KISIP 2024
Konferensi Ilmu Sosial dan Ilmu Politik

Research Paper

Filling the Loopholes in Media Literacy: A Rapid Evidence Assessment on Effective Countermeasures Against Election Disinformation

Panel 2
Technological Challenges and Innovations in Combating Disinformation

Elsa Hestriana

University College London (UCL), London, United Kingdom

✉ elsa.hestriana@protonmail.com

Elsa holds an MPA in Digital Technologies and Policy from University College London (UCL). Prior to UCL, she completed her Bachelor's degree in International Relations at the University of Indonesia. Her research focus includes tech policy, privacy, data protection, and online disinformation. Elsa's practical insights have been featured in renowned media outlets like The Jakarta Post.

This working paper is circulated for discussion and comment purposes. It has not been peer-reviewed or been subject to internal review by SAIL, CSIS, or Google. The views expressed here are solely those of the author(s) and do not represent an official position of SAIL, CSIS, Google, or any other organization. Feedback is welcome as the author(s) continue to develop these ideas for future formal publication. Please contact the author(s) directly with any comments or questions.

Editor: Dandy Rafitrandi

Abstracts

Online disinformation has put democracies worldwide to the test, disrupting elections across the globe. Today, advanced technologies such as AI, deepfakes, and social media bots have further exacerbated the difficulty of discerning truth from fiction. From media literacy to platform regulation, various interventions have been proposed and analyzed through academic research. In particular, media literacy has been a widely recommended method suggested by academics. However, even the most favored strategy comes with a trade-off and cannot be the sole solution. Therefore, this REA attempts to look for effective strategies beyond media literacy to counter disinformation, particularly in the context of elections. This research aims to help policymakers weigh the pros and cons of different interventions and make well-informed decisions on effective strategies for combating election disinformation. Using the Rapid Evidence Assessment (REA) method, this research synthesizes various literature on a wide range of strategies deployed to counter election disinformation. To meet the research objective, the search strings employed included terms addressing disinformation, digital technologies that enable it, strategies, interventions, and the political or electoral context. Out of 2,086 documents initially retrieved, 27 academic articles met the criteria to be included for synthesis. The synthesis concluded that there is no silver bullet in combating disinformation and every type of intervention has its benefits as well as side effects. While media literacy remains a fundamental strategy, it also carries a significant caveat: it does not only increase one's skepticism towards fake news but also real news. Although there is no one-size-fits-all solution in disinformation war, each type of intervention serves as a piece of Swiss cheese defense, with every layer complementing the other and where a combination of multiple interventions is highly suggested. This research provides evidence-based insights to help policymakers conduct impact assessments against potential policy interventions and ensure that election disinformation can be effectively tackled with minimum unintended consequences and without undermining democratic values.

Keywords: *Disinformation, Election, Media Literacy, Digital Technologies*

Introduction

Disinformation has plagued elections across the globe – from the 2016 US Presidential Election (Grinberg et al. 2019, 374-378; Thompson 2020, 182-184) to 2019 Indonesian Presidential Election (Rumata and Nugraha 2020, 352; Duile and Tamma 2021, 81). Moreover, the advancement of digital technologies, like DeepFake and other AI-generated fake content, has made it easier to generate convincing disinformation that is harder to distinguish from real information (Kertysova 2018, 63-68). This development poses tougher challenges for policymakers to come up with effective strategies against disinformation. Without swift actions, disinformation could undermine the democratic values of elections and erode public trust in policymakers.

In the quest for effective strategies against election disinformation, researchers have continued to analyze different intervention types, such as media literacy, fact-checking, nudging, and platform regulation. Media literacy has been a particularly popular method recommended by academics, journalists, and other experts as a primary means to combat disinformation (Medeiros and Singh 2020, 288; Kahne and Bowyer 2017, 15). However, not all experts agree that media literacy is the most effective strategy to combat fake news. Critics have argued that media literacy alone does not address the "psychological and social-identity-based factors" that often influence one's ability to distinguish between truth and fake news (Benkler, Faris, and Roberts 2018, 378). Media literacy is deemed insufficient to prevent the spread of false news if it solely focuses on the user's technical abilities to fact check and not their critical thinking skills or ability to understand the nuances of ideologies, political economy, and power (Banaji and Bhat 2019, 27).

Moreover, this REA also found evidence that media literacy may come with a significant trade-off: enhancing critical thinking skills through media literacy can result in not only increased skepticism towards fake news but also factual news (Guess et al. 2020, 15541; Moore and Hancock 2022, 7). This finding will be discussed further in a later part of the article.

Therefore, this REA attempts to look for effective strategies beyond media literacy to counter disinformation, particularly in the context of elections. This research aims to help policymakers weigh the pros and cons of different interventions and make well-informed decisions on effective strategies for combating election disinformation, based on rigorous evidence. This REA attempts to do so by assessing evidence on strategies that have been employed to counter disinformation. Ultimately, this research provides evidence-based insights to help policymakers conduct Impact Assessments (IAs) of potential policy interventions and ensure that election disinformation can be effectively tackled with minimum unintended consequences and without undermining democratic values.

Methodology

This REA followed a Systematic Review Protocol to ensure the rigor and reliability of the REA. By including carefully selected search databases, relevant keywords and subject areas, as well as specific inclusion criteria, the protocol was designed to capture the most relevant literature to answer the research questions. Moreover, this section serves as a transparent record of the search methodology. The protocol outlined below describes how the search for evidence on this REA was conducted.

Research Questions

With the aim to assist policymakers in searching for effective strategies to combat election mis- and disinformation, this REA was guided by research questions below:

Main question: What is the effective strategy to counter election disinformation?

The main question is supported by a set of relevant sub-questions that will provide a comprehensive understanding of the current state of disinformation counter strategies.

- Sub-question 1: What are the past and/or existing strategies to counter election-related disinformation?
- Sub-question 2: What did or did not work from these strategies?
- Sub-question 3: What are the challenges to creating effective counter disinformation strategies?

Assessment Scope

To address the policy problem and effectively answer the formulated research questions, the scope of this REA follows criteria below:

- This REA focuses on disinformation, which is defined as "false information that is *purposely* spread to deceive people" (Lazer et al. 2018, 1094), emphasizing on the intention. While the term is not identical with misinformation,[2] the two overlap with each other in definition and, thus, the latter was used as a synonym in the Search Strategy to reach more potentially relevant evidence.
- This REA looks at the effectiveness of interventions against disinformation. Therefore, this REA focuses on literature that analyses and evaluates disinformation strategies, including the outcomes and trade-offs.
- This REA prioritizes literature discussing disinformation within the context of election. However, this REA also acknowledges that some interventions have a universal nature that can be implemented across different contexts.
- The discussion on this REA focuses on disinformation enabled by digital technologies, like the use of bots and social media to spread fake news, as well as AI-generated fake content.

[2] Lazer et al. (2018) define misinformation as simply "false or misleading information."

This scope is also reflected in the keywords and terms formulation to search for relevant evidence and the inclusion criteria for the evidence assessed (see the following section).

Search Strategy

This REA used three different platforms to search for evidence documents:

- Scopus
- Web of Science
- ProQuest

The three platforms were carefully selected based on four main considerations. First, their Advanced Search features use the Boolean Operators, allowing more focused, specific, and relevant results. Secondly, they offer extensive reach of documents across different disciplines, particularly around Social Sciences and Computer Sciences, which are the focus of this search. Thirdly, they record comprehensive details of documents they index, which allows more effective filtering and analysis of search results. Finally, the three platforms can be widely accessible across different academic and research institutions, ensuring the replicability of this REA. Table 1 below breaks down the combination and formulation of keywords and terms to search for relevant evidence.

Table 1: Keywords Combinations and Formulation

Component 1		Component 2		Component 3		Component 4
Online OR Social media OR Artificial Intelligence OR AI OR Deepfake OR Bot*	AND	Disinformation OR Misinformation OR Fake News	AND	Strateg* OR Policy OR Intervention*	AND	Election OR Political OR Democrac*

Combining different keywords and synonyms for each component ensured that the REA would capture a wide yet specific range of literature relevant to the policy topic and research questions. In Component 1, for example, this REA combined keywords and terms that represent digital technologies known to be used to produce and spread disinformation, as aligned with the REA's focus on disinformation enabled by digital technologies.

Moreover, the different synonyms also helped this REA cover different possible vocabularies for discussing the issue. In Component 2, for example, this REA employed synonyms commonly used to describe misleading or false information (Lazer et al. 2018, 1094). While this REA focuses on disinformation, the author acknowledges that "disinformation" and "misinformation" overlap with each other in definition and, thus, the term "misinformation" is included in the search string to reach more potentially relevant evidence.

With the combination of keywords above, here is the search string deployed on the selected search databases:

> ((online OR "social media" OR "artificial intelligence" OR AI OR deepfake OR bot*) AND (disinformation OR misinformation OR "fake news") AND (strateg* OR policy OR intervention*) AND (election* OR political OR democrac*))

Inclusion Criteria

To determine the relevant core literature to be synthesized, this REA applied a set of inclusion criteria, which is divided into two types: technical and substantive.

Technical criteria

- Literature published between 2016-2023. The year 2016 was determined to be the starting point considering the intensity of discussions around fake news that increased after the 2016 US Presidential Election (Rahman and Tang 2022, 155).
- Literature within the subject of Social Science and its branches. Thus, the search also included subject areas of Communications, Government Law, Public Administration, Psychology, Sociology, and Education. Social Sciences as a focus area is determined considering the REA's scope in the context of elections. Additionally, this REA also included literature within the subject of Computer Science and its branches, such as Telecommunication and Information Technology. Including studies from the perspective of Computer Sciences in the literature search is crucial given the REA's scope in disinformation enabled by digital technologies.
- Given the academic nature of this REA, the evidence included for synthesis was limited to academic literature only. This includes journal articles, books, and book chapters.
- This REA limited its review to literature in English only. The author acknowledges this as part of the REA's limitations.

Substantive criteria

- All studies discussing disinformation in the context of elections. Therefore, the search excluded all literature discussing the issue in the context of the COVID-19 pandemic, which overlaps with the 2020 US Presidential Election (Chen et al. 2021, 1-3).

- Studies that analyze past and/or existing strategies to counter disinformation. This REA also included literature analyzing an alternative approach to past or existing strategies. This criterion is not to be mistaken with strategies to spread mis- and disinformation.
- Given its focus on election-related disinformation, the search only included research that involves voters as the research subject. Therefore, it excluded studies on media literacy for primary school pupils, for example.

Screening and Study Prioritization

The literature resulting from the search followed two primary steps. First, the search results were filtered using the criteria laid out above. The first round of filtering used the technical criteria. Then, the selected literatures were further sifted according to the substantive criteria. This was done by assessing the title, abstract, and keywords of each literature. Next, a further assessment took place by screening the full texts to determine the core papers that would be synthesized. Second, the selected literatures were coded according to the intervention category that the article discusses to map the main strategies for countering disinformation.

Results

The combination of keywords laid out in the previous resulted in a total of 2,086 documents from the three different platforms. Then, 1,797 documents were eliminated once the technical criteria were applied and duplicates were removed along with those which full texts cannot be fully retrieved. This left us with 289 titles and abstracts to screen based on the substantive criteria. After the titles and abstracts were screened, 57 full texts were assessed more thoroughly using the same substantive criteria to identify the core papers. A series of filtering, sifting, and screening ultimately brought us to 27 key literatures to synthesize. Figure 3.1 shows the PRISMA flow diagram summarizing the process of the evidence screening and identification process.

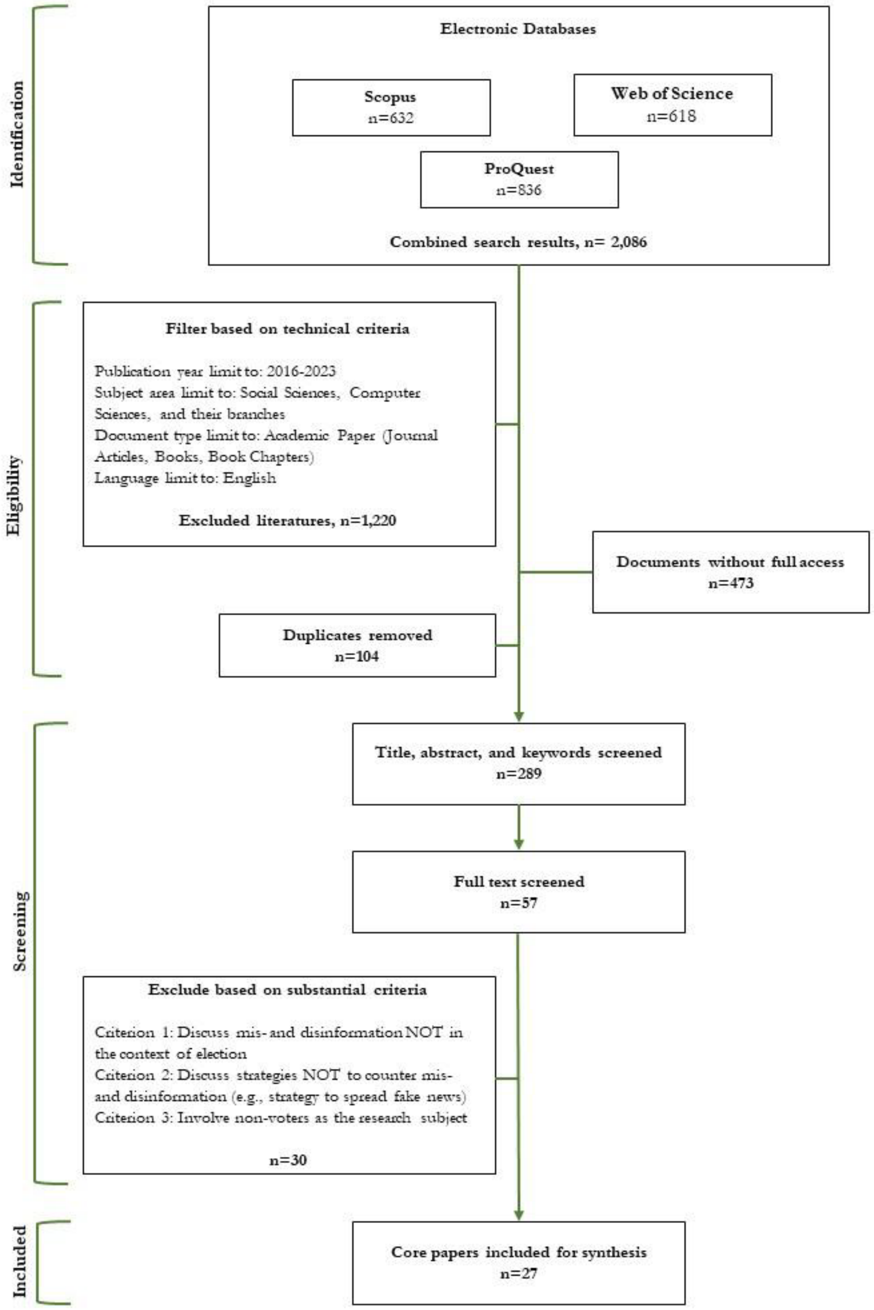
Identification
Electronic Databases
Scopus
n=632
Web of Science
n=618
ProQuest
n=836
Combined search results, n= 2,086
Eligibility
Filter based on technical criteria
Publication year limit to: 2016-2023
Subject area limit to: Social Sciences, Computer Sciences, and their branches
Document type limit to: Academic Paper (Journal Articles, Books, Book Chapters)
Language limit to: English
Excluded literatures, n=1,220
Documents without full access
n=473
Duplicates removed
n=104
Screening
Title, abstract, and keywords screened
n=289
Full text screened
n=57
Exclude based on substantial criteria
Criterion 1: Discuss mis- and disinformation NOT in the context of election
Criterion 2: Discuss strategies NOT to counter mis- and disinformation (e.g., strategy to spread fake news)
Criterion 3: Involve non-voters as the research subject
n=30
Included
Core papers included for synthesis
n=27

Figure 1: PRISMA flow diagram demonstrating the REA's process of evidence screening and identification process.

To help answer the research questions, the 27 core papers were coded based on the intervention categories discussed in the article. This REA noted that individual literature can contain discussion about multiple interventions.

As shown in Figure 3.2, most of the literature discusses law or regulatory intervention as a strategy to combat election disinformation (37%). The types of regulatory intervention discussed across the papers are diverse – from social media platform regulation to international law against state-sponsored disinformation. Then, fact-checking, both carried out by social media companies and non-profit organizations, comes second (30%).

Media literacy remains a popular method of countering disinformation (18.5%). This intervention type is discussed within different approaches across the selected papers – from digital media literacy designed for older adults to the gamification of media literacy. On top of those, this REA also gathered evidence on nudging (18.5%), access blocking (7%), and filtering (4%).

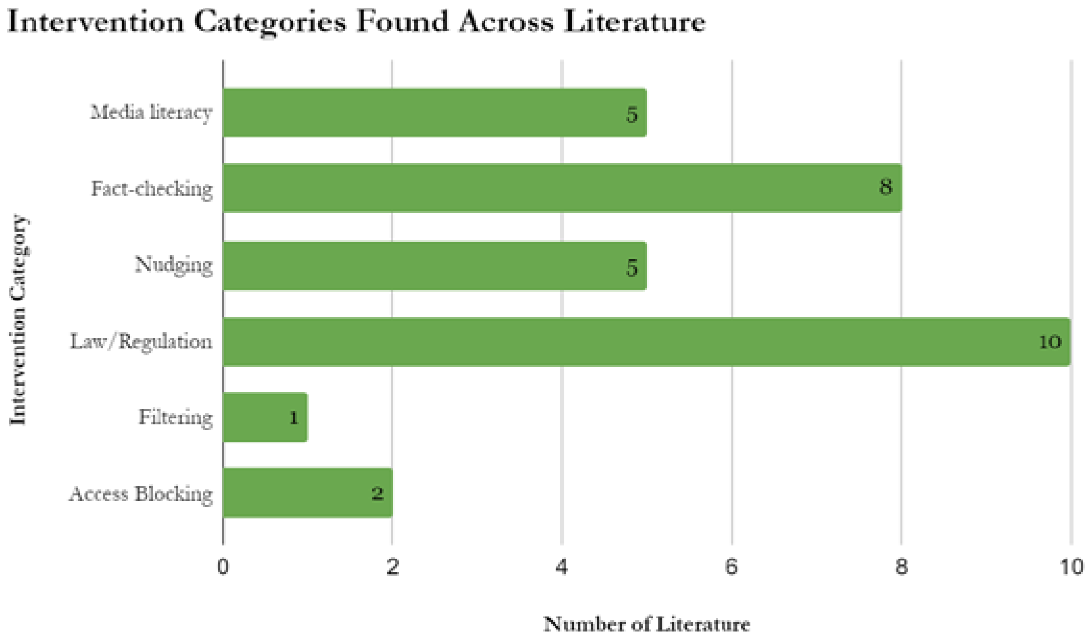

Figure 2: A bar chart summarizing the distribution across the core papers based on intervention category

Findings

This section elaborates the findings from the synthesis process.

Revisiting Media Literacy

Evidence synthesized from different research suggests that media literacy is proven to have successfully improved people's ability to identify fake news. This success has been demonstrated through different experiments and research methods. Among others, Kahne and Bowyer (2017) surveyed young voters to investigate whether political knowledge and media literacy improve youth's ability to accurately judge false information. According to their research, a young voter's political knowledge does not enhance their capacity to accurately identify misinformation. The youth are found to be more likely to label information as inaccurate when it conflicts with their political beliefs. In contrast, however, youth who had media literacy learning opportunities showed more accurate judgment of truth and misinformation, even when both information aligned with their political beliefs. This quality is also known as critical loyalty or the ability to adopt a critical stance when evaluating an argument regardless of their partisan preference (Kahne and Bowyer 2017, 20-27).

The effectiveness of media literacy in countering disinformation is also found in research done by Moore and Hancock (2022) which experimented the intervention with older adults, age 60 and above. Their experiment found that digital media literacy significantly improved older adults' ability to accurately discern fake news and truth, from 64% to 85% (Moore and Hancock 2022, 4-8). Similarly, surveys by Guess et al. (2020) among voters in the United States (US) and India showed that respondents' ability to differentiate between mainstream and untrustworthy news increased by more than 26% in the US sample and 17% in the Indian sample. Moreover, this improvement is independent of whether the claims made in the headlines align with the respondents' political inclinations (Guess et al. 2020, 15542).

Additionally, Roozenbeek and Linden (2019) took a creative step by gamifying media literacy as a form of psychological intervention against fake news. The game preemptively exposing, warning, and familiarizing the participants with the strategies used in the production of fake news, employing an inoculation metaphor to create cognitive immunity when individuals are exposed to misinformation (Roozenbeek and Linden 2019, 2-5). They, too, concluded the success of their media literacy model in improving people's ability to spot and resist misinformation, regardless of age, education, political views, and cognitive style.

Findings by Kahne and Bowyer (2017), Guess et al. (2020), and Roozenbeek and Linden (2019) rebut the criticisms that media literacy does not address the psychological, social, and identity factors that influence one's ability accurately judge fake news (Benkler, Faris, and Roberts 2018, 378).

Despite proven effectiveness of media literacy, it comes with significant caveats. Guess et al. (2020) emphasize that "increased skepticism of false news headlines may come at the expense of decreased belief in mainstream news headlines." In other words, media literacy does not only increase people's skepticism towards fake news but also real news. Moore and Hancock (2022) also point out the same concern. Their findings indicated that individuals were more adept at accurately identifying false news than true news (Moore and Hancock 2022, 7).

Furthermore, as an educational effort, the impact of media literacy would decay over time if the individual does not continuously exercise what they learn. Consequently, some researchers have suggested that educators, social media companies, and journalists should reinforce media literacy lessons on a recurring basis (Guess et al. 2020, 15542) and conduct them at the local grassroots level. Discussing the circulation of misinformation on the messenger app WhatsApp, Medeiros and Singh (2020) argue that platforms should consider investing in locally grounded media literacy initiatives. This is to ensure that the learning can be tailored to engage with the unique regional tensions and demographics (Medeiros and Singh 2020, 295).

The caveats of media literacy indicate that media literacy ultimately cannot be the sole weapon in the disinformation war. Even the strong supporters of media literacy have called for media literacy to be combined with other types of intervention (Medeiros and Singh 2020, 289-290). The following subsections will discuss other types of intervention suggested by researchers, that may help fill the loopholes in media literacy.

Fact-Checking

Fact-checking emerged as one of the most commonly employed interventions against disinformation. It can be and has been done at many levels and by different actors – individuals, social media platforms, media outlets, non-profit organizations, and government (Gupta et al. 2022, 78277-80; López-García, Vizoso, and Pérez-Seijo 2019, 625; Bak-Coleman et al. 2022, 1374). For example, the UK's news outlet BBC started BBC Reality Check in 2015 and France's Le Monde began actively running Les Décodeurs in 2012 as their fact-checking spaces (López-García, Vizoso, and Pérez-Seijo 2019, 625). At the platform level, Facebook, Google, and Instagram have deployed efforts such as third-party checking and the implementation of misrepresentations, impersonation, and spam detection systems (Gupta et al. 2022, 78279). Meanwhile, at the government level, Singapore released Factually, a state-operated fact-checking website as early as 2012, while its neighboring Southeast Asian countries Malaysia launched Sebenarnya.my in 2017 and Thailand announced its Anti-Fake News Center in late 2019 (Schuldt 2021, 341).

While fact-checking tools and sites can help users verify information, their existing flaws could significantly undermine the effectiveness of fact-checking as an intervention against disinformation. First, the speed and volume of fake news production overpower the ability of human reviewers (Pierri and Ceri 2019, 20; Shao et al. 2018, 75331). Moreover, even when automated technology is involved, the sharing of fact-checking results typically lags the spread of fake news by about one day. In addition, the fake news themselves are often more popular than their corresponding debunking (Shao et al. 2018, 75338).

Secondly, AI-powered fact-checking tools, such as the use of Natural Language Processing (NLP), have been proposed and utilized to address the first concern. However, automated fact-checking has been criticized for its tendency to bias in its claims (Das et al. 2023, 6; Gupta et al. 2022, 78283). Notably, human's sensitivity in understanding nuances and different cultural and social contexts are yet to be replaced by machines. Gupta et al. (2022) points out that one of the technical challenges of AI-powered fake news detection lies in cultural diversity, where "what is constructed as satire in one region of the world may be considered offensive in another and fake news in another" (Gupta et al. 2022, 78276).

Thirdly, a lot of online disinformation is spread through private platforms such as WhatsApp (Medeiros and Singh 2020, 276-8). Consequently, real-time fact-checking would require privacy intrusions, which do not only violate the users' privacy rights but would also create potential user backlash (Gupta et al. 2022, 78279). If backed by the government, this practice could increase the concern that government-led fact-checking may turn into a propaganda tool, positioning the government as the arbiter of truths (Schuldt 2021, 357).

Lastly, and more importantly, the success of fact-checking as an intervention against disinformation requires the users themselves to take the initiative and actively utilize the available fact-checking tools or visit fact-checking sites (Gupta et al. 2022, 78278). In other words, the act of fact-checking at the user level relies on the user's understanding of media literacy; their ability to use the right tools. This further emphasizes the significance of media literacy as the fundamental intervention against disinformation.

Regulation

Across different academic papers synthesized for this REA, one key agreement persists: that regulating information to combat fake news involves complex and multi-layered issues beyond just information. The regulation(s) should also uphold transparency, data protection, and ensure the right to free expression (Nenadic 2019, 1-15; Helberger 2020, 848-50; Marsden, Meyer, and Brown 2020, 2-18; Craufurd-Smith 2019, 58-63).

In this context, transparency refers to the clarity for users about where the information comes from and why they receive certain content when an algorithmic system is applied. It should be demanded from media companies, including social media, and the political actors involved in electoral campaigns (Craufurd-Smith 2019, 62; Nenadic 2019, 10; Marsden, Meyer, and Brown 2020, 8). For example, the EU Code of Practice on Disinformation was created to be a self-regulatory tool to encourage online platforms to be more transparent about political advertising and prevent the spread of disinformation through automation. Meanwhile, for political campaigners, the European approach requires political advertisers to be transparent about their spending on political advertising and who funds them (Nenadic 2019, 6-14).

Transparency is also closely linked to privacy and data protection as online political campaigns are largely driven by users' personal data. The Cambridge Analytica scandal in 2018 has significantly put a brighter spotlight on this issue.[3] Hence, social scientists have argued that protection against personal data abuse is crucial in ensuring fair elections, especially with the increasing use of AI-enabled political advertising (Nenadic 2019, 6; Marsden, Meyer, and Brown 2020, 16; Shattock 2019, 220-21).

However, the biggest criticism towards anti-fake news regulations remains about its potential impact on freedom of expression. Many researchers argue that restrictions on false speech can affect a much broader range of speech and

[3] The Guardian reported that the data analytics firm gathered and used personal data of Facebook users without authorization to profile individual US voters and target them with personalized political advertisements. Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," *The Guardian*, March 18, 2018, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

expression (Medeiros and Singh 2020, 277-85; Craufurd-Smith 2019, 81; Shattock 2019, 212).

This issue becomes increasingly complex with the involvement of AI-powered or automated content policing employed by platform companies (Marsden, Meyer, and Brown 2020, 6-8). Many social media companies began implementing such technology to avoid sanctions from strict laws that require them to quickly take down content containing misinformation. For example, Germany introduced NetzDG in 2017 which obliges platform companies to remove inappropriate content within 24 hours (López-García, Vizoso, and Pérez-Seijo 2019, 624). France has also adopted its version of the NetzDG, requiring platforms to delete certain types of content within an hour (Helberger 2020, 844).

Applying AI in content moderation means that some content or event accounts can be removed without human intervention to review the judgment's accuracy. Social scientists strongly advise against this, suggesting a hybrid model that combines AI and human review to ensure value judgments (Marsden, Meyer, and Brown 2020, 16-17). Moreover, today's social media landscape is largely dominated by just a few companies, such as Meta and Alphabet, providing them with not only economic but also political power to police online expressions (Iosifidis and Andrews 2020, 222).

Thus, instead of handing over the power to platforms through self-regulation, Marsden, Meyer, and Brown (2020) argue that content regulation should not be the responsibility of only one party. Instead, it should be co-regulated where platform companies collectively regulate their users and co-regulation must be approved and monitored by state regulators (Marsden, Meyer, and Brown 2020, 9-10). Additionally, Işik, Bildik, and Molla (2022) propose that if a malicious act of disinformation is backed by foreign actors, as was the case during the 2016 US presidential election, the issue should be brought to the international level using international law (Işik, Bildik, and Molla 2022, 106-20).

Nudging

Nudging refers to prompting or warning users on social media that the information they see may be misleading. In this sense, nudging can also be in the form of flagging or labelling a content as misleading or false. Generally, nudging aims to reduce the likelihood of users sharing false news (Pennycook and Rand 2022, 153-162). Different experiments have indicated that adding nudges or general warnings help shift users' perceived accuracy towards information they see on social media (Bhuiyan et al. 2021, 22; Pennycook and Rand 2022, 162; Thornhill et al. 2019, 7-8; Clayton et al. 2019, 1073-91). In fact, a simulation by Bak-Coleman et al (2022) found that nudges resulted in reductions in misinformation sharing and engagement (Bak-Coleman et al. 2022, 1375).

Moreover, nudging can also act as a complement to fact-checking. In cases where misleading claims contain partly true information and require more time to assess, nudging can be implemented as a warning that the content may be false (Bak-Coleman et al. 2022, 1374-5). In other words, nudges offer a faster and more efficient alternative to fact-checking and can be used to help trigger skepticism towards potentially misleading news (Pennycook and Rand 2022, 153; Thornhill et al. 2019, 8).

However, nudging is found to carry a similar trade-off to media literacy: general warnings do not only appear to decrease belief in fake news but also factual news, which can be a “potential hazard” (Clayton et al. 2019, 1091-2).

Access Blocking

This REA found two academic papers discussing policy interventions against fake news that fall under the category of access blocking: internet shutdown amid the threat of post-election riots and account banning by social media. Both interventions generally aim to prevent or stop the circulation of misinformation. Generally, access blocking raises a serious threat to freedom of expression. In particular, internet shutdown is heavily criticized as it seriously undermines democratic values and is deemed not the proper nor constitutional way to combat fake news (Rahman and Tang 2022, 151-3). On the other hand, simulations by Bak-Coleman et al. (2022) revealed that account banning successfully reduced total engagement with misinformation by 30% (Bak-Coleman et al. 2022, 1376). However, account banning may not completely halt the spread of disinformation as current practices tend to concentrate on accounts with a large number of followers, leaving smaller accounts free to continue disseminating false information (Bak-Coleman et al. 2022, 1375-6).

Filtering

Filtering of misinformation has not been widely discussed compared to other types of interventions like media literacy and fact-checking. Using the technical and substantive criteria described in the previous section, this REA includes one academic research exploring the filtering method. The study by Dave et al. (2022) proposes a new mechanism that implements the GNE (General Nash Equilibrium) to efficiently filter misleading information on social media. This mechanism is designed to incentivize the application of misinformation filtering across social media platforms (Dave et al. 2022, 2633). However, the lack of sufficient academic papers on filtering poses a vital challenge in coming up with a rigorous synthesis on the intervention.

Conclusion

This REA is centered around the research question, “What is the effective strategy to counter election disinformation?” Based on the findings synthesized in the previous section, this REA concluded that there is no silver bullet in combating disinformation and every type of intervention has its benefits as well as side effects. Media literacy is indeed a fundamental strategy that lies as the foundation of many other types of interventions. This is evident from how the success of interventions such as fact-checking and nudging relies on the individual’s ability to navigate the information they receive and to utilize relevant tools. Moreover, the effectiveness and positive impact of media literacy have been largely evaluated and proven through different experiments and research methods. However, media literacy also comes with a risk. Multiple studies have discovered that media literacy potentially does not only increase one's skepticism towards fake news, but also real news. A similar risk is also found in nudging. In the long run, this could pose a serious hazard.

While there is no silver bullet in combating fake news, each type of intervention serves as a piece of Swiss cheese defense, with every layer complementing the other. As suggested by papers included in this REA, the ideal strategy to combat disinformation would be to combine multiple interventions and involve all stakeholders in the chain – from individual users to government – in the process (Thompson 2020, 182-8; Bak-Coleman 2022, 1376-77; López-García, Vizoso, and Pérez-Seijo 2019, 629).
In reaching this conclusion, we also addressed the sub-questions below:

1. **What are the past and/or existing strategies to counter election-related disinformation?**

   This REA identified at least six different intervention categories that have been analyzed and evaluated through academic research: media literacy, fact-checking, regulations, nudging, access blocking, and filtering.

2. **What did or did not work from these strategies?**

   The academic evidence synthesized in this REA shows that all six interventions demonstrated at least some extent of effectiveness in reducing or preventing the spread of disinformation. However, every intervention also has its flaws. For example, while AI helps speed up and automate fact-checking, the technology tends to be biased in its claims and lacks the human ability to understand nuances in different cultural contexts.

3. **What are the challenges to creating effective counter disinformation strategies?**

   The findings of this REA demonstrate that it has been a great challenge to come up with a balanced strategy that can effectively address disinformation issues without creating any countervailing risk, or even undermining democratic values. For example, some research expresses concern that overregulating content moderation against disinformation will eliminate public's right to free expression and aggressive fact-checking may intrude users' rights to privacy.

Recommendations for Policymakers

To make an informed and careful decision in combatting disinformation, this paper highly recommends conducting IAs that weigh the benefits and flaws of each intervention type. Such assessments should include a careful consideration of the rationale, cost and benefits analysis, and monitoring frameworks for each policy option.

The insights obtained from this REA can be highly valuable for policymakers in conducting IAs against possible policy interventions and determining appropriate strategies to combat election disinformation. It offers a more in-depth analysis of previous implementations and their impact, enabling policymakers to identify potential risks and obstacles, as well as obtain a better understanding of the consequences – both positive and negative, intended and unintended – of potential policy interventions.

Irrespective of the policy decisions taken, this REA emphasizes the importance of maintaining a balance between combating disinformation and safeguarding free and fair elections, preserving democratic values, and upholding individuals' rights. In doing so, policymakers must thoughtfully assess whether new policy interventions are necessary or if the answer lies in strengthening the enforcement of existing ones. For example, data protection laws like the EU General Data Protection Regulation (GDPR) can also act as a regulatory tool to prevent the exploitation of social media users' personal data for the purpose of micro targeted disinformation.

By following these insights, policymakers can navigate the challenges and protect the integrity of our elections while respecting individual rights and democratic principles. Whether through new interventions or reinforcing existing ones, the path to a more secure and democratic future begins with a measured and well-informed approach.

References

Bak-Coleman, JB, I Kennedy, M Wack, A Beers, JS Schafer, ES Spiro, K Starbird, and JD West. 'Combining Interventions to Reduce the Spread of Viral Misinformation'. *NATURE HUMAN BEHAVIOUR* 6, no. 10 (October 2022): 1372-+. https://doi.org/10.1038/s41562-022-01388-6.

Banaji, Shakuntala, and Ram Bhat. "WhatsApp Vigilantes: An Exploration of Citizen Reception and Circulation of WhatsApp Misinformation Linked to Mob Violence in India." *LSE Blogs*, 2019. https://eprints.lse.ac.uk/104316/1/Banaji_whatsapp_vigilantes_exploration_of_citizen_reception_published.pdf

Benkler, Yochai, Robert Faris, and Hal Roberts. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press, 2018. doi: 10.1093/oso/9780190923624.001.0001.

Bhuiyan, M.M., M. Horning, S.W. Lee, and T. Mitra. 'NudgeCred: Supporting News Credibility Assessment on Social Media through Nudges'. *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (2021). https://doi.org/10.1145/3479571.

Cadwalladr, Carole, and Emma Graham-Harrison. "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach." *The Guardian*, March 18, 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Chen, Emily, Herbert Chang, Ashwin Rao, Kristina Lerman, Geoffrey Cowan, and Emilio Ferrara. "COVID-19 misinformation and the 2020 US presidential election." *The Harvard Kennedy School Misinformation Review* (2021). doi: 10.37016/mr-2020-57.

Clayton, Katherine, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, et al. 'Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media'. *Political Behavior* 42, no. 4 (2020): 1073–95. https://doi.org/10.1007/s11109-019-09533-0.

Craufurd Smith, R. 'Fake News, French Law and Democratic Legitimacy: Lessons for the United Kingdom?' *Journal of Media Law* 11, no. 1 (2019): 52–81. https://doi.org/10.1080/17577632.2019.1679424.

Das, Anubrata, Houjiang Liu, Venelin Kovatchev, and Matthew Lease. 'The State of Human-Centered NLP Technology for Fact-Checking'. *Information Processing & Management* 60, no. 2 (2023): 103219. https://doi.org/10.1016/j.ipm.2022.103219.

Dave, A., I.V. Chremos, and A.A. Malikopoulos. 'Social Media and Misleading Information in a Democracy: A Mechanism Design Approach'. *IEEE Transactions on Automatic Control* 67, no. 5 (2022): 2633–39. https://doi.org/10.1109/TAC.2021.3087466.

Duile, Timo, and Sukri Tamma. "Political language and fake news: Some considerations from the 2019 election in Indonesia." *Indonesia and the Malay world* 49, no. 143 (2021): 82-105. https://doi.org/10.1080/13639811.2021.1862496.

Guess, Andrew M., Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 'A Digital Media Literacy Intervention Increases Discernment between Mainstream and False News in the United States and India'. *Proceedings of the National Academy of Sciences - PNAS* 117, no. 27 (2020): 15536–45. https://doi.org/10.1073/pnas.1920498117.

Gupta, A., N. Kumar, P. Prabhat, R. Gupta, S. Tanwar, G. Sharma, P.N. Bokoro, and R. Sharma. 'Combating Fake News: Stakeholder Interventions and Potential Solutions'. *IEEE Access* 10 (2022): 78268–89. https://doi.org/10.1109/ACCESS.2022.3193670.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. "Fake news on Twitter during the 2016 US presidential election." *Science* 363, no. 6425 (2019): 374-378. https://doi.org/10.1126/science.aau2706.

Helberger, N. 'The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power'. *Digital Journalism*, 2020, 842–54. https://doi.org/10.1080/21670811.2020.1773888.

Iosifidis, Petros, and Leighton Andrews. 'Regulating the Internet Intermediaries in a Post-Truth World: Beyond Media Policy?' *The International Communication Gazette* 82, no. 3 (2020): 211–30. https://doi.org/10.1177/1748048519828595.

Işik, Irem, Ömer F. Bildik, and Tayanç T. Molla. 'Securing Elections Through International Law: A Tool for Combatting Disinformation Operations?' *Journal of Strategic Security* 15, no. 4 (2022): 106–25. https://doi.org/10.5038/1944-0472.15.4.2033.

Kahne, J., and B. Bowyer. 'Educating for Democracy in a Partisan Age: Confronting the Challenges of Motivated Reasoning and Misinformation'. *American Educational Research Journal* 54, no. 1 (2017): 3–34. https://doi.org/10.3102/0002831216679817.

Kertysova, Katarina. "Artificial intelligence and disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered." *Security and Human Rights* 29, no. 1-4 (2018): 55-81. doi: https://doi.org/10.1163/18750230-02901005.

Lazer, David MJ, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger et al. "The science of fake news." *Science* 359, no. 6380 (2018): 1094-1096. https://doi.org/10.1126/science.aao2998

López-García, Xosé, Ángel Vizoso, and Sara Pérez-Seijo. 'Verification Initiatives in the Scenario of Misinformation. Actants for Integrated Plans with Multi-Level Strategies'. *Brazilian Journalism Research* 15, no. 3 (December 2019): 614–35. https://doi.org/10.25200/BJR.v15n3.2019.1215.

Marsden, C., T. Meyer, and I. Brown. 'Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?' *Computer Law and Security Review* 36 (2020). https://doi.org/10.1016/j.clsr.2019.105373.

Medeiros, B, and P Singh. Addressing Misinformation on Whatsapp in India through Intermediary Liability Policy, Platform Design Modification, And Media Literacy. *Journal Of Information Policy* 10 (2020): 276–98. https://doi.org/10.5325/jinfopoli.10.2020.0276.

Moore, Ryan C, and Jeffrey T Hancock. 'A Digital Media Literacy Intervention for Older Adults Improves Resilience to Fake News'. *Scientific Reports (Nature Publisher Group)* 12, no. 1 (2022). https://doi.org/10.1038/s41598-022-08437-0.

Nenadic, I. ‘Unpacking the “European Approach” to Tackling Challenges of Disinformation and Political Manipulation’. *Internet Policy Review* 8, no. 4 (2019). https://doi.org/10.14763/2019.4.1436.

Palomo, Bella, and Jon Sedano. ‘Cross-Media Alliances to Stop Disinformation: A Real Solution?’ *Media and Communication* 9, no. 1 (2021): 239–50. https://doi.org/10.17645/mac.v9i1.3535.

Pennycook, G, and DG Rand. ‘Nudging Social Media toward Accuracy’. *Annals of the American Academy of Political and Social Science* 700, no. 1 (March 2022): 152–64. https://doi.org/10.1177/00027162221092342.

Pierri, F, and S Ceri. ‘False News on Social Media: A Data-Driven Survey’. *SIGMOD RECORD* 48, no. 2 (June 2019): 18–32. https://doi.org/10.1145/3377330.3377334.

Rahman, R.A., and S.-M. Tang. ‘Fake News and Internet Shutdowns in Indonesia: Symptoms of Failure to Uphold Democracy’. *Constitutional Review* 8, no. 1 (2022): 151–83. https://doi.org/10.31078/consrev816.

Roozenbeek, J., and S. van der Linden. ‘Fake News Game Confers Psychological Resistance against Online Misinformation’. *Palgrave Communications* 5, no. 1 (2019). https://doi.org/10.1057/s41599-019-0279-9.

Rumata, Vience Mutiara, and Fajar Kuala Nugraha. "An analysis of fake narratives on social media during 2019 Indonesian presidential election." *Jurnal Komunikasi: Malaysian Journal of Communication* 36, no. 4 (2020): 351-368. https://doi.org/10.17576/JKMJC-2020-3604-22.

Schuldt, L. ‘Official Truths in a War on Fake News: Governmental Fact-Checking in Malaysia, Singapore, and Thailand’. *Journal of Current Southeast Asian Affairs* 40, no. 2 (2021): 340–71. https://doi.org/10.1177/18681034211008908.

Shao, C., P.-M. Hui, P. Cui, X. Jiang, and Y. Peng. ‘Tracking and Characterizing the Competition of Fact Checking and Misinformation: Case Studies’. *IEEE Access* 6 (2018): 75327–41. https://doi.org/10.1109/ACCESS.2018.2881037.

Shattock, E. ‘Fake News, Free Elections, and Free Expression: Balancing Fundamental Rights in Irish Policy Responses to Disinformation Online’. *Publicum* 5, no. 2 (2019): 201–31. https://doi.org/10.12957/publicum.2019.47210.

Thompson, Terry L. ‘No Silver Bullet: Fighting Russian Disinformation Requires Multiple Actions’. *Georgetown Journal of International Affairs* 21 (Fall 2020): 182–94. https://doi.org/10.1353/gia.2020.0033.

Thornhill, C., Q. Meeus, J. Peperkamp, and B. Berendt. ‘A Digital Nudge to Counter Confirmation Bias’. *Frontiers in Big Data* 2 (2019). https://doi.org/10.3389/fdata.2019.00011.