

O Fenômeno da Desinformação no WhatsApp, Telegram e Outras Plataformas de Mensagens Instantâneas: Uma Revisão Sistemática da Literatura

João Francisco Hecksher Olivetti, Ana Flávia Bomfim, Marcos Oliveira,
Helom Felipe Marques, Juan Avelar, Julio C. S. Reis, Philipe Melo
joao.olivetti@ufv.br, ana.bomfim@ufv.br, matto.m@outlook.com.br
helom.alves@ufv.br, juan.avelar@ufv.br, jreis@ufv.br, philipe.freitas@ufv.br
Universidade Federal de Viçosa - UFV
Florestal/Viçosa, MG, Brasil

ABSTRACT

This literature review examines the current state of research on misinformation within Instant Messaging Platforms (IMPs) such as WhatsApp, Telegram and WeChat. By conducting comprehensive searches across multiple repositories, we identified 110 relevant studies published between 2018 and 2025, covering more than 20 countries and 7 distinct messaging platforms. These works were systematically analyzed to: (i) map internal and external factors of IMPs that facilitate the creation and spread of misinformation; (ii) identify its main impacts and negative consequences reported to our society; and (iii) examine strategies proposed or implemented to mitigate and to combat misinformation within this environment. Our findings reveal recurrent themes across studies employing diverse methodologies, indicating a growing convergence in this still-fragmented research field, while also exposing points of disagreement. The analysis highlights how specific architectural features make IMPs uniquely susceptible to misinformation. Unlike mainstream social networks, their information flow is governed directly by user interactions rather than feed algorithms. In this context, end-to-end encryption, high virality, and limited moderation or fact-checking emerge as internal components that sustain the misinformation ecosystem. Externally, factors such as low digital literacy, social polarization, and political bias further reinforce the problem. Notably, some elements such as distrust in institutions and political polarization are reported both as causes and as consequences of misinformation, suggesting a potential self-reinforcing cycle. Crucially, our review shed light on a paradoxical communication architecture in IMPs, where encrypted and seemingly personal messages coexist seamlessly with large-scale, viral dissemination. This duality blurs the distinction for users between intimate communication and mass circulation, making it harder to assess the origin and credibility of information.

KEYWORDS

RSL, Revisão Sistemática da Literatura, Desinformação, *fake news*, Plataformas de mensagem instantâneas, WhatsApp, Telegram

1 INTRODUÇÃO

A desinformação é um fenômeno antigo, anterior à própria Internet [31], mas, com a evolução das tecnologias informacionais contemporâneas, foi ampliado exponencialmente seu alcance e impacto [102]. A pandemia da COVID-19 evidenciou de maneira ainda mais expressiva esse problema, com a proliferação de boatos e teorias conspiratórias sobre a doença, que causaram medo na população. Isso levou ao que a Organização Mundial da Saúde (OMS) chamou de *infodemia* [123], uma sobrecarga informacional onde conteúdos imprecisos ou falsos se disseminam rapidamente, dificultando respostas coordenadas e provocando consequências sociais e sanitárias [19].

No contexto digital atual, as **Plataformas de Mensagens Instantâneas** (PMIs) como WhatsApp, Telegram e WeChat, ocupam um papel central na comunicação cotidiana, com bilhões de usuários ao redor do mundo. A chegada dessas plataformas, embora facilite e agilize a comunicação interpessoal, também tem sido apontada como vetor central na disseminação de desinformação no país [26, 106] e no mundo. Essa constatação já levou inclusive a alterações técnicas pelas empresas responsáveis, como a implementação de limites no encaminhamento de mensagens no WhatsApp¹ visando reduzir a viralidade de conteúdos potencialmente falsos.

A relevância única das PMIs, em especial do WhatsApp, para a circulação da desinformação se deve tanto a fatores técnicos quanto sociais. Diferentemente de redes sociais tradicionais, que operam com algoritmos de recomendação pública e visibilidade ampla, as PMIs funcionam com base em relações interpessoais, muitas vezes privadas e criptografadas, tornando o monitoramento e a verificação de conteúdos muito mais desafiador [68, 101]. Além disso, essas plataformas passaram por um processo de popularização massiva no Brasil, substituindo com rapidez antigos meios de comunicação como SMS e e-mail. Por dependerem apenas de conexão à Internet, e não de créditos pagos por mensagem, elas democratizaram o acesso à comunicação síncrona digital em larga escala. Isso criou um ecossistema comunicacional ímpar, que hoje domina o território nacional com alta capilaridade inclusive entre populações de menor renda e de faixas etárias que outras redes não atingem [30].

O tema do espalhamento de conteúdo enganoso tem se tornado cada vez mais relevante em escala global, sendo inclusive apontado pela UNESCO como um dos principais desafios globais para a sociedade em 2025 [67]. A Organização para Cooperação e Desenvolvimento Econômico (OCDE) aponta a desinformação como

¹<https://blog.whatsapp.com/more-changes-to-forwarding>

um desafio de alto impacto para diversas áreas na política pública, incluindo saúde, segurança nacional e confiança em instituições democráticas [87]. Porém, apesar da conhecida emergência das PMIs como meios centrais neste contexto desinformativo, ainda há uma carência de consolidação robusta do conhecimento científico quanto ao papel que essas plataformas exercem na disseminação de desinformação. A literatura científica permanece ainda fragmentada e, muitas vezes, conduzidos de forma desarticulada, dificultando uma melhor compreensão dos riscos envolvidos, dos impactos causados por essas plataformas e, consequentemente, também do estabelecimento de meios de mitigação do problema.

Além disso, é comum que pesquisas agrupem plataformas como WhatsApp e Telegram simplesmente como redes sociais tradicionais, ignorando diferenças estruturais e sociotécnicas fundamentais entre esses ambientes. Essa questão se agrava num cenário onde ainda há utilização inconsistente de termos e conceitos importantes como *misinformation*, *disinformation* e *fake news*, frequentemente tratados como intercambiáveis, o que compromete a comparabilidade entre os estudos e dificulta a consolidação do campo [11].

Desta forma, neste trabalho realiza-se uma **Revisão Sistemática da Literatura** (RSL) dedicada exclusivamente a investigação da desinformação no ecossistema das Plataformas de Mensagens Instantâneas. O objetivo é reunir e organizar o conhecimento disperso sobre o tema, identificando padrões, lacunas e direções promissoras de pesquisa com os principais desafios e oportunidades sobre o tema. Adotamos, ao longo do texto, o termo “desinformação” de forma ampla, referindo-se a conteúdos falsos ou enganosos que circulam nesses meios e que são objeto de investigação nos estudos, sem pretensão de estabelecer uma definição final para o conceito.

O restante deste artigo está organizado da seguinte forma: na Seção 2, contextualizamos o tema e discutimos trabalhos correlatos. A Seção 3 apresenta a metodologia utilizada na condução da revisão sistemática. Em seguida, na Seção 4, descrevemos os resultados encontrados, organizados pelas questões de pesquisa. A Seção 5 discute os principais achados e limitações do estudo apresentando a taxonomia proposta sobre desinformação em PMIs. Por fim, a Seção 6 apresenta as considerações finais e sugestões para trabalhos futuros.

2 TRABALHOS RELACIONADOS

O fenômeno da desinformação tem atraído uma crescente atenção acadêmica, especialmente a partir da pandemia de COVID-19, sendo abordado por distintas áreas do conhecimento e sob múltiplas perspectivas analíticas. De maneira fundacional, uma parcela da literatura tem se dedicado à definição conceitual do problema, buscando distinguir termos como “*fake news*”, “*misinformation*” e “*disinformation*”. Um exemplo é o trabalho de [33], que identifica uma, já mencionada, falta de padronização no uso dos termos, que apresenta diversas inconsistências. Como resultado, eles definem desinformação como “informações equivocadas cujo propósito é enganar”, destacando o fenômeno a partir de um caráter intencional e até malicioso, sendo menos sobre a simples disposição de informações incorretas a cerca de um tema. Apesar destes esforços taxonômicos, a aplicação desses conceitos na literatura científica ainda ocorre de maneira heterogênea, comprometendo a consistência entre os estudos [11].

Dada a relevância do tema, diversos esforços, incluindo outras Revisões Sistemáticas da Literatura (RSLs), têm sido realizados para tratar deste problema. O estudo de [8] apresenta uma revisão ampla sobre desinformação em mídias sociais, abordando desafios técnicos e estratégias de prevenção. Da mesma forma, o trabalho de [60] foca em populações específicas, analisando a disseminação da desinformação entre membros da comunidade negra. Já a revisão de [96] oferece uma visão abrangente do campo ao analisar 756 publicações entre 2014 e 2022. A partir dessas análises, os autores mapeiam definições, canais, agentes de disseminação e lacunas de pesquisa, destacando como o tópico da desinformação ameaça não apenas a verdade factual, mas também a coesão democrática em contextos marcados por populismo, nacionalismo e crises globais.

Embora esses trabalhos contribuam para o entendimento geral do fenômeno, eles não focam nas particularidades do ecossistema informativo específico de Plataformas de Mensagens Instantâneas (PMIs). Estudos como os de [2, 127] abordam a desinformação em redes sociais digitais, mas o fazem sob uma lente que agrupa indiscriminadamente plataformas como Twitter, Instagram, YouTube e WhatsApp. Essa abordagem ignora aspectos cruciais das PMIs, como sua natureza criptografada, o caráter fechado das conversas, e as limitações técnicas para moderação e checagem de conteúdo.

Por outro lado, existe um volume crescente de artigos focados em PMIs, já existindo inclusive revisões da literatura associada a eles, especialmente em relação ao WhatsApp. Estas, no entanto, não abordam de maneira aprofundada questões de desinformação, sendo mais voltadas para cenários que exploram o uso destas mídias sociais em ambientes de aprendizagem/ensino [85] ou mesmo em de aplicações em chatbots [98].

3 METODOLOGIA

Esta pesquisa compreende a realização de uma Revisão Sistemática da Literatura (RSL) com o objetivo de identificar, organizar e sintetizar o conhecimento disponível sobre o recente fenômeno da desinformação em PMIs. A opção metodológica por seguir uma RSL se justifica, como mencionado nas seções anteriores, pela necessidade de reunir as abundantes, porém dispersas evidências científicas relacionadas ao tema. Dessa maneira, avaliando criticamente os estudos preexistentes e promovendo uma visão consolidada sobre as particularidades dessas plataformas no ecossistema da desinformação. Assim, a metodologia adotada segue as diretrizes propostas por Kitchenham e Charters [59], reconhecidas na área de Ciência da Computação como referência para a condução de revisões sistemáticas. De forma complementar, utilizamos as recomendações do PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) [90] para organizar os resultados e, assim, garantir a transparência, reprodutibilidade e qualidade do relato.

Mais especificamente, esta revisão foi realizada entre janeiro e junho de 2025, por uma equipe composta por 5 pesquisadores independentes, divididos em subgrupos de trabalho para garantir múltiplas visões nas etapas de busca, triagem e análise dos estudos. A Figura 1 apresenta uma visão geral das etapas realizadas, organizadas em sete fases principais: (i) definição das questões de pesquisa, (ii) construção das expressões de busca; (iii) definição dos critérios de inclusão e exclusão (iv) seleção dos repositórios e bases

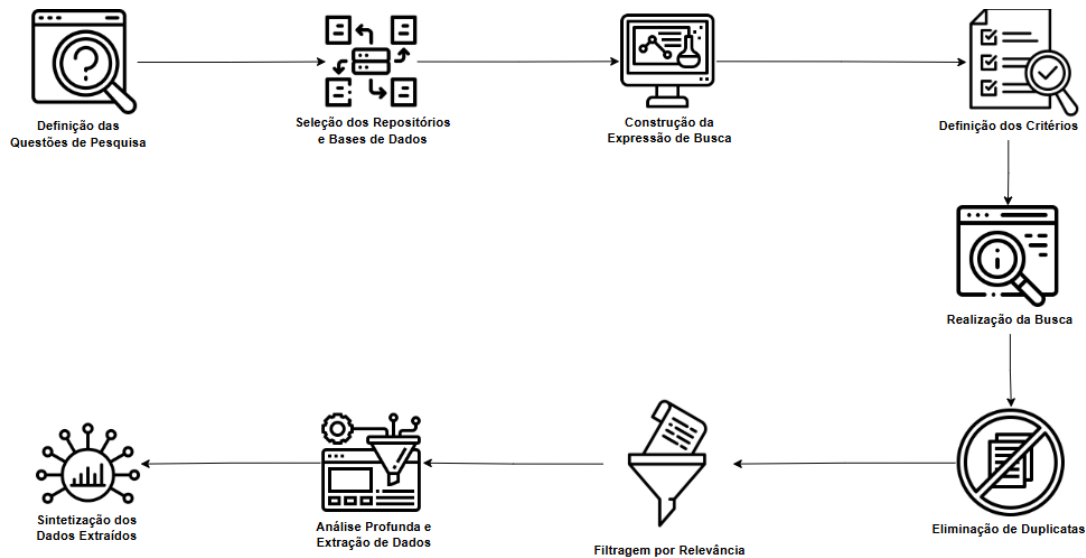


Figura 1: Diagrama geral das etapas metodológicas da Revisão Sistemática da Literatura seguidas pelo estudo.

de artigos (v) recuperação e consolidação dos estudos, com eliminação de duplicatas; (vi) triagem por critérios de inclusão e exclusão; e (vii) leitura analítica e categorização dos estudos selecionados.

As seções seguintes detalham as etapas ilustradas pelo fluxograma na Figura 2, incluindo as estratégias de busca utilizadas, os critérios de elegibilidade adotados, os procedimentos de extração e análise dos dados, bem como as ações tomadas para mitigar possíveis vieses e fortalecer a confiabilidade dos resultados. O material complementar e os artefatos com os metadados gerados por meio desta metodologia estão disponíveis no repositório do projeto ².

3.1 Definição das Questões de Pesquisa

A primeira etapa metodológica consiste na formulação de questões de pesquisa que direcionem bem todas as demais fases do processo, desde a construção das expressões de busca até a análise dos resultados. Essa prática está em conformidade com as diretrizes do PRISMA, que recomendam a explicitação clara e fundamentada dos objetivos e das perguntas investigadas, de modo a garantir transparência, foco temático e reprodutibilidade da revisão [90].

Neste trabalho, as questões de pesquisa foram formuladas com base na lacuna identificada de que embora haja uma produção crescente sobre desinformação em plataformas digitais, observamos a ausência de uma sistematização dedicada especificamente nas particularidades das PMIs. Dessa forma, investigamos os fatores que contribuem para o espalhamento da desinformação nesse ecossistema. Portanto, as seguintes questões guiam esta revisão:

Essas questões foram definidas de modo a cobrir de forma ampla os principais eixos temáticos identificados sobre o tema: os vetores do fenômeno (QP1), seus impactos e consequências (QP2) e as possíveis ações de respostas (QP3). Essa etapa é crítica e serve de base para a seleção dos estudos, a categorização das evidências e a estrutura da análise apresentada nas seções seguintes.

QUESTÕES DE PESQUISA (QP)

- **QP1:** Quais características internas (sociotécnicas) e externas (contextuais, regulatórias, culturais) às plataformas são apontadas como facilitadoras da disseminação de desinformação?
- **QP2:** De que maneiras a desinformação em Plataformas de Mensagens Instantâneas afeta a sociedade, em termos individuais, coletivos e institucionais?
- **QP3:** Quais são os principais mecanismos e abordagens propostos na literatura para a detecção, contenção ou combate à desinformação nesse ecossistema?

3.2 Seleção dos Repositórios e Bases de Dados

A seleção dos repositórios de artigos é uma etapa importante do processo, pois impacta diretamente a abrangência e a diversidade do material recuperado. Para este estudo, foi definido como foco principal trabalhos mais voltados para áreas computacionais e tecnológicas. Assim, foram escolhidas bases de dados científicas amplamente reconhecidas e complementares entre si, considerando critérios como relevância acadêmica, qualidade e acesso a artigos que fossem revisados por pares.

A escolha por múltiplas bases com diferentes escopos permite uma recuperação mais panorâmica de publicações relevantes, minimizando o risco de viés de indexação e maximizando a cobertura do estado da arte. As bases selecionadas foram:

- **ACM Digital Library**³: principal base de dados da área de Computação e Tecnologia, com grande volume de artigos sobre mídias digitais, redes sociais e desinformação, identificado como central para o nosso estudo;

²<https://github.com/Phlop/WhatsAppMisInfo>

³<https://dl.acm.org/>

- **IEEE Xplore**⁴: repositório de grande relevância de publicações técnico-científicas em engenharia elétrica, ciência da computação e tecnologias de comunicação;
- **Periódicos CAPES**⁵: portal brasileiro que agrega conteúdo de diversas bases nacionais e internacionais, garantindo acesso integrado à produção de diversas áreas. Sua inclusão permite uma maior variedade de resultados.
- **PubMed**⁶: base de dados especializada em literatura biomédica e de saúde, com relevância para estudos sobre desinformação em contextos de saúde pública. Mesmo sendo de outra área, foi adicionado devido a quantidade de artigos relevantes publicados neste campo, especialmente após o período da pandemia de COVID-19.
- **Google Scholar**⁷: ferramenta de busca ampla que permite complementar as demais bases com uma literatura alternativa, em periódicos e anais diferentes e, principalmente, de trabalhos interdisciplinares. Sendo assim, se torna um ótimo repositório para identificar trabalhos de grande relevância que eventualmente não foram recuperados a partir das buscas realizadas nas demais bases.

3.3 Construção da Expressão de Busca

A construção da *string* de busca representa uma etapa central em RSLs, uma vez que determina a abrangência, a clareza/transparência e a reprodutibilidade do processo de identificação de estudos relevantes. Nesta etapa, também seguimos as recomendações PRISMA [90]. A formulação da busca foi diretamente guiada pelas QPs previamente estabelecidas (i.e., QP1, QP2 e QP3). Inicialmente, foi realizada uma etapa exploratória com o objetivo de identificar os principais termos, sinônimos e variações semânticas comumente utilizados sobre o tema. Esta análise preliminar orientou a construção das expressões booleanas de busca, que buscaram combinar duas dimensões principais: 1) **Fenômeno** – termos relacionados a desinformação, como *misinformation*, *disinformation*, *fake news*, e; 2) **Canal**: nomes de PMIs específicas (e.g., WhatsApp, Telegram, WeChat) e termos genéricos como *instant messaging*, *messaging services*, *messaging platforms*.

As expressões de busca foram adaptadas conforme as particularidades sintáticas e operacionais de cada base de dados utilizada. Além disso, adotou-se a filtragem por título, resumo e palavras-chave nos campos disponíveis de cada repositório, a fim de maximizar a relevância dos resultados. A *string* de busca final é:

STRING DE BUSCA

(misinformation OR disinformation OR infodemic OR fake news) AND (WhatsApp OR Telegram OR WeChat OR 'Instant Message' OR 'Messaging Service' OR 'Messaging Platform' OR 'Messaging System')

3.3.1 Primeira Etapa de Triagem. A primeira etapa de coleta consistiu portanto na execução de buscas diretas nas cinco bases científicas

⁴<https://ieeexplore.ieee.org/>

⁵<https://www.periodicos.capes.gov.br/>

⁶<https://pubmed.ncbi.nlm.nih.gov/>

⁷<https://scholar.google.com.br/>

Tabela 1: Critérios de inclusão.

#	Critério de Inclusão
I01	Foram publicados a partir de 2016;
I02	Abordam diretamente o fenômeno da desinformação, fake news ou misinformation;
I03	Focam em plataformas de mensagens instantâneas específicas
I04	Apresentam os termos-chave no título, resumo ou palavras-chave;
I05	Utilizam dados empíricos, bases de dados ou métodos computacionais para análise;
I06	Estão escritos em inglês ou português;
I07	Estão disponíveis em texto completo para leitura e análise;
I08	São artigos acadêmicos ou científicos publicados e revisados.

Tabela 2: Critérios de exclusão.

#	Critério de Exclusão
E01	Publicados antes de 2016
E02	Texto completo indisponível para acesso
E03	Redigidos em idiomas diferentes de inglês ou português
E04	Mencionam desinformação apenas de forma superficial ou tangencial
E05	Tratam redes sociais apenas de maneira genérica, sem especificar plataformas
E06	Focam exclusivamente em contextos educacionais, médicos ou jurídicos (não computacionais nem comunicacionais)
E07	Foco centrado apenas no comportamento do usuário, sem análise do conteúdo da desinformação
E08	Não abordam mídias sociais ou plataformas de mensagens instantâneas

selecionadas, com o objetivo de recuperar o maior número possível de estudos potencialmente relevantes. Para isso, foi utilizada a *string* de busca definida, entre janeiro e fevereiro de 2025, respeitando a sintaxe (e.g., operadores lógicos aceitos) e as particularidades de cada base, dado que algumas permitiam aplicação antecipada de filtros por data de publicação, tipo de documento (por exemplo, artigos revisados por pares), e campos específicos como título, resumo e palavras-chave. Outras, como o Google Acadêmico, apresentam uma estrutura mais livre, portanto nos restringimos somente aos primeiros 120 resultados retornados para maior controle e equilíbrio com as demais bases. Ao final dessa etapa, obtivemos um total bruto de **3.837 registros** para o início das etapas de filtragem.

3.4 Definição dos Critérios

Para garantir relevância dos estudos selecionados nesta revisão sistemática para o tema proposto, nesta próxima etapa foram definidos critérios de inclusão e exclusão de trabalhos, com base nas recomendações metodológicas do PRISMA e nas diretrizes de Kitchenham e Charters [59]. Esses critérios foram aplicados de forma sistemática pelos cinco rotuladores durante as fases de triagem e filtragem, com o intuito de selecionar apenas os estudos que abordassem de maneira substantiva o fenômeno da desinformação em plataformas de mensagens instantâneas.

Os critérios de inclusão apresentados na Tabela 1 consideram aspectos formais mais gerais (como ano de publicação e idioma), além de critérios conceituais, como o foco explícito na desinformação e o uso de métodos empíricos ou computacionais para análise e

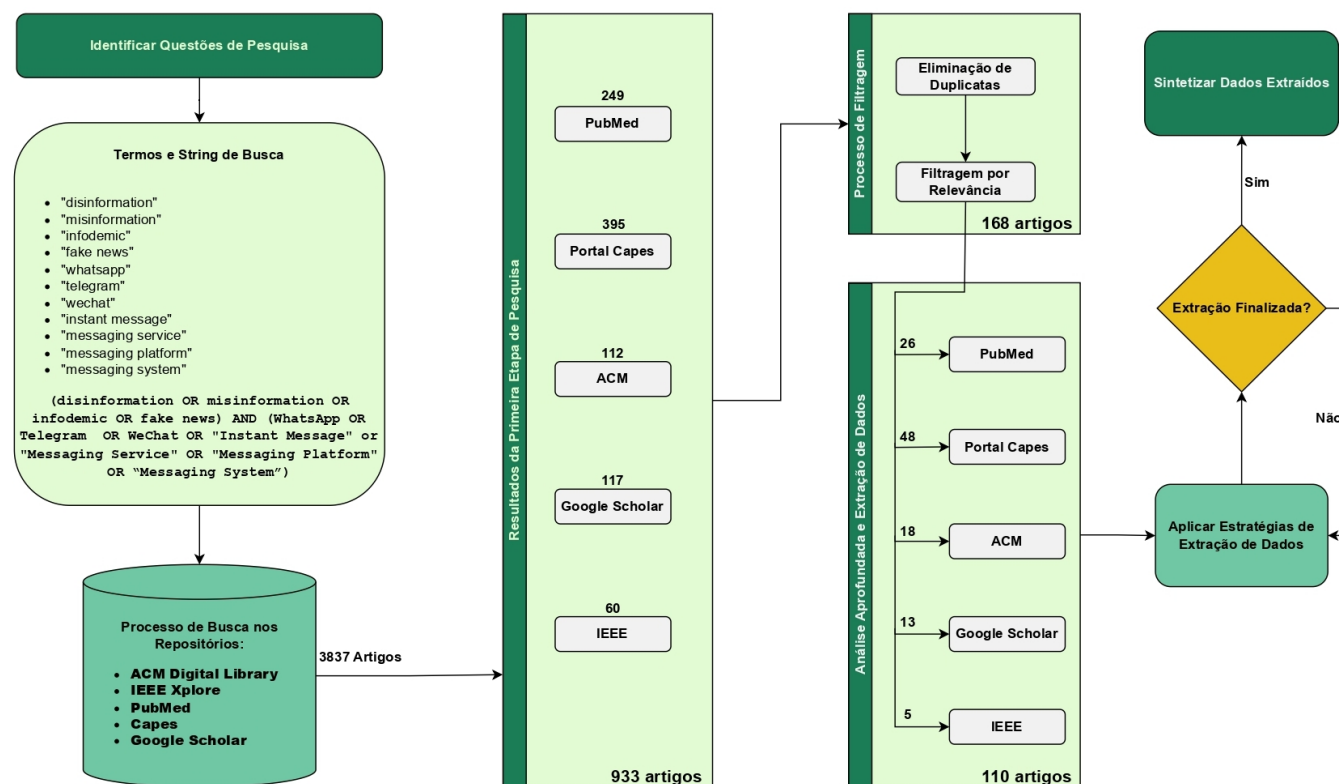


Figura 2: Fluxograma com a execução de cada etapa da metodologia da revisão sistemática.

interpretação dos dados coletados. Além disso, é crucial definir um critério que exigisse a inclusão de alguma plataforma de mensagem (e.g., WhatsApp, Telegram, WeChat, LINE, Viber, Signal) dado o objetivo do trabalho. A escolha do recorte temporal a partir de 2016 se justifica pela intensificação do uso deste tipo de plataforma após esse período, conforme apontado por estudos sobre sua popularização global e crescente papel na circulação de informações [61, 68].

Já os critérios de exclusão descritos na Tabela 2 foram elaborados para eliminar trabalhos que, embora tangenciem o tema de alguma maneira, não se enquadram com os objetivos desta revisão. Foram excluídos estudos genéricos, com foco superficial ou indireto sobre a desinformação (E04), bem como aqueles que não especificam plataformas digitais (E05) ou abordam contextos muito distantes das plataformas sociais, como educação ou medicina, sem articulação computacional ou comunicacional clara (E06). Além disso, também foram desconsiderados estudos que não atendiam ao recorte temporal definido (E01), que estavam indisponíveis para leitura completa (E02) ou que não estavam escritos nos idiomas de interesse (E03).

3.4.1 Segunda Etapa de Triagem. Concluída a coleta inicial dos resultados brutos provenientes das cinco bases de dados, a segunda etapa consistiu na aplicação preliminar dos filtros dos critérios de inclusão e exclusão (Tabelas 1 e 2) por meio da leitura exclusivamente dos títulos de todos os trabalhos recuperados, realizada manualmente por uma equipe de seis pesquisadores, que atuaram como rotuladores independentes. Esta etapa, devido ao grande volume de dados, teve como objetivo inicial remover os trabalhos mais

claramente fora do escopo da revisão. Como resultado, foram selecionados **933 artigos**, distribuídos entre as bases da seguinte forma: PubMed (249), Capes (395), ACM Digital Library (112), Google Scholar (117) e IEEE Xplore (60), conforme ilustrado pela Figura 2. Essa filtragem representou uma redução significativa do conjunto inicial, com artigos mais alinhados com os objetivos da revisão.

3.5 Processo de Filtragem

Após a seleção inicial, iniciamos o processo de filtragem refinada com o objetivo de garantir que apenas artigos pertinentes, únicos e completos fossem mantidos para análise detalhada. Essa etapa, a mais longa do processo, envolveu múltiplas rodadas sucessivas de filtragem e avaliação, conforme descrito a seguir.

3.5.1 Eliminação de Duplicatas. Inicialmente nesta etapa, foi realizada a remoção de duplicatas, considerando que muitos artigos estavam indexados em mais de uma das bases de dados consultadas. Esse procedimento foi essencial para consolidar o conjunto de artigos de interesse e garantir a unicidade dos registros.

3.5.2 Terceira Etapa de Triagem. Em seguida, os títulos, resumos e palavras-chave das centenas de artigos pré-selecionados foram minuciosamente analisados. Essa etapa visou remover estudos que, embora tivessem passado pelas triagens anteriores, apresentavam formatos inadequados (como *preprints* não revisados), não eram artigos completos ou apresentavam escopo impreciso. De maneira

notável, nenhum artigo foi excluído por estar fora do período estipulado (anterior a 2016). Dada a grande quantidade e diversidade de publicações sobre o tema e o objetivo de centralizar os estudos mais significativos, os rotuladores conduziram uma avaliação mais rigorosa e crítica dos metadados, verificando a relevância de cada estudo às questões de pesquisa da revisão. Ao final desta etapa, o conjunto foi reduzido a **168 artigos**.

O passo seguinte desta triagem teve como foco a validação definitiva dos artigos selecionados. Aqui os artigos foram então lidos em sua totalidade e reavaliados com base em sua conformidade com os critérios definidos nas Tabelas 1 e 2, sua contribuição efetiva ao escopo da revisão e sua capacidade de responder às questões de pesquisa formuladas (Seção 3.1). Como resultado, **110 artigos** foram mantidos, distribuídos da seguinte forma: PubMed (26), Capes (48), ACM Digital Library (18), Google Scholar (13) e IEEE Xplore (5).

3.6 Análise Aprofundada e Extração de Dados

A etapa final da revisão consistiu em uma análise minuciosa dos 110 artigos selecionados, com o objetivo de encontrar evidências e padrões temáticos capazes de responder às questões de pesquisa estabelecidas. Para isso, foram definidos algumas características a serem extraídas, com base em dados recorrentes observados durante a leitura inicial dos trabalhos: (i) **Características internas das plataformas**: aspectos sociotécnicos das PMIs que favorecem a disseminação da desinformação; (ii) **Fatores sociais e externos**: elementos contextuais fora do controle direto das plataformas que impulsionam o compartilhamento de conteúdo falso; (iii) **Impactos sociais da desinformação**: consequências observadas ou discutidas nos estudos; (iv) **Soluções e estratégias de combate**: propostas técnicas, políticas ou educativas para mitigação do problema, incluindo o uso de inteligência artificial, parcerias e políticas públicas;

Vale destacar que essa fase de extração de dados foi realizada de maneira colaborativa e iterativa, nas quais diferentes membros da equipe de pesquisa trabalhavam o mesmo artigo, na tentativa de assegurar que o máximo de informações fosse extraídas do mesmo e que os critérios de Inclusão e Exclusão fossem cumpridos.

3.7 Limitações

Em relação às limitações desta revisão, é importante destacar a exclusão de estudos que não estavam disponíveis em inglês. Embora o inglês seja o idioma predominante nas publicações científicas, essa escolha pode ter limitado o acesso a pesquisas relevantes produzidas em países de outras regiões. Ainda assim, a análise demonstra que nosso estudo capturou uma grande parte da produção científica sobre desinformação em PMIs em diversos países, inclusive que não falam inglês. No que se refere aos procedimentos metodológicos da revisão, optamos por uma divisão do trabalho entre diferentes membros da equipe, de forma a reduzir vieses individuais. Contudo, essa estratégia pode ter introduzido variações na interpretação e classificação dos artigos. Para mitigar esse risco, o processo foi acompanhado por momentos de validação após cada etapa, em que os revisores chegaram a um consenso, além de uma etapa iterativa de revisão cruzada.

4 RESULTADOS

Após a aplicação da metodologia descrita anteriormente, 110 artigos foram selecionados para compor o conjunto final de estudos

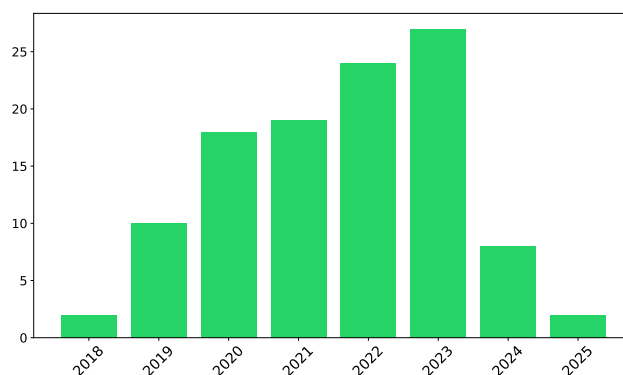


Figura 3: Total de artigos incluídos por ano de publicação.

desta revisão. Esses estudos compreendem os períodos entre 2018 e 2025, representando diferentes abordagens, contextos geográficos e metodologias voltadas à análise da desinformação em Plataformas de Mensagens Instantâneas (PMIs). A Tabela 3 apresenta a lista completa dos artigos selecionados, acompanhada de seus principais metadados (i.e., título, autores, local de publicação e repositório). Nela é possível observar, já nos títulos das pesquisas, que o termo *misinformation* aparece 41 vezes, enquanto *fake news* tem 26 ocorrências e *disinformation* 9. A tabela com mais informações sobre cada artigo está disponível no **GitHub**⁸.

4.1 Evolução Temporal

Como ponto de partida, realizamos uma análise temporal da produção científica selecionada, com o objetivo de observar a evolução do interesse pelo tema ao longo dos anos. A Figura 3 apresenta a distribuição dos artigos por ano de publicação, permitindo visualizar tendências e eventos relevantes no campo. A partir desta retrospectiva, notamos o crescimento das publicações a partir de 2018, com um pico notável entre 2020 e 2023, durante o período da pandemia de COVID-19. Observa-se que, apesar do critério de inclusão permitir artigos publicados a partir de 2016, os dois estudos mais antigos encontrados/selecionados foram publicados apenas em 2018, demonstrando que o tema é relativamente recente e como a intensificação do debate sobre desinformação em PMIs aconteceu somente após esse período. Ambos representam estudos realizados na Índia, com foco em modelagem da desinformação, e inclusão do WhatsApp em sua linha de pesquisa [4, 61].

Além disso, dentre os estudos analisados, 31 mencionam diretamente o termo “COVID” em seus títulos, coincidindo também com o período com maior número de artigos, revelando a importância da pandemia como evento catalisador de investigações científicas sobre o tema. Esses trabalhos abordam desde a caracterização e os impactos da desinformação em contextos nacionais [9, 28, 51] até sua detecção e combate com ferramentas computacionais [35, 55].

A observação temporal também mostra que, apesar de uma queda após o período pandêmico, este campo de estudo continuou ativo, com 2024 sendo representado em oito artigos e 2025, ano de realização desta revisão, sendo representado em dois. Neste par de pesquisas recentes, encontram-se propostas para o combate da desinformação através de chatbots em grupos privados no LINE [71]

⁸<https://github.com/Phlop/WhatsAppMisInfo>

Tabela 3: Lista resultante dos 110 trabalhos avaliados e selecionados após aplicação das etapas de filtragem.

Citação	Título	Autores	Local de Publicação	Repositório
[129]	"Fake or not, I'm sharing it": teen perception about disinformation in social network	Zozaya-Durazo et al.	Young Consumers - Emerald Insights	ACM
[74]	A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections.	Machado et al.	ACM Web Conference	ACM
[119]	Accost, Accede, or Amplify: Attitudes towards COVID-19 Misinformation on WhatsApp in India	Varanasi et al.	Conference on Human Factors in Computing Systems	ACM
[106]	Analyzing Textual (Mis)Information Shared in WhatsApp Groups	Resende et al.	Web Science Conference	ACM
[3]	Conversational Agents to Facilitate Deliberation on Harmful Content in WhatsApp Groups	Agarwal et al.	Human-Computer Interaction	ACM
[104]	Helping Fact-Checkers Identify Fake News Stories Shared through Images on WhatsApp	Reis et al.	WebMedia Conference	ACM
[71]	Countering Misinformation in Private Messaging Groups: Insights From a Fact-checking Chatbot	Lee and Fussell	Human-Computer Interaction	ACM
[23]	IARA - An Architectural Model to Assist the Development of Advising Bots for Misinformation Detection	Cacabro et al.	WebMedia Conference	ACM
[95]	WhatsApp and false information: a value-oriented evaluation	Puska et al.	Brazilian Symposium on Human Factors in Computing Systems	ACM
[47]	On the Globalization of the QAnon Conspiracy Theory Through Telegram	Hosini et al.	Web Science Conference 2023	ACM
[114]	Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps?	Sundar et al.	Journal of Computer-Mediated Communication	ACM
[103]	Supervised Learning for Misinformation Detection in WhatsApp	Reis and Benevenuto	Brazilian Symposium on Multimedia and the Web	ACM
[107]	(Mis)information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures	Resende et al.	The World Wide Web Conference	ACM
[24]	Characterizing Attention Cascades in WhatsApp Groups	Carlano et al.	ACM Conference on Web Science	ACM
[55]	COVID-19 Fake News Detection in Malaysia - A Supervised Approach	Kalaimagal et al.	International Conference on Software and Computer Applications	ACM
[109]	Farewell to Big Data? Studying Misinformation in Mobile Messaging Applications	Rossini	Political Communication	ACM
[17]	Misinformation Campaigns through WhatsApp and Telegram in Presidential Elections in Brazil	Benevenuto and Melo	Communications of the ACM	ACM
[92]	Social Debunking of Misinformation on WhatsApp: The Case for Strong and In-group Ties	Pasquetto et al.	ACM on Human-Computer Interaction	ACM
[30]	Can WhatsApp Counter Misinformation by Limiting Message Forwarding?	de Freitas Melo et al.	Complex Networks and Their Applications	Capes
[7]	Content Analysis of Persuasion Principles in Mobile Instant Message Phishing	Ahmad et al.	International Symposium on Human Aspects of Information Security and Assurance	Capes
[48]	Perceived Information Overload and Unverified Information Sharing on WeChat Amid the COVID-19 Pandemic: A Moderated Mediation Model of Anxiety and Perceived Herd	Huang et al.	Frontiers in Psychology	Capes
[89]	How Disinformation on WhatsApp Went From Campaign Weapon to Governmental Propaganda in Brazil	Ozawa et al.	Social Media + Society	Capes
[5]	Interactive Web App for Fake News Detection	Agarwal, Spursh et al.	ITM Web of Conferences	Capes
[115]	Messaging Apps: A Rising Tool for Informational Autocrats	Trauthig et al.	Political Research Quarterly	Capes
[61]	Sir Model for Fake News Spreading Through Whatsapp	Khurana and Kumar	International Conference on Internet of Things and Connected Technologies	Capes
[11]	Tackling fake news in socially mediated public spheres: A comparison of Weibo and WeChat	Pang et al.	Technology in Society	Capes
[142]	Use of Chatbots for News Verification	Arias Jimenez et al.	Communication and Applied Technologies	Capes
[63]	WhatsApp audios and the remediation of radio: Disinformation in Brazilian 2018 presidential election	Kischinevsky et al.	Radio Journal: International Studies in Broadcast & Audio Media	Capes
[110]	A Comparative Study on Various Machine Learning Algorithms for the Prediction of Fake News Detections Using Bring Feed New Data Sets	Senthilkumar and Kumar	Int. Journal of Scientific Research in Computer Science, Engineering and Information Technology	Capes
[57]	Bias in X (Twitter) and Telegram Based Intelligence Analysis: Exploring Challenges and Potential Mitigating Roles of AI	Karakikes et al.	SN Computer Science	Capes
[35]	Detection of Misinformation About COVID-19 in Brazilian Portuguese WhatsApp Messages	Forte Martins et al.	International Conference on Applications of Natural Language to Information Systems	Capes
[117]	How to improve the rumor-confutation ability of official rumor-refuting account on social media: A Chinese case study	Tu et al.	Computer Communications	Capes
[10]	Legitimizing falsehood in social media: A discourse analysis of political fake news	Igwebuike and Chimunya	Discourse & Communication	Capes
[122]	Weak Supervision for Fake News Detection via Reinforcement Learning	Wang et al.	AAAI Conference on Artificial Intelligence	Capes
[78]	"Fake Elections"? Cyber Propaganda, Disinformation and the 2017 General Elections in Kenya	Mwera	African Journalism Studies	Capes
[77]	A study of misinformation in audio messages shared in whatsapp groups	Maros et al.	Multidisciplinary International Symposium on Disinformation in Open Online Media	Capes
[54]	Deep reinforcement learning-based approach for rumor influence minimization in social networks	Jiang et al.	Applied Intelligence	Capes
[124]	Disinformation networks: A quali-quantitative investigation of antagonistic Dutch-speaking Telegram channels	Willært et al.	First Monday	Capes
[94]	Exploring fake news identification using word and sentence embeddings	Priyanga et al.	Journal of Intelligent & Fuzzy Systems	Capes
[66]	Fake News Detection using Machine Learning	Kulkarni et al.	International Conference on Automation, Computing and Communication	Capes
[70]	Connectivity Between Russian Information Sources and Extremist Communities Across Social Media Platforms	Leahy et al.	Fontiers in Political Science	Capes
[28]	Interactive mapping of Covid-19 misinformation in Ibero-America	Colmenero-Ruiz et al.	El profesional de la información	Capes
[16]	The Pushshift Telegram Dataset	Baumgartner et al.	AAAI Conference on Web and Social Media	Capes
[13]	#RumorsCOVID-19: Predicting the Forwarding of Online Rumors in Wuhan, China and in Israel	Ariel et al.	The International Communication Gazette	Capes
[108]	A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects	Robers et al.	IEEE Journals & Magazine	Capes
[41]	Chat Apps and Cascade Logic: A Multi-Platform Perspective on India, Mexico, and the United States	Gursky et al.	Social Media + Society	Capes
[116]	Detection of Fake Information in Spanish Using Machine Learning Techniques	Tretiakov et al.	Lecture Notes in Computer Science	Capes
[53]	Fake face image detection using feature network	Jayaram et al.	Int. Journal of Health Sciences	Capes
[45]	Misinformation and professional news on largely unmoderated platforms: the case of telegram	Herasimenka et al.	Journal of Information Technology & Politics	Capes
[69]	Implementing A Passive Aggressive Classifier To Detect Fake Information	Lajurkar and Chandhari	Int. Journal of Advanced Research in Computer & Communication Engineering	Capes
[127]	Medical and Health-Related Misinformation on Social Media: Bibliometric Study of the Scientific Literature	Yeung et al.	Journal of Medical Internet Research	Capes
[126]	Misinformation messages shared via WhatsApp in Mexico during the COVID-19 pandemic: an exploratory study	Wirtz et al.	Health Promotion International	Capes
[105]	A Dataset of Fact-Checked Images Shared on WhatsApp During the Brazilian and Indian Elections	Reis et al.	AAAI Conference on Web and Social Media	Capes
[49]	Narrative-based misinformation in India about protection against Covid-19: Not just another "noo-point"	Indian Journal of medical ethics	First Monday	Capes
[99]	Discursive strategies for disinformation on WhatsApp and Twitter during the 2018 Brazilian presidential election	Recuero et al.	First Monday	Capes
[43]	Hold Your Fingers: The Communication Ethics on WhatsApp Based on the Hadith	Hasanah and Busro	Jurnal Komunikasi: Malaysian Journal of Communication	Capes
[25]	WhatsApp and audio misinformation during the Covid-19 pandemic	Cardoso et al.	ISCTE-IUL	Capes
[86]	A hierarchical network-oriented analysis of user participation in misinformation spread on WhatsApp	Nobre et al.	Information Processing & Management	Capes
[42]	Analyzing Misinformation Claims During the 2022 Brazilian General Election on WhatsApp, Twitter, and Kwai	Hale et al.	International Journal of Public Opinion Research	Capes
[58]	COVID-19-Related Social Media Fake News in India	Al-Zaman	Journalism and Media	Capes
[9]	Detection of fake images on whatsapp using socio-temporal features	Kaur et al.	Social Network Analysis and Mining	Capes
[81]	Don't Break the Chain: Measuring Message Forwarding on WhatsApp	Melo et al.	AAAI Conference on Web and Social Media	Capes
[118]	Fake News Detection With the Help of Computation Time to Increase Accuracy	Umasankarasari et al.	IGI Global Scientific Publishing	Capes
[4]	Predicting the Veracity of Fake Information on Smart Media for Reducing Misinformation Diffusion	Agarwal et al.	International Journal of Computational Intelligence & IoT	Capes
[97]	Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News	Qayyum et al.	IT Professional	Capes
[80]	WhatsApp Monitor: A Fact-Checking System for WhatsApp	Melo et al.	AAAI Conference on Web and Social Media	Capes
[51]	A First Look at COVID-19 Messages on WhatsApp in Pakistan	Javed et al.	International Conference on Advances in Social Network Analysis and Mining	IEEE
[36]	Unsupervised WhatsApp Fake News Detection using Semantic Search	Gagliani et al.	International Conference on ent Computing and Control Systems	IEEE
[84]	Analyzing Public Opinion and Misinformation in a COVID-19 Telegram Group Chat	Ng and Luke	IEEE Internet Computing	IEEE
[64]	A Survey on Role of Machine Learning and NLP in Fake News Detection on Social Media	Agrawal et al.	International Conference on Computing, Power and Communication Technologies	IEEE
[44]	Content Characteristics and Transmission Strategies of Social Media Rumors in China: Big Data Analysis of WeChat Rumors	He et al.	International Conference on Behavioral, Economic and Socio-Cultural Computing	IEEE
[10]	Assessing the Validity of Health Messages Used by the Saudi Public in WhatsApp	Alfaris et al.	Patient Preference and Adherence	PubMed
[20]	Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe	Bowles et al.	PLoS ONE	PubMed
[111]	COVID-19-Associated Misinformation Across the South Asian Diaspora: Qualitative Study of WhatsApp Messages	Sharma et al.	JMIR Infodemiology 2023	PubMed
[35]	Demographic Factors Influencing the Impact of Coronavirus-Related Misinformation on WhatsApp: Cross-sectional Questionnaire Study	Bapaye and Bapaye	JMIR Novel Coronavirus (COVID-19)	PubMed
[14]	COVID-19 Information Dissemination Using the WeChat Communication Index: Retrospective Analysis Study	Fan et al.	Journal of Medical Internet Research (JMIR)	PubMed
[125]	Debunking highly prevalent health misinformation using audio dramas delivered by WhatsApp: evidence from a randomised controlled trial in Sierra Leone	Winters et al.	BMJ Global Health	PubMed
[38]	Fake News and vaccine hesitancy in the COVID-19 pandemic in Brazil	Galhardi et al.	Fiocruz	PubMed
[18]	Misinformation on social networks during the novel coronavirus pandemic: a quali-quantitative case study of Brazil	Biancovilli et al.	BMC Public Health	PubMed
[37]	Social Media and the Influence of Fake News on Global Health Interventions: Implications for a Study on Dengue in Brazil	Gagnon-Dufresne et al.	International Journal of Environmental Research and Public Health	PubMed
[82]	The Electoral Misinformation Nexus: How News Consumption, Platform Use, and Trust in News Influence Belief in Electoral Misinformation	Mont'Alveme et al.	Public Opinion Quarterly	PubMed
[52]	A deep dive into COVID-19-related messages on WhatsApp in Pakistan	Javed et al.	Social Network Analysis and Mining	PubMed
[72]	Hot Topic Recognition of Health Rumors Based on Anti-Rumor Articles on the WeChat Official Account Platform: Topic Modeling	Li et al.	Journal of Medical Internet Research	PubMed
[128]	Mechanism of WeChat's Impact on Public Risk Perception During COVID-19	Zhuang et al.	Risk Management and Healthcare Policy	PubMed
[21]	COVID-19 Infodemic: A study on the Fragile Five countries	Boyacı Yıldırım	Journal of Public Affairs	PubMed
[14]	COVID-19, a tale of two pandemics: novel coronavirus and fake news messaging	Alehortsu and Patino	Health Promotion International	PubMed
[121]	Dynamics of social corrections to peers sharing COVID-19 misinformation on WhatsApp in Brazil	Vijaykumar et al.	Journal of the American Medical Informatics Association	PubMed
[1]	How people interact with a chatbot against disinformation and fake news in COVID-19 in Brazil: The CoronaAI case	Abonizio et al.	International Journal of Medical Informatics	PubMed
[65]	Wildfire-Like Effect of a WhatsApp Campaign to Mobilize a Group of Predominantly Health Professionals With a University Degree on a Health Issue: Infodemiology Study	Kopilas and Gajović	Journal of Medical Internet Research	PubMed
[112]	Everyday non-partisan fake news: Sharing behavior, platform specificity, and detection	Shephard et al.	Frontiers in Psychology	PubMed
[75]	A Relationship-Centered and Culturally Informed Approach to Studying Misinformation on COVID-19	Malhotra	Social Media + Society	PubMed
[76]	Antecedents and Consequences of Misinformation Sharing Behavior among Adults on Social Media during COVID-19	Malik et al.	Sage Open	PubMed
[29]	Covid-19 Protesters and the Far Right on Telegram: Co-Conspirators or Accidental Bedfellows?	Curley et al.	Social Media + Society	PubMed
[83]	Exploratory study of the hoaxes spread via WhatsApp in Spain to prevent and/or cure COVID-19	Moreno-Castro et al.	Gaceta Sanitaria	PubMed
[113]	The importance of social media users' responses in tackling digital COVID-19 misinformation in Africa	Stewart et al.	Digital Health	PubMed
[39]	Fact or Fake? An analysis of disinformation regarding the Covid-19 pandemic in Brazil	Galhardi et al.	Ciencia & Saude Coletiva	PubMed
[73]	How Official Social Media Affected the Infodemic among Adults during the First Wave of COVID-19 in China	Liu et al.	International Journal of Environmental Research and Public Health	PubMed
[27]	A real-time platform to monitoring misinformation on telegram	Claudio et al.	International Conference on Enterprise Information Systems	Google Scholar
[64]	Collective social correction: addressing misinformation through group practices of information verification on WhatsApp	Kluger-Vilenchik	Digital Journalism (Taylor & Francis)	Google Scholar
[22]	FakeWhatsApp, BR: NLP and Machine Learning Techniques for Misinformation Detection in Brazilian Portuguese WhatsApp Messages.	Cabral et al.	International Conference on Enterprise Information Systems	Google Scholar
[62]	Research of fake news spreading through WhatsApp	Khurana et al.	International Journal of Innovative Technology and Exploring Engineering	Google Scholar
[79]	Addressing misinformation on WhatsApp in India through intermediary liability policy, platform design modification, and media literacy	Medeiros and Singh	Journal of Information Policy	Google Scholar
[100]	Can WhatsApp benefit from debunked fact-checked stories to reduce misinformation?	Reis et al.	Harvard Kennedy School (HKS) Misinformation Review	Google Scholar
[91]	Sharing fake news about health in the cross-platform messaging app WhatsApp during the COVID-19 pandemic: A pilot study	Pinto et al.	International Journal of Scientific Research and Management	Google Scholar
[88]	Social media and fake news in Nigeria: A speech act analysis of WhatsApp messages on coronavirus	Owojoko	Studies in Pragmatics and Discourse Analysis	Google Scholar
[120]	Information, misinformation, and disinformation about COVID-19: A content study of closed-cross platform messaging using WhatsApp	Verna et al.	Int. Journal of Advance Science and Technology	Google Scholar
[56]	Unraveling WhatsApp group dynamics to understand the threat of misinformation in messaging apps	Kalogeropoulos and Rossini	New Media & Society	Google Scholar
[32]	Audio misinformation on WhatsApp: A case study from Lebanon	El-Masri et al.	Harvard Kennedy School Misinformation Review	Google Scholar
[40]	Images and misinformation in political groups: Evidence from WhatsApp in India	Garmella and Eckles	Harvard Kennedy School Misinformation Review	Google Scholar
[46]	Teens' motivations to spread fake news on WhatsApp	Herrero-Di et al.	Social Media + Society	Google Scholar

e a verificação de dinâmicas de grupo que podem exercer influência nesse processo [56].

4.2 Plataformas de Mensagem Investigadas

Outro dado interessante se deriva das plataformas utilizadas para análise nos trabalhos. Embora o WhatsApp seja o mais popular no

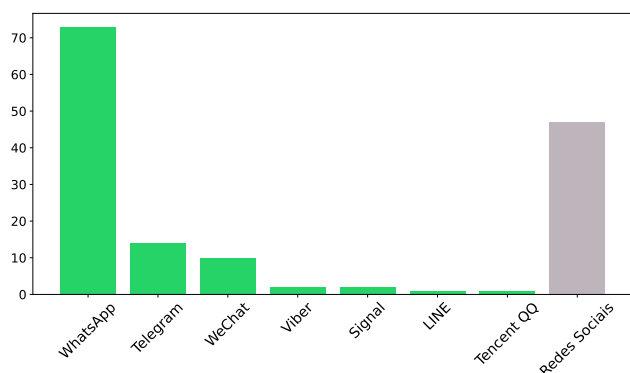


Figura 4: Quantidade de estudos por PMI investigada.

Brasil, existem diversas outras PMIs com características semelhantes e que também foram identificadas como vetores de desinformação pelos estudos, como mostrado pela Figura 4. De forma geral, o WhatsApp é a plataforma mais estudada, sendo analisada em 72 dos estudos incluídos e contendo também a maior quantidade de pesquisas dedicadas unicamente a si. Comparativamente, as duas outras plataformas mais discutidas são o Telegram, que é diretamente investigado em 13 dos estudos, e o WeChat, que se encontra presente em 10 deles. Além das três plataformas incluídas diretamente na string de busca, existem menções a outros aplicativos de Mensagens Instantâneas, como LINE [71], Viber e o Signal, [41, 115], além das Redes Sociais tradicionais como Instagram, Twitter e o Facebook, que muitas vezes são analisadas de maneira conjunta com as PMIs [38, 57, 78, 112].

Como mencionado, a segunda PMI mais investigada foi o Telegram, serviço que apresenta características similares ao WhatsApp. Em ambas, a facilidade no encaminhamento, a natureza das comunicações privadas e as dificuldades em implementar moderação interna e denúncia de mensagens foram apontadas facilitadoras do espalhamento da desinformação. A terceira plataforma mais frequente, o WeChat, é popular principalmente na China. Além de diferenças em relação ao número de funcionalidades se comparado com as demais, também verificaram-se diferenças na natureza dos estudos que investigam o WeChat, com sua maioria sendo dedicada a análises de natureza social em relação ao impacto e a caracterização da desinformação.

Os resultados encontrados sugerem que o ecossistema criado por sistemas de plataformas de mensagens instantâneas (e.g., WhatsApp, Telegram, WeChat, LINE, Signal) compartilha de um conjunto de características estruturais que favorecem a disseminação da desinformação. A arquitetura fechada dessas plataformas, com comunicação predominantemente privada, criptografada e centrada em grupos ou conversas interconectados, limita mecanismos de moderação e dificulta o rastreamento do conteúdo, criando um ambiente propício para a circulação de mensagens falsas sem detecção imediata. Além disso, a facilidade técnica de encaminhamento, presente em plataformas como WhatsApp e Telegram, potencializa a viralização de conteúdos, especialmente quando aliada à confiança social implícita nas redes interpessoais que operam nesses espaços.

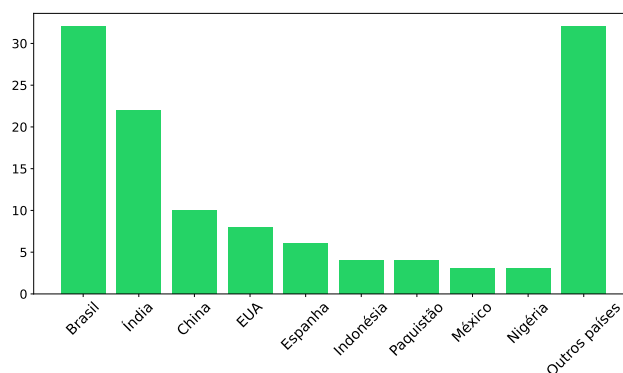


Figura 5: Distribuição por país de realização do estudo.

4.3 Distribuição Geográfica

Parte dos estudos considerados foca na realidade de um país, região geográfica ou grupo populacional específico, colaborando para a construção de uma visão mais local do fenômeno da desinformação. Portanto, também investigamos os artigos que mencionam explicitamente os países analisados, desconsiderando, portanto, estudos com escopo continental ou mais disperso, como o de Sharma et al. [111]. Como visto na Figura 5, os países com maior número de estudos são o Brasil e a Índia, ambos com mais de vinte artigos dedicados ao tema. Em seguida, aparecem a China, com 10 artigos, os Estados Unidos com 8 e a Espanha, com 6. A predominância de estudos focados no Brasil e na Índia parece refletir a alta penetração do WhatsApp nessas regiões com centenas de milhões de usuários em cada um desses países. Outro motivo é a intensidade do uso político e polarizado da plataforma em processos eleitorais recentes nesses locais, bem como os desafios enfrentados por autoridades locais no combate à desinformação [17, 50, 89].

Um aspecto relevante é a concentração das pesquisas no chamado “sul global”, incluindo na lista países como Indonésia (3), México (3), Paquistão (2) e Nigéria (2). Essa distribuição sugere que, embora a desinformação seja um fenômeno global, seu impacto e suas dinâmicas assumem contornos particulares em países com desigualdades acentuadas, sistemas de informação mais vulneráveis e alta dependência das PMIs como meio primário de comunicação digital. Além disso, diversos estudos apontam que essas regiões possuem um ambiente propício à circulação de boatos, teorias conspiratórias e campanhas coordenadas de desinformação, dada a combinação entre conectividade crescente e baixo letramento digital. A categoria “Outros Países” inclui um total de 23 nações com uma ou duas ocorrências cada, o que demonstra também o problema atinge em escala global. Essa desigualdade geográfica da produção científica também evidencia a necessidade de mais investigações comparativas para compreender melhor as características regionais da desinformação mediada por aplicativos de mensagens.

4.4 Abordagens Metodológicas dos Estudos

Ao analisar os estudos, percebemos uma diversidade nas abordagens metodológicas utilizadas para tratar o fenômeno da desinformação. Isso reflete os desafios inerentes à investigação do fenômeno da desinformação e, especialmente, do ecossistema específico das PMIs que, devido à arquitetura fechada, impõe barreiras significativas a

uma coleta mais sistemática de dados. Diante disso, compreender as abordagens utilizadas pelos pesquisadores é essencial para avaliar as oportunidades e desafios existentes na literatura.

Entrevistas e outras abordagens qualitativas: Dadas as dificuldades na aquisição de dados associadas a estas plataformas, muitos pesquisadores optam por estratégias que os colocam em contato direto com usuários finais, de forma que através de seus relatos seja possível entender o cenário investigado. Uma dessas abordagens é a realização de entrevistas, usadas para entender problemas percebidos pelos usuários e a eficácia de métodos para informá-los [3, 41, 75, 115]. Existe também a utilização de questionários, havendo maior padronização de respostas e potencialmente menos foco em especificidades individuais [13, 48, 121, 128]. Existem estudos que combinam os dois métodos [92] ou utilizam de um deles somado a métodos menos qualitativos [82].

Etnografia digital: Uma das formas de se adquirir dados de aplicativos fechados é a do uso das informações de determinado grupo que as fornece com esse propósito, ou mesmo do conteúdo presente nas contas pessoais de um autor usuário da plataforma, fornecendo um recorte individualizado mas potencialmente informativo da questão investigada. Essa abordagem é vista em [7, 10, 64, 65, 88, 93, 111, 120, 126], normalmente menores em escala e abrangência devido a limitação inerente em se usar dados de um único grupo ou usuário. No entanto, também existem autores que a combinam métodos de coleta massiva de dados com os dados etnográficos fornecendo um ponto de comparação com os originados por outros métodos [32].

Coletas de dados de grupos públicos: Visando análises quantitativas, a principal estratégia observada é a da coleta de dados de grupos públicos nas plataformas investigadas. Dada a natureza social da comunicação instantânea, é comum que usuários se organizem em grupos dedicados a assuntos específicos, criando comunidades públicas. Algumas dessas se organizam esperando um fluxo de novos usuários, divulgando o acesso ao grupo em *websites* dedicados a este propósito. A partir destes grupos públicos, que podem ser incluídos com o propósito de analisar setores específicos ou de fornecer uma visão geral do fluxo de comunicação, é possível realizar a coleta massiva de dados, como centenas de milhões de mensagens [16], ou ainda dados de outras modalidades, como imagens [40, 52, 58, 104, 105] e áudios [25, 32, 77]. Essa metodologia é também vista em [17, 22, 24, 27, 29, 30, 35, 42, 51, 63, 74, 81, 83, 84, 86, 99, 106, 107, 124]. Um ponto importante desta abordagem é o desafio metodológico apontado para aquisição de dados destas plataformas, que esbarram tanto em questões éticas quanto também em dificuldades técnicas, pois lidam com dados sensíveis, uma arquitetura fechada e de difícil acesso. Por outro lado, uma vantagem deste método é o grande volume de dados analisados que possibilitam a criação de *datasets* públicos que podem ser aproveitados em outras pesquisas, como ocorre em [16, 45, 105].

5 TAXONOMIA DA DESINFORMAÇÃO

Para estruturar a análise e sintetizar as evidências encontradas relativas as questões de pesquisa definidas neste estudo (Seção 3.1), também conduzimos uma organização temática dos resultados obtidos. A partir deste processo, desenvolvemos uma estrutura hierárquica que permitiu a sistematização dos resultados em eixos

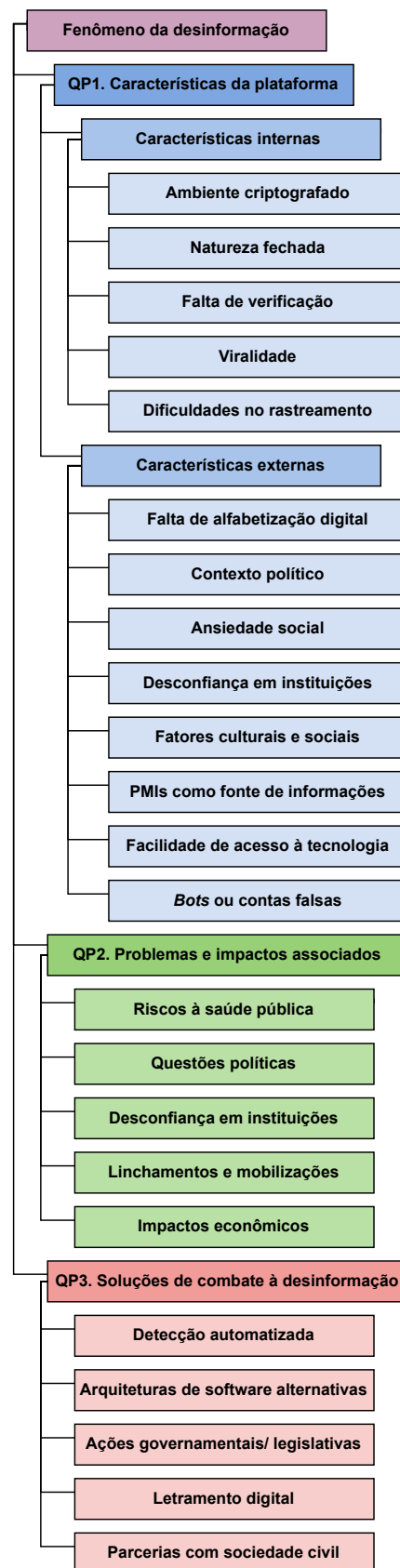


Figura 6: Representação hierárquica.

temáticos. Essa taxonomia surgiu a partir da sintetização do conteúdo dos artigos e foi refinada por meio de discussões entre os membros da equipe de pesquisa, buscando capturar as características principais e os diferentes aspectos do fenômeno investigado. A Figura 6 apresenta essa representação hierárquica, que sumariza os principais achados da literatura em torno das três questões de pesquisa previamente definidas: (1) os fatores facilitadores da desinformação em PMIs; (2) os impactos sociais relacionados; e, (3) as estratégias de combate propostas ou avaliadas. A seguir detalhamos cada uma das categorias criadas, apontados os artigos e discutindo a respeito dos resultados.

5.1 QP1: Características Internas e Externas das Plataformas de Mensagem

Características internas: Na Figura 6, no subtópico referente às características internas, são apresentadas as funcionalidades das plataformas que exercem maior influência no envio de conteúdo desinformativo, indicando a necessidade de uma análise mais aprofundada. As principais características internas identificadas como potencialmente favoráveis à criação e disseminação de desinformação nas PMIs estão associadas à arquitetura e à proposta central desses aplicativos, que para além de simples conversas *one-to-one* (“um pra um”) comuns em plataformas mais antigas, existe também o formato *many-to-many* (“muitos pra muitos”) através de grupos de conversa, ou ainda “um pra muitos”, como é o caso de canais de transmissão *broadcast*, populares no Telegram e relacionados a “criação de narrativas antagonísticas”, sendo essas representadas em teorias da conspiração que se inter-conectam, como “QAnon” e o chamado “Great Reset” [124]. Em todas essas modalidades de conversa, existe uma expectativa de privacidade onde os envolvidos tem algum conhecimento sobre os demais participantes, na forma de números ou nomes de usuário, mas não-membros não possuem acesso a tais dados. Tal expectativa marca uma diferença profunda para com as Redes Sociais Tradicionais (RSTs), cujo foco em *feeds* e recomendação algorítmica leva a um maior grau de contato com usuários fora da rede principal do usuário. Enquanto algoritmos sofisticados coordenam o que um usuário visualiza em plataformas como o Twitter, Instagram ou Facebook, em outras como Telegram e WhatsApp, são os próprios usuários dos grupos e *chats* que filtram, selecionam e encaminham os conteúdos que seus pares receberão no fluxo da conversa, como apontado em [109].

A característica mais citada como facilitadora da desinformação é a **viralidade**, que muitas vezes se relaciona com facilidade no encaminhamento de mensagens, sendo esse um apontamento que já havia sido feito em 2019 nos estudos [24, 107]. No entanto, esse assunto atingiu proeminência nos anos seguintes, com a ocorrência da pandemia e a urgência no entendimento de um fenômeno que poderia estar agravando-a, e mesmo após ela, como visto em [9, 13, 16–18, 20–22, 29, 30, 37, 41, 43, 47, 48, 50, 53, 54, 58, 71, 76, 77, 81, 92, 95, 99, 103, 107, 111, 113, 114, 120, 121, 124].

A viralidade e a facilidade no encaminhamento de conteúdo não são propriedades exclusivas dos aplicativos de mensagens, mas tem suas consequências potencialmente agravadas num meio privado. Essa consideração ganha destaque ao observar que as próximas duas categorias mais referidas neste contexto descrevem o

ambiente criptografado e a **natureza fechada** destas plataformas, propriedades mais relacionadas a aplicativos de mensagens. O ambiente criptografado está presente em todas as plataformas identificadas na *string* de busca, incluindo o WeChat, e é mencionado em [3, 16, 17, 24, 25, 36, 41, 56, 58, 61, 63, 64, 75, 79, 81, 86, 89, 92, 93, 95, 105, 107, 113–115, 119, 126]. Diversos estudos apontam para a chamada “natureza fechada” das PMIs como sendo facilitadora do espalhamento da desinformação, sendo essa uma categoria mais abrangente que a anterior, abrangendo arquiteturas não diretamente criptográficas mas que contribuem para o agravamento do fenômeno [3, 16, 17, 24, 25, 36, 41, 56, 58, 61, 63, 64, 75, 79, 81, 86, 89, 92, 93, 95, 105, 107, 113–115, 115, 119, 126].

Ainda considerando fatores agravantes mais unicamente relacionados à PMIs, observa-se a dificuldade na moderação [3, 9, 18, 24, 29, 41, 43, 47, 51, 52, 57, 70, 70, 74, 92, 94, 107, 113, 124] e na implementação de mecanismos que facilitem e/ou forneçam checagem dos fatos compartilhados [51, 71, 74, 76, 83, 95, 99, 107, 117] em ambientes deste tipo, o que leva à **falta de verificação**. Quando existente, a moderação de comunidades é feita por seus próprios membros, havendo pouco ou nenhum recurso [100] além da remoção de usuários. A habilidade de denunciar conteúdos, comum em redes sociais tradicionais, é um dos recursos não disponíveis que representa uma barreira para que haja moderação adequada.

Outro fator agravante é o da **dificuldade no rastreamento** da origem de conteúdos recebidos [17, 29, 58, 58, 71, 84, 89, 111, 114], havendo o encaminhamento sem atribuição ao criador. Em contraste com outras redes sociais, cujo grande foco consiste em autoria e influência, nas PMIs, por vezes, o conteúdo de terceiros é encaminhado para o usuário final através de uma pessoa que participa de sua rede imediata de contatos, podendo herdar a credibilidade atribuída a essa pessoa de maneira equivocada.

Uma propriedade muito comum tanto em PMIs quanto em outras redes sociais tradicionais é a organização em comunidades, que ocorre de diferentes maneiras mas se mantém definida por um interesse comum e pela interação conjunta de seus usuários. Nas PMIs, comunidades ou grupos, no geral, são ambientes onde todos seus participantes podem ler e enviar mensagens, com exceção dos já mencionados Canais *one-way*, onde os proprietários fazem a emissão de conteúdo sem que haja a possibilidade de interagir diretamente. Dada a distinção não tão clara entre comunicação privada e massiva em tais ambientes, existem indícios de sua ação como facilitador da desinformação em um número considerável de estudos [3, 17, 22, 24, 27, 47, 50, 56, 84, 86, 93, 93, 103–106, 119, 121].

Características externas: O estudo dos fatores externos revelou uma diversidade ainda maior em comparação aos fatores internos, como pode ser observado pela quantidade de categorias presentes em cada tópico da Figura 6, um resultado esperado partindo-se desta questão mais ampla. A característica mais mencionada nesta categoria é o **contexto político** marcado, muitas vezes, pela polarização [3, 13, 17, 24, 27, 37, 38, 47, 49–51, 56–58, 64, 76, 82, 88, 95, 99, 103–106, 113, 119, 127], por vezes associada diretamente a pandemia do COVID-19. Ela indica a forma como, mediante a crises e desafios sociais, acaba por haver maior disposição em se acreditar em conteúdo desinformativo. Num tópico relacionado, observa-se também que grande quantidade de estudos classifica o uso automatizado e sistemático das plataformas como potencial facilitador da desinformação, algo visto em

[14, 18, 24, 29, 49, 50, 56, 70, 74, 81, 89, 92, 99, 104, 111, 116, 120]. Tal uso (normalmente de cunho político) é representado pela utilização de **bots ou contas falsas** numa escala fora do alcance para usuários comuns, tendo o objetivo de manipular a opinião pública. Ainda em relação a política, diversos estudos apresentam constatações acerca do papel da **desconfiança em instituições**, sejam elas científicas, públicas ou de mídia [9, 14, 17, 18, 37, 43, 64, 73, 78, 82, 92, 127], que se soma ao já referido problema de campanhas organizadas de uso de *bots* e contas falsas que não necessariamente ocorre apenas no contexto político [78, 89, 107].

Levando em conta a pandemia do COVID-19 e demais cenários de crises sanitárias, políticas ou econômicas, outro dos principais fatores levantados nesta análise foi o da **ansiedade social** perante a eventos disruptivos [1, 15, 21, 34, 48, 48, 51, 73, 76, 111–113, 116, 120], onde a preocupação generalizada leva a busca de informações sem o foco necessário na verificação. Fenômenos psicológicos compõe grande parte desta categoria, como nos apontamentos acerca de vieses de confirmação [41, 49, 57, 64, 76, 78, 83, 114, 120, 127], por vezes agravados pelo contexto político e ansiedade social, e também nas menções ao efeito da familiaridade entre usuários na comunicação por mensagens instantâneas [20, 89, 93, 95, 106, 113, 114, 121], onde informação externa e não verificada pode adquirir aparência de legitimidade. De maneira adjacente, são apresentados fatores como hierarquias sociais [3, 75, 119] e a preferência de se usar **PMIs como fonte de informação** [10, 56, 78, 126].

Outra consideração frequente relata a forma como a **falta de alfabetização digital** pode facilitar o espalhamento da desinformação [1, 9, 17, 38, 43, 51, 53, 58, 69, 79, 81, 93, 111, 113, 114], trazendo para a discussão uma perspectiva educacional. Em paralelo, alguns estudos apontam que a **facilidade de acesso à tecnologia** pode também contribuir para a difusão de informações falsas [13, 56, 61, 62, 69, 78], com usuários pouco experientes sendo mais vulneráveis. Também relacionados estão os estudos que investigam o efeito da faixa etária do usuário, onde dois apontam que idosos costumam compartilhar mais desinformação [15, 93] enquanto outro indica a vulnerabilidade de jovens, que são excessivamente confiantes em sua habilidade de detectá-las [129]. Esses fatores podem acabar se somando a dificuldade de encontrar/falta de informações oficiais acerca dos assuntos em pauta, especialmente mediante a situações que trazem insegurança ao público, como apontado em [13, 21, 84, 116]. Além disso, a forma como determinados grupos lidam com a informação varia em função de **fatores sociais e culturais** [43, 47, 49, 71, 76, 79].

Pode-se observar ainda que estes temas estão largamente interconectados, de forma que mesmo constatações encontradas de maneira única se relacionam com as demais categorias. Um exemplo pode ser visto na consideração do papel da homofilia, um conceito de Redes Complexas que explica a tendência de união entre iguais, tido como contribuinte para o fenômeno da desinformação [126], dado que num grupo de similares, existe uma chance menor de que haja discordância para com uma informação divulgada. Conceitos de outras áreas, como “Pós-verdade” e o fenômeno “Dr. Google” também são apontados como contribuintes [28], estando ambos relacionados com a forma como o avanço de tecnologias de comunicação mudam o relacionamento do público com as fontes de informação.

5.2 QP2: Problemas e Impactos Associados ao Espalhamento de Desinformação

Em seguida, focamos em investigar as consequências negativas e/ou impactos reais que o problema desinformação compartilhada em PMIs tem causado na sociedade. A intenção não é apenas listar efeitos isolados, mas sintetizar evidências empíricas encontradas na literatura sobre como mensagens falsas afetam decisões individuais e coletivas, por exemplo, gerando riscos à saúde pública (como durante a pandemia de COVID-19) e enfraquecimento de processos democráticos e campanhas eleitorais. Os estudos revisados também mostraram que esses efeitos variam segundo o contexto sociopolítico e a penetração da plataforma no país, podendo gerar tanto consequências imediatas quanto dinâmicas de longo prazo (e.g., erosão de confiança e polarização política). A seguir, organizamos as evidências por categorias de impacto para facilitar a leitura e o confronto entre achados empíricos.

A crença em informações equivocadas dificulta a tomada de decisão, trazendo consequências para indivíduos e para a sociedade como um todo. Esse fenômeno foi posto em destaque com a ocorrência da pandemia, quando a urgência de sua compreensão se comprova em estudos que o apontam como fator de **risco à saúde pública** [1, 3, 4, 9, 10, 14, 15, 18, 20, 22, 23, 25, 28, 34, 35, 37, 38, 43, 49, 51, 52, 54, 64, 75, 76, 79, 83, 84, 93, 95, 108, 111–113, 116, 119, 121, 125–127] e gerador de **impactos econômicos** [13, 21, 113, 117]. A desinformação é apontada como tendo potencial para interferir em **questões políticas**, ferindo o processo democrático, conforme evidenciado pela alta incidência de estudos que a relacionam a campanhas de manipulação eleitoral e ao aumento da polarização política [3, 4, 12, 17, 21–24, 27, 36, 38, 41, 50, 53, 54, 56, 58, 63, 74, 77–79, 89, 92, 95, 97, 103, 106, 108, 115, 116]. O fato da polarização ser destacada nessa seção demonstra a primeira ocorrência de uma característica sendo referida tanto como facilitadora da desinformação (QP1) quanto como causada por ela (QP2), trazendo a tona a necessidade de entender se isto de fato caracteriza uma relação circular e auto-agravante.

Outra ocorrência do tipo se dá nas referências relacionadas à **desconfiança em instituições**, [14, 22, 23, 28, 43, 45, 48, 49, 57, 72, 76–78, 82, 99, 108, 114, 121, 122, 126], sejam elas públicas, como agências sanitárias governamentais, ou privadas, como jornais e empresas de *fact-checking*. Após estas sofrerem ataques desafiando a legitimidade de seu conteúdo, algo agravado pela dificuldade de verificação da corretude e origem das informações, ocorre que parcelas da população passam a desconfiar de fontes reputáveis.

Por fim, a possibilidade da existência de dinâmicas recursivas que levam ao aumento da desinformação e suas consequências se tornam ainda mais problemáticas ao considerar a ampla gama de estudos que aponta resultados violentos, como **mobilizações**, revoltas populares e até mesmo **linchamentos** públicos, causados diretamente por boatos espalhados em PMIs [1, 3, 14, 21, 29, 45, 47, 50, 53, 56, 64–66, 74, 79, 88, 97, 105, 114, 116, 120, 124].

5.3 QP3: Soluções de Combate à Desinformação

Considerando os problemas já referenciados, é natural que na literatura também se encontrem propostas para solução do problema. Estas propostas lidam com os diferentes aspectos do fenômeno da desinformação, muitas vezes trabalhando diretamente com as

características internas e externas, apresentadas na Figura 6, e destacando seus próprios desafios. Um exemplo disso está nos trabalhos que sugerem campanhas de **letramento digital** [14, 21, 28, 43, 62, 64, 79, 95, 113, 114, 121, 127] como forma de mitigar o problema, novamente reforçando a consideração feita na QP1 que aponta a falta de alfabetização digital como uma característica externa as PMIs que contribui para o espalhamento da desinformação.

Outra categoria de solução que mantém o diálogo direto com os fatores apontados pela QP1 é representada nos esforços envolvendo **arquiteturas de software alternativas**, sejam em relação a modificações internas nas plataformas associadas ou soluções externas que possam ser aplicadas a elas. Considerando as propostas internas, o principal apontamento sobre a redução da viralidade é a da realização de alterações no sistema *forwarding* (encaminhamento) de mensagens [17, 25, 30, 36, 56, 63, 71, 77, 92, 95, 104, 114, 115]. Dentre as medidas externas, uma porção expressiva se dedica a implementação de *chatbots* com o propósito de esclarecer dúvidas e fornecer informações verificadas para os usuários [1, 12, 49, 71, 92, 120], indicando a importância de fontes de informação confiáveis e facilmente acessíveis, com *bots* podendo ser adicionados a grupos e agir de maneira passiva ou ativa na verificação de conteúdo, eliminando a necessidade de sair do aplicativo para confirmar fatos.

Apesar de em isolamento não se tratar de um mecanismo de combate, os esforços relacionados à **detecção automatizada** possibilitam a aplicação e aumentam a abrangência dos demais métodos apresentados. Dessa forma, no contexto da QP3 é importante observar tais estudos, onde modelos de aprendizado de máquina são empregados com o objetivo de distinguir uma informação falsa das demais. Iniciativas desta natureza podem ser aplicadas diretamente em PMIs, ou feitas de maneira externa, mostrando novamente sua importância. Elas são representadas em [3, 4, 22, 35, 36, 53, 54, 57, 58, 66, 69, 72, 103–105, 108, 110, 116, 122]. Em outro campo, destacam-se as soluções políticas, nas quais **ações governamentais e legislativas** são empregadas com o objetivo de conter a criação e a disseminação da desinformação [9, 13, 29, 37, 82, 91, 99, 117, 124], sendo esse o tipo de medida mais apontada como já estando em aplicação no mundo real. Além de soluções de caráter exclusivamente governamental, também se observam propostas que estabelecem **parcerias com a sociedade civil**, envolvendo jornalistas, pesquisadores e empresas para oferecer checagem de fatos de maneira dinâmica [9, 10, 21, 38, 83, 91, 99, 119, 126, 127]. O reconhecimento da presença de dinâmicas sociais leva inclusive a sugestões acerca da efetividade de correções feitas por membros do próprio grupo onde a desinformação ocorre [56, 92, 125]. Existem também artigos dedicados exclusivamente ao enfrentamento da desinformação, trazendo propostas em várias das direções já delimitadas [79].

6 DISCUSSÃO E CONSIDERAÇÕES FINAIS

Os resultados obtidos neste estudo apontam para um campo de estudo em expansão que, embora ainda fragmentado metodológica e geograficamente, apresenta um número significativo de trabalhos convergentes acerca dos principais desafios e dinâmicas da desinformação nas Plataformas de Mensagens Instantâneas (PMIs). Ainda que as pesquisas tenham sido realizadas em contextos sociais, culturais e políticos diversos, observamos uma forte presença de comunalidades, com destaque para fatores como a viralização facilitada, a opacidade da arquitetura desses sistemas e a dificuldade de

moderação. Por outro lado, o levantamento também revelou pontos de divergência, como as diferentes conclusões sobre o papel da faixa etária dos usuários e sobre a efetividade de medidas específicas de mitigação, como a limitação do encaminhamento de mensagens. Entretanto, essas discordâncias não representam um problema, mas refletem a própria evolução do campo, em que diferentes hipóteses ainda estão sendo testadas e contextualizadas, muitas vezes com base em realidades locais específicas. Tal cenário reforça a importância da continuidade da produção científica sobre o tema, incluindo mais abordagens comparativas e interdisciplinares.

Um dos achados mais importantes deste trabalho diz respeito à **arquitetura comunicacional paradoxal presente em PMIs**, em particular no WhatsApp, plataforma dominante nos estudos analisados. Embora o aplicativo funcione com criptografia ponta-a-ponta e se apresente como um ambiente de comunicação privada, particular e direta, seu *design* também favorece o espalhamento de conteúdos em uma comunicação em larga escala e de forma viral, por meio de grupos amplos e do mecanismo de encaminhamento rápido. Assim, convivem dentro da mesma plataforma dois modos de circulação informacional indistinguíveis aos olhos do usuário: um pessoal, confidencial e protegido; outro massivo, viral e público. Essa ambiguidade é particularmente sensível no contexto da desinformação: o usuário, ao receber uma mensagem, tende a interpretá-la como vinda de uma fonte confiável, um familiar ou amigo próximo, mesmo quando ela já percorreu um longo caminho de disseminação impessoal. A interface não oferece pistas claras o suficiente sobre a origem e o alcance prévio da mensagem, ou as oferece de maneira facilmente contornável [81], enfraquecendo a percepção de risco e contribuindo para a desinformação. Além disso, nossas descobertas sugerem que a privacidade garantida por esse tipo de plataforma, embora fundamental em termos de direitos digitais, também dificulta a fiscalização e o combate à conteúdos danosos. Esse ponto evidencia uma tensão central no debate contemporâneo: o equilíbrio entre liberdade individual e segurança de dados pessoais. Não há solução simples para esse impasse, mas reconhecê-lo é essencial para a formulação de políticas públicas, diretrizes e estratégias de combate mais eficazes.

Por fim, este estudo contribui para consolidar o campo de desinformação sobre PMIs ao oferecer uma síntese crítica e estruturada das evidências disponíveis na literatura. Ao organizar os pontos em comum e lacunas existentes, abrimos espaço para pesquisas futuras que aprofundem temas ainda pouco explorados, como a eficácia de intervenções específicas em contextos locais e/ou globais, desinformação em tópicos além de política e saúde, a dinâmica de desinformação em outras plataformas emergentes, ou mesmo os possíveis ciclos viciosos entre desinformação, desconfiança institucional e polarização política. Dessa forma, compreender melhor esses mecanismos que regem a desinformação em PMIs é um passo fundamental para o fortalecimento de sociedades democráticas, pluralistas e informacionalmente saudáveis.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) Identificador APQ-03115-24 e do INCT-TILD-IAR – Grant #408490/2024-1.

REFERÊNCIAS

- [1] Hugo Queiroz Abonizio, Ana Paula Ayub da Costa Barbon, Renne Rodrigues, Mayara Santos, Vicente Martínez-Vizcaino, Arthur Eumann Mesas, and Sylvio Barbon Junior. 2023. How people interact with a chatbot against disinformation and fake news in COVID-19 in Brazil: The CoronaAI case. *Intl. Journal of Medical Informatics* 177 (2023), 105134.
- [2] Funmi Adebisin, Hanlie Smuts, Tendani Mawela, George Maramba, and Marie Hattingh. 2023. The Role of Social Media in Health Misinformation and Disinformation During the COVID-19 Pandemic: Bibliometric Analysis. *JMIR Infodemiology* 3 (20 Sep 2023), e48620.
- [3] Dhruv Agarwal, Farhana Shahid, and Aditya Vashistha. 2024. Conversational Agents to Facilitate Deliberation on Harmful Content in WhatsApp Groups. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 491 (Nov. 2024), 32 pages.
- [4] Isha Agarwal, Dipti P Rana, and Anjali S More. 2018. Predicting the Veracity of Fake Information on Smart Media for Reducing Misinformation Diffusion. *Intl. Journal of Computational Intelligence & IoT* 1, 2 (2018).
- [5] Agarwal, Sparsh, Varun, Malempati, and Prabakeran, S. 2023. Interactive Web App for Fake News Detection. *ITM Web Conf.* 53 (2023), 03003.
- [6] Chetan Agrawal, Anjana Pandey, and Sachin Goyal. 2021. A Survey on Role of Machine Learning and NLP in Fake News Detection on Social Media. In *2021 IEEE 4th Intl. Conf. on Computing, Power and Communication Technologies*. 1–7.
- [7] Rufai Ahmad, Sotirios Terzis, and Karen Renaud. 2023. Content Analysis of Persuasion Principles in Mobile Instant Message Phishing. In *Human Aspects of Information Security and Assurance*, Steven Furnell and Nathan Clarke (Eds.). Springer Nature Switzerland, Cham, 324–336.
- [8] Esma Aïmeur, Sabrina Amri, and Gilles Brassard. 2023. Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining* 13, 1 (2023), 30.
- [9] Md Sayeed Al-Zaman. 2021. COVID-19-related social media fake news in India. *Journalism and Media* 2, 1 (2021), 100–114.
- [10] Eiad Alfaris, Yasser Alhazzani, Abdullah Alkhenizan, Farhana Irfan, Naif Al-moneef, Nada Alyousefi, Huda Alfaris, Khitam Alodhaibi, and Abdullah MA Ahmed. 2023. Assessing the Validity of Health Messages Used by the Saudi Public in WhatsApp. *Patient Preference and Adherence* 17 (2023), 67–73.
- [11] Sacha Altay, Manon Berriche, Hendrik Heuer, Johan Farkas, and Steven Rathje. 2023. A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. *Harvard Kennedy School Misinformation Review* 4, 4 (2023), 1–34.
- [12] B. Arias Jiménez, C. Rodríguez-Hidalgo, C. Mier-Sanmartín, and G. Coronel-Salas. 2023. Use of Chatbots for News Verification. In *Communication and Applied Technologies*, Paulo Carlos López-López, Daniel Barredo, Ángel Torres-Toukoumidis, Andrea De-Santis, and Óscar Avilés (Eds.). Springer Nature Singapore, Singapore, 133–143.
- [13] Yaron Ariel, Vered Elishar-Malka, Shuo Seah, Dana Weimann-Saks, and Gabriel Weimann. 2022. #RumorsCOVID-19: Predicting the Forwarding of Online Rumors in Wuhan, China and in Israel. *Intl. Communication Gazette* 84, 6 (2022), 550–569.
- [14] Nelson A Atehortua and Stella Patino. 2021. COVID-19, a tale of two pandemics: novel coronavirus and fake news messaging. *Health Promotion Intl.* 36, 2 (01 2021), 524–534. [arXiv:https://academic.oup.com/heapro/article-pdf/36/2/524/37083837/daaa140.pdf](https://academic.oup.com/heapro/article-pdf/36/2/524/37083837/daaa140.pdf)
- [15] Jay Amol Bapaye and Harsh Amol Bapaye. 2021. Demographic Factors Influencing the Impact of Coronavirus-Related Misinformation on WhatsApp: Cross-sectional Questionnaire Study. *JMIR Public Health Surveill* 7, 1 (30 Jan 2021), e19858.
- [16] Jason Baumgartner, Savvas Zannettou, Megan Squire, and Jeremy Blackburn. 2020. The Pushshift Telegram Dataset. *Proc. of the Intl. AAAI Conference on Web and Social Media* 14, 1 (May 2020), 840–847.
- [17] Fabrício Benevenuto and Philippe Melo. 2024. Misinformation Campaigns through WhatsApp and Telegram in Presidential Elections in Brazil. *Commun. ACM* 67, 8 (Aug. 2024), 72–77.
- [18] Priscila Biancovilli, Lilla Makszin, and Claudia Jurberg. 2021. Misinformation on social networks during the novel coronavirus pandemic: a quali-quantitative case study of Brazil. *BMC Public Health* 21, 1 (2021), 1200.
- [19] Dinur Blum, Stacy L. Smith, and Adam G. Sanford. 2025. *Infodemics*. John Wiley & Sons, Ltd, 1–3.
- [20] Jeremy Bowles, Horacio Larreguy, and Shelley Liu. 2020. Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe. *PLOS ONE* 15, 10 (10 2020), 1–11.
- [21] Merve Boyacı Yıldırım. 2023. COVID-19 Infodemic: A study on the Fragile Five countries. *Journal of Public Affairs* 23, 1 (2023), e2846.
- [22] Lucas Cabral, José Maria Monteiro, José Wellington Franco da Silva, César Lincoln C Mattos, and Pedro Jorge Chaves Mourao. 2021. FakeWhastApp. BR: NLP and Machine Learning Techniques for Misinformation Detection in Brazilian Portuguese WhatsApp Messages.. In *ICEIS (I)*. 63–74.
- [23] Macaio Cacabro, Wellington Franco, José Maria Monteiro, and Javam Machado. 2023. IARA - An Architectural Model to Assist the Development of Advising Bots for Misinformation Detection. In *Proc. of the 29th Brazilian Symposium on Multimedia and the Web (WebMedia '23)*. ACM, NY, USA, 168–176.
- [24] Josemar Alves Caetano, Gabriel Magno, Marcos Gonçalves, Jussara Almeida, Humberto T. Marques-Neto, and Virgílio Almeida. 2019. Characterizing Attention Cascades in WhatsApp Groups. In *Proc. of the 10th ACM Conference on Web Science (Boston, Massachusetts, USA) (WebSci '19)*. ACM, NY, USA, 27–36.
- [25] Gustavo Cardoso, Rita Sepúlveda, and Inês Narciso. 2022. WhatsApp and audio misinformation during the Covid-19 pandemic. (2022).
- [26] Leticia Cesarino. 2020. Como vencer uma eleição sem sair de casa: a ascensão do populismo digital no Brasil. *Internet & sociedade* 1, 1 (2020), 91–120.
- [27] Ivandro Claudino, Thiago Gadelha, Tiago Vinuto, José Franco, José Monteiro, and Javam Machado. 2023. A real-time platform to monitoring misinformation on telegram. In *Intl. Conference on Enterprise Information Systems*. 271–278.
- [28] Maria-Jesús Colmenero-Ruiz, Francisco-Carlos Paletta, and Audilio Gonzales-Aguilar. 2023. Interactive mapping of Covid-19 misinformation in Ibero-America. *El Profesional de la Información* 32, 5 (2023).
- [29] Cliona Curley, Eugenia Siapera, and Joe Carthy. 2022. Covid-19 Protesters and the Far Right on Telegram: Co-Conspirators or Accidental Bedfellows? *Social Media + Society* 8, 4 (2022), 20563051221129187.
- [30] Philippe de Freitas Melo, Carolina Coimbra Vieira, Kiran Garimella, Pedro O. S. Vaz de Melo, and Fabrício Benevenuto. 2020. Can WhatsApp Counter Misinformation by Limiting Message Forwarding?. In *Complex Networks and Their Applications VIII*, Hocine Cherifi, Sabrina Gaito, José Fernando Mendes, Esteban Moro, and Luis Mateus Rocha (Eds.). Springer, Cham, 372–384.
- [31] Caroline Delmazo and Jonas CL Valente. 2018. Fake news nas redes sociais online: propagação e reações à desinformação em busca de cliques. *Mídia & jornalismo* 18, 32 (2018), 155–169.
- [32] Azza El-Masri, Martin J Riedl, and Samuel Woolley. 2022. Audio misinformation on WhatsApp: A case study from Lebanon. *Harvard Kennedy School Misinformation Review* 3, 4 (2022), 1–13.
- [33] Don Fallis. 2015. What is disinformation? *Library trends* 63, 3 (2015), 401–426.
- [34] Zina Fan, Wenqiang Yin, Han Zhang, Dandan Wang, Chengxin Fan, Zhongming Chen, Jinwei Hu, Dongping Ma, and Hongwei Guo. 2021. COVID-19 Information Dissemination Using the WeChat Communication Index: Retrospective Analysis Study. *J Med Internet Res* 23, 7 (16 Jul 2021), e28563.
- [35] Antônio Diogo Forte Martins, Lucas Cabral, Pedro Jorge Chaves Mourão, José Maria Monteiro, and Javam Machado. 2021. Detection of Misinformation About COVID-19 in Brazilian Portuguese WhatsApp Messages. In *Natural Language Processing and Information Systems*, Elisabeth Métais, Farid Meiane, Helmut Horacek, and Epaminondas Kapetanios (Eds.). Springer, Cham, 199–206.
- [36] Jaynil Gaglani, Yash Gandhi, Shubham Gogate, and Aparna Halbe. 2020. Unsupervised WhatsApp Fake News Detection using Semantic Search. In *2020 4th Intl. Conference on Intelligent Computing and Control Systems (ICICCS)*. 285–289.
- [37] Marie-Catherine Gagnon-Dufresne, Mayana Azevedo Dantas, Kellyanne Abreu Silva, Jean Souza dos Anjos, Delano Pessoa Carneiro Barbosa, Rebeca Porto Rosa, William de Luca, Monica Zahreddine, Andrea Caprara, Valéry Ridde, et al. 2023. Social media and the influence of fake news on global health interventions: implications for a study on dengue in Brazil. *Intl. Journal of Environmental Research and Public Health* 20, 7 (2023), 5299.
- [38] Cláudia Pereira Galhardi, Neyson Pinheiro Freire, Maria Clara Marques Fagundes, Maria Cecilia de Souza Minayo, and Isabel Cristina Kowal Olm Cunha. 2022. Fake News and vaccine hesitancy in the COVID-19 pandemic in Brazil. *Ciência & Saúde Coletiva* 27 (2022), 1849–1858.
- [39] Cláudia Pereira Galhardi, Neyson Pinheiro Freire, Maria Cecilia de Souza Minayo, and Maria Clara Marques Fagundes. 2020. Fact or fake? An analysis of disinformation regarding the Covid-19 pandemic in Brazil. *Ciencia & saude coletiva* 25 (2020), 4201–4210.
- [40] Kiran Garimella and Dean Eckles. 2020. Images and Misinformation in Political Groups: Evidence from WhatsApp in India. *Harvard Kennedy School (HKS) Misinformation Review* (2020). <https://doi.org/10.37016/mr-2020-030>
- [41] Jacob Gursky, Martin J. Riedl, Katie Joseff, and Samuel Woolley. 2022. Chat Apps and Cascade Logic: A Multi-Platform Perspective on India, Mexico, and the United States. *Social Media + Society* 8, 2 (2022), 20563051221094773.
- [42] Scott A Hale, Adriano Belisario, Ahmed Nasser Mostafa, and Chico Camargo. 2024. Analyzing misinformation claims during the 2022 Brazilian general election on WhatsApp, Twitter, and Kwai. *Intl. Journal of Public Opinion Research* 36, 3 (2024), ead032.
- [43] Uswatun Hasanah and Busro Busro. 2023. Hold Your Fingers: The Communication Ethics on WhatsApp Based on the Hadith. *Jurnal Komunikasi: Malaysian Journal of Communication* 39, 3 (2023), 239–254.
- [44] Lingnan He, Jing Gu, Dan Li, and Kaisheng Lai. 2019. Content Characteristics and Transmission Strategies of Social Media Rumors in China: Big Data Analysis of WeChat Rumors. In *2019 6th Intl. Conference on Behavioral, Economic and Socio-Cultural Computing (BESCom)*. 1–5.
- [45] Aliaksandr Herasimenka, Jonathan Bright, Aleks Knuutila, and Philip N. Howard. 2023. Misinformation and professional news on largely unmoderated platforms: the case of telegram. *Journal of Information Technology & Politics* 20, 2 (2023), 198–212.

- [46] Paula Herrero-Diz, Jesús Conde-Jiménez, and Salvador Reyes de Cózar. 2020. Teens' Motivations to Spread Fake News on WhatsApp. *Social Media + Society* 6, 3 (2020), 2056305120942879.
- [47] Mohamad Hoseini, Philippe Melo, Fabricio Benevenuto, Anja Feldmann, and Savvas Zannettou. 2023. On the Globalization of the QAnon Conspiracy Theory Through Telegram. In *Proc. of the 15th ACM Web Science Conference 2023* (Austin, TX, USA) (*WebSci '23*). ACM, NY, USA, 75–85.
- [48] Qing Huang, Sihan Lei, and Binbin Ni. 2022. Perceived Information Overload and Unverified Information Sharing on WeChat Amid the COVID-19 Pandemic: A Moderated Mediation Model of Anxiety and Perceived Herd. *Frontiers in Psychology* Volume 13 - 2022 (2022).
- [49] Beth Hurford, Abhishek Rana, and Rohan Sachan. 2022. Narrative-based misinformation in India about protection against COVID-19: not just another "moo-point". *Indian journal of medical ethics* 7, 1 (2022), 1–10.
- [50] Ebuha Elias Igwebuike and Lily Chimuanya. 2021. Legitimizing falsehood in social media: A discourse analysis of political fake news. *Discourse & Communication* 15, 1 (2021), 42–58.
- [51] R. Tallal Javed, Mirza Elaaf Shuja, Muhammad Usama, Junaid Qadir, Waleed Iqbal, Gareth Tyson, Ignacio Castro, and Kiran Garimella. 2020. A First Look at COVID-19 Messages on WhatsApp in Pakistan. In *2020 IEEE/ACM Intl. Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 118–125.
- [52] R Tallal Javed, Muhammad Usama, Waleed Iqbal, Junaid Qadir, Gareth Tyson, Ignacio Castro, and Kiran Garimella. 2022. A deep dive into COVID-19-related messages on WhatsApp in Pakistan. *Social Network Analysis and Mining* 12 (2022), 1–16.
- [53] D. Jayaram, M. Venu Gopalachari, S. Rakesh, J. Shiva Sai, and G. Kiran Kumar. 2022. Fake Face Image Detection Using Feature Network. *Intl. Journal of Health Sciences* 6, S5 (2022), 3027–3039.
- [54] Jiajian Jiang, Xiaoliang Chen, Zexia Huang, Xianrong Li, and Yajun Du. 2023. Deep reinforcement learning-based approach for rumor influence minimization in social networks. *Applied Intelligence* 53, 17 (2023), 20293–20310.
- [55] Ramakrishnan Kalaimagal, Balakrishnan Vimala, and Soo Mun Chong. 2023. COVID-19 Fake News Detection in Malaysia – A Supervised Approach. In *Proc. of the 2023 12th Intl. Conference on Software and Computer Applications* (Kuantan, Malaysia) (*ICSCA '23*). ACM, NY, USA, 165–170.
- [56] Antonis Kalogeropoulos and Patricia Rossini. 2025. Unraveling WhatsApp group dynamics to understand the threat of misinformation in messaging apps. *New Media & Society* 27, 3 (2025), 1625–1650.
- [57] Alexandros Karakikes, Panagiotis Alexiadis, and Konstantinos Kotis. 2024. Bias in X (Twitter) and Telegram Based Intelligence Analysis: Exploring Challenges and Potential Mitigating Roles of AI. *SN Computer Science* 5, 5 (2024), 574.
- [58] Manmeet Kaur, Priyanka Daryani, Megha Varshney, and Rishabh Kaushal. 2022. Detection of fake images on whatsapp using socio-temporal features. *Social Network Analysis and Mining* 12, 1 (2022), 58.
- [59] Staffs Keele et al. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report. Technical report, ver. 2.3 ebse technical report. ebse.
- [60] Janet Kemei, Dominic A Alaazi, Mia Tulli, Megan Kennedy, Modupe Tunde-Biyass, Paul Bailey, Ato Sekyi-Otu, Sharon Murdoch, Habiba Mohamud, Jeanne Lehman, et al. 2022. A scoping review of COVID-19 online mis/disinformation in Black communities. *Journal of global health* 12 (2022), 05026.
- [61] Pooja Khurana and Deepak Kumar. 2018. Sir model for fake news spreading through whatsapp. In *Proc. of 3rd Intl. conference on Internet of Things and connected technologies (ICIoTCT)*, 26–27.
- [62] Pooja Khurana, Deepak Kumar, and Sanjeev Kumar. 2019. Research of fake news spreading through WhatsApp. *Intl. Journal of Innovative Technology and Exploring Engineering (IJITEE)* 8, 6S4 (2019).
- [63] Marcelo Kischinhevsky, Itala Maduell Vieira, João Guilherme Bastos dos Santos, Viktor Chagas, Miguel de Andrade Freitas, and Alessandra Aldé. 2020. WhatsApp audios and the remediation of radio: Disinformation in Brazilian 2018 presidential election. *radio journal: Intl. studies in broadcast & audio media* 18, 2 (2020), 139–158.
- [64] Neta Kligler-Vilenchik. 2022. Collective Social Correction: Addressing Misinformation through Group Practices of Information Verification on WhatsApp. *Digital Journalism* 10, 2 (2022), 300–318.
- [65] Vanja Kopilaš and Srećko Gajović. 2020. Wildfire-Like Effect of a WhatsApp Campaign to Mobilize a Group of Predominantly Health Professionals With a University Degree on a Health Issue: Infodemiology Study. *J Med Internet Res* 22, 8 (10 Aug 2020), e17051.
- [66] Prasad Kulkarni, Suyash Karwande, Rhucha Kesar, Prashant Kale, and Sumitra Iyer. 2021. Fake news detection using machine learning. In *ITM Web of Conferences*, Vol. 40. EDP Sciences, 03003.
- [67] Paula Laboissière. 2025. Desinformação é principal risco global para 2025, afirma Unesco. Agência Brasil. <https://agenciabrasil.ebc.com.br/geral/noticia/2025-05/desinformacao-e-principal-risco-global-para-2025-e-anos-subsequentes-Online>. Acessado em: 2025-07-20.
- [68] Ivan Ka Wai Lai and Guicheng Shi. 2015. The impact of privacy concerns on the intention for continued use of an integrated mobile instant messaging and social network platform. *International Journal of Mobile Communications* 13, 6 (2015), 641–669.
- [69] Alkesh S Lajurkar and AU Chaudhari. 2023. Implementing A Passive Aggressive Classifier To Detect False Information. *Intl. Journal of Advanced Research in Computer and Communication Engineering* 12, 4 (2023), 767–774.
- [70] Rhys Leahy, Nicholas Johnson Restrepo, Richard Sear, and Neil F. Johnson. 2022. Connectivity Between Russian Information Sources and Extremist Communities Across Social Media Platforms. *Frontiers in Political Science* 4 (2022).
- [71] Tuan-He Lee and Susan R. Fussell. 2025. Countering Misinformation in Private Messaging Groups: Insights From a Fact-checking Chatbot. *Proc. ACM Hum.-Comput. Interact.* 9, 1, Article GROUP10 (Jan. 2025), 30 pages.
- [72] Ziyu Li, Xiaoqian Wu, Lin Xu, Ming Liu, and Cheng Huang. 2023. Hot Topic Recognition of Health Rumors Based on Anti-Rumor Articles on the WeChat Official Account Platform: Topic Modeling. *J Med Internet Res* 25 (2023), e45019.
- [73] Huan Liu, Qiang Chen, and Richard Evans. 2022. How official social media affected the infodemic among adults during the first wave of COVID-19 in China. *Journal of Environmental Research and Public Health* 19, 11 (2022), 6751.
- [74] Caio Machado, Beatriz Kira, Vidya Narayanan, Bence Kollanyi, and Philip Howard. 2019. A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. In *Companion Proc. of The 2019 World Wide Web Conference*. ACM, NY, USA, 1013–1019.
- [75] Pranav Malhotra. 2020. A Relationship-Centered and Culturally Informed Approach to Studying Misinformation on COVID-19. *Social Media + Society* 6, 3 (2020), 2056305120948224.
- [76] Ammara Malik, Faiza Bashir, and Khalid Mahmood. 2023. Antecedents and Consequences of Misinformation Sharing Behavior among Adults on Social Media during COVID-19. *SAGE Open* 13, 1 (2023), 21582440221147022.
- [77] Alexandre Maros, Jussara M. Almeida, and Marisa Vasconcelos. 2021. A Study of Misinformation in Audio Messages Shared in WhatsApp Groups. In *Disinformation in Open Online Media*, Jonathan Bright, Anastasia Giachanou, Viktoria Spaiser, Francesca Spezzano, Anna George, and Alexandra Pavliuc (Eds.). Springer, Cham, 85–100.
- [78] Jacinta Mwende Maweu. 2019. "Fake Elections"? Cyber Propaganda, Disinformation and the 2017 General Elections in Kenya. *African Journalism Studies* 40, 4 (2019), 62–76.
- [79] Ben Medeiros and Pawan Singh. 2020. Addressing Misinformation on Whatsapp in India Through Intermediary Liability Policy, Platform Design Modification, and Media Literacy. *Journal of Information Policy* 10 (05 2020), 276–298.
- [80] Philippe Melo, Johnnatn Messias, Gustavo Resende, Kiran Garimella, Jussara Almeida, and Fabricio Benevenuto. 2019. WhatsApp Monitor: A Fact-Checking System for WhatsApp. *Proc. of the Intl. AAAI Conference on Web and Social Media* 13, 01 (Jul. 2019), 676–677.
- [81] Philippe de Freitas Melo, Mohamad Hoseini, Savvas Zannettou, and Fabricio Benevenuto. 2024. Don't Break the Chain: Measuring Message Forwarding on WhatsApp. *Proc. of the Intl. AAAI Conference on Web and Social Media* 18 (May 2024), 1054–1067.
- [82] Camila Mont'Alverne, Amy Ross Arguedas, Sayan Banerjee, Benjamin Toff, Richard Fletcher, and Rasmus Kleis Nielsen. 2024. The Electoral Misinformation Nexus: How News Consumption, Platform Use, and Trust in News Influence Belief in Electoral Misinformation. *Public Opinion Quarterly* 88, SI (07 2024), 681–707. [arXiv:https://academic.oup.com/poq/article-pdf/88/SI/681/58740401/nfae019.pdf](https://academic.oup.com/poq/article-pdf/88/SI/681/58740401/nfae019.pdf)
- [83] Carolina Moreno-Castro, Empar Vengut-Climent, Lorena Cano-Orón, and Isabel Mendoza-Poudereux. 2022. Exploratory study of the hoaxes spread via WhatsApp in Spain to prevent and/or cure COVID-19. *Gaceta sanitaria* 35 (2022), 534–540.
- [84] Lynnette Hui Xian Ng and Jia Yuan Loke. 2021. Analyzing Public Opinion and Misinformation in a COVID-19 Telegram Group Chat. *IEEE Internet Computing* 25, 2 (2021), 84–91.
- [85] Pinky Kusuma Ningtyas, Hayuni Retno Widarti, and Parlan Parlan. 2024. Exploring the Use of Social Media in Science Learning Environments: A Systematic Literature Review. *Journal of Science Learning* 7, 2 (2024), 178–186.
- [86] Gabriel Peres Nobre, Carlos H.G. Ferreira, and Jussara M. Almeida. 2022. A hierarchical network-oriented analysis of user participation in misinformation spread on WhatsApp. *Information Processing & Management* 59, 1 (2022), 102757.
- [87] OECD. 2024. Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity. OECD Publishing. <https://doi.org/10.1787/d909ff7a-en> Acessado em: 2025-07-20.
- [88] Godwin Ayigbo Owojecho. 2021. Social Media and Fake News in Nigeria: A Speech Act Analysis of WhatsApp Messages on Coronavirus. *Studies in Pragmatics and Discourse Analysis* 2, 1 (Jun. 2021), 56–66.
- [89] Joao V. S. Ozawa, Samuel C. Woolley, Joseph Straubhaar, Martin J. Riedl, Katie Joseff, and Jacob Gursky. 2023. How Disinformation on WhatsApp Went From Campaign Weapon to Governmental Propaganda in Brazil. *Social Media + Society* 9, 1 (2023), 20563051231160632.
- [90] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw,

- Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021). <https://doi.org/10.1136/bmj.n71> arXiv:<https://www.bmj.com/content/372/bmj.n71.full.pdf>
- [91] Hua Pang, Jun Liu, and Jiahui Lu. 2022. Tackling fake news in socially mediated public spheres: A comparison of Weibo and WeChat. *Technology in Society* 70 (2022), 102004.
- [92] Irene V. Pasquetto, Eaman Jahani, Shubham Atreja, and Matthew Baum. 2022. Social Debunking of Misinformation on WhatsApp: The Case for Strong and In-group Ties. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 117 (April 2022), 35 pages.
- [93] David Araújo Pinheiro, Mariana Gomes Leitão De Araújo, Keilla Barbosa De Souza, BDS Campos, EM De Oliveira, RSM Lima, GA Ferreira, ACA De Freitas, CB Toledo, GB De Souza, et al. 2020. Sharing fake news about health in the cross-platform messaging app WhatsApp during the COVID-19 pandemic: A pilot study. *Int. J. Sci. Res. Manag* 8 (2020), 403–410.
- [94] V.T Priyanga, J.P Sanjanasri, Vijay Krishna Menon, E.A Gopalakrishnan, and K.P Soman. 2021. Exploring fake news identification using word and sentence embeddings. *Journal of Intelligent & Fuzzy Systems* 41, 5 (2021), 5441–5448.
- [95] Alisson Puska, Luiz Adolpho Baroni, Máira Codo Canal, Lara S. G. Piccolo, and Roberto Pereira. 2020. WhatsApp and false information: a value-oriented evaluation. In *Proc. of the 19th Brazilian Symposium on Human Factors in Computing Systems* (Diamantina, Brazil) (IHC '20). ACM, NY, USA, Article 37, 10 pages.
- [96] Marta Pérez-Escobar, Darren Lilleker, and Alejandro Tapia-Frade. 2023. A Systematic Literature Review of the Phenomenon of Disinformation and Misinformation. *Media and Communication* 11, 2 (2023), 76–87.
- [97] Adnan Qayyum, Junaid Qadir, Muhammad Umar Janjua, and Falak Sher. 2019. Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News. *IT Professional* 21, 4 (2019), 16–24.
- [98] Muhammad Rahulil, Yuni Yamasari, Ricky Eka Putra, I Made Suartana, and Anita Qoiriah. 2025. A Systematic Literature Review on Chatbot Development For Whatsapp: Programming Language, Method, And Utility. *Jurnal Serambi Engineering* 10, 3 (2025).
- [99] Raquel Recuero, Felipe Bonow Soares, Otávio Vinhas, et al. 2021. Discursive strategies for disinformation on WhatsApp and Twitter during the 2018 Brazilian presidential election. *First Monday* 26, 1 (2021).
- [100] Julio CS Reis, Philippe Melo, Kiran Garimella, and Fabricio Benevenuto. 2020. Can WhatsApp benefit from debunked fact-checked stories to reduce misinformation? *Harvard Kennedy School (HKS) Misinformation Review* (2020). <https://doi.org/10.37016/mr-2020-035>
- [101] Julio CS Reis, Philippe Melo, Kiran Garimella, and Fabricio Benevenuto. 2020. Detecting misinformation on WhatsApp without breaking encryption. *Association for the Advancement of Artificial Intelligence* (2020).
- [102] Julio CS Reis, Philippe Melo, Márcio Silva, and Fabricio Benevenuto. 2023. Desinformação em plataformas digitais: Conceitos, abordagens tecnológicas e desafios. *Jornada de Atualização em Informática (JAI), Sociedade Brasileira de Computação* (2023).
- [103] Julio C. S. Reis and Fabricio Benevenuto. 2021. Supervised Learning for Misinformation Detection in WhatsApp. In *Proc. of the Brazilian Symposium on Multimedia and the Web* (Belo Horizonte, Minas Gerais, Brazil) (WebMedia '21). ACM, NY, USA, 245–252.
- [104] Julio C. S. Reis, Philippe Melo, Fabiano Belém, Fabricio Murai, Jussara M. Almeida, and Fabricio Benevenuto. 2023. Helping Fact-Checkers Identify Fake News Stories Shared through Images on WhatsApp. In *Proc. of the 29th Brazilian Symposium on Multimedia and the Web* (WebMedia '23). ACM, NY, USA, 159–167.
- [105] Julio C. S. Reis, Philippe Melo, Kiran Garimella, Jussara M. Almeida, Dean Eckles, and Fabricio Benevenuto. 2020. A Dataset of Fact-Checked Images Shared on WhatsApp During the Brazilian and Indian Elections. *Proc. of the Intl. AAAI Conference on Web and Social Media* 14, 1 (May 2020), 903–908.
- [106] Gustavo Resende, Philippe Melo, Julio C. S. Reis, Marisa Vasconcelos, Jussara M. Almeida, and Fabricio Benevenuto. 2019. Analyzing Textual (Mis)Information Shared in WhatsApp Groups. In *Proc. of the 10th ACM Conference on Web Science* (Boston, USA) (WebSci '19). ACM, NY, USA, 225–234.
- [107] Gustavo Resende, Philippe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabricio Benevenuto. 2019. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *The World Wide Web Conference (WWW '19)*. ACM, NY, USA, 818–828.
- [108] Dhiren Rohera, Harshal Shethna, Keyur Patel, Urvi Thakker, Sudeep Tanwar, Rajesh Gupta, Wei-Chiang Hong, and Ravi Sharma. 2022. A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects. *IEEE Access* 10 (2022), 30367–30394.
- [109] Patricia Rossini. 2023. Farewell to Big Data? Studying Misinformation in Mobile Messaging Applications. *Political Communication* 40, 3 (2023), 361–366.
- [110] G Senthilkumar and D Ashok Kumar. 2023. A Comparative Study on Various Machine Learning Algorithms for the Prediction of Fake News Detections Using Bring Feed New Data Sets. *Intl. Journal of Scientific Research in Computer Science, Engineering and Information Technology* (2023).
- [111] Anjana E Sharma, Kiran Khosla, Kameswari Potharaju, Arnab Mukherjee, and Urmimala Sarkar. 2023. COVID-19–Associated Misinformation Across the South Asian Diaspora: Qualitative Study of WhatsApp Messages. *JMIR Infodemiology* 3 (5 Jan 2023), e38607.
- [112] Mark P. Shephard, David J. Robertson, Narisong Huhe, and Anthony Anderson. 2023. Everyday non-partisan fake news: Sharing behavior, platform specificity, and detection. *Frontiers in Psychology* Volume 14 - 2023 (2023).
- [113] Ruth Stewart, Andile Madonsela, Nkululeko Tshabalala, Linda Etale, and Nicola Theunissen. 2022. The importance of social media users' responses in tackling digital COVID-19 misinformation in Africa. *DIGITAL HEALTH* 8 (2022), 20552076221085070.
- [114] S Shyam Sundar, Maria D Molina, and Eugene Cho. 2021. Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps? *Journal of Computer-Mediated Communication* 26, 6 (08 2021), 301–319. arXiv:<https://academic.oup.com/jcmc/article-pdf/26/6/301/41139661/zmab010.pdf>
- [115] Inga K. Trauthig, Zelly C. Martin, and Samuel C. Woolley. 2024. Messaging Apps: A Rising Tool for Informational Autocrats. *Political Research Quarterly* 77, 1 (2024), 17–29.
- [116] Arsenii Tretiakov, Alejandro Martín, and David Camacho. 2022. Detection of False Information in Spanish Using Machine Learning Techniques. In *Intelligent Data Engineering and Automated Learning – IDEAL 2022*. Hujun Yin, David Camacho, and Peter Tino (Eds.), Springer, Cham, 42–53.
- [117] Yan Tu, Linqi Cheng, Liyi Liu, Zongmin Li, and Benjamin Lev. 2023. How to improve the rumor-confutation ability of official rumor-refuting account on social media: A Chinese case study. *Computer Communications* 209 (2023), 331–348.
- [118] P Umamaheswari, N Umasankari, and Selvakumar Samuel. 2023. Fake News Detection With the Help of Computation Time to Increase Accuracy. In *Risk Detection and Cyber Security for the Success of Contemporary Computing*. IGI Global, 218–240.
- [119] Rama Adithya Varanasi, Joyojeet Pal, and Aditya Vashistha. 2022. Accost, Accede, or Amplify: Attitudes towards COVID-19 Misinformation on WhatsApp in India. In *Proc. of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). ACM, NY, USA, Article 256, 17 pages.
- [120] Amit Verma, Maninder Kumar Singh, and Aditi Pareek. 2020. Information, misinformation, and disinformation about COVID-19: A content study of closed-closed platform messaging using Whatsapp. *Intl. Journal of Advance Science and Technology* 29, 10 (2020), 7797–7804.
- [121] Santosh Vijaykumar, Daniel T Rogerson, Yan Jin, and Mariella Silva de Oliveira Costa. 2021. Dynamics of social corrections to peers sharing COVID-19 misinformation on WhatsApp in Brazil. *Journal of the American Medical Informatics Association* 29, 1 (11 2021), 33–42. arXiv:<https://academic.oup.com/jamia/article-pdf/29/1/33/41955567/ocab219.pdf>
- [122] Yaqing Wang, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, and Jing Gao. 2020. Weak Supervision for Fake News Detection via Reinforcement Learning. *Proc. of the AAAI Conference on Artificial Intelligence* 34, 01 (Apr. 2020), 516–523.
- [123] WHO. 2025. Infodemic. World Health Organization – Health Topics. https://www.who.int/health-topics/infodemic#tab=tab_1 Online. Acessado em: 2025-07-14.
- [124] Tom Willaert, Stijn Peeters, Jasmin Seijbel, and Nathalie Van Raemdonck. 2022. Disinformation networks: A qualitative investigation of antagonistic Dutch-speaking Telegram channels. *First Monday* 27, 5 (Sep. 2022).
- [125] Maïke Winters, Ben Oppenheim, Paul Sengeh, Mohammad B Jalloh, Nance Webber, Samuel Abu Pratt, Bailah Leigh, Helle Molsted-Alvesson, Zangin Zebari, Carl Johan Sundberg, et al. 2021. Debunking highly prevalent health misinformation using audio dramas delivered by WhatsApp: evidence from a randomised controlled trial in Sierra Leone. *BMJ global health* 6, 11 (2021), e006954.
- [126] Veronika J Wirtz, Gabriel Millán-Garduño, Jennifer Hegewisch-Taylor, Anahi Dreser, Andrea Anaya-Sanchez, Tonatiuh T González-Vázquez, Ricardo Escalera, and Pilar Torres-Pereda. 2023. Misinformation messages shared via WhatsApp in Mexico during the COVID-19 pandemic: an exploratory study. *Health Promotion Intl.* 38, 3 (05 2023), daad041. arXiv:<https://academic.oup.com/heapro/article-pdf/38/3/daad041/50200070/daad041.pdf>
- [127] Andy Wai Kan Yeung, Anela Tosevska, Elisabeth Klager, Fabian Eibensteiner, Christos Tsagkaris, Emil D Parvanov, Faisal A Nawaz, Sabine Völkl-Kernstock, Eva Schaden, Maria Kletecka-Pulker, Harald Willschke, and Atanas G Atanasov. 2022. Medical and Health-Related Misinformation on Social Media: Bibliometric Study of the Scientific Literature. *J Med Internet Res* 24, 1 (25 Jan 2022), e28152.
- [128] Yue Zhuang, Tiantian Zhao, and Xuanrong Shao. 2021. Mechanism of WeChat's Impact on Public Risk Perception During COVID-19. *Risk Management and Healthcare Policy* 14 (2021), 4223–4233.
- [129] Luisa Dolores Zozaya-Durazo, Charo Sádaba-Chalezquer, and Beatriz Feijoo-Fernández. 2024. "Fake or not, I'm sharing it": teen perception about disinformation in social networks. *Young Consumers* 25, 4 (2024), 425–438.