incorporados nesta pesquisa ajudaram a compreender tais materiais como dispositivos de controle no interior do regime de informação. São regras, normas e políticas, formais ou informais, que delimitam possibilidades de circulação e estabelecem limites de atuação para atores e artefatos (Gómez, 2012). É importante notar que esses dispositivos não apenas regulam, mas também são influenciados pela atuação dos atores e pela natureza dos artefatos, impactando diretamente as ações informacionais.

Ao serem examinados em conjunto, revelam um traço recorrente de reatividade. Em vez de antecipar problemas, tanto as políticas de plataformas quanto as tentativas de regulação tendem a surgir como resposta a crises públicas ou pressões externas, o que limita seu caráter preventivo. Essa constatação está de acordo com o que aborda Schneider (2022) sobre a fraqueza regulatória e sua lentidão como fatores estruturais que permitem que a Desinformação Digital em Rede se estabeleça e persista no regime de informação contemporâneo. Na Tabela 8, estão sintetizadas essas diretrizes normativas e seus limites, evidenciando justamente o caráter reativo e a insuficiência preventiva dos dispositivos de controle identificados.

Tabela 8 – Dispositivos de controle e suas falhas

<b>Categoria do dispositivo</b>	<b>Papel no regime</b>	<b>Evidências / Análise</b>	<b>Relevância no regime</b>
Políticas e normas das plataformas (Google, Meta, TikTok e X)	As regras internas das plataformas (termos de uso, políticas de conteúdo) deveriam coibir a desinformação. Na prática, a moderação de anúncios pagos é o dispositivo mais falho.	A repetição de golpes financeiros idênticos na Meta (Facebook/Instagram), mesmo após múltiplas denúncias, evidencia a falha sistêmica deste dispositivo.	Sua falta de moderação é a principal condição de possibilidade para a circulação da desinformação monetizada.
Lógica de lucro das <i>big techs</i>	O modelo de negócio das plataformas, baseado na monetização da atenção e no lucro com publicidade direcionada. Este dispositivo cria um incentivo perverso.	O fato de que os agentes de desinformação <i>pagam</i> para impulsionar anúncios fraudulentos e as plataformas <i>lucram</i> com isso, transforma a desinformação em parte do modelo de negócio.	Este dispositivo permite e incentiva economicamente a circulação dos artefatos fraudulentos.

Categoria do dispositivo	Papel no regime	Evidências / Análise	Relevância no regime
Algoritmos de recomendação	O sistema que define a visibilidade dos conteúdos. Prioriza o engajamento sobre a veracidade, amplificando o alcance dos artefatos.	A rápida viralização de imagens de alto apelo emocional, como as da tragédia no RS, é uma evidência da ação deste dispositivo, que promove o que é "engajante", não o que é verdadeiro.	O amplificador do regime. Garante que os artefatos mais potentes atinjam uma audiência de forma orgânica ou semi-orgânica.
Marco regulatório (Em construção)	O conjunto de leis e normas estatais. No Brasil, há ausência de uma lei específica para IA em vigor.	O PL 2338/2023 é a principal tentativa de construir este dispositivo. Seus artigos sobre identificação de conteúdo sintético (Art. 19) e responsabilidade (Cap. V) visam justamente conter o regime, mas ainda não são uma realidade.	Sua ausência garante um ambiente de baixo risco para os agentes de desinformação.

Fonte: elaborado pela autora (2025)

Em relação ao documento analisado Projeto de Lei nº 2338/2023, há três artigos particularmente significativos que ajudaram na análise deste resultado e que merecem destaque: o Artigo 8º, que trata da supervisão humana em sistemas de alto risco; o Artigo 6º, que trata dos direitos de revisão e contestação; e o Artigo 19º, que estabelece a obrigatoriedade de identificação desses conteúdos.

Especialmente no Artigo 8º (BRASIL, p. 7), há a discussão sobre a responsabilização e o controle dos sistemas de alto risco:

Art. 8º A supervisão humana de sistemas de IA de alto risco buscará prevenir ou minimizar os riscos para direitos e liberdades das pessoas ou grupos afetados que possam decorrer de seu uso normal ou de seu uso em condições de utilização indevida razoavelmente previsíveis, viabilizando que as pessoas responsáveis pela supervisão humana possam, nos termos de regulamento, compreender, interpretar, decidir e intervir nos sistemas de IA, bem como priorizar o gerenciamento de riscos e impactos irreversíveis.

Este trecho é um dispositivo de controle fundamental, pois estabelece um mecanismo de freio (a intervenção humana) contra a opacidade e a autonomia dos sistemas de alto risco, ligando-se diretamente à necessidade de mitigar impactos irreversíveis.

Enquanto o Artigo 8º foca no mecanismo de freio contra a autonomia da IA (o controle humano), os direitos de contestação e revisão estão estabelecidos no Artigo 6º, inciso II, estabelece diretamente o direito de o cidadão solicitar a revisão e a contestação dos resultados do sistema de IA.

Art. 6º: A pessoa ou grupo afetado por sistema de IA de alto risco tem os seguintes direitos: I – direito à explicação sobre a decisão, a recomendação ou a previsão feitas pelo sistema; II – direito de contestar e de solicitar a revisão de decisões,

recomendações ou previsões de sistema de IA; III – direito à revisão humana das decisões, levando-se em conta o contexto, o risco e o estado da arte do desenvolvimento tecnológico.

No caso do Artigo 19º (BRASIL, p. 14) , em que determina o conteúdo sintético, visa criar um dispositivo técnico para combater o ultrarrealismo e a falta de transparência dos artefatos:

Quando o sistema de IA gerar conteúdo sintético, deverá, considerando o estado da arte do desenvolvimento tecnológico e o contexto de uso, incluir identificador em tais conteúdos para verificação de autenticidade ou de características de sua proveniência, modificações ou transmissão, conforme regulamento.

Na prática, os casos verificados nesta pesquisa demonstram que a ausência de rotulagem ou de mecanismos de autenticação facilita a circulação de anúncios enganosos e de peças ultrarrealistas criadas com IA, explorando justamente a dificuldade de distinguir o real do sintético. Trata-se de tentativa de corrigir uma das lacunas mais críticas do regime de informação, mas sua eficácia depende não apenas da aprovação legislativa, como também da capacidade técnica e da vontade política de implementá-lo de forma consistente.

Schneider (2022) destaca que um dos elementos centrais na Desinformação Digital em Rede é justamente a “escassa e difícil regulação em termos técnicos e jurídicos”, condição que favorece a expansão das narrativas falsas em rede. Ou seja, a ausência de dispositivos eficazes de controle, ou a fragilidade daqueles existentes, constitui parte estruturante dessa dinâmica. Nesse mesmo sentido, Bezerra (2023) observa que as plataformas digitais criam normas e políticas não como instrumentos neutros de regulação, mas como engrenagens orientadas pela lógica do lucro e da acumulação de capital, nas quais o engajamento é privilegiado em detrimento da qualidade informacional.

Em relação às normas internas das plataformas digitais para lidar com conteúdos sintéticos analisados nesta pesquisa, revelam tanto avanços quanto limitações dos dispositivos corporativos de controle:

- A. Google (em relação ao YouTube)<sup>46</sup>: determinou que “os criadores têm de incluir divulgações quando o conteúdo [...] faz com que uma pessoa real pareça dizer ou fazer algo que nunca fez, altera as filmagens de um evento ou um local real ou gera uma cena realista que nunca aconteceu na realidade”. Essa exigência mostra a preocupação com a verossimilhança dos conteúdos, embora delegue a responsabilidade ao próprio criador, o que fragiliza sua efetividade.

---

<sup>46</sup> Disponível em: [support.google.com/youtube/answer/14328491](https://support.google.com/youtube/answer/14328491)

- B. Google (em relação a anúncios)<sup>47</sup>: discutiu as normas em duas frentes: de um lado, proibindo conteúdos sintéticos sexualmente explícitos ou de nudez; de outro, exigindo, em anúncios políticos, a divulgação clara quando houver manipulação digital: “os anunciantes têm de divulgar os anúncios eleitorais com conteúdo sintético ou alterado digitalmente que represente de forma não autêntica pessoas ou eventos reais ou realistas”.
- C. Meta (Facebook, Instagram e Threads)<sup>48</sup>: a empresa anunciou em 2024 que “começaremos a adicionar rótulos de ‘informações de IA’ a uma gama mais ampla de conteúdo de vídeo, áudio e imagem quando detectarmos indicadores de imagem de IA padrão do setor ou quando as pessoas revelarem”. A política sinaliza a busca por maior transparência, mas, na prática, mantém a circulação do material manipulado, limitando-se a inserir advertências contextuais.
- D. TikTok<sup>49</sup>: proíbe a utilização de “figura pública quando usada para apoio político ou comercial” e adotou postura semelhante às demais ao estabelecer que “os criadores devem identificar todo o conteúdo gerado por IA que contenha imagens, áudio e vídeo realistas”. A regra toca diretamente em um dos problemas centrais observados nesta pesquisa, que é a apropriação da imagem de autoridades públicas em anúncios fraudulentos.
- E. X (antigo Twitter)<sup>50</sup>: incorporou em poucas linhas em sua política de autenticidade que o usuário “não pode compartilhar mídia não autêntica, incluindo mídia manipulada ou fora de contexto que possa resultar em confusão generalizada sobre questões públicas, impactar a segurança pública ou causar danos graves (‘mídia enganosa’)”. Apesar da formulação ampla, a aplicação prática dessa diretriz mostrou-se limitada frente à velocidade de disseminação da desinformação.

O que foi observado é que as plataformas analisadas convergem na rotulagem ou identificação, mas divergem quanto à força e moderação da aplicação e aos contextos proibidos. Isso enfraquece a contenção da desinformação, pois as rotulagens sugeridas,

---

<sup>47</sup> Disponível em: [support.google.com/adspolicy/answer/14720423](https://support.google.com/adspolicy/answer/14720423) e em: [support.google.com/adspolicy/answer/15142358](https://support.google.com/adspolicy/answer/15142358)

<sup>48</sup> Disponível em: <https://about.fb.com/br/news/2024/04/nossa-abordagem-para-rotular-conteudos-criados-com-ia-e-midias-manipuladas/>

<sup>49</sup> Disponível em: [https://support.tiktok.com/pt\\_BR/using-tiktok/creating-videos/ai-generated-content](https://support.tiktok.com/pt_BR/using-tiktok/creating-videos/ai-generated-content)

<sup>50</sup> Disponível em: <https://help.x.com/pt/rules-and-policies/authenticity>

embora sejam uma tentativa de transparência, não removem o artefato desinformativo que explora o engajamento.

Assim, os dispositivos de controle analisados nesta pesquisa, sejam corporativos, como as políticas de plataformas, ou institucionais, como o PL nº 2338/2023, buscam responder à pressão social e política por responsabilização, mas permanecem condicionados a interesses econômicos que moldam sua efetividade, especialmente nos casos que envolvem anúncios pagos nas plataformas digitais de caráter desinformativo, conforme discutidos nas subseções anteriores.

A análise também demonstra que, mesmo quando dispositivos de controle estão formalmente instituídos, seu alcance é limitado por três fatores centrais:

- lentidão e insuficiência na aplicação: conteúdos falsos permanecem disponíveis tempo suficiente para atingir grandes audiências;
- prioridade aos interesses econômicos: a moderação efetiva de anúncios pagos colide diretamente com o modelo de negócios das plataformas;
- ausência de regulamentação específica: a falta de marcos legais para IA e desinformação cria um ambiente de baixo risco para agentes maliciosos.

Como trazido por Silveira (2024), a fragilidade desses dispositivos é agravada pela opacidade dos modelos de IA e pela concentração de poder nas mãos das *big techs*, tornando a explicabilidade e a auditabilidade (Kaufman, 2021; Coeckelbergh, 2019; 2021) distantes para o usuário comum. A lentidão na resposta regulatória e a definição de critérios de moderação por interesses corporativos demonstram que, nesse regime de informação, os dispositivos, ao invés de proteger, podem inadvertidamente amplificar os riscos informacionais.

As políticas da plataforma frequentemente se mostram mais simbólicas do que efetivas e a promessa de um marco regulatório em andamento no Brasil mantém um cenário de baixo risco para os agentes da Desinformação Digital em Rede. Combinados, esses elementos revelam que os dispositivos de controle estabelecem limites apenas formais, mas deixam lacunas operacionais e estratégicas que permitem que atores maliciosos e artefatos sofisticados continuem a operar e a se adaptar.

#### **6.4 Ações de informação**

As ações de informação são as práticas realizadas pelos atores por meio dos artefatos que, como propõe Gómez (2003), podem ter diferentes finalidades (mediação, formativa, relacional). No regime de informação contemporâneo, essas ações impulsionam a circulação

da falsidade ou buscam contestá-la, revelando a dinâmica intrínseca a este regime. Essas ações são o elo que conecta os atores, os artefatos e os dispositivos, manifestando a complexidade do regime de informação da desinformação por IA generativa. A tabela a seguir categoriza as principais ações identificadas no *corpus*.

Tabela 9 – Ações de informação

Tipo de ação	Descrição da prática	Exemplo de caso (do <u>Apêndice A</u> )	Padrão identificado
<b>Ações que sustentam o regime</b>	<ol style="list-style-type: none"> <li>1) Criar e disseminar golpes baseados em falsas promessas de dinheiro fácil, produtos milagrosos ou benefícios sociais.</li> <li>2) Criar e disseminar conteúdo para atacar reputação de adversários ou para fabricar apoio a candidatos e pautas ideológicas.</li> <li>3) Produzir conteúdo de alto impacto emocional em contextos de crise ou tragédia para maximizar o engajamento e a viralização.</li> </ol>	<ol style="list-style-type: none"> <li>1) Golpe da falsa indenização usando a imagem de Lula. Falso tratamento para diabetes usando imagens de famosos.</li> <li>2) Falso apoio de Bolsonaro a Pablo Marçal. Montagem de Lula com cabeça enfaixada para simular agressão.</li> <li>3) Vídeo de incêndios catastróficos falsamente atribuídos a Los Angeles. Imagens falsas de corpos boiando na enchente do Rio Grande do Sul em 2024.</li> </ol>	<ol style="list-style-type: none"> <li>1) Ação mais recorrente, nos casos analisados.</li> <li>2) Ação estratégica com picos de ocorrência em períodos pré-eleitorais.</li> <li>3) Padrão oportunista que explora a vulnerabilidade emocional do público.</li> </ol>

Fonte: elaborado pela autora (2025)

Percebe-se que esse regime de informação é mantido por um conjunto de ações estratégicas e recorrentes, com destaque para aquelas voltadas ao ganho financeiro. A repetição de formatos de golpes sugere um *modus operandi* industrializado. Em contrapartida, a ação de verificação, embora fundamental, atua de forma reativa, analisando o conteúdo desinformativo apenas depois que ele já circulou e causou impacto. Essa assimetria temporal entre as ações de produção e as de contestação é uma das razões da eficácia do regime de informação. A observação dos casos também demonstrou que a desinformação não se encerra apenas com a checagem, pois os conteúdos já desmentidos frequentemente voltam a circular com pequenas alterações ou novos enquadramentos, evidenciando as estratégias de recirculação e a fragilidade das ações corretivas.

Chegando a este ponto dos resultados, vale destacar que a leitura combinada dos quatro componentes evidencia que o regime de informação da Desinformação Digital em Rede não se sustenta apenas pela presença isolada de atores maliciosos, artefatos persuasivos, dispositivos falhos e ações recorrentes, mas por sua conexão dinâmica e adaptativa. Ou seja,

em diversos casos, a atuação inicial de um agente anônimo, associada ao uso de IA generativa para criação de artefatos ultrarrealistas, encontra oportunidade na arquitetura das plataformas, cujos algoritmos privilegiam o engajamento. A ausência de dispositivos regulatórios eficazes e a lentidão nas ações de contestação completam o ciclo, permitindo que narrativas falsas atinjam públicos amplos antes de qualquer mitigação.

Os resultados que compõem esse cenário complexo também se refletem na articulação dos eixos temáticos identificados nos casos analisados: golpes de natureza financeira, manipulação da imagem de figuras públicas e desinformação com fins políticos e eleitorais. Importante reforçar que é comum que um mesmo caso combine a manipulação de uma figura pública (ator/artefato) para promover um golpe financeiro (ação/categoria temática) ou para disseminar uma narrativa política (outra categoria temática), demonstrando que essas categorias não são mutuamente exclusivas, mas sim estratégias interligadas que potencializam o impacto dessa desinformação.

Há, também, a ideia de que a informação nesse regime é construída e interpretada de forma contextual, pois seu significado é moldado por interesses e relações de poder que definem o que é reconhecido como verdade (Capurro; Hjørland, 2007). Esse reconhecimento é relacionado à IA generativa que, ao facilitar a criação massiva de conteúdos sintéticos ultrarrealistas desinformativos, potencializa as ações de manipulação e “semeadura de dúvida” (Westerlund, 2019), tornando essa distinção um desafio constante para a sociedade. Diante disso, a eficácia das ações de contestação depende, além da tecnologia, da capacidade humana de discernimento e da responsabilidade dos agentes de responsabilidade (Coeckelbergh, 2019; 2021) em programar sistemas que lidem com a incerteza sobre o que os humanos realmente querem (Russell, 2021).

Essa dinâmica também revela que a Desinformação Digital em Rede, como apontado por Schneider (2022), se consolida por sua capacidade de persistir e se adaptar, desafiando a noção de que a verdade, uma vez restabelecida, elimina o engano. Em suma, essa desinformação no regime de informação contemporâneo se manifesta como uma prática que é amplificada pela IA generativa, onde a hiperrealidade dos conteúdos sintéticos enganosos muitas vezes se sobrepõe à veracidade.

## 7 CONSIDERAÇÕES FINAIS

O interesse por esta pesquisa começou com a observação da projeção global da inteligência artificial (IA) quando, a partir de 2022, ferramentas como o ChatGPT e outras plataformas similares tornaram-se acessíveis a qualquer pessoa com conexão à internet. Como reforçam os estudos da McKinsey (2023), a ferramenta da OpenAI reuniu mais de 100 milhões de usuários nos primeiros meses de lançamento. A partir desse momento, os usuários, de estudantes curiosos a profissionais de múltiplas áreas, passaram a explorar suas funcionalidades de forma recreativa, desde um pedido de escrita de poesias, criação de imagens ou produção de vídeos improváveis entre personagens históricos. Esse uso lúdico, amplificado por redes digitais, ajudou a consolidar a presença da IA generativa no cotidiano das pessoas, transformando-a em um fenômeno cultural e tecnológico.

Porém, como já ocorreu em outras transições tecnológicas como *Adobe Photoshop* mencionado ao longo desta investigação, a mesma ferramenta que possibilita usos criativos e inovadores também abre espaço para usos maliciosos. A facilidade de criar textos, imagens, áudios e vídeos ultrarrealistas sem exigir conhecimentos técnicos avançados introduziu novas formas de manipulação informacional, agora mais sofisticadas e difíceis de detectar. Por isso, investigar a desinformação gerada por IA se apresentou como um campo de interesse acadêmico nesta pesquisa com a necessidade para compreender e enfrentar os riscos que essa tecnologia impõe ao regime de informação contemporâneo.

Os resultados obtidos reforçam essa leitura. A análise dos documentos e dos casos verificados confirmou a presença de um padrão recorrente em que as plataformas digitais tratam a desinformação gerada por IA com medidas pontuais e tardias. Políticas de rotulagem, remoção e transparência foram observadas, mas sempre sob a lógica de mitigação do dano e não de prevenção. Nos casos avaliados neste documento, como as campanhas com imagens manipuladas de figuras públicas, a permanência do conteúdo esteve diretamente relacionada aos mecanismos de engajamento das próprias plataformas. Esses resultados evidenciam a fragilidade regulatória do regime de informação contemporâneo e reforçam a leitura de que a desinformação sintética é consequência estrutural desse sistema, e não uma exceção a ele.

Diante desse cenário, a análise desenvolvida nesta dissertação partiu do reconhecimento de que a desinformação gerada por IA generativa é um fenômeno que faz parte do regime de informação contemporâneo. Ao transformar a produção e circulação de conteúdos ultrarrealistas, a IA generativa potencializa a lógica da Desinformação Digital em Rede (DDR), como formulada por Schneider (2022), integrando-se às mesmas engrenagens

que sustentam os modelos de negócios das plataformas digitais. Inclusive, a escolha da DDR como referencial não foi casual, pois esse caminho permitiu compreender a desinformação como uma força ativa dentro do regime de informação, articulada a atores, artefatos, dispositivos e ações que se retroalimentam. Investigar esse fenômeno, portanto, foi essencial para revelar como o acesso a ferramentas como ChatGPT, Gemini e Veo amplia a escala, a velocidade e o realismo de narrativas enganosas, influenciando processos políticos, econômicos e culturais.

Como mostram os resultados consolidados, este regime de informação se sustenta a partir de uma tecnologia inserida em um ambiente que envolve atores maliciosos, artefatos sintéticos, dispositivos de circulação permissivos e uma lógica de plataformas que privilegia o engajamento acima da verdade. Esse fenômeno também se amplia pela acessibilidade das ferramentas de IA generativa, cuja disponibilidade pública e gratuita permite que qualquer usuário, sem conhecimentos técnicos avançados, possa produzir conteúdos com alto grau de realismo e com narrativas fraudulentas ou difamatórias.

As plataformas digitais são espaços que, como aponta a análise, lucram com a circulação e a viralização desses conteúdos, especialmente por meio de anúncios pagos, uma lógica que converge com a perspectiva de autores como Westerlund (2019) e Bezerra (2023), que apontam para os modelos de negócio das *big techs* como facilitadores da desinformação. É neste ponto que a formulação de Frohmann (1995) se torna central sobre como a informação não é uma entidade neutra, mas uma "prática discursiva situada em relações de poder".

Ao adotar o conceito de regime de informação, esta pesquisa se inspirou nas formulações de González de Gómez (2003; 2012), que o compreende como um sistema dinâmico e disputado onde a informação legítima é constantemente negociada, e na proposta analítica de Delaia e Freire (2010). A partir dessa base, a investigação direcionou seu olhar para a Desinformação Digital em Rede (DDR), que integra o regime de informação contemporâneo, pois os resultados demonstraram que a DDR, potencializada pela IA generativa, opera como um modo de produção informacional dominante. Sua facilidade na criação de conteúdos sintéticos ultrarrealistas e a exploração da confiança que eles geram se contrapõem diretamente à fragilidade dos mecanismos de verificação e resposta, conforme verificado na análise dos quatro elementos do regime de informação contemporâneo.

Esse contexto reforça o que Schneider (2022) descreve como um fenômeno inédito, com custo relativamente baixo de suas operações, alcance imenso e customizado, com escassa e difícil regulação. A desinformação, nesse processo, deixa de ser ocasional, se impõe como

estrutura e se aproveita da lógica da pós-verdade, onde a persuasão e a emoção se sobrepõem à veracidade. A pesquisa, portanto, colabora para o entendimento de que essa dinâmica está no regime de informação contemporâneo, no qual a desinformação se consolida como um elemento central, e não apenas um desvio.

Retomando a problemática, atenta-se para a desinformação digital em rede como parte constitutiva do regime de informação contemporâneo, sustentado por plataformas digitais orientadas ao engajamento e ao lucro.

Retomando a problemática discutida ao longo desta pesquisa, cabe reforçar que a popularidade de ferramentas como ChatGPT, Copilot e Gemini se mostrou um fenômeno de alcance global que, ao mesmo tempo, trouxe novas possibilidades criativas e ampliou os riscos da desinformação digital. A partir da ótica da Ciência da Informação, foi possível perceber que os achados da pesquisa levam a questão para além da simples verificação. Mesmo um fato sendo verdade ou não, revela-se um problema de dimensões democráticas e culturais, pois o impacto se consolida na circulação dos conteúdos manipulados, principalmente, de autoridades públicas com rapidez, explorando a confiança das pessoas.

Por isso, a definição do objetivo central de compreender como se configura o regime de informação da desinformação gerada por IA generativa, permitiu mostrar a mobilização feita por corporações que privilegiam engajamento e lucro em detrimento da veracidade. Isso reforça a lógica da Desinformação Digital em Rede e expõe a fragilidade dos mecanismos tradicionais de checagem e moderação.

Diante da dimensão e da rápida evolução tecnológica do objeto em questão, é preciso reconhecer que a complexidade do tema exigiu um recorte metodológico focado e, ao centrar a análise em uma única agência de checagem e em um intervalo de tempo específico, o estudo oferece um retrato aprofundado, mas não exaustivo, do fenômeno. Por isso, importante reforçar que os padrões e as frequências identificados nesta pesquisa, especialmente em relação aos atores e artefatos da desinformação gerada por IA generativa, são resultados diretos da análise de um *corpus* específico (os casos verificados por agências de *fact-checking* no período delimitado). Embora esses achados ofereçam uma compreensão aprofundada das dinâmicas operantes dentro dessa amostra, eles não têm a pretensão de representar a exata proporção da desinformação que circula no cenário brasileiro. No entanto, ao trazer luz às estratégias e as articulações entre os componentes do regime de informação neste recorte empírico, este estudo contribui significativamente para o debate sobre os desafios impostos pela IA generativa e a Desinformação Digital em Rede, abrindo caminhos para futuras investigações que possam expandir essa compreensão para um panorama mais amplo.

Portanto, apesar dessa delimitação, a abordagem se mostra necessária, como aponta Araújo (2014), já que um dos desafios centrais da Ciência da Informação é, justamente, "acompanhar criticamente os impactos de tecnologias emergentes". Este documento responde a essa perspectiva ao investigar um fenômeno recente como a IA generativa.

Para dar conta da complexidade do fenômeno, a pesquisa propõe a construção de um diálogo teórico entre diferentes campos do conhecimento, tendo a Ciência da Informação como alicerce e de onde partem os conceitos de regime de informação de Gómez (2003; 2012) e da Desinformação Digital em Rede (Schneider, 2022) que estruturam a análise, um caminho contextualizado pelo cenário da pós-verdade, discutido por Carlos Alberto Araújo (2020a, 2020b). Para a dimensão tecnológica, o estudo mobilizou os trabalhos de Wardle e Derakhshan (2023) e Marques e Laipelt (2024) sobre a desordem informacional e o uso massivo da ferramenta, bem como as reflexões de Dora Kaufman (2021) e seus estudos com Lucia Santaella (Kaufman; Santaella, 2024) sobre a ética e a semiótica dos conteúdos sintéticos. Acredita-se, portanto, que a articulação dessas perspectivas constitui um caminho teórico-metodológico robusto e pertinente para alcançar os objetivos propostos.

Por fim, vale ressaltar que a presente investigação, ao focar na configuração do regime de informação contemporâneo, pode abrir caminhos para futuras pesquisas que explorem outras perspectivas críticas do fenômeno. Uma ideia, como aponta Silveira (2024), seria investigar as condições materiais de produção da IA generativa, incluindo seus impactos ambientais e trabalhistas, que são dilemas éticos para além do conteúdo. Outro caminho promissor seria analisar a eficácia das ações de combate à desinformação, mensurando se elas de fato alteram o comportamento dos usuários, e explorar estratégias para mitigar a sobrecarga cognitiva diante do volume de informações, considerando a fragilidade humana nesse ecossistema. Por fim, a partir do diagnóstico inicial realizado, a Ciência da Informação pode contribuir para o desenvolvimento de sistemas de IA humildes ou beneficentes, como propõe Russell (2021), que aprendam com as preferências humanas em vez de otimizar cegamente objetivos que se provam destrutivos.

Não se trata, portanto, de buscar uma resposta definitiva, mas de levantar perguntas críticas diante de um cenário em que a distinção entre o autêntico e o sintético se torna cada vez mais difícil. Se “viver é muito perigoso” (Guimarães Rosa, 1994), como levantado na introdução deste documento, talvez hoje o perigo esteja menos no caminho e mais na manipulação dos sinais que guiam nosso caminho. Ao seguir na leitura crítica dos regimes informacionais, esta pesquisa espera contribuir, além para o campo da Ciência da Informação, com a construção de uma cidadania reflexiva e menos vulnerável à desordem informacional.



## REFERÊNCIAS

- ALI, S. *et al.* Explainable Artificial Intelligence: What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, p. 101805, 2023. Disponível em: <https://doi.org/10.1016/j.inffus.2023.101805>. Acesso em: 27 jan. 2025.
- ALMEIDA JUNIOR, O. F.; et al.. Mediação da informação e múltiplas linguagens. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 2, n. 1, 2009.
- ARAÚJO, C. A. A. **O que é ciência da informação**. Belo Horizonte: KMA, 2018.
- ARAÚJO, C. A. A. A pós-verdade como desafio central para a ciência da informação contemporânea. **Em Questão**, Porto Alegre, v. 27, n. 1, p. 13–29, 2020a. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/101666>. Acesso em: 20 jan. 2025.
- ARAÚJO, C. A. A. O fenômeno da pós-verdade e suas implicações para a agenda de pesquisa na Ciência da Informação. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis/SC, Brasil, v. 25, p. 01–17, 2020b. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e72673>. Acesso em: 20 jan. 2025.
- ARAUJO, R. F. Atores e ações de informação em redes sociais na internet: pensando os regimes de informação em ambientes digitais. **DataGramZero**, v. 15, n. 3, 2014. Disponível em: <https://brapci.inf.br/v/8239>. Acesso em: 22 jan. 2025.
- BARRETO, K. Lula sanciona projeto que aumenta a pena para crimes contra mulheres com uso de IA. G1, 24 abr. 2025. **G1 Política**. Disponível em: <https://g1.globo.com/politica/noticia/2025/04/24/lula-sanciona-projeto-que-aumenta-a-pena-p-ara-crimes-contra-mulheres-com-uso-de-ia.ghtml>. Acesso em: 20 mai. 2025.
- BBC BRASIL. Demissão de pesquisadora do Google acende debate sobre ética e viés em inteligência artificial. **BBC News Brasil**, [S. l.], 4 dez. 2020. Disponível em: <https://www.bbc.com/portuguese/geral-55195078>. Acesso em: 01 jan. 2025.
- BELUZZO, R. C. B. **A competência em informação no Brasil**: cenários e espectros. São Paulo: ABECIN Editora, 2018.
- BEZERRA, A. C. O regime de desinformação das Big Techs. Entrevista concedida ao canal II Diálogo sobre Bibliotecas (In)formadoras – DBI. **YouTube**, 9 ago. 2023b. Disponível em: <https://www.youtube.com/live/SUogoN2yrHk>. Acesso em: 20 mai. 2025.
- BEZERRA, E. *et al.* Regime de informação: abordagens conceituais e aplicações práticas. **Em Questão**, Porto Alegre, v. 22, n. 2, p. 60–86, 2016. DOI: 10.19132/1808-5245222.60-86. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/57935>. Acesso em: 21 jan. 2025.
- BRASIL. Senado Federal. Projeto de Lei nº 2338/2023. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília: Senado Federal. 17 mai. 2025. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233> Acesso em: 24 jun. 2024.
- BRISOLA, A. **Competência crítica em informação como resistência à sociedade da desinformação sob um olhar freiriano**. Tese (Doutorado) – Universidade Federal do Rio de

Janeiro, Rio de Janeiro, 2021. Disponível em: <https://ridi.ibict.br/handle/123456789/1165>  
Acesso em: 20 jul. 2024.

BRISOLA, A.; BEZERRA, A. C. Desinformação e circulação de “fake news”: distinções, diagnóstico e reação. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 19., 22-26 out. 2018, Londrina. **Anais [...]**. Marília: Unesp, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/102819>. Acesso em: 27 mai. 2024.

CAPURRO, R. HJORLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, [S. l.], v. 12, n. 1, 2007. Disponível em:  
<https://periodicos.ufmg.br/index.php/pci/article/view/22360>. Acesso em: 6 jun. 2025.  
CHOMSKY, N. **Mídia: Propaganda Política e Manipulação**. São Paulo: Martins Fontes, 2014.

COECKELBERGH, M. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. **AI & Society**, v. 34, n. 4, 2019. Disponível em:  
<link.springer.com/article/10.1007/s11948-019-00146-8>. Acesso em: 5 jun. 2024.

COECKELBERGH, M. Narrative Responsibility and Artificial Intelligence: How AI Challenges Human Responsibility and Sense-making. **Ethics and Information Technology**, v. 23, n. 2, 2021. Disponível em:  
<https://link.springer.com/article/10.1007/s00146-021-01375-x>. Acesso em: 5 jun. 2024.

CONVERSION. **Pesquisa nacional: inteligência artificial no comportamento do consumidor brasileiro**. São Paulo: Conversion, 2025. Disponível em:  
<https://www.conversion.com.br/wp-content/uploads/2025/05/Conversion-Pesquisa-IA.pdf>.  
Acesso em: 26 jun. 2025.

COUTINHO, S. Artigo de reitor da UFRJ repudia censura à ciência. **UFRJ**, Rio de Janeiro, 12 jun. 2024. Disponível em:  
<https://ufrj.br/2024/06/artigo-de-reitor-da-ufrj-repudia-censura-a-ciencia/>. Acesso em: 21 jun. 2024.

DELAIA, C. R.; FREIRE, I. M. Subsídios para uma política de gestão de informação da Embrapa Solos: à luz do regime de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 15, n. 3, p. 107-130, set./dez. 2010. Disponível em:  
<https://www.scielo.br/j/pci/a/bRhBtd5yC77tyjRjFTRSqL/>. Acesso em: 25 nov. 2024.

FLORIDI, L. **The Fourth Revolution: How the Infosphere Is Reshaping Human Reality**. New York: Oxford University Press, 2014.

FROHMANN, B. Talking information policy beyond information science: applying the actor network theory. In: OLSON, H. A.; WARD, D. B. (Eds.). **ANNUAL CONFERENCE OF THE CANADIAN ASSOCIATION FOR INFORMATION SCIENCE**, 23., 1995, Edmonton, Alberta, Proceedings 1995.

FURTADO, M. Facebook: estudo aponta 643 anúncios falsos sobre o Desenrola Brasil e o Voa Brasil. **Extra.globo.com**, 15 out. 2024. Disponível em:  
<https://extra.globo.com/economia/noticia/2024/10/facebook-estudo-aponta-643-anuncios-falsos-sobre-o-desenrola-brasil-e-o-voa-brasil.ghtml>. Acesso em: 03 mai. 2025.

G1. VEO 3: IA de vídeos realistas do Google 'bomba' nas redes. **G1 Tecnologia**, 3 jun. 2025. Disponível em:

<https://g1.globo.com/tecnologia/noticia/2025/06/03/veo-3-ia-de-videos-realistas-do-google-bo-mba-nas-redes-mas-preocupa-usuarios-imagina-nas-eleicoes.ghtml>. Acesso em: 16 jun. 2025.

GOES, B. Casos de deepfake chegam à Justiça Eleitoral, que cobra WhatsApp. **O Globo**, 24 jan. 2024. Disponível em:

<https://oglobo.globo.com/politica/noticia/2024/01/24/casos-de-deep-fake-chegam-a-justica-eleitoral-que-cobra-whatsapp.ghtml>. Acesso em: 29 jan. 2025.

GOLDMAN SACHS. **Gen AI**: too much spend, too little benefit? New York: Goldman Sachs Global Investment Research, jun. 2024. Disponível em:

<https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit/>. Acesso em: 21 jun. 2025.

GOMES, H. F. Mediação da informação e suas dimensões dialógica, estética, formativa, ética e política: um fundamento da Ciência da Informação em favor do protagonismo social.

**Informação & Sociedade**, [S. l.], v. 30, n. 4, p. 1–23, 2020. Disponível em:

<https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/57047>. Acesso em: 21 out. 2024.

GONZÁLEZ DE GÓMEZ, M. N. Regime de informação: construção de um conceito.

**Informação & Sociedade**: Estudos, João Pessoa, v. 22, n. 3, p.43-60, set. 2012. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/14376/8576>. Acesso em: 12 dez. 2024.

GONZÁLEZ DE GÓMEZ, M. N. Escopo e abrangência da Ciência da Informação e a pós-graduação na área: anotações para uma reflexão. **Transinformação**, v. 15, n. 1, p. 31-43, 2003. Disponível em:

<https://www.scielo.br/j/tinf/a/FwJWGzhN77SSYWNqwHHyYgw/?format=pdf&lang=pt>. Acesso em: 12 dez. 2024.

GOOGLE; IPSOS. **Our life with AI: from innovation to application**. [S.l.]: Google/Ipsos, jan. 2025. Disponível em:

[https://publicpolicy.google/resources/our\\_life\\_with\\_ai\\_google\\_ipsos\\_report.pdf](https://publicpolicy.google/resources/our_life_with_ai_google_ipsos_report.pdf). Acesso em: 31 jan. 2025.

GUIMARÃES ROSA, J. **Grande sertão**: veredas. Rio de Janeiro: Nova Aguilar, 1994.

HABIB LANTYER, V., The Phantom Menace: Generative AI Hallucinations and Their Legal Implications. **SSRN**, 2025. Disponível em: <https://ssrn.com/abstract=5167036>. Acesso em: 20 jan. 2025.

INTERESTELAR. Direção: Christopher Nolan. Produção: Emma Thomas, Christopher Nolan e Lynda Obst. Estados Unidos: Paramount Pictures; **Warner Bros. Pictures**, 2014. HBO Max (169 min).

JANIAKI, L. U.; SANTOS, L. C. dos. Aos Fatos, Lupa e Fato ou Fake: metodologias de agências de checagem no combate à desinformação. *In*: INTERCOM. 47º Congresso Brasileiro de Ciências da Comunicação, 5–6 set. 2024, Balneário Camboriú. **Anais [...]**. São Paulo: Intercom, 2024. Disponível em:

<https://sistemas.intercom.org.br/pdf/submissao/nacional/17/06282024172510667f1c260e45d.pdf>. Acesso em: 21 mai. 2025.

JIANG M., BELLI, L. **Digital Sovereignty in the BRICS Countries**: How the Global South and Emerging Power Alliances Are Reshaping Digital Governance. [S.l.]: Cambridge

University Press; 2025. Disponível em: <https://doi.org/10.1017/9781009531085>. Acesso em: 20 fev. 2025.

KAUFMAN, D. Inteligência Artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA. **PAULUS**: Revista de Comunicação da FAPCOM, v. 5, n. 9, 2021. Disponível em: Disponível em: <https://revista.fapcom.edu.br/index.php/revista-paulus/article/view/453>. Acesso em: 03 jun. 2024.

KEMP, S. Digital 2024: Brazil. **DataReportal**, 21 fev. 2024. Disponível em: <https://datareportal.com/reports/digital-2024-brazil>. Acesso em: 6 jun. 2025.

KUMAR, N. Grok AI statistics. **Demand Sage**, 16 jun. 2025. Disponível em: <https://www.demandsage.com/grok-ai-statistics/>. Acesso em: 18 jun. 2025.

LAVILLE, C.; DIONNE, J. **A construção do saber**: manual de metodologia da pesquisa em ciências humanas. Porto Alegre: Artmed; Belo Horizonte: Editora UFMG, 1999.

LORENZ, P., PERSET, K. BERRYHILL, J. **Initial policy considerations for generative artificial intelligence**. OECD Artificial Intelligence Papers, No. 1, OECD Publishing, Paris, <https://doi.org/10.1787/fae2d1e6-en>. Disponível em: [https://www.oecd.org/en/publications/initial-policy-considerations-for-generative-artificial-intelligence\\_fae2d1e6-en.html](https://www.oecd.org/en/publications/initial-policy-considerations-for-generative-artificial-intelligence_fae2d1e6-en.html) Acesso em: 20 jan. 2025.

LE COADIC, Y. F. **A Ciência da Informação**. tradução de Maria Yêda F. S. de Filgueiras Gomes. Brasília: Briquet de Lemos, 1996. *Perspectivas Em Ciência Da Informação*, 1(2).

MARQUES, J. F. **Das práticas desinformacionais ao regime de desinformação**: as narrativas do governo Bolsonaro na pandemia de Covid-19. 2023. Tese (Doutorado) - Universidade Federal da Paraíba, João Pessoa, 2023. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/27112>. Acesso em: 20 nov. 2024.

MARQUES, J. F.; ALVES, E. C. Entre regimes de informação e desinformação: modos de produção informacional na contemporaneidade. **Perspectivas em Ciência da Informação**, v. 29, n., 2024. Disponível em: <https://brapci.inf.br/v/338395>. Acesso em: 20 jan. 2024.

MARQUES, S. D.; LAIPELT, R. do C. F. Pós-realidade e Teoria da Desinformação: inquietações sobre o uso massivo de IA Generativa. *In: FÓRUM DE ESTUDOS EM INFORMAÇÃO, SOCIEDADE E CIÊNCIA*, 5., 2023, Porto Alegre. **Anais...** Porto Alegre, Rio Grande do Sul: 2023, p. 132-140. Disponível em: <https://lume.ufrgs.br/handle/10183/265448>. Acesso em: 20 nov. 2024.

MCKINSEY. **What's the future of generative AI**: An early view in 15 charts. [S.l.]: McKinsey & Company, ago. 2023. Disponível em: <https://www.mckinsey.com/featured-insights/mckinsey-explainers/whats-the-future-of-generative-ai-an-early-view-in-15-charts>. Acesso em: 14 jan. 2025.

NETLAB. Danos causados pela publicidade enganosa na Meta. **NetLab/UFRJ**, 25 mar. 2024. Disponível em: <https://netlab.eco.ufrj.br/post/danos-causados-pela-publicidade-enganosa-na-meta>. Acesso em: 20 jan. 2025.

NICOCELI, A. Youtubers ensinam como lucrar com uso de inteligência artificial para criar vídeos e livros infantis 'em minutos'. **G1**, 16 abr. 2024. Disponível em:

<https://g1.globo.com/educacao/noticia/2024/04/16/youtubers-ensinam-como-lucrar-com-uso-de-inteligencia-artificial-para-criar-videos-e-livros-infantis-em-minutos.ghhtml>. Acesso em: 20 jan. 2025.

OLIVEIRA, M. L. P. de. **Competência crítica em informação e fake news**: das metodologias de fact-checking à auditabilidade do sujeito comum. 2020. Tese (Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2020. Disponível em: Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/18200> Acesso em: 24 jun. 2024.

O'NEIL, C. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. 1 ed. Santo André, SP: Rua do Sabão, 2021.

OPENAI. Security on the path to AGI. San Francisco, 26 mar. 2025. **OpenAI**. Disponível em: <https://openai.com/index/security-on-the-path-to-agi>. Acesso em: 21 mai. 2025.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **The OECD Truth Quest Survey: Methodology and findings**. OECD Digital Economy Papers, No. 369, OECD Publishing, Paris, 2024. <https://doi.org/10.1787/92a94c0f-en>. Disponível em: [https://www.oecd.org/en/publications/2024/06/the-oecd-truth-quest-survey\\_a1b1739c.html](https://www.oecd.org/en/publications/2024/06/the-oecd-truth-quest-survey_a1b1739c.html) Acesso em: 20 jan. 2025.

PICHAU, S. Google I/O 2025: da pesquisa à realidade – 20 de maio de 2025. **Google Blog**, 20 maio 2025. Disponível em: <https://blog.google/intl/pt-br/novidades/google-io-2025-da-pesquisa-a-realidade-20-de-maio-d-e-2025/>. Acesso em: 01 jun. 2025

PIMENTA, R. M. Por que humanidades digitais na ciência da informação? Perspectivas progressas e futuras de uma prática transdisciplinar comum. **Informação & Sociedade: Estudos**, João Pessoa, v. 30, n. 2, 2020. DOI: 10.22478/ufpb.1809-4783.2020v30n2.52122. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/52122>. Acesso em: 8 jul. 2025.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2ed. Novo Hamburgo: Feevale, 2013.

RUSSELL, S. **Inteligência artificial a nosso favor**: como manter o controle sobre a tecnologia. Tradução: Berilo Vargas. São Paulo: Companhia das Letras, 2021.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 3. ed. New Jersey: Prentice Hall, 2009.

SANTAELLA, L.; KAUFMAN, D. A Inteligência artificial generativa como quarta ferida narcísica do humano. **MATRIZES**, São Paulo, Brasil, v. 18, n. 1, p. 37–53, 2024. DOI: 10.11606/issn.1982-8160.v18i1p37-53. Disponível em: <https://www.revistas.usp.br/matrizes/article/view/210834>. Acesso em: 24 jun. 2024.

SCHNEIDER, M. **A era da desinformação**: pós-verdade, fake news e outras armadilhas. Rio de Janeiro, Garamond, 2022.

SEVERINO, A. J. **Metodologia do trabalho científico**. 24 ed. São Paulo: Cortez, 2016.

SICHMAN, J. S. **Inteligência artificial e sociedade: avanços e riscos.** *Estudos Avançados*, São Paulo, v. 35, n. 101, p. 37-50, 2021. DOI: <https://doi.org/10.1590/s0103-4014.2021.35101.004>. Acesso em: 24 jun. 2025.

SILVA, T. *In: Racismo Algorítmico: inteligência artificial e discriminação nas redes digitais.* São Paulo: Edições Sesc, 2022.

SILVEIRA, S. A. Responsabilidade algorítmica, personalidade eletrônica e democracia. **Revista Eletrônica Internacional de Economia Política da Informação da Comunicação e da Cultura**, São Cristóvão, v. 22, n. 2, p. 83–96, 2020. Disponível em: <https://periodicos.ufs.br/eptic/article/view/12021>. Acesso em: 21 jan. 2025.

SLOTA, D. Daily active users of Deepseek. **Statista**, 2024. Disponível em: <https://www.statista.com/statistics/1561128/deepseek-daily-active-users/>. Acesso em: 01 jun. 2025.

SIMON, F. *et al.* Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. **Harvard Kennedy School (HKS) Misinformation Review**, v. 4, n. 5, 2023. Disponível em: <https://misinfoeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/>. Acesso em: 20 jan. 2025.

SPITALE, G. *et al.* AI model GPT-3 (dis)informs us better than humans. **Science Advances**, v. 9, n. 26, 2023. Disponível em: <https://www.science.org/doi/10.1126/sciadv.adh1850>. Acesso em: 20 jul. 2024.

STANFORD UNIVERSITY Human-Centered Artificial Intelligence. **Artificial Intelligence Index Report 2025**. Stanford, 2025. Disponível em: <https://hai.stanford.edu/ai-index/2025-ai-index-report>. Acesso em: 6 jun. 2025.

TERRA, V., Imagens de IA polarizam e aproximam humor e ódio. **Agência Lupa**, 12 dez. 2024. Disponível em: <https://lupa.uol.com.br/educacao/2024/12/12/imagens-de-ia-polarizam-e-aproximam-humor-e-odio>. Acesso em: 20 jan. 2025.

TRACY, S. J.; HINRICHS, M. M. Big tent criteria for qualitative quality. In: MATTHES, Jörg (ed.). **The International Encyclopedia of Communication Research Methods**. Hoboken: John Wiley & Sons, 2017. Disponível em: <https://doi.org/10.1002/9781118901731.iecrm0016>. Acesso em: 20 jan. 2025.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação.** São Paulo: Atlas, 1987

TURING, A. M. **Computing machinery and intelligence.** *Mind*, v. 59, n. 236, p. 433–460, 1950. Disponível em: <https://doi.org/10.1093/mind/LIX.236.433>. Acesso em: 20 maio 2024.

UNESCO. **Recomendação sobre a ética da inteligência artificial.** Paris: UNESCO, 2022. 43 p. (SHS/BIO/PI/2021/1). Disponível em: [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000381137_por). Acesso em: 21 jan. 2025.

UNESCO. **Synthetic content and its implications for AI policy**: a primer. Paris: UNESCO, 2024. 41 p. Disponível em: <https://doi.org/10.54678/URJF3412>. Acesso em: 21 jan. 2025.

VANIAN, J. Zuckerberg: Meta AI reaches one billion monthly users. **CNBC**, 28 maio 2025. Disponível em: <https://www.cnbc.com/2025/05/28/zuckerberg-meta-ai-one-billion-monthly-users.html>. Acesso em: 01 jun. 2025.

WARDLE, C.; DERAKHSHAN, H. **Desordem informacional**: para um quadro interdisciplinar de investigação e elaboração de políticas públicas. 1. ed. Campinas: UNICAMP, Centro de Lógica, Epistemologia e História da Ciência, 2023. (Coleção CLE, 92).

WESTERLUND, M. The Emergence of Deepfake Technology: A Review. **Technology Innovation Management Review**, Ottawa, v. 9, n. 11, p. 39–52, nov. 2019. Disponível em: <https://timreview.ca/article/1282>. Acesso em: 21 jul. 2024.

WINNER, L. Do artifacts have politics? In: WINNER, Langdon. **The whale and the reactor**: a search for limits in an age of high technology. Chicago: The University of Chicago Press. 1986.

ZUBOFF, S. **The age of surveillance capitalism**: the fight for a human future at the new frontier of power. London: Profile Books, 2019.

**APÊNDICE A – LISTAGEM DE NOTÍCIAS DE CASOS COM USO DE IA  
VERIFICADAS PELA AGÊNCIA LUPA NO PERÍODO DE MARÇO DE 2024 A  
MARÇO DE 2025**

Casos previamente identificados e verificados pela plataforma brasileira Agência Lupa como desinformação gerada por Inteligência Artificial Generativa. Total: 46 casos.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2025/03/24/imagem-de-giovanna-ewbank-e-manipulada-com-ia-para-venda-de-produto-milagroso-">https://lupa.uol.com.br/jornalismo/2025/03/24/imagem-de-giovanna-ewbank-e-manipulada-com-ia-para-venda-de-produto-milagroso-</a>	Imagem de Giovanna Ewbank é manipulada com IA para venda de “produto milagroso”	Giovanna Ewbank	Vídeo deepfake	Instagram e Facebook.	Conteúdo usava imagem alterada da atriz e texto falso atribuindo a ela a venda de produto de emagrecimento milagroso.	Celebridade usada como autoridade falsa; artefato com alta verossimilhança; plataformas não sinalizaram conteúdo como falso. Uso comercial e emocional da imagem.
<a href="https://lupa.uol.com.br/jornalismo/2025/03/21/deepfake-usa-video-de-lula-para-promover-golpe-da-falsa-indenizacao">https://lupa.uol.com.br/jornalismo/2025/03/21/deepfake-usa-video-de-lula-para-promover-golpe-da-falsa-indenizacao</a>	Deepfake usa vídeo de Lula para promover golpe da falsa indenização	Luiz Inácio Lula da Silva	Vídeo deepfake	WhatsApp, Facebook	Golpistas usaram vídeo manipulado do presidente Lula para divulgar falsa indenização de R\$ 15 mil.	Atores maliciosos exploram figura política com alta credibilidade. Plataforma favorece viralização. Ausência de marcação de conteúdo.
<a href="https://lupa.uol.com.br/jornalismo/2025/02/03/deepfake-de-nikolas-ferreira-e-usada-em-golpe-de-falsa-indenizacao">https://lupa.uol.com.br/jornalismo/2025/02/03/deepfake-de-nikolas-ferreira-e-usada-em-golpe-de-falsa-indenizacao</a>	Deepfake de Nikolas Ferreira é usada em golpe de falsa indenização	Nikolas Ferreira	Vídeo <i>deepfake</i>	WhatsApp	Vídeo falso atribui ao deputado a promoção de resgate financeiro automatizado.	Atores e artefatos semelhantes ao caso anterior; repetição do formato de golpe com autoridade política jovem; direcionamento para páginas de captura de dados.
<a href="https://lupa.uol.com.br/jornalismo/2025/01/21/imagem-de-michelle-e-trump-com-placa-bolsonaro-2026-foi-criada-por-ia">https://lupa.uol.com.br/jornalismo/2025/01/21/imagem-de-michelle-e-trump-com-placa-bolsonaro-2026-foi-criada-por-ia</a>	Imagem de Michelle e Trump com placa "Bolsonaro 2026" foi criada por IA	Michelle Bolsonaro, Donald Trump	Imagem manipulada	X (Twitter)	Montagem apresentava os dois segurando placa com mensagem política falsa.	Ação de manipulação simbólica com conotação eleitoral; uso de figuras internacionais e nacionais para mobilização política; artefato altamente compartilhável.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2025/01/20/video-de-nikolas-ferreira-sobre-pix-foi-manipulado-com-ia-para-aplicar-golpe">https://lupa.uol.com.br/jornalismo/2025/01/20/video-de-nikolas-ferreira-sobre-pix-foi-manipulado-com-ia-para-aplicar-golpe</a>	Vídeo de Nikolas Ferreira sobre PIX foi manipulado com IA para aplicar golpe	Nikolas Ferreira	Vídeo deepfake	Facebook	Vídeo atribuído ao deputado declarações sobre nova funcionalidade do PIX e bônus falso do governo.	Padrão de desinformação financeira persistente; plataforma usada para enganar com aparência de veracidade institucional; ausência de checagem nas redes primárias.
<a href="https://lupa.uol.com.br/jornalismo/2025/01/20/deepfake-usa-fatima-bernardes-para-vender-falso-tratamento-ocular">https://lupa.uol.com.br/jornalismo/2025/01/20/deepfake-usa-fatima-bernardes-para-vender-falso-tratamento-ocular</a>	Deepfake usa Fátima Bernardes para vender falso tratamento ocular	Fátima Bernardes	Vídeo deepfake	Facebook	A jornalista é mostrada promovendo produto falso de recuperação de visão.	Uso comercial de autoridade midiática; exploração de confiabilidade da apresentadora; artefato envolvia site falso e estética de propaganda legítima.
<a href="https://lupa.uol.com.br/jornalismo/2025/01/17/video-em-que-advogado-fala-sobre-fgts-foi-manipulado-com-uso-de-ia">https://lupa.uol.com.br/jornalismo/2025/01/17/video-em-que-advogado-fala-sobre-fgts-foi-manipulado-com-uso-de-ia</a>	Vídeo em que advogado fala sobre FGTS foi manipulado com uso de IA	Advogado (não é figura pública)	Vídeo deepfake	TikTok	Vídeo distorce fala de suposto advogado para criar golpe sobre resgate de FGTS.	Artefato genérico explorando suposta autoridade jurídica; contexto financeiro e social para criar urgência e engajamento rápido.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/20/video-de-mulher-que-deu-a-luz-13-filhos-foi-gerado-por-ia">https://lupa.uol.com.br/jornalismo/2024/12/20/video-de-mulher-que-deu-a-luz-13-filhos-foi-gerado-por-ia</a>	Vídeo de mulher que deu à luz 13 filhos foi gerado por IA	Mulher (não é figura pública)	Vídeo deepfake	Facebook, TikTok	Vídeo apresenta falsa narrativa de parto múltiplo com imagens geradas por IA, viralizando com tom emocional.	Narrativa apelativa, sem personagem real; uso de IA para gerar empatia e engajamento; desinformação emocional.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/19/video-nao-traz-musk-anunciando-investimento-de-us-15-milhoes-na-educacao-no-brasil">https://lupa.uol.com.br/jornalismo/2024/12/19/video-nao-traz-musk-anunciando-investimento-de-us-15-milhoes-na-educacao-no-brasil</a>	Vídeo não traz Musk anunciando investimento de US\$ 15 milhões na educação no Brasil	Elon Musk	Vídeo deepfake	Facebook	Fala atribuída a Musk é sintetizada para simular apoio a programa educacional falso no Brasil.	Estratégia de empréstimo de autoridade internacional; artefato técnico sofisticado; vídeo sem marcas de verificação.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/18/drauzio-varella-nao-recomendou-remedio-para-disfuncao-sexual">https://lupa.uol.com.br/jornalismo/2024/12/18/drauzio-varella-nao-recomendou-remedio-para-disfuncao-sexual</a>	Drauzio Varella não recomendou remédio para disfunção sexual –	Drauzio Varella	Vídeo deepfake	Facebook, Instagram	Drauzio aparece recomendando remédio milagroso para disfunção, sem	Uso indevido de figura médica pública; estética publicitária falsa; risco à saúde pública por desinformação.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
funcao-sexual-video-e-deepfake	vídeo é deepfake				nunca ter feito isso.	
<a href="https://lupa.uol.com.br/jornalismo/2024/12/17/e-golpe-anuncio-que-promete-registro-de-cac-em-minutos">https://lupa.uol.com.br/jornalismo/2024/12/17/e-golpe-anuncio-que-promete-registro-de-cac-em-minutos</a>	É golpe: anúncio que promete registro de CAC em minutos	Nenhuma	Site falso com imagem gerada	Facebook, sites de anúncios	Página usa estética oficial e simula facilitação de registro de arma de fogo com promessa irreal.	Uso de site clonado e visual sintético; conteúdo impulsionado por anúncios pagos; tema sensível explorado com IA.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/13/deepfake-com-imagem-de-advogada-e-golpe-para-promover-falsa-indenizacao">https://lupa.uol.com.br/jornalismo/2024/12/13/deepfake-com-imagem-de-advogada-e-golpe-para-promover-falsa-indenizacao</a>	Deepfake com imagem de advogada é golpe para promover falsa indenização	Advogada não identificada (não é figura pública)	Imagem deepfake estática	Facebook, Instagram	Golpistas criaram identidade visual de falsa advogada oferecendo resgates de indenização.	Criação de persona fake com aparência realista; disseminação em redes e sites; engano promovido com visual profissional falso.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/13/drauzio-varella-nao-recomendou-remedio-para-zumbido-no-ouvido-ou-video-e-deepfake">https://lupa.uol.com.br/jornalismo/2024/12/13/drauzio-varella-nao-recomendou-remedio-para-zumbido-no-ouvido-ou-video-e-deepfake</a>	Drauzio Varella não recomendou remédio para zumbido no ouvido – vídeo é deepfake	Drauzio Varella	Vídeo deepfake	YouTube, Facebook	Repetição da estratégia com uso da imagem do médico para vender produto ineficaz.	Exploração sistemática de figura confiável; variação do mesmo golpe; algoritmo das plataformas favorece reprodução.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/13/imagem-de-lula-com-cabeca-enfaixada-foi-manipulada-por-ia">https://lupa.uol.com.br/jornalismo/2024/12/13/imagem-de-lula-com-cabeca-enfaixada-foi-manipulada-por-ia</a>	Imagem de Lula com cabeça enfaixada foi manipulada por IA	Luiz Inácio Lula da Silva	Imagem manipulada	Twitter, grupos de WhatsApp	Imagem do presidente é adulterada para simular agressão ou estado grave de saúde.	Técnica de manipulação para gerar medo, ódio ou confusão; conteúdo circula em bolhas políticas; impacto simbólico relevante.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/13/e-falsa-imagem-com-pedido-de-certidao-de-obito-de-lula">https://lupa.uol.com.br/jornalismo/2024/12/13/e-falsa-imagem-com-pedido-de-certidao-de-obito-de-lula</a>	É falsa imagem com pedido de certidão de óbito de Lula	Luiz Inácio Lula da Silva	Imagem	WhatsApp, Facebook, X	Circula imagem com pedido de certidão de óbito de Lula, sugerindo falsamente que ele teria morrido. A Lupa verificou que se trata de uma montagem sem nenhum	Caso típico de desinformação política baseada em montagem gráfica. Alinha-se a estratégias de difamação e caos informacional. Possível uso de IA para forjar documentos oficiais. Utiliza figura de autoridade para gerar choque e viralização.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
					fundamento.	
<a href="https://lupa.uol.com.br/jornalismo/2025/01/17/video-manipula-imagens-de-nikolas-ferreira-e-cesar-tralli-para-aplicar-golpe">https://lupa.uol.com.br/jornalismo/2025/01/17/video-manipula-imagens-de-nikolas-ferreira-e-cesar-tralli-para-aplicar-golpe</a>	Vídeo manipula imagens de Nikolas Ferreira e Cesar Tralli para aplicar golpe	Nikolas Ferreira, Cesar Tralli	Vídeo deepfake	WhatsApp, Facebook	Montagem combina trechos dos dois para simular apoio político e financeiro a programa falso.	Técnica de fusão visual e narrativa; manipulação dupla de personalidades públicas de esferas distintas para dar verossimilhança ao golpe.
<a href="https://lupa.uol.com.br/jornalismo/2025/01/16/e-falso-que-fernando-haddad-disse-que-gosta-de-taxar-pobres">https://lupa.uol.com.br/jornalismo/2025/01/16/e-falso-que-fernando-haddad-disse-que-gosta-de-taxar-pobres</a>	É falso que Fernando Haddad disse que gosta de taxar pobres	Fernando Haddad	Texto com imagem manipulada	Twitter, grupos de WhatsApp	Frase inventada é atribuída ao ministro com arte gráfica simulando notícia.	Conteúdo apócrifo replicado com fins políticos; IA utilizada na criação visual e textual para reforçar narrativas de desinformação econômica e eleitoral.
<a href="https://lupa.uol.com.br/jornalismo/2025/01/14/video-viral-de-incendios-em-los-angeles-foi-produzido-com-ia">https://lupa.uol.com.br/jornalismo/2025/01/14/video-viral-de-incendios-em-los-angeles-foi-produzido-com-ia</a>	Vídeo viral de incêndios em Los Angeles foi produzido com IA	Evento ambiental	Vídeo gerado artificialmente	TikTok, Instagram	Simula cenas de incêndio em escala catastrófica como se fossem reais, causando pânico.	Conteúdo emocional e sensacionalista; apelo ao medo; artefato difícil de identificar como falso sem verificação técnica.
<a href="https://lupa.uol.com.br/jornalismo/2024/12/06/deepfake-usa-imagem-de-ratinho-em-golpe-de-falsa-oferta-do-mercado-livre">https://lupa.uol.com.br/jornalismo/2024/12/06/deepfake-usa-imagem-de-ratinho-em-golpe-de-falsa-oferta-do-mercado-livre</a>	Deepfake usa imagem de Ratinho em golpe de falsa oferta do Mercado Livre	Ratinho	Vídeo deepfake	YouTube, Facebook	Apresentador aparece promovendo falso cupom de desconto com <i>link</i> fraudulento.	Mistura de confiança midiática com plataforma de comércio eletrônico; prática de phishing com rosto famoso; ausência de moderação prévia.
<a href="https://lupa.uol.com.br/jornalismo/2024/11/07/whatsapp-nao-vazou-dados-video-sobre-indenizacao-de-r-30-mil-e-deepfake">https://lupa.uol.com.br/jornalismo/2024/11/07/whatsapp-nao-vazou-dados-video-sobre-indenizacao-de-r-30-mil-e-deepfake</a>	WhatsApp não vazou dados – vídeo sobre indenização de R\$ 30 mil é deepfake	Nenhuma	Vídeo deepfake	WhatsApp	Vídeo engana usuários com falsa promessa de indenização relacionada a vazamento de dados.	Padrão clássico de golpe em massa via WhatsApp; uso de narrativa jurídica com tom alarmista e simulando canal oficial.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2024/11/06/deepfake-usa-imagem-de-famosos-para-aplicar-golpe-em-de-famosos-para-aplicar-golpe-resgata-brasil">https://lupa.uol.com.br/jornalismo/2024/11/06/deepfake-usa-imagem-de-famosos-para-aplicar-golpe-em-de-famosos-para-aplicar-golpe-resgata-brasil</a>	Deepfake usa imagem de famosos para aplicar golpe “Resgata Brasil”	Vários artistas brasileiros	Vídeo e imagem manipulados	Instagram, Facebook	Falsas falas e imagens de artistas são usadas para dar legitimidade a golpe de resgate de valores.	Multiplicidade de atores e formatos; repetição de estratégia de “autoridade falsa”; conteúdo viral por identificação emocional.
<a href="https://lupa.uol.com.br/jornalismo/2024/10/26/video-sobre-anuncio-de-secretariado-de-candidato-de-pl-em-santos-e-falso-audio-foi-manipulado">https://lupa.uol.com.br/jornalismo/2024/10/26/video-sobre-anuncio-de-secretariado-de-candidato-de-pl-em-santos-e-falso-audio-foi-manipulado</a>	Vídeo sobre anúncio de secretariado de candidata do PL em Santos é falso – áudio foi manipulado	Candidata não identificada	Áudio manipulado com imagem	WhatsApp, Facebook	Montagem atribui à candidata anúncio de cargos inexistentes com interesses eleitorais.	Exemplo de manipulação política localizada; IA usada para alterar contexto eleitoral municipal; desinformação altamente compartilhável em redes de bairro e grupos locais.
<a href="https://lupa.uol.com.br/jornalismo/2024/10/22/deepfake-usa-video-de-luciano-hang-para-promover-golpe-de-resgate-de-valores">https://lupa.uol.com.br/jornalismo/2024/10/22/deepfake-usa-video-de-luciano-hang-para-promover-golpe-de-resgate-de-valores</a>	Deepfake usa vídeo de Luciano Hang para promover golpe de resgate de valores	Luciano Hang	Vídeo deepfake	Todas as redes sociais mencionadas	Empresário aparece anunciando resgate falso de valores com <i>link</i> malicioso.	Autoridade empresarial usada para dar credibilidade; formato recorrente de golpe financeiro; forte circulação entre públicos pró-governo.
<a href="https://lupa.uol.com.br/jornalismo/2024/10/11/e-falso-video-de-elon-musk-falando-sobre-jogo-sujo-no-desbloqueio-do-x">https://lupa.uol.com.br/jornalismo/2024/10/11/e-falso-video-de-elon-musk-falando-sobre-jogo-sujo-no-desbloqueio-do-x</a>	É falso vídeo de Elon Musk falando sobre jogo sujo no desbloqueio do X	Elon Musk	Vídeo deepfake	Todas as redes sociais mencionadas	Vídeo simula denúncia de censura por Musk envolvendo desbloqueios na plataforma X.	Narrativa ideológica de censura; IA aplicada à criação de fala falsa em vídeo legítimo; apelo emocional para público politizado.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2024/10/08/e-falsa-a-imagem-de-elon-musk-fazendo-o-m-de-marcal-imagem-foi-gerada-por-ia">https://lupa.uol.com.br/jornalismo/2024/10/08/e-falsa-a-imagem-de-elon-musk-fazendo-o-m-de-marcal-imagem-foi-gerada-por-ia</a>	É falsa a imagem de Elon Musk fazendo o “M” de Marçal – imagem foi gerada por IA	Elon Musk	Imagem manipulada	Twitter	Foto artificial mostra Musk com gesto associado a pré-candidato brasileiro.	Montagem visual como forma de endosso político; IA como recurso de propaganda simbólica internacional; desinformação simbólica.
<a href="https://lupa.uol.com.br/jornalismo/2024/10/01/datenavira-alvo-de-deepfake-video-usa-ia-para-mostrar-suposto-caso-de-assedio-sexual">https://lupa.uol.com.br/jornalismo/2024/10/01/datenavira-alvo-de-deepfake-video-usa-ia-para-mostrar-suposto-caso-de-assedio-sexual</a>	Datena vira alvo de deepfake – vídeo usa IA para mostrar suposto caso de assédio sexual	José Luiz Datena	Vídeo deepfake	Instagram, grupos privados	Jornalista aparece em situação falsa de assédio, com edição e voz simuladas.	Grave desinformação com risco à reputação; artefato visual e sonoro; potencial de dano à imagem pública e judicialização.
<a href="https://lupa.uol.com.br/jornalismo/2024/09/18/jair-bolsonaro-nao-declarou-apoio-a-pablo-marcal-video-foi-feito-com-ia">https://lupa.uol.com.br/jornalismo/2024/09/18/jair-bolsonaro-nao-declarou-apoio-a-pablo-marcal-video-foi-feito-com-ia</a>	Jair Bolsonaro não declarou apoio a Pablo Marçal – vídeo foi feito com IA	Jair Bolsonaro	Vídeo deepfake	WhatsApp, X (Twitter)	Bolsonaro aparece endossando candidatura em vídeo que nunca existiu.	Conteúdo eleitoral fraudulento; criação de apoio falso para influenciar decisão política; IA como motor de manipulação eleitoral direta.
<a href="https://lupa.uol.com.br/jornalismo/2024/09/12/e-deepfake-video-em-que-lula-e-janja-fazem-comentarios-de-teor-sexual">https://lupa.uol.com.br/jornalismo/2024/09/12/e-deepfake-video-em-que-lula-e-janja-fazem-comentarios-de-teor-sexual</a>	É deepfake vídeo em que Lula e Janja fazem comentários de teor sexual	Lula e Janja	Vídeo deepfake	WhatsApp, redes alternativas	Casal presidencial é alvo de montagem com insinuações sexuais.	Desinformação baseada em escândalo fabricado; IA utilizada para afetar imagem pública e promover discurso de ódio ou ridicularização.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2024/09/11/bolsonaro-e-kid-bengala-nao-trocaram-selinho-vidео-foi-criado-por-ia">https://lupa.uol.com.br/jornalismo/2024/09/11/bolsonaro-e-kid-bengala-nao-trocaram-selinho-vidео-foi-criado-por-ia</a>	Bolsonaro e Kid Bengala não trocaram selinho – vídeo foi criado por IA	Jair Bolsonaro, Kid Bengala	Vídeo deepfake humorístico	Grupos de WhatsApp	Montagem grotesca usada para viralização cômica ou ofensiva com viés político.	Deepfake com intenção humorística, mas com viés de ridicularização; debate entre sátira e desinformação política; difícil regulação nas plataformas.
<a href="https://lupa.uol.com.br/jornalismo/2024/09/05/deepfake-usa-imagem-de-pablo-marçal-para-promover-golpe-indeniza-brasil">https://lupa.uol.com.br/jornalismo/2024/09/05/deepfake-usa-imagem-de-pablo-marçal-para-promover-golpe-indeniza-brasil</a>	Deepfake usa imagem de Pablo Marçal para promover golpe “Indeniza Brasil”	Pablo Marçal	Imagem e vídeo manipulados	Instagram, Facebook	Candidato aparece promovendo falso programa de indenização com <i>link</i> fraudulento.	Uso eleitoral e financeiro simultâneo; IA aplicada para aumentar autoridade e engajamento; combinação de elementos visuais e narrativos sintéticos.
<a href="https://lupa.uol.com.br/jornalismo/2024/09/03/é-golpe-site-que-oferece-resgate-de-valores-para-usuários-de-cartão-de-crédito-ou-débito">https://lupa.uol.com.br/jornalismo/2024/09/03/é-golpe-site-que-oferece-resgate-de-valores-para-usuários-de-cartão-de-crédito-ou-débito</a>	É golpe site que oferece resgate de valores para usuários de cartão de crédito ou débito	Celso Russomanno	Vídeo deepfake	WhatsApp, anúncios patrocinados na Meta	Anúncio de falso portal simula resgate financeiro via cartão para roubar dados bancários.	Engenharia social com vídeo IA, visual e textual; uso de publicidade para impulsionar golpe; ausência de moderação prévia.
<a href="https://lupa.uol.com.br/jornalismo/2024/09/02/video-de-lula-e-alexandre-de-moraes-se-beijando-e-falso-e-foi-criado-com-ia">https://lupa.uol.com.br/jornalismo/2024/09/02/video-de-lula-e-alexandre-de-moraes-se-beijando-e-falso-e-foi-criado-com-ia</a>	Vídeo de Lula e Alexandre de Moraes se beijando é falso e foi criado com IA	Lula, Alexandre de Moraes	Vídeo deepfake	WhatsApp, X (Twitter)	Conteúdo simula situação íntima entre autoridades com viés político-satírico.	Humor ácido com intenção de humilhação pública; IA como recurso de ataque simbólico; fronteira entre sátira e desinformação.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2024/08/20/videos-manipulam-imagens-de-famosos-para-promover-medidor-de-glicose-a-laser-ineficaz">https://lupa.uol.com.br/jornalismo/2024/08/20/videos-manipulam-imagens-de-famosos-para-promover-medidor-de-glicose-a-laser-ineficaz</a>	Vídeos manipulam imagens de famosos para promover medidor de glicose a laser ineficaz	Celebridades não identificadas	Vídeos manipulados	YouTube, Facebook, Instagram	Personalidades supostamente endossam produto médico falso para diabetes.	Risco à saúde pública; exploração de confiança em figuras públicas; ausência de marcação e controle por parte das plataformas.
<a href="https://lupa.uol.com.br/jornalismo/2024/08/14/e-falso-que-elon-musk-anunciou-sua-primeira-namorada-roboto">https://lupa.uol.com.br/jornalismo/2024/08/14/e-falso-que-elon-musk-anunciou-sua-primeira-namorada-roboto</a>	É falso que Elon Musk anunciou sua primeira namorada robô	Elon Musk	Imagem e texto manipulados	Sites de curiosidades, Twitter	Publicações simulam anúncio sensacionalista com uso de IA visual.	Mistura de ficção com desinformação; apelo sensacionalista e viral; difícil de moderar por parecer sátira.
<a href="https://lupa.uol.com.br/jornalismo/2024/08/02/e-falso-que-alto-comando-militar-reconheceu-que-Maduro-perdeu-eleicoes">https://lupa.uol.com.br/jornalismo/2024/08/02/e-falso-que-alto-comando-militar-reconheceu-que-Maduro-perdeu-eleicoes</a>	É falso que alto comando militar reconheceu que Maduro perdeu eleições	Nicolás Maduro, militares venezuelanos	Texto com montagem de documento	WhatsApp, grupos políticos	Montagem simula documento militar sobre derrota eleitoral de Maduro.	Estratégia política transnacional; IA usada para criar aparência de prova oficial; circulação em redes polarizadas.
<a href="https://lupa.uol.com.br/jornalismo/2024/07/26/video-de-bolsonaro-feito-por-ia-mente-que-pf-negou-venda-de-joias-nos-eua">https://lupa.uol.com.br/jornalismo/2024/07/26/video-de-bolsonaro-feito-por-ia-mente-que-pf-negou-venda-de-joias-nos-eua</a>	Vídeo de Bolsonaro feito por IA mente que PF negou venda de joias nos EUA	Jair Bolsonaro	Video deepfake	WhatsApp, X (Twitter)	Ex-presidente aparece negando fatos sobre investigação oficial.	Desinformação de defesa política; uso de IA para tentar restaurar imagem pública; relação direta com tema jurídico de interesse nacional.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2024/06/11/imagem-de-corpos-boiando-em-canoas-rs-foi-manipulada-com-uso-de-ia">https://lupa.uol.com.br/jornalismo/2024/06/11/imagem-de-corpos-boiando-em-canoas-rs-foi-manipulada-com-uso-de-ia</a>	Imagem de corpos boiando em Canoas (RS) foi manipulada com uso de IA	Nenhuma figura pública	Imagem manipulada	Twitter, Facebook	Fotografia falsa simula corpos em enchente para gerar comoção e viralização.	Uso de tragédia para criar desinformação emocional; IA como ferramenta de impacto visual; conteúdo amplamente compartilhado em contextos de crise.
<a href="https://lupa.uol.com.br/jornalismo/2024/06/10/videos-manipulados-por-ia-promovem-golpe-do-dinheiro-acumulado-no-cpf">https://lupa.uol.com.br/jornalismo/2024/06/10/videos-manipulados-por-ia-promovem-golpe-do-dinheiro-acumulado-no-cpf</a>	Vídeos manipulados por IA promovem golpe do dinheiro acumulado no CPF	Luiz Inácio Lula da Silva, Fernando Haddad, Luciano Hang, Michel Temer (MDB), Marina Silva (REDE)	Vídeo deepfake	Instagram e Facebook.	Figuras explicam falsa oportunidade de resgate de valores do CPF.	Golpe padronizado com estética institucional; IA reforça tom de credibilidade; ação fraudulenta com narrativa pública falsa.
<a href="https://lupa.uol.com.br/jornalismo/2024/05/21/e-falsa-imagem-viral-de-homem-resgatando-bebe-em-rua-alagada-no-rs">https://lupa.uol.com.br/jornalismo/2024/05/21/e-falsa-imagem-viral-de-homem-resgatando-bebe-em-rua-alagada-no-rs</a>	É falsa imagem viral de homem resgatando bebê em rua alagada no RS	Homem (não é figura pública)	Imagem gerada artificialmente	WhatsApp	Cena emocionante de resgate foi inteiramente fabricada com IA.	Exploração emocional durante tragédia; desinformação como mobilizador afetivo; imagem amplamente viral.
<a href="https://lupa.uol.com.br/jornalismo/2024/05/16/e-falsa-a-imagem-viral-que-mostra-helicoptero-da-havan-em-resgate-no-rs">https://lupa.uol.com.br/jornalismo/2024/05/16/e-falsa-a-imagem-viral-que-mostra-helicoptero-da-havan-em-resgate-no-rs</a>	É falsa a imagem viral que mostra helicóptero da Havan em resgate no RS	Empresa Havan (instituição)	Imagem manipulada	Facebook, grupos pró-empresa	Fotografia forjada com helicóptero adesivado da empresa simulando ajuda no RS.	Reputação corporativa manipulada via IA; simbologia empresarial utilizada em contexto de calamidade; engajamento emocional e político.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2024/05/16/golpistas-usam-nome-da-havan-para-vender-produtos-em-prol-do-rs">https://lupa.uol.com.br/jornalismo/2024/05/16/golpistas-usam-nome-da-havan-para-vender-produtos-em-prol-do-rs</a>	Golpistas usam nome da Havan para vender produtos em prol do RS	Empresa Havan (instituição)	Sites falsos + imagens manipuladas	Facebook	Criação de site falso simulando parceria da Havan para vender produtos e arrecadar fundos.	Engenharia social empresarial; IA usada para falsificar identidade visual e capturar confiança; golpe oportunista em contexto humanitário.
<a href="https://lupa.uol.com.br/jornalismo/2024/05/02/e-falso-que-elon-musk-anunciou-sinal-gratuito-de-internet-para-o-brasil">https://lupa.uol.com.br/jornalismo/2024/05/02/e-falso-que-elon-musk-anunciou-sinal-gratuito-de-internet-para-o-brasil</a>	É falso que Elon Musk anunciou sinal gratuito de internet para o Brasil	Elon Musk	Texto e imagem manipulados	Facebook	Publicação falsa sugere que Musk anunciou internet grátis para brasileiros.	Mistura de ficção tecnológica com desejo popular; desinformação atrativa e compartilhável; baixa verificação entre usuários.
<a href="https://lupa.uol.com.br/jornalismo/2024/04/15/imagem-de-elon-musk-segurando-bandeira-do-brasil-foi-gerada-por-ia">https://lupa.uol.com.br/jornalismo/2024/04/15/imagem-de-elon-musk-segurando-bandeira-do-brasil-foi-gerada-por-ia</a>	Imagem de Elon Musk segurando bandeira do Brasil foi gerada por IA	Elon Musk	Imagem manipulada	Instagram, Twitter	Imagem simula apoio do empresário ao Brasil em contexto político.	Imagem nacionalista construída com IA; desinformação visual como arma de convencimento ideológico; alta viralização.
<a href="https://lupa.uol.com.br/jornalismo/2024/04/04/e-falso-que-mcdonalds-criou-lanche-do-diabo-que-acompanha-esttua-de-baphomet">https://lupa.uol.com.br/jornalismo/2024/04/04/e-falso-que-mcdonalds-criou-lanche-do-diabo-que-acompanha-esttua-de-baphomet</a>	É falso que McDonald's criou lanche do diabo que acompanha estátua de Baphomet	McDonald's (marca citada)	Imagem e texto manipulados	Facebook	Montagem mostra sanduíche temático com simbologia satânica.	Desinformação com apelo religioso e moral; IA usada para criar imagem ofensiva com fins ideológicos ou conspiratórios.
<a href="https://lupa.uol.com.br/jornalismo/2024/04/02/videos-manipulados-por-ia-promovem-falso-tratamento-contradiabetes">https://lupa.uol.com.br/jornalismo/2024/04/02/videos-manipulados-por-ia-promovem-falso-tratamento-contradiabetes</a>	Vídeos manipulados por IA promovem falso tratamento contra diabetes	ator José Loreto, o cantor Zeca Pagodinho e o médico Drauzio Varella	Vídeo deepfake	Instagram e Youtube	Vídeos apresentam produto inexistente com discurso médico manipulado.	Risco à saúde pública por desinformação médica; IA usada para simular fala médica; estética de credibilidade sem base científica.

<b>Link / Fonte</b>	<b>Título da checagem</b>	<b>Figura pública envolvida</b>	<b>Formato do conteúdo sintético</b>	<b>Plataforma de circulação</b>	<b>Resumo do conteúdo verificado</b>	<b>Observações</b>
<a href="https://lupa.uol.com.br/jornalismo/2024/03/15/e-golpe-anuncio-no-facebook-com-indenizacao-de-r-30-mil-da-serasa">https://lupa.uol.com.br/jornalismo/2024/03/15/e-golpe-anuncio-no-facebook-com-indenizacao-de-r-30-mil-da-serasa</a>	Deepfake usa vídeo de Lula para promover golpe da falsa indenização	Luiz Inácio Lula da Silva, Serasa (instituição mencionada)	Vídeo deepfake	Facebook	Lula é mostrado em vídeo falso promovendo uma falsa indenização.	Alta capacidade de engajamento pela imagem presidencial; IA usada com foco financeiro; golpe replicado com estética governamental.

## APÊNDICE B – DISPOSITIVOS ANALISADOS

Listagem dos trechos destacados nas políticas institucionais de plataformas digitais como Meta, Google, TikTok e X e o Projeto de Lei nº 2338/2023.

Plataforma/ Instituição	Documento analisado	Trecho destacado	Elemento do regime (ação, artefato, ator ou dispositivo)	Tipo de abordagem (restritiva, permissiva, omissa)	Observações analíticas
Google	Política de conteúdo sobre IA no YouTube e em anúncios	<p>“Os criadores têm de incluir divulgações quando o conteúdo [...] faz com que uma pessoa real pareça dizer ou fazer algo que nunca fez, altera as filmagens de um evento ou um local real ou gera uma cena realista que nunca aconteceu na realidade.”</p> <p>“Os anunciantes têm de divulgar os anúncios eleitorais com conteúdo sintético ou alterado digitalmente que represente de forma não autêntica pessoas ou eventos reais ou realistas.”</p>	Dispositivo	Permissiva	Para o YouTube, "exige que os criadores incluam divulgações", colocando o ônus da rotulagem no produtor do conteúdo. A fiscalização e a iniciativa dependem do criador. A política para anúncios só proíbe conteúdo sintético se ele for "sexualmente explícito ou contiver nudez".
Meta	Política de anúncios com IA (2025)	“Começaremos a adicionar rótulos de ‘informações de IA’ a uma gama mais ampla de conteúdo de vídeo, áudio e imagem quando detectarmos indicadores de imagem de IA padrão do setor ou quando as pessoas revelarem que estão enviando conteúdo gerado por IA.”	Dispositivo / Ação	Permissiva	O dispositivo limita-se a inserir advertências contextuais (rótulos), em vez de bloquear ou moderar efetivamente a desinformação sintética. Isso reforça a ideia de que os dispositivos de controle corporativos funcionam mais como medidas simbólicas do que como barreiras eficazes, em

					consonância com sua análise do regime informacional.
TikTok	Termos de uso – IA generativa	“Os seguintes tipos de conteúdo gerado por IA que contêm a semelhança (visual ou sonora) de uma pessoa real ou fictícia não são permitidos (...): Uma figura pública quando usada para apoio político ou comercial. Uma figura privada.”	Dispositivo / Artefato	Restritiva	Regras claras, mas sem detalhamento sobre mecanismos de detecção
X	Política de manipulação > Mídia Sintética e Manipulada	“Poderemos utilizar a nossa própria tecnologia ou receber denúncias através de parcerias com terceiros, a fim de determinar se os meios de comunicação foram manipulados ou apresentados fora de contexto. Em situações em que não formos capazes de determinar com segurança se a mídia está enganando, não poderemos tomar medidas.”	Dispositivo/Ação	Permissiva	Na incerteza (que é comum com a sofisticação da IA generativa), o X não agirá.
Senado Federal	PL nº 2338/2023	Artigo 8º: “Art. 8º A supervisão humana de sistemas de IA de alto risco buscará prevenir ou minimizar os riscos para direitos e liberdades das pessoas ou grupos afetados que possam decorrer de seu uso normal ou de seu uso em condições de utilização indevida razoavelmente previsíveis, viabilizando que as pessoas responsáveis pela supervisão humana possam, nos termos de regulamento, compreender, interpretar, decidir e intervir nos	Dispositivo / Ação	Restritiva	Regulação ainda em tramitação; falta previsão operacional de aplicação

		<p>sistemas de IA, bem como priorizar o gerenciamento de riscos e impactos irreversíveis.”</p> <p>Artigo 6º, inciso II (você inseriu como complemento):          “A pessoa ou grupo afetado por sistema de IA de alto risco tem os seguintes direitos: (...) II – direito de contestar e de solicitar a revisão de decisões, recomendações ou previsões de sistema de IA.”</p> <p>Artigo 19º:          “Quando o sistema de IA gerar conteúdo sintético, deverá, considerando o estado da arte do desenvolvimento tecnológico e o contexto de uso, incluir identificador em tais conteúdos para verificação de autenticidade ou de características de sua proveniência, modificações ou transmissão, conforme regulamento.”</p>			
--	--	---	--	--	--