



## PSICOPATOLOGIAS VIRAIS: AUTODIAGNÓSTICO E DESINFORMAÇÃO EM SAÚDE MENTAL

**Adriano Alves da Silva**

Centro Universitário Luterano de Palmas (CEULP/ULBRA)

**Élida Souza Silva**

Centro Universitário Luterano de Palmas (CEULP/ULBRA)

**Mayelle Batista da Silva**

Centro Universitário Luterano de Palmas (CEULP/ULBRA)

### RESUMO

As redes sociais remodelaram a circulação de saberes em saúde mental, expondo usuários a conteúdos de alta difusão e acurácia variável. Este artigo examina, à luz de teorias da comunicação e da psicologia, como mecanismos algorítmicos e vieses cognitivos favorecem desinformação e autodiagnóstico. Realizamos uma revisão integrativa (2010–2025) com buscas em PsycINFO, PubMed/MEDLINE, Scopus, Web of Science e SciELO, incluindo apenas estudos revisados por pares em português e inglês. A síntese indica que, em plataformas de vídeo curto, combinações de saliência e enquadramento com repetição, alta fluência e prova social sustentam efeitos psicológicos como illusory truth, além de mediações por nocebo e cibercondria; em paralelo, dinâmicas de câmaras de eco e algoritmos silenciadores estreitam a diversidade informacional percebida e ampliam a plausibilidade de listas inespecíficas de sintomas, favorecendo autodiagnósticos sem avaliação clínica. Discutimos implicações para prática clínica, saúde pública e desenho de intervenções baseadas em evidências – com destaque para prebunking e accuracy nudges – e apontamos lacunas relativas a estudos longitudinais, auditorias algorítmicas e validação transcultural. Propomos, por fim, uma matriz psicocomunicacional integrativa para orientar pesquisas e políticas que mitiguem autodiagnósticos indevidos, preservando o valor do apoio entre pares nos ambientes digitais.

**Palavras-chave:** Redes sociais; Desinformação; Autodiagnóstico; Saúde mental; Vieses cognitivos.

### INTRODUÇÃO

As redes sociais reconfiguraram a circulação de sentidos sobre saúde mental ao combinar formatos breves, métricas públicas de popularidade e recomendação orientada por indicadores de engajamento. Nesse ecossistema, a saliência de temas e a forma de enquadramento das mensagens deixam de ser apenas atributos editoriais dos meios e passam a ser co-produzidas por sistemas de ranqueamento, com efeitos sobre o que se torna visível e crível para diferentes públicos (McCombs; Shaw, 1972; Entman, 1993). A hipótese clássica da espiral do silêncio segue útil para explicar por que percepções de “clima de opinião” inibem o dissenso; em ambientes plataformizados, tais percepções são amplificadas por sinais sociais (curtidas, comentários) e por arquiteturas que distribuem conteúdos de modo seletivo, produzindo agrupamentos de afinidade e redundância informacional (Noelle-Neumann, 1974; Cinelli et al., 2021).



Neste trabalho, adotamos o termo “algoritmos silenciadores” tal como cunhado por Silva e Pôrto Júnior (2024), para designar o efeito combinado de sistemas de recomendação e métricas de engajamento que reduzem a visibilidade de conteúdos dissonantes nos clusters, reforçam a aparência de consenso e desestimulam a exposição a alternativas informadas; trata-se de um mecanismo probabilístico emergente de regras de ranqueamento centradas em engajamento e homofilia, não de uma intenção normativa necessária das plataformas (Cinelli et al., 2021; Vosoughi; Roy; Aral, 2018; Silva; Pôrto Júnior, 2024).

No plano psicológico, julgamentos sob incerteza são guiados por heurísticas e vieses. O efeito de verdade ilusória mostra que a simples repetição eleva a veracidade percebida, mesmo quando a pessoa conhece a resposta correta (Fazio et al., 2015). Em contextos de saúde, cibercondria descreve a busca online excessiva por informação, associada a hipervigilância e ansiedade; já o nocebo evidencia como expectativas negativas podem induzir ou intensificar sintomas subjetivos (Starcevic; Berle, 2013; Colloca, 2024). Revisões recentes indicam que crenças desinformativas persistem pela conjunção de fluência, viés de confirmação e prova social (indícios observáveis de aprovação coletiva - p. ex., curtidas, comentários, compartilhamentos - que sinalizam aceitabilidade e orientam julgamento e ação), resistindo a correções pontuais (Ecker et al., 2022). No debate público, a expansão semântica de categorias de dano - concept creep - contribui para a rotulação patologizante de experiências comuns, especialmente quando rótulos funcionam como marcadores identitários em comunidades online (Haslam, 2016; Haslam, 2021). Esses processos são modulados por literacia em saúde e eHealth literacy, que condicionam a capacidade de julgar fontes e aplicar informação qualificada (Nutbeam, 2000; Norman; Skinner, 2006).

Aplicadas ao recorte de saúde mental nas plataformas, tais dinâmicas ajudam a explicar por que conteúdos experienciais e listas “guarda-chuva” de sintomas, embora frequentemente imprecisos, alcançam maior engajamento do que comunicações clínicas mais densas. Em TDAH, análises de vídeos populares apontam proporções elevadas de afirmações enganosas e generalizações, com predomínio de criadores leigos entre os mais influentes (Yeung et al., 2022). Em transtornos alimentares, convivem materiais pró-recuperação e conteúdos potencialmente desencadeantes, com associação entre popularidade e baixa qualidade informacional (Lookingbill; Mohammadi; Cai, 2023). Em depressão/ansiedade, o chamado TherapyTok é dominado por relatos em primeira pessoa e conselhos genéricos, de utilidade percebida, porém com heterogeneidade de acurácia (Samuel; Kuijpers; Bleakley, 2024). À luz do quadro teórico acima, algoritmos silenciadores operam como condição de possibilidade para que a exposição repetida, o enquadramento emotivo e a prova social sustentem crer para ver, favorecendo o autodiagnóstico sem mediação profissional.

Para investigar como mecanismos de saliência, enquadramento e distribuição algorítmica - sintetizados pelo conceito de algoritmos silenciadores - interagem com heurísticas, vieses, cibercondria e expectativas negativas para incrementar a probabilidade de desinformação e estimular autodiagnósticos indevidos, realizamos uma revisão integrativa (2010–2025), com buscas em PsycINFO, PubMed/MEDLINE, Scopus, Web of Science e SciELO, incluindo clássicos seminais anteriores quando indispensáveis à fundamentação; consideramos artigos revisados por pares em português e inglês, registrando delineamentos, medidas e desfechos relevantes para o tema.



## 1 REFERENCIAL TEÓRICO

A compreensão do autodiagnóstico em saúde mental mediado por redes sociais requer um arcabouço que articule mecanismos comunicacionais de visibilidade e sentido com processos psicológicos de crença e decisão. No plano da comunicação, a função de agenda-setting explica como sistemas midiáticos, hoje plataformizados, modulam a saliência do que entra em pauta pública (McCombs; Shaw, 1972), enquanto o framing mostra que enquadramentos (metáforas causais, ênfases, soluções) orientam inferências e preferências do público (Entman, 1993). A hipótese da espiral do silêncio acrescenta que percepções de clima de opinião tendem a inibir dissensos (Noelle-Neumann, 1974). Em ecossistemas regidos por ranqueamento algorítmico e métricas de engajamento, tais processos se combinam com homofilia e câmaras de eco, produzindo exposição redundante a conteúdos congruentes e reforçando crenças (Cinelli et al., 2021). Nesse contexto, adotamos o conceito analítico de “algoritmos silenciadores” (Silva; Pôrto Júnior, 2024) para designar o efeito probabilístico pelo qual sistemas de recomendação e sinais de engajamento reduzem a visibilidade de conteúdos dissonantes em clusters, reforçando aparência de consenso e desestimulando a busca ativa por contrapontos, o que se configura como uma dinâmica compatível com achados sobre a vantagem difusiva de mensagens novas e emocionais (Vosoughi; Roy; Aral, 2018).

Entendemos “algoritmos silenciadores” como rótulo heurístico para a convergência entre (i) objetivos de maximização de engajamento (tempo de visualização, compartilhamentos), (ii) feedbacks de prova social (curtidas, comentários), (iii) homofilia estrutural da rede e (iv) personalização de recomendações. Essa convergência prioriza conteúdos de alta fluência e apelo identitário e subprioriza comunicações técnicas com maior custo cognitivo, estreitando a variedade informacional em cada bolha. Em saúde mental, isso eleva a probabilidade de que listas “guarda-chuva” e testemunhos emocionais ganhem espaço frente a mensagens clínicas mais nuançadas (Cinelli et al., 2021; Vosoughi; Roy; Aral, 2018; Silva; Pôrto Júnior, 2024).

A literatura distingue misinformation (informação enganosa compartilhada sem intenção de enganar), disinformation (enganosa com intenção) e malinformation (verdadeiro fora de contexto, gerando dano) (Ecker et al., 2022). No domínio da saúde mental, as três formas aparecem em checklists inespecíficas, conselhos não validados e narrativas que extrapolam casos individuais. Essa taxonomia orienta a análise de acurácia e potencial de dano, sobretudo quando rótulos diagnósticos são expandidos para abranger experiências cotidianas (Haslam, 2016; Haslam, 2021).

Sob incerteza e sobrecarga informacional, julgamentos dependem de heurísticas que economizam esforço ao custo de vieses (Ecker et al., 2022). O efeito de verdade ilusória demonstra que a repetição aumenta a veracidade percebida mesmo quando o indivíduo conhece a resposta correta, explicando porque slogans e “listas virais” parecem confiáveis após múltiplas exposições (Fazio et al., 2015). Em saúde, cibercondria descreve busca online excessiva associada a hipervigilância e ansiedade, favorecendo interpretações patologizantes de sinais inespecíficos (Starcevic; Berle, 2013). O nocebo mostra que expectativas negativas podem induzir ou intensificar experiências subjetivas (Colloca, 2024). Esses mecanismos se acoplam às dinâmicas comunicacionais: algoritmos silenciadores aumentam exposição seletiva e repetição, framing emotivo e prova social elevam fluência, e o conjunto sustenta crenças imprecisas resistentes a correções (Ecker et al., 2022).



A noção de concept creep argumenta que conceitos de dano e patologia tendem a expandir suas fronteiras semânticas, deslocando limites entre normalidade e transtorno (Haslam, 2016; Haslam, 2021). Em ambientes de rede, rótulos diagnósticos podem funcionar como marcadores identitários (“neuroidentidades” – que pode ser interpretado como rótulos identitários baseados em categorias clínicas – p. ex., TDAH, TEA – adotados por usuários para fins de pertencimento e auto-descrição), estabilizados por recompensas sociais (pertencimento, reconhecimento). A aprendizagem social e a modelagem por figuras de referência, inclusive influenciadores, contribuem para a observação imitativa e para a legitimação de atribuições diagnósticas em primeira pessoa (Bandura, 2001). Sob a lente de uses & gratifications, parte do público busca informação, afiliação e confirmação identitária, o que ajuda a compreender a vantagem atencional de narrativas pessoais frente a conteúdos técnico-científicos (Ruggiero, 2000).

Formatos como TikTok/Reels/Shorts combinam tempo reduzido, edição veloz, música de fundo e legendas imperativas que funcionam como andaimes atencionais. Em termos cognitivos, aumentam fluência e recordação; socialmente, acoplam-se a prova social visível; comunicacionalmente, favorecem enquadramentos simplificadores. Esse triplo acoplamento explica por que conteúdos experienciais e checklists, frequentemente imprecisos, podem superar mensagens clínicas em engajamento e alcance, como mostram estudos em TDAH, transtornos alimentares e ansiedade/depressão (Yeung et al., 2022; Lookingbill; Mohammadi; Cai, 2023; Samuel; Kuijpers; Bleakley, 2024).

A literacia em saúde (competências funcional, interativa e crítica) e a eHealth literacy (mensurável, por exemplo, via eHEALS) moderam a suscetibilidade a desinformação e a propensão ao autodiagnóstico: níveis mais altos associam-se a melhor triagem de fontes e menor adesão a conteúdos de baixa qualidade (Nutbeam, 2000; Norman; Skinner, 2006). Idade e fase do desenvolvimento também importam: adolescentes e jovens adultos, em processo de formação identitária, tendem a valorizar pares e influenciadores, o que pode potencializar os efeitos de algoritmos silenciadores sobre a percepção de consenso e sobre a aceitação de rótulos diagnósticos.

A inoculation theory sugere que prebunking, isto é, expor preventivamente o público a técnicas retóricas enganosas, uma estratégia de inoculação que antecipa técnicas comuns de manipulação, fortalecendo a resistência cognitiva antes da exposição ao conteúdo enganoso, pode fortalecer “anticorpos cognitivos” (van der Linden; Lewandowsky; Ecker, 2017). Nudges de acurácia, de modo a tornar saliente o critério de veracidade no momento da avaliação/compartilhamento, apresentam efeitos replicáveis em reduzir intenção de compartilhar informações falsas (Pennycook; Rand, 2021; 2022). Integrando o quadro, derivam-se hipóteses: (H1) exposição seletiva e repetição mediadas por algoritmos silenciadores aumentam crenças imprecisas; (H2) framing emotivo e prova social elevam fluência e veracidade percebida; (H3) literacia e eHealth literacy atenuam esses efeitos, ao passo que busca identitária e afinidade com influenciadores os amplificam.

Em síntese, o modelo psicocomunicacional que orienta este estudo postula um encadeamento: plataforma/algoritmo → ecologia informacional (saliência, enquadramentos, redundância/eco) → processamento (heurísticas, illusory truth, cibercondria, nocebo) → desfechos (crença, autodiagnóstico), moderados por literacia, idade e identidade. O conceito de algoritmos silenciadores ajuda a nominar a engrenagem que estreita a diversidade informacional e eleva a plausibilidade de autodiagnósticos quando listas inespecíficas e narrativas emocionais dominam o fluxo,



cenário em que mitigações baseadas em evidência tornam-se componentes críticos de governança informacional em saúde mental (Cinelli et al., 2021; Vosoughi; Roy; Aral, 2018; Ecker et al., 2022).

## 2 MATERIAIS E MÉTODOS

Este estudo configura-se como revisão integrativa da literatura, concebida para sintetizar e articular evidências de psicologia (julgamento/decisão; psicologia da saúde/mental) e comunicação (teorias clássicas e comunicação digital) a fim de explicar como as redes sociais contribuem para a desinformação em saúde mental e de que modo essa dinâmica estimula autodiagnósticos sem avaliação profissional. A opção por revisão integrativa favorece a incorporação de delineamentos heterogêneos (análises de conteúdo, estudos observacionais e experimentais, revisões sistemáticas/metanálises) e a construção de uma síntese explicativa (Baethge; Goldbeck-Wood; Mertens, 2019).

Para atender a um recorte temporal e linguístico, buscou-se literatura publicada entre 2010 e 2025, em português e inglês. Clássicos anteriores foram incluídos apenas quando seminais à fundamentação teórica (p. ex., agenda-setting e framing). O escopo empírico contemplou plataformas com feed algorítmico e vídeo curto (TikTok, Instagram Reels, YouTube/Shorts) e, quando pertinente, X/Twitter, Reddit e Facebook.

Para assegurar transparência e reprodutibilidade das buscas, realizaram-se consultas nas bases PsycINFO, PubMed/MEDLINE, Scopus, Web of Science e SciELO, utilizando combinações de descritores em português e inglês com operadores booleanos e, quando aplicável, termos controlados (MeSH/DeCS). As cadeias nucleares integraram três eixos semânticos: contexto/plataforma (“social media”, TikTok, Instagram, YouTube, Twitter/X, Reddit, “redes sociais”), fenômeno informacional (misinformation, disinformation, “desinformação”, “conteúdo enganoso”) e constructos clínico-psicológicos (“mental health”/“saúde mental”, “self-diagnos”/autodiagnóstico, cyberchondria/cibercondria, “illusory truth”, nocebo). Exemplos de consultas foram: (“social media” AND misinformation AND “mental health”); (autodiagnóstico AND “redes sociais”); (cyberchondria AND “social media”); (“illusory truth” AND framing); e (accuracy nudges AND misinformation AND “mental health”). Por “Accuracy nudges”, entende-se a sinalização de se tornar necessária a verificação de veracidade no momento de avaliar/compartilhar conteúdo, deslocando a atenção para a acurácia. As estratégias foram pilotadas e refinadas por dois avaliadores, visando equilíbrio entre sensibilidade (recuperação ampla do tema) e especificidade (redução de ruído).

Para estabelecer elegibilidade, incluíram-se estudos revisados por pares que abordassem redes sociais e ao menos um dos eixos: desinformação em saúde/saúde mental, autodiagnóstico (self-diagnosis) ou constructos psicológicos associados (p. ex., cibercondria, nocebo, illusory truth, literacia em saúde) no contexto de redes; aceitaram-se estudos empíricos (observacionais/experimentais) e revisões (sistemáticas/narrativas/metanálises) com método declarado, em PT/EN, publicados entre 2010–2025. Excluíram-se preprints, editoriais, cartas sem dados, relatórios não revisados, livros/capítulos sem peer review, teses/dissertações e estudos cujo “autodiagnóstico” não se relacionasse à saúde mental ou ao uso de redes.

Para reduzir o viés de seleção, a triagem ocorreu em duas etapas independentes (títulos/resumos → texto completo) conduzidas por dois avaliadores; divergências foram resolvidas por consenso e, quando necessário, por terceiro avaliador. O relato do fluxo de seleção e os motivos de exclusão no texto integral seguiram recomendações



de transparência do PRISMA 2020, reconhecendo-se o caráter integrativo (não sistemático) da revisão (Page et al., 2021). Foram incluídos 13 estudos publicados entre 2010 e 2025, com foco em TikTok, Instagram Reels, YouTube/Shorts e Twitter/X. A amostra compreendeu 23% análises de conteúdo, 15% estudos observacionais e 31% experimentos, além de 31% revisões (sistemáticas/narrativas).

Para garantir consistência conceitual, elaborou-se um codebook prévio que operacionalizou os constructos: desinformação como guarda-chuva para misinformation (sem intenção), disinformation (com intenção) e malinformation (verdadeira fora de contexto); autodiagnóstico como atribuição de rótulo clínico sem avaliação profissional baseada em conteúdo consumido em redes; “algoritmos silenciadores” como categoria analítica (Silva; Pôrto Júnior, 2024) que codifica redução de visibilidade de conteúdos dissonantes por ranqueamento personalizável e sinais de engajamento – com proxies observáveis como exposição seletiva, redundância em clusters, homofilia e vantagem difusiva de conteúdos emotivos/identitários; mecanismos psicológicos (p. ex., illusory truth, cibercondria, nocebo, vieses/heurísticas); e moderadores (literacia em saúde/eHealth – instrumentos como eHEALS, quando reportados – , idade/grupo etário, tipo de fonte e formato de mensagem). Como indicadores de qualidade informacional e potencial de dano, quando disponíveis nos estudos incluídos, foram observadas métricas como DISCERN, JAMA benchmark e PEMAT, bem como a presença de “checklists guarda-chuva” e desfechos de crença, intenção de autodiagnóstico e busca de cuidado.

Para apreciação crítica da robustez metodológica, empregaram-se instrumentos adequados ao delineamento: MMAT 2018 para métodos mistos/qualitativos/quantitativos (Hong et al., 2018), checklists JBI para observacionais (Aromataris; Munn, 2020) e AMSTAR 2 para revisões sistemáticas (Shea et al., 2017). A força inferencial foi ponderada por uma hierarquia analítica (metanálises/revisões sistemáticas > experimentos/ensaios > coortes > transversais > análises de conteúdo), utilizada apenas para calibrar a discussão (não como critério de exclusão). Como verificação de confiabilidade, calculou-se coeficiente Kappa em aproximadamente 30% da amostra (meta  $\geq 0,70$ ).

Para conduzir a síntese, adotou-se abordagem narrativa com análise temática reflexiva (Braun; Clarke, 2006), organizando os achados em eixos: qualidade/acurácia informacional, padrões de engajamento e linguagem do vídeo curto, mecanismos psicológicos de crença e decisão, arquiteturas sociotécnicas/“algoritmos silenciadores” e estratégias de mitigação. A interpretação buscou encadear níveis (mensagem–usuário–plataforma) e explicitar a força da evidência (associativa vs. experimental), em consonância com boas práticas de relato (Page et al., 2021).

### 3 RESULTADOS E DISCUSSÕES

A síntese integrativa do corpus (2010–2025) evidenciou padrões consistentes entre plataformas com feed algorítmico e forte presença de vídeo curto (TikTok, Instagram Reels, YouTube/Shorts), convergindo para dois achados gerais: (i) acurácia informacional variável, com vantagem de alcance para conteúdos experienciais e “listas guarda-chuva” de sintomas; e (ii) mecanismos psicocomunicacionais que explicam a adesão e a persistência de crenças imprecisas, com ênfase no papel de arquiteturas de recomendação que estreitam a diversidade informacional percebida. Esses resultados dialogam diretamente com os objetivos de compreender como as redes contribuem



para a desinformação em saúde mental e de que modo essa dinâmica estimula autodiagnósticos sem avaliação profissional.

No eixo da acurácia informacional, análises de conteúdo em TDAH, transtornos alimentares e ansiedade/depressão mostram proporções relevantes de informações enganosas ou descontextualizadas, acompanhadas de maior engajamento de narrativas pessoais e conselhos genéricos quando comparadas a comunicações clínicas (Yeung et al., 2022; Lookingbill; Mohammadi; Cai, 2023; Samuel; Kuijpers; Bleakley, 2024). O padrão é ambivalente: comunidades online oferecem apoio entre pares e redução de estigma, mas esse mesmo ambiente favorece a circulação de rótulos diagnósticos amplos e de checklists inespecíficas, condições que podem induzir autoatribuições precipitadas. A ambivalência sugere que engajamento não é proxy de qualidade, o que tem implicações tanto para avaliação crítica por parte dos usuários quanto para diretrizes de ranking e rotulagem informacional.

A linguagem do vídeo curto emerge como elemento estruturante desses efeitos. Formatos breves com edição veloz, trilha sonora e legendas imperativas funcionam como andaimes atencionais que aumentam fluência de processamento (grau de facilidade com que uma mensagem é processada; maior fluência tende a elevar familiaridade e veracidade percebida) e recordação; associados à prova social visível (curtidas, comentários), elevam a veracidade percebida de mensagens simples e fortemente afetivas (Fazio et al., 2015). Em termos comunicacionais, enquadramentos (framing) que oferecem explicações unívocas e soluções rápidas tendem a competir com mensagens técnicas com maior nuance (Entman, 1993). Em termos psicológicos, a combinação entre repetição, viés de confirmação e fluência responde pela resistência à correção que a literatura vem documentando (Ecker et al., 2022). O resultado prático é a maior plausibilidade subjetiva de listas “guarda-chuva” e testemunhos emocionais, sobretudo quando funcionam como sinais identitários dentro das comunidades.

As arquiteturas sociotécnicas ajudam a explicar a escala e a persistência desses fenômenos. Em ambientes de rede, homofilia e câmaras de eco organizam audiências em clusters de afinidade, com exposição redundante a conteúdos congruentes (Cinelli et al., 2021). Em paralelo, estudos de difusão em larga escala mostram que conteúdos novos e emocionais tendem a se propagar mais rápido e mais longe do que conteúdos precisos, mas menos excitantes (Vosoughi; Roy; Aral, 2018). À luz desses achados, adotamos o conceito analítico de algoritmos silenciadores para nomear o efeito probabilístico pelo qual recomendação personalizada e sinais de engajamento reduzem a visibilidade de conteúdos dissonantes dentro de clusters, reforçando aparência de consenso e desestimulando o contato com contrapontos informados (Silva; Pôrto Júnior, 2024). Importa sublinhar que o termo descreve mecanismo emergente de regras de ranqueamento e não pressupõe intencionalidade normativa das plataformas; ainda assim, seus efeitos combinam-se com agenda-setting (saliência) e framing (Enquadramento) para estreitar as possibilidades de interpretação pública (McCombs; Shaw, 1972; Entman, 1993).

No plano psicológico-clínico, o conjunto de evidências sustenta que cibercondria e efeitos nocebo são vias plausíveis de mediação entre consumo digital e autopercepção sintomática. A busca ansiosa por informação em saúde aumenta a vigilância a sinais inespecíficos e favorece interpretações patologizantes, enquanto expectativas negativas podem induzir ou intensificar experiências subjetivas (Starcevic; Berle, 2013; Colloca, 2024). Esses processos, quando acoplados a exposição seletiva, repetição e prova social, ajudam a explicar por que indivíduos passam a reconhecer-se em descrições vagas e a adotar rótulos sem avaliação clínica. Do ponto de vista



sociocultural, a expansão semântica de categorias de dano - concept creep - oferece uma lente para compreender a psiquiatrização do cotidiano em arenas públicas, nas quais rótulos funcionam também como marcadores de pertencimento (Haslam, 2016; Haslam, 2021).

A análise indica ainda moderadores individuais e contextuais. Literacia em saúde e eHealth literacy atuam como amortecedores: níveis mais altos associam-se a melhor triagem de fontes e menor adesão a conteúdos de baixa qualidade (Nutbeam, 2000; Norman; Skinner, 2006). Idade e fase do desenvolvimento importam, pois adolescentes e jovens adultos tendem a valorizar pares e influenciadores, intensificando a percepção de consenso local e, por consequência, a plausibilidade de autodiagnósticos identitários. Fonte (profissional vs. leigo) e formato (vídeo curto vs. texto longo) também modulam efeitos; no corpus analisado, conteúdos de profissionais aparecem menos e, quando presentes, concorrem em desvantagem atencional, o que reforça a necessidade de adaptação retórica sem perda de rigor.

Quanto às estratégias de mitigação, duas famílias de intervenções apresentam evidência replicável: inoculação (prebunking) - que treina o reconhecimento de técnicas retóricas enganosas antes da exposição - e nudges de acurácia - que tornam saliente o critério de veracidade no momento de avaliar/compartilhar (van der Linden; Lewandowsky; Ecker, 2017; Pennycook; Rand, 2021; 2022). Esses resultados sustentam que ajustes leves no ambiente decisório podem reduzir intenção de compartilhar e melhorar julgamentos de veracidade. Persistem, entretanto, desafios de transferência para vídeos ultracurtos e públicos adolescentes, bem como questões de contexto cultural que exigem adaptações locais. À medida que se reconhece a operação de algoritmos silenciadores, emerge a necessidade de testar intervenções in-feed que conciliem fricções ao compartilhamento de alegações médicas com vias de referência a fontes de alta qualidade, sem produzir efeitos adversos de reatância.

Em síntese interpretativa, os objetivos propostos foram contemplados: a revisão mostrou como mecanismos de saliência, enquadramento e distribuição algorítmica se articulam a heurísticas, efeitos de repetição, cibercondria e expectativas negativas para aumentar a probabilidade de desinformação e estimular autodiagnósticos. O conceito de algoritmos silenciadores funcionou como peça de ligação entre os níveis comunicacional e psicológico, iluminando a estreita janela de conteúdos que muitos usuários efetivamente experimentam. À luz da força da evidência (predominantemente associativa, com suporte experimental para mecanismos cognitivos e para algumas intervenções), as implicações práticas envolvem triagem clínica de autodiagnóstico e exposição digital, educação midiática situada e testes A/B de prebunking e accuracy prompts em formatos compatíveis com vídeo curto. Esses passos, ancorados nos resultados, apontam caminhos realistas para mitigar autodiagnósticos indevidos sem suprimir o valor do apoio entre pares no ecossistema informacional contemporâneo (McCombs; Shaw, 1972; Entman, 1993; Noelle-Neumann, 1974; Cinelli et al., 2021).

## CONSIDERAÇÕES FINAIS

O objetivo central deste estudo foi explicar como mecanismos comunicacionais e psicológicos - mediados por arquiteturas de recomendação - contribuem para a desinformação em saúde mental e de que modo essa dinâmica estimula autodiagnósticos sem avaliação profissional. À luz dos resultados, esse objetivo foi alcançado ao demonstrar que a articulação entre saliência temática (associável ao agenda-setting), enquadramentos persuasivos (relacionados ao framing), repetição



com elevada fluência e estreitamento da diversidade informacional por algoritmos silenciadores aumenta a plausibilidade subjetiva de listas inespecíficas de sintomas e de narrativas experienciais, favorecendo a autoatribuição diagnóstica. Em termos interpretativos, trata-se de efeitos probabilísticos: ecologias de feed personalizadas, combinadas a prova social visível e formatos de vídeo ultracurto, elevam a chance de que sinais vagos sejam reinterpretados como patologia, especialmente em contextos de baixa literacia em saúde.

Os achados também permitem delimitar implicações práticas coerentes com a força da evidência. No plano clínico-assistencial, recomenda-se incluir na anamnese perguntas sobre exposição digital e autodiagnóstico, oferecendo orientações claras de quando e por que buscar avaliação profissional. No desenho de plataformas, os resultados apoiam o teste de intervenções leves no fluxo, por exemplo, prebunking e accuracy nudges/accuracy prompts, que aumentem a saliência do critério de veracidade antes do compartilhamento, ao mesmo tempo em que criam vias de referência a fontes qualificadas sem produzir reatância. No campo educacional, programas de literacia (incluindo eHealth) dirigidos a adolescentes, responsáveis e educadores podem reduzir a vulnerabilidade a conteúdos de baixa qualidade e mitigar trajetórias de autodiagnóstico.

Reconhecemos limitações que condicionam o alcance das conclusões: trata-se de revisão narrativa integrativa, com heterogeneidade de delineamentos e métricas; prevalece evidência associativa para vários desfechos; há viés linguístico para publicações em inglês; e a volatilidade de políticas e algoritmos dificulta a generalização temporal. Portanto, as inferências devem ser lidas como convergências robustas e não como demonstrações causais. Ainda assim, a consistência entre diferentes métodos, análises de conteúdo, estudos observacionais e experimentais sobre mecanismos cognitivos, sustenta o encadeamento proposto entre enquadramentos, repetição, prova social e algoritmos silenciadores.

Como agenda específica de aprofundamento, destacamos seis lacunas prioritárias: (i) harmonização de medidas para avaliação de acurácia em saúde mental, com um núcleo mínimo comparável entre estudos (p. ex., convergência entre DISCERN, JAMA benchmark e PEMAT, acrescida de itens específicos para vídeo curto); (ii) desfechos clínicos downstream, testando se a exposição repetida a conteúdos de baixa qualidade se associa a autodiagnóstico confirmado (ou não) em entrevistas estruturadas e à trajetória de busca/adesão ao cuidado; (iii) ensaios in-feed, isto é, intervenções testadas dentro do próprio fluxo de uso (short-form), com monitoramento de impacto em intenção de compartilhamento e veracidade percebida; (iv) replicações em português (Brasil e demais contextos lusófonos), reduzindo o viés anglófono e verificando validade transcultural dos efeitos e das intervenções; (v) transparência algorítmica por meio de auditorias com pré-registro e telemetria ética, capazes de estimar a contribuição efetiva da recomendação personalizada para o estreitamento informacional; e (vi) trilhas de literacia, inclusive eHealth, com módulos breves e mensuráveis para adolescentes, responsáveis e educadores, visando ganhos verificáveis de checagem de veracidade e redução de autodiagnósticos indevidos.

Em síntese, os resultados confrontados com os objetivos indicam que algoritmos silenciadores operam como peça de ligação entre dinâmicas de visibilidade e processos de crença, estreitando a janela informacional de muitos usuários e aumentando a probabilidade de autodiagnósticos indevidos. Mitigar esse risco exige combinar ação clínica, educação midiática e ajustes de arquitetura informacional,



ancorados em evidência e sensíveis às especificidades dos ambientes digitais contemporâneos.

## REFERÊNCIAS

- AROMATARIS, E.; MUNN, Z. (eds.). **JBI Manual for Evidence Synthesis**. Adelaide: JBI, 2020. DOI: 10.46658/JBIMES-20-01. Link: <https://doi.org/10.46658/JBIMES-20-01>
- BANDURA, A. Social cognitive theory of mass communication. **Media Psychology**, v. 3, n. 3, p. 265–299, 2001. DOI: 10.1207/S1532785XMEP0303\_03. Link: [https://doi.org/10.1207/S1532785XMEP0303\\_03](https://doi.org/10.1207/S1532785XMEP0303_03)
- BAETHGE, C.; GOLDBECK-WOOD, S.; MERTENS, S. SANRA: a scale for the quality assessment of narrative review articles. **Research Integrity and Peer Review**, v. 4, 5, 2019. DOI: 10.1186/s41073-019-0064-8. Link: <https://doi.org/10.1186/s41073-019-0064-8>
- BRAUN, V.; CLARKE, V. Using thematic analysis in psychology. **Qualitative Research in Psychology**, v. 3, n. 2, p. 77–101, 2006. DOI: 10.1191/1478088706qp063oa. Link: <https://doi.org/10.1191/1478088706qp063oa>
- CINELLI, M. et al. The echo chamber effect on social media. **Proceedings of the National Academy of Sciences**, v. 118, n. 9, e2023301118, 2021. DOI: 10.1073/pnas.2023301118. Link: <https://doi.org/10.1073/pnas.2023301118>
- COLLOCA, L. The nocebo effect. **Annual Review of Pharmacology and Toxicology**, v. 64, p. 171–190, 2024. DOI: 10.1146/annurev-pharmtox-022723-112425. Link: <https://doi.org/10.1146/annurev-pharmtox-022723-112425>
- SILVA, Adriano Alves da; PÔRTO JÚNIOR, Francisco Gilson Rebouças. Algoritmos silenciadores: desinformação e espiral do silêncio na era da inteligência artificial. **Organicom**, São Paulo, v. 21, n. 44, p. 147–158, 2024. DOI: 10.11606/issn.2238-2593.organicom.2024.221079. Disponível em: <https://revistas.usp.br/organicom/article/view/221079>. Acesso em: 10 jun. 2025.
- ECKER, U. K. H. et al. The psychological drivers of misinformation belief and its resistance to correction. **Nature Reviews Psychology**, v. 1, p. 13–29, 2022. DOI: 10.1038/s44159-021-00006-y. Link: <https://doi.org/10.1038/s44159-021-00006-y>
- ENTMAN, R. M. Framing: Toward clarification of a fractured paradigm. **Journal of Communication**, v. 43, n. 4, p. 51–58, 1993. DOI: 10.1111/j.1460-2466.1993.tb01304.x. Link: <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- FAZIO, L. K. et al. Knowledge does not protect against illusory truth. **Journal of Experimental Psychology: General**, v. 144, n. 5, p. 993–1002, 2015. DOI: 10.1037/xge0000098. Link: <https://doi.org/10.1037/xge0000098>
- HASLAM, N. Concept creep: Psychology's expanding concepts of harm and pathology. **Psychological Inquiry**, v. 27, n. 1, p. 1–17, 2016. DOI: 10.1080/1047840X.2016.1082418. Link: <https://doi.org/10.1080/1047840X.2016.1082418>



HASLAM, N. Concept creep and psychiatrization. **Frontiers in Sociology**, v. 6, 806147, 2021. DOI: 10.3389/fsoc.2021.806147. Link: <https://doi.org/10.3389/fsoc.2021.806147>

HONG, Q. N. et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. **Education for Information**, v. 34, n. 4, p. 285–291, 2018. DOI: 10.3233/EFI-180221. Link: <https://doi.org/10.3233/EFI-180221>

LOOKINGBILL, V.; MOHAMMADI, E.; CAI, Y. Assessment of accuracy, user engagement, and themes of eating disorder content in social media short videos. **JAMA Network Open**, v. 6, n. 4, e238897, 2023. DOI: 10.1001/jamanetworkopen.2023.8897. Link: <https://doi.org/10.1001/jamanetworkopen.2023.8897>

MCCOMBS, M. E.; SHAW, D. L. The agenda-setting function of mass media. **Public Opinion Quarterly**, v. 36, n. 2, p. 176–187, 1972. DOI: 10.1086/267990. Link: <https://doi.org/10.1086/267990>

NOELLE-NEUMANN, E. The spiral of silence: A theory of public opinion. **Journal of Communication**, v. 24, n. 2, p. 43–51, 1974. DOI:10.1111/j.1460-2466.1974.tb00367.x. Link: <https://doi.org/10.1111/j.1460-2466.1974.tb00367.x>

NORMAN, C. D.; SKINNER, H. A. eHEALS: The eHealth Literacy Scale. **Journal of Medical Internet Research**, v. 8, n. 4, e27, 2006. DOI: 10.2196/jmir.8.4.e27. Link: <https://doi.org/10.2196/jmir.8.4.e27>

NUTBEAM, D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies in the 21st century. **Health Promotion International**, v. 15, n. 3, p. 259–267, 2000. DOI: 10.1093/heapro/15.3.259. Link: <https://doi.org/10.1093/heapro/15.3.259>

PAGE, M. J. et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. **BMJ**, v. 372, n71, 2021. DOI: 10.1136/bmj.n71. Link: <https://doi.org/10.1136/bmj.n71>

PENNYCOOK, G.; RAND, D. G. Shifting attention to accuracy can reduce misinformation online. **Nature**, v. 592, p. 590–595, 2021. DOI: 10.1038/s41586-021-03344-2. Link: <https://doi.org/10.1038/s41586-021-03344-2>

PENNYCOOK, G.; RAND, D. G. Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. **Nature Communications**, v. 13, 2333, 2022. DOI: 10.1038/s41467-022-30073-5. Link: <https://doi.org/10.1038/s41467-022-30073-5>

SHEA, B. J. et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. **BMJ**, v. 358, j4008, 2017. DOI: 10.1136/bmj.j4008. Link: <https://doi.org/10.1136/bmj.j4008>

STARCEVIC, V.; BERLE, D. Cyberchondria: Towards a better understanding of excessive health-related Internet searching. **Expert Review of Neurotherapeutics**, v. 13, n. 2, p. 205–213, 2013. DOI: 10.1586/ern.12.162. Link: <https://doi.org/10.1586/ern.12.162>



VAN DER LINDEN, S.; LEWANDOWSKY, S.; ECKER, U. K. H. Inoculating the public against misinformation. **Global Challenges**, v. 1, n. 2, 1600008, 2017. DOI: 10.1002/gch2.201600008. Link: <https://doi.org/10.1002/gch2.201600008>

VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. **Science**, v. 359, n. 6380, p. 1146–1151, 2018. DOI: 10.1126/science.aap9559. Link: <https://doi.org/10.1126/science.aap9559>

YEUNG, A. et al. TikTok and attention-deficit/hyperactivity disorder: A cross-sectional study of social media content quality. **The Canadian Journal of Psychiatry**, v. 67, n. 12, p. 899–906, 2022. DOI: 10.1177/07067437221082854. Link: <https://doi.org/10.1177/07067437221082854>

---

### INFORMAÇÕES AUTORAIS

1. **Adriano Alves da Silva**, acadêmico de Psicologia, Centro Universitário Luterano de Palmas (CEULP/ULBRA).
2. **Élida Souza Silva**, acadêmica de Psicologia, Centro Universitário Luterano de Palmas (CEULP/ULBRA).
3. **Mayelle Batista da Silva**, Psicóloga, Especialista, Docente em Psicologia, Centro Universitário Luterano de Palmas (CEULP/ULBRA).