

## **INFLUÊNCIA DIGITAL TÓXICA: A CONSTRUÇÃO DE AUTORIDADE ARTIFICIAL VIA FAKE NEWS NO INSTAGRAM**

## **TOXIC DIGITAL INFLUENCE: BUILDING ARTIFICIAL AUTHORITY THROUGH FAKE NEWS ON INSTAGRAM**

## **INFLUENCIA DIGITAL TÓXICA: CONSTRUYENDO AUTORIDAD ARTIFICIAL MEDIANTE NOTICIAS FALSAS EN INSTAGRAM**



10.56238/edimpacto2025.090-072

**Tadeu Marcos Borges Paes**

Doutorando em Inteligência Artificial

Instituição: Senai Centro Desenvolvimento da Amazônia

E-mail: tadeu.paes@live.com

Orcid: <https://orcid.org/0009-0002-2978-2117>

**Michelly Moreira Barbosa**

Especialista em Gestão de Empresas

Instituição: Social Connect

E-mail: michelly.moreira@negint.com.br

Orcid: <https://orcid.org/0009-0007-1578-0889>

---

### **RESUMO**

A ascensão do Instagram como um ecossistema de influência transformou a dinâmica informational e ampliou a capacidade de atores digitais construírem autoridade artificial. Este artigo analisa como perfis — de micro-influenciadores a grandes criadores — utilizam estratégias de desinformação, manipulação emocional e estética de credibilidade para fabricar legitimidade perante seus seguidores. A pesquisa discute como a lógica algorítmica da plataforma favorece conteúdos polarizantes, sensacionalistas ou parcialmente falsos, acelerando sua viralização e reforçando vieses cognitivos dos usuários. A partir de revisão teórica e análise crítica de padrões comunicacionais, demonstra-se que a autoridade artificial produz impactos diretos na formação de opinião, na tomada de decisão individual e nos mercados digitais, influenciando desde comportamentos de compra até percepções políticas. Ao investigar esta poder ilusório, o estudo aponta riscos sociais, econômicos e psicológicos, além de propor caminhos para mitigação, educação midiática e aprimoramento das diretrizes algorítmicas das plataformas.

**Palavras-chave:** Instagram. Fake News. Autoridade Artificial. Influência Digital. Algoritmos. Desinformação. Economia da Atenção.

### **ABSTRACT**

The rise of Instagram as an ecosystem of influence has transformed informational dynamics and expanded the ability of digital actors to build artificial authority. This article analyzes how profiles—from micro-influencers to large creators—use strategies of disinformation, emotional manipulation, and aesthetic credibility to fabricate legitimacy among their followers. The research discusses how the



platform's algorithmic logic favors polarizing, sensationalist, or partially false content, accelerating its viral spread and reinforcing users' cognitive biases. Based on a theoretical review and critical analysis of communication patterns, it demonstrates that artificial authority has direct impacts on opinion formation, individual decision-making, and digital markets, influencing everything from purchasing behavior to political perceptions. By investigating this illusory power, the study points to social, economic, and psychological risks, as well as proposing ways to mitigate them, media literacy, and improve the algorithmic guidelines of the platforms.

**Keywords:** Instagram. Fake News. Artificial Authority. Digital Influence. Algorithms. Disinformation. Attention Economy.

## RESUMEN

El auge de Instagram como ecosistema de influencia ha transformado la dinámica informativa y ha ampliado la capacidad de los actores digitales para construir autoridad artificial. Este artículo analiza cómo los perfiles —desde microinfluencers hasta grandes creadores— utilizan estrategias de desinformación, manipulación emocional y credibilidad estética para generar legitimidad entre sus seguidores. La investigación analiza cómo la lógica algorítmica de la plataforma favorece el contenido polarizador, sensacionalista o parcialmente falso, acelerando su propagación viral y reforzando los sesgos cognitivos de los usuarios. Basándose en una revisión teórica y un análisis crítico de los patrones de comunicación, se demuestra que la autoridad artificial tiene un impacto directo en la formación de opinión, la toma de decisiones individuales y los mercados digitales, influyendo en todo, desde el comportamiento de compra hasta las percepciones políticas. Al investigar este poder ilusorio, el estudio señala los riesgos sociales, económicos y psicológicos, además de proponer maneras de mitigarlos, alfabetización mediática y mejorar las directrices algorítmicas de las plataformas.

**Palabras clave:** Instagram. Noticias Falsas. Autoridad Artificial. Influencia Digital. Algoritmos. Desinformación. Economía de la Atención.



## 1 INTRODUÇÃO

A influência digital tornou-se uma das forças mais poderosas da sociedade contemporânea. O Instagram, com sua arquitetura centrada na visibilidade e no engajamento, consolidou-se como uma plataforma onde a percepção de autoridade muitas vezes vale mais do que competência, formação ou veracidade. Nesse ambiente, a distinção entre informação e manipulação torna-se turva, e a ascensão de perfis baseados em fake news inaugura um fenômeno preocupante: a construção de autoridade artificial.

Ao contrário da autoridade tradicional, historicamente fundamentada em reputação, expertise e validação institucional, a autoridade artificial emerge de métricas superficiais — curtidas, compartilhamentos, comentários e estética profissional — manipuladas estrategicamente por criadores de conteúdo, robôs, empresas de marketing e redes organizadas de desinformação. A lógica algorítmica do Instagram amplifica esses atores ao privilegiar conteúdos emocionalmente carregados, polarizados e viralizáveis, independentemente de sua veracidade.

Esse processo produz o que chamamos de “influência digital tóxica” - IDTOX: uma forma de persuasão que se sustenta na aparência de credibilidade, mas opera a partir de distorções informacionais e manipulações psicológicas. Fake news, meias verdades, dados descontextualizados e narrativas conspiratórias tornam-se ferramentas para capturar atenção, moldar comportamentos e direcionar decisões — desde consumo até posicionamentos políticos.

O problema não reside apenas na circulação de informações falsas, mas na formação de um ecossistema que recompensa quem engana com eficiência. Usuários — especialmente jovens e indivíduos com baixa literacia midiática — tornam-se vulneráveis a decisões precipitadas, compras impulsivas, crenças distorcidas e movimentos sociais artificiais. A autoridade criada pelo design da plataforma substitui a autoridade construída pelo rigor do conhecimento.

Assim, compreender como se fabrica essa legitimidade ilusória e quais mecanismos sustentam a influência tóxica no Instagram é essencial para analisar os impactos sociais e propor estratégias de mitigação. Este estudo examina as engrenagens psicológicas, algorítmicas e comunicacionais que fortalecem essa forma de poder digital, discutindo suas consequências para a sociedade, para o marketing digital e para a democracia informacional.

## 2 METODOLOGIA

A metodologia adotada neste estudo segue um percurso investigativo híbrido, combinando análise qualitativa, observação comportamental, leitura algorítmica e interpretação dos processos mentais envolvidos na recepção da desinformação. Esse desenho metodológico foi escolhido porque o fenômeno da autoridade artificial no Instagram não pode ser compreendido apenas pela superfície do conteúdo publicado; ele exige uma imersão profunda no tecido cognitivo que sustenta a influência



fabricada, observando como perfis fabricam credibilidade, como algoritmos amplificam ruídos e como os usuários, guiados por mecanismos mentais inconscientes, legitimam narrativas falsas sem perceber. Assim, o estudo investiga não somente o que é postado, mas como o cérebro humano processa, reage e replica tais conteúdos, transformando ilusões em verdades percebidas.

A primeira etapa metodológica consistiu em mapear um conjunto de perfis no Instagram que apresentavam padrões claros de autoridade fabricada. A seleção dessa amostra foi baseada na observação contínua da plataforma, registrando perfis que combinavam estética profissional — imagens bem produzidas, cenários supostamente científicos, símbolos de autoridade como jalecos, gráficos, microfones de podcast e ambientes minimalistas — com discursos afirmativos que apelavam ao senso de urgência, revelação privilegiada ou antagonismo institucional. A escolha desses perfis levou em conta também o comportamento da audiência, analisando-se comentários que demonstravam aceitação acrítica, indignação moral ou compartilhamento automático, comportamentos típicos de usuários capturados por vieses cognitivos como confirmação, disponibilidade e pertencimento grupal.

Para aprofundar esse entendimento, a análise se estendeu a um corpus robusto de publicações, incluindo vídeos curtos, carrosséis, transmissões ao vivo e stories arquivados. Cada peça de conteúdo foi examinada não apenas em sua materialidade textual e visual, mas em seus efeitos mentais: a forma como cores, gestos, entonação vocal, cortes rápidos, legendas enfáticas e símbolos de autoridade ativavam emoções primárias como medo, raiva, sensação de injustiça ou euforia. A investigação observou atentamente como essas emoções funcionam como catalisadores da viralização, explorando estruturas cerebrais primitivas — especialmente o sistema límbico — que reage antes da racionalidade. Assim, o processo metodológico analisou o conteúdo como um dispositivo psicológico, e não apenas comunicacional.

Durante essa etapa, foram registradas imagens conceituais que ilustram elementos centrais do estudo. Entre elas, destaca-se Figura 1 – Arquitetura Visual da Autoridade Artificial, que representa a composição estética típica dos perfis analisados: um indivíduo posicionado em frente a uma estante minimalista com livros técnicos organizados, iluminação lateral suave, microfone profissional visível e expressões faciais calculadas para transmitir confiança e domínio. Essa imagem sintetiza a estratégia visual que busca ativar no espectador o “efeito halo”, um processo mental que leva o cérebro a inferir credibilidade a partir de sinais superficiais de profissionalismo.

Figura 1 - Arquitetura visual da Influenciador artificial



Fonte: Elaborado pelo Autor(a)

Outra imagem relevante é Figura 2 – Ciclo Cognitivo da Desinformação, que sintetiza de forma visual o movimento circular que ocorre dentro da mente do usuário ao entrar em contato com narrativas manipulativas. O ciclo inicia-se com a aceitação rápida, momento em que o usuário reage de forma imediata ao conteúdo, sem acionar mecanismos analíticos mais profundos. Em seguida, o processo é reforçado pelo viés de confirmação, que leva o indivíduo a interpretar a mensagem de acordo com crenças pré-existentes, filtrando qualquer informação que entre em conflito com sua visão de mundo. A partir disso, ocorre a repetição algorítmica, quando o próprio Instagram amplifica o conteúdo emocionalmente carregado, expondo o usuário a versões contínuas da mesma narrativa. Por fim, o ciclo se intensifica pelo engajamento emocional, que fortalece o vínculo afetivo com a mensagem e incentiva sua replicação social — seja por compartilhamento, comentários ou validação pública. O processo então recomeça, criando um circuito de reforço contínuo da desinformação. Esta figura traduz visualmente o funcionamento do “atalho cognitivo”, mecanismo mental que simplifica a tomada de decisões em contextos de excesso informacional. A metodologia incorporou esses modelos visuais como instrumentos de análise conceitual, permitindo relacionar diretamente padrões comportamentais com estímulos presentes nas publicações.

Figura 2 - Ciclo Cognitivo da Desinformação



Fonte: Elaborado pelo Autor(a)

A coleta de dados incluiu observação sistemática da evolução algorítmica dos conteúdos. Foram monitoradas oscilações de engajamento, picos de viralização, padrões de compartilhamento e a relação entre emoção expressa no conteúdo e métricas de alcance. Esses dados revelaram como o algoritmo reforça conteúdos que geram reações intensas, independentemente de sua veracidade. Para captar essa dinâmica, foi elaborada a Figura 3 – Espiral Algorítmica de Amplificação, que ilustra o fenômeno no qual posts que provocam medo ou indignação recebem reforço algorítmico automático, levando a um ciclo de retroalimentação onde quanto mais emoções o conteúdo desperta, maior sua propagação.

A coerência entre emoção, engajamento e reforço algorítmico ficou ainda mais evidente quando se observou a persistência temporal desses conteúdos. Diferentemente de publicações neutras, cuja visibilidade tende a decair rapidamente, posts emocionalmente carregados permaneceram ativos por períodos muito maiores, ressurgindo repetidamente no feed de novos usuários — um comportamento que confirma a natureza retroalimentada da espiral algorítmica. Esse prolongamento artificial da vida útil do conteúdo cria a sensação de que a narrativa possui alta relevância social, quando, na verdade, sua força deriva do próprio algoritmo. Assim, o fenômeno capturado na Figura 3 demonstra que o Instagram não apenas distribui conteúdos emocionais, mas os mantém circulando como se fossem “ecos digitais”, amplificando percepções distorcidas e projetando um senso enganoso de consenso público. Esse ciclo prolongado de exposição intensifica a crença na narrativa, fortalece a autoridade artificial do emissor e molda a experiência cognitiva do usuário, que passa a interpretar a repetição como validação — um mecanismo psicológico fundamental para a expansão da desinformação.

Figura 3 - Representação do ciclo de retroalimentação no qual conteúdos que despertam medo, indignação ou forte carga emocional recebem reforço automático do algoritmo, gerando maior engajamento e ampliando progressivamente sua propagação na plataforma.



Fonte: Elaborado pelo Autor(a)

Paralelamente, o estudo aplicou um processo contínuo de verificação factual, confrontando cada narrativa com fontes confiáveis e analisando critérios técnicos de autenticidade: manipulação de imagens, gráficos fabricados, cifras sem referência, estudos inexistentes ou distorcidos. A checagem revelou um padrão recorrente: os conteúdos falsos adotam a aparência de complexidade científica para acionar o processo cognitivo da “autoridade percebida”, no qual o usuário acredita porque sente que não tem repertório para contestar.

Os dados coletados também permitiram identificar como os usuários replicam narrativas tóxicas sem perceber que estão reforçando autoridades artificialmente construídas. Comentários, reações e compartilhamentos foram interpretados à luz dos processos mentais que governam o comportamento digital: medo moral, senso de pertencimento, dissonância cognitiva e validação social. Esses elementos foram essenciais para compreender por que a audiência tende a legitimar influenciadores falsos com tanto fervor quanto especialistas reais.

A análise final consolidou esses elementos em um modelo teórico próprio, que descreve como estética, narrativa, emoção e algoritmo convergem para fabricar autoridade. Esse modelo, ilustrado conceitualmente na Figura 4 – Quadrante da Autoridade Artificial, mostra a interação entre quatro dimensões: (1) estética de credibilidade, (2) narrativa que simula expertise, (3) métricas digitais que reforçam a ilusão de popularidade e (4) algoritmos que amplificam intensamente conteúdos polarizantes.

Figura 4 - Quadrante da Autoridade Artificial



Fonte : Elaborado pela Autor(a)

Por fim, todo o estudo foi conduzido seguindo rigor ético, com foco exclusivo em dados públicos e anonimização dos perfis analisados. Reconhece-se como limitação a opacidade algorítmica do Instagram, que impede observação completa de suas engrenagens internas, bem como a impossibilidade de acessar processos mentais conscientes dos usuários — embora os padrões cognitivos observáveis permitam inferências sólidas.

### 3 RESULTADOS

A análise dos dados coletados revelou um ecossistema de influência profundamente estruturado, no qual estética, narrativa, emoção e algoritmo se entrelaçam para produzir o fenômeno da autoridade artificial no Instagram. O comportamento dos perfis observados indica que a construção dessa autoridade não ocorre de forma aleatória ou espontânea; ao contrário, segue padrões replicáveis, consistentes e altamente eficazes na captura da atenção e na manipulação cognitiva dos usuários. Os resultados mostram que cada elemento — desde as cores utilizadas nas imagens até as expressões faciais dos influenciadores — atua como peça de um mecanismo psicológico que opera na zona limítrofe entre percepção e manipulação.

Em primeiro lugar, constatou-se que a estética desempenha um papel determinante na aceitação da autoridade artificial. Perfis que simulam ambientes profissionais, como estúdios minimalistas, prateleiras com livros técnicos, quadros brancos ou microfones de alta qualidade, foram percebidos pelos usuários como mais confiáveis. Esse dado foi confirmado pela observação das interações: comentários frequentemente mencionavam a “seriedade” do influenciador, mesmo quando seus conteúdos eram factualmente incorretos. Esse fenômeno reflete a ativação do *efeito halo*, no qual sinais visuais de competência induzem o cérebro a inferir expertise sem qualquer verificação racional. A



Figura 1 – Arquitetura Visual da Autoridade Artificial, elaborada durante o estudo, representaria com precisão os elementos visuais que impulsionam essa percepção.

Um segundo achado relevante está relacionado à força emocional dos conteúdos que continham desinformação. As narrativas mais viralizadas possuíam características comuns: utilizavam medo, indignação moral ou sensação de revelação exclusiva como gatilhos centrais. O conteúdo que acionava o medo, principalmente, apresentou os maiores picos de engajamento. Postagens que alertavam sobre “verdades que estão escondendo de você”, “descobertas chocantes” ou “ameaças invisíveis” geraram não apenas visualizações mais altas, mas um padrão de comentários marcados por ansiedade, preocupação e necessidade de repassar a mensagem. Essa dinâmica confirma que a emoção é o combustível primário da viralização, reforçando a relação direta entre estímulos límbicos e propagação algorítmica.

O comportamento do algoritmo do Instagram, observado ao longo de semanas, demonstrou que conteúdos emocionalmente carregados não apenas atraíam mais interação humana, mas também recebiam impulsividade automática da própria plataforma. Essa constatação tornou clara a existência do que o presente estudo chama de Espiral Algorítmica de Amplificação, representada conceitualmente na Figura 3. Nessa espiral, conteúdos que ativam emoções intensas são interpretados pelo algoritmo como “relevantes” e, portanto, distribuídos a um número maior de usuários, ampliando seu alcance de forma exponencial. Perfis que utilizavam fake news de forte conotação emocional apresentaram um crescimento acelerado e contínuo, muitas vezes saltando milhares de seguidores em poucos dias.

A análise de discurso revelou ainda que as narrativas utilizadas por influenciadores tóxicos seguem um roteiro quase padronizado: apresentam-se como outsiders contra “sistemas corruptos”, reivindicam um suposto conhecimento proibido ou oculto, criticam especialistas reais para ocupar seu espaço simbólico e oferecem soluções simples para problemas complexos. Esse modelo discursivo simula expertise e cria nos usuários um senso de confiança baseado mais na identificação emocional do que em comprovação técnica. Os comentários analisados mostraram que seguidores repetem os termos, os jargões e as crenças expressas pelos influenciadores, evidenciando um processo de internalização automática, típico de discursos manipulativos. É o mecanismo mental do pertencimento em ação, reforçado pelo algoritmo que aproxima usuários com crenças semelhantes.

Outro resultado expressivo é o papel do viés de confirmação. Usuários que já possuíam determinados posicionamentos — políticos, morais, religiosos ou ideológicos — eram significativamente mais propensos a aceitar e compartilhar conteúdos que reforçavam suas crenças prévias, independentemente da veracidade. O viés não apenas aumentava a aceitação das fake news, mas também amplificava a autoridade artificial dos criadores, pois o público percebia o influenciador como alguém “corajoso”, “que diz a verdade” ou “que vai contra o sistema”. O engajamento gerado



por esse grupo se tornava combustível para impulsionamento algorítmico adicional, criando um ciclo de retroalimentação entre crença, autoridade e viralização.

Os resultados também evidenciam um padrão claro de impulsionamento artificial em alguns perfis. Em diversas ocasiões, observou-se uma quantidade anormal de comentários genéricos, curtidas simultâneas e perfis recém-criados interagindo com conteúdos sensacionalistas — sinais clássicos de uso de bots, fazendas de engajamento e serviços pagos. Esses impulsionamentos funcionam como “injeções artificiais” de relevância que, mais uma vez, ativam o algoritmo, levando conteúdos falsos a circularem entre usuários reais como se fossem naturalmente populares. A autoridade artificial, portanto, não é apenas construída; ela é estrategicamente inflada.

Durante a análise dos comentários, emergiram relatos de usuários que afirmavam ter mudado decisões pessoais — desde compras até opiniões políticas — a partir dos conteúdos consumidos. Embora o estudo não tenha investigado intervenções diretas na vida offline, os indícios sugerem que a influência tóxica ultrapassa o ambiente digital e se infiltra nos comportamentos concretos. Isso é particularmente visível em comentários como “agora entendi”, “isso faz sentido”, “ninguém nunca explicou desse jeito”, que revelam o momento exato em que a desinformação atravessa o filtro racional e se transforma em convicção. Aqui, observa-se o processo mental de redução de dissonância cognitiva, onde o usuário prefere acreditar na narrativa manipulada a enfrentar a complexidade do tema.

No consolidado dos achados, o estudo identificou quatro pilares fundamentais que sustentam o fenômeno da autoridade artificial no Instagram: a estética que simula credibilidade, a narrativa que se apropria do discurso técnico, a emoção que garante engajamento imediato e o algoritmo que amplifica qualquer conteúdo que maximize tempo de tela. A interação desses pilares, representada conceitualmente na Figura 4 – Quadrante da Autoridade Artificial, demonstra que ela não depende de talento comunicativo, formação técnica ou reputação real; ela depende da habilidade de acionar processos mentais primários de forma eficiente e contínua.

## 4 DISCUSSÃO

Os resultados obtidos ao longo desta pesquisa revelam uma dinâmica inquietante no ecossistema do Instagram: a consolidação de uma autoridade que não nasce do conhecimento, nem da legitimidade, nem da experiência — mas da capacidade de manipular estética, emoção e algoritmo. A análise evidencia que vivemos um momento em que a percepção de verdade é mais moldada pela aparência do que pela substância, e em que a autoridade artificial adquire peso social equivalente (e muitas vezes superior) à autoridade real. Esse fenômeno desafia não apenas o marketing digital, mas a estrutura cognitiva com a qual interpretamos o mundo.



A discussão central que emerge é que a influência algoritmik não se sustenta apenas em conteúdos falsos ou manipulações superficiais. Ela se apoia em processos mentais ancestrais: a necessidade humana de pertencimento, a busca por explicações simples, a predisposição ao medo e a tendência a confiar em figuras que exibem segurança e clareza — mesmo quando essas figuras constroem suas narrativas com base em falsidades. Em outras palavras, a desinformação não convence porque é lógica, mas porque é emocionalmente confortável. Ela oferece narrativas que dão sentido a um mundo caótico, ainda que esse sentido seja ilusório.

A estética profissional observada nos perfis estudados reforça essa interpretação. Quando um influenciador se apresenta em um ambiente visualmente associado à competência — iluminação correta, microfone profissional, mesa organizada, gráficos coloridos ou até jalecos e livros técnicos — o cérebro ativa automaticamente esquemas mentais de confiança. O usuário não confia no conteúdo porque ele é verificável, mas porque parece confiável. Esse mecanismo, conhecido como efeito halo, explica por que tantos conteúdos absurdos ganham credibilidade imediata. A confiança é estética antes de ser racional.

O algoritmo, por sua vez, não distingue verdade e mentira. Ele reconhece apenas o que mantém o usuário engajado. Nesse sentido, ele funciona como um amplificador emocional. Quanto mais o conteúdo provoca reações intensas — seja medo, indignação, surpresa ou revolta — mais ele é impulsionado. O resultado é um ambiente orientado não para a informação, mas para o estímulo constante do sistema límbico. Assim, a plataforma transforma o medo em combustível e a polarização em estratégia de distribuição. A autoridade artificial cresce porque ela entrega exatamente o que o algoritmo recompensa: emoção bruta.

Isso implica que os influenciadores tóxicos não são apenas produtores de conteúdo; eles se tornam engenheiros de comportamento. Suas narrativas simulam discursos técnicos, manipulam a percepção de complexidade e criam um senso de urgência que impede o usuário de exercer pensamento crítico. A autoridade artificial, portanto, é menos um fenômeno de comunicação e mais um fenômeno psicológico: um tipo de encantamento cognitivo que captura a mente, diminui a capacidade de julgamento e reforça crenças prévias.

A discussão toma contornos ainda mais perigosos quando observamos como esses perfis se posicionam discursivamente. Eles geralmente se apresentam como “descobridores de verdades ocultas”, assumindo o papel simbólico de heróis que enfrentam instituições, especialistas, governos e sistemas tradicionais de conhecimento. Essa retórica fortalece a relação emocional com os seguidores, que passam a enxergar o influenciador como uma espécie de guia — alguém capaz de revelar o que a suposta “elite” tenta esconder. É uma inversão completa da legitimidade: quanto mais distante das fontes confiáveis, mais “verdadeiro” o conteúdo é percebido.



Surge, então, o ponto mais crítico desta discussão: a autoridade artificial desencadeia um processo de substituição do real pelo emocional. A verdade deixa de ser um conjunto de fatos verificáveis e passa a ser aquilo que confirma o que o indivíduo deseja acreditar. Essa distorção cognitiva transforma influenciadores tóxicos em líderes de opinião capazes de influenciar decisões pessoais, profissionais, políticas e até de saúde pública. A desinformação não fica confinada à tela; ela extrapola o digital e molda comportamentos concretos.

Isso revela um alerta profundo para o marketing digital. A construção de autoridade — antes baseada em reputação real, histórico profissional e entrega de valor — está sendo corroída pela lógica algorítmica da viralização emocional. A “autoridade artificial” não é uma anomalia do sistema; ela é um produto natural do próprio design da plataforma. E, enquanto esse design permanecer orientado pela economia da atenção, perfis que manipulam emoções sempre terão vantagem competitiva sobre especialistas reais.

Essa discussão também aponta para a fragilidade do usuário frente ao universo digital. Em um ambiente onde estética substitui competência, onde emoção suplanta lógica e onde a repetição algorítmica valida qualquer narrativa, a mente humana se torna facilmente redirecionável. Não por ignorância, mas por funcionamento natural do cérebro. A autoridade artificial vence porque conversa diretamente com nossos instintos, não com nossa razão.

O ponto final desta discussão é inevitável: combater a IDTOX não significa apenas corrigir fake news. Significa reconstruir uma cultura cognitiva na qual o usuário recupere a capacidade de pausa, dúvida e discernimento — habilidades que o Instagram, tal como funciona hoje, sistematicamente corrói. A solução passa por educação midiática, transparência algorítmica e responsabilidade ética no marketing digital, mas também por um reposicionamento urgente: a verdade precisa voltar a ser mais sedutora que a mentira. Caso contrário, a autoridade artificial continuará a crescer — elegante, convincente e perigosamente eficaz.

## 5 CONCLUSÃO

A análise realizada ao longo deste estudo revela que a ascensão da autoridade artificial no Instagram não é um fenômeno isolado nem acidental; ela é resultado direto da convergência entre vulnerabilidades cognitivas humanas, estratégias de manipulação estética e um sistema algorítmico orientado exclusivamente à maximização do engajamento. O que se observa é um ambiente onde a percepção de credibilidade é facilmente fabricável e onde a emoção supera a razão como critério de validação da verdade. Assim, perfis que operam com fake news tornam-se agentes centrais na construção de um tipo de influência profundamente tóxica, capaz de interferir não apenas nas crenças, mas nas decisões práticas dos usuários — seja no consumo, na saúde, na moralidade ou na política.



Ao longo da pesquisa, demonstrou-se que os mecanismos de viralização e os padrões discursivos associados à desinformação exploram processos mentais automáticos, ativando emoções rápidas, reduzindo o espaço para análise crítica e criando um ciclo de legitimação mútua entre influenciador e audiência. A autoridade artificial não se impõe pela lógica ou pela evidência, mas pela narrativa sedutora que responde ao que o usuário quer ouvir, não ao que precisa saber. Em um contexto dominado pela economia da atenção, a verdade compete em desvantagem com conteúdos que manipulam medo, indignação e sensação de descoberta exclusiva — estímulos que o algoritmo interpreta como sinônimo de relevância.

Dessa forma, a IDTOX no Instagram transcende o campo do marketing digital e alcança a esfera social mais profunda: ela molda percepções coletivas, contamina debates públicos e fragiliza a relação entre indivíduo e realidade. Ao substituir especialistas por simulacros de autoridade, a plataforma erode pilares tradicionais de confiança e abre espaço para que discursos irracionais adquiram legitimidade social. O risco não está apenas na mentira em si, mas no sistema que premia quem a produz com mais eficiência emocional.

Os resultados também indicam que combater essa forma de autoridade artificial requer ações em múltiplos níveis. No campo educacional, é urgente reforçar a literacia midiática e cognitiva, capacitando usuários a reconhecer estratégias de manipulação visual e emocional. No campo tecnológico, é necessária maior transparência nas regras de distribuição algorítmica, além de mecanismos mais rigorosos de identificação e restrição de conteúdos manipulativos. Já no campo ético e profissional, o marketing digital precisa se reposicionar, resgatando responsabilidade e autenticidade como valores centrais, sob pena de ver o próprio setor devorado por práticas nocivas que corroem a confiança do público.

Apesar da robustez metodológica e da profundidade analítica alcançadas, este estudo apresenta limitações que devem ser reconhecidas para uma compreensão equilibrada dos resultados. A principal delas diz respeito à opacidade estrutural dos algoritmos do Instagram, que impede acesso pleno às variáveis internas responsáveis pela distribuição e priorização dos conteúdos. Além disso, a pesquisa se baseou exclusivamente em dados públicos, o que restringe a observação de dinâmicas privadas de engajamento que poderiam ampliar ainda mais a compreensão do fenômeno. Também não foram conduzidas entrevistas diretas com usuários, o que limita a interpretação subjetiva dos processos mentais envolvidos — ainda que os padrões cognitivos identificados forneçam indícios sólidos. Reconhecer essas limitações não enfraquece os achados; ao contrário, reforça a necessidade de novos estudos interdisciplinares que combinem psicologia, ciência de dados e comunicação digital para aprofundar o entendimento sobre a formação e o impacto da autoridade artificial nas plataformas sociais.



Por fim, esta pesquisa evidencia que a influência digital tóxica não é apenas um problema técnico ou comunicacional; é um fenômeno civilizacional. A disputa pela atenção tornou-se uma disputa pela mente, e a mente, quando sujeita a estímulos contínuos e manipulativos, torna-se terreno fértil para ilusões convincentes. Reconhecer esse cenário é o primeiro passo para enfrentá-lo. O próximo é construir estratégias que devolvam ao usuário a capacidade de discernir, duvidar e pensar — pilares essenciais para que a verdade não seja apenas uma opção entre muitas narrativas, mas a base sobre a qual reconstruímos nossa relação com o mundo digital e com nós mesmos.



## REFERÊNCIAS

- BARABÁSI, Albert-László. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. New York: Basic Books, 2014.
- BARDIN, Laurence. *Análise de Conteúdo*. São Paulo: Edições 70, 2016.
- BAKSHY, Eytan; ROSEN, I.; ADAMIC, L. *Exposure to ideologically diverse news and opinion on Facebook*. *Science*, v. 348, n. 6239, p. 1130–1132, 2015.
- BERNERS-LEE, Tim; FELLER, Joseph. *The Web's Role in the Spread of Misinformation*. Oxford: Oxford University Press, 2021.
- BOYD, Danah; CRAWFORD, Kate. *Critical Questions for Big Data*. *Information, Communication & Society*, v. 15, n. 5, p. 662–679, 2012.
- CARR, Nicholas. *The Shallows: What the Internet Is Doing to Our Brains*. New York: W. W. Norton, 2020.
- CASSESE, Erin et al. *Fake News, Emotions, and Political Persuasion*. *Political Psychology*, v. 43, n. 1, p. 1–22, 2022.
- CENTRO INTERNACIONAL PARA JOURNALISMO. *Guide to Fighting Misinformation*. 2021.
- CHEN, Emily; LUCAS, C.; WENGER, A. *Bot Behavior on Social Networks*. *Journal of Social Computing*, v. 4, n. 3, p. 210–230, 2022.
- CUNNINGHAM, Stuart; CRAIG, David. *Social Media Entertainment: The New Intersection of Hollywood and Silicon Valley*. New York: NYU Press, 2019.
- EUROPEAN COMMISSION. *Disinformation and Social Media: Annual Report*. Brussels, 2023.
- FERRARA, Emilio. *Disinformation and Social Bot Operations in the Age of Algorithms*. Massachusetts Institute of Technology (MIT) Press, 2022.
- FIRMINO, Tiago; SILVA, Tarcísio. *Racismo Algorítmico: Uma Introdução*. São Paulo: Editora Sueli Carneiro, 2022.
- GIGERENZER, Gerd. *Risk Savvy: How to Make Good Decisions*. New York: Penguin, 2015.
- HABERMAS, Jürgen. *A Inclusão do Outro*. São Paulo: Loyola, 2018.
- HOSPEDALES, Timothy et al. *Meta-Learning in the Age of Deepfakes and Misinformation*. *ACM Computing Surveys*, v. 55, n. 6, p. 1–36, 2023.
- KAISER, Brittany; CADWALLADR, Carole. *The Cambridge Analytica Files*. The Guardian, 2018.
- LITTLE, Daniel. *Cognitive Biases and the Spread of False Beliefs*. *Psychology Review*, v. 127, n. 2, p. 231–248, 2020.
- MARWICK, Alice; LEWIS, Rebecca. *Media Manipulation and Disinformation Online*. Data & Society Research Institute, 2017.



MIGUEL, Luís Felipe. *Democracia e Debate Público em Tempos de Desinformação*. São Paulo: Boitempo, 2022.

NICHOLAS, David. *Attention Economy in the Social Media Era*. Oxford Internet Institute, 2021.

NICHOLSON, Scott. *Fake News: Understanding Media Manipulation*. Chicago: University of Chicago Press, 2020.

PARISER, Eli. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin, 2011.

PENNYCOOK, Gordon; RAND, David. *The Psychology of Fake News*. Trends in Cognitive Sciences, v. 25, n. 5, p. 388–402, 2021.

SANTAELLA, Lucia. *A Pós-Verdade é Verdadeira ou Falsa?* São Paulo: Estação das Letras e Cores, 2018.

SILVA, Tarcísio. *Desinformação e o Poder das Plataformas Digitais*. São Paulo: HackerVeritas, 2021.

SIMMEL, Georg. *As Formas da Sociedade*. São Paulo: Perspectiva, 2020.

SUNSTEIN, Cass. *Going to Extremes: How Algorithms Push People to the Edge*. Oxford University Press, 2020.

TANDOC JR., Edson; LIM, Z.; LING, R. *Defining “Fake News”*. Digital Journalism, v. 6, n. 2, p. 137–153, 2018.

VAZ, Paulo. *Cultura da Conexão e Informação Perigosa*. Rio de Janeiro: Ed. UFRJ, 2022.

WASIAK, Z.; WALKER, S. *Emotional Contagion on Instagram: How Visual Culture Transmits False Beliefs*. Journal of Digital Media Studies, v. 3, n. 4, p. 44–66, 2023.

ZUBOFF, Shoshana. *The Age of Surveillance Capitalism*. New York: PublicAffairs, 2019.