

## **Regras e ruídos: o impacto das diretrizes de moderação de Instagram, Facebook e X na disseminação da desinformação no início de 2025<sup>1</sup>**

Universidade Tuiuti do Paraná, Curitiba, PR

### **Resumo**

Este estudo analisa o impacto das diretrizes de moderação das plataformas Instagram, Facebook e X na disseminação da desinformação no Brasil no início de 2025. Justifica-se pela relevância social e acadêmica do tema, abordando ambiguidades dessas políticas e suas implicações democráticas e sociais. O objetivo é examinar a eficácia das estratégias automatizadas e humanas adotadas por essas plataformas, por meio de análise crítica dos documentos oficiais chamados Padrões da Comunidade, além da revisão bibliográfica sobre o tema. Realiza-se também um estudo de caso específico sobre transferências financeiras digitais em janeiro de 2025. Os resultados indicam eficácia parcial nas estratégias híbridas adotadas pela Meta, com limitações práticas evidenciadas pelo caso analisado. A plataforma X, baseada em regras mais flexíveis, demonstrou fragilidade significativa frente à circulação de informações falsas. Conclui-se pela necessidade urgente de políticas transparentes, claras e culturalmente adaptadas, integrando automação e intervenção humana, para equilíbrio entre liberdade de expressão e segurança informacional.

**Palavras-chave:** desinformação; redes sociais; moderação

### **Corpo do trabalho**

A desinformação é um fenômeno que tem suas raízes em práticas históricas de manipulação da informação, mas que encontra no ambiente digital um que favorece sua disseminação. Pascual Serrano (2010) argumenta que a desinformação não se limita à mentira explícita, mas abrange um conjunto de ações que descontextualizam, fragmentam e manipulam informações para criar cenários intencionalmente enganosos. Esse processo, muitas vezes, apaga a historicidade dos fatos e distorce a realidade, configurando uma prática que vai além do falso: trata-se de apresentar verdades parciais ou distorcidas como se fossem absolutas.

---

<sup>1</sup> Trabalho apresentado na Sessão Temática Organizações, comunicação e consumo na sociedade de risco, atividade integrante do XIX Congresso Brasileiro Científico de Comunicação Organizacional e Relações Públicas.

A noção de desinformação, que se manifesta de forma contemporânea no meio digital, remonta a períodos históricos como a Guerra Fria, em que o termo disinformation foi cunhado para descrever o vazamento proposital de informações enganosas (Zattar, 2020). A desinformação, potencializada pelas redes digitais, torna-se um problema informativo, ético e cultural, afetando profundamente a formação de opiniões e identidades sociais (Brisola, 2017).

No passado os meios de comunicação de massa eram os principais vetores de manipulação informativa (Serrano, 2010); hoje, plataformas digitais como Instagram, Facebook e X potencializam a disseminação de conteúdos por meio de algoritmos que priorizam o engajamento. Segundo Teixeira (2018), a internet ampliou o papel de criador e disseminador de conteúdos para qualquer usuário, intensificando a propagação de fake news e conferindo nova potência ao fenômeno. Essa amplificação ocorre porque o ambiente digital, caracterizado por redes de inter-relações entre atores sociais e dispositivos técnicos (Santaella; Lemos, 2010), gera fluxos descentralizados. Como observa Recuero (2009), as redes sociais são tanto meios de conexão entre indivíduos quanto sistemas que estruturam relações sociais por meio da comunicação mediada por computador. Esse cenário gera o que Serrano (2010) denomina de "superinformação", um excesso de conteúdos que, em vez de ampliar a consciência crítica, dificulta o acesso a informações relevantes e sobrecarrega os usuários.

Diante do impacto das redes sociais na disseminação da desinformação, as políticas de moderação, representadas pelos "Padrões da Comunidade", surgem como uma tentativa de regular as interações e os conteúdos nas plataformas. Essas diretrizes enfrentam desafios para equilibrar a liberdade de expressão com a responsabilidade de mitigar danos sociais. Segundo Moisés de Lemos Martins (2017), vivemos uma crise de cultura e identidade em que a ética e a verdade são muitas vezes secundarizadas pela hegemonia do pensamento uniformizado. A moderação de conteúdos reflete interesses institucionais e, frequentemente, ambivalências em sua aplicação, conforme ressalta Lévy (1999).

Em 2025, a desinformação é marcada pela crescente centralidade das redes sociais no fluxo de informações e pela complexidade de suas políticas de moderação. Mudanças nas diretrizes de plataformas como Instagram, Facebook e X (antigo Twitter)

suscitaram debates globais. Recentemente, a Meta anunciou o fim da checagem independente de fatos em suas plataformas, optando por um modelo semelhante às "notas da comunidade" do X, em que os próprios usuários contribuem para adicionar contexto às publicações (G1, 2025). Embora a medida tenha sido justificada pelo CEO Mark Zuckerberg como uma forma de mitigar o viés político dos moderadores profissionais, especialistas apontam que tal mudança pode enfraquecer o combate à desinformação, favorecendo a circulação de conteúdos enganosos.

O avanço tecnológico e a sofisticação dos algoritmos de curadoria agravam os obstáculos à moderação. Como observa Recuero (2009), as redes sociais são espaços dinâmicos onde fluxos informacionais são moldados por interações humanas e estruturas técnicas, tornando a desinformação um problema estrutural difícil de conter. Nesse contexto, plataformas buscam promover a liberdade de expressão, mas a ausência de políticas robustas e imparciais compromete a qualidade do ambiente informacional.

As mudanças nas diretrizes de moderação destacam a relevância do tema. Este estudo propõe investigar como os "Padrões da Comunidade" de Instagram, Facebook e X influenciam a circulação de informações enganosas, considerando as implicações sociais, políticas e tecnológicas deste fenômeno.

Para a elaboração deste estudo, foi realizada uma análise crítica dos documentos oficiais – os "Padrões da Comunidade" – por meio de revisão bibliográfica e estudo de caso, sem qualquer alteração na redação original dos textos analisados. Além disso, foram coletados dados provenientes de publicações oficiais, notícias e relatórios que documentam as mudanças nas políticas de moderação das plataformas. Essa abordagem permitiu identificar as discrepâncias entre as diretrizes teóricas e sua aplicação prática.

A análise dos documentos foi realizada utilizando, especificamente, a análise de conteúdo e a análise crítica do discurso, métodos que possibilitaram uma compreensão aprofundada das nuances presentes nos 'Padrões da Comunidade'. Quanto ao estudo de caso, os dados referentes às transferências PIX ocorridas em janeiro de 2025 foram coletados a partir de registros oficiais e notícias, tendo o critério de seleção baseado na relevância e representatividade desse caso para ilustrar as vulnerabilidades na

moderação de conteúdo. A validação dos dados deu-se por meio da triangulação com a revisão bibliográfica e da verificação cruzada com fontes oficiais e acadêmicas.

Observou-se que, enquanto a Meta demonstra eficácia parcial em sua estratégia híbrida – integrando métodos automatizados e intervenção humana –, com limitações práticas evidenciadas pelo estudo de caso sobre transferências financeiras digitais em janeiro de 2025, a plataforma X, fundamentada em regras mais flexíveis, apresentou vulnerabilidade significativa na contenção da disseminação de informações falsas. Tais constatações foram corroboradas pela análise dos dados coletados, que revelaram inconsistências na execução das políticas e a necessidade de revisão dos critérios de moderação.

A integração dos dados permitiu, ainda, a identificação de fatores contextuais e culturais que influenciam a eficácia dos mecanismos de moderação. Dessa forma, a análise enfatiza que a adaptação das estratégias de moderação às especificidades regionais e ao contexto de rápida evolução tecnológica é imprescindível para aprimorar os resultados e garantir um ambiente informacional equilibrado.

Conclui-se que as políticas de moderação adotadas pelas plataformas digitais, embora essenciais para mitigar a disseminação de desinformação, apresentam lacunas importantes em sua aplicação prática. A análise dos “Padrões da Comunidade” revelou inconsistências na forma como as diretrizes são implementadas, evidenciando a necessidade de políticas mais transparentes e adaptadas às realidades locais. O estudo destaca que a combinação de estratégias automatizadas com intervenção humana – de forma integrada e sensível às especificidades culturais – é fundamental para superar as limitações observadas.

Os resultados apontam para uma urgência na reformulação dos processos de moderação, de modo a fortalecer a segurança informacional sem comprometer a liberdade de expressão. Em síntese, a ampliação dos mecanismos de controle e a revisão dos critérios de moderação emergem como medidas necessárias para enfrentar os desafios impostos pela rápida propagação de informações falsas nas redes sociais.

## Referências

BRISOLA, A.; BEZERRA, A. C. Desinformação e circulação de “fake news”: Distinções, diagnóstico e reação. Enancib, 2018. Disponível em:  
[http://enancib.marilia.unesp.br/index.php/XIX\\_ENANCIB/xixenancib/paper/view/1219/1636](http://enancib.marilia.unesp.br/index.php/XIX_ENANCIB/xixenancib/paper/view/1219/1636). Acesso em: 27 jan. 2025.

LÉVY, P. Cibercultura. Tradução de Carlos Irineu da Costa. São Paulo: Editora 34, 1999.

RECUERO, R. Redes Sociais na Internet. Porto Alegre: Sulina, 2009. (Coleção Cibercultura).

SANTAELLA, L.; LEMOS, R. Redes sociais digitais: a cognição conectiva do Twitter. São Paulo: Paulus, 2010. (Coleção Comunicação).

SERRANO, P. Desinformação: como os meios de comunicação ocultam o mundo. Rio de Janeiro: Espalhafato, 2010.

TEIXEIRA, Adriana. Fake news contra a vida: desinformação ameaça vacinação de combate à febre amarela. 2018. 97 f. Dissertação (Mestrado em Comunicação e Semiótica) – Pontifícia Universidade Católica, São Paulo, 2018.

ZATTAR, Mariana. Competência em Informação e Desinfodemia no contexto da pandemia de Covid-19. Liinc em Revista, Rio de Janeiro, v. 16, n. 2, e5391, 2020. DOI:  
<https://doi.org/10.18617/liinc.v16i2.5391>. Disponível em:  
<http://revista.ibict.br/liinc/article/view/5391>. Acesso em: 20 jan. 2025.

FORNI, João José. Gestão de Crises e Comunicação. São Paulo: Atlas, 2019.