



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

ORLANDO VERDASCA ACETO

**DESINFORMAÇÃO EM REDES SOCIAIS: análise do crescimento e estratégias
de combate**

Recife
2025

ORLANDO VERDASCA ACETO

**DESINFORMAÇÃO EM REDES SOCIAIS: análise do crescimento e estratégias
de combate**

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Engenharia da Computação da
Universidade Federal de Pernambuco,
como requisito parcial para obtenção do
título de bacharel em Engenharia da
Computação

Orientador: Cleber Zanchettin

Recife
2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Aceto, Orlando Verdasca.

Desinformação em redes sociais: análise do crescimento e estratégias de
combate / Orlando Verdasca Aceto. - Recife, 2025.
42 p., tab.

Orientador(a): Cleber Zanchettin

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro de Informática, Engenharia da Computação - Bacharelado,
2025.

Inclui referências.

1. Desinformação. 2. Redes sociais. 3. Engenharia de software. I. Zanchettin,
Cleber. (Orientação). II. Título.

000 CDD (22.ed.)

ORLANDO VERDASCA ACETO

**DESINFORMAÇÃO EM REDES SOCIAIS: análise do crescimento e estratégias
de combate**

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Engenharia da Computação da
Universidade Federal de Pernambuco,
como requisito parcial para obtenção do
título de bacharel em Engenharia da
Computação

Aprovado em: 28/07/2025

BANCA EXAMINADORA

Prof. Cleber Zanchettin (Orientador)

Universidade Federal de Pernambuco

Prof. Adriano Lorena Inácio de Oliveira (Examinador Interno)

Universidade Federal de Pernambuco

AGRADECIMENTOS

Agradeço a todas as pessoas que passaram pela minha vida nos últimos anos e de alguma forma contribuíram para a história da realização deste trabalho, estejam eles hoje ainda presentes ou não. Em particular, agradeço a meus familiares, minha psicóloga, meu orientador e todos os professores que tive no CIn desde o primeiro semestre até o último.

RESUMO

O fenômeno da desinformação na internet é um assunto que ganhou espaço no debate público nos últimos anos. Muitos pesquisadores, desenvolvedores de redes sociais e jornalistas tentam conter o problema. Com o acontecimento da pandemia de Covid-19 e em diversas eleições e debates públicos pelo mundo, isso provou ser mais difícil que o esperado. Este trabalho analisa os meios pelos quais essa desinformação se espalha, as diversas estratégias propostas para contê-la e as implicações disso para a sociedade, o jornalismo e a engenharia de software, concluindo que apesar do problema se mostrar ainda em crescimento, em particular com o surgimento de novas tecnologias baseados em IA, as medidas contra ele não o acompanha, o que deixa lacunas abertas para o agravamento do problema e fazendo-se necessário que novas estratégias e tecnologias para seu combate sejam desenvolvidas.

Palavras-chave: desinformação; engenharia de software; redes sociais; jornalismo.

ABSTRACT

The phenomenon of online disinformation is a subject that gained space in the public discourse in the last few years. Many researchers, social network developers and journalists try to solve the problem. With the covid-19 pandemic, many elections around the world and other political discussions, this proved to be harder than expected. This essay analyzes the means by which disinformation spreads, the many strategies proposed to contain it and the implications for society, journalism and software engineering, concluding that in spite of the problem's continued growth, in particular with the introduction of new AI based technologies, the measures taken against it seems to have stagnated, leaving gaps which allow the problem to get worse and making new strategies and technologies to fight against it necessary.

Keywords: disinformation; software engineering; social networks; journalism.

LISTA DE GRÁFICOS

Gráfico 1 –	Proporção das pessoas que dizem que redes sociais são sua principal fonte de notícias	11
Gráfico 2 –	Proporção das pessoas que usaram cada uma como fontes de notícias na última semana	11
Gráfico 3 –	Número de fact-checkers ativos por ano	15
Gráfico 4 –	Proporção do tema vacinas nas “fake news” publicadas em redes sociais	21
Gráfico 5 –	Engajamento no Facebook das 20 notícias mais proeminentes	22
Gráfico 6 –	Proporção de americanos acreditando em teorias da conspiração históricas	23
Gráfico 7 –	Número de usuários mensais das redes sociais mais populares (em milhões)	26

SUMÁRIO

1	INTRODUÇÃO	8
1.1	Objetivo geral	10
1.2	Objetivos específicos	11
2	METODOLOGIA	12
3	DESINFORMAÇÃO	13
3.1	Como definir o que é desinformação?	13
3.2	Como detectar desinformação	14
3.2.1	<i>Agências de Fact Checking</i>	14
3.2.2	<i>Fact-checking automático</i>	16
3.2.3	<i>IA Generativa</i>	17
3.3	Exemplos	19
3.3.1	<i>Covid 19 e vacinas</i>	20
3.3.2	<i>Eleições</i>	22
3.3.3	<i>Palestina e Israel</i>	24
4	REDES SOCIAIS	26
4.1	Políticas acerca de desinformação	27
4.1.1	<i>Meta</i>	27
4.1.2	<i>Youtube</i>	29
4.1.3	<i>TikTok</i>	30
4.2	Ações específicas	31
4.3	Análise dos achados	34
5	CONSIDERAÇÕES FINAIS	36
5.1	Limitações	37
5.2	Trabalhos futuros	37
	REFERÊNCIAS	38

1 INTRODUÇÃO

Em Novembro de 1997, a ACM (Association for Computing Machinery), e a IEEE (Institute of Electrical and Electronic Engineers) publicaram o primeiro código de ética da Engenharia de Software. Em sua versão atualizada, ela tem como preceito 1: “Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing”[1] (ACM, 2018). Tal preceito já constava desde a primeira versão.

A ideia de que o engenheiro de software deve priorizar o interesse da sociedade não é nova. No entanto, a crescente importância da tecnologia e a forma como molda a sociedade atual apenas tornou essa ideia ainda mais indispensável. A importância da internet e do smartphone na comunicação, as tecnologias educacionais, a automatização dos sistemas de transporte são apenas alguns dos inúmeros benefícios que a engenharia de software e a tecnologia nos proporcionam.

É de se esperar, portanto, que com o crescimento da importância do software em nossas vidas, cresce também a responsabilidade que o Engenheiro de Software carrega, bem como das empresas, governos e organizações que os empregam.

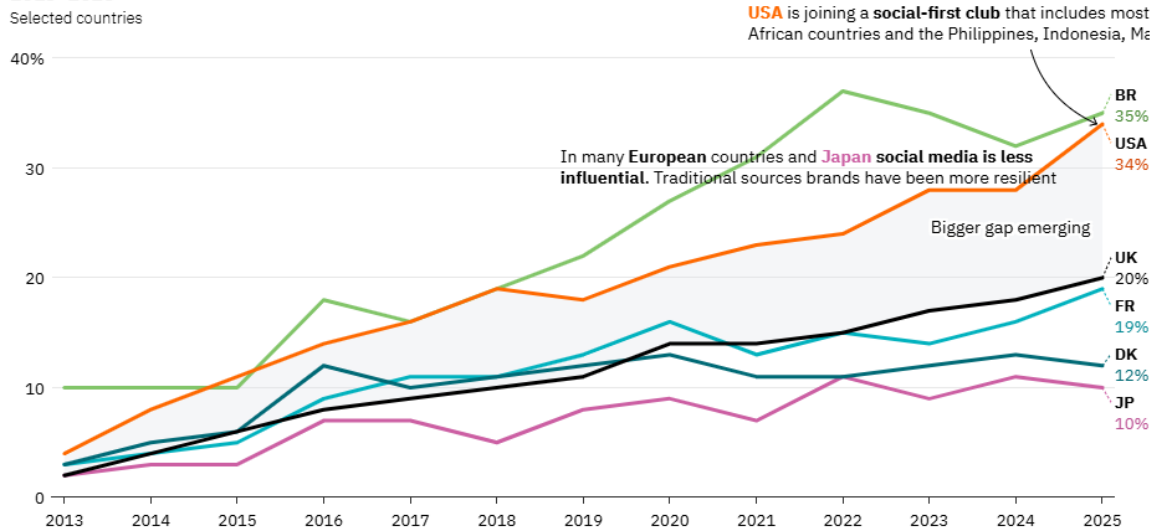
De todos os impactos causados pelo novo paradigma que o advento da computação trouxe à sociedade, certamente um dos maiores foi na área da comunicação. Especialmente o crescimento da importância das redes sociais na forma como as pessoas se informam é notável, como mostra pesquisa publicada pela Reuters.

[1] Em tradução livre: Contribuir para a sociedade e o bem-estar humano, entendendo que todas as pessoas são partes interessadas na computação

Gráfico 1 – Proporção das pessoas que dizem que redes sociais são sua principal fonte de notícias

Proportion that say social media is their main source of news**2013–2025**

Selected countries



Q4. You say you've used these sources of news in the last week, which would you say is your MAIN source of news? Base: All that used a source of news in the last week in each country-year ≈ 1900. Note: No data for 2014. There was a sampling and weighting change from 2021 onwards.

Fonte: Reuters Institute (2025)

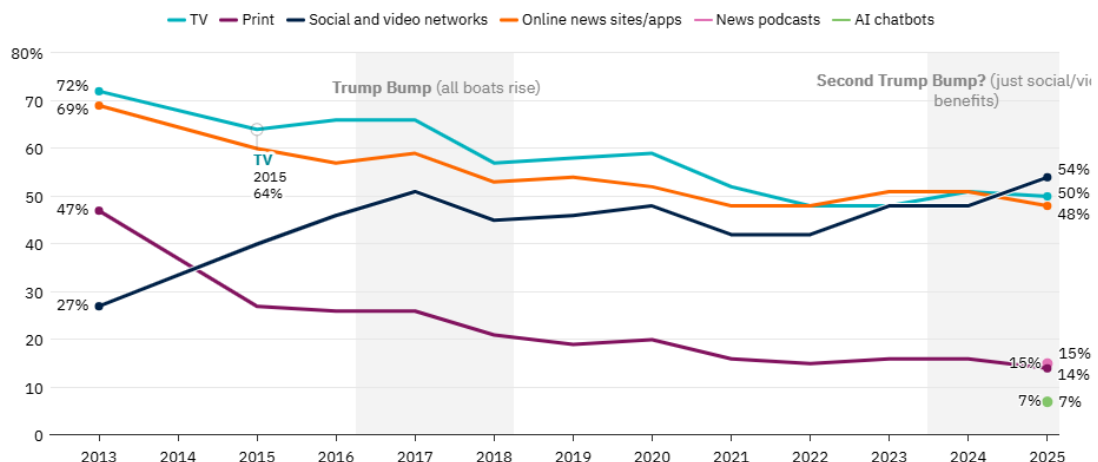
Na mesma pesquisa, foi feito um comparativo do crescimento do consumo de notícias através das novas mídias com outras mídias tradicionais:

Gráfico 2 – Proporção das pessoas que usaram cada uma como fontes de notícias na última semana

Proportion that used each as a source of news in the last week**2013–2025**

United States

→ TV down **-22 points** | → Print down **-33** | → Social and video networks above TV for first time | → Online news sites down **-21** | → News podcasts = **15%**, more than Radio = **13%** | → AI news emerging



Q3. Which, if any, of the following have you used in the last week as a source of news? Base: Total sample in each year ≈ 2000. Note: No data for 2014. There was a sampling and weighting change from 2021 onwards.

Fonte: Reuters Institute (2025)

Um espaço que, anos atrás, era quase completamente feito por mídias tradicionais como TV, jornal e rádio, hoje tem se tornado cada vez mais dominado por novas tecnologias, em especial as redes sociais e sites de vídeos, que em 2025, após anos sendo a segunda ou terceira fonte mais utilizada, se tornaram a principal fonte de informação nos Estados Unidos.

Mídias tradicionais sempre tiveram seus próprios problemas, como por exemplo o conflito de interesse que ocorre com o dilema entre informar e gerar lucro, o monopólio das organizações de notícias nas mãos de grandes empresas e bilionários e a dependência da distribuição da informação por partes de algumas poucas empresas e conglomerados. Porém, com a mudança nos hábitos da população e na infraestrutura da difusão de informação, surgem problemas, alguns novos e alguns familiares. Dentre eles, se destaca um específico: a desinformação.

Pensando nesse problema e entendendo que o problema está cada vez mais atrelado às redes sociais, surge a seguinte pergunta: Quais políticas, mecanismos e práticas de moderação adotados pela Meta, YouTube e TikTok demonstram maior eficácia na contenção e mitigação da desinformação, e quais lacunas permanecem nesses sistemas?

Nesta pesquisa, será feita uma análise de artigos publicados acerca da desinformação, dados coletados por pesquisadores e jornalistas e informações disponibilizadas pelos responsáveis pelas maiores redes sociais do mundo com os seguintes objetivos:

1.1 Objetivo geral

Entender as formas como a divulgação de desinformação acontece no contexto geral e, em especial, nas maiores redes sociais do mundo em números de usuários, assim como delinear a forma como o problema é tratado por elas, a fim de propor futuros caminhos para um combate mais efetivo e para futuros estudos na área.

1.2 Objetivos específicos

- Elencar alguns dos caminhos propostos por pesquisadores e jornalistas para um melhor combate à divulgação da desinformação;
- Estudar como é feito o fact-checking por terceiros para identificação e combate à desinformação e como as redes sociais podem aprender através deles;
- Mapear as políticas utilizadas pelas maiores redes sociais acerca de desinformação a fim de entender os pontos comuns e as disparidades entre elas;
- Analisar a efetividade das políticas utilizadas atualmente no combate à desinformação utilizando-se da informação disponibilizada por elas e por pesquisadores e jornalistas;
- Identificar problemas nas políticas utilizadas;
- Propor possíveis melhorias nas políticas e possíveis iniciativas que poderiam ser tomadas para combater a desinformação de uma forma mais eficiente.

2 METODOLOGIA

Para chegar aos objetivos propostos, foi feito primeiramente uma revisão bibliográfica acerca do tema desinformação, englobando estudos e publicações jornalísticas feitas principalmente a partir de 2016 sobre o tema, ano em que ganhou maior relevância, dando preferência aos artigos mais relevantes para o contexto buscado, com maiores números de citações ou feitas por jornais de maior reconhecimento a nível mundial, sempre que possível.

Os motores de buscas utilizados na pesquisa foram principalmente o Google e o Google Scholar, usando em geral o primeiro para artigos jornalísticos e o segundo quando buscando informações mais adequadamente apresentadas em artigos científicos.

As pesquisas foram feitas principalmente em inglês, considerada língua franca no mundo, e em português quando buscando informações mais atreladas ao contexto brasileiro. Algumas das principais palavras chave e frases utilizadas nas buscas foram: *misinformation, disinformation, fake news, fact-checking, how to detect disinformation, fact checking agencies, how to automate fact-checking, covid 19, disinformation in politics, fake news eleições 2018 brasil*

Após a rápida revisão bibliográfica acerca de desinformação e os principais temas atrelados a ela, num segundo momento a pesquisa passou a focar especificamente em ações e políticas das 5 maiores redes sociais em números de usuários: Facebook, Whatsapp, Instagram (geralmente pesquisados juntos no guarda-chuva Meta), Youtube e TikTok. Nessa parte da pesquisa, foi também considerado informações publicadas pelas próprias empresas, mas sem considerar as informações como absolutas, buscando contrapontos em publicações jornalísticas ou acadêmicas quando possível.

3 DESINFORMAÇÃO

A desinformação, ou como ficou popularmente conhecida inclusive em países não anglófonos, “fake news”, ganhou holofotes nas eleições presidenciais americanas de 2016 (BBC, 2016). Um fenômeno que, se não novo, se tornou muito maior e passou a causar muito mais impacto desde então. Em 2020, com a pandemia mundial de COVID-19, houve também uma “pandemia de desinformação” (Tagliabue, Fabio et al., 2020).

Redes sociais e redes de vídeos passaram a ser utilizadas, às vezes de forma intencional e às vezes por usuários desavisados, para espalhar desinformação em uma escala cada vez maior. Um problema que é inerente a essas tecnologias, já que tem conteúdo gerado pelos próprios usuários e, portanto, não tem a figura do editor, comum em mídias tradicionais e não necessariamente sofrem consequências pelas suas ações como uma rede de TV ou jornal poderia sofrer caso deliberadamente espalhasse notícias falsas.

3.1 Como definir o que é desinformação?

É difícil definir o que exatamente se encaixa no conceito de desinformação. Poderíamos definir como qualquer informação que não é verdade. Esse conceito, porém, não ajuda em nada a entender o problema, afinal, o conceito de “verdade” é um dos problemas mais antigos da Filosofia. Em artigo publicado na Internet Encyclopedia of Philosophy, o problema é bem ilustrado:

This philosophical problem of truth has been with us for a long time. In the first century AD, Pontius Pilate (John 18:38) asked “What is truth?” but no answer was forthcoming. The problem has been studied more since the turn of the twentieth century than at any other previous time. (Dowden e Swartz, [s.d])

Para algo ser considerado desinformação, portanto, não basta que seja uma informação mentirosa. No inglês, existe a distinção entre dois conceitos diferentes, porém parecidos: “misinformation” e “disinformation”. A American Psychological Association faz a seguinte distinção: “Misinformation is false or inaccurate information—getting the facts wrong. Disinformation is false information which is

deliberately intended to mislead—intentionally misstating the facts.” (American Psychological Association, [s.d]). No português, ambos os conceitos são chamados de desinformação. Apesar de fenômenos psicologicamente diferentes, muitas vezes têm consequências e mecanismos de difusão parecidos. Neste trabalho, portanto, não faremos essa distinção a não ser que seja relevante para o contexto discutido.

3.2 Como detectar desinformação

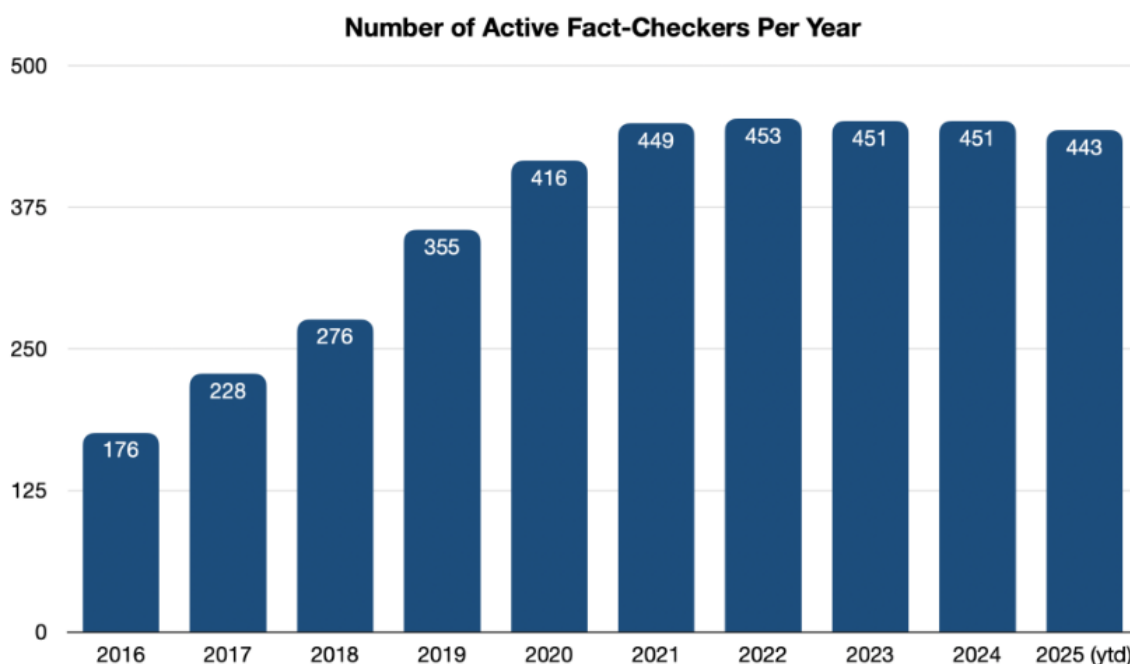
Com um fenômeno difícil de definir com precisão, é natural, então, que muitas vezes seja de difícil identificação. Nem tudo que foge do discurso mais aceito é necessariamente desinformação. O conflito de ideias é natural e o debate deve existir. Com o crescente problema, porém, algumas tentativas de separar o que é desinformação do que não é foram feitas.

3.2.1 Agências de fact checking

Uma das iniciativas mais conhecidas para combater a desinformação desde que o problema se tornou um pilar do discurso público no mundo foi o crescimento de agências de fact checking. Segundo a AFP, detentora de uma das maiores redes de fact-checking: “Fact-checking involves verifying content that has already been published, particularly on social media, and establishing what is known—or not—about a subject.”(AFP, [s.d]).

Fact checking é feita majoritariamente por agências de jornalismo e sites independentes. A Duke Reporters’ Lab mantém uma base de dados de agências de fact-checking em todo o mundo e no censo de 2025 é possível ver claramente como o crescimento se deu a partir de 2016 até 2020 e depois se manteve estável.

Gráfico 3 – Número de fact-checkers ativos por ano



Fonte: Duke Reporters' Lab (2025)

No mesmo censo, a autora deixa claro o que é perceptível nos números: apesar do crescimento passado, o número de fact-checkers estagnou e o futuro é incerto, principalmente com a crescente resistência por parte de políticos e empresas com iniciativas para combater desinformação desde o final de 2024.

No artigo ““Fact-checking” fact checkers: A data-driven approach” (LEE, S. et al., 2023), os autores fazem uma análise comparativa de dados dos veredictos de diferentes agências para determinar se os resultados são consistentes entre si e concluem que, ao ajustar os dados para levar em consideração a diferença entre os sistemas de classificação de cada agência, quase nunca se encontra uma contradição clara, com apenas um caso em mais de 10 mil analisados por cada uma delas.

Apesar de ter a confiança de muitos, o trabalho de fact-checking manual apresenta problema de escalabilidade, já que alguns milhares de jornalistas ficam encarregados de checar conteúdo gerado por potencialmente bilhões de usuários. No artigo “The Quest to Automate Fact-Checking”, os autores delineiam o problema com o fact-checking manual, feito por jornalistas:

The growth of fact-checking has been hampered by the nature of the work. It is time-consuming to find claims to check. Journalists have to spend hours going through transcripts of speeches, debates and interviews to identify claims they will research.

Also, fact-checking requires advanced research techniques. While ordinary journalism can rely on simple “on-the-onehand, on-the-other-hand” quotations, a fact-check requires more thorough research so the journalist can determine the accuracy of a claim. (Hasan et al, 2015, p.1)

Dadas essas limitações, é natural que cientistas busquem maneiras de automatizar o fact-checking.

3.2.2 *Fact checking automático*

Pensando no problema de escalabilidade, outra solução proposta para o problema de detecção de desinformação é a automatização do fact-checking, com vários engenheiros e cientistas estudando soluções. Como esperado, ainda não existe uma solução perfeita para esse problema, já que existem inúmeros problemas ao se tentar automatizar uma tarefa tão humana como essa, tanto relacionado com a questão jornalística, como computacional. Ainda no artigo “The quest to automate fact-checking”, mencionado na seção anterior, é discutido os principais problemas para a área computacional:

On the computational side, there are mainly two fundamental challenges. One is to understand what one says. Computer scientists have made leaps and bounds in speech recognition and Natural Language Processing (NLP). But these technologies are far from perfect. The other challenge lies in our capability of collecting sufficient evidence for checking facts. (Hassan et al., 2015, p. 2)

Segundo os autores do mesmo artigo é possível que se chegue muito perto de uma automatização completa, mas que ela nunca seja totalmente alcançada. Para eles, o “santo graal”, ou seja, uma solução perfeita, teria que ter as seguintes características:

Fully automated: It checks facts without human intervention. It takes as input the video/audio signals and texts of a political discourse and returns factual claims and a truthness rating for each claim (e.g., the Truth-O-Meter by PolitiFact).
 Instant: It immediately reaches conclusions and returns results after claims are made, without noticeable delays.
 Accurate: It is equally or more accurate than any human fact-checker.
 Accountable: It self-documents its data sources and analysis, and makes the process of each fact-check transparent.
 This process can then be independently verified, critiqued, improved, and even extended to other situations (Hassan et al., 2015, p. 2)

Ao longo dos últimos anos, múltiplas soluções para automatização do processo de fact-checking foram propostas. Em artigo de 2017, intitulado “Fully

Automated Fact Checking Using External Sources”, Karadzhkov et al. propõem uma solução utilizando redes neurais:

Our framework uses a deep neural network with LSTM text encoding to combine semantic kernels with task-specific embeddings that encode a claim together with pieces of potentially relevant text fragments from the Web, taking the source reliability into account. (Karadzhkov et al, 2017, p.1)

Outra solução desenvolvida, focando em uma parte específica do processo de fact-checking foi introduzida em 2017 e apresentada no artigo “Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster” (Hasan et al., 2017). Nomeado de claim-buster, a solução foca especificamente na identificação de afirmações que precisam ser checadas. O funcionamento deste sistema é explicado da seguinte forma:

[..] we have been building ClaimBuster, an end-to-end system for computer-assisted fact-checking. ClaimBuster uses machine learning, natural language processing, and database query techniques to aid fact-checking. It monitors live discourses (e.g., interviews, speeches and debates), social media, and news to catch factual claims, detects matches with a curated repository of fact-checks from professionals, and delivers the matches instantly to readers and viewers (Hasan et. al, 2017, p. 1)

Em artigo de 2021, intitulado “Automated Fact-Checking for Assisting Human Fact-Checkers”, Nakov et al. identificam e analisam vários sistemas focados em diferentes partes do processo de fact-checking. Listados no artigo, várias soluções já em utilização são listadas: AFCNR, BRENDA, ClaimPortal, Squash e Full Fact’s. Algumas dessas soluções utilizadas por jornalistas no processo de fact-checking e outras focadas no leitor das notícias, para que se facilite o acesso a fatos já checados mas que o usuário possa não estar ciente.

3.2.3 IA generativa

Uma das maiores mudanças recentes na forma como a população consome informação é o advento do uso de IAs generativas pelo público geral. Extremamente controverso, essa mudança de paradigma trouxe benefícios e novos problemas para a forma como as pessoas se informam, popularmente chamadas apenas de IA, modelos como ChatGPT, desenvolvida pela OpenAI e Claude, da Google, competem para conquistar cada vez mais usuários.

Em artigo de 2023, intitulado “Generative AI”, Feuerriegel et al. definem modelos de IA generativa como:

A generative AI model is a type of machine learning architecture that uses AI algorithms to create novel data instances, drawing upon the patterns and relationships observed in the training data. A generative AI model is of critically central yet incomplete nature, as it requires further fine-tuning to specific tasks through systems and applications.(Feuerriegel et al., 2023)

O uso de IA generativa no contexto de notícias vem se tornando cada vez mais prevalente, O Digital News Report 2025, feito pela Reuters institute, constatou que em 2025, 7% dos respondentes utilizaram IA como fonte de notícias na semana pesquisada, mas esse número cresce para 15% quando se considera apenas os participante menores de 25 anos. Ela também vem crescendo no uso para o jornalismo, como constatado nesta mesma pesquisa:

Over the last year more journalists have been using generative AI in newsrooms to support their work via research, transcription, suggested headlines/summaries, and other purposes. We have also started to see more cases where AI is generating stories automatically. The UK's largest regional publisher Reach, for example, employs an AI tool called Gutenbot to assist its journalists in rewriting stories for different websites within its network, while the German tabloid Express.de has used an AI bot called 'Klara Indernach' to author more than 1,500 stories, accounting for 10% of stories read. (Reuters Institute, 2025)

Para fact-checking, há controvérsias sobre a efetividade do uso da IA Generativa. Porém, nos últimos anos, uma frase ficou famosa na rede social Twitter: “Ei, @grok, isso é verdade?”. Usuários do X (ex-Twitter), ainda que muitas vezes de maneira jocosa, passaram a utilizar a funcionalidade de IA Generativa para checar se afirmações são verdadeiras. Em artigo de 2025 publicado pela Deutsche Welle, é reportado:

"Hey, @Grok, is this true?" Ever since Elon Musk's xAI launched its generative artificial intelligence chatbot Grok in November 2023, and especially since it was rolled out to all non-premium users in December 2024, thousands of X (formerly Twitter) users have been asking this question to carry out rapid fact checks on information they see on the platform. (Deutsche Welle, 2025)

No entanto, nem sempre as respostas dadas pelo Grok são satisfatórias e nem sempre fazem sentido, mostrando limitações no uso dessa ferramenta para checar veracidade de afirmações. Muitas vezes, inclusive, resultados obtidos são inesperados e de extremo mau-gosto, como reportado nesse mesmo artigo:

But how accurate and reliable are the chatbots' responses? Many people have asked themselves this question in the face of Grok's recent statements about "white genocide" in South Africa. Apart from Grok's problematic stance on the topic, X users were also irritated

about the fact that the bot started to talk about the issue when it was asked about completely different topics[.] (Deutsche Welle, 2025)

O Grok, no entanto, apesar de ter dado origem à frase que ilustra o uso de IA generativa para checagem de informação, não está entre os chatbots mais utilizados, tendo apenas registrado cerca de 0.8% do market share em julho de 2025 (FirstPageSage, 2025). Em estudo recente, jornalistas da BBC fizeram um experimento para testar a fidelidade das respostas dadas pelos chatbots de IA mais utilizados, com resultados reproduzidos a seguir:

51% of all AI answers to questions about the news were judged to have significant issues of some form
 19% of AI answers which cited BBC content introduced factual errors – incorrect factual statements, numbers and dates
 13% of the quotes sourced from BBC articles were either altered or didn't actually exist in that article.(BBC, 2025)

Muitos defensores dizem que as IA Generativas ainda estão em sua infância e é esperado que, com o tempo, a tecnologia evolua e passe a ser mais confiável. Porém, com os dados que temos hoje e resultados tão negativamente alarmantes, é impossível considerá-las como uma forma eficaz de detectar informações falsas, sendo inclusive responsável por gerar ainda mais informações falsas em muitos casos.

3.3 Exemplos

Como discutido anteriormente, é difícil, em alguns casos, dizer com certeza se algo é verdadeiro ou falso, inclusive com fact-checkers profissionais tendo diferentes metodologias de classificação para tentar capturar todas as nuances necessárias para um tema tão complexo. Muitas vezes, por exemplo, algo pode estar parcialmente correto, pode estar fora de contexto ou apenas ser uma notícia verdadeira porém antiga. Existem muitas maneiras de informar e, conversamente, muitas maneiras de desinformar.

Podemos, no entanto, apresentar temas que são alvos constantes de tentativas de distorção, principalmente na política. A AFP, uma das maiores redes de fact checkers apresenta, em seu site, uma lista dos temas mais comuns em notícias analisados por ela no ano de 2024 e a quantidade de histórias analisadas para cada um desses temas. São eles:

In 2024, 7 536 fact-checks were published, including:

713 on the Israel-Hamas war

524 on the Russian invasion of Ukraine

383 on Covid-19

346 on climate issues

540 on the US elections

513 on American politics

342 on vaccines

221 on artificial intelligence (AFP, 2025)

Como é possível ver, muitos dos temas utilizados para desinformação são temas polêmicos ou de grande repercussão, principalmente quando são altamente divididos entre campos políticos e ideológicos diferentes: direita e esquerda, pró ciência e anti ciência, conservadores e progressistas, etc.

Isso significa que quanto maior poder destrutivo e divisório de um tema, mais frequente é sua utilização na produção de desinformação. As motivações podem ser várias e é difícil determinar quais são elas com precisão. Analisemos, então, alguns desses temas e o impacto causado pela desinformação acerca deles.

3.3.1 Covid-19 e vacinas

Em 2020, ocorreu um dos acontecimentos mais impactantes do século: a pandemia de Covid-19. Até o ano de 2025, estimativas da Organização Mundial de Saúde são de que ocorreram cerca de 7 milhões de mortes diretas pela doença. O impacto global da pandemia, porém, foi muito além desse número, a pandemia alterou a organização da sociedade de diferentes formas.

Notícias e informações falsas sobre a doença, formas de evitar o contágio e, principalmente, sobre as vacinas, que vieram a ser uma das mais importantes ferramentas para restaurar a saúde e a ordem global, foram constantes durante os anos em que a pandemia perdurou, marcando uma das maiores campanhas de desinformação da história.

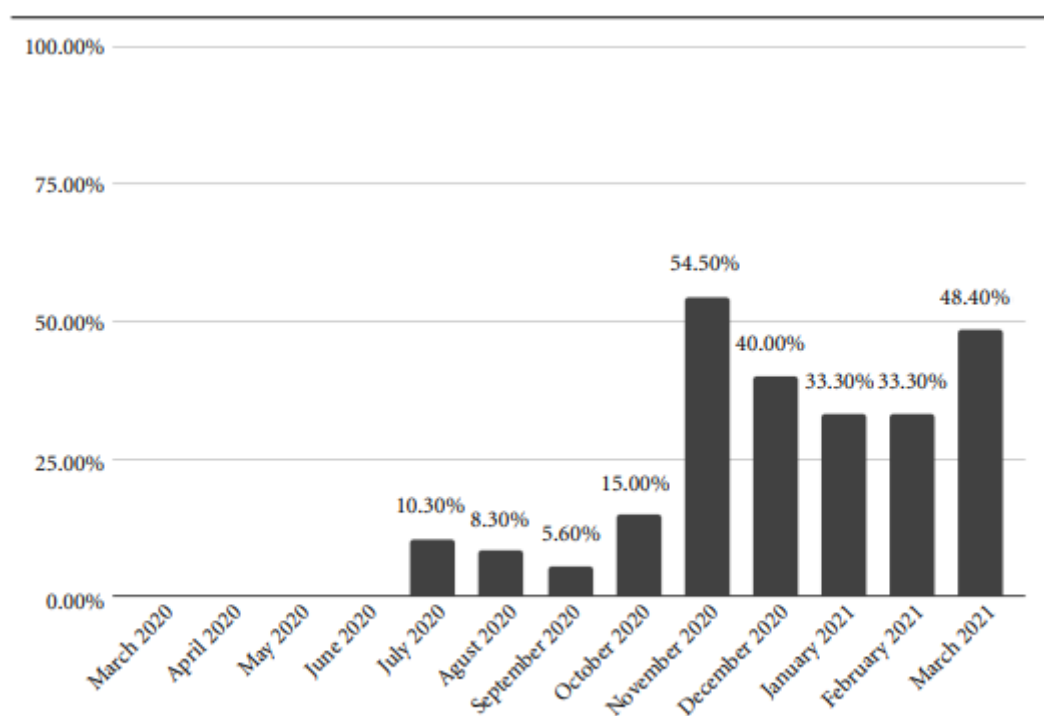
Com os perigos óbvios trazidos pela pandemia, a sociedade teve que lidar com mais essa questão, que dividiu famílias, semeou desconfiança e acabou resultando em, provavelmente, uma grande exacerbação do problema através de recusa por parte de muitas pessoas a tomar medidas preventivas, lentidão das medidas do governo no combate à doença e consequente prolongamento do período.

No artigo intitulado “The disinformation Pandemic-19”, Tagliabue et al. mostram algum dos efeitos diretos da desinformação durante a pandemia:

In some cases, patients refused to take ibuprofen or other anti-inflammatory drugs because of the erroneous idea that they could increase the chances of getting infected with the coronavirus. Misleading information about treatment for COVID-19 has resulted in an increasing number of vitamin D abuse and even mass poisoning from methanol intake. After the lockdown, in countries where social distancing and the use of face masks were mandated, news of correlation between cancer and mask coverings appeared on social networks. (Tagliabue et al., 2020)

Como mencionado no trecho, as redes sociais foram um grande veículo utilizado para desinformação durante a pandemia. No artigo “Fake News and vaccine hesitancy in the COVID-19 pandemic in Brazil”, Galhardi et al. fazem um estudo para mostrar o quanto as redes sociais foram utilizadas durante o período para publicação de informações falsas sobre vacina e concluem que isto contribuiu significativamente para a hesitação por parte da população no uso de vacinas contra covid. Mostram também que o tema vacinas foi um dos principais temas das “fake news” disseminadas em redes sociais:

Gráfico 4 - Proporção do tema vacinas nas “fake news” publicadas em redes sociais



Graph 1. Proportion of fake news disseminated on social networks and messaging apps about vaccines, compared to the total of fake news verified in the Eu Fiscalizo app, from March 26, 2020, to March 31, 2021.

Fonte: Galhardi, Cláudia Pereira, et al.(2022)

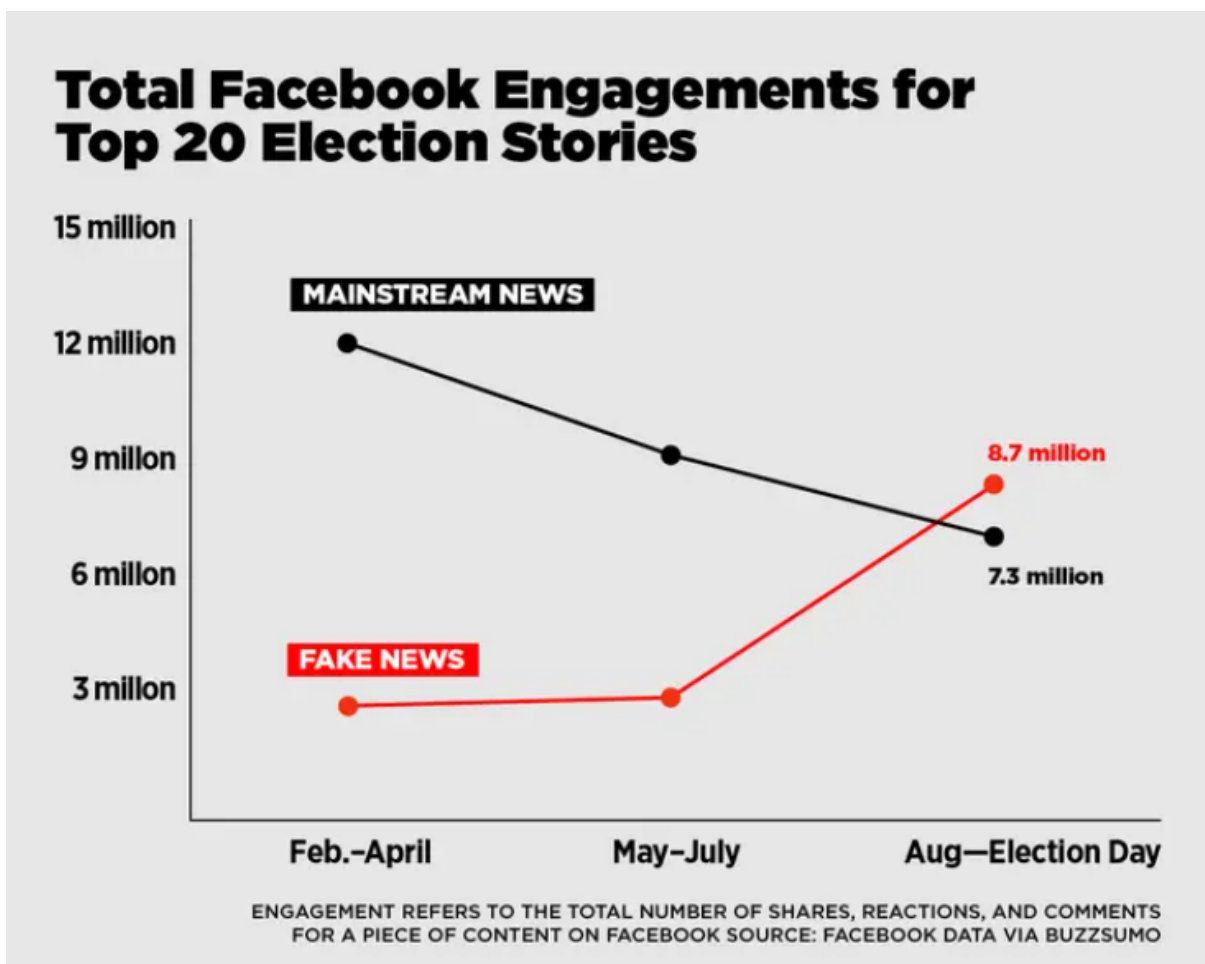
O período da pandemia foi uma das maiores demonstrações do poder destrutivo da desinformação e uma das maiores campanhas de desinformação que se tem notícia, tendo como único “rival” nesse quesito temas relacionados à política.

3.3.2 Eleições

Ainda que a desinformação na internet exista desde muito antes, um grande “boom” na preocupação pública com o tema ocorreu durante as eleições americanas de 2016. É provável que tenha sido uma das primeiras vezes que boa parte da população começou a ter contato com esta discussão, dada a magnitude que o assunto tomou a partir deste marco.

Em artigo de 2016, Craig Silverman/Buzzfeed News, publicou uma análise mostrando que, pela primeira vez registrada, notícias falsas relacionadas às eleições americanas estavam obtendo mais engajamento no Facebook que as publicadas por todos os 19 maiores portais de notícias juntos.

Gráfico 5 - Engajamento no Facebook das 20 notícias mais proeminentes



Fonte: BuzzFeed News (2016)

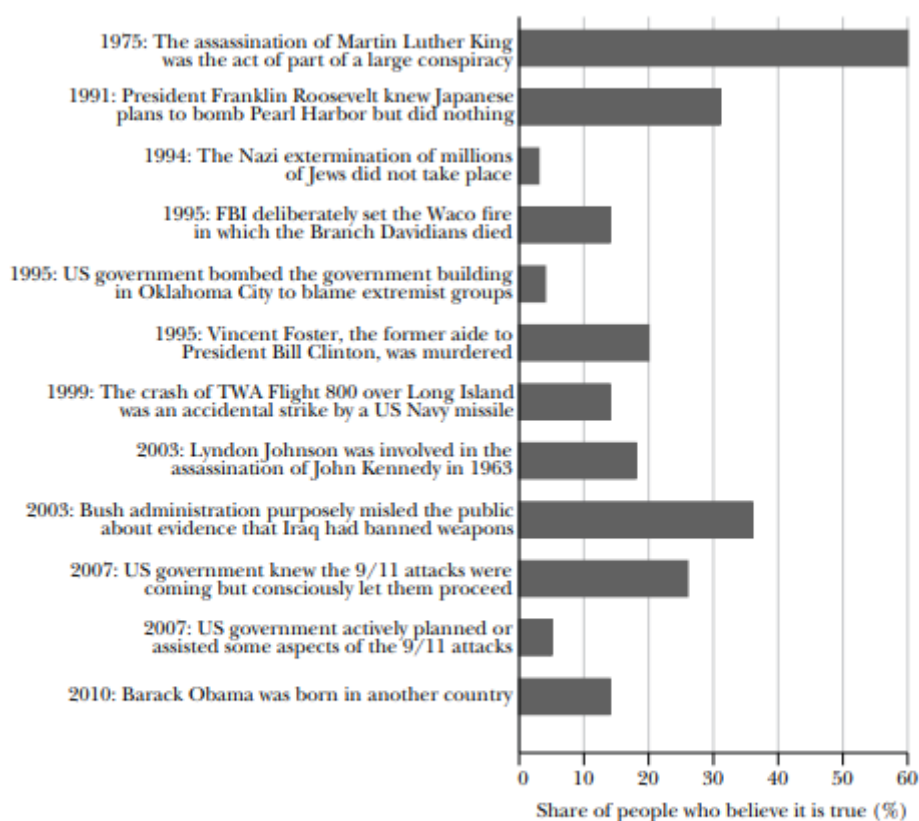
No mesmo artigo, foram elencadas as principais notícias falsas que obtiveram mais engajamento no período. É importante mencionar que todas elas são

afirmações objetivamente falsas, como exemplo, a mais acessada delas, aqui traduzido do inglês: “Papa Francisco Choca o Mundo, Endossa Donald Trump Para Presidente, Divulga Comunicado”. Tal endosso nunca ocorreu.

Teorias da conspiração não eram novidades na política americana e tampouco em eleições. No artigo “Social Media and Fake News in the 2016 Election” (2017), Hunt Allcott e Matthew Gentzkow oferecem uma base teórica e empírica para contextualizar o debate acerca da influência das “fake news” na política americana durante e após as eleições de 2016, e mostram, no seguinte gráfico, algumas das teorias da conspiração com maior proporção de convictos dentre os americanos:

Gráfico 6: Proporção de americanos acreditando em teorias da conspiração históricas

Share of Americans Believing Historical Partisan Conspiracy Theories



Note: From polling data compiled by the American Enterprise Institute (2013), we selected all conspiracy theories with political implications. This figure plots the share of people who report believing the statement listed, using opinion polls from the date listed.

Fonte: Hunt Allcott e Matthew Gentzkow (2017)

Apesar de nem toda teoria da conspiração poder ser facilmente classificada como desinformação, já que nem sempre existem evidências para provar definitivamente se a história é verdadeira ou falsa, é possível perceber que, dentre essas teorias, já existiam histórias feitas, provavelmente, com interesses eleitorais (“Barack Obama nasceu em outro país”), história veiculada em 2010, quando o então presidente americano ainda estava em seu primeiro mandato. À época, um dos defensores dessa teoria era o futuro candidato presidencial (e futuro presidente) Donald Trump (CBS News, 2011). Em 2016, porém, ele admitiu que Barack Obama tinha nascido nos Estados Unidos (CBS News, 2016).

O caso das eleições americanas de 2016 é icônico porque, além de envolver o candidato ganhador, que tinha histórico de espalhar desinformação, foi uma das primeiras vezes que uma campanha de tal tipo conseguiu mudar a opinião pública de forma tão clara e uma das primeiras vezes que as redes sociais foram utilizadas sistematicamente para tal tipo de campanha.

Não é surpresa que, desde então, tal *modus-operandi* tenha inspirado outras campanhas políticas pelo mundo. Em 2018, apenas dois anos após, foi a vez do Brasil passar por algo semelhante. As eleições presidenciais daquele ano foram marcadas por ondas de desinformação veiculadas nas redes sociais. A pesquisadora Mariana Dourado, em sua tese “Fake News na eleição presidencial de 2018 no Brasil”, identificou 346 diferentes histórias falsas compartilhadas nas redes sociais, sendo a maioria delas pró-Bolsonaro, e conclui que o candidato, que veio a ser o ganhador, foi o maior beneficiado pelo compartilhamento dessas histórias falsas.

3.3.3 Palestina e Israel

Apesar de ser um conflito que se arrasta por décadas, em 2023, as tensões entre Israel e Palestina chegaram a um momento crítico, deflagrando uma guerra que já se estende por quase 2 anos. Rapidamente, se tornou o principal conflito geopolítico da atualidade, ainda que outros conflitos de grande porte e visibilidade estejam ocorrendo, como a guerra entre Rússia e Ucrânia.

Como muitos dos outros grandes acontecimentos nos últimos anos, especialmente os mais polarizantes, a guerra na Palestina também veio acompanhada de grandes quantidades de desinformação sendo compartilhadas em redes sociais.

Um dos casos mais emblemáticos aconteceu quando notícias de que o Hamas havia decapitado 40 bebês ganhou holofotes na mídia, afirmação sem nenhuma evidência, mas que se espalhou na mídia tradicional, chegando até a ser citado por Joe Biden, então presidente americano (Le Monde, 2023). O impulsionamento da notícia não verificada teve também grande uso de redes sociais.

Marc Owen Jones, an academic researching disinformation in the Middle East, told MEE that tweets about the “40 murdered babies” story had picked up 44 million impressions by mid-morning on Wednesday. Jones said he did not know if there was an organised campaign behind the spread of the story but said that narratives around murdered babies were “emotive” and had been used in propaganda campaigns since World War One to demonise enemies.(Middle East Eye, 2023)

Por ser mais recente, as características e a forma como tem sido espalhada desinformação acerca de Gaza tem algumas diferenças se comparado àqueles de anos atrás. Em um comunicado em 2024, a empresa OpenAI, responsável pelo desenvolvimento do ChatGPT, um dos chatbots de IA mais usados no mundo, revelou que campanhas de desinformação originadas na China, Rússia, Irã e Israel utilizaram a plataforma para espalhar desinformação. Eles afirmam, no entanto, que nenhuma delas obteve sucesso e não conseguiu alcançar grandes audiências (The Guardian, 2024). O comunicado original da OpenAI não está mais acessível.

Em artigo publicado pela Al Jazeera Media Institute em 2025 intitulado “Weaponized Artificial Intelligence: The Unseen Threat to Fact-Checking”, o jornalista Ahmad Al-Arja alerta para a magnitude da campanha de desinformação acerca do genocídio que vem ocorrendo contra os Palestinos e o uso de IA generativa para campanhas de desinformação, o que torna mais difícil o trabalho de fact-checking feito da forma tradicional:

Over the past nine and a half months, hundreds of fake images, stories, news reports, and pieces of misinformation have flooded social media and gained global traction amid widespread coverage of the genocide initiated by the Israeli occupation on October 7, following Hamas's 'Al-Aqsa Flood' operation targeting settlements in the Gaza envelope.

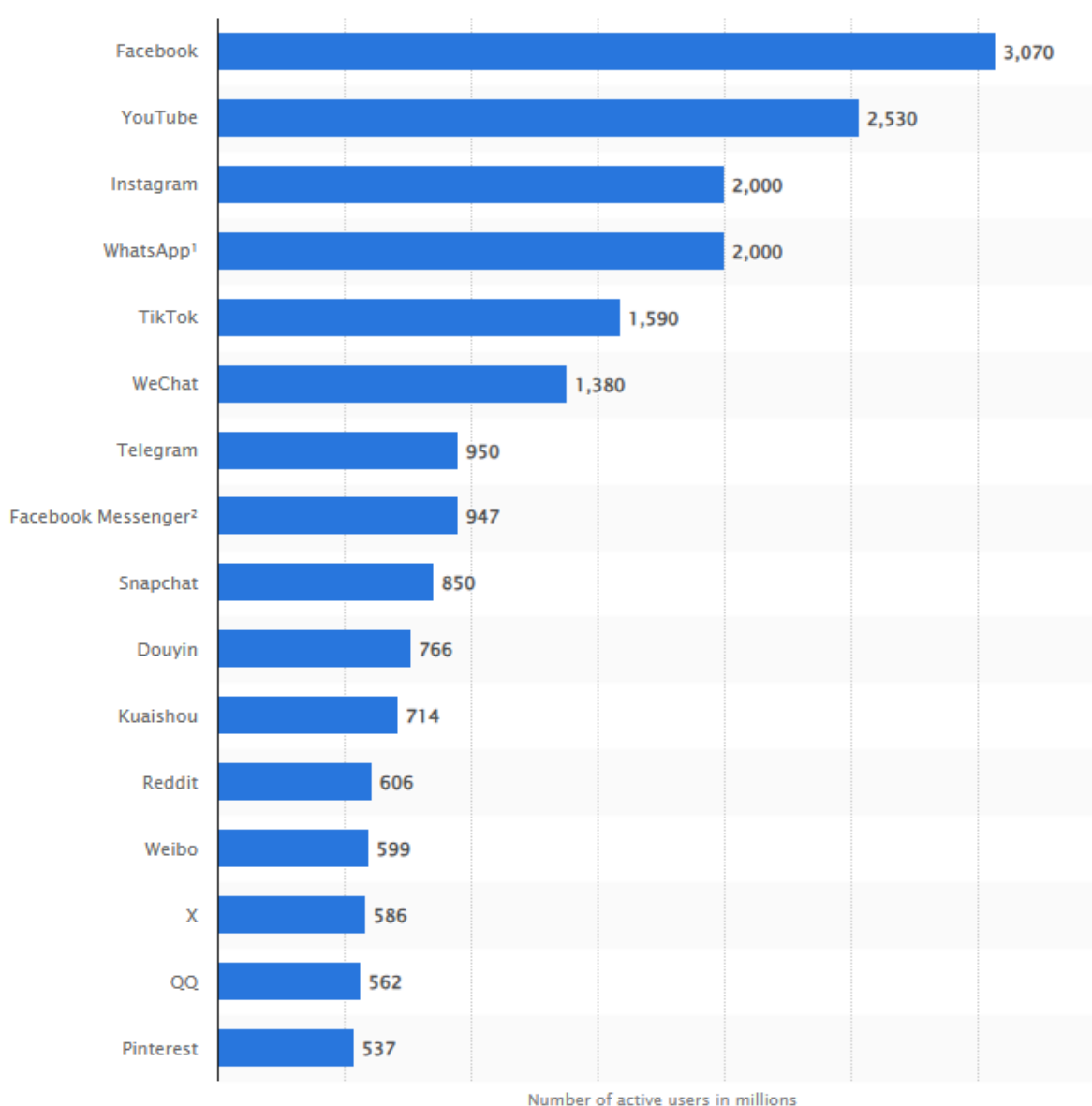
[...]

Fact-checking platforms have monitored hundreds of claims related to the ongoing war, yet many have not relied on traditional verification methods commonly used during previous Israeli assaults on Gaza. The current war of extermination has marked a turning point in the spread of misinformation, with artificial intelligence playing a central role.(Al Jazeera Media Institute, 2025)

4 REDES SOCIAIS

Desde os primórdios da internet, diversas aplicações de comunicação entre usuários nasceram e morreram. No entanto, nenhuma delas se compara, tanto em números de usuários, quanto em influência na sociedade quanto as redes sociais de hoje. Seis das maiores plataformas ultrapassam 1 bilhão de usuários mensais, com algumas outras próximas a esse número.

Gráfico 7: Número de usuários mensais das redes sociais mais populares (em milhões)



Fonte: Statista (2025)

Através dos dados agregados pela Statista e reproduzidos acima, é possível dizer que 5 dessas redes sociais se destacam, com mais de 1.5 bilhões de usuários mensais: Facebook, Instagram e Whatsapp, operados pela Meta, Youtube, operado pela Google e o TikTok, mais jovem dessas redes e operado pela chinesa ByteDance.

Também segundo a Statista, estima-se que no total, 5.42 bilhões de pessoas no mundo utilizam alguma rede social. Com tamanha influência, é de se esperar que as decisões tomadas pelos responsáveis por essas redes tenham consequências globais. O conteúdo permitido ou proibido de ser veiculado por elas tem grande influência na forma como a sociedade se informa e forma opiniões por todo o mundo.

Das maiores redes sociais em números de usuários, a maioria é originária dos Estados Unidos, sendo 3 delas de uma mesma empresa, a Meta. Mais recentemente, algumas redes chinesas como WeChat e TikTok também disputam o mercado global, o que tem causado inclusive tensões diplomáticas, como ficou demonstrado no episódio do banimento do TikTok nos Estados Unidos, citando preocupações com a segurança nacional (Minges, M, American University, 2025).

O que se tem, então, é um grande ecossistema de empresas com grandes responsabilidades e diversos interesses envolvidos. Com a crise de desconfiança global causada pelo aumento da veiculação de desinformação com uso dessas redes, é importante entender quais as políticas adotadas por elas e quais as medidas elas tomam para manter a integridade da forma como a sociedade consome informação.

4.1 Políticas das redes sociais acerca de desinformação

4.1.1 Meta

Responsável pelo Facebook, a maior rede em números de usuários, e também pelo Instagram e Whatsapp, a Meta dá a seguinte explicação para suas políticas a respeito de desinformação:

Misinformation is different from other types of speech addressed in our Community Standards because there is no way to articulate a comprehensive list of what is prohibited. With graphic violence or hateful conduct, for instance, our policies specify the content we prohibit, and even persons who disagree with those policies can follow them. With misinformation, however, we cannot provide such a

line. The world is changing constantly, and what is true one minute may not be true the next minute. (Meta Transparency Center, 2025)

Ela justifica, com razão, que nem sempre é possível determinar exatamente o que é permitido e o que não é quando o tema é desinformação, já que os temas e as formas de falar sobre eles evoluem, assim como a informação disponível para a população. Ela deixa claro, no entanto, que é possível delimitar alguns casos em que certos tipos de desinformação são proibidos e consequentemente são removidos:

We remove the following types of misinformation:

- I. Physical Harm or Violence [...]
- II. Harmful Health Misinformation [...]
 - Misinformation about vaccines. [...]
 - Misinformation about health during public health emergencies. [...]
 - Promoting or advocating for harmful miracle cures for health issues [...]
- III. Voter or Census Interference [...]

For the following content, we include an informative label:

Manipulated Media

Content Digitally Created or Altered that May Mislead. [...] (Meta, 2025)

Em geral, as políticas sobre conteúdo falso da Meta parecem estar alinhadas com os acontecimentos mais problemáticos dos últimos anos no mundo da desinformação, incluindo a preocupação com o atual e futuro uso do uso de IA generativa para criar imagens falsas que podem iludir usuários desavisados. A Meta, no entanto, mudou sua política de fact-checking e moderação em Janeiro de 2025, após a segunda vitória de Donald Trump para a presidência americana. Um artigo da MSU Today explica as mudanças impostas por Zuckerberg, CEO da Meta:

Meta's previous content moderation policy relied on third-party fact-checking organizations, which brought problematic content to the attention of Meta staff. Meta's U.S. fact-checking organizations were AFP USA, Check Your Fact, Factcheck.org, Lead Stories, PolitiFact, Science Feedback, Reuters Fact Check, Televisa, Univision, The Dispatch and USA Today.

Meta CEO Mark Zuckerberg highlighted that content moderation at Meta would shift to a community labeling model similar to X, formerly Twitter. X's Community Notes is a crowdsourced fact-checking approach that allows users to write notes to inform others about potentially misleading posts. (MSU Today, 2025)

Segundo a análise presente neste mesmo artigo, as críticas acerca dessa mudança tem fundamento, já que estudos não chegaram a um consenso se iniciativas de moderação feita pelos próprios usuários, como também adotada no X, antigo twitter, tem efetividade no combate a desinformação e a velocidade necessária para conter o problema na fase inicial e mais viral de divulgação.

4.1.2 Youtube

O Youtube, segunda maior rede por números de usuários e que pertence à Google, tem as seguintes políticas acerca de desinformação:

Misinformation policies

Certain types of misleading or deceptive content with serious risk of egregious harm are not allowed on YouTube. This includes certain types of misinformation that can cause real-world harm, certain types of technically manipulated content, or content interfering with democratic processes.

[..]

Don't post content on YouTube if it fits any of the descriptions below.

Suppression of census participation: Content aiming to mislead census participants about the time, place, means, or eligibility requirements of the census, or false claims that could materially discourage census participation.

Manipulated content: Content that has been technically manipulated or doctored in a way that misleads users (usually beyond clips taken out of context) and may pose a serious risk of egregious harm.

Misattributed content: Content that may pose a serious risk of egregious harm by falsely claiming that old footage from a past event is from a current event. (Google, 2025)

[...]

Elections misinformation policies

These policies prohibit certain types of content relating to free and fair democratic elections. Don't post elections-related content on YouTube if it fits any of the descriptions noted below.

Voter suppression: Content aiming to mislead voters about the time, place, means, or eligibility requirements for voting, or false claims that could materially discourage voting.

Candidate eligibility: Content that advances false claims related to the technical eligibility requirements for current political candidates and sitting elected government officials to serve in office. Eligibility requirements considered are based on applicable national law, and include age, citizenship, or vital status.

Incitement to interfere with democratic processes: Content encouraging others to interfere with democratic processes. This includes obstructing or interrupting voting procedures.

Election integrity: Content advancing false claims that widespread fraud, errors, or glitches occurred in certain past elections to determine heads of government. Or, content that claims that the certified results of those elections were false. This policy currently applies to:

The 2014, 2018, and 2022 Brazilian Presidential elections

[..]

Medical misinformation policy

Don't post content on YouTube if it includes any of the following:

Prevention misinformation: We do not allow content that promotes information that contradicts health authority guidance on the prevention or transmission of specific health conditions, or on the safety, efficacy or ingredients of currently approved and administered vaccines.

Treatment misinformation: We do not allow content that promotes information that contradicts health authority guidance on treatments for specific health conditions, including promotion of specific harmful substances or practices that have not been approved by local health authorities as safe or effective, or that have been confirmed to cause severe harm. (Youtube, 2025)

Similar às políticas da Meta, as do Youtube também parecem ser feitas pensadas nos tipos de desinformação mais comuns divulgadas através da plataforma no passado e já apresentadas neste trabalho. Ela, inclusive, chega a especificar as eleições brasileiras (provavelmente porque a página foi acessada do Brasil) e os tipos mais difundidos de afirmações falsas feitas acerca delas no

passado, como por exemplo alegações de que as urnas eletrônicas adulteraram o resultado dessas eleições (CNN Brasil, 2023). Há também um aviso, não incluído no trecho destacado que, em 2023, as políticas acerca de informações sobre Covid 19 mudaram e passaram a ser incluídas nas políticas gerais sobre doenças infecciosas.

4.1.3 *TikTok*

Dentre as maiores redes sociais a mais nova e, diferentemente das duas citadas anteriormente, originada na China, o TikTok tem políticas relacionadas a desinformação parecidas com os competidores:

Misinformation

In a global community, it is natural for people to have different opinions, but we seek to operate on a shared set of facts and reality. We do not allow misinformation that may cause significant harm to individuals or society, regardless of intent. We rely on independent fact-checking partners, guidance from public health authorities, and our database of previously fact-checked claims to help assess the accuracy of content.

Content is ineligible for the FYF if it contains misinformation that may cause moderate harm, such as certain health content, conspiracy theories, repurposed media, or misrepresented authoritative sources. To be cautious, unverified information about emergencies and content temporarily under review by fact-checkers is also ineligible for the FYF.

To help you manage your TikTok experience, we may apply warning labels to content that has been assessed by our fact-checking partners and cannot be verified as accurate. We may also send prompts to reconsider sharing such content.

Civic and Election Integrity

Elections are important events and are often the subject of intense discussion and analysis. We try to balance enabling these discussions, while also being a place that brings people together and does not cause division. We do not allow paid political promotion, political advertising, or fundraising by politicians and political parties (for themselves or others). Our political advertising policy includes both traditional paid advertisements and creators receiving compensation to support or oppose a candidate for office.

We want to enable the informed exchange of civic ideas in a way that fosters productive dialogue. We do not allow misinformation or content about civic and electoral processes that may result in voter interference, disrupt the peaceful transfer of power, or lead to off-platform violence.

Content may be ineligible for the FYF if it contains misinformation that can hinder the ability of a voter to make an informed decision. To be cautious, unverified claims about an election and content temporarily under review by fact-checkers may also be ineligible for the FYF.

To help you manage your TikTok experience, we may apply warning labels to content that has been assessed by our fact-checking partners and cannot be verified as accurate. Learn more about our election integrity work, and Government, Politician and Political Party accounts.

Edited Media and AI-Generated Content (AIGC)

We welcome the creativity that new artificial intelligence (AI) and other digital technologies may unlock. However, AI and other digital editing technologies can make it difficult to tell the difference between fact and fiction, which may mislead individuals or harm society. We require you to label AIGC or edited media that shows realistic-appearing scenes or people. This can be done using the AIGC label, or by adding a clear caption, watermark, or sticker of your own. Even when appropriately labeled, AIGC or edited media may still be harmful. We do not allow content that shares or shows fake authoritative sources or crisis events, or falsely shows public figures in certain contexts. This includes being bullied, making an endorsement, or being endorsed.

We are committed to protecting people's privacy. We do not allow content that contains the likeness of young people, or the likeness of adult private figures used without their permission.

(TikTok, 2024)

É possível perceber que, em alguns casos, o TikTok tem uma política mais branda, já que o conteúdo não necessariamente é removido, apenas não é apto a aparecer na lista de conteúdos sugeridos para outros usuários. Porém, é a única das três políticas aqui apresentadas que menciona sua própria rede parceira de fact-checkers e menciona nominalmente conteúdo gerado por IA.

4.2 Ações específicas

Para combater a desinformação nas plataformas, ter políticas claras e bem pensadas para cada situação é apenas o começo. É necessário, também, que sejam tomadas ações efetivas contra a divulgação desses conteúdos. É difícil obter dados, já que as plataformas costumam divulgar apenas o que lhes convêm e muitas vezes não são transparentes o suficiente. Porém, podemos analisar algumas das medidas tomadas por elas ao longo dos anos.

Uma das iniciativas comuns tomadas pelas plataformas é a prática de identificar conteúdos potencialmente enganosos e adicionar contexto para que o usuário se informe acerca do assunto que ele trata.

Em 2018, o Youtube anunciou que iria começar a adicionar links da Wikipedia em conteúdos relacionados a teorias da conspiração, conforme publicação da CNBC:

Google-owned YouTube is trying to combat the amount of misinformation spread on its site, announcing that it will link videos that promote conspiracy theories to “fact-based” sites like Wikipedia pages.

YouTube CEO Susan Wojcicki made the announcement Tuesday at the South by Southwest Festival in Austin, Texas.

[...]

In coming months, Wojcicki said, conspiracy videos will start including text boxes, dubbed “information cues,” that link to third-party sources that debunk the hoaxes in question.

Wojcicki said the new feature will only be used on conspiracies causing “significant debate” on YouTube, like those about chemtrails or the moon landing. After the school shooting massacre in Parkland, Florida, last month, a video that theorized that one of the survivors was a crisis actor made it into YouTube’s trending section. Because that conspiracy theory took off in a matter of days, it isn’t clear whether a Wikipedia page disputing that theory would even be available yet. (CNBC, 2018)

A medida foi recebida com ceticismo, mas foi uma das primeiras do gênero e pode ser considerada um avanço, ainda que pequeno, no combate à desinformação nas redes sociais. No entanto, a maior crítica feita à época e ainda relevante até hoje, é que o problema central estava no algoritmo de recomendação do Youtube, que sugeria a usuários conteúdos conspiratórios (MIT Technology Review, 2018).

Uma medida similar, focada em adicionar contexto junto a conteúdo potencialmente falso, porém gerado a partir da contribuição dos próprios usuários,

foi tomada pelo Twitter (hoje chamado de X) e intitulada Birdwatch (hoje chamado de Community Notes). Inicialmente disponível para apenas alguns usuários, a medida foi expandida em 2022. Conforme publicado pela Reuters naquele ano:

Twitter Inc (TWTR.N), will expand its community fact-checking project called Birdwatch, the social media company said on Wednesday, deepening its novel approach to a new form of content moderation.

Birdwatch launched last year, allowing some Twitter users to debunk misleading tweets by attaching notes to the content in order to provide context or point to accurate sources.

[...]

Until now, Birdwatch was a limited experiment with 15,000 contributors writing fact-checking notes. Twitter said it will now add about 1,000 new contributors per week.

[...]

The project has produced encouraging results, Twitter said. People are 15% to 35% less likely to "like" or retweet content that has a Birdwatch note attached to it. They are also 20% to 40% less likely to agree with a potentially misleading tweet after reading a Birdwatch note about it. (Reuters, 2022)

A medida, apesar de apresentada inicialmente acompanhada de bons números, como pode ser constatado no trecho destacado, foi alvo de controvérsias ao longo dos anos após sua expansão, e sua efetividade foi colocada em cheque (Poynter, 2023). Mais recentemente, estudos estatísticos foram feitos e os resultados não foram animadores: "As our main contribution, we demonstrate that the roll-out of Community Notes had no statistically significant effect on reducing engagement with misinformation on X/Twitter in terms of retweet count and like count" (Chuai et al., 2024, p. 26).

Em 2025, em medida criticada por especialistas, a Meta resolveu acabar com sua política de ter seus próprios fact checkers dedicados e passar a utilizar um sistema parecido com o Community Notes, com checagem feita pelos próprios usuários:

Shortly after the January 6th Capitol riots in 2021, which were fuelled by false claims of a rigged US presidential election, Mr Zuckerberg gave testimony to Congress. The billionaire boasted about Meta's "industry-leading fact checking program".

It drew, he pointed out, on 80 "independent third-party fact checkers" to curb misinformation on Facebook and Instagram.

Four years on, that system is no longer something to brag about.

"Fact checkers have just been too politically biased and have destroyed more trust than they've created, especially in the US," Mr Zuckerberg said earlier in January.

Taking their place, he said, would be something totally different: a system inspired by X's "community notes", where users rather than experts adjudicate on accuracy.

Many experts and fact checkers questioned Mr Zuckerberg's motives.

"Mark Zuckerberg was clearly pandering to the incoming administration and to Elon Musk," Alexios Mantzarlis, the director of the Security, Trust and Safety Initiative at Cornell Tech, told the BBC. (BBC, 2025)

Ainda é cedo para determinar se o novo sistema, também intitulado Community Notes pela Meta, resultou em alguma melhora real na qualidade do

combate à desinformação em seus produtos. Porém, as críticas de especialistas têm fundamento, já que, ainda que haja possíveis argumentos a favor da solução baseada em contribuição do usuário, ela não necessariamente precisava ter vindo somada ao fim do uso de fact checkers profissionais independentes. A decisão parece ter sido, de fato, para agradar a nova administração americana.

Apesar de todas as maiores redes sociais alegarem que removem conteúdo que desrespeita suas políticas internas, incluindo as relacionadas à desinformação, é difícil determinar o quanto elas realmente cumprem o que se propõem.

As plataformas têm um incentivo financeiro para manter conteúdo que gera engajamento no ar e a quantidade de publicações que se mantém nas plataformas, ainda que aparentemente desrespeitem as políticas delineadas para o combate à desinformação não é um bom sinal.

Em carta aberta publicada em 2022 e endereçada à CEO do Youtube, a organização “The International Fact Checking Network”, composta por fact checkers de todo o mundo, deixa claro o problema da aparente falta de cumprimento das próprias políticas do Youtube no que diz respeito à remoção de conteúdo falso:

The examples are too many to count. Many of those videos and channels remain online today, and they all went under the radar of YouTube’s policies, especially in non-English speaking countries and the Global South. We are glad that the company has made some moves to try to address this problem lately, but based on what we see daily on the platform, we think these efforts are not working — nor has YouTube produced any quality data to prove their effectiveness.

Your company platform has so far framed discussions about disinformation as a false dichotomy of deleting or not deleting content. By doing this, YouTube is avoiding the possibility of doing what has been proven to work: our experience as fact-checkers together with academic evidence tells us that surfacing fact-checked information is more effective than deleting content. (Poynter, 2022)

Sob a perspectiva desse grupo, a falsa dicotomia entre remover ou não conteúdo é menos importante do que o que outras possíveis ações que as plataformas poderiam tomar. Eles alegam, porém, que o Youtube não tem feito o suficiente para conter o problema.

Outro caso que mostra que basear o combate à desinformação apenas na remoção de conteúdo não é o suficiente é o do TikTok. A própria plataforma divulgou, em 2024, em cumprimento à Lei Europeia números acerca dos seus esforços para combater desinformação, em particular acerca das Eleições Europeias de 2024:

With roughly 3,300 data points, including at country level, the report highlights how we’re continuing to strengthen our efforts to combat mis- and disinformation and platform

manipulation across Europe, and the progress we're making in this space: Notably, in the first half of 2024, 97% of videos violating our misinformation policies were removed proactively, before any user report.

[...]

In the four weeks leading up to and including the elections, our efforts resulted in the removal of 43,000 pieces of content in violation of our misinformation policies and 2,600 pieces of civic- and election-integrity related content, with 96% of this content removed before it was reported and over 80% before receiving a single view. (TikTok, 2024)

A publicação menciona também outras medidas, como o trabalho conjunto com fact checkers e a inclusão de informações acerca das eleições dentro do próprio app, com o objetivo de disponibilizar informações confiáveis aos usuários.

É possível, porém, que essas medidas tenham sido uma resposta a críticas recebidas acerca da ineficácia da plataforma em combater divulgação de desinformação através dela. Apenas alguns meses antes, reportagem da BBC mostrou que usuários do TikTok consumiram conteúdo enganoso a respeito das eleições gerais do Reino Unido:

But a BBC project to investigate the content promoted by social media algorithms has found - alongside funny montages - young people on TikTok are being exposed to misleading and divisive content. It is being shared by everyone from students and political activists to comedians and anonymous bot-like accounts.

Videos which have racked up hundreds of thousands of views have promoted unfounded rumours that a major scandal prompted Rishi Sunak to call an early election and the baseless claim that Sir Keir Starmer was responsible for the failure to prosecute serial paedophile Jimmy Savile.

Satirical, fake AI-generated clips show Rishi Sunak declaring, "Please don't vote us out, we would be proper gutted!" and making unevidenced claims about how the Conservative leader is spending public money - including how he will send his "mates loads of dosh". (BBC, 2024)

4.3 Análise dos achados

É possível notar que a tomada de medidas no combate à desinformação por parte delas parece estar sempre atrasada, com ações quase sempre focadas em problemas ocorridos no passado, o que dá impressão que estão fazendo apenas o necessário para manter sua boa imagem ao público, sem que necessariamente haja uma preocupação real com o futuro do combate à desinformação.

Números publicados pelas próprias redes sociais dão a entender que iniciativas baseadas em auto policiamento pelos usuários, como o community notes, do X/Twitter e mais recentemente adotado pela Meta, são efetivos na diminuição da divulgação de desinformação, um estudo subsequente não encontrou efeito estatisticamente significativo na redução do alcance de desinformação. (Reuters, 2022; Chuai et al., 2024), o que sugere possível cherry-picking nos números oficiais divulgados pelas empresas.

Pode-se afirmar também que as métricas disponibilizadas pelas empresas responsáveis pelas redes são pouco específicas, sem informações estatísticas importantes como taxas de falso positivo para conteúdo removido e taxas de falso negativo (ou seja, conteúdo que deveria ter sido removido de acordo com suas políticas mas foi mantido no ar), fato criticado pelo jornalismo especializado no assunto (Poynter, 2024) e que dificulta a checagem independente das afirmações feitas pelas empresas acerca da efetividade de suas políticas e ações.

Apesar das plataformas utilizarem de escudo a subjetividade intrínseca ao problema da desinformação para justificar eventual inércia no combate à desinformação, estudos mostram que, ao harmonizar sistemas de classificação, as maiores agências de fact-checking raramente apresentam contradições claras, o que sugere maior convergência metodológica do que se supõe (Lee et al., 2023), e implicando que na maioria dos casos é possível dizer de forma clara que um conteúdo é falso e portanto poderia ser removido.

Ainda que seja possível notar exemplos de adequações regionais nas políticas das empresas (como o Youtube incluindo especificamente as eleições brasileiras), preocupações expressadas pelas maiores agências de fact-checking pelo mundo deixam claro que ainda há lacunas no monitoramento de desinformação em idiomas de mercados menores, com o problema sendo agravado no sul global (Poynter, 2022)(Poynter, 2024).

Um dos elementos mais recentes no ecossistema de redes sociais é o uso massivo de IA Generativa. Ainda que políticas de redes como a do TikTok mostrem preocupação com essas tecnologias, incluindo indicação obrigatória por parte do usuário ao utilizar conteúdo gerado por IA, é difícil garantir que tais políticas sejam sempre respeitadas, principalmente frente aos rápidos avanços na área. Isto também contribui para que usuários menos familiarizados com essas ferramentas tenham dificuldade em perceber a natureza de tais conteúdos.

Paralelo a isso, o uso dessas ferramentas na checagem de informação apresenta problemas tanto tecnológicos quanto éticos, dada a eventual tendência a “alucinações” apresentadas por esses sistemas (New York Times, 2025).

5 CONSIDERAÇÕES FINAIS

Este trabalho investigou o fenômeno da desinformação na Internet, um assunto que ganhou espaço no debate público nos últimos anos. Muitos pesquisadores, desenvolvedores de redes sociais e jornalistas tentam conter o problema. O trabalho analisou os meios pelos quais essa desinformação se espalha, as diversas estratégias propostas para contê-la e as implicações disso para a sociedade, o jornalismo e a engenharia de software.

Não há dúvidas que, por se tratar de um tema complexo, a crise de desinformação mundial está longe de ter solução. Num tempo em que é possível perceber que o problema ainda cresce, muitos dos principais atores parecem desacelerar a busca por melhores soluções. Na contramão disso, a população usa a internet para se comunicar, se relacionar e se informar cada vez mais, se tornando mais vulnerável e mais exposta a desinformação.

Ademais, o surgimento de novas tecnologias baseadas em IA se prova tanto um possível aliado quanto mais uma ferramenta a ser usada de forma mal intencionada, tornando ainda mais complexo o ecossistema de informação e desinformação por todo o mundo.

Muitas das principais inovações nesse combate tem vindo de jornalistas e pesquisadores da área, mas as empresas responsáveis pelas maiores redes se mostram inertes e, em alguns casos, retrocedem no quesito combate à desinformação, como fica bem demonstrado no caso da Meta.

Interesses políticos e a busca pelo constante aumento do lucro podem ser os empecilhos por parte das empresas, mas há caminhos para mitigação do problema sem depender apenas delas. Os usuários das redes também podem ser aliados nesse combate, principalmente através do papel do jornalismo de informar acerca dos perigos e vulnerabilidades que pessoas mal intencionadas buscam explorar.

A inovação no desenvolvimento de novos softwares voltados ao público para tornar mais factível o uso de redes sociais com chance menor de ser enganado é outro caminho possível. Pesquisas futuras na área podem elucidar ainda mais os mecanismos responsáveis pela difusão de desinformação nas redes, abrindo caminhos para tais soluções inovadoras.

5.1 Limitações

Apesar de ser possível tirar algumas conclusões a partir deste estudo, é importante primeiro salientar algumas limitações:

- Não foi feita análise empírica, como levantamento de dados próprio, utilizando-se apenas o que foi feito nos últimos anos por outros autores
- A pesquisa foi feita apenas em inglês e português, não atingindo publicações feitas em outras línguas
- Muitos dos dados acerca de ações feitas pelas redes sociais foram publicados pelas próprias empresas, ainda que haja cuidado para não tomá-las como verdade absoluta, é impossível afirmar quais desses dados são precisos ou estatisticamente relevantes
- O escopo limitado deste trabalho significa que nem todas as informações potencialmente relevantes foram exploradas a fundo, e poderiam, por vezes, mudar certas conclusões
- A natureza do problema analisado é de constante mudança nos últimos anos, o que pode agravar eventuais problemas de viés de recência, assim como entendimentos e desenvolvimentos futuros poderiam mudar o rumo da pesquisa
- O uso de fontes jornalísticas pode apresentar viés de seleção, ainda que tenha sido utilizado publicações mais conhecidas e de diferentes linhas editoriais quando possível, não há como garantir imparcialidade

5.2 Trabalhos Futuros

Dados os achados deste trabalho, alguns caminhos para futuras pesquisas podem ser sugeridos:

- Impactos do uso massivo de IA generativa no ecossistema de desinformação nas redes sociais
- Importância da educação dos usuários para diminuir a divulgação de desinformação através de redes sociais;
- Análises empíricas das atuais políticas das redes sociais no impacto da divulgação de desinformação;
- Geração de desinformação causada por alucinações de IAs generativas;
- Novas maneiras de combate à desinformação sem depender de ações específicas dos responsáveis pelas redes sociais.

REFERÊNCIAS

Don Gotterbarn, Keith Miller, and Simon Rogerson. 1997. **Software engineering code of ethics**. Commun. ACM 40, 11 (Nov. 1997), 110–118.
<https://doi.org/10.1145/265684.265699>

ACM. **ACM Code of Ethics and Professional Conduct**. Disponível em:
<https://www.acm.org/code-of-ethics>>. Acesso em: 02/07/2025.

Reuters Institute. **Overview and key findings of the 2025 Digital News Report**. Disponível em:
<https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary>>. Acesso em: 02/07/2025

BBC. **The rise and rise of fake news**. Disponível em:
<https://www.bbc.com/news/blogs-trending-37846860>>. Acesso em: 02/07/2025

Tagliabue, F., Galassi, L., & Mariani, P. (2020). **The "Pandemic" of Disinformation in COVID-19**. *SN comprehensive clinical medicine*, 2(9), 1287–1289.
<https://doi.org/10.1007/s42399-020-00439-1>

Dowden, B., Swartz, N., [s.d.]. **Truth**. Disponível em:
<https://iep.utm.edu/truth/>>. Acesso em: 03/07/2025

American Psychological Association, [s.d.]. **Misinformation and Disinformation**. Disponível em:
<https://www.apa.org/topics/journalism-facts/misinformation-disinformation>>. Acesso em: 03/07/2025

AFP. **What is fact-checking?** Disponível em:
<https://www.afp.com/en/fact-checking/fact-checking-afp/what-fact-checking>>. Acesso em: 05/07/2025

Erica Ryan, Duke Reporters' Lab, 2025.. **2025 census: Fact-checkers persevere as politicians, platforms turn up heat**. Disponível em:
<https://reporterslab.org/2025/06/19/fact-checkers-persevere-as-politicians-platforms-turn-up-heat/>>. Acesso em 05/07/2025

Lee, S., Xiong, A., Seo, H., & Lee, D. (2023). **"Fact-checking" fact checkers: A data-driven approach**. Harvard Kennedy School (HKS) Misinformation Review.
<https://doi.org/10.37016/mr-2020-126>

Hassan, Naeemul, et al. **"The quest to automate fact-checking."** Proceedings of the 2015 computation+ journalism symposium. New York: The Computation+ Journalism Symposium, 2015.

Karadzhov, Georgi, et al. **"Fully automated fact checking using external sources."** arXiv preprint arXiv:1710.00341 (2017).

Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. **Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster**. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17). Association for Computing Machinery, New York, NY, USA, 1803–1812.
<https://doi.org/10.1145/3097983.3098131>

Nakov, Preslav, et al. "Automated fact-checking for assisting human fact-checkers." arXiv preprint arXiv:2103.07769 (2021).

Feuerriegel, S., Hartmann, J., Janiesch, C. et al. **Generative AI**. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

DW. **Fact check: How trustworthy are AI fact checks?** Disponível em: <<https://www.dw.com/en/fact-check-hey-grok-is-this-true-how-trustworthy-are-ai-fact-checks/a-72539345>>. Acesso em: 07/07/2025

BBC. **Groundbreaking BBC research shows issues with over half the answers from Artificial Intelligence (AI) assistants**. Disponível em: <<https://www.bbc.com/mediacentre/2025/bbc-research-shows-issues-with-answers-from-artificial-intelligence-assistants>>. Acesso em: 07/07/2025

WHO. **WHO COVID-19 dashboard**. Disponível em: <<https://data.who.int/dashboards/covid19/deaths>>. Acesso em: 08/07/2025

Skaflé I, Nordahl-Hansen A, Quintana D, Wynn R, Gabarron E
Misinformation About COVID-19 Vaccines on Social Media: Rapid Review
 J Med Internet Res 2022;24(8):e37367
 URL: <https://www.jmir.org/2022/8/e37367>
 DOI: 10.2196/37367

Galhardi, Cláudia Pereira, et al. "Fake News and vaccine hesitancy in the COVID-19 pandemic in Brazil." *Ciência & Saúde Coletiva* 27 (2022): 1849-1858.

FirstPageSage. **Top Generative AI Chatbots by Market Share – July 2025**. Disponível em: <<https://firstpagesage.com/reports/top-generative-ai-chatbots/>>. Acesso em: 08/07/2025

Buzzfeed News. **This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook**. Disponível em: <<https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>>. Acesso em: 09/07/2025

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.
 DOI: 10.1257/jep.31.2.211

CBS News. **Donald Trump keeps pushing "birther" claims** (2011). Disponível em: <<https://www.cbsnews.com/news/donald-trump-keeps-pushing-birther-claims/>>. Acesso em: 10/07/2025

CBS News. **Donald Trump: "Obama was born in the United States. Period."** (2016). Disponível em: <<https://www.cbsnews.com/news/donald-trump-birther-obama-announcement/>>. Acesso em: 10/07/2025

Dourado, Tatiana Maria Silva Galvão. **Fake News na eleição presidencial de 2018 no Brasil** (2020). Disponível em: <<https://repositorio.ufba.br/handle/ri/31967>>

The Guardian. **OpenAI says Russian and Israeli groups used its tools to spread disinformation.** Disponível em: <<https://www.theguardian.com/technology/article/2024/may/30/openai-disinformation-russia-israel-china-iran>> Acesso em: 10/07/2025

Le Monde. **'40 beheaded babies': Deconstructing the rumor at the heart of the information battle between Israel and Hamas.** Disponível em: <https://www.lemonde.fr/en/les-decodeurs/article/2024/04/03/40-beheaded-babies-the-itinerary-of-a-rumor-at-the-heart-of-the-information-battle-between-israel-and-hamas_6667274_8.html> Acesso em: 11/07/2025

Middle East Eye. **Israel-Palestine war: How unverified reports of Hamas 'beheading babies' filled front pages.** (2023). Disponível em: <<https://www.middleeasteye.net/news/israel-palestine-war-beheading-babies-how-unverified-reports-hamas-killings-filled-front-pages>> Acesso em: 11/07/2025

Al Jazeera Media Institute. **Weaponized Artificial Intelligence: The Unseen Threat to Fact-Checking.** (2025). Disponível em: <<https://institute.aljazeera.net/en/ajr/article/2861>> Acesso em: 11/07/2025

Statista. **Most popular social networks worldwide as of February 2025, by number of monthly active users** (2025). Disponível em: <<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>>. Acesso em: 12/07/2025

American University. **National Security and the TikTok Ban** (2025). Disponível em: <<https://www.american.edu/sis/news/20250123-national-security-and-the-tik-tok-ban.cfm>>. Acesso em: 12/07/2025

Meta Transparency Center. **Misinformation.** Disponível em: <<https://transparency.meta.com/policies/community-standards/misinformation>>. Acesso em: 12/07/2025

MSU Today. **Ask the expert: What Meta's new fact-checking policies mean for misinformation and hate speech.** (2025). Disponível em:

<<https://msutoday.msu.edu/news/2025/ask-the-expert-what-meta-new-fact-checking-policies-mean-for-misinformation-and-hate-speech>>

Acesso em: 12/07/2025

Google Support. **Misinformation policies**. Disponível em:

<https://support.google.com/youtube/topic/10833358?hl=en&ref_topic=2803176&sjid=9043948871589383540-SA>. Acesso em: 12/07/2025

CNN Brasil. **Relembre vezes em que Jair Bolsonaro questionou o sistema eleitoral**.

(2023). Disponível em:

<<https://www.cnnbrasil.com.br/politica/relembre-vezes-em-que-jair-bolsonaro-questionou-o-sistema-eleitoral/>>. Acesso em: 12/07/2025

TikTok. **Integrity and Authenticity**. Disponível em:

<<https://www.tiktok.com/community-guidelines/en/integrity-authenticity>>. Acesso em: 13/07/2025

CNBC. **YouTube will add Wikipedia links debunking conspiracy theory videos** (2018).

Disponível em:

<<https://www.cnbc.com/2018/03/13/youtube-wikipedia-links-debunk-conspiracy.html>>. Acesso em: 13/07/2025

MIT Technology Review. **Linking to Wikipedia articles from conspiracy videos won't solve YouTube's core issue**. (2018). Disponível em:

<<https://www.technologyreview.com/2018/03/14/67150/linking-to-wikipedia-articles-on-conspiracy-videos-wont-solve-youtubes-core/>>. Acesso em: 13/07/2025

Reuters. **Twitter to expand its community fact-checking experiment, Birdwatch**.

(2022). Disponível em:

<<https://www.reuters.com/technology/twitter-expand-its-community-fact-checking-experiment-birdwatch-2022-09-07/>>. Acesso em: 13/07/2025.

Poynter. **Why Community Notes mostly fails to combat misinformation**. (2023).

Disponível em:

<<https://www.poynter.org/fact-checking/2023/why-twitters-community-notes-feature-mostly-fails-to-combat-misinformation/>>. Acesso em: 13/07/2025

Yuwei Chuai, Haoye Tian, Nicolas Pröllochs, e Gabriele Lenzini. 2024. **Did the Roll-Out of Community Notes Reduce Engagement With Misinformation on X/Twitter?** Proc. ACM Hum.-Comput. Interact. 8, CSCW2, Article 428 (November 2024), 52 pages. <https://doi.org/10.1145/3686967>

BBC. **Meta is ditching fact checkers for X-style community notes. Will they work?**. (2025). Disponível em:

<<https://www.bbc.com/news/articles/c4g93nvrzd7o>>.

Acesso em: 15/07/2025

Meta. **Introducing Community Notes**. Disponível em:

<<https://www.meta.com/technologies/community-notes/>>. Acesso em: 15/07/2025.

Poynter. **An open letter to YouTube's CEO from the world's fact-checkers.** (2022). Disponível em:

<<https://www.poynter.org/fact-checking/2022/an-open-letter-to-youtubes-ceo-from-the-worlds-fact-checkers/>>. Acesso em: 15/07/2025

BBC. **TikTok users being fed misleading election news, BBC finds.** (2024).

Disponível em: <<https://www.bbc.com/news/articles/c1ww6vz1l81o>>. Acesso em: 15/07/2025

Poynter. **Two years after an open letter to YouTube, fact-checkers remain dissatisfied with the platform's inaction.** (2024). Disponível em:

<<https://www.poynter.org/ifcn/2024/fact-checking-youtube-misinformation/>>. Acesso em: 24/07/2025

New York Times. **A.I. Is Getting More Powerful, but Its Hallucinations Are Getting Worse.** (2025). Disponível em:

<<https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html>>. Acesso em: 24/07/2025