Misinformation Battle Revisited

Counter Strategies from Clinics to Artificial Intelligence

Fard, Amir Ebrahimi; Lingeswaran, Shajeeshan

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Misinformation Battle Revisited: Counter Strategies from Clinics to Artificial Intelligence

Amir Ebrahimi Fard
TU Delft
Delft, The Netherlands
a.ebrahimifard@tudelft.nl

Shajeeshan Lingeswaran
TU Delft
Delft, The Netherlands
shajeeshan.lingeswaran@outlook.com

ABSTRACT

The spread of misinformation is one of the severe challenges that societies have been dealing with for many years. However, the rapid growth of social media has accelerated the creation and circulation of such information and turned it into a potential threat to the main societal institutions such as peace and democracy. Although many of iconic figures, policymakers, business leaders and researchers have warned us of serious repercussions of misinformation, a clear course of action is not yet visible. To tackle such an issue, the preliminary step would be the evaluation of the as-is situation, which allows us to identify the deficiencies of existing solutions. This issue has been addressed in this study by a comprehensive analysis over decades of societal efforts against misinformation. In this analysis, quelling strategies from organisational and government perspectives are explained. Then they are investigated from efficacy level and governance mode. Our analyses show that, despite a seemingly suitable setting for confronting misinformation, there is a major shortcoming in governance mode of current quelling strategies.

1 INTRODUCTION

Although for many "fake news" is a phenomenon of the past couple of years, misinformation is as old as journalism itself. During the first world war, to convince China to join the Entente Powers, the Allied published fake news in the English-language press "North China Daily News" that the Germans are using their fallen soldiers to extract fat to make candles, human soap and boat dubbing. However, it is true that only after the US election in 2016, "fake news" made it on the political agenda of many countries. During the US election, teenagers in a small village in Macedonia made a small fortune by creating more than 100 pro-Trump websites and sharing them in Facebook with headlines such as "Pope Francis Shocks World, Endorses Donald Trump for President" and "FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide". These entries created over more than 1 million

shares, reaction and comments on Facebook. The rapid spreading of misinformation via online social networks (OSNs) such as Twitter or Facebook, was not only used during the US election but it is also increasingly used by nationalists and populists around the world. This is an alarming trend which can lead to severe threats to main institutions of societies such as peace and democracy [18, 26, 52].

The spread of false and unverified information has been studied by academia since the early twentieth century [28, 48]. From the beginning, curbing the misinformation was of utmost importance and was one of their topics of interest [43]. However, the mass spread of misinformation has not been systematically approached in societal level until almost 80 years ago with the emergence of rumour clinics [2]. Those ideas were followed up later by control centres [41] and artificial intelligence [55] based solutions. Despite tremendous efforts by years of research and practise, diffusion of misinformation not only has not shrunk but also escalated and turned into a high-priority issue in individual, organisational, and societal level.

In order to effectively quell misinformation, in addition to developing novel techniques, it is crucial to be aware of the existing strategies and built capabilities. It works as a bird-eye view on the quelling strategies and allows us to understand what aspects of misinformation diffusion has been targeted extensively and what aspects are highly neglected. It also prevents us from reinventing the wheel and proposing the ideas that have already been exercised.

In this study, we exercise the same approach by comprehensive analysis over the quelling strategies in three steps. First, we review quelling strategies from organisational and state point of view. Then we analyse the lasting capability of the strategies. For this part, we use the epidemic control framework to understand which strategies have temporary and which have more lasting effects on their corresponding actors. Finally, we analyse the governance mode of each strategy. This allows us to understand to what extent the enforcement of strategies is guaranteed.

This paper is organised as follows. Section 2 gives an overview of the dynamics of misinformation from emergence to diffusion. Section 3 elaborates on quelling strategies against misinformation. In Section 4 two theoretical frameworks which are used to evaluate strategies from two perspectives of efficacy and execution are introduced. In Section 5 we report the evaluation of the quelling strategies, and in Section 6 we conclude this research by discussing the results and giving some recommendations.

2 MISINFORMATION DYNAMICS

The notion of misinformation is referred to any piece of information which is wrong or incorrect. Misinformation is not a newly emerged phenomenon; however, the rapid growth of social media

turned it to a problem with severe consequences. Misinformation is created either deliberately as part of an information operation or emerged when people share their opinions and comments during a conversation.

The dynamics of misinformation including creation and dissemination are captured by misinformation machine model [46]. This model introduces five elements as essential components of misinformation dynamics: publishers, authors, articles, audience, and rumours. Publishers run content distribution platforms. They vary from highly professional ones with codes of conduct, style guides, and journalistic guidelines to the ones without any guideline or standard. Authors are content providers of the platforms. Depending on the quality of the articles, they appear on different platforms. The audience is the target market of publishers and primarily interact with platforms through articles. There are two primary misinformation moments that are incorporated into this model. The first one is the audience-article interaction or when the audience is confronted with articles. The way they react (believing the misinformation or denying it) determines if the misinformation could succeed in misleading users. The other misinformation moment is called rumours and happen when the audience discuss with each other and circulate their impressions, interpretations, or reactions.
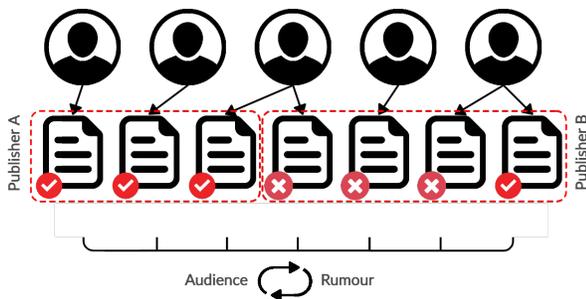

Figure 1: Emergence and spread of misinformation [46].

3 MITIGATION OF MISINFORMATION

In this section, we elaborate on quelling strategies against misinformation. We review strategies from organisational and state perspectives. We assign a strategy in either of these two categories based on their main execution body. Figure 2 summarises all the strategies that are discussed in this section.

3.1 Organisational-Level Strategies

3.1.1 **Platform Strategies**. Social media platforms aim to comply with two competing goals: (i) Keeping the platform free and open to a broad spectrum of ideas and opinions, and (ii) reducing the spread of misinformation.

By the prevalence of social media platforms, they have turned into a hotbed for the spread of misinformation. Due to the number of active users and high-degree of connectivity between them, such information can spread deeply, widely and rapidly throughout the network. Therefore it is of the essence to develop exclusive solutions to control the diffusion of misinformation in social media.

Social network companies have practised different approaches to combat the diffusion of misinformation across a wide range of services in their platforms. A.I. is the first and foremost strategy that have been taken into account by platforms in order to improve the news-feed algorithm, recommendation systems, and content and user filtering [18, 55]. This approach is genuinely appealing to the platform owners as it is fast, cheap, and scalable; however, it suffers from serious flaws which makes it unreliable to be seen as a standalone approach to control the diffusion of misinformation [20]. Design solution is the other approach which is adopted by platform owners to reduce the chance of misinformation emergence or diffusion in their platform. The third main strategy pursued by the platforms is collaboration with third parties, including news organizations, NGOs, universities, and the social media crowd. In the following, we discuss every strategy in more details.

Artificial Intelligence. One of the most popular and controversial ways of approaching misinformation in social media is artificial intelligence (AI). AI tactics let social media platforms tackle the spread of misinformation at scale, across languages and time zones, without relying on reactive reports. Giant social network companies such as Google, Facebook, and Twitter have already utilised AI for managing misinformation and incorporated a wide range of machine learning techniques into their services. In the following two major approaches of social media companies to counter misinformation are discussed.

Filtering. Every minute millions of messages are transmitted across online messaging applications, and thousands of posts are published on social network platforms. Among this massive flow of information, there are some problematic contents which have to be filtered out before appearing on the platforms and applications. Due to the massive inflow of information manual inspection of all the contents seems impossible. What usually platforms do is to first inspect the contents using some machine learning models. In this step, if the machine can decide with very high confidence, there is no need for the second opinion, and machine decision is deemed as final, otherwise, it sent to the next step which is human judgment. Most of the social network platforms follow a similar approach in their active and passive content monitoring.

For instance, Facebook uses AI to enforce its policy guideline called Community Standards. Using AI, Facebook can proactively detect bad contents before anyone reports them and sometimes before a few people, if any, even see them. This works better in some areas such as graphic violence than hate speech where the language need to be understood in order to assess the intent and context of the post. Twitter is also heavily dependent on AI-driven filtering techniques. They have developed machine learning tools that identify and take action on networks of spammy or automated accounts automatically and proactively. Google also utilises AI to automatically detect and remove the contents that violate community guidelines across its services. For instance, in YouTube when a video's content or metadata (title, description or tags) violates guidelines, it will be removed.
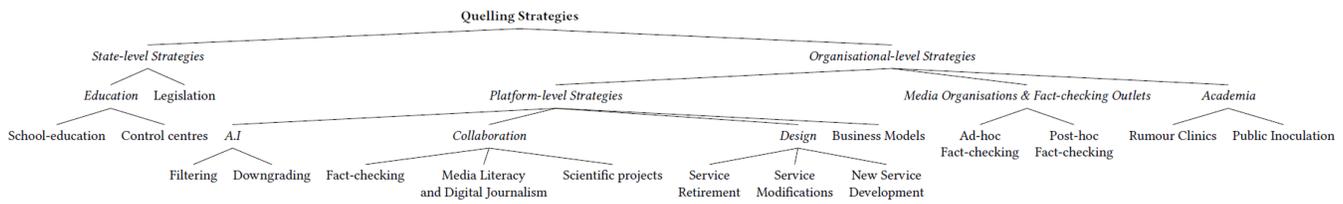
Figure 2: An overview on quelling strategies of misinformation.

Downgrading. We are living in an information-rich world, which wealth of information comes with dearth of attention [47]. Misinformation misuse this simple principle quite often and represent themselves as novel and appealing contents in order to catch eyeballs. In such a setting, the role of ranking and recommendation systems is crucial. If misinformation could manipulate such systems and elevate in search results or timelines, then they will most likely spread rapidly. However, if ranking systems can identify misinformation and relegate them, their visibility and subsequently chance of spread reduces. Platforms have also taken this into account and tried to incorporate signals to their rankings systems in order to make it sensitive to misinformation.

For instance, Facebook has realized most of the false information are financially motivated, so using AI approach they have built classifiers to identify clickbait, ad-farms, and other tactics of financially motivated spammers to reduce the ranking of such stories in news-feed. This action reduces the likelihood of viewing them, clicking them, and making money out of them which ultimately disrupt the incentive for creating such contents in the first place[1]. Twitter has also incorporated behavioural signals into the timeline, search and conversation in order to downrank tweets which neither violate Twitter's policies nor considered as healthy[2]. Google also has incorporated additional signals regarding inaccurate content or debunked conspiracy theories into its recommendation system for queries related to YMYL topics[3]. Beyond specific types of content, some contexts such as breaking news events are also more prone to the dissemination of misinformation. In fact, such contexts work like a magnet for bad behaviour by malicious actors. To reduce the visibility of misinformation, Google designed its systems in a way to prefer authority over factors such as recency or exact word matches while a crisis is developing [4].

One of the crucial discussions regarding filtering and downgrading approaches is difficulty of judging the truthfulness level in a piece of information. It is an extremely difficult and controversial task even for humans. That is why even Facebook as one of the main stakeholders in this domain, tries to stay away from the controversy and not define misinformation in its community standards. This means that downgrading is a safer option for platforms compared to filtering.

Collaboration. Although AI and computational approach can help to filter out or downgrade misinformation in a timely manner, there are still plenty of cases that need human judgement. Additionally, AI is an interventional approach without long-lasting impact on online communities. In order to address such concerns, social media platforms started to collaborate with individuals and organizations for the purpose of fact-checking, media literacy and digital journalism, and scientific projects. In the following, we elaborate over each of those collaboration types.

Fact-checking. Truth finding is a crucial step to mitigate misinformation in online social networks. There are controversial articles circulating in social media and users have no clue about their truthfulness. If those articles could get verified by evidence, they might not spread widely and would quell quickly. Fact-checking also helps AI solutions since it provides valuable training samples for the machine learning algorithms and makes them more accurate. Facebook is practising the same strategy through collaboration with its userbase and independent third-party fact-checkers all around the world. If some shared contents need to be verified by the experts, independent third-party fact-checkers come to the rescue and assess the flagged contents. The downside of such a fact-checking system is the lack of scalability. There are too many items that need to be verified, while the resources are limited. To address this problem, Facebook is going to expand its collaboration with fact-checking organizations and crowdsource it to individual fact-checkers by giving the fact-checking privilege to some of its users [5].

Media Literacy and Digital Journalism. Media literacy and digital journalism are key to empowering citizens with the critical thinking essential to building a resilient society against misinformation. Social media platforms have the potential to raise awareness about the misinformation and reduce media illiteracy through educating people and promoting quality journalism.

Facebook also is engaged in several projects regarding media literacy and digital journalism. One of these projects, which is a collaboration between Facebook and First Draft, aims to build an educational tool to help people spot false news [6]. News Integrity Initiative is the other project launched by 25 partners, including tech industry leaders, academic institutions, non-profits and third party organizations. The initiative's mission is to advance news

---

[1]https://newsroom.fb.com/news/2018/05/hard-questions-false-news/
[2]https://blog.twitter.com/en_us/topics/product/2018/Serving_Healthy_Conversation.html
[3]YMYL is an abbreviation for Your Money Your Life. It is an information category introduced by Google in 2014 and includes financial transaction or information pages, medical and legal information pages, as well as news articles, and public and/or official information pages that are important for having an informed citizenry.
[4]https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/How_Google_Fights_Disinformation.pdf

[5]https://newsroom.fb.com/news/2019/04/tackling-more-false-news-more-quickly-https-newsroom-fb-com-news-2019-04-tackling-more-false-news-more-quickly/
[6]https://newsroom.fb.com/news/2017/04/a-new-educational-tool-against-misinformation/

literacy, to increase trust in journalism around the world and to better inform the public conversation [7].

Twitter is also involved in several projects regarding media literacy and digital journalism. Twitter works with Common Sense Media, the National Association for Media Literacy, the Family Online Safety Institute and Connect Safely, to craft materials and conduct workshops to help users learn how to process online information and understand which sources of news have integrity [8]. Twitter also is in partnership with UNESCO to develop a network of Media and Information Literate Cities [9]. Twitter also has partnerships with EU Disinfo Lab, First Draft, and Atlantic Council's DFRL Lab to name but a few in order to study manipulation techniques and disinformation.

Google has recently established an initiative called Google News Initiative (GNI) to promote media literacy and support quality journalism. In GNI, Google follows a ternary approach. First, it has a partnership with many industry organizations and nonprofits such as First Draft, Poynter, and Local Media Consortium to name but a few to solve important business and industry-wide challenges[10]. Second, it develops programs to meet the needs of freelance journalists and news organizations to succeed on the web.[11] Third, Google builds products to help news organizations to grow their digital businesses[12].

Scientific Projects. In an academic-corporate relationship, collaboration is of the essence for both sides. It helps the academia to ensure industrial relevance in its research [53], and on the other hand, it provides the opportunity of knowledge complementary and risk sharing with the corporates [1]. In the misinformation case also both social media platforms and academia can benefit the collaboration in multiple ways. Diffusion of misinformation is a multifaceted phenomenon which originates in several scientific disciplines such as psychology, neuroscience, and computer science, to name but a few. Collaboration with academia allows platforms to have access to expert human capital in a wide range of disciplines. On the other hand, by accessing unique datasets from social media platforms, academics will be able to test not only old social theories and hypotheses but also propose new ones.

For instance, Twitter began collaborating with the non-profit research centre Cortico and the Massachusetts Institute of Technology Media Lab on exploring how to measure aspects of the health of the public sphere in 2018 [13]. To get more specific indicators for conversational health on Twitter, they expanded their collaboration domain by opening up a request for proposal (RFP) [14] process to cast the widest net possible for great ideas and implementations [15]. In addition to collaboration with academia, Twitter shows interest in open innovation as well. In this regard, they have released several datasets of tweets that resulted from potentially state-backed information operations on Twitter.

__Design Oriented Solutions__. The design solution is another approach to reduce the likelihood of platform abuse and mitigate misinformation in online social networks. For this approach, the social media companies either phase out their operational services, change the access rights to particular services, or develop new services. In the following, we explain how major social media platforms applied it to themselves.

Service Retirement. One of the categories of design oriented solutions is service retirement. When the platform owners see more trouble than benefit coming from a service, they decide to retire that service. It is the most severe yet naive approach regarding a service. When the platform cuts the service misuse from the source, there will not be any abuse from that particular service. On the other hand, there are many researchers and practitioners that their work depends on those services and shutting them down can create a huge mess in their work. For instance, in 2018 and after Cambridge Analytica scandal, Facebook started shutting down some of its APIs such as, Events API, Search API, and App Insights API[16].

Service Modifications. The more intelligent way of approaching platform misuse is service modification. Changing services in terms of closing back doors, and making it more restricted reduce the chance of platform misuse, while keeping it open for legitimate users. However, closing all the breaches and making a service abuse-proof takes huge effort and much time. For instance, WhatsApp recently announced that based on their new policy, a message could be forwarded to a maximum of five recipients [31]. They have also changed their "adding to a group" policy, and users can determine who is allowed to add them to the groups [42]. Twitter also made several changes in its operational services. It reduced follow per day from 1000 accounts to 400 accounts for its users [15].They have also introduced a new registration process for developers requesting access to the APIs [17]. As another change, Twitter has required new accounts to confirm either an email address or phone number when they sign up to Twitter [18]. Twitter also has facilitated reviewing the tweets reports by allowing sharing more information regarding the tweets that are reported [19]. Additionally, Twitter updated the timeline personalisation setting to allow users to select a strictly reverse-chronological experience, without recommended content and recaps [20].

New Service Development. The other design oriented approach is developing new services that have not been existed. The newly designed services are essentially developed to reduce the chance of

---

misinformation emergence or propagation. For example, one of the services that can effectively impact the circulation of misinformation in social media is content labelling which refers to providing meta-information and context regarding a piece of information. This is much the same way we label food in terms of its calories content, its origin, and how it is produced. On the internet and specifically social media, we receive no information regarding the origin, financial resource, and political orientation of an article when we read it in social media. This helps readers to make an informed and smart decision on reading and sharing that article. Although this seems an ideal solution, it comes with many difficulties and ambiguities. One of the key questions that need to be addressed before implementation of such as system is, how are the decisions regarding the content of the labels made? [3, 22, 52].

Facebook has recently developed a similar service called context bottom, which brings more information about the articles that are shared on Facebook. This information include the publisher's Wikipedia entry, other recent articles they have published, information about how many times the article has been shared on Facebook, where it is has been shared, as well as an option to follow the publisher's page. When a publisher does not have a Wikipedia entry, Facebook indicates that the information is unavailable, which can also be helpful context [21]. If that article is reviewed by a fact-checker, Facebook shows more articles from the fact-checker in the related articles unit. If someone tries to post or share that content, Facebook pops up a warning screen to make the user aware that additional reports on this article exist. Additionally, if the article is shared in the past and before it is reviewed by the fact-checkers, Facebook notifies those who previously shared the story on Facebook [22].

Google has also developed new services across its products. For instance knowledge panel in Google Search, full coverage function in Google News, why this ad on Google Ad, breaking news shelves, top news shelves, and developing news information panels on YouTube are some of the services developed to bring more context to the users.

Non-attention Base Economy. One of the other approaches to combat misinformation is changing the internet business model and designing incentive mechanisms to dissuade users from the diffusion of misinformation. The current internet business model works based on a simple principle: attention. This means the contents with more novelty gets more attention from users, which means it is more likely for those contents to be clicked on, and more clicks is interpreted as more money. In such a market, novel contents attract human attention, and being unconstrained by reality makes it easy to come up with any unsubstantiated contents. In other words, it is easy to be surprising and novel when disseminating factually accurate information does not matter [3, 52]. Therefore, generating misinformation can be considered as an effective strategy for gaining attention and subsequently earning money. Thus it is crucial to break this business model or at least disincentivise engagement in such an ill-economy. Implementation of such an approach needs joint effort from major internet companies as the significant share of this market belong to them, and they are shaping this market.

3.1.2 *Fact-checking.*
Media organisations are in the front-line of combating misinformation. They are basically the first provider of the news for their audience, thus they have a huge responsibility in sharing concise and unbiased contents. To fulfill such a promise, media organisations started pay a special attention to fact-checking.

Although the factual writing was always an essential principle in journalism, it was in the early 1900s that the American journalism industry began to really focus on facts. "The professionalization of the business included codifying ethics and creating professional organizations. And, as objective journalism caught on, ideals of accuracy and impartiality began to matter more than ever" [21]. This led to the creation of a new role in media organisations just for inspecting the veracity of information before making public. This role, later on, was called fact-checker [23]. Fact-checking as a systematic practice was started in 1920 when TIME began hiring people to check the articles for accuracy before publication. Shortly after TIME, The New Yorker and Newsweek also launched their fact-checking in 1927 and 1933 respectively. A few decades later, the responsibility of accuracy shifted from independent fact-checkers to writers. As a result, the number of fact-checkers who were solely fact-checker shrank [21].

In the wake of deceptive ads that populated 1988's U.S. presidential race, a new kind of fact-checking started to emerge. It was called political fact-checking which was devoted to analysing the factual accuracy of politicians' statements. Unlike the traditional fact-checking which aims to correct mistakes before making public, political fact-checking's goal is to correcting the misinformation once it is already out there in public sphere[24] [25]. In addition to traditional fact-checking which is a key element in the news industry, since the 80s news organisation started taking political fact-checking into account. A few years later, some non-profit organisations and academic institutions such as Snopes and FactCheck also started launching their initiatives on political fact-checking. Most of those initiatives are associated with a website and communicate with their audience via the internet. Fact-checking is a growing practice among media outlets, non-profit organisations, and academic institutions across the world. Currently, there are more than 200 fact-checking initiatives in more than 68 countries in the world [25]. From the institutional form, based on the research done by The International Fact-Checking Network (IFCN), 64.3% of fact-checking initiatives are non-profit organisations, 28.6% media outlets, and 7.1% are academic initiatives.

3.1.3 *Academia.*
The strategies which are initiated, organized and implemented by the academia fall into this category. They mostly have preemptive nature and aim to educate people against misinformation. In the following, we discuss rumour clinics and public inoculation as two significant strategies by academia to combat misinformation.

Rumour Clinics. The rumour clinic was a response to the growing demand for a strategic solution to the problem of rumour dissemination during World War II. It is mostly inspired by Gordon

---

Allport, a psychology professor at Harvard University. At its core, rumour clinic is referred to a group of technical, representative, and prestige advisors to collect and analyse significant rumours [2, 23]. The initial plan was to deploy and organise the clinics under governmental supervision; however, the fear of sharing the rumours with the public and the following inevitable repercussions make them extremely prudent in such a way that they ultimately withdraw from the project. The government's withdrawal from the project provided social scientists, journalists, and politicians with the opportunity of proceeding in the absence of governmental impediments. Rumour clinics started to proliferate in an unforeseeable rate. One of the most successful rumour clinics was the Boston Herald Rumour Clinic managed by Allport's graduate student, Robert H. Knapp. In the clinic, the rumours reported by the official agencies or non-anonymous readers were analysed. The focal point of this clinic was a newspaper column published every Sunday and its distributions in the high schools. Despite the initial withdrawal of government from rumour clinics, they could not stand the idea of public rumour revelation and in less than a year launched a concrete effort to bury clinics. Their effort paid off in 1943 when the rumour clinic column no longer appeared in the newspapers.

Public Inoculation. Public inoculation is the other approach to confront misinformation. The history of using inoculation goes back to fifty years ago when the inoculation theory was introduced as a means for gaining resistance against persuasion. Inoculation is a metaphor borrowed from biology and refers to the process of injecting the weakened doses of the virus which trigger the immune system to build up resistance against future infection. Inoculation theory also works the same, but instead of a virus, it confers resistance against influence and persuasion. This approach works like a misinformation vaccine and aims to inoculate enough individuals, so the virus (misinformation) does not have a chance to spread [50, 51].

Research on inoculation theory was traditionally focused on maintaining a positive perspective toward cultural truisms, a setting that most people had supportive preexisting attitudes towards the issues [45, 51]. In recent years, scholars started taking the case of misinformation into account. Unlike the traditional setting that the inoculation has been practised, for misinformation prone topics people are likely to hold polarised opinions [45]. The primary reason for choosing this approach to confront dissemination of misinformation was that debunking efforts and correcting the misinformation is not sufficient to stem the flow of online misinformation.

This approach is composed of two primary steps. The first one is similar to biological metaphor and comes with exposing the people to a weakened virus which is in this case, information that challenges their existing beliefs or behaviours. It is worth noting that, as in the vaccination process, the weakened virus should not be so strong as to overwhelm the body's immune system [50]. Then, in the second step, one or more of the presented examples are directly refuted in a process called "refutational pre-emption" or "prebunking" [13, 50]. To improve the effectiveness of this approach, the public should also be vaccinated against the sources of misinformation, by drawing more explicit attention to exactly who is behind those information [22]. Although public inoculation studies have

been focused in particular domains such as health, political campaigning, and climate change, the hypothesis is umbrella protection against the misinformation regardless of the context [44].

Although public inoculation sounds a promising strategy to preempt misinformation campaigns, the implementation of it is shrouded in mystery. One of the most common implementation approaches is the collaboration of academics with reporters to echo the inoculation messages by their media. This approach can also be reinforced if elites and thought-leaders play an active role in the dissemination of inoculation messages [22]. Recently it is shown that serious gaming is a promising vehicle to inoculate public against misinformation [45]. Additionally, in another research, critical thinking was introduced as an effective approach for public inoculation [13]. Currently, public inoculation is mostly in experimental settings in academia. Extension of this strategy beyond that entails future research on different aspects of public inoculation.

3.2 State-level Strategies

In addition to organisations, governments also provide mechanisms to reduce the spread of misinformation in social media. One of the central governments' policies against misinformation is education which is slow but has a long-lasting impact on society. The more invasive approach with faster results is legislation which punishes false news spreaders by means of law enforcement. In the following, we elaborate on each approach.

3.2.1 Education. Education is one of the most promising policies against misinformation. Open Society Institute calls education "the best all-round solution" and mentions "high-quality education and having more and more educated people is a prerequisite for tackling the negative effects of fake news and post-truth" [34].

School Education. This policy aims to inject training of critical-information skills into primary and secondary schools; however, educational policies can be tricky, as too much emphasis on fake-news might backfire and come with unintended consequences of undervaluing the real news outlets [18, 29]. So far, multiple countries have considered incorporating media literacy and journalism principles to the students curriculum. For instance, in Finland, the digital literacy program, designed by one of the Finish fact-checking organisation concentrates on good research skills and critical thinking. It outlines three areas to be aware of: misinformation, disinformation, and malinformation [10]. In Italy, the educational program is designed by the government in cooperation with leading digital companies, including Facebook. The students learn not to share unverified news, ask for sources and evidence, remember that the internet and social networks can be manipulated, and to be always a bit sceptic regarding the digital information [30]. Taiwan is also running an educational program against misinformation. This program helps students develop critical thinking when using social media and focuses on deciphering propaganda and sources of information.

Control Centres. One of the earliest systematic attempts against the mass spread of misinformation was a telephone service called rumour control centre (RCC). It was in fact a follow-up on the rumour clinics which retired in 1943[26] [23]. RCCs were also recognised

---

[26]Please refer to Section 3.1.3

by other names such as "Rumour Central", "Verification Centre", or "Fact Factory" depending on their emergence location. They appeared amid the racial conflicts in the U.S in the 1960s to serve three purposes: riot control, riot prevention, and provision of information service to the general public. "Citizens were encouraged to call RCCs if they heard a rumour that suggested social tensions were increasing in their area. Working with local police and intelligence units, staff would try to locate the source of the rumour and test its veracity. The police could then take preemptive measures to address the unfolding situation by monitoring or arresting suspected agitators and spreading counter-information in areas of the city where the rumour was circulating" [54].

RCCs were inexpensive organisations to set-up. In its most basic form, what they required was a telephone, a few people, and some advertising[27]. From an organisational point of view, in addition to rumour control staff to operates the telephones, there was a communication hookup with police and fire departments as well as other city agencies to get the most recent updates regarding the incidents [32, 41]. From an institutional point of view, RCCs were adopting a similar institutional form by embodying within the government [54].

3.2.2 Legislation. One of the oldest state-level strategies against misinformation is legislation. The earliest form of legislation is the defamation law which may cause lawsuit against defamers. The main purpose of defamation law is to protect the reputation of an entity against libel. Although defamation laws, in particular in US, UK and Canada, are well established and thus a very effective (and obvious) instrument to confront and discourage the spread of misinformation directly, they are also very costly and have a long time horizon. Another hindrance is that defamation laws only protect mostly the reputation of an individual, rather than prevent the spread of false statements. However, lawsuits can draw media attention and thus it can strengthen public awareness by revealing the latent actors and their techniques in the diffusion of misinformation. One of the important points that need to be taken into account is to protect those who purse and win the lawsuits against negative repercussions from the defendant. [22].

After 2016 social media crisis, some countries started to initiate legislation to protect their societies against misinformation. Although legislation is a very powerful tool in the hands of government to punish those who circulate misinformation, it might wittingly or unwittingly clamp down freedom of speech. For instance, in September 2015, as a response to the increasing hate in the social media and the worry of foreign powers using fake news to influence the forthcoming federal elections, Federal Ministry of Justice in Germany organised the formation of a task force consisting of social media platforms, civil society organizations, and institutions of media control. The goal of this task force was to mitigate misinformation in online social media; however, in 2016, the hate speech was still increasing in number. Besides, the first assessment report showed the promise to remove the majority of reported unlawful hate messages within 24 hours is not yet honoured by any company.

Seeing this trend and the inability of the social media platform to tackle hate speech, in March 2017, Federal Justice Minister Heiko

Maas proposed a draft bill called Network Enforcement Act (NetzDG), which sets binding standards for effective and transparent complaints management and obligates the operators of social networks to publicly report quarterly on the handling of complaints about criminally relevant content. Any violations of this law would cause social media platforms with heavy fines. Two months before the 2017 federal elections, this law passed and enforced by Bundestag [8, 11, 12, 16, 17, 40].

France was the other country which initiated a legislation against misinformation. Regarding the massive misinformation campaigns in presidential election 2017, Emmanuel Macron vowed to introduce a law to ban fake news on the internet. This desire is reconciled into two bills which were adapted on the 20 November 2018:

- Organic Law Against Manipulation of Information, No. 772
- Bill on the Fight Against the Manipulation of Information, No. 799.

Those bills essentially cover four main aspects. First of all, those bills will allow judges to take appropriate steps against internet service providers and hosts to curb the spread of inaccurate or misleading allegations or accusations of a fact. On the flipside, internet service providers and hosts are now obligated to inform users as well as authorities about potential misinformation on their platforms. Furthermore, they have to ensure transparency of advertisers on their platform and for whom they work for. Moreover, those bills allow the Conseil Supéreur de l'Audiovisuel (CSA)[28] to prevent, withhold or stop the broadcasting of television services controlled by a foreign state in the case of a malicious interference of the French state's interests. Before its adoption, the Senate rejected those bills in the first and second reading by pointing out that "the weaknesses inherent in texts drafted in a hurry, without prior assessment of the real gaps or failures of our legislation and current regulations" but also that "the dangers of infringements on freedom of expression" is existent [6, 7, 49].

4 EVALUATION FRAMEWORKS

In this section we explain two main platforms that we use to analyse the quelling strategies.

4.1 Strategies Efficacy Level

The spread of misinformation bears many similarities to the evolution and transmission of contagious diseases [33]. Almost half a century ago Goffman and Newill [24] directed attention to the analogy between the spread of infectious diseases and dissemination of information [14]. They argued that transmission of ideas need not to be restricted to infectious diseases but is a more general process that might be applied to many contexts. For example, the development of the psychoanalytic movement in the early twentieth century was no less an epidemic than the outbreak of influenza in 1917 and 1918. As another example, Darwin and evolution, Cantor and set theory, Newton and mechanics, and so on are the examples of epidemics in the world of scientific thoughts. The analogy is not restricted to only scientific context; for example, Christ, Buddha, Moses, and Mohammad can be cited in the religious field, and many other examples can be given almost endlessly [24]. This similarity

---

[27]This set-up varies by the size of the organisation

[28]French regulation agency for various electronic media such as television or radio.

between biological and intellectual epidemics is even caused the same modelling paradigm to be adopted in order to explain the dynamic of propagation [14, 36].

Due to the similarity between infectious diseases and information dissemination, we decided to analyze the existing solutions from the epidemic perspective. In the epidemiology, three approaches are adopted to control epidemics [14]: Education, immunization and screening, and quarantine. The first two tend to prevention and the third one has more intervention nature. Education is one of the simplest and cheapest ways to control epidemics. It is mostly about raising awareness about dos and don'ts regarding a particular disease. For example, in the case of AIDS epidemics, the educational campaigns in February 1987 tried to discourage risk-prone behaviours such as unprotected sex or needle exchange for drug users. The campaign was successful by reducing the spread of viruses in countries where educational campaigns were organized by the state or other organizations.

Immunization is one of the most effective strategies to control epidemics. For example, in the case of smallpox, a worldwide vaccination campaign succeeded in eradicating the disease. The third way to control the epidemic is screening and quarantine. This is an interventional approach as it is exercised when an epidemic has already started to spread. During the epidemics, the susceptible individuals are screened, and the ones who are thought to pose a risk will be quarantined.

4.2 Strategies Execution Mode

One of the key aspects of the confrontation against misinformation is the actual will to combat this malicious phenomenon. This work uses strategy governance mode as a proxy to measure the will for the implementation of a strategy. The strategy governance mode shows to what extent the implementation of a strategy is backed by the government. These policies can be categorized into three forms of institutional regulations:

- **Self-Regulation**: Non-state actor regulates the behaviour of its members [5], [39]
- **Co-Regulation**: Non-state actor regulates the behaviour of its members with some superintendence by governmental agencies [5], [39]
- **Statutory-Regulation**: State actor regulates the behaviour of organizations and its members by implementing and enforcing legislation [5], [39]

In contrast to the common understanding of regulations ("a government intervention to control or constraint"), it actually "enables, facilitates, or adjusts activities, with no restrictions" [38]. Therefore, statutory regulations are not only used by governments to protect the rights of their citizens but also to generate outcomes which would not occur otherwise. However, statutory regulation is considered as insufficient in coping with the increasing complexity of socio-economic challenges. As a consequence, self-regulation gained popularity increasingly over the last couple of decades, and it has already a long tradition in the media sector. Self-regulation is in particular efficient in environments with strong competition and when information is largely in the public domain. Whereas opponents of self-regulation argue that the private firms drive for profits will inevitably lead to exploitation of non-existing legislation

on the shoulders of consumers, proponents claim that the use of self-regulation can overcome the fundamental pitfalls of statutory regulation (i.e., government regulators' lack of understanding about volatile domains). However, there is very little empirical evidence whether statutory-regulation or self-regulation is more effective in practice. An approach to potentially overcome the pitfalls and also to combine the strengths of the two regulatory approaches is co-regulation. Co-regulation means you do not trust the companies to regulate themselves (self-regulation). It also does not mean that you will impose state-law that says you will do X, Y, and Z (statutory-regulation). Co-regulation says, you will do X or we will do Y. In other words, you will demonstrate your ability to regulate fake-news or we will do it for you. It is threatening the companies with action if they do not engage in proper regulation themselves [19, 35]. Similarly to [9], [27] or [4] categorization of self-regulation, this article is differentiating between three shades of co-regulation:

- "*Mandated*" Co-Regulation: self-regulatory framework is determined by the government
- "*Sanctioned*" Co-Regulation: policies are approved by a governmental entity
- "*Coerced*" Co-Regulation: self-regulated policies are developed in reaction to a governmental menace to impose statutory regulation;

5 STRATEGIES EVALUATION

In this section we analyse the quelling strategies against misinformation from two perspectives. First, using epidemic control framework, we investigate the efficacy level of those strategies. Then, using governance framework we look at the execution level of different strategies. As Table 1 displays, strategies have different purposes. Some of them aim to raise awareness and educate people. Platforms collaboration, state educational policy, public inoculation, and fact-checking as well as the obsolete rumour clinics and control centres are all strategies to raise awareness and educate people. This group has a more fundamental approach toward misinformation and consequently takes a long time to be implemented.

The second approach works like vaccination, which makes the social media immune against misinformation even if users are not educated for misinformation. Part of AI strategy works with immunization approach and detect problematic contents before anyone reports them and sometimes before few people, if any, even see them. In design-oriented solution, the goal of removal, modification, or introduction of service is to reduce the chance of platform abuse. In other words, design solutions are like vaccination of the platforms to make creation and spread of misinformation difficult. Fact-checking outlets can work as a vaccine by flagging and correcting the misinformation; however, some research show it is not always the case and sometimes correction of the misinformation backfire [37]. Legislation can also work as a vaccine since it creates an environment that people do not dare to contribute to the dissemination of misinformation.

The third approach is more like damage control and tries to find and remove all tracks of misinformation. In addition to immunization, A.I strategy also used to quarantine and remove (if it is necessary) the problematic contents. The legal approach is also used to capture misinformation and their agents.

Table 1: Analysis of the quelling strategies against epidemic control framework.

		Epidemic Control Platform		
		Education	Immunization	Screening & Quarantine
Strategies	Organisational-level	Public Inoculation Platforms Collaboration Fact checking Rumour Clinics	Platforms AI Platforms Design Fact checking	Platforms AI Fact checking
	State-level	School education Control centres	Legal	Legal

In the second step we look at the execution mode of each approach using governance framework. As Table 2 tabulates, without any surprise, all the organisational strategies, including platform strategies, public inoculation, fact-checking, and rumour clinics, fall into self-regulation category. All those strategies are implemented within the organisational boundary; hence the government has nothing to do with them. Similarly, most of the state-level strategies fall into statutory regulation. To the best of our knowledge, the only exception is the case of NetzDG in Germany (discussed in Section 3.2.2) which is co-regulation. In this case, the initial council works based on mandated co-regulation. It is composed of social media platforms, civil society organizations, and institutions of media control and had sufficient authority to improve the situation of misinformation diffusion; however, after the first round of evaluation, the government was not satisfied by the results; thus they switched to coerced co-regulation, and government took a major control and passed the NetzDG law.

6 DISCUSSION

Our analysis shows that the existing strategies against misinformation cover all levels of efficacy. Although both prevention and intervention strategies have been taken into account in organisational and state level, there is more emphasis on prevention strategies. Additionally, we can observe that the government is much less engaged in a confrontation with misinformation than organisations. We can also observe a polarised sphere from the governance point of view between self-regulation and statutory regulation. In fact most of the strategies fall into either of these two governance modes. The self-regulated strategies are mostly implemented by the organisations which are naturally close to the problem domain, and correspondingly have a very good sense of it. However, for-profit organisations are driven by a key principle of profit maximization, which leads to priorities that are not always aligned with public expectations. Therefore there is no guarantee that those strategies are executed if they clash with companies priorities. Unlike the self-regulated strategies, statutory-regulated strategies are mostly implemented by the government, which is relatively far from problem scope; therefore have less information about it. However, the government goal is to protect its citizens not to maximize any kind of monetary profit and to fulfil its goal, the government has power and authority.

As we can see, there is a paradoxical situation here. From one hand, self-regulated strategies are closer to the problem, but there is no neutral party with enough authority and power that monitor and evaluate them and decide whether they have met specific goals or

not. On the other hand, statutory-regulated strategies have enough power and authority, but they are far from the problem and most likely do not know problems details. This is a classic information asymmetry problem which brings pros and cons for both sides at the same time.

To tackle this issue, we need to find a middle ground in governance modes with sufficient information and power & authority at the same time. Luckily there is such a governance mode, and it is called, co-regulation. Co-regulation has benefits of both self- and statutory-regulation at the same time. It has a very good sense of problem and enforcement guarantee due to the engagement of both organisations and the government. More specifically, Organisations bring details and knowledge about the problem while the government provides legislative support. This approach can be practised in the case of existing quelling strategies by starting from mandated co-regulation which has the least governmental intervention and gradually shift to sanctioned and then coerced regulation if the organisational part could not fulfil the expected outcomes.

REFERENCES

[1] Omar Al-Tabbaa and Samuel Ankrah. 2019. Engineered University-Industry Collaboration: A Social Capital Perspective. *European Management Review* 16, 3 (2019), 543–565.

[2] Gordon Willard Allport and Leo Joseph Postman. 1947. *The psychology of rumor.* H. Holt and company. 247 pages.

[3] Sinan Aral. 2018. MIT Sloan Experts Series – Sinan Aral: The Truth About Fake News. https://www.youtube.com/watch?v=T4enR0sWrLw&t=1s

[4] Ian Ayres and John Braithwaite. 1994. Responsive Regulation: Transcending the Deregulation Debate. *Journal of Legal Studies Education* 12, 2 (6 1994), 345–353.

[5] Robert Baldwin, Martin. Cave, and Martin Lodge. 2012. *Understanding regulation : theory, strategy, and practice.* Oxford University Press. 548 pages.

[6] Gwenaelle Barzic and Sudip Kar-Gupta. 2017. Facebook, Google join drive against fake news in France. https://www.reuters.com/article/us-france-election-facebook-idUSKBN15L0QU

[7] BBC. 2017. Fact-checking fake news in the French election. https://www.bbc.com/news/world-europe-39495635

[8] Sarah Berning. 2015. Charlie Hebdo attack: German lawmakers warn against xenophobia. https://www.dw.com/en/charlie-hebdo-attack-german-lawmakers-warn-against-xenophobia/a-18180826

[9] Julia Black. 1996. Constitutionalising Self-Regulation. *The Modern Law Review* 59, 1 (1 1996), 24–55.

[10] Emma Charlton. 2019. How Finland is fighting fake news - in the classroom. https://www.weforum.org/agenda/2019/05/how-finland-is-fighting-fake-news-in-the-classroom/

[11] Kate Conolly. 2015. Germany greets refugees with help and kindness at Munich central station. https://www.theguardian.com/world/2015/sep/03/germany-refugees-munich-central-station

[12] Naomi Conrad. 2015. German populist groups AfD, Pegida look for common ground. https://www.dw.com/en/german-populist-groups-afd-pegida-look-for-common-ground/a-18180581

[13] John Cook, Peter Ellerton, and David Kinkead. 2018. Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters* 13, 2 (2018), 024018.

Table 2: Analysis of the quelling strategies against governance mode framework.

		Governance Mode Platform		
		Self-regulation	Co-regulation	Statutory-regulation
Strategies	Organisational-level	Platforms AI Platforms Collaboration Platforms Design Fact checking Rumor Clinics Public Inoculation		
	State-level		Legal	School education Control centres Legal

[14] D. J. Daley and D. G. Kendall. 1964. Epidemics and Rumours. *Nature* 204, 4963 (12 1964), 1118–1118.

[15] AJ Dellinger. 2019. Twitter limits daily follows to combat spammers. https://www.engadget.com/2019/04/08/twitter-400-follows-per-day-limit/

[16] DW. 2015. German justice minister to set up task force on Internet hate speech. https://www.dw.com/en/german-justice-minister-to-set-up-task-force-on-internet-hate-speech/a-18714334

[17] DW. 2016. German justice minister: AfD uses hate speech online. https://www.dw.com/en/german-justice-minister-afd-uses-hate-speech-online/a-35961820

[18] David Lazer et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[19] Euronews. 2019. How can Europe tackle fake news in the digital age? https://www.euronews.com/2019/01/09/how-can-europe-tackle-fake-news-in-the-digital-age

[20] Euronews. 2019. world news How can Europe tackle fake news in the digital age? https://www.euronews.com/2019/01/09/how-can-europe-tackle-fake-news-in-the-digital-age

[21] Fabry, Merril. 2017. Here's How the First Fact-Checkers Were Able to Do Their Jobs Before the Internet. https://time.com/4858683/fact-checking-history/

[22] Justin Farrell, Kathryn McConnell, and Robert Brulle. 2019. Evidence-based strategies to combat scientific misinformation. *Nature Climate Change* 9, 3 (3 2019), 191–195.

[23] Cathy Faye. 2007. Governing the grapevine: The study of rumor during World War II. *History of psychology* 10, 1 (2007), 1.

[24] William Goffman and Vaun A. Newill. 1964. Generalization of Epidemic Theory: An Application to the Transmission of Ideas. *Nature* 204, 4955 (10 1964), 225–228.

[25] Lucas Graves. 2016. *Deciding what's true: The rise of political fact-checking in American journalism.* Columbia University Press.

[26] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363, 6425 (1 2019), 374–378.

[27] Neil Gunningham and Joseph Rees. 1997. Industry Self-Regulation: An Institutional Perspective. *Law & Policy* 19, 4 (10 1997), 363–414.

[28] Bernard Hart. 1916. The psychology of rumour.

[29] HGG Hazenberg, MJ van den Hoven, S Cunningham, MR Alfano, H Asghari, EE Sullivan-Mumm, A Ebrahimi Fard, and Elsa Turcios Rodriguez. 2018. Micro-Targeting and ICT media in the Dutch Parliamentary system: Technological changes in Dutch Democracy. (2018).

[30] Jason Horowitz. 2017. In Italian Schools, Reading, Writing and Recognizing Fake News. https://www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html?linked=google&auth=linked-google

[31] Jacob Kastrenakes. 2019. WhatsApp limits message forwarding in fight against misinformation. https://www.theverge.com/2019/1/21/18191455/whatsapp-forwarding-limit-five-messages-misinformation-battle

[32] Terry Ann Knopf. 2017. *Rumors, Race and Riots.* Routledge.

[33] Adam Kucharski. 2016. Study epidemiology of fake news. *Nature* 540, 7634 (12 2016), 525–525.

[34] Marin Lessenski. 2018. *Common Sense Wanted: Resilience to 'Post-Truth' and its Predictors in the New Media Literacy Index 2018.* Technical Report. Open Society Institute. https://osis.bg/wp-content/uploads/2018/04/MediaLiteracyIndex2018_publishENG.pdf

[35] Christopher T Marsden. 2010. *Net neutrality: Towards a co-regulatory solution.* A&C Black.

[36] Yamir Moreno, Maziar Nekovee, and Amalio F. Pacheco. 2004. Dynamics of rumor spreading in complex networks. *Physical Review E* 69, 6 (6 2004), 066130.

[37] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[38] B Orbach. 2012. What is regulation? *Yale Journal of Regulation* 30, 1 (2012), 1–10.

[39] OSCE. 2004. *The Media Freedom Internet Cookbook - Chapter: Self-regulation, Co-regulation, State-regulation.* Technical Report. Organization for Security and Co-operation in Europe, Amsterdam. https://www.osce.org/fom/13844

[40] Alexander Pearson. 2018. German opposition parties call to replace online hate speech law. https://www.dw.com/en/german-opposition-parties-call-to-replace-online-hate-speech-law/a-42058030

[41] J. Rick Ponting. 1973. Rumor Control Centers. *American Behavioral Scientist* 16, 3 (1 1973), 391–401.

[42] Jon Porter. 2019. WhatsApp now lets you control who can add you to groups. https://www.theverge.com/2019/4/3/18293441/whatsapp-group-block-private-invitation-misinformation

[43] Jamuna Prasad. 1935. The psychology of rumour: A study relating to the great Indian earthquake of 1934. *British Journal of Psychology. General Section* 26, 1 (1935), 1–15.

[44] Jon Roozenbeek and Sander van der Linden. 2019. The fake news game: actively inoculating against the risk of misinformation. *Journal of Risk Research* 22, 5 (2019), 570–580.

[45] Jon Roozenbeek and Sander van der Linden. 2019. Fake news game confers psychological resistance against online misinformation. *Palgrave Communications* 5, 1 (2019), 12.

[46] Derek Ruths. 2019. The misinformation machine. *Science* 363, 6425 (2019), 348.

[47] Herbert A Simon. 1996. Designing organizations for an information-rich world. *International Library of Critical Writings in Economics* 70 (1996), 187–202.

[48] L William Stern. 1902. *Zur psychologie der aussage: experimentelle untersuchungen über erinnerungstreue.*

[49] The Local. 2017. Battle begins to stop 'fake news' from impacting the French presidential election. https://www.thelocal.fr/20170228/can-could-fake-news-influence-the-french-election

[50] Sander Van der Linden, Anthony Leiserowitz, Seth Rosenthal, and Edward Maibach. 2017. Inoculating the public against misinformation about climate change. *Global Challenges* 1, 2 (2017), 1600008.

[51] Sander Van Der Linden, Edward Maibach, John Cook, Anthony Leiserowitz, and Stephan Lewandowsky. 2017. Inoculating against misinformation. *Science* 358, 6367 (2017), 1141–1142.

[52] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (3 2018), 1146–1151.

[53] Claes Wohlin, Aybuke Aurum, Lefteris Angelis, Laura Phillips, Yvonne Dittrich, Tony Gorschek, Hakan Grahn, Kennet Henningsson, Simon Kagstrom, Graham Low, et al. 2011. The success factors powering industry-academia collaboration. *IEEE software* 29, 2 (2011), 67–73.

[54] Stephen Young, Alasdair Pinkerton, and Klaus Dodds. 2014. The word on the street: Rumor,"race" and the anticipation of urban unrest. *Political Geography* 38 (2014), 57–67.

[55] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and Resolution of Rumours in Social Media. *Comput. Surveys* 51, 2 (2 2018), 1–36.