



DISINFORMATION IN THE DIGITAL AGE: IMPACTS ON DEMOCRACY AND STRATEGIES FOR MITIGATION

THIERRY WARIN

PR

2024PR-03
FOR REFLECTION

The papers For Reflection... are working papers published for discussion and comment and are the sole responsibility of their authors. These publications are based on research findings.

Les documents Pour Réflexion... sont des documents de travail publiés pour susciter échanges et commentaires et n'engagent que leurs auteurs. Ces publications s'appuient sur des résultats de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO Partners - Les partenaires du CIRANO

Corporate Partners - Partenaires corporatifs

Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque nationale du Canada
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Énergir
Hydro-Québec
Innovation, Sciences et Développement économique Canada
Intact Corporation Financière
Investissements PSP
Manuvie Canada
Ministère de l'Économie, de l'Innovation et de l'Énergie
Ministère des finances du Québec
Mouvement Desjardins
Power Corporation du Canada
Ville de Montréal

Academic Partners - Partenaires universitaires

École de technologie supérieure
École nationale d'administration publique
HEC Montréal
Institut national de la recherche scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

CIRANO collaborates with many centers and university research chairs; list available on its website. *Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.*

© Juillet 2024. Thierry Warin All rights reserved. *Tous droits réservés.* Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source. *Reproduction partielle permise avec citation du document source, incluant la notice ©.*

The observations and viewpoints expressed in this publication are the sole responsibility of the authors; they do not represent the positions of CIRANO or its partners. *Les idées et les opinions émises dans cette publication sont sous l'unique responsabilité des auteurs et ne représentent pas les positions du CIRANO ou de ses partenaires.*

ISSN 2563-7266 (online version)

Disinformation in the Digital Age: Impacts on Democracy and Strategies for Mitigation

Thierry Warin* Professor at HEC Montréal Researcher and Fellow at CIRANO and OBVIA Fellow, Digital, Data and Design (D³) Institute at Harvard Business School

* Profil ORCID disponible à: <https://orcid.org/0000-0002-5921-3428> Email: thierry.warin@hec.ca

Abstract / Résumé

Disinformation has become a substantial threat to democratic institutions and societal stability, intensified by the proliferation of social media. Traditionally spread through media like newspapers and television, information was controlled by gatekeeping mechanisms. However, the rise of social media has changed this dynamic, allowing rapid, widespread dissemination without traditional checks. Algorithms prioritizing engagement amplify sensational content, facilitating the spread of falsehoods.

This paper examines the extensive impact of disinformation, including the erosion of public trust, distortion of democratic processes, and manipulation of electoral outcomes. It traces the evolution of disinformation from traditional media to digital platforms, emphasizing the need for scientific research to develop detection technologies and effective policies. Strategies to combat disinformation include enhancing digital literacy, increasing transparency of information sources, and implementing regulatory frameworks for social media accountability. AI-driven tools and international cooperation are essential to safeguard democratic integrity. It is crucial to reflect on and discuss these issues to develop comprehensive and effective solutions.

La désinformation est devenue une menace importante pour les institutions démocratiques et la stabilité de la société, intensifiée par la prolifération des médias sociaux. Traditionnellement diffusées par des médias tels que les journaux et la télévision, les informations étaient contrôlées par des mécanismes de régulation. Toutefois, l'essor des médias sociaux a modifié cette dynamique, en permettant une diffusion rapide et à grande échelle sans les contrôles traditionnels. Les algorithmes qui privilégient l'engagement amplifient le contenu sensationnel, facilitant ainsi la diffusion de fausses informations.

Ce « Rapport pour réflexion » examine l'impact considérable de la désinformation, notamment l'érosion de la confiance du public, la distorsion des processus démocratiques et la manipulation des résultats électoraux. Il retrace l'évolution de la désinformation, des médias traditionnels aux plateformes numériques, en soulignant la nécessité de la recherche scientifique pour développer des technologies de détection et des politiques efficaces. Les stratégies de lutte contre la désinformation comprennent le renforcement de la culture numérique, l'amélioration de la transparence des sources d'information et la mise en œuvre de cadres réglementaires pour la responsabilité des médias sociaux. Les outils pilotés par l'IA et la coopération internationale sont essentiels pour préserver l'intégrité démocratique. Il est essentiel de réfléchir à ces questions et d'en débattre afin d'élaborer des solutions globales et efficaces.

Keywords: Disinformation; Social media; Artificial Intelligence; Democratic Institutions; Public Trust; Digital Literacy / Désinformation, Médias sociaux, Intelligence artificielle, Institutions démocratiques, Confiance du public, Culture numérique.

To quote this document: Warin, T. (2024). Disinformation in the Digital Age: Impacts on Democracy and Strategies for Mitigation (2024PR-03, For Reflection, CIRANO) <https://doi.org/10.54932/GQWB1497>

1. Introduction

The rapid advancement of generative artificial intelligence (AI) technologies presents both incredible opportunities and formidable challenges. Among the latter, one of the most pressing is the potential for AI to exacerbate the spread of disinformation, with profound implications for democratic societies. Imagine scrolling through your social media feed and encountering a seemingly authentic video of a prominent political figure making controversial statements, only to discover that it is an expertly crafted deepfake. This unsettling scenario is not just a hypothetical; it is a reality we are increasingly confronting.

In recent months, high-profile incidents have underscored the dangers posed by AI-driven disinformation. For instance, in a bid to undermine an opponent, Florida Governor Ron DeSantis shared hyper-realistic images of Donald Trump fabricated to appear as if he were embracing Anthony Fauci. Similarly, a pro-DeSantis Political Action Committee (PAC) circulated a synthetic audio clip of Trump criticizing a fellow Republican.¹ These deepfakes, produced with advanced AI tools, are not only convincing but also challenging to detect, raising alarms about their potential to manipulate public opinion and disrupt democratic processes.

Disinformation, defined as deliberately misleading or false information disseminated with the intent to deceive, has emerged as a significant threat to democratic institutions and societal stability. Although in some circumstances it comes with great benefits, overall, the advent of social media has magnified this issue, creating an environment where false narratives can proliferate rapidly, influencing public opinion and behavior in unprecedented ways. Without going too far in History, before

the Internet, information was disseminated through traditional media channels such as newspapers, radio, and television, which had inherent gatekeeping mechanisms to help contain and regulate the spread of false information, at least in some countries in the world. Disinformation campaigns were then often state-sponsored and used as tools of propaganda, particularly during periods of political tension. The scale of dissemination was limited by the reach and influence of these traditional media outlets.

With the rise of social media, the dynamics of disinformation have fundamentally changed. Social media platforms, with their vast reach and real-time communication capabilities, enable the rapid and widespread distribution of content. The decentralized nature of these platforms bypasses traditional gatekeeping mechanisms, allowing true and false information to spread unchecked. Algorithms that prioritize engagement often amplify sensational or emotionally charged content, regardless of its veracity, creating an environment where disinformation can proliferate more easily and widely. Humans can generate and disseminate disinformation at an unprecedented scale, targeting specific audiences with tailored falsehoods.

AI has further transformed the landscape of disinformation. With machine learning-based AI (ML), recommender systems have allowed to target the perfect audience for an information. With the latest generative AI systems, we have the potential to put together super creators of false content. Generative AI technologies can produce highly realistic and persuasive text, images, and videos, including deepfakes. For instance, a recent Newsweek article highlights a deepfake video of Kamala Harris making a bumbling and awkward speech,² demonstrating the advanced capabilities of AI-generated disinformation. Such AI-generated pieces of disinformation are harder to detect and can be produced at scale, making them potent tools for spreading false narratives. The ability of AI—both ML and generative AI—to create convincing disinformation poses significant challenges for detection and mitigation, as traditional fact-checking methods struggle to keep pace with the sophistication and volume of AI-generated content. The use of generative AI by both humans and robots to generate disinformation compounds the challenge of managing and mitigating its impact. While generative AI can also be used to generate scientifically based content, this study focuses specifically on disinformation.

This study aims to discuss the following research question: How have the rise of social media and advancements in AI technology transformed the dynamics of disinformation, and what are the implications for democratic institutions and electoral integrity?

Science plays a crucial role in understanding, detecting, and combating disinformation. Scientific research can provide insights into the psychological and sociological mechanisms that make disinformation effective. It also helps develop advanced technologies for detecting disinformation, such as AI-driven fact-checking tools and deepfake detection systems. Furthermore, science-based policies and regulations can help create frameworks that mitigate the spread and impact of disinformation.

This paper examines the need for scientific research to develop detection technologies and effective policies. The paper reflects on the concepts and dynamics of disinformation, delving into how it spreads and influences public perception. It examines the legal aspects of disinformation, exploring how laws and regulations can be crafted to mitigate its adverse effects while preserving democratic principles. Through this multifaceted approach, the paper aims to foster a deeper understanding and stimulate thoughtful discussions on the urgent needs to address the challenges posed by disinformation.

2. Literature Review

Disinformation, as a concept, has been explored extensively in the academic literature. It is generally defined as false information that is deliberately spread to deceive people (Wardle & Derakhshan, 2017). The intentional nature of disinformation distinguishes it from misinformation, which is false information spread without harmful intent. The literature highlights several motives behind disinformation: political manipulation – including geo-political disruption -, financial gain, and social influence (Tandoc et al., 2018). Disinformation has been a longstanding issue, historically spread through traditional media channels such as newspapers, radio, and television. A seminal study by Romerstein (2001) explored how disinformation was used as a tool of psychological warfare during the Cold War. The study emphasized that disinformation campaigns were meticulously crafted to manipulate public perception and weaken enemy morale, often involving the dissemination of false narratives through controlled media outlets (Romerstein, 2001). Key findings from Romerstein's work indicate that traditional media, despite its gatekeeping mechanisms, was susceptible to being co-opted for the spread of state-sponsored disinformation. He illustrated how strategic falsehoods were propagated to achieve specific political and military objectives. This historical context underscores the persistent challenge of disinformation, which has only been magnified in the digital age.

2.1 The Role of Social Media in Disinformation

The advent of social media has revolutionized the dissemination of information, including disinformation. Social media platforms allow for the rapid and widespread distribution of content, often bypassing traditional gatekeeping mechanisms (Allcott & Gentzkow, 2017). Allcott and Gentzkow (2017) examine the influence of fake news during the U.S. presidential election in 2016, revealing that fake news stories were widely shared on social media platforms. Their study estimates that the average American saw several fake news stories during the election period, with pro-Trump fake news significantly

outnumbering pro-Clinton fake news. This dissemination of false information likely played a role in shaping voter perceptions and potentially influencing electoral outcomes. They highlight the mechanisms by which fake news spreads, focusing on social media algorithms – including recommender systems – that prioritize engaging content. Allcott and Gentzkow (2017) explain that these algorithms inadvertently amplify sensational and misleading stories, as they tend to generate higher engagement than factual news. They underscore the critical role of social media platforms in the propagation of fake news and the need for more stringent content moderation policies.

To mitigate the impact of fake news, Allcott and Gentzkow (2017) propose several options. They recommend improving digital literacy among users to help them critically evaluate the credibility of online information. Additionally, they suggest that social media companies should enhance their efforts to detect and remove fake news, possibly by using advanced algorithms and human review systems. Policymakers are also encouraged to consider regulatory measures that promote transparency and accountability in digital news dissemination. This has led to an environment where disinformation can spread more easily and quickly, often with significant real-world consequences. Studies have shown that false information spreads faster – the question of velocity and more widely on social media than true information (Vosoughi et al., 2018).

2.2 Impacts of Disinformation and Strategies to Combat Disinformation

The impacts of disinformation are multifaceted. Politically, disinformation can influence voter behavior, distort public opinion, and undermine trust in democratic institutions (Bennett & Livingston, 2018). Socially, exacerbated by recommender systems, it can contribute to polarization, exacerbate conflicts, and spread fear and uncertainty (Guess et al., 2019). Economically, disinformation can damage reputations, influence markets, and disrupt economic activities (Tambini, 2017). The potential catastrophic risks of disinformation have become a focal point in recent research. A catastrophic risk is one with a very low probability of occurrence, but a very high consequence when it realizes. These risks include the erosion of democratic processes, the exacerbation of social divisions, and the potential for disinformation to incite violence or unrest (Bradshaw & Howard, 2018). The literature suggests that the interconnected and global nature of social media amplifies these risks, creating a volatile information environment where disinformation can have far-reaching and unpredictable effects (Wardle & Derakhshan, 2017). Social media has significantly altered the dynamics of elections. Social media platforms have been used to spread disinformation during election campaigns, aiming to influence voter perceptions and behaviors (Persily, 2017). The Cambridge Analytica scandal is a prominent example, where data from millions of Facebook users were used to create targeted political advertisements based on psychological profiles, demonstrating the sophisticated and pervasive nature of modern disinformation tactics (Isaak & Hanna, 2018).

Media literacy has emerged as a crucial tool in combating disinformation. Bulger and Davison (2018) highlight that media literacy efforts focus on enhancing critical thinking skills and promoting informed participation in the media landscape. They emphasize the importance of youth participation, teacher training, parental support, policy initiatives, and the construction of an evidence base to evaluate the effectiveness of media literacy programs. These initiatives have shown positive outcomes in helping individuals critically assess information and recognize disinformation (Bulger & Davison, 2018).

Despite its promise, media literacy faces several challenges. Bulger and Davison (2018) note a lack of comprehensive evaluation data, which hampers the ability to measure the long-term impact of media literacy programs. Additionally, there is a risk of overconfidence among individuals who may feel they can discern all disinformation but still fall prey to sophisticated false narratives. The authors recommend ongoing assessment and adaptation of media literacy efforts to ensure they remain effective in the evolving media landscape. Looking forward, Bulger and Davison (2018) advocate for a multifaceted approach to media literacy that includes collaboration among educators, policymakers, technologists, and philanthropists. By integrating media literacy into educational curricula and providing continuous support and resources for educators, media literacy can be more effectively implemented. They also emphasize the need for policy initiatives that support media literacy at a systemic level, ensuring it is a sustained priority in the fight against disinformation.

2.3 The Need for Comprehensive Strategies

The literature also underscores the urgent need for comprehensive strategies to combat disinformation. Effective measures include improving digital literacy, enhancing the transparency of information sources, and developing robust regulatory frameworks to hold social media platforms accountable (Guess et al., 2019). Policymakers and civil servants must stay informed about the evolving nature of disinformation and its impacts to devise effective responses and safeguard democratic institutions.

Understanding the prevalence and predictors of fake news dissemination on social media is critical for developing effective strategies to combat disinformation. Guess et al. (2019) conducted a detailed study examining how fake news spreads on Facebook. Their findings indicate that although fake news constitutes a small fraction of the overall news consumption, it disproportionately affects certain demographics, particularly older adults. The study highlights that users with lower levels of digital literacy are more likely to share false information. These insights underscore the importance of targeted digital literacy programs and algorithmic adjustments by social media platforms to reduce the spread of fake news (Guess et al., 2019).

2.4 Ethical and Regulatory Frameworks

Enhancing the transparency of information sources involves making the origins and intentions behind information clear, which can help individuals assess credibility (Wardle & Derakhshan, 2017). Developing robust regulatory frameworks can hold social media platforms accountable for the content they host and spread, ensuring that disinformation is quickly identified and addressed (Floridi et al., 2018). Addressing the challenges of disinformation in the digital age requires robust ethical frameworks, particularly as AI plays an increasingly prominent role. Floridi et al. (2018) provide a comprehensive ethical framework for the development and deployment of AI, emphasizing opportunities and risks associated with AI technologies. Their study identifies five core ethical principles: beneficence, non-maleficence, autonomy, justice, and explicability. These principles are crucial for ensuring that AI systems are designed and used in ways that promote human dignity, protect individual rights, and enhance societal cohesion (Floridi et al., 2018).

Key findings from the study highlight the dual advantage of adopting an ethical approach to AI. By aligning AI development with these ethical principles, organizations can not only leverage AI's potential to benefit society but also mitigate risks associated with its misuse. This includes the ethical deployment of AI in detecting and combating disinformation, where transparency, accountability, and human oversight are paramount to preserving public trust and preventing harm. Furthermore, policymakers and civil servants need to stay informed about the evolving nature of disinformation. Of course, this is no easy task. This involves keeping up with the latest research, understanding the technological advancements that enable disinformation, and being aware of the socio-political contexts in which disinformation thrives. By doing so, they can devise effective responses to safeguard democratic institutions and ensure the integrity of public discourse.

3. Concepts and Dynamics of Disinformation

The rise of social media has fundamentally transformed the landscape of information dissemination, creating both opportunities and challenges for public discourse. Social media platforms, with their vast reach and real-time communication capabilities, have become powerful tools for spreading information. Volume and velocity, but what about veracity? However, these same features make them particularly susceptible to the rapid spread of disinformation. The algorithms that drive social media prioritize engagement, often amplifying sensational or emotionally charged content, regardless of its veracity. This environment allows false information to spread more quickly and widely than ever before, posing significant risks to public trust, electoral integrity, and social cohesion. Disinformation campaigns on social media have been linked to various harmful outcomes, including undermining democratic processes, fostering polarization, and inciting violence.

3.1 Disinformation and Propaganda

Disinformation and propaganda are distinct yet interrelated phenomena that play critical roles in shaping public perception and influencing political and social landscapes. Disinformation is characterized by the intentional dissemination of false or misleading information to deceive an audience, obscuring the truth, creating confusion, and manipulating beliefs and behaviors. It spreads through various channels, including traditional media, social media, and word of mouth, often exploiting emotional and psychological triggers to achieve its goals.

Propaganda, while sometimes involving disinformation, is a broader concept encompassing the strategic dissemination of information—whether true, partially true, or false—to promote a specific political agenda or ideology. It aims to shape public opinion and behavior in favor of the propagator's objectives. Historically, propaganda has been used by states and political groups to mobilize support, justify actions, and suppress dissent. In contemporary contexts, the rise of digital media has amplified the reach and impact of both disinformation and propaganda, allowing these tools to influence public discourse and electoral processes globally. Understanding these dynamics is essential for developing effective strategies to counteract their adverse effects on democracy and societal cohesion.

3.2 The Strategic Omission of Facts

Disinformation is typically associated with the deliberate dissemination of false or misleading information. However, it is crucial to recognize that the strategic omission of facts can also constitute a form of disinformation, ie avoiding the survival bias in Statistics. This practice, often subtle and insidious, can shape perceptions and influence decisions by presenting an incomplete or skewed version of reality. Indeed, the concept of survival bias, a cognitive bias where one focuses on successful outcomes while overlooking failures, provides a valuable framework for understanding how the omission of facts can mislead. Survival bias occurs when only the surviving elements are considered, leading to overly optimistic conclusions. For instance, during World War II, military analysts initially focused on the bullet holes in returning aircraft to determine which areas needed reinforcement. They overlooked the aircraft that did not return, which likely had critical hits in areas that appeared unscathed on the surviving planes. This oversight led to potentially flawed conclusions about aircraft vulnerabilities (Wald, 1943).

Similarly, in the context of information dissemination, emphasizing successful outcomes or favorable data while ignoring failures or negative data can create a misleading narrative. In the realm of disinformation, omitting relevant facts can be as damaging as fabricating falsehoods. By selectively presenting information, propagators of disinformation can construct

narratives that serve specific agendas, manipulate public opinion, and obscure the truth. This method is particularly effective because it leverages the audience's trust in the presented information's accuracy while subtly guiding them toward a biased conclusion. For example, a news report highlighting the economic benefits of a new policy without mentioning its adverse effects on specific communities engages in “disinformation through omission”. This practice can distort public perceptions and lead to poorly informed decisions, ultimately eroding democratic processes and diminishing social trust.

When critical facts are omitted, the audience receives an incomplete picture, leading to skewed perceptions and misguided conclusions. This can affect public opinion, policy support, and even personal decisions. The omission of facts can reinforce existing biases and beliefs. When people are presented only with information that confirms their preconceptions, they are less likely to seek out or accept contradictory evidence (Nickerson, 1998). Consistent omission of important information can erode trust in information sources. When the public becomes aware that they are being misled by omission, it can diminish trust in media, government, and other institutions (Lazer et al., 2018). The study by Lazer et al. (2018) delves into the mechanisms behind the rapid spread of fake news on social media platforms. The authors highlight that false information spreads more quickly and broadly than true information due to its novel and sensational nature. The study emphasizes that fake news stories are more likely to be shared because they evoke strong emotional reactions, which prompts users to disseminate them further. This underscores indeed the significant challenge that fake news poses to maintaining an informed public and the integrity of the information ecosystem in its multiple dimensions.

3.3 The Spread and Impact of Fake News

As aforementioned, Lazer et al. (2018) discuss the role of algorithms in the spread of fake news. Social media platforms utilize algorithms designed to maximize user engagement by promoting content that generates high levels of interaction, the recommender systems. These algorithms will often lead to sensational and emotionally charged stories being more spread and focused to certain groups, which can include false information. The study calls for greater transparency in how these algorithms operate and suggests that adjustments are necessary to reduce the amplification of fake news and promote more reliable sources of information. An important issue is also polarization.

Lazer et al. (2018) propose several strategies to combat the spread of fake news. They advocate for improved digital literacy among the public to help individuals critically evaluate the credibility of information sources. They also recommend the development and deployment of advanced detection technologies, such as machine learning algorithms, to identify and flag fake news stories. Additionally, the study suggests that social media platforms should take a more active role in curating content and ensuring that reliable information is more prominently displayed. These combined efforts are crucial for mitigating the impact of fake news on public discourse and democratic processes. To counteract disinformation through omission, several strategies can be implemented. Encouraging media outlets and information sources to provide balanced and comprehensive coverage, including both positive and negative aspects of a story, can promote comprehensive reporting. Fostering critical thinking skills among the public can enhance skepticism and encourage the questioning of incomplete narratives. Supporting independent fact-checking organizations and promoting transparency in information dissemination can ensure that critical facts are not omitted. We can also build new indicators to represent the level of veracity of a story. We could also use large scientific databases (commercial or open source), in particular in this era of generative AI.

Generative AI represents a new frontier in the evolution of disinformation. These advanced AI systems can create highly realistic and persuasive content, including text, images, and videos. While generative AI holds immense potential for positive applications, such as in creative industries and personalized services, it also presents new challenges in the realm of disinformation. AI-generated deepfakes – realistic but fake audio or video recordings – can be used to create convincing false narratives, making it increasingly difficult for individuals to distinguish between genuine and fabricated content. The ability of generative AI to produce high-quality disinformation at scale poses a significant threat to information integrity and public trust. Addressing these challenges requires a multifaceted approach, including the development of robust detection technologies, the promotion of digital literacy, and the implementation of effective regulatory frameworks to mitigate the misuse of generative AI for malicious purposes. We can also use generative AI to promote scientifically based content. By understanding the dynamics of social media, the nature of disinformation, and the implications of generative AI, policymakers and civil servants can develop more effective strategies to protect democratic institutions and promote a well-informed public.

3.4 Extremist Ideologies and Disinformation

The online expression of extremist ideologies plays a significant role in the dissemination of disinformation. Holt et al. (2022) examine how far-right extremist forum users propagate their beliefs through coordinated campaigns. Their study reveals that these groups exploit social media to spread false narratives, recruit members, and engage in harassment tactics like trolling and doxing. The findings underscore the importance of monitoring online extremist activities and developing strategies to counteract the spread of disinformation rooted in extremist ideologies (Holt et al., 2022). Social media algorithms have significantly contributed to the rise and visibility of alt-right ideologies. Daniels (2018) explores how these algorithms, designed to maximize user engagement, often prioritize sensational and polarizing content. This mechanism creates an environment

where extremist narratives can flourish, as they tend to generate high levels of interaction. The study illustrates that platforms like YouTube and Twitter inadvertently amplify the reach of such content, accelerating the radicalization process of individuals exposed to it. The algorithms used by social media platforms are not neutral; they are tailored to keep users engaged by feeding them content that aligns with their existing beliefs and interests. This is the naïve purpose of recommender systems.

Daniels (2018) found that this leads to the formation of echo chambers, where users are repeatedly exposed to similar extremist views without encountering opposing perspectives. This reinforcement of ideology not only solidifies users' beliefs but also encourages the spread of disinformation and extremist propaganda. The study underscores the need for a more responsible approach to content moderation by these platforms. Daniels (2018) emphasizes the urgent need for regulatory frameworks to address the unintended consequences of algorithm-driven content promotion. The proliferation of alt-right content due to these algorithms poses significant risks to social cohesion and democratic processes.

This is the issue of polarization. Policymakers must consider strategies to mitigate these effects, such as enhancing transparency in algorithmic processes and promoting digital literacy among users. By doing so, society can better manage the influence of social media algorithms on public discourse and curb the spread of extremist ideologies. Geopolitical influences on disinformation are significant, with several countries engaging in state-sponsored disinformation campaigns to influence geopolitical outcomes. Russia, for example, has been notorious for its use of disinformation to interfere in the political processes of other nations, including the 2016 U.S. presidential election (Rid, 2020). These campaigns often involve the use of bot networks and fake accounts to spread false information and sow discord among the population of the target country (Howard & Kollanyi, 2016). Disinformation is a key component of modern information warfare, where states use false or misleading information as a strategic tool to achieve military or political objectives (Singer & Brooking, 2018).

In various contexts, disinformation has significant implications. Disinformation within health communication poses significant public health risks. Broniatowski et al. (2018) explore how Twitter bots and Russian trolls have weaponized the vaccine debate, amplifying both pro- and anti-vaccine messages to create division and confusion. Their study found that these actors contribute to the spread of disinformation by promoting discord and undermining public trust in vaccination, ultimately affecting public health outcomes. This highlights the need for robust countermeasures to protect public health information from malicious interference (Broniatowski et al., 2018). Economic disinformation can influence markets and investor behavior. False information about companies or economic conditions can lead to stock market fluctuations and financial instability (Tambini, 2017).

Another example can be climate change. Environmental disinformation, such as climate change denial, hampers efforts to address global environmental challenges by spreading doubt about scientific consensus, delaying policy action, and undermining public support for environmental initiatives (Oreskes & Conway, 2011). In this context, disinformation can significantly distort the policy-making process. When policymakers rely on false or misleading information, they may enact policies that are ineffective or even harmful. For example, disinformation regarding climate change has profound implications for public understanding and policy action. McCright and Dunlap (2010) discuss how the American conservative movement has systematically undermined climate science and policy through anti-reflexive strategies. These strategies involve promoting doubt about scientific consensus, influencing public opinion, and delaying policy interventions. The study highlights the need for robust countermeasures, including improving public understanding of scientific methods and strengthening policy frameworks to resist disinformation campaigns (McCright & Dunlap, 2010). Public support for policies can also be manipulated through disinformation, leading the public to support or oppose measures that do not align with their best interests (van der Linden et al., 2017). Government resources may be misallocated due to disinformation. During health crises like the COVID-19 pandemic, disinformation about treatments and preventive measures can lead to wasted resources on ineffective solutions, while effective measures are overlooked or underfunded (Broniatowski et al., 2018).

Disinformation undermines trust in public institutions, including governments, media, and scientific communities. When citizens cannot discern reliable information from falsehoods, trust in these institutions diminishes, weakening societal consensus (Lazer et al., 2018). This erosion of trust can lead to increased polarization, as different groups of people base their beliefs and actions on different sets of “facts,” often perpetuated by disinformation (Sunstein, 2017). Disinformation on social media is further exacerbated by algorithms that create filter bubbles and echo chambers. Flaxman et al. (2016) investigate how these algorithms influence news consumption patterns. Their study reveals that social networks and search engines can increase ideological segregation, but also expose users to a broader range of perspectives than previously thought. They found that most online news consumption is still driven by direct visits to mainstream news sites, tempering the impact of these algorithms on ideological segregation (Flaxman et al., 2016). Disinformation poses a direct threat to democratic processes by distorting electoral outcomes and undermining the integrity of elections. When voters make decisions based on false information, the legitimacy of elected officials and the policies they implement can be called into question (Persily, 2017). This undermines the social contract, as the foundational principle of democracy—governance by informed consent of the governed—is compromised. Persistent disinformation can destabilize societies by fostering division, resentment, and conflict. It can amplify societal tensions, leading to unrest and violence, as seen in various contexts worldwide (Bennett & Livingston, 2018). A

fractured society struggles to maintain a cohesive social contract, as mutual trust and shared values are essential for social stability and cooperation. In our day and age, we are observing this effect too often.

4. Disinformation and the Role of Science

So, what about science? Disinformation in the realm of science involves the deliberate spread of false or misleading information about scientific facts, theories, or research findings. This can include denial of scientific consensus, manipulation of data, or promotion of pseudoscience (Lewandowsky et al., 2017). Common targets of scientific disinformation include climate change, vaccination, and genetic engineering due to their complexity and emotional charge. Misrepresentation of scientific consensus is a tactic used to create the illusion of controversy, undermining public trust and delaying policy action (Oreskes & Conway, 2011). Cherry-picking data selectively presents information supporting a particular narrative while ignoring contradictory data, often using out-of-context quotes or isolated studies (Boykoff & Boykoff, 2004). Pseudoscientific claims promote theories or products not supported by rigorous scientific evidence, typically for ideological or financial gain (Hansson, 2017).

4.1 Impacts of Scientific Disinformation

The impacts of scientific disinformation are profound. Disinformation about vaccines and medical treatments can lead to public health crises, as seen with the MMR vaccine controversy, which resulted in reduced vaccination rates and measles outbreaks (Larson, 2018). Climate change denial and misinformation hinder efforts to mitigate environmental damage, delaying necessary policies and innovations (McCright & Dunlap, 2010). Persistent disinformation erodes public trust in scientific institutions and experts, complicating the communication of accurate information and the implementation of evidence-based policies (Gauchat, 2012).

4.2 The Role of Social Media and Bots

Disinformation campaigns leverage social media's vast reach and real-time communication capabilities to spread false narratives rapidly. This has been exacerbated by the activity of social bots, which are automated accounts that disseminate low-credibility content, significantly influencing public opinion and the spread of false information. A study by Shao et al. (2018) highlights the role of these social bots in amplifying disinformation, showing how they contribute to the rapid spread of low-credibility content across platforms, complicating efforts to maintain information integrity (Shao et al., 2018). The study found that social bots play a critical role in the early spreading of low-credibility content, often by targeting influential users who can further amplify these messages to a broader audience. Specifically, the results indicated that bots are responsible for a significant portion of the traffic surrounding low-credibility sources, disproportionately affecting the visibility and perceived credibility of disinformation. This activity not only increases the reach of false information but also accelerates its dissemination speed, outpacing efforts by fact-checkers and legitimate news sources to correct the record. Addressing the challenges posed by social bots requires advanced detection technologies, such as machine learning algorithms that can identify bot behavior patterns, and robust regulatory frameworks to mitigate their impact on public discourse.

4.3 Technological and Educational Solutions

So, what about education? Disinformation in the digital age presents significant challenges, particularly due to the widespread use of social media platforms. Lazer et al. (2018) in their comprehensive study on fake news, highlight the critical role science plays in understanding and combating this phenomenon. The study identifies that false news spreads faster and more widely than true news, largely due to its novelty and the emotional reactions it evokes. Social bots further exacerbate this issue by amplifying disinformation, making it difficult for fact-checkers and reliable news sources to keep up (Lazer et al., 2018). Key findings from the study indicate that the nature of social media algorithms, which prioritize engagement, contributes significantly to the virality of disinformation. Moreover, the authors stress the importance of developing advanced technological solutions, such as AI-driven fact-checking tools, to identify and mitigate the spread of false information effectively. This underscores the need for a multifaceted approach, combining technological, educational, and regulatory measures to combat the spread of disinformation and protect the integrity of public discourse.

4.4 Enhancing Public Understanding and Transparency

Addressing scientific disinformation requires enhancing public understanding of scientific methods and consensus to make individuals less susceptible to false information (National Academies of Sciences, Engineering, and Medicine, 2016). Establishing robust fact-checking mechanisms and quickly addressing false claims with accurate information are crucial, with initiatives like FactCheck.org and Snopes playing significant roles (Lewandowsky et al., 2017). Promoting transparency in scientific research, including open access to data and methodologies, can build public trust and allow for independent verification of findings (Nosek et al., 2015). Scientists and research institutions should engage with the public and media to communicate their findings clearly and effectively, countering disinformation with credible information (Scheufele & Krause, 2019). The scientific community should also advocate for policies that protect and promote the integrity of science, including regulations against the deliberate spread of disinformation (Douglas, 2015).

4.5 The Role of Policy and Regulation

Policymakers play a crucial role in supporting science-based decision-making. This includes enacting policies that ensure funding for scientific research and communication efforts, as well as implementing regulations that hold purveyors of disinformation accountable. By creating an environment that values and supports evidence-based reports and scientific integrity, policymakers can help safeguard public health, environmental sustainability, and societal wellbeing. To comprehensively tackle scientific disinformation, it is vital to integrate these strategies into broader efforts to promote scientific literacy, transparency, and public engagement. This holistic approach can help restore public trust, ensure accurate information dissemination, and foster a more informed society capable of making decisions based on sound scientific evidence.

4.6 Engaging with the Public

The role of science in combating disinformation extends beyond the mere provision of facts. Scientists play a critical role in actively engaging with the public to debunk myths and clarify misconceptions. This involves a proactive approach to public communication, leveraging various platforms to reach a broad audience. Effective science communication requires not only the dissemination of accurate information but also an understanding of how to communicate complex scientific concepts in an accessible and engaging manner. By building relationships with the media and utilizing social media platforms, scientists can directly address misinformation and provide the public with reliable information.

4.7 Fostering Transparency and Accountability

Scientific institutions must also prioritize transparency in their operations. Open science practices, such as sharing data and research methodologies, enhance credibility and allow for independent verification of scientific claims. Transparency fosters a culture of accountability and trust, making it more difficult for disinformation to take root. Moreover, institutions should support and collaborate with fact-checking organizations and engage in continuous monitoring of misinformation trends to respond swiftly and effectively.

5. Disinformation Detection

Can technology help? Disinformation detection involves several sophisticated methods, each targeting different aspects of false information dissemination. The rapid spread of disinformation on digital platforms necessitates advanced technological solutions for detection and mitigation. Conroy et al. (2015) examine various methods for automatic deception detection, highlighting the efficacy of machine learning algorithms and Natural Language Processing (NLP) techniques. These methods can analyze text to identify deceptive patterns, thereby aiding in the identification of fake news. Tools such as ClaimBuster and Full Fact's automated system employ these methods (Hassan et al., 2017; Conroy et al., 2015). Research and development in disinformation detection also drive technological advancements in AI, NLP, and data analysis, which have broader applications beyond combating disinformation (Zhou & Zafarani, 2018). These technologies can enhance various fields, including cybersecurity, healthcare, and market analysis, demonstrating the multifaceted benefits of investing in advanced detection systems.

5.1 Network and Behavioral Analysis

Network analysis focuses on mapping and analyzing information spread across social networks to identify disinformation sources and influential nodes. Techniques for bot detection involve identifying automated accounts that amplify disinformation using behavioral patterns and machine learning (Shao et al., 2018; Ferrara et al., 2016). These techniques are crucial in pinpointing and mitigating the influence of bots that propagate false information, thereby protecting the integrity of online discussions.

5.2 Image and Video Verification

Image and video analysis, including deepfake detection, employs AI to identify manipulated media by analyzing inconsistencies in facial movements, lighting, and audio-visual synchronization. Reverse image search tools, such as Google Reverse Image Search, are used to verify the origin and authenticity of images circulating online (Agarwal et al., 2019). These tools are essential for combating visual disinformation, which can be particularly convincing and damaging.

5.3 Cross-Referencing and Source Verification

Cross-referencing and source verification involve comparing information across multiple reputable sources to identify inconsistencies and verify accuracy. Analyzing metadata of digital content helps trace the origin and authenticity of the information. These traditional journalistic practices are enhanced by digital tools that can handle large volumes of data quickly and efficiently, aiding in the fight against disinformation.

5.4 Challenges and Benefits in Disinformation Detection

However, disinformation detection faces several challenges. The rapid evolution of disinformation tactics, including AI-generated deepfakes, makes it difficult for detection methods to keep pace (Chesney & Citron, 2019). The sheer volume and velocity of content generated on social media platforms complicate real-time monitoring and verification (Vosoughi et al., 2018). Ambiguity and context also play roles in complicating the task of distinguishing between disinformation, misinformation, and genuine content (Pennycook & Rand, 2019). Implementing detection methods, particularly those involving data analysis and surveillance, raises ethical and privacy concerns (Floridi et al., 2018).

Despite these challenges, effective disinformation detection can restore public trust in media and online platforms by reducing the spread of false information and ensuring the integrity of information sources (Lazer et al., 2018). It can protect the integrity of democratic processes by curbing disinformation's influence, particularly during election cycles, ensuring that voters make informed decisions (Bennett & Livingston, 2018). The development and promotion of disinformation detection tools can raise public awareness about the prevalence of disinformation and enhance digital literacy, encouraging individuals to critically evaluate information (Tully et al., 2020).

5.5 News Literacy and Public Education

Tully et al. (2020) emphasize the crucial role of news literacy in combating the spread of disinformation on social media. Their study focuses on developing and testing effective news literacy messages that can be disseminated on social media platforms to help users critically evaluate the information they encounter. The authors highlight that well-designed news literacy interventions can significantly improve users' ability to identify and resist fake news, thereby enhancing the overall quality of information in the digital environment. The study presents a detailed analysis of different strategies for creating news literacy messages. Tully et al. (2020) found that messages emphasizing the importance of critical thinking and providing clear guidelines on how to verify information were particularly effective. They also note the importance of using relatable and engaging content to capture users' attention and encourage them to apply these critical skills in their daily media consumption. The research suggests that a combination of educational and motivational elements in news literacy messages can lead to better outcomes.

To ensure the effectiveness of news literacy messages, Tully et al. (2020) conducted rigorous testing across various social media platforms. Their findings indicate that messages tailored to the specific characteristics and user behaviors of each platform were more successful in improving news literacy. The authors advocate for a collaborative approach, involving educators, media organizations, and platform designers, to implement these strategies widely. By doing so, it is possible to create a more informed public that is better equipped to navigate the challenges of the digital information landscape.

6. Disinformation and Legal Aspects

Disinformation presents complex legal challenges involving the balance between freedom of speech and protecting society from harm. Legal frameworks to combat disinformation vary globally. The European Union's Code of Practice on Disinformation (<https://digitalstrategy.ec.europa.eu/en/policies/code-practice-disinformation>) promotes transparency and accountability for online platforms. This code emphasizes the responsibility of social media companies to prevent the spread of false information by implementing measures such as transparency in political advertising and the closure of fake accounts. Additionally, it encourages collaboration between governments, tech companies, and civil society to address the issue comprehensively.

In the United States, the protection of free speech under the First Amendment poses significant challenges for regulating disinformation. The First Amendment protects even false speech unless it falls under exceptions such as defamation, fraud, or incitement to violence. This broad protection creates a complex environment where any regulatory attempts must carefully navigate constitutional boundaries. There are increasing calls for regulation, particularly on social media platforms, where disinformation spreads rapidly. These calls often focus on enhancing platform accountability and transparency without infringing on free speech rights. Efforts in the U.S. have seen proposals for greater oversight and the imposition of duties on social media companies to monitor and remove harmful content. However, these proposals face resistance due to fears of government overreach and the potential stifling of legitimate expression. The challenge remains to find a balance where harmful disinformation can be curtailed without compromising the fundamental right to free speech.

6.1 Defining Disinformation and Legal Challenges

One primary challenge in regulating disinformation is defining it without infringing on free speech. Disinformation is often subjective, as it involves false or misleading information intended to deceive. However, distinguishing between harmful disinformation and legitimate, albeit controversial, speech can be difficult. The subjective nature of truth complicates the legal definition, as what one party considers disinformation, another might see as an opinion or a differing perspective. The potential for government abuse of disinformation laws is a significant concern. Governments might exploit such laws to suppress dissent and stifle free expression under the guise of combating disinformation. Historical and contemporary examples show that

regimes can label opposition and critical journalism as "disinformation" to silence dissenting voices, posing a risk to democratic principles and freedom of the press.

Additionally, the rapid evolution of disinformation tactics, such as AI-generated deepfakes, complicates the establishment of effective legal frameworks. Deepfakes, which are highly realistic but entirely fabricated images, audio, and videos, represent a new frontier in disinformation. Their sophisticated nature makes detection and regulation challenging, as they can be used to create false narratives that are difficult to debunk.

The dynamic and evolving nature of digital technology further exacerbates these challenges. Legal frameworks must adapt quickly to new forms of disinformation and malicious technologies while ensuring they do not become obsolete. This requires continuous monitoring, research, and the development of flexible legal instruments capable of addressing emerging threats without infringing on fundamental rights.

To tackle these issues, a nuanced approach is essential. This includes developing clear, precise definitions of disinformation, ensuring transparency in the creation and enforcement of disinformation laws, and fostering public dialogue to build consensus on the boundaries of free speech and the need for regulation. Legal measures should be complemented by technological solutions and public education to enhance media literacy and critical thinking.

6.2 Case Studies and Legal Precedents

Several legal cases highlight these complexities. In *"United States v. Alvarez"* (2012), the U.S. Supreme Court struck down the Stolen Valor Act, which criminalized false claims about military decorations, citing free speech protections. This case underscores the tension between combating disinformation and protecting constitutional rights. The Stolen Valor Act was intended to preserve the integrity of military honors by criminalizing false statements about receiving such awards. However, the Supreme Court ruled that the Act violated the First Amendment. The Court's decision was based on the principle that even false speech is protected under the First Amendment, unless it directly causes harm, such as fraud or defamation. The ruling highlighted the difficulty in creating laws that punish disinformation without infringing on free speech rights. This case exemplifies the broader legal challenges in regulating disinformation. It illustrates how laws designed to combat falsehoods can conflict with the fundamental right to free speech. The subjective nature of determining what constitutes harmful disinformation versus protected speech makes it challenging to enforce such laws without risking censorship or abuse.

Furthermore, *"United States v. Alvarez"* reveals the potential consequences of overreaching disinformation laws. If such laws are too broadly defined or applied, they can suppress legitimate expression and dissent. This case sets a precedent that cautions against broad legislative measures to control disinformation, emphasizing the need for precise and narrowly tailored approaches that respect constitutional protections.

Addressing these legal challenges requires a careful balance. Legal frameworks must be designed to target the most harmful types of disinformation, such as those that incite violence or cause significant public harm, while safeguarding free speech. This balance is critical to ensuring that efforts to combat disinformation do not undermine the democratic principles they aim to protect.

6.3 International Approaches

Different countries have varying approaches to regulating disinformation. Germany's Network Enforcement Act (NetzDG) requires social media platforms to remove illegal content, including disinformation, within specific timeframes or face significant fines. This law aims to combat hate speech, fake news, and other illegal content by holding platforms accountable for quickly addressing flagged issues. The Act mandates transparency reports from social media companies, outlining how they handle content complaints and enforce their terms of service.

In contrast, China implements strict state-controlled measures to manage information dissemination. These measures often involve comprehensive surveillance, censorship, and control over online content, aiming to maintain social stability and national security. However, these strict regulations have been widely criticized for suppressing free expression and dissent. Chinese authorities regularly monitor, and censor content deemed politically sensitive or harmful to social order, limiting the ability of citizens to freely express their views and access diverse information sources.

These contrasting approaches highlight the global diversity in handling disinformation. While Germany focuses on holding private companies accountable within a legal framework that respects democratic principles, China's approach reflects a centralized, state-controlled model that prioritizes control over freedom. Both methods have sparked debates on the effectiveness and ethical implications of disinformation regulation.

In 2018, the Standing Committee on Access to Information, Privacy and Ethics of the House of Commons of the Canadian Parliament released the report *"Democracy Under Threat: Risks and Solutions in the Era of Disinformation and Data Monopoly"* (House of Commons, Standing Committee on Access to Information, Privacy and Ethics, 2018). The 26

recommendations outlined include enhancing the transparency of online political advertisements, addressing foreign funding during election periods, verifying algorithms, and improving digital literacy (Assemblée nationale du Québec, 2022).

Understanding these international differences is crucial for developing balanced global strategies to combat disinformation. Effective regulation must protect public interests and uphold democratic values, ensuring that measures to curb disinformation do not become tools for censorship or political control. International cooperation and dialogue are essential to share best practices, address cross-border disinformation challenges, and establish standards that safeguard both information integrity and freedom of expression.

7. Conclusion

Disinformation poses a significant threat to democratic processes, public trust, and societal cohesion. Addressing disinformation requires a multifaceted approach, combining legal regulation, technological solutions, and public education and international cooperation. Legal regulation involves creating robust frameworks that hold platforms accountable for the content they host while protecting free speech. This includes implementing clear guidelines on what constitutes disinformation and establishing enforcement mechanisms that ensure compliance without overreaching into censorship.

Technological solutions are essential to detect and mitigate disinformation. This involves developing advanced AI and machine learning tools to identify and flag false content in real-time. These technologies can analyze patterns, detect deepfakes, and monitor the spread of disinformation across platforms, providing a proactive approach to content moderation and detect deepfakes. Studies by Conroy et al. (2015) and Shao et al. (2018) demonstrate the effectiveness of machine learning algorithms and network analysis in detecting deceptive patterns and disinformation spread by social bots. Implementing these technologies widely can help maintain the integrity of online information.

Public education is critical in fostering a well-informed society capable of critically evaluating information. This includes integrating media literacy programs into educational curricula to teach individuals how to discern credible sources from false information. Research by Tully et al. (2020) shows that well-designed news literacy messages can improve users' ability to discern false information. Such initiatives should be integrated into school curricula and adult education programs to build a more informed and discerning public. Enhancing digital literacy is another crucial component.

International cooperation and the development of global standards are also crucial. Disinformation often transcends borders, making it a global issue that requires a coordinated response. Countries need to collaborate on establishing international standards and best practices for disinformation regulation, share intelligence and technological innovations, and support initiatives that promote information integrity globally. Support for independent fact-checking organizations and promoting open science and data transparency can also help build public trust and reduce the impact of disinformation. The work of independent fact-checkers, as highlighted by Graves and Cherubini (2016), emphasizes the importance of reliable verification processes. The National Assembly of Quebec's report highlights the need for modernizing laws related to public information access and emphasizes the importance of media literacy education to address the challenges posed by widespread disinformation (Assemblée nationale du Québec, 2022). By combining these strategies – legal, technological, educational, and international cooperation – societies can build a resilient information ecosystem that effectively counters disinformation while upholding democratic values and protecting free expression.

Policymakers should implement robust regulations that define and penalize the spread of disinformation without infringing on free speech. Legal frameworks, such as Germany's Network Enforcement Act and the EU's Code of Practice on Disinformation, provide valuable models for holding social media platforms accountable. These laws should mandate timely removal of false content and ensure transparency in the operations of digital platforms. As Lazer et al. (2018) argue, a concerted effort involving transparency in algorithmic processes and improved digital literacy can significantly curb the spread of fake news. The future of a well-informed public and robust democratic institutions depends on our collective ability to address and overcome the challenges posed by disinformation.

This study invites reflection on disinformation by examining two key dimensions: the vehicle of dissemination (traditional media versus social media) and the role of creators (human versus AI). By considering these dimensions, we can gain a comprehensive understanding of how disinformation has evolved and how science can address these challenges. Traditional media, constrained by gatekeeping mechanisms, results in a slower and more limited spread of disinformation. In contrast, social media enables rapid, widespread distribution, driven by algorithms that often amplify sensational or emotionally charged content, bypassing traditional regulation.

The creators of disinformation also warrant careful consideration. Historically, human-generated disinformation involved traditional propaganda, political manipulation, and targeted misinformation. However, the advent of AI has introduced a new level of sophistication in disinformation creation. Generative AI can produce highly realistic and persuasive false content, such as deepfakes, on a large scale, making it a powerful tool for disinformation campaigns. Both humans and AI can now generate

and spread disinformation more effectively than ever before through social media. Reflecting on these dimensions aims to inform policy decisions and legislative actions to effectively mitigate the risks associated with both traditional and AI-generated disinformation in the digital age. This approach encourages thoughtful discussion and comprehensive strategies to address the complex challenges posed by modern disinformation.

References

- Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2019). Protecting world leaders against deep fakes. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 38–45). IEEE.
https://openaccess.thecvf.com/content_CVPRW_2019/html/Media_Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.html
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236.
<https://doi.org/10.1257/jep.31.2.211>
- Assemblée nationale du Québec. (2022, October 31). *Matière à réflexion—La désinformation : risques pour la démocratie et pistes de solution*. <https://premierelecture.bibliotheque.assnat.qc.ca/2022/10/31/matiere-a-reflexion-la-desinformation-risques-pour-la-democratie-et-pistes-de-solution/>
- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139. <https://doi.org/10.1177/0267323118760317>
- Boykoff, M. T., & Boykoff, J. M. (2004). Balance as bias: Global warming and the US prestige press. *Global Environmental Change*, 14(2), 125–136. <https://doi.org/10.1016/j.gloenvcha.2003.10.001>
- Bradshaw, S., & Howard, P. N. (2018). The global organization of social media disinformation campaigns. *Journal of International Affairs*, 71(1.5), 23–32. <https://jia.sipa.columbia.edu/news/global-organization-social-media-disinformation-campaigns>
- Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108(10), 1378–1384. <https://doi.org/10.2105/AJPH.2018.304567>
- Bulger, M., & Davison, P. (2018). The promises, challenges, and futures of media literacy. *Journal of Media Literacy Education*, 10(1), 1–21. <https://doi.org/10.23860/JMLE-2018-10-1-1>
- Chesney, R., & Citron, D. K. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147–155. <https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war>
- Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4. <https://doi.org/10.1002/prs.2015.145052010082>
- Daniels, J. (2018). The algorithmic rise of the “alt-right.” *Contexts*, 17(1), 60–65. <https://doi.org/10.1177/1536504218766547>
- Douglas, H. (2015). Politics and science: Untangling values, ideologies, and reasons. *The ANNALS of the American Academy of Political and Social Science*, 658(1), 296–306. <https://doi.org/10.1177/0002716214557237>
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320. <https://doi.org/10.1093/poq/nfw006>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American Sociological Review*, 77(2), 167–187. <https://doi.org/10.1177/0003122412438225>
- Graves, L., & Cherubini, F. (2016). *The rise of fact-checking sites in Europe*. Reuters Institute for the Study of Journalism. <https://doi.org/10.60625/risi-tdn4-p140>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), Article eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Hansson, S. O. (2017). Science denial as a form of pseudoscience. *Studies in History and Philosophy of Science Part A*, 63, 39–47. <https://doi.org/10.1016/j.shpsa.2017.05.002>
- Hassan, N., Arslan, F., Li, C., & Tremayne, M. (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* (pp. 1803–1812). Association for Computing Machinery. <https://doi.org/10.1145/3097983.3098131>
- Holt, T. J., Freilich, J. D., & Chermak, S. M. (2022). Examining the online expression of ideology among far-right extremist forum users. *Terrorism and Political Violence*, 34(2), 364–384. <https://doi.org/10.1080/09546553.2019.1701446>
- House of Commons, Standing Committee on Access to Information, Privacy and Ethics. (2018). *Democracy under threat: Risks and solutions in the era of disinformation and data monopoly* (Report 17, 42nd Parliament, 1st Session). <https://www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-17/>
- Howard, P. N., & Kollanyi, B. (2016). *Bots, #StrongerIn, and #Brexit: Computational propaganda during the UK–EU referendum*. SSRN. <https://doi.org/10.2139/ssrn.2798311>
- Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, 51(8), 56–59. <https://doi.org/10.1109/MC.2018.3191268>

- Larson, H. J. (2018). The state of vaccine confidence. *The Lancet*, 392(10161), 2244–2246. [https://doi.org/10.1016/S0140-6736\(18\)32608-4](https://doi.org/10.1016/S0140-6736(18)32608-4)
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- McCright, A. M., & Dunlap, R. E. (2010). Anti-reflexivity. *Theory, Culture & Society*, 27(2–3), 100–133. <https://doi.org/10.1177/0263276409356001>
- National Academies of Sciences, Engineering, and Medicine. (2016). *Science literacy: Concepts, contexts, and consequences*. National Academies Press. <https://doi.org/10.17226/23595>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Oreskes, N., & Conway, E. M. (2011). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Press. <https://www.bloomsbury.com/us/merchants-of-doubt-9781608193943/>
- Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7), 2521–2526. <https://doi.org/10.1073/pnas.1806781116>
- Persily, N. (2017). The 2016 U.S. election: Can democracy survive the internet? *Journal of Democracy*, 28(2), 63–76. <https://doi.org/10.1353/jod.2017.0025>
- Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux. <https://us.macmillan.com/books/9780374718657/activemeasures/>
- Romerstein, H. (2001). Disinformation as a KGB weapon in the Cold War. *Journal of Intelligence History*, 1(1), 54–67. <https://doi.org/10.1080/16161262.2001.10555046>
- Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, 116(16), 7662–7669. <https://doi.org/10.1073/pnas.1805871115>
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9, Article 4787. <https://doi.org/10.1038/s41467-018-06930-7>
- Singer, P. W., & Brooking, E. T. (2018). *LikeWar: The weaponization of social media*. Houghton Mifflin Harcourt. <https://www.harpercollins.com/products/likewar-p-w-singerremerson-t-brooking?variant=39936274858018>
- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press. <https://doi.org/10.1515/9781400884711>
- Tambini, D. (2017). *Fake news: Public policy responses* (Media Policy Brief 20). London School of Economics and Political Science. <https://eprints.lse.ac.uk/73015/>
- Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2018). Defining “fake news”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153. <https://doi.org/10.1080/21670811.2017.1360143>
- Tully, M., Vraga, E. K., & Bode, L. (2020). Designing and testing news literacy messages for social media. *Mass Communication and Society*, 23(1), 22–46. <https://doi.org/10.1080/15205436.2019.1604970>
- van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2), Article 1600008. <https://doi.org/10.1002/gch2.201600008>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Wald, A. (1943). *A method of estimating plane vulnerability based on damage of survivors* (Statistical Research Group memorandum). Columbia University. (Reprinted by the Center for Naval Analyses). <https://apps.dtic.mil/sti/tr/pdf/ADA091073.pdf>
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe. <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>
- Zhou, X., & Zafarani, R. (2018). *Fake news: A survey of research, detection methods, and opportunities*. arXiv. <https://doi.org/10.48550/arXiv.1812.00315>