Countering Disinformation: Using Systems Thinking to Develop an Integrated Approach

by

Brendan Tan Weijian

B. Eng. (Hons) Electrical Engineering
National University of Singapore, 2009

Submitted to the System and Design Management Program
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management
at the
Massachusetts Institute of Technology

February 2020

Signature redacted

Signature of Author______________________________

Brendan Tan Weijian
MIT System and Design Management Program
December 16, 2019

Certified by____________________ Signature redacted

Donna H. Rhodes
Principal Research Scientist, Sociotechnical Systems Research Center
Thesis Supervisor

Accepted by__________________ Signature redacted

Joan Rubin
Executive Director, System Design and Management Program

This page is intentionally left blank

Countering Disinformation:
Using Systems Thinking to Develop an Integrated Approach

by

Brendan Tan Weijian

Submitted to the MIT System and Design Management (SDM) program on December 16, 2019 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Engineering and Management.

Abstract

Although disinformation has been used by malicious actors throughout the ages, technological advances and social changes in the Internet era have made it a more potent problem than ever before. Several key elections in 2016 were heavily influenced by disinformation – including the Brexit referendum and US presidential elections – which made it evident that information is increasingly being used as a weapon. Fittingly, the Center for European Policy Analysis has highlighted that the "age of information is fast becoming the age of disinformation". Although disinformation campaigns can have a significant impact on the population, the manpower and costs required to carry them out are disproportionately low. It is thus imperative that nations are adequately prepared to deal with this asymmetric and dangerous threat using a suite of countermeasures, as there is no one silver bullet to tackle this problem. Countermeasures that have been proposed or implemented around the globe are analyzed, and a systems thinking methodology is used to develop an integrated approach to deal with this complex issue at the national-level. To guide the entire thought process, the ARchitecting Innovative Enterprise Strategy (ARIES) framework is utilized. Finally, a case study on the annexation of Crimea by Russia serves to qualitatively validate the proposed system and ascertain its applicability in a real scenario. This particular case study is chosen as it is a prime example of how disinformation can be a powerful tool in the hands of adversaries.

Thesis Supervisor: Donna H. Rhodes
Title: Principal Research Scientist, Sociotechnical Systems Research Center

This page is intentionally left blank

Acknowledgements

First and foremost, I would like to thank my wife, Ai Leen, for her unwavering support and understanding throughout my studies in MIT. It was definitely not easy taking care of our young daughter while also being pregnant with my son, and for that I am forever grateful.

To my three-year-old daughter, thank you for being as understanding as you possibly could, whenever I was too busy to play with you or keep you company. I would also like to thank my mum for coming over for four months to help look after our two young children.

Next, I would like to thank my thesis advisor, Dr. Donna Rhodes, for her guidance and invaluable advice throughout the course of my work. I was always amazed with her prompt responses despite her busy schedule.

I would also like to thank my organization, the Defence Science and Technology Agency, for sponsoring my degree and giving me this opportunity to study at this prestigious university. In particular, I would like to express my gratitude to my supervisor, Ms. Wong Hsin Min, for supporting my scholarship application.

To all faculty members and staff of the SDM program, thank you for putting together this enriching program and letting me be a part of it. I have certainly learnt a lot during my time here.

I am also indebted to our family in Singapore, for looking after our dog while we are away and visiting us whenever time permitted.

This page is intentionally left blank

Table of Contents

This page is intentionally left blank

List of Figures

This page is intentionally left blank

List of Tables

This page is intentionally left blank

List of Acronyms and Abbreviations

ARIES	ARchitecting Innovative Enterprise Strategy
BBC	British Broadcasting Corporation
CfA	Campaign for Accountability
IREX	International Research & Exchange Board
NGO	Non-Governmental Organization
OPD	Object-Process Diagram
R&T	Research and Technology
UK	United Kingdom
US	United States
WHO	World Health Organization

This page is intentionally left blank

Chapter 1: Introduction

This thesis aims to understand the growing threat of disinformation and to propose an integrated approach to deal with the issue. This chapter lays out the background, motivation, aim, approach, scope, and structure of the thesis.

1.1 Background

It has been said that "disinformation is an old story, fueled by new technology". While manipulation of information has been used as far back as the times of ancient Rome [1], technological advances and social changes in the Internet era have made it a more potent problem than ever before. The Internet has not only "democratized information" but also "democratized its weaponization", making digital tools and techniques easily accessible to average civilians [2]. Consequently, this "age of information is fast becoming the age of disinformation" [3]. In 2016, several events made it evident that information is increasingly being used as a weapon – Philippines' President Rodrigo Duterte's rise to power, which was spurred by intensive Facebook activity; the unexpected results during the Brexit referendum in June and the US presidential elections in November [4], both of which were thought to have been significantly influenced by the prevalence of fake news [5]. Technological advances such as the emergence of automation and micro-targeting have also made it easier for disinformation agents to leverage regular social media users to spread harmful messages and fuel disinformation campaigns aiming to manipulate public opinion [4]. This "communications revolution" has accelerated the spread of lies, misinformation and dubious claims [6], making it a serious problem for nations worldwide.

1.2 Motivation

Disinformation campaigns can have a significant impact on the population, while the manpower and costs required to carry them out are disproportionately low. In addition, disinformation operations undertaken by state actors carry a low risk of detection and allow aggressor states to disclaim responsibility [2]. Hence it is imperative that nations are adequately prepared to deal with this asymmetric and dangerous threat using a suite of countermeasures. It is widely accepted that there is no one silver bullet to tackle this problem and a multi-pronged approach is necessary [2] [3] [7]. It is also important to address the complex socio-technical challenges posed by the prevalence of disinformation [8]. Consequently, the author sees value in using a systems thinking methodology to develop an integrated approach to deal with this complex issue.

1.3 Aim

The author finds the "*To-By-Using*" framing [9] to be particularly useful in articulating the aim of the thesis and it can be stated as:

To prevent the spread of disinformation
By developing an integrated approach at the national-level
Using Systems Thinking methodology

1.4 Approach

The primary research question revolves around "how" to address the threat of disinformation. Firstly, however, it is necessary to understand "what" the threat is and "what" are the possible countermeasures that are being adopted or in the process of being adopted by nations worldwide. Hence a literature review was conducted to better understand the threat – the means in which disinformation is generated and spread, the actors involved and their objectives, the impact to society, the challenges faced in countering disinformation, as well as possible countermeasures and their respective pros and cons.

Secondly, systems thinking methodologies and tools are used to dissect the problem and subsequently piece together the various elements into an integrated approach that can prevent the spread of disinformation. This includes analyzing stakeholders and their needs, identifying key architectural decisions, and defining the appropriate abstractions to best represent the entities of the system [9]. The ARchitecting Innovative Enterprise Strategy (ARIES) framework [10] is used to guide the entire thought process. A brief explanation of the framework and its application to this thesis is provided in Appendix A.

Lastly, a historical case study approach is used to qualitatively validate the proposed system and ascertain its applicability in a real scenario.

1.5 Scope

This thesis focuses on disinformation, i.e., information that is both false and disseminated with an intent to harm. It also focuses on the online dissemination of disinformation, as opposed to other means, since this is how falsehoods can be generated and spread quickly to a wide reach of audience. The author has also made a conscious decision to avoid analysing and developing a response for a specific country, but rather to create a broad framework that can then be customized to suit each country's specific needs.

1.6 Thesis Structure

Chapter 1 lays out the background, motivation, aim, approach, scope, and structure of the thesis. Chapter 2 provides an overview of the problem of disinformation, the objectives of disinformation agents, as well as the impact on society and challenges faced in combatting disinformation. Chapter 3 describes the current landscape, stakeholders, and current architecture by looking at it from different perspectives. Chapter 4 covers possible countermeasures that are being proposed or have been adopted by countries worldwide. Chapter 5 builds on the list of possible countermeasures to develop a proposed system architecture for combating disinformation. Chapter 6 qualitatively validates the proposed system and ascertain its applicability in a real scenario by using a case study. Chapter 7 summarizes the findings of the research, and identifies both the limitations and possible future research related to the topic.

This page is intentionally left blank

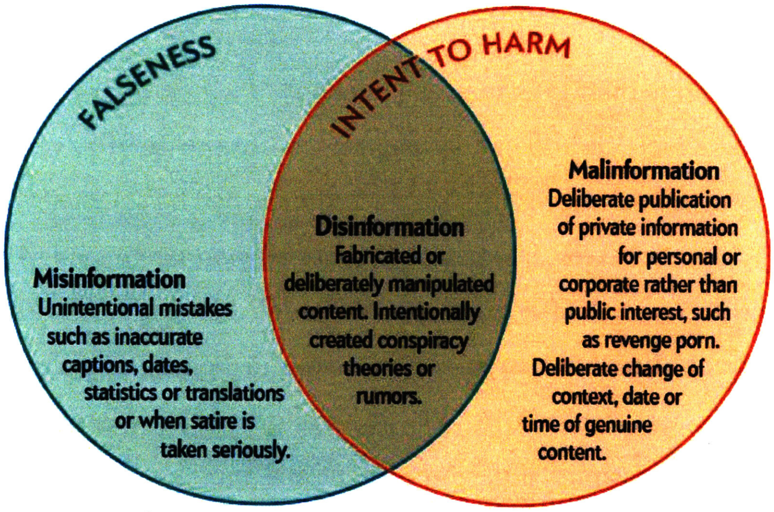
Chapter 2: Overview of Disinformation

This chapter aims to provide a better understanding of the problem at hand and describe how the emergence of the internet and social technology have made disinformation a more pressing and potent issue. This includes a review of the various types of disinformation agents and their objectives, as well as the impact on societies and challenges faced in combatting disinformation. Actual examples are used to illustrate each case in point whenever possible.

2.1 Definition of Disinformation

To better understand the complexity of the information ecosystem, there is a need for a common interpretation of the term "disinformation". There are several categories of information disorder and this is explained in Figure 1. This thesis will focus on disinformation, i.e., when both falseness and intent to harm are present. However, it is important to note that misinformation – i.e. the spread of false content by a person unintentionally – is one way in which disinformation can be spread unwittingly by people. It is hence crucial to address this issue as well.

Figure 1: Three Categories of Information Disorder **[4]**.


FALSENESS
INTENT TO HARM
Misinformation
Unintentional mistakes such as inaccurate captions, dates, statistics or translations or when satire is taken seriously.
Disinformation
Fabricated or deliberately manipulated content. Intentionally created conspiracy theories or rumors.
Malinformation
Deliberate publication of private information for personal or corporate rather than public interest, such as revenge porn. Deliberate change of context, date or time of genuine content.

According to the definition, disinformation includes both fabricated or deliberately manipulated content. As opposed to wholly fabricated content, deliberately manipulated content that has a kernel of truth to it and is merely misleading, has been deemed to be the most effective type of disinformation [4]. It has been said that disinformation messages "must at least partially correspond to reality or generally accepted views" [11]. Hence influence agents have resorted to reframing genuine content and using hyperbolic headlines [4] as one of the primary means of disinformation.

2.2 Effects and Implications of Digital Technologies

The emergence of the Internet and social technology have brought about fundamental changes to the way in which information is produced, communicated, and distributed. The widely accessible, cheap and sophisticated editing and publishing technology, has made it easier than ever for anyone to create and distribute content; while information can be disseminated and passed between trusted peers at an unprecedented pace [12]. This can be described using the analogy of applying Moore's Law to the distribution of disinformation: "an exponential growth of available technology coupled with a rapid collapse of costs" [13]. Furthermore, the technology, tools and services available to malicious actors are continuously improving. Anyone, not just well-resourced States, can carry out impactful disinformation campaigns [2]. This section further elaborates on how digital technologies have not only accelerated the dissemination of disinformation but increased the accessibility of tools to potential disinformation agents.

2.2.1 Accelerating the Dissemination of Disinformation

Targeted Advertising. Online platforms such as Facebook, Twitter, and Google offer user-friendly and affordable targeted advertising tools that anyone can use to send advertisements to specific users based on their known preferences [2]. Target audience can be selected based on criteria such as location, demographics, behavior, interests, connections, and language [14] [15] [16]. This is a form of "micro-targeting" where different groups of people are targeted with tailored messaging [2]. On its marketing site, Facebook even cited its role in Pennsylvania senator Pat Toomey's extremely narrow win in November 2016 [17]. It highlighted that custom advertisements on its platform were able to "significantly shift voter intent and increase favorability" for Toomey by double digits among key demographics [18]. Targeted advertisements were also allegedly purchased by Russian trolls during the 2016 US presidential election [19] and Facebook estimated that 126 million Americans viewed propaganda generated by the Internet Research Agency – a "troll farm" with known Kremlin ties [20]. The advertisements were targeted against specific groups such as professed gun lovers, fans of Martin Luther King Jr., supporters of Trump, supporters of Clinton, and residents of certain states [21].

Wide Reach of Audience. The borderless nature of the Internet means one can reach anyone anywhere in the world. The sheer size of some social media platforms provides a huge potential audience for disinformation agents to act on [2]. This ability of social media to directly reach large numbers of people while simultaneously micro-targeting individuals with personalized messages, makes social media platforms very attractive not only to advertisers, but also to political operatives and foreign adversaries [22].

Easy Amplification of Falsehoods. Falsehoods may be spread further and faster using basic, everyday social media functions, such as posting, "sharing," "liking", re-tweeting, hyper-linking and hash-tagging [2]. For example, during Germany's national election in 2017, anonymous troll accounts and bots were able to push vote-rigging claims via Twitter using re-tweets and hash-tags [23]. Such functions have become so common that people are accustomed to sharing information without knowing its original source. This is exacerbated by the fact that on social media, information is often shared amongst trusted peers without any verification [2]. Through targeted advertising, social networks allow "atoms" of propaganda to be directly targeted at users who are more likely to share a particular message. The next person who sees it in their social feed probably trusts the original poster, and goes on to share it themselves. These "atoms" then rocket through the information ecosystem at high speed, powered by trusted peer-to-peer networks [24].

Inauthentic Accounts. Inauthentic social media accounts may be used to artificially amplify online falsehoods. Fake social media accounts are easily created, due to the lack of stringent verification requirements [2]. This issue is widespread and almost all 48 countries surveyed by a University of Oxford study, displayed evidence of fake accounts [22]. These accounts usually seek to attract followers, to boost the size of their social network and audience [2]. For example, a fake Tennessee Republican Party Twitter account had more than 152,000 followers, which is significantly more than the 13,800 followers of the real Tennessee Republican Party account [25]. Fake social media accounts may be run either by humans, known as "trolls", or by algorithms, known as "bots". While human trolls work in a coordinated manner to rapidly amplify a particular online falsehood [2]; bots work by strategically posting particular keywords, in order to game algorithms and cause certain content to trend, as well as flooding hashtags with automated messages [22]. An example was the Twitter activity around the hashtag #MacronLeaks[1] in 2017, which targeted then-presidential candidate Emmanuel Macron. One bot on Twitter posted 294 tweets on #MacronLeaks in three-and-a-half hours [26].

Social Media Algorithms. Falsehoods are given a further boost by the algorithms of social media platforms, which are designed to automatically promote popular posts [2]. While the original intent was to "surface the most newsworthy, relevant information in the midst of a vast sea of content", these algorithms could also inadvertently drive bot-fueled hashtag campaigns promoting gun rights to the top of Twitter Trends [27]. Ironically, an algorithm change apparently designed to keep fake news out was precisely what put a conspiracy video at the top of YouTube's Trending section. The video falsely claimed that a survivor of the shooting at a school in Parkland, Florida was not a genuine victim but an actor [27]. Although the video had "no catchy headline" and "no recognizable personality", it still "blasted through YouTube's safeguards and somehow kept going" [28]. YouTube later clarified that its system had "misclassified" the conspiracy video "because the video contained footage from an authoritative news source" [27].

Multiple Platforms to Spread Disinformation. Besides social media, there are multiple platforms that can spread disinformation in the online sphere. One of the iconic hoaxes of the 2016 US presidential election was that Pope Francis had endorsed Donald Trump. The Facebook post earned close to 1 million Facebook engagements and was the single biggest fake news hit of the election. It was discovered that the website, which first published the story, was part of a network of at least 43 websites that published more than 750 fake news articles in total [29]. The range of platforms on which disinformation is carried out is also growing, with evidence of cyber troop activity on chat applications or other platforms including Instagram, LINE, SnapChat, Telegram, Tinder, WeChat, and WhatsApp [22].

2.2.2 Increasing the Accessibility of Disinformation Tools

Low Cost and High Impact Tools. User-friendly and cheap tools for creating audio and visual content are readily available. This has enabled relatively unskilled users to manipulate and distort

[1] The hashtag referred to e-mails leaked after Macron's account was hacked.

videos in ways that are difficult to detect [2]. For example, in 2016, a video featuring a speech of the then-incumbent Jakarta governor was edited and subsequently uploaded online. A vital part of the speech was edited out of the video, creating the perception that his remarks were aimed at the Islamic holy book, rather than opponents who had misquoted Quranic verses to support their political agenda [30]. The doctored video went viral and successfully drew hundreds of thousands of Muslims to participate in two rallies in Jakarta [31]. One of the rallies turned violent, leaving one man dead and dozens of police and demonstrators injured [32]. To make matters worse, there is also the looming security threat of "deepfakes" – the artificial intelligence-powered imitation of speech and images that can make someone appear to be saying or doing things they never said or did. Such fabricated videos could put words and expressions onto the face and mouth of a politician and influence elections [33].

Availability of Bots and Services. Online disinformation has become a profitable industry. Digital tools such as bots are cheap and easy to deploy [34], while services that need more manpower or skill are also available for a price. This includes "hired guns" who perform online manipulation of public opinion and voting outcomes. These online influence tools and services cost significantly less than conventional advertising and marketing, and is able to obtain the same or even greater reach [2]. With easy access to such tools and services, even ordinary people can engage in sophisticated online disinformation campaigns, and spread falsehoods quickly and widely [2] [34].

2.3 Actors and Objectives

Disinformation agents, or actors, may be foreign or local, state-sponsored or merely civilians. They may have different motivations and seek to achieve a myriad of objectives, as will be discussed in this section. The objectives of the different actors may sometimes align, either knowingly or unknowingly.

2.3.1 Foreign State Actors

Falsehoods systematically spread by foreign states, i.e. disinformation operations, is of grave concern as they have the ability to harm the national sovereignty and security of the target state. Such operations are now an established part of the military arsenal, and serve to further the aggressor state's broader geopolitical interests, as well as to undermine the social resilience of the target state. It has been said that such "non-kinetic" warfare have become just as important as the conventional "kinetic" warfare. In fact, these non-kinetic means may even be more important in current geopolitical conflicts, especially when up against a militarily superior adversary [2]. For example, when President Vladimir Putin came to power, Russia began leveraging influence campaigns and cyberwarfare to compensate for its weaker military compared to the US. Both these methods were cheap, easy to deploy, and hard for an open and networked society such as the US to defend against [35].

Disinformation operations are attractive for a couple of reasons. Firstly, the costs and manpower needed are disproportionately low when considering its potential impact. In addition, due to its effectiveness, the need to deploy kinetic means of warfare could even be reduced. Secondly, disinformation operations carry a low risk of detection and allow aggressor states to disclaim responsibility, especially if they are carried out by agents not paid by the state and appear as ordinary citizens. Hence these operations can be conducted persistently and permanently [2], without the need to distinguish between peacetime and wartime. To put it simply, information operations allow aggressor states to wage war without ever announcing it officially and opens wide asymmetrical possibilities to reduce the enemy's fighting potential [11]. Subsequent paragraphs dive deeper into the various objectives of foreign state actors.

Influence Government Policy. Disinformation can make it almost impossible for governments to develop policies to deal with issues such as immigration [2]. For example, Russian state media ran reports on the alleged abduction and rape of an ethnic Russian girl by Middle Eastern migrants in Berlin, and accusations of a cover-up, well after the Berlin prosecutor's office had rejected the allegations as false [36]. The aim was to influence the European debate on immigration [37].

Discredit the Government. There have been several instances of one-sided reporting that aimed to discredit governments. For example, programs were aired by the Russian broadcaster, RT, on events in Ukraine in 2014, and events in Turkey in 2016, during periods when the Russian government was at loggerheads with the countries in question. The broadcaster would interview multiple speakers who made grave accusations against the Ukrainian and Turkish governments, without providing adequate comments from those governments [36]. It has been alleged that the Kremlin-run broadcaster, RT, focuses on making the West, and especially the US, look bad [38].

Achieve an Election Outcome. Perhaps one of the most notable incidents is Russia's alleged influence campaign during the 2016 US presidential election, where the goals were to undermine public faith in the US democratic process, denigrate Secretary Clinton, and harm her electability and potential presidency [39].

Fuel Discontent within Society. A report by the US Intelligence Community discussed Russian efforts to fuel discontent in the US, with the longstanding desire to undermine the US-led liberal democratic order [39]. This was apparently done by targeting already divisive issues, such as race, LGBT rights, gun control, and immigration, with the objective of turning groups against each other [2]. For example, an army of well-paid "trolls" from Russia were alleged to have spread a fake video of an unarmed black woman who was shot dead by police, with the hashtag #shockingmurderinatlanta [40] [41]. Its purpose was to widen the divide between the African-American community and the police, as well as to undermine the police as an institution [2]. Public health issues, such as vaccination, have also been used by foreign powers to serve as a political wedge issue [42]. Russian troll accounts allegedly posted both pro- and anti-vaccination messages to create "false equivalency", used polarizing language, and linked vaccination to controversial statements about race, class, and government legitimacy [43].

2.3.2 Local Actors

Influence Government Policy. Disinformation may be spread to manipulate those who do not share the same political or ideological beliefs [2]. For example, in the UK, several media companies published articles in 2011 reporting that new sentencing guidelines would allow those supplying drugs to avoid a custodial sentence, if they were playing a "subordinate" role in a criminal gang. The reality is that the new sentencing guidelines made no changes at all to the approach to sentencing those involved in the supply of drugs. Nonetheless, the stories fit within an established agenda to resist a general "softening" in criminal sentencing [44]. In the US, after the school shooting in Parkland, Florida in Feb 2018, far-right websites spread false accusations that students interviewed were paid "crisis actors" and that the shooting never occurred [45]. Their objective was to shore up support for gun rights [2].

Discredit the Government. Independent of elections, online hoaxes spearheaded by domestic groups in Indonesia have reportedly aimed to discredit Indonesian President Jokowi. One common claim was that President Jokowi had communist affiliations, with the objective of tapping on entrenched anti-communist sentiments [2]. During the 2019 protests in Hong Kong that crippled the city, protest supporters often demonized the police and the government. One article claimed that a 15 year old girl was killed by police or government agents for participating in the protests, despite her death being an apparent suicide [46].

Achieve an Election Outcome. During the 2016 US presidential election campaign, the domestic alt-right group allegedly used several false narratives to harm the Clinton campaign and boost that of Trump [2]. Another notable example, the Brexit referendum campaign, is best remembered for the lies told by leading campaigners. A local British newspaper ran a front-page story with the headline "Queen Backs Brexit", which was purely based on anonymous sources and turned out to be false. Although corrections were subsequently published on inside pages of the newspaper, the false story had already been ingrained in the collective consciousness of the readers [6].

Fuel Discontent within Society. Anti-Muslim falsehoods have been spread in the US and the UK by domestic far-right groups. For example, following the terrorist attack in Paris in 2017, a far-right political leader in the UK posted a video and described it as showing Muslims celebrating the attack. It was in fact a video of British Pakistanis celebrating a cricket match victory back in 2009 [47]. Indonesian authorities have discovered an extensive and politically well-connected network known as the Muslim Cyber Army. This group has spread disinformation to inflame sentiments against gay men and lesbians, communists, Chinese, and the government, as well as to promote a hardline Islamist stance. Their actions were coordinated through a central WhatsApp group, and used bots to amplify the disinformation [2].

Financial Gains. Digital advertising models allow website owners to earn advertising revenue depending on the amount of user interaction with the advertisements placed on or linked to their websites. This has encouraged content producers to disregard the truth to attract more "clicks" and earn more revenue [2]. For example, American Paul Horner, who claimed that he hated Trump, wrote false stories that could have influenced the elections in Trump's favor just "for the money". He published a series of fake URLs that cloaked themselves to appear to be being coming from major media companies, such as ABC and CNN, and then disseminated fake news stories across Facebook [48]. He apparently earned about $10,000 a month doing so [49].

2.3.3 Foreign Non-State Actors

Influence Government Policy. China apparently has a growing army of Internet trolls to attack voices perceived to be hostile to Beijing's interest [50]. This army comprises volunteers who are willing to work without pay and prowl social media to rebut criticism of their homeland [51].

Discredit the Government. Foreign non-State actors such as NGOs and media organizations may spread disinformation to de-legitimize a government [2]. For example, prior to the annexation of Crimea, Russian media allegedly painted a negative picture of Ukraine. One pro-Kremlin claim was that Ukraine's ethnic Russians were being oppressed and attacked. A video on Russian television allegedly showed the Ukrainian military firing on civilians, when instead, the men filmed were actually Russian militia [52].

Achieve an Election Outcome. During Germany's national election in 2017, right-wing groups in the United States were alleged to have tried to sway public opinion in favor of German right-wing parties [53]. A Colombian, Andrés Sepúlveda, manipulated social media to create false waves of enthusiasm and derision to favor right-winged candidates. He apparently influenced presidential elections across Latin America – Nicaragua, Panama, Honduras, El Salvador, Colombia, Mexico, Costa Rica, Guatemala, and Venezuela – for almost a decade [54].

Fuel Discontent within Society. Racist and other such prejudiced agendas are not limited by national borders, and disinformation supporting such agendas may be disseminated online to audiences around the world [2]. For example, British Darren Osburne was influenced by material from the US "conspiracy theory and fake news website" InfoWars, in the weeks prior to carrying out a van attack against Muslim worshippers in Finsbury Park in 2017 [55].

Financial Gain. Perhaps the most notable example is that of the many fake news websites that sprang up during the 2016 US presidential election. These websites were traced to a small city in Macedonia, where teenagers generated sensationalized stories just to earn money through digital advertising [56].

Radicalize. Terrorist organizations, such as ISIL, have used online disinformation to radicalize people around the world [2]. For example, the terror group released a propaganda video in 2017 featuring a Singaporean fighter who called on viewers to join its fight [57]. The video was deemed to be "full of distortions and falsehood, deliberately designed to mislead Muslim viewers into sympathizing with ISIS" [58].

2.3.4 Alignment of Objectives

A summary of the actors and their objectives is depicted in Table 1. It shows that the objectives of the different types of actors may overlap and such alignment of interests may be deliberate or unwitting. In such cases, the threat they pose is greater [2]. Governments and political parties have worked in conjunction with private industry, civil society organizations, Internet subcultures, youth groups, hacker collectives, fringe movements, social media influencers, and volunteers who ideologically supported their cause. For example, formal cooperation between industry and political parties have apparently occurred in Austria, Brazil, Colombia, Ecuador, India, Kyrgyzstan, Malaysia, Mexico, Nigeria, Philippines, Poland, South Africa, the UK, and the US, where political parties and campaign managers directly hired PR or consulting firms to help spread propaganda during elections [22].

Table 1: Summary of actors and their objectives.

Actor / Objective	Influence Government Policy	Discredit the Government	Achieve an Election Outcome	Fuel Discontent within Society	Financial Gain	Radicalize
Foreign State Actor	✓	✓	✓	✓		
Local Actor	✓	✓	✓	✓	✓	
Foreign Non-State Actor	✓	✓	✓	✓	✓	✓

2.4 Impact of Disinformation

Disinformation could have short-term or long-term effects. It could also endanger national security, the democratic system we live in, and even individuals. These are discussed in the following paragraphs.

2.4.1 Immediate versus "Slow Drip" Effects

The impact of disinformation could be evident immediately or over a long term. An example of an immediate impact is the fake tweet about an explosion at the White House that caused the S&P 500 index to plunge and wipe off about US$137B in the space of three minutes. This fake tweet was a result of the Associated Press' Twitter account being hacked, an act claimed by the Syria Electronic Army [59].

On the other hand, "slow drip" effects may take a longer time to be evident. Such falsehoods may gradually inflame tensions and exploit "slow burn" issues, such as communal tensions, to cause more serious crises in the future. For example, in Myanmar, falsehoods are spread on Facebook to stoke violence against the Rohingya, a Muslim ethnic minority group [60]. Accusations of sexual assault have led to communal clashes between Buddhists and Muslims in Myanmar, and resulted in deaths [61].

2.4.2 Threat to National Security

Foreign state actors can undermine a nation's sovereignty when they attempt to influence government policies, discredit governments, or fuel discontent within society, so as to pave the way for the foreign state to assume control. These objectives have been described above in Chapter 2.3.1. Regardless of whether a foreign state is behind it, disinformation can affect national security if they undermine social cohesion, incite public unrest or violence, or cause public alarm [2].

Undermine Social Cohesion. Falsehoods can polarize and divide society by exploiting "fault lines" [2]. For example, in the US, it is alleged that Russian-linked bots have rallied around divisive issues such as gun control. After the mass shooting in Parkland, Florida, those bots promoted the conspiracy theory that the shooting had never happened and was in fact a secret government operation [62].

Incite Public Unrest and Violence. The rupturing of societal "fault lines" due to falsehoods can also lead to public unrest and violence in extreme cases. For example, in India, a photoshopped image that depicted two men acting indecently atop the Kaaba, a sacred Islamic building in Mecca, led to violence between the Muslim and Hindu community in India. This resulted in one dead and dozens injured [63]. Other images have also been intentionally selected from the Internet to fuel tensions and incite violence between the two communities, including screenshots from movies that were being passed off as actual incidents [64]. In the US, the debunked "pizzagate" conspiracy theory, which alleged that Hillary Clinton and prominent Democrats ran a child sex ring out of a pizzeria's basement, led to shots being fired inside the restaurant and a separate arson incident [65]. It was also discovered that two Russian-linked Facebook groups intentionally organized opposing protests that took place simultaneously outside an Islamic center in Houston, in order to make it appear as a protest and a counter-protest, with the aim to "tear apart" society. Interactions between the two groups eventually escalated into confrontations and verbal attacks [66].

Cause Public Alarm. Disinformation can also cause unnecessary public alarm that could impact people's lives. For example, in China, following the nuclear disaster in neighboring Japan, false rumors that nuclear plumes were spreading throughout Asia and that iodized salt could ward off radiation poisoning were circulated. This led to panic amongst the public and worried shoppers stripped stores of salt throughout the country [67].

2.4.3 Harm to Democratic System

In March 2018, the European Commission's High Level Group on fake news and online disinformation issued a report concluding that disinformation is harmful to democratic values [68]. Facts matter for democracy and people should be well-informed when called upon to make potentially life-changing decisions [6]. Hence arguably one of the biggest threats that disinformation poses is the potential harm to the cornerstones of a well-functioning and democratic society – citizen engagement in public discourse, trust in public institutions, and the right of citizens to have a representative government [2].

Impede Citizen Engagement in Public Discourse. Firstly, falsehoods tend to appeal to emotions and cause people to react with anger, thus making it more difficult to engage in rational debate and lead to ugly and uncivilised public discourse [2]. Secondly, and perhaps one of the most salient dangers associated with disinformation, is that it "devalues and delegitimizes voices of expertise, authoritative institutions, and the concept of objective data – all of which undermines society's ability to engage in rational discourse based upon shared facts" [69]. Lastly, falsehoods can cause citizens to disengage from public discourse if they stop believing in facts altogether [2].

Erode Trust in Public Institutions. Citizen trust has long been viewed as a critical component of a functioning democracy [70]. Forms of disinformation such as conspiracy theories have been found to make people less likely to accept official information [71] and adversely affects trust in public institutions, including those unconnected to the false allegations [70]. This erosion of trust in public institutions thus diminishes their ability to respond effectively to threats and crises, and to govern and introduce policies [2]. For example, in 2016, the Netherlands held a referendum on whether to approve a trade agreement between the EU and Ukraine, during which about two-thirds voted against the agreement. A Ukrainian foreign ministry official suggested that one of the reasons for the failure was that some Dutch voters had the impression that Ukraine was responsible for the shooting down of Malaysia Airlines Flight 17 [72], which killed 193 Dutch citizens. This was a common and proven false theme of Russian propaganda [73]. Another example is the selfie of a Syrian refugee with German Chancellor Angela Merkel, which went viral and became a symbol of Merkel's refugee policy. Unfortunately, the photo subsequently appeared in numerous false stories on social media linking the refugee to terrorist attacks across Europe. This served as a tool for right-wing supporters to claim that the policy had allowed dangerous people into the country [74].

Undermine Citizen's Right to a Representative Government. Informed voting is critical for a government to have legitimacy, for stable and strong governance [2]. Falsehoods have been used to cast doubts on the legitimacy of the outcome of a vote. For example, pro-Russian Internet trolls and automated "bot" accounts allegedly fuelled claims that Scotland's independence referendum in 2014 was rigged, and amplified demands for a revote [75]. A study by researchers from Ohio State University found that fake news probably played a significant role in depressing Hillary Clinton's support during the 2016 US presidential elections, and suggested that about 4 percent of President Barack Obama's 2012 supporters were dissuaded from voting for Clinton as a result [76]. Tweets with images in English and Spanish were also disseminated to encourage Clinton supporters to vote online, by phone, or by text. These were invalid ways of voting and the objective was to depress valid votes in favour of Clinton [77].A separate study by a team from Oxford University on Twitter data found that during the presidential elections, there were more misinformation, polarizing and conspiratorial content than professionally produced news; and the average levels of misinformation were higher in swing states than in uncontested states [78]. The significant amount of disinformation spread over social media made it difficult for voters to be equipped with what they needed to make good decisions [79], thus casting doubt over the legitimacy of the outcome of the elections.

2.4.4 Harm to Individuals

Threaten Health. Disinformation has eroded public trust in vaccination, exposing us all to the risk of infectious diseases. In Aug 2018, the World Health Organization (WHO) said cases of measles in Europe had hit a record high, with experts blaming this surge in infections on a drop in the number of people being vaccinated [43].

Cause Harassment. Individuals may suffer public humiliation or harassment as a result of falsehoods. For example, in Singapore, a false article was circulated alleging that a Singaporean citizen had regretted taking up citizenship and wanted to return back home. This led to him and his family receiving xenophobic messages [80].

2.5 Challenges in Combatting Disinformation

There are several key challenges in combatting disinformation that are difficult to overcome. This include our very own cognitive tendencies, the weakness of truth compared to falsehoods, and the further and faster reach of falsehoods. Exacerbating this problem is the fact that the digital technologies that facilitate the creation of falsehoods is continuously improving, as well as the disruption of the news ecosystem that has lowered the barrier of entry for news sources.

2.5.1 Human Cognitive Tendencies

Although individuals are their own first line of defense against disinformation, most people are unable to fully assess every piece of information due to the significant volume presented to individuals on a daily basis. Instead, people rely on cognitive shortcuts, such as the perceived trustworthiness and attractiveness of the information source, individual past experiences, as well as what others think. As a result, individuals typically do not interpret information in a rational, neutral, and objective manner, but succumb to their own biases when processing information they encounter. For example, people tend to uncritically favor information that confirms their existing beliefs and dismiss counter-attitudinal information regardless of its truthfulness [34].

2.5.2 Weakness of Truth Compared to Falsehoods

Truth is generally weaker than falsehoods due to human psychology and online conditions.

Human Psychology. Correcting disinformation may not work, as the truth could differ significantly from one's ideological preference. In the worst case, it may even strengthen the misperception [81]. In addition, any repetition of misinformation, even in the context of refuting it, can be harmful. This is because of the "illusory truth effect", whereby prior exposure to a statement increases the likelihood that one would judge it to be accurate. Repeated exposure to any falsehood could thus increase its perceived accuracy [82]. As a result, exposure to misinformation can have long-term effects, while corrections may only be short-lived [83]. In some sense, rebutting dubious claims or downright lies only serves to draw attention to the untruth rather than debunk it [6].

Biases Worsened by Online Conditions. Social media has become a powerful tool for sharing and disseminating information online, and has become fertile soil for the spread of hateful ideas and motivates real-life action [84]. This problem is exacerbated by "filter bubbles" and "echo chambers". Social media platforms use algorithms to personalize information flow and results in "filter bubbles", which surrounds users with suggested content that fit their pre-analyzed preferences and avoids divergent content [85], thus reinforcing their worldview in the absence of conflicting ideas [34]. Selective exposure to content is the primary driver of content diffusion and generates the formation of homogeneous clusters, i.e., "echo chambers" [86]. Consequently, these users are disproportionately exposed to like-minded information and that information also reaches like-minded users more quickly [87].

2.5.3 Further and Faster Reach of Falsehoods

Falsehoods Spread Faster Than Truths. The novel content of falsehoods attracts human attention and is a key enabler of its diffusion [88]. A study by researchers from the Massachusetts Institute of Technology found that falsehoods diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, especially for false political news [89]. A separate study also found that in the last few months of the election, the top 20 false news stories had greater engagement on Facebook than the top 20 legitimate news stories [90]. In Michigan, one of the highly contested states of the 2016 US presidential elections, the amount of falsehoods shared "outperform[ed]" professionally researched news [91].

Corrections Lag Behind. Falsehoods generally cause damage long before it is discovered and corrected. For example, it often takes experts hours to conclude if a photo is fake or authentic, during which time, the news would already have disseminated across social media channels around the world [92]. During the 2017 Catalan independence referendum, a fake photograph was posted of police pushing back against demonstrators, under an enormous Catalan flag. Within one and a half hours, a Spanish fact-checking group was able to determine its inauthenticity and tweeted the correction. However, the fake photo was eventually retweeted over 12,600 times while the correction was only retweeted 3,700 times [93]. Hence one should not "expect to counter the firehose of falsehood with the squirt gun of truth" [94].

Corrections Unlikely to Reach Those Exposed. A study on the consumption of fake news during the 2016 US presidential elections found that fact-checks of fake news never reached its consumers [95]. This finding is similar to that of a study on the 2017 France presidential election. It found that the audience for a rumor, that then-Presidential candidate Emmanuel Macron was funded by Saudi Arabia, had almost no overlap with the audience of its debunk [12]. It would seem that misinformed audiences do not seek out corrections, and efforts to provide corrections could merely be "preaching to the choir" [7].

2.5.4 Digital Technologies are Improving Continuously

The digital technologies available to disinformation agents are improving continuously, thus making the problem increasingly difficult to tackle. In front of a joint session of several US senate committees, Facebook CEO Mark Zuckerberg said that his company was in an "arms race" with operators seeking to exploit the social network [96]. In July 2018, Facebook removed more than 30 accounts and pages believed to have been set up to influence the US mid-term elections [97]. They added that the perpetrators went to much greater lengths to obscure their true identities than the Russian-based Internet Research Agency did to disrupt the 2016 US presidential elections [98]. Facebook's chief security officer remarked that this was one of the fundamental limitations of attribution, and that organizations would seek to improve their techniques once they have been uncovered [97].

2.5.5 Disruption of the News Ecosystem

Fundamental disruptions to the news ecosystem have facilitated the spread of disinformation in several ways.

Lower Barrier of Entry. With the advent of digital platforms, users globally are able to provide commentary on news topics, contribute to the production of news, and even disseminate them. This has enabled participation of individuals as both consumers and producers of news [8]. Algorithm-driven news distribution platforms have also contributed to reduced market entry costs and widened the market reach for news publishers and readers [88]. This lower barrier of entry has enabled non-professional sources of news, such as individuals whom are not subjected to editorial oversight, to disseminate information and gain popularity on social media. Some users also exploit the anonymity of the Internet to share information via their social media accounts without thinking about the potential consequences [2]. Social media platforms have thus facilitated the spread of the "most engaging" content through their algorithms, enabling the "speed of news transmission to take precedence over the truth" [8].

Expectation of Real-Time Updates. Historically, journalists would fact-check their work and have their work reviewed by experienced editors. However, this has little value in the current landscape as if one publishes late, one might as well not publish at all [69]. The unprecedented connectivity of the Internet has led the public to expect real-time updates of news, possibly at the expense of robust verification. This has unfortunately facilitated the spread of rumors and conflicting information [2].

Disruption of Business Model. The business model of newspapers has been disrupted as advertising revenue is being diverted to social media platforms, which provide news aggregation and digital advertising. This has put a strain on the financial resources of traditional media organizations [2]. Online platforms such as Google and Facebook have essentially "grown rich by using technology to impoverish" traditional media organizations [6].

This page is intentionally left blank

Chapter 3: Understanding the Current Landscape

The previous chapter discusses in detail the threat of disinformation and the key challenges in combatting it. This chapter focuses on using ten "unique lenses" to look at the enterprise that is key to the proposed system, as well as begin applying the process of the ARIES framework [10]. The objective is to provide a better understanding of the current landscape, the stakeholders and current architecture.

3.1 System and Enterprise Definition

A *system* can be defined as a set of entities and their relationships, whose functionality is greater than the sum of the individual entities [9]. With this in mind, a key component of the *system* – that would be designed to counter the threat of disinformation – is a collective group of organizations at the national-level, each tasked with distinct roles and responsibilities. This collective group would be the national-level *enterprise* that would deal with the threat of disinformation. Key entities of this enterprise would include the public, the government, online platforms, traditional media organizations, fact-checking organizations, educational and research institutes, and other non-governmental organizations; all of which have an important role to play in preventing the spread of disinformation. The following paragraphs in this section attempt to discuss the reasons for their importance as part of this national-level enterprise.

3.1.1 The Public

Every single person plays a crucial part in this ecosystem. Every time someone passively accepts information without double-checking, or shares a post, image, or video before verifying it, that person is adding to the noise and confusion [24]. Hence it is crucial that people are better equipped with the necessary skills to become their own first line of defense against disinformation.

3.1.2 The Government

Logically, governments would have to be part of the enterprise as they have a duty to protect their citizens from fake news [68], in order to better advance their citizens' interests and resolve social issues through informed public deliberation [8]. The 2018 Reuters Institute Digital News Report also found that people were generally supportive of government intervention to tackle disinformation, especially in Europe (60%) and Asia (63%). By contrast, only 41% of Americans thought that the government should do more, perhaps because of a strong commitment to the First Amendment and freedom of speech [99]. The government would comprise the legislative, executive, and judicial branches, including the relevant government agencies.

3.1.3 Online Platforms

Chapter 2.2 explains how the emergence of the Internet and its associated technologies have made disinformation a more potent problem than ever before. Besides social media, disinformation can also spread via other platforms such as search engines and instant messaging [77]. Hence it is only logical that the online platforms play a role in this enterprise.

Search Engines. Search engines are the quintessential tool of the Internet, yet it has been shown that search ranking bias can be undetectable and shift the voting preference of undecided voters by 20% or more. These companies thus have the power to influence the results of elections [100] and undoubtedly need to participate in the solution against disinformation. In the aftermath of the 2017 mass shooting in Las Vegas, shortcomings in Google's algorithm meant that a search for the name of the man wrongly accused of the shootings, would bring up the false claim to the prominent "Top Stories" box at the top of the search results page [101]. This is yet another piece of evidence that search engines can spread disinformation if left unchecked.

Social Media Platforms. Studies show that more than 50% of people get their news from social media these days, and in the United States, this figure is more than 60% according to the Pew Research Center [6]. Significant content does not even masquerade as news but appear as memes, videos, and social posts on social media platforms such as Facebook and Instagram. These platforms play a significant role in encouraging the sharing of information because they are designed to be performative in nature. Slowing down to check whether content is true before sharing it, is far less compelling than reinforcing to one's "audience" that you love or hate a certain policy. The business model for these platforms is based on this idea because it encourages one to spend more time on their site [4]. Facebook's product manager for civic engagement also admitted in a blog post that social media can be used to "spread misinformation and corrode democracy"

[102]. It is evident how the advent of social media has drastically changed the news landscape and these social media giants are a key stakeholder to engage.

Closed Messaging Platforms. Much of India's false stories are spread through WhatsApp. This include rumors of salt shortages in November 2016 that led to a rush for salt in four Indian states, and a rumor about vaccines that thwarted a government immunization drive [103]. A survey conducted in Singapore found that 50% of respondents came across fake news via WhatsApp, slightly more than via Facebook (46%) [104]. Hence closed messaging platforms have a key role to play as well.

Online Video Sites. Google discovered that tens of thousands of dollars were spent on advertisements by Russian agents, in an attempt to influence the 2016 presidential elections via their products, including YouTube. Google runs the world's largest online advertising business and YouTube is the world's largest online video site. This discovery is significant because the advertisements were not from the same Kremlin-linked troll farm that bought advertisements on Facebook – a sign that the Russian disinformation effort was a much broader problem than initially thought [105].

Discussion Forums. Burgeoning discussions about the previously mentioned "pizzagate" conspiracy theory were found on online bulletin boards such as Reddit and 4chan. Reddit even closed one of its site forums on the conspiracy, as it was revealing private information about people from the restaurant [106]. These online discussion forums that allow people to post stories and comments are also a potential source of disinformation.

3.1.4 Traditional Media Organizations

Traditional media organizations used to safeguard the accuracy of information in the public sphere through their rigorous checking mechanisms and editorial oversight. Although their influence has waned due to the rise of alternative news sources [2], there is still relatively higher trust in "legacy printed and broadcasted news publishers" than "algorithm-driven news distribution channels such as aggregators and social media" [88]. For example, the 2019 Edelman Trust Barometer Global Report found that more people trusted traditional media (65%) as compared to social media (43%) [107]. The 2018 Reuters Institute Digital News Report also found that 75% of respondents still believed that traditional media organizations should do more to separate what is true and false on the Internet [99]. It is evident that there is still a role to play for these organizations.

3.1.5 Fact-Checking Organizations

Although fact-checking is a resource intensive endeavor [108], studies have shown that fact-checking is able to influence people's assessments of the accuracy of political advertisements [109] and is thus critical to combatting disinformation. As media practitioners have highlighted that hiring more resources for increased fact-checking is a key challenge for them [108], there is a growing need for independent fact-checking organizations to provide support in this area. In fact, Facebook launched a fact-checking program in December 2016 and now has 54 fact-checking partners working in 42 languages [110]. It is evident that these organizations are a critical part of the enterprise.

3.1.6 Educational and Research Institutes

Universities, and educational institutes in general, are ideally placed to help students and the wider public decide on the authenticity of news. For example, universities aim to build students' critical skills for academic work, but these skills have more relevance than merely checking one's references or the authority of the information when preparing a piece of work, as these same skills can be applied to authenticating pieces of news [111]. The roles of educational institutes can even be extended to build up news literacy amongst the public. This would enable people to better separate fact from fiction, potentially limiting the spread of false information [99]. Research conducted in universities or independent research institutes can also help develop technologies to combat the spread of disinformation.

3.1.7 Non-Governmental Organizations (NGOs)

Think Tanks. Think tanks can serve as watchdogs and provide advice or ideas to tackle disinformation. For example, the European Values Think Tank based in the Czech Republic has a program called "Kremlin Watch", which "aims to expose and confront instruments of Russian influence and disinformation operations focused against Western democracies" [112].

Grant-Making Foundations. Initiatives to combat disinformation – whether it is to develop new technologies, perform fact-checking, or enhance media literacy – would require sources of funding. Hence grant-making foundations could play a role to provide the needed funding. For example, Facebook's News Integrity Initiative, which aims to improve news literacy, is funded by Facebook and various foundations including the Craig Newmark Philanthropic Fund, the Ford Foundation, and the Democracy Fund [113].

3.2 External Landscape: Ecosystem Factors

An analysis of the external factors that would influence the enterprise's ability to combat the spread of disinformation is summarized in Table 2. The factors are ranked according to their impact to the enterprise. These exogenous factors shift over time in response to a changing world and can affect the enterprise positively or negatively [10]. Subsequent paragraphs provide an elaboration of each of these factors.

Table 2: Analysis of external landscape - ecosystem factors.

Factors	**Description**	**Impact of Change**	**Level of Change**	**Pace of Change**
Technology	▪ Emerging low cost, high impact technology such as deepfakes [33] [114] ▪ Increasingly sophisticated bots [115] [22]	High	High	High
Regulation	▪ Increase in regulations can reduce the spread of disinformation but could also impact free speech	High	High	High
Competitive	▪ Technology companies in an "arms race" with operators seeking to spread disinformation [25], e.g. bot detection and creation ▪ Lower barrier of entry [2] and reduced market entry costs for news publishers [88]	High	High	Moderate
Market	▪ Most people feel that more should be done to combat disinformation [99] ▪ Online disinformation campaigns have evolved into a profitable industry [2]	High	Moderate	Moderate
Economic	▪ Financial resources of media organizations are strained as business model is disrupted [2] ▪ Economic downturns may reduce funds available for combatting disinformation	High	Moderate	Moderate

Geo-Political	▪ Erosion of democracy due to lack of trust in facts [92] ▪ Increased tensions between nations could lead to an increase in disinformation campaigns	High	Moderate	Low
Resource	▪ Fact-checking is resource intensive and manpower availability is an issue [108]	Moderate	Moderate	Low

Technology. There is the looming security threat of "deepfakes". Within a few years, we may be watching moving images and speeches without anyone being able to tell whether they are real or fabrications [33]. Research organization Future Advocacy illustrated the potential of deepfakes to undermine democracy, by creating a fake video of rivals Boris Johnson and Jeremy Corbyn endorsing one another for Britain's prime minister [116]. The low cost of using such technology is also worrying – a New York Times reporter managed to create a "deepfake" using a free application downloaded online and by renting a Google Cloud Platform to provide the required processing power for only US$86 [114]. Bots are becoming increasingly sophisticated and some can engage in more complex types of interactions, such as having conversations with other people, commenting on their posts, and answering their questions. Such bots can easily infiltrate a population of unaware humans and manipulate them to affect their perception of reality, with unpredictable results [115]. Such bots have been termed as "cyborg" accounts, as they combine automation with elements of human curation to appear as legitimate accounts [22].

Regulation. An increase in regulations would certainly help to reduce the spread of disinformation but could come at the expense of freedom of speech if they are overly restrictive. For example, in Jan 2018, Germany announced that it would begin to enforce a law, the Network Enforcement Act, which requires social media sites to remove hate speech and fake news within 24 hours or face severe fines of up to €50M [68]. The law was introduced to "enable prompt, effective action" as the online platforms were not removing illegal posts within 24 hours, despite committing to do so [117]. On the other hand, First Amendment protections for false speech present a key obstacle for introducing similar regulations in the US [118].

Competitive. In front of a joint session of several US senate committees, Facebook CEO Mark Zuckerberg said that his company was in an "arms race" with operators seeking to exploit the social network [96]. For example, bot detection will always be a cat-and-mouse game in which a large, but unknown, number of human-like bots may go undetected. Any success at detection, in turn, will inspire future countermeasures by bot producers. Identification of bots will therefore be a major ongoing research challenge [119]. The barrier of entry for sources of news have been lowered as social media platforms have become a popular source for information [2]. Algorithm-driven news distribution platforms have reduced market entry costs and widened the market reach for news publishers and readers [88].

Market. People are highly concerned about disinformation in the news and most feel that more should be done to combat it. Across the 23 markets surveyed in the 2018 Reuters Institute Digital News Report, 75% felt that media companies should do more to separate what is true and false on the internet. The figure was slightly lower (71%) for online platforms such as Facebook and Google, while 61% believed that the government should do more [99]. These figures show the high demand for more initiatives to combat disinformation. At the same time, Section 2.2.2 highlights the widespread availability of tools and service that allow even the ordinary citizen to engage in disinformation campaigns. This illustrates how online disinformation campaigns have evolved into a profitable industry [2].

Economic. The business model of newspapers has been disrupted as advertising revenue is being diverted to social media platforms, thus putting a strain on the financial resources of traditional media organizations [2]. Economic downturns may reduce funds available for combatting disinformation, and thus could affect the development of capabilities if they are not prioritized.

Geo-Political. Computer generated video and audio recordings could undermine credibility of all recordings. Viewers may no longer believe anything they read, hear, or see online. Such lack of trust in facts could erode democracy [92]. Any increase in tensions between nations could also lead to a corresponding increase in disinformation campaigns.

Resource. Established news organisations are experiencing the pressure of having to respond quickly to falsehoods, when verifying and cross-checking information online is in fact a heavily resource-intensive one [108]. Hence the availability of manpower in the workforce is a source of concern.

3.3 Internal Landscape: Eight View Elements

An analysis of the current enterprise by looking at it from eight different view elements [10], or perspectives, is summarized in Table 3. Subsequent paragraphs provide an elaboration of each of these factors.

Table 3: Analysis of internal landscape - eight view elements.

Element	**Description**
Strategy	Overarching strategy of many governments, including the US and European Union, is to avoid over-regulating the industry [88] and to encourage self-regulation by online platforms as a start [120]
Information	Access to data have been closely-guarded by online platforms [121] and more data needs to be shared to address the threat of disinformation [122]
Infrastructure	Secured means to exchange sensitive data and information between entities would be needed
Products	Fact-checking websites to debunk falsehoods and release corrections, and tools to allow public to report false stories
Services	Preventing public exposure to falsehoods, and providing educational programs to enhance media literacy of the public
Process	Depending on the type of legislation available in the country, curbing the spread of disinformation could range from imposing fines, ordering the removal of false information, use of anti-defamation laws [123], or merely self-regulation by online platforms; internal bureaucratic processes of a country may also impede the introduction of effective countermeasures
Organization	Loosely coupled organization that would require joint committees to facilitate collaboration between the entities
Knowledge	Possess competency in analyzing text but lacking similar ability for images or videos [4]

Strategy. Many governments, including the US and the European Union tend to favour against over-regulating the industry. A European Commission report concluded that government regulation of disinformation can be a "blunt and risky instrument" [88], and has recommended a self-regulatory approach as a start [120].

Information. Access to data has been "zealously guarded" by online platforms [121] and those outside of the company do not have access to the data unless the company chooses to release it [124]. The amount of data currently being shared with researchers has been deemed inadequate [12] [125] and there have been recommendations to share more data with trusted fact-checking organizations and academia to address the issue of disinformation [122].

Infrastructure. Being a group of entities, secured means of communication or a well-defined process to exchange any sensitive data and information between the parties is needed.

Products. There are several existing fact-checking websites that serve to determine the veracity of news and provide corrections. Prominent examples include Fact Check, Politifact, and Snopes [126]. The public are also equipped with tools to report falsehoods they encounter. For example, Facebook has made it easier for users to report a false story for further analysis and checks [113].

Services. A key service provided by the enterprise is to prevent people from being exposed to falsehoods in the first place. Examples of such behind-the-scenes work include Facebook's initiative to show false content lower in the news feed to reduce the number of people exposed to it [127], and removing accounts linked to "coordinated inauthentic behavior" to prevent the manipulation of users [128]. There are also ongoing educational initiatives aimed at improving media and digital literacy at the national-level, as highlighted in parliamentary reports from the UK and Singapore [2] [124]. An important part of such literacy is the understanding of how social media works, and how algorithms are responsible for the content that is provided to each user [124]. However, there is significant room for improvement in this area, as the 2018 Reuters Institute Digital News Report found that only 29% of respondents were aware that computer algorithms were responsible for the news shown to them on social media [99]. In addition, more than 55% of the sample across 38 countries in the 2019 Reuters Institute Digital News Report remain concerned about their ability to separate what is true or false on the Internet [129].

Process. The process of curbing the spread of disinformation is dependent on the type of legislation available in the country. In countries such as Germany, Israel, and Russia, where there is new and more focused legislation, platforms can be ordered to remove false information and even be imposed with fines. Other countries such as the UK, Canada, and Japan, rely on existing laws to regulating the media or anti-defamation laws [123]. In the US, online platforms are given some degree of freedom and allowed to self-regulate. In a surprising turn of events, in March 2019, Facebook's CEO even invited the US Congress to regulate his company [130], calling for a "more active role for governments and regulators" [131]. The internal bureaucratic processes of the country could also impede the introduction of effective countermeasures. For example, it has been said that "bureaucratic inertia" left the US vulnerable to Russian interference in the 2016 US presidential elections. Although the US saw some warning signs of Russian meddling in Europe and the US, they didn't appreciate the dangers and struggled to develop an effective response, due to its domestic politics and a "misguided belief in the resilience of American society and its democratic institutions" [35].

Organization. Being a loosely coupled group of entities, the enterprise will require joint committees involving the various partners, to facilitate collaboration and provide regular updates on the progress of the initiatives.

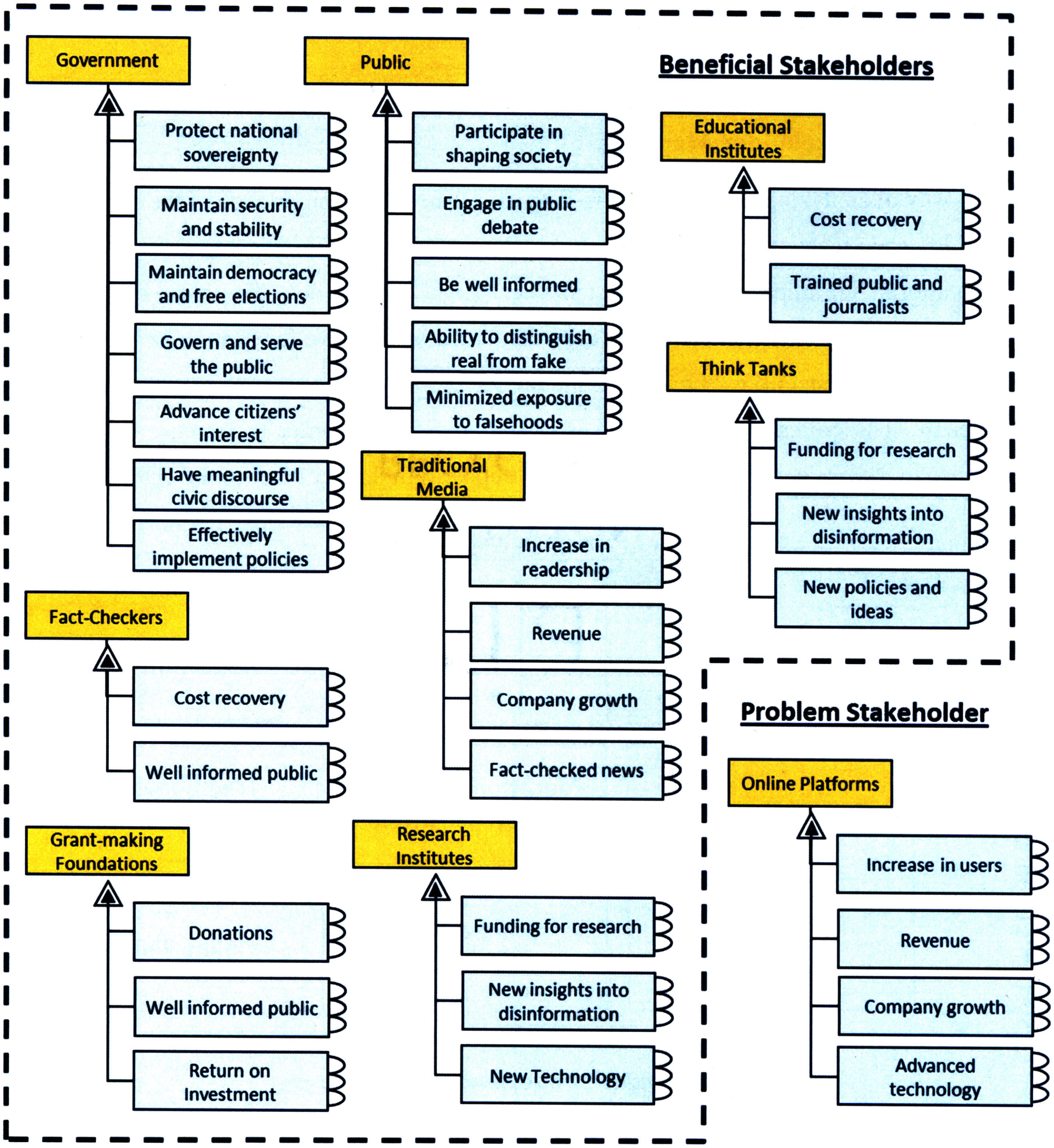
Knowledge. Currently, techniques for studying text such as natural-language processing are far more advanced than techniques for studying images or videos memes. Unfortunately, the most effective misinformation is that which will be shared, and memes tend to be much more shareable than text. In fact, during the 2016 US presidential elections, many of the Russian-created posts and advertisements on Facebook were memes [4]. This is one area where the capability and competency needs to be further developed.

3.4 Stakeholder Analysis

3.4.1 Stakeholder and Needs

A *stakeholder* is defined as one that has a stake in the system, and could be both internal or external to the enterprise. Stakeholders can also be distinguished into *Beneficial Stakeholders*, who both receive valued outputs from the system and provide valued inputs to the system; as well as *Problem Stakeholders*, whom you require something from but there is nothing they need that you can provide in return [9]. An analysis of the stakeholders and their key needs is shown in Figure 2.

Figure 2: Stakeholders and their key needs.

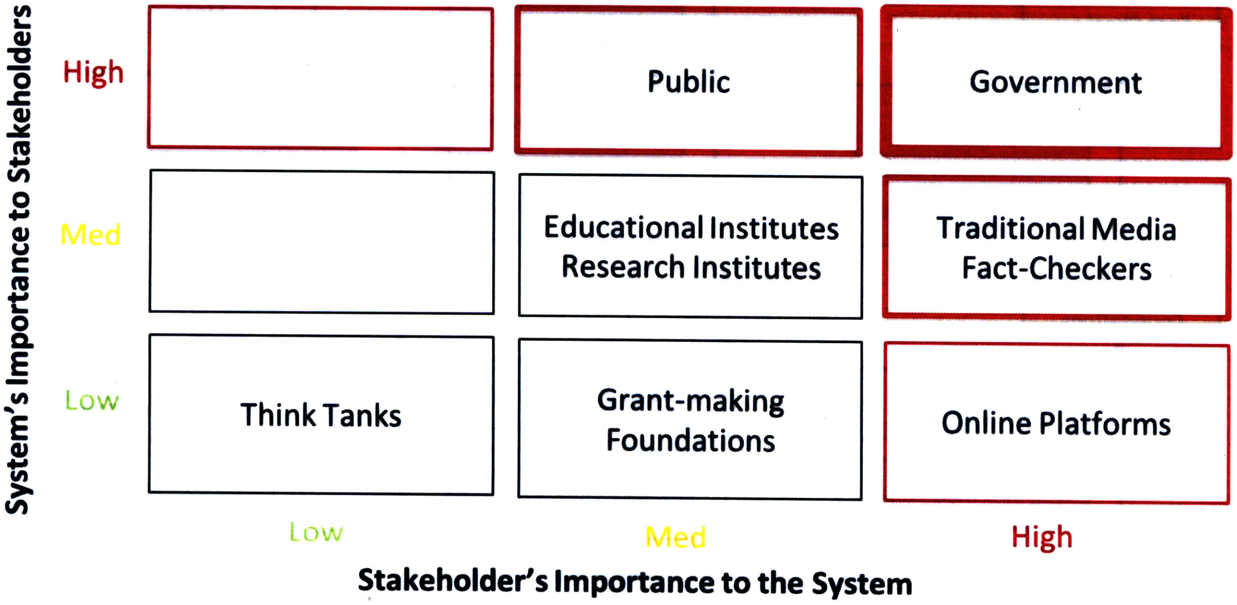


A key observation is that a critical part of the enterprise – the online platforms – have been classified as a problem stakeholder because there is little to offer them in exchange for their commitment to making the system work. This is not surprising as the business models of these companies rely on revenue coming from the sale of advertisements, which desire user attention or engagement. Users are typically drawn to negative or sensational content, and because profits are critical to the online platforms, such content will always be prioritized by the platforms' algorithms. This would facilitate the spread of disinformation [124] [132]. Hence, this relationship with the online platforms needs to be altered and addressed.

3.4.2 Stakeholder Mapping

Another way of analyzing the stakeholders is to map the relative importance of the system to the stakeholder, and vice versa. This is illustrated in Figure 3. Once again, the key observation is that online platforms are important to the system (i.e. "High"), given how they are a key source of falsehoods. However, as explained previously, there is little incentive for them to be committed to the success of the system (i.e. "Low"). Online platforms thus fall into the lower right hand corner of the chart. It is critical that something is done about this, in order to increase the system's importance to them (i.e. moving them towards the upper right hand corner of the chart).

Figure 3: Stakeholder mapping.



System's Importance to Stakeholders

High		Public	Government
Med		Educational Institutes Research Institutes	Traditional Media Fact-Checkers
Low	Think Tanks	Grant-making Foundations	Online Platforms
	Low	Med	High

Stakeholder's Importance to the System

A brief explanation of why the other stakeholders are mapped as such is provided in the subsequent paragraphs.

The Government. The government has the most at stake in the system (i.e. "High") as it is their duty to maintain national sovereignty and democratic stability, as well as to govern and serve the public. At the same time, they also play a critical role in the system (i.e. "High") as the government is able to introduce legislation and policies that can enhance the nation's defense against disinformation campaigns.

Traditional Media and Fact-Checkers. It is critical that the public is given a reliable source of information, and access to corrections of falsehoods. Hence these two entities play a very important role to the system (i.e. "High). However, traditional media organizations may have other priorities and needs – such as being the first news agency to report a breaking story, and to maintain the financial viability of the company. Fact-checkers may also be involved in multiple other projects and thus may not be able to provide their full attention to the system. Consequently, the importance of the system to these two entities was deemed as "Medium".

The Public. This system is critical (i.e. "High") to the public as it would ensure that they are adequately protected from disinformation campaigns. However, the system cannot be fully reliant on the public to stop the spread of disinformation (i.e. "Medium"), as even the most well-trained citizen could be prone to mistakes and inadvertently share false information, which could then propagate quickly via trusted peer-to-peer sharing.

Educational and Research Institutes. These two organizations have a "Medium" importance to the system as they serve to educate the public and to develop new technologies respectively. At the same time, given the other priorities and commitments of these entities, the system is also deemed to be of "Medium" importance to them.

Grant-Making Foundations. Funding sources are important to the system (i.e. "Medium") as investments are needed to create the technological solutions needed. However, as these foundations have limited funds and could be involved in other projects, the system could be of "Low" importance to them.

Think Tanks. Proposed changes to existing policies and new ideas may be useful to the enhancement of the system, but not all think tanks have the required clout at the national-level to influence policies. Hence they are deemed to be of comparatively "Low" importance to the system. As think tanks may have other ongoing research areas, their interest in combatting disinformation may also be relatively diluted (i.e. "Low").

This page is intentionally left blank

Chapter 4: Overview of Possible Countermeasures

This chapter provides an overview of possible countermeasures that are being proposed or have been adopted by countries worldwide. The objective is to use these countermeasures to develop an integrated approach, or system architecture, from the ground-up. Using a "2-down-1-up" [9] approach[2], each individual countermeasure is analyzed before grouping, aggregating, and abstracting them to the various categories (which form the headings of this chapter). The principle of "7 ± 2"[3] is also applied to keep the number of headings (and sub-headings) manageable [9]. These countermeasures include both short and long-term efforts, as well as the involvement of multiple stakeholders.

4.1 Fact-Checking

Promoting fact-checking initiatives that correct falsehoods in a timely manner, and inform the public of the facts is critical. Despite the limitations of correcting falsehoods that Sections 2.5.2 and 2.5.3 discuss, leaving them uncorrected is not a viable option. This is because falsehoods become harder to debunk the longer it goes unchallenged, and the more likely people will start to believe it [133]. Hence fact-checkers are an integral element in the value chain, and a strong and independent network of fact-checkers is essential for a healthy digital ecosystem [122]. It is also critical to restrict access to falsehoods once they have been detected to curb their spread.

Forming a Coalition. Fact-checking initiatives can be run by dedicated fact-checking organizations, traditional media organizations, volunteers or community-driven, and even by governments. However, it is worthwhile to form a fact-checking coalition involving different parties such as traditional media organizations, online platforms, and fact-checkers, whereby they can work collaboratively to optimize manpower resources and prevent duplication of efforts [2] [12]. Examples of such coalitions include StopFake, First Draft, and the International Fact

[2] This principle of system decomposition states that the decomposition at Level 1 cannot be evaluated until one descends to Level 2, i.e. one level deeper.

[3] This principle states that the human brain can only reason about a finite number of things simultaneously, while remaining able to understand their interaction. This manageable number is conventionally thought of as seven +/– two.

Checking Network [134]. Facebook launched a fact-checking program in December 2016 and now has 54 fact-checking partners working in 42 languages [110].

Public Reporting. The public can also be provided with tools or portals to report disinformation. This could be an effective means to identify falsehoods, as a study, co-authored by an MIT professor, found that crowdsourced judgements about the quality of news sources may help to identify false news stories [135]. For example, Italy developed a portal that allows the public to report falsehoods for investigation by law enforcement, in the lead up to the next election [136]. Facebook has also made it easier for users to report a hoax, in a bid to leverage community efforts to detect more falsehoods [137].

Tagging Falsehoods and Sources. Once an article has been determined to be false, it can be tagged to notify and warn people. Facebook claimed that tagged articles received 80 percent fewer "impressions", while an independent research by Dartmouth College found that the percentage of readers believing a piece of false news was reduced by 13 percentage points when it was tagged as "false" [138]. Another benefit of tagging is to potentially decrease the sharing of falsehoods, as one may feel embarrassed sharing news that is perceived to be false [83]. Websites known to carry falsehoods could also be tagged to warn visitors to the website [2]. While tagging false stories have shown to moderately reduce the perceived accuracy of false headlines, a potential downside is the implied truth effect – whereby false headlines that fail to get tagged are considered validated and thus are seen as more accurate [139].

Reducing Access. It is important to find means to "dilute the visibility of disinformation" [122] and limit the circulation of falsehoods. In Facebook's system, fact-checking partners rate content with labels. Facebook would then downgrade the distribution of certain items and cause them to appear lower in users' News Feed [110] [137]. Google also opts to demote disinformation to reduce the chance of people seeing it, rather than removing it altogether [132]. Facebook will also prompt users if they insist on sharing disputed articles [137]. For closed messaging platforms, one possible solution would be to prevent the forwarding of proven falsehoods [2].

Removing Access. It is also important to remove or block access to the falsehood to prevent further exposure. This could include aggressive enforcement of terms of service agreements with Internet providers and online platforms, to interfere with the ability of disinformation agents to broadcast their messages [94]. Countries such as Brazil, Germany, and South Korea have established or are proposing laws that require online platforms to take down illegal content or face steep fines [136]. The Indian government has disconnected Internet connectivity more than 100 times within a year to stop the spread of rumors via WhatsApp, while Indonesia blocked access to various social media features in May 2019 following violent riots that occurred after the presidential elections [140]. A wider range of information warfare capabilities, such as jamming, could also be employed [94]. In Ukraine, special technology has been installed to reduce the strength of TV signals from a foreign state [2]; access to social media networks owned or linked to Russia has also been blocked as part of sanctions [141].

Broadcasting Corrections. A correction needs to be circulated with "equal vigor", i.e. equal coverage, to prevent the influence of the falsehood from persisting [142]. Individual notifications could also be sent to people exposed to falsehoods. For example, Twitter decided to send emails to notify almost 680,000 Americans, who were exposed to Russian disinformation efforts during the 2016 US presidential elections [143]. Facebook adds fact-checkers' articles to a controversial story's "Related Articles" to give people more context about a story, including what information in the story is false [144]; and notifies users if they try to share or have previously shared a false story [110]. Facebook would also notify specific people that had been targeted by "sophisticated attackers", and even proactively notify people who have yet to be targeted but are at risk of being so [145]. Corrections could be broadcast over multiple platforms [2] to ensure greater reach.

Watchdog and Challenge Disinformation Narratives. Think tanks and other NGOs should watchdog and scrutinize institutions and politicians that lobby on behalf of foreign powers. Any disinformation narratives should also be challenged publicly and done on a regular basis [37]. For example, France is expanding the obligations of media watchdogs to scrutinize non-governmental institutions [136]. A counterpart to organizations such as Global Witness and Transparency International could also investigate disinformation campaigns [3]. As many of these organizations are underfinanced, Governments and grant-making foundations should contribute funds to support their work [37]. Traditional media organizations should also "call out fake news and disinformation" by relying on their in-house professionals and fact-checkers [146].

4.2 Research and Technology (R&T)

Section 2.5.4 discusses how the digital technologies available to disinformation agents are improving continuously, and there is a constant "arms race" between technology companies and disinformation agents. Hence there is a need to continually invest and develop new technologies to counter the threat of disinformation. To better utilize available resources, there should be increased collaborations between academics and practitioners to identify key focus areas for R&T development.

Automating Fact-Checking. As there are more than one billion pieces of content posted everyday just on Facebook, it would not be possible for fact-checkers to review every single article [147] that is shared online. Hence there is a need to build tools to support fact-checking and verification of content [12]. There is significant research being done to automatically assess the credibility of articles through various means – data-driven approaches using machine learning, model-driven approaches using multiple criteria to estimate an overall credibility, and graph-based approaches that analyze the relationships between various actors or entities [148]. For example, Facebook has used machine learning to identify duplicate articles that had been rated as false by fact-checkers [147]; while an independent fact-checking initiative, Full Fact, uses natural language processing and statistical analysis to automatically fact-check claims [34]. In Singapore, DSO National Laboratories aims to develop an artificial intelligence system that could determine the authenticity of stories using crowdsourcing, content analysis, and source profiling [149]. However there is more work to be done in this area, as contents are usually only labelled false after a story has already gone viral and the damage done [150]. Hence there is a need for better and more advanced "real-time monitoring" systems to identify falsehoods early [2].

Identifying Sources. Besides identifying false content, there is also a need to enhance investigative capabilities to trace and identify the source of the false content for accountability [2] [122]. Facebook has also attempted to use machine learning technology to identify pages that are "likely" to spread falsehoods based on past behavior (e.g. frequently copying and pasting content from other sources) and other signals (e.g. websites covered in low-quality advertisements and page administrators that target people in other countries). Facebook would then reduce the "reach" of such pages [151].

Removing Inauthentic Accounts. The ease of creating and maintaining fake accounts is a major enabler of disinformation [121]. A popular tactic used by disinformation agents is the use of bots or the sophisticated coordination of passionate supporters and paid trolls, or a combination of both, to make it appear that a person or policy has considerable grassroots support [4] with numerous unrelated individuals affirming the information [77]. Some actors even use sophisticated means to mask the fact that the accounts are part of a coordinated operation, including redirecting through "proxies" to hide their original location [152].

- Closing Fake Accounts. It is thus important that online platforms "intensify and demonstrate the effectiveness of efforts to close fake accounts" [122]. In response to this growing threat, Facebook has employed analytical techniques, including machine learning, to identify and remove inauthentic accounts [145].

- Authenticating Accounts. Another option would be to strengthen accountability of users by reducing the ability of disinformation agents to remain anonymous on the Internet. This can be achieved by conducting user authentication and encouraging content creators to digitally sign and verify their content [2]. This approach could also help prevent the spread of fake video and audio clips, especially if social media platforms amend their algorithms to favor and promote only verified and signed content [8]. Legislation to compel online platforms to authenticate accounts and their origins, as well as remove inauthentic accounts, would help limit the influence of malicious actors [121].

- Labelling Bots. On a related note, online platforms should also be obligated to label bots that they provide or are maintained on their platforms, in order to prevent the use of bots in amplifying disinformation [121]. These bots should be clearly marked and rules established to ensure that their activities would not be confused with human interaction [122].

Improving Image, Video, and Audio Analytics. Much of the focus thus far has been on text-based disinformation. However, fabricated or manipulated visual content are more pervasive than textual falsehoods [12]. This is because the most effective misinformation is that which will be shared, and memes tend to be much more shareable than text, as was the case during the 2016 US presidential elections. Unfortunately, techniques for studying text such as natural-language processing are far more advanced than techniques for studying images or videos memes [4]. Fabricated audio is also expected to increasingly become a problem. To address these emerging threats, water-marking technologies could be used to authenticate original material [12].

Enhancing Algorithms. Two-thirds of news consumption online is via algorithm-driven platforms such as search engines, news aggregators, and social media [88]. Hence algorithms essentially decide what information people should consume [125], by making millions of "editorial decisions" and systematically targeting users to receive specific content [153]. Consequently, a powerful tool to manipulate and amplify falsehoods is the use of bots that leverage trending algorithms of the platforms themselves to disseminate information automatically [154]. In an ideal situation, people would still be free to express what they want, but information that is designed to mislead, incite hatred, or cause physical harm would not be amplified by algorithms and allowed to trend on online platforms [4]. To help address this, Google has amended its algorithms to "surface more authoritative pages and demote low quality content" to enhance the quality of its search results [155]. However, more needs to be done to better address this issue. Hence it is important that online platforms be more transparent about algorithm changes, and allow their impact to be independently measured and assessed by the larger community to prevent unintended consequences [12].

Sharing Data for Research. The lack of good datasets and the difficulty of collecting and mining large volumes of data, have been cited as a key challenge in ongoing research initiatives aimed at automatically assessing the credibility of articles [148].

- Online Platform Data. Online platforms should collaborate closely with trusted partners in the ecosystem to share data and co-develop advanced solutions. Anonymized platform data should be made accessible to allow independent researchers to analyze and monitor disinformation dynamics and their impact to society. This would enable the early identification of potential problems and misuses on platforms, and can be done while still respecting user privacy and intellectual property [121] [122]. For example, one of Facebook's third-party fact-checking partners, UK-based charity Full Fact, issued a report in July 2019 urging Facebook to provide more data on how flagged content is shared over time, to see how quickly false information is spreading and assess how fact-checks are containing the spread, so as to better curb the spread of disinformation [110]. This would also enable researchers to evaluate the effectiveness of measures implemented by online platforms to better enhance the "integrity of public communication spaces" [12].

- Online Platform Algorithms. Online platforms should allow third parties to monitor and report the effects of their algorithms. This is not about releasing their algorithms but merely the results of their algorithms. Independent monitoring is critical as editorial decisions of such algorithms are opaque and can take weeks for someone to find out what has been disseminated by the algorithms [153]. This would also enable outputs of the algorithms to be evaluated for efficacy and fairness, and possible hidden biases [121].

- Government Data. Federal governments typically maintain significant volume of data, in multiple domains across their various agencies, that may be valuable for researchers. Hence it has been proposed that federal datasets be opened up to university researchers and qualified small businesses or start-ups [121].

Dedicating Resources and Support. In order to conduct research and develop new technologies, sufficient resources and support are required from various stakeholders.

- Online Platforms. Online platforms have profited significantly from their business model and should shoulder some responsibility for preventing the spread of disinformation via their platforms. For example, more than a billion dollars were spent on digital political advertisements during the 2016 US presidential elections [156]. Hence it would seem appropriate that online platforms reinvest some of their profits into R&T development.

- Governments. With a duty to protect their citizens, Governments should facilitate the development of technologies by start-ups and other companies that can "ensure the integrity" of the "online information ecosystem" [2]. They should also support research into how disinformation is created and spread across the internet [124]. For example, the US National Science Foundation has provided funding for projects such as ClaimBuster, which leverages natural language processing techniques to identify factual claims [134]. There is also a need to set aside sufficient resources for cyber forensics and intelligence agencies to enhance their capabilities, such as the ability to trace the origins of falsehoods

[2]. This can include a dedicated budget line to firmly put such efforts on the foreign and security policy agenda [37].

- Grant-Making Foundations. Another source for funding are grant-making foundations, which can lend support to start-ups to design, test, and innovate solutions [12].

4.3 Business Model

The engagement-driven advertising business model of online platforms tend to promote contentious or emotionally-charged articles, while ignoring the authenticity of the content. This has contributed to the fast-paced dissemination of falsehoods [8]. Section 2.2.1 discusses how targeted advertising can be used to send advertisements to specific users, and such tailored messages can significantly influence individuals. The stakeholder analysis in Section 3.4.1 also highlights that the business models of online platforms rely on revenue coming from the sale of advertisements, which may indirectly promote the spread of disinformation. Hence it is critical that measures related to the digital advertising industry are put in place, to disrupt the financial incentives of disinformation agents and ensure that the digital advertising industry does not inadvertently provide incentives to distribute falsehoods.

Disrupting Financial Incentives. There is a need for online platforms to "reduce revenues for purveyors of disinformation" [122]. This is the "fight with banks, not tanks" approach and could involve freezing assets of individuals or organizations responsible [37]. Facebook has rolled out several initiatives in this respect. Firstly, stories that have been flagged as disputed would be prevented from being made into advertisements or promoted [137]. Secondly, pages found to have repeatedly shared falsehoods would "lose their ability to make money on the site" [157]. Thirdly, the ability to spoof domains when purchasing advertisements has been eliminated, in order to reduce the prevalence of websites that impersonate well-known news organizations [137].

Increasing Transparency. Issues-based advertising is often used to propagate falsehoods, leveraging polarizing issues to amplify existing social divides and partisan disagreements [8].

- Financing of Sponsored Content. There is a need to ensure transparency in sponsored content, particularly those related to political and issue-based advertising. Information such as sponsor identity, amount spent, and targeting criteria should be provided to help the public understand why they were targeted [122]. The country of origin should also be made available. All this information should be easily accessible for academia and research institutes to conduct analysis and highlight trends [124]. For example, the US, France, and Ireland require online platforms to reveal information to users about who paid for the advertisement, and to disclose information about the targeted audience [136]. In Australia, political advertisements are required to reveal the author and funder [140].

- Financing of Political Parties. Strict legislation also needs to be in place to ensure that any funding received by political parties are fully transparent [37]. This is to prevent foreign countries from attempting to influence local politics.

Limiting Political Advertisements. Both the European Commission and a UK parliamentary committee highlighted the need for online platforms to restrict targeting options for political advertising [122] [124]. In Oct 2019, Twitter announced that it would stop accepting political advertisements around the world, in response to growing concerns about disinformation spread by politicians that could influence votes and affect the lives of millions. However, Facebook seems unlikely to follow this path as their policy essentially allows political advertisements to be exempted from fact-checking [158].

Preventing Targeted Advertisements. In practice, it is difficult for users to protect their data due to "complicated and lengthy terms and conditions, small buttons to protect our data and large buttons to share our data" [124]. This has facilitated micro-targeting and "dark ads" that let organizations target posts at certain people. As these advertisements do not sit on the organization's main page, it makes it difficult for researchers or journalists to track what posts are being targeted at different groups of people. This is particularly concerning during elections [4]. The Cambridge Analytica incident – where personal information was taken without authorization and used to target affected individuals with personalized political advertisements [159] – illustrated how data can be used to disseminate manipulative and polarizing information. Twitter also recently admitted that user email addresses and phone numbers may have been "inadvertently" used for targeted advertising [160]. Such data breaches highlight the importance of data security [136]. Hence there needs to be greater transparency about access permissions granted to third-party applications that use data for profiling and micro-targeting. Users need to better view, understand, and manage such permissions [8]. They should also be given the option to opt-out of receiving targeted advertisements [8] [124], or advertisements from certain pages, accounts, or regions [8].

Rewarding Quality Over Engagement. Pages or publishers who consistently maintain higher journalistic standards could be rewarded financially, to encourage others to adopt similar high standards. This would incentivize publishing accurate and factual stories as opposed to retroactive fact-checking [8]. Another proposal was to shift towards a virtuous cycle where watchdogs award media with a seal-of-approval for quality news, which would encourage donors to support these media outlets in purchasing the most popular entertainment formats. This would in turn, attract more advertising [3].

4.4 Quality Journalism

Quality journalism is a "pillar of a society's information ecosystem", which needs to be "continually supported and nurtured" [2]. This "healthy Fourth Estate" should also be "independent of public authorities" [146], and remain the "most trusted medium available" [3]. This would provide the public with a reliable and trustworthy source of information, and dissuade the public from seeking less reliable sources of information to validate their perceptions [34]. It is thus critical that credible news sources do not unintentionally spread falsehoods [2].

Training Journalists. A recent report from the Institute for the Future, found that only 15 percent of US journalists had been trained to report misinformation more responsibly [4]. As the "gatekeepers of information" for society, it is important that journalists are trained in fact-checking, and be well-versed with how disinformation campaigns work [2] [37]. The curriculum should also encompass computational monitoring and forensic verification techniques for authenticating content [12], as well as how to distinguish and treat calculated political falsehoods versus legitimate alternative viewpoints differently [7]. Traditional media organizations, online platforms, and institutes of higher learning should collaborate to provide courses and workshops to journalists from all backgrounds, including online journalists [2]. In terms of funding and support, Governments should contribute and make it a national security priority [37], while additional funding can be sought from grant-making foundations.

Tightening Journalistic Controls. Several governments have proposed or implemented tighter controls over their national media, as well as online news and social media platforms.

- Introducing Legislation. For example, the US' enforcement of the Foreign Agents Registration Act, aim to enhance quality journalism and improve transparency on information sources [136]. In China, social media platforms are allowed to only publish news articles from registered news media [140].

- Abiding by Common Standard. To improve the quality of online news platforms, one proposal was to encourage or require these alternative media to subscribe to a code that defines acceptable behavior and standards of accuracy [71], and be held to similar standards as traditional media organizations [2]. Those that sign up to the charter should be supported by governments and grant-making foundations to participate in regular exchange programs with other partners, such as journalists and academics [3]. National and international professional journalistic associations should also maintain and enforce a code of conduct, to ensure that their members are kept in check [37].

- Strengthening Community Standards. A UK parliamentary committee called for social media platforms to jointly develop a "professional global Code of Ethics" with "governments, academics, and interested parties", which would be "continually referred to when developing new technologies and algorithms" as well as define "what is and what is not acceptable" behavior of social media users [124].

Conducting Independent Audits. Similar to how finances of companies are audited and scrutinized, independent audits should be conducted on non-financial aspects of online platforms – including algorithms and security features – to ensure that they are operating responsibly [124]. Online platforms should also be transparent and provide regular voluntarily reports on the nature and extent of disinformation being spread on their platforms, and the effectiveness of their responses [2].

Relieving Financial Pressure. As financial pressures have led to "news deserts" in certain areas, it has been suggested that governments find means to support quality journalism initiatives at the local, regional, and national levels [12], perhaps even through subsidies for local news outlets [83].

Channeling more financial resources to quality news sources could help prevent the proliferation of low quality news sources. Such an initiative was practiced in the past, when newspaper distribution subsidies were common in the era of printed newspapers, usually via subsidies of postal services. This enabled news outlets to have a wider reach [88]. Alternatively, a possible business model to adopt for traditional media organizations is that of the British Broadcasting Corporation (BBC) in Britain [2]. The BBC is primarily funded by a television license fee charged to households or organizations that watch or record any live television program [161].

Creating "Whitelists" and "Blacklists". A list of trusted news sources could be developed and maintained, based either on user ratings or that of an independent organization [69]. Conversely, media advocacy and civil society groups can also compile lists of online sites that have a record of broadcasting falsehoods [162], as what an associate professor of communication and media at Merrimack College in Massachusetts had done [163]. French news outlet Le Monde also has a database of more than 600 news sites that have been labelled as "satire", "real" or "fake" [146]. Some governments, like Russia, have also proposed setting up a database of news sources that disseminate falsehoods [140].

Promoting Credible Content. Articles from more authoritative sources or publishers, or other indicators to determine quality of the source or content, should be used to prioritize articles via the algorithms of online platforms [8], while proven falsehoods should be deprioritized [2]. Social media users should also be given the option to choose if they wish to receive news from "accredited sources" only or from any source [88].

4.5 Education

It is precisely our willingness to share without thinking that agents of disinformation will use as a weapon. Hence, it is critical that every person recognizes how he or she, is involved in the spread of disinformation and develop a set of skills to navigate communication online [4]. To tackle disinformation in the long-term, it is thus necessary to have an informed and discerning public who are aware of the threat and equipped with the critical thinking skills required to protect against malicious influence [121]. Unfortunately, many studies have indicated a lack of news literacy among the public. For example, Stanford researchers found that students from middle school, high school, and college, had difficulties judging the credibility of online information despite being digitally savvy [164]. In Singapore, a government survey found that two in three Singaporeans could not discern falsehoods when they first see it, and a quarter of respondents admitted to sharing information that were later proven to be false [165].

Developing Essential Skills. Public education refers to the build-up of media and digital literacy, and critical thinking skills [2]: (a) how disinformation campaigns work and how they can be spotted [71]; (b) how news is made, who makes it, how it is selected, and how it is financed [99]; (c) how algorithms determine what users see by prioritizing posts that "spark conversations and meaningful interactions between people" [4]; and (d) practical skills in browsing the Internet for information and evaluating the quality of the content [136].

Tailoring to the Audience. Although news literacy is often associated with the young, it is equally important to also educate adults and the elderly [34]. Educational initiatives should encompass all segments of the population, with the material and modality tailored to suit their specific needs.

- Students and Youths. Critical thinking and digital literacy should be part of the school curriculum for students and youths as they can be easily influenced [2] [37]. This will enable them to critically assess the veracity of news from a young age. Students should also be aware of the threat of disinformation operations [2]. Countries like Canada, Italy, Sweden, and Taiwan have already amended their school curricula to enable students to spot false news articles and critically evaluate sources [134] [166].

- Working-Age Adults. For working adults, companies should incorporate such modules into their training program. For non-working adults, seminars and workshops should be curated for them [2].

- Elderly. There should be initiatives that reach out to the elderly as they can be vulnerable to disinformation. This could include talks [2] and other outreach programs.

- Key Government Personnel. Interior ministries and counter-intelligence agencies should train key personnel in the government to identify and resist disinformation operations, especially politicians, diplomats, and high-level state bureaucrats, who are obvious targets [2] [37]. Highly specialized experts would also be needed for the government's strategic communications unit. Knowledge in law, international relations, security policies, defense matters, social media and media environment, and political science are needed. Governments would need to work with universities and institutes of higher learning to develop a specially curated program for this purpose [37].

Using Various Means. Education should include both formal and non-formal means [122]. For example, full-page advertisements in newspapers [103], posters distributed in various languages to local neighborhoods [2], and free online courses [111] can all be used to explain how to spot fake news. Traditional media organizations can also have dedicated programs to raise public awareness about disinformation campaigns and their implications [12] [37], as well as teach audiences how to assess content they consume and the verification process to debunk falsehoods [12]. It can also take the form of entertaining TV or online content and include humor, fun, liveliness, and other qualities that audiences are attracted to. However, to reach a wider audience, such programs should be part of mainstream programming rather than a separate show. The challenge is introducing such themes into talk shows, sitcoms, popular dramas, and cartoons [3]. Other engaging methods such as interactive games can also be used to captivate audiences, especially the young people [2].

"Inoculating" the Public. This is based on "inoculation theory", which prepares individuals for falsehoods by exposing them to information containing weakened false arguments. This would help individuals develop a resistance against more persuasive attacks in the future [34]. An "inoculation message" typically consists of two parts: (1) a direct and explicit warning that one's belief on an issue is under threat, and (2) a refutation of a possible argument raised by opponents [167]. Using an example on climate change, the warning could state how certain people aim to convince the public that there are disagreements among scientists about climate change. This false claim can then be refuted by reiterating scientific research that there are no disagreements among climate scientists that humans are causing climate change [168]. In a separate study, a group of researchers from University of Cambridge developed an online game in which people take on the role of propaganda producers, to raise awareness about real-world disinformation. It was found that the game increased their "psychological resistance" to falsehoods like a "psychological vaccine" [169]. Such initiatives associated with "forewarning" is possibly more effective than refutation, and can be likened to "[putting] raincoats on those at whom the firehose of falsehood is being directed" [94].

Switching to "Slow Thinking". Media literacy also involves educating the public to switch from "fast to slow thinking" when assessing news articles [88]. In the same way that you're told to wait 20 minutes before you reach for a second helping of food, because you need to wait for your brain to catch up with your stomach, the same is true with information. Hence one should wait a couple of minutes before clicking the share button [24]. To do so, people need to be taught to develop cognitive "muscles" in emotional skepticism and trained to withstand the onslaught of content designed to trigger base fears and prejudices. Understanding how each one of us is subject to such campaigns – and might unwittingly participate in them – is a crucial first step to fighting back [4].

Ensuring a Coordinated Approach. There should be a national framework to coordinate and guide the various public education initiatives, in order to ensure coverage of all segments of society. A university or research institute could devise an index to measure the public's ability to identify falsehoods. The curriculum should also be regularly updated based on the latest research and knowledge about the threat of disinformation [2].

Providing Funding and Support. The government, educational institutes, trustworthy media entities, and private sector technology companies should collaborate to build media literacy from an early age [121], while encouraging ground-up or community-based projects and initiatives. Grant-making foundations can also support such programs to help the public navigate the information ecosystem [12]. For example, the governments of Croatia [136] and Canada have set aside funds meant to increase public awareness of online disinformation [140]. Facebook has also launched the News Integrity Initiative, which is a group of over 25 funders and participants – including tech industry leaders, academic institutions, non-profits, and third party organizations – with the objective of helping the public make informed judgements about the news read online [113]. An educational levy on online platforms has also been suggested, as a means to finance a comprehensive education framework [124].

4.6 Social Cohesion and Trust

Disinformation agents seek to exploit fault lines and underlying conditions present in society, such as those between racial and religious groups, between citizens and immigrants, and between people of different socio-economic backgrounds. One way to mitigate the potential effects is to strengthen social cohesion and trust – both among the different communities and groups, and in the government [2]. Reestablishing trust in a democratic society can serve as a countermeasure against the systematic efforts to devalue truth [69].

Between Communities. Initiatives to reinforce social cohesion and trust between people and communities seek to counter the effects of disinformation, rather than the disinformation itself [94]. Such initiatives should take into consideration several key principles to ensure their effectiveness. Firstly, there is a need for in-person interactions and communications. Secondly, community leaders, including cultural and religious leaders, should be involved. Thirdly, there should be "safe spaces" to discuss different perspectives on divisive and sensitive issues such as race or religion, in the presence of trained moderators. Lastly, government-led initiatives need to be complemented with ground-up initiatives [2].

Trust in Government. It is crucial that the public has trust in its government so that public institutions can intervene effectively when called upon. To reinforce trust in the government, it is important to have effective real-time communication. This includes responding to disinformation promptly and releasing the appropriate facts to the public, as well as preemptively inoculating the public against potential vulnerabilities [2]. Firstly, a crisis communication plan should be developed to provide an immediate response to disinformation operations. There should also be regular inter-agency scenario planning and mock exercises, to ensure the plans stay relevant and prepare government agencies for an actual incident [134]. Secondly, the government needs to ensure there are adequate efforts and initiatives to demonstrate its transparency and accountability. An example was to implement a Freedom of Information Act to allow the public to request access to government records [2]. Thirdly, the government needs to sufficiently engage the public on their strategies against disinformation operations through outreach activities [2] [37]. This would mitigate the perception of the "establishment against people" [37].

Identifying Vulnerabilities. It is important that the government supports or conducts research on vulnerabilities of its society. This would involve monitoring, identifying, and understanding the trigger points in society and its vulnerabilities to foreign disinformation, of both the general population and key personnel in the government. Such key personnel would include politicians, as well as those in the economic, energy, financial, transportation, security, and information sectors [2] [37]. Think tanks and academia can conduct such research and examine both short and long-term scenarios, to identify vulnerabilities that need to be addressed. Grant-making foundations could also be asked to sponsor national polls [37].

Encouraging Acceptance of Diverse Views. Several digital tools have been developed to reduce the impact of filter bubbles and thus expose people to more diverse views. The Times launched a Facebook messenger bot called "Filter-bubble Buster", which aimed to provide people with a balanced view of information ahead of the 2017 UK elections [170]. "FlipFeed", a Twitter plug-in, replaces one's regular Twitter feed with that of a random, anonymous user of a different political inclination. A similar plug-in, "Escape Your Bubble", seeds one's Facebook feed with opposing political views [171]. Users should also be allowed to customize their feed or search algorithms to enable them to see more diverse content if they wish, thus minimizing the impact of filter bubbles [12]. A key challenge to such initiatives would be determining what is a suitable "oppositional voice" [69], which is sufficiently different to burst the filter bubble, yet without being too different and appear as offensive to the user [71].

4.7 Legislation and Regulation

The stakeholder analysis in Section 3.4 highlights that online platforms play a critical role in combatting disinformation. However, as there is currently little incentive for them to be committed to the success of the system, the introduction of legislation and other forms of regulation appear necessary to provide this "incentive" through legally binding orders, and making them accountable for developing better solutions and technology. Such action will also act as a form of deterrence against creating or spreading disinformation, to ensure that the Internet does not remain a "wild west" [124]. As legislation would serve to encourage or even compel online platforms to implement or support many of the countermeasures from Sections 4.1 to 4.6, those related countermeasures are not repeated in this section. Instead, this section highlights some of the benefits of legislation and how they can serve to overcome existing shortcomings or deficiencies in the absence of legislation.

Protecting Democracy. There are arguments that legislation or regulations could impede freedom of speech. However, it can be said that falsehoods harm democracy by providing the public with false information, and legislation actually serves to protect democracy and preserves these ideals. In fact, disinformation agents have claimed to be upholding "free speech" as a means to disguise their malicious objectives [172]. A famous metaphor introduced by Justice Oliver Wendell Holmes in 1919 may help to differentiate free speech from falsehoods that need to be addressed. He said that even "the most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic", and it is important to consider whether the falsehood would "create a clear and present danger". People in the theatre would not know that the statement was false and would also not have the time to check. Hence this would cause a panic, even if everyone was acting rationally [173].

Serving as Punishment and Deterrence. Disinformation agents that have created or intentionally spread falsehoods to harm the public should be punished accordingly through criminal sanctions [2]. It was also suggested that defamation suits against intermediaries be permitted, to encourage intermediaries to more carefully screen content on their platforms [69]. These proceedings can be launched by a public regulator, individuals, or organizations [124]. Such punishments can also serve as a form of deterrence. However, it is important that such measures be imposed only when certain criteria are met, such as having the intent or knowledge and causing a certain level of harm (e.g., incite hatred, lead to public disorder, interfere with elections, or impact national security) [2]. Countries such as Australia, Indonesia, Ireland, Italy, and the Philippines impose criminal penalties and fines for producing or sharing disinformation, or for developing a bot campaign targeting political issues [136].

Ensuring Prompt Response. Several cases have demonstrated that online platforms have not dealt with disinformation promptly in the absence of legislation. The UK's Home Affairs Committee called Google's weak and delayed response to dealing with illegal neo-Nazi propaganda "dreadful" and questioned the need to strengthen law and enforcement mechanisms in this area [174]. The Sri Lankan government also criticized Facebook for failing to control rampant hate speech that apparently contributed to the deadly anti-Muslim riots in 2018, adding that it took days for Facebook to review and remove pages [175]. Sri Lanka had to resort to blocking access to Facebook, as well as two other platforms owned by the company – WhatsApp and Instagram – in an attempt to stem the violence directed at the Muslim community [176].

Enabling Consistency in Application. It can be argued that allowing online platforms to self-regulate can also lead to inconsistencies in fact-checking of articles. For example, Facebook has said that it would fact-check advertisements from political groups but not politicians. Facebook CEO Mark Zuckerberg explained that this policy was to avoid stifling political speech, drawing anger from Democratic candidates running in the 2020 presidential election such as former Vice President Joe Biden and Senator Elizabeth Warren [177]. Senator Warren demonstrated the loophole in such a policy by successfully purchasing an advertisement, which falsely stated that Zuckerberg had endorsed Trump for president [178].

Encouraging Increased Scrutiny. When faced with the threat of fines, online platforms would be more motivated to do all they can to prevent the spread of disinformation that could harm society, and be more meticulous in scrutinizing and identifying potential disinformation agents. The non-profit watchdog, Campaign for Accountability (CfA), exposed how Google has thus far done little to prevent foreign actors from influencing US elections. This is despite Google's claims that strict new rules had been put in place to stop foreign actors from purchasing political advertisements that can spread divisive propaganda. In an experiment, CfA managed to purchase advertisements on Google's Russian advertisement platform to target US Internet users, despite waving "obvious red flags throughout" in an effort to trigger Google's apparent safeguards. For example, CfA used a Russian IP address, supplied details and promoted websites of the previously indicted Russian troll farm, and paid for the advertisements in Russian currency using Russia's largest online payment service. However, Google made no attempts to verify the account and approved the advertisements within 24 hours [179]. The European Commission also acknowledged the need for online platforms to "significantly improve the scrutiny of advertisement placements" to "reduce revenues for purveyors of disinformation" [122].

This page is intentionally left blank

Chapter 5: Developing the Proposed Architecture

This chapter aims to develop the proposed architecture by building on the possible countermeasures from Chapter 4. The chapter starts by envisioning the future through the eyes of various stakeholders after the system has become a reality. These visions of the future will serve as a guide when selecting the preferred options for key architectural decisions[4] [9], which each government would need to consider when developing its customized solution. The chapter concludes with the proposed system architecture.

5.1 Envisioned Future

This section describes the envisioned future through the eyes of key stakeholders[5] through "vignettes". These vignettes help to illustrate, at a more personal level, how stakeholders contribute to and benefit from the system [10].

Member of Public. I am confident that the majority of the information and news I read online are genuine due to the multiple safeguards that have been put in place. However, I do not let my guard down and am confident in my own ability to identify falsehoods, if I should come across them. I can also rest assured that even if I should be unwittingly exposed to falsehoods, I would be notified of corrections in due time. There are also tools and websites that enable me to easily report any falsehoods that I encounter, allowing me to play my part in the effort against disinformation. During elections, I have faith in the democratic process and am equipped with the necessary information to vote wisely. Hence I have trust in the government that has been elected, and that the public institutions have our best interests at heart.

Senior Government Official. Through the various initiatives and support that we have provided over the years, we are delighted to have an information ecosystem that is now relatively free of falsehoods. The nation remains safe and secured, and our democratic system and society is adequately protected from state-sponsored disinformation operations, as well as other malicious

[4] Architectural decisions are the subset of design decisions that are most impactful. They determine the performance envelope and encode the key trade-offs of the system.

[5] These are the stakeholders that lie in the "red boxes" of Figure 3.

actors that aim to influence our policies or fuel discontent in society. The high level of mutual respect and trust between our communities makes it harder for disinformation agents to exploit fault lines in our society. At the individual level, citizens have easy access to trusted sources and accurate facts. We are also confident that the education system has given rise to a population that is well equipped to critically assess the veracity of information they encounter, and would not be easily swayed by falsehoods.

CEO of an Online Platform. We no longer have to be the arbiters of truth as there are now clear definitions on what constitute a falsehood that needs to be addressed. We have reaped benefits from sharing data with our trusted partners, as we have gained valuable insights into how disinformation is spread on our platform and the effectiveness of the measures that we have implemented. Our investments in technology have also enhanced our capabilities greatly and made us more effective in combatting disinformation on our platform, to the delight of both the authorities and our users. This has served as a catalyst for growth in both our user base and revenue, as we are now a more trusted organization in the eyes of the public.

Editor from Traditional Media. The increase in readership and support provided to us from various sources have helped to relieve us from financial pressures. This has enabled us to focus on what we do best – produce high quality news promptly. We are now more widely respected and trusted by the public. It would seem that this increase in trust has led to a corresponding increase in readership as people have started to avoid less trustworthy sources, leading to a virtuous cycle that has spurred the revival of our industry. We are also able to better utilize and allocate our manpower resources as we no longer have to perform fact-checking on our own, but rather with the help of a coalition of fact-checkers. With the extensive training provided to our journalists, I have the utmost confidence in their ability to identify and report falsehoods accordingly.

Fact-Checker. Fact-checking used to be an extremely resource intensive process. Fortunately, the development of automated fact-checking tools have greatly reduced the time needed to review and verify articles, thus making my job much easier. I used to take hours just to verify the authenticity of an image or video. Now, with the advanced tools provided, I can do it in a matter of seconds. The close partnerships with other stakeholders in the information ecosystem and other fact-checking organizations, have enabled our teams to optimize our limited resources and prevent any duplication in efforts.

5.2 Architectural Decisions

This sections lists the key architectural decisions and possible options for each decision. As stated in Section 1.5, the author intends to avoid developing an architecture for a specific country. Hence, following a discussion on the pros and cons of each option, a preferred option that is likely to be applicable to most nations is proposed for each architectural decision. In reality, the architect could develop a set of evaluation criteria based on what is deemed to be important to the specific nation, taking into consideration the needs of all major stakeholders and prioritizing them accordingly. This would facilitate a more detailed assessment on the various architectural options before deciding on an architecture that is most suited to the nation's specific needs and circumstances.

5.2.1 Legislation

This decision will set the stage and eventually determine the amount of leverage that the government would have over online platforms, as well as the level of deterrence against malicious actors. Hence, this is arguably the most important architectural decision and would likely impact the effectiveness of the system significantly. Since 2016, more than 40 countries have proposed or implemented regulations specifically designed to handle disinformation campaigns [136]. However, this has engendered backlash from various segments of society due to fears of censorship and lack of freedom of expression [34].

Using Existing Legislation. Several countries – Canada, Japan, Sweden, and the UK – are relying on existing legislation that regulates the media and elections, and anti-defamation laws. However, current legal frameworks were developed during a period when disinformation operations could not undermine the basic values of democracies so aggressively [37], and may not reflect current technological and telecommunications developments [123]. Hence such laws may be inadequate to deal with disinformation, which can be disseminated almost instantaneously and with great ease. A UK parliamentary committee also recognized that current electoral laws are "not fit for purpose for the digital age" and needs to be updated [124], in order to keep up with evolving technology [121].

Introducing New Legislation. Since 2016, more than 30 countries have introduced legislation aimed to combat falsehoods on the Internet [22]. This include imposing sanctions on social media companies that disseminate falsehoods, such as fines, and ordering the removal of false information. For example, Germany, Israel, Russia, and China have enacted such laws [123]. Countries such as the Ukraine, the Czech Republic, and France have initiated reviews of their legislation to ensure that they are adequate to deal with the threat of disinformation [2]. New legislation can also send a signal to society that this is an important focus area and that facts matter.

Introducing Indirect Regulation. There is concern that any direct regulation imposed by governments could lead to suppression of dissenting views and thus undermine free speech [166]. An alternative is to make necessary amendments to existing legislation, and allow tort lawsuits alleging defamation by those directly harmed by the false story. This could also pressure online platforms to intervene more regularly to avoid any liability in the spread of disinformation [119].

Allowing Voluntary Regulation. This option would remove the risk of "overboard government regulation" that could grant the government too much power to control speech [69]. The main downside of voluntary regulation is that online platforms may not abide with their commitments. For example, in Germany, despite making voluntary commitments to remove criminal content within 24 hours, it was found that Facebook only acted in 46% of the cases and Twitter only 1%. This led to the introduction of the Network Enforcement Act in Germany, as increased pressure was deemed necessary [180].

Proposed Option: Introducing New Legislation. As the UK's Digital, Culture, Media and Sport Committee had stated in their interim report on disinformation and "fake news", online platforms should not be placed in a position to "[mark] their own homework" [124]. There also seems to be a fundamental conflict of interest between self-regulating false content and their business model of rewarding engaging content [2] [124]. Hence legislation and regulation would ensure that online platforms undertake or contribute to the measures proposed in Chapter 4. Legislation would also likely enable a swift response to combat the spread of disinformation. Given that existing legislation may not be adequate to deal with this new threat, it is proposed that new legislation be introduced to specifically combat disinformation. However, it is critical that any form of legislation or regulation is calibrated and measured, with sufficient checks and balances. As legislation may need to be applied across various online platforms, it is important to be respectful of communications that are personal, private, and of limited circulation [2]. Any legislation should also avoid overly broad language that could affect a range of platforms and services and put decisions about what is illegal into the hands of private companies, which may then be inclined to over-censor to avoid potential fines [181].

5.2.2 Decision-Maker

This decision selects the entity that determines what is true or false, and thus significantly impacts the overall system effectiveness.

The Executive. There may be situations where only the Executive would have the necessary facts to make a decision. For example, issues related to national security, public order, and public institutions. It can also be argued that a judicial process may not be fast enough to deal with disinformation as falsehoods can spread at breakneck speeds. However, there would be concerns about the credibility of Executive action as there could be abuse of power [2]. Singapore is one country that has introduced measures enabling executive decisions to be made on content moderation and to disrupt digital advertising revenue [136].

The Judiciary. It can be problematic when government entities become arbiters of what is true and what is fake. Logically, governments would have a duty to protect audiences from fake news. Yet, both the executive and legislative branches might be perceived as self-interested if they tried to evaluate truth or falsity. The courts may thus be the best alternative [68]. For example, France leaves the legal interpretation of falsehoods to its judiciary, whereby judges would rule on a case-by-case basis for prominent content [136].

Government Department. For example, Philippines' proposed law on "false content" would enable a government department, the Cybercrime Office in the Justice Department, to decide whether any material is permissible [182]. However, it is not clear if an office within a government department would have sufficient credibility in the eyes of the public.

Independent Committee. An independent committee comprising representatives from various fields of expertise could be better positioned to make decisions, especially for difficult or controversial cases, where there are conflicting viewpoints on the need for intervention. This could also mitigate any concerns about governmental abuse of power [2].

Online Platforms. In this model, online platforms would decide whether to remove material based on user notifications. For example, in Germany, social media networks have to remove "clearly illegal" content within 24 hours, but can take seven days or even more to determine the truthfulness of the information [183]. A UK parliamentary committee has also proposed that technology companies act against harmful and illegal content on their platforms which have been flagged by users, or content that is easily identifiable by the companies themselves [124].

Proposed Option: The Judiciary. It is important that the decision-maker has credibility and is able to make decisions effectively. Online platforms may not be the best option due to potential conflict of interests and should not be ultimately responsible for making decisions that could impact society. However, they could still serve as the first line of defense by removing "clearly illegal" content based on notifications from users or fact-checkers. While an independent committee could be unbiased, it is unclear if they would have sufficient authority or experience to make such important decisions. Between the various government branches and entities, it would seem that the judiciary is best placed to be the final arbiter of truth given its apparent independence.

5.2.3 Accountable Party

This decision selects the party that would be held accountable for the creation or spread of disinformation, and hence determines the reach of any proposed legislation or regulation.

Social Media Platforms. Germany's Network Enforcement Act is aimed at social media platforms with more than 2 million members – such as Facebook, YouTube, and Twitter [140]. The UK's Digital, Culture, Media and Sport Committee also recommended to "tighten" the liabilities of social media platforms as they influence what information people receives [124].

Online Platforms. In Russia, legislation exempts mainstream news organizations and is aimed at online news outlets. Websites that have a commenting feature and amass more than 100,000 visitors every day are also required to remove false comments within 24 hours [140]. This is an expansion of the option above and seems to target other online platforms such as online news outlets and discussion forums.

Individuals and Online Platforms. In May 2019, Singapore passed a law that criminalizes the dissemination of falsehoods online. Any "malicious actor" can be punished with fines and jail time. Online platforms like Facebook are also subject to similar punishments for their roles in the dissemination of falsehoods. Even closed messaging applications are impacted by the law. It has been said that these measures are among the most comprehensive anti-disinformation laws in the world. Besides Singapore, other countries such as Sri Lanka are also considering the expansion of existing legislation to prosecute people that spread false statements or hate speech that "hinders the peace among communities and national security" [140]. In Egypt, any account with more than 5,000 followers on social media platforms will be treated like a media outlet and have to abide by existing laws [184].

Proposed Option: Individuals and Online Platforms. While this may seem over-reaching and harsh, individuals should not be exempted from criminal sanctions, especially when they had intended to cause serious harm to society. Examples of how individuals have caused significant harm are mentioned in Section 2.3. However, it is important that such sanctions should only be imposed under certain circumstances, e.g. based on a threshold of criminal intent and harm caused. It is also important to subject all online platforms to the same legislation, as disinformation can be spread through any online channel.

5.2.4 Degree of Government Involvement in Fact-Checking Initiatives

Fact-checking is a key component of the system as it serves to sieve out the falsehoods. This decision determines the degree of government involvement, and hence the perceived independence of fact-checking initiatives.

Completely Independent from Government. This could increase the credibility of the initiative as the public may perceive any government involvement to be spreading propaganda. It would also allow the initiative to objectively analyze issues related to politics or governance [2]. Non-governmental initiatives will also likely provide a quicker response to disinformation campaigns as they would not be hindered by bureaucratic demands [134]. For example, Facebook has partnered fact-checking organizations to prevent the spread of disinformation on its platform [34].

Government Involvement Only when Necessary. The government may need to be involved in exceptional circumstances, particularly when government-backed information is required. For example, for cases where national security is at stake, the government might need to get involved to debunk the falsehoods [2].

Independent, but with Government Funding. While the initiative remains independent, the government could provide a degree of funding to support the work [2]. However, there will be some concern that any form of government funding and incentives could allow government actors to interfere and determine what is true or worthy [69]. Hence it is important for such initiatives to prove and showcase their independence from government influence.

Collaboration between Government and other Stakeholders. A joint initiative between the government with other stakeholders such as independent fact-checkers, traditional media organizations, and online platforms, could increase the effectiveness and efficiency of information verification. Having a diverse set of skills and knowledge would also aid in building credible narratives against disinformation campaigns [134].

Government-Run Initiative. On the other end of the spectrum, fact-checking initiatives could be fully-run and owned by the government. This might be effective in countries where trust in government is relatively high such as China, Indonesia, India, Singapore, and Malaysia [107]. For example, Singapore's government launched a website "Factually", which aims to dispel and clarify falsehoods that captured the public's attention [34] on matters of government policy or public concern that can "harm Singapore's social fabric". Topics are selected based on feedback from sources both within and outside of the government [185]. Malaysia also launched a similar fact-checking website to curb the spread of false information [186]. However, government-led initiatives could be perceived as pro-government bias, especially if every single post supports the government's action, e.g. Fact Checking Turkey [187].

Proposed Option: Government Involvement Only when Necessary. A fact-checking initiative should be as independent from the government as possible, to increase the credibility of the initiative. However, there may be situations where government-backed information is required, particularly when national security is at stake. Hence while the fact-checking initiative could remain fully independent, it should reach out to the government when the need arises.

5.2.5 Anonymity on Internet

This decision will determine the ease at which malicious actors can be identified by law enforcement agencies, and hence how swiftly they can be taken to task.

Zero Anonymity. China enacted and enforced a "real name registration" scheme, which requires internet-related business operators to obtain evidence of the "true identity" of internet users in China. This aims to deter the public from breaching regulations as they can be easily identified by law enforcement agencies [188]. However, such an approach could come at the cost of user privacy and increase the vulnerability of at-risk users such as dissidents and journalists [121].

Verification of Influential Accounts. Facebook noted that disinformation operations tend to be led by people operating inauthentic accounts rather than merely an automated process [145]. Hence to prevent the misuse of inauthentic accounts, influential accounts whose posts reach a significant number of people could be asked to verify their accounts [8] [122]. Thresholds could be based on criteria such as the number of followers or posts that have been re-shared [8].

Anonymity Allowed. This is the other end of the spectrum where people can remain anonymous and retain their privacy. Unfortunately, anonymity on the Internet has enabled malicious actors to hide their identities while disseminating falsehoods [121].

Proposed Option: Verification of Influential Accounts. While there are merits to privacy, this has been exploited by malicious actors and facilitated the spread of falsehoods. Over-regulation would also not be ideal and the ordinary citizen should not have to be subjected to such requirements. Hence the proposed option is to require verification of influential accounts, which have the potential to be misused and disseminate falsehoods to a large enough audience.

5.2.6 Structure of National Agency

Section 3.4.2 identified the government as a critical stakeholder to the system. In fact, a whole-of-government approach is needed to counter the asymmetric threat [121]. The government would require a multi-disciplinary team at the national-level to provide a coordinated approach against disinformation. This would require expertise from the various branches such as foreign policy, defense, homeland security, military, intelligence agencies, as well as media and communication.

The team would need to react in real-time to disinformation operations, publish regular updates on disinformation trends, conduct research, and coordinate with other partners [37]. This architectural decision is to determine how this team should be structured within the government.

Separate Agency. Several governments have established new agencies or mandated existing organizations to handle the threat of disinformation [22]. For example, in 2018, Sweden announced that it would establish a new "psychological defense" authority to counter disinformation and foreign influence campaigns. Rather than attempting to directly combat false information, the agency is tasked to promote factual content [140]. The UK government has established a dedicated national security communications unit that would be tasked with combating disinformation by state actors and others [189]. Indonesia also set up an agency to combat fake news, and among its tasks would be to monitor news circulating online [190]. In the US, after the end of the Cold War, the Bill Clinton administration and Congress closed down the US Information Agency, which was tasked with influencing foreign populations [35]. Some are calling for its reconstruction [3].

Unit within Another Entity. Governments have set up cybersecurity and information security units within their militaries to tackle foreign influence in elections [136]. The Indonesian government has also established a team of 70 engineers constantly monitoring social media traffic to identify online falsehoods, ahead of its 2019 presidential elections. This team is part of the Ministry of Communications [191] and has the authority to remove posts containing falsehoods [140].

Inter-Agency Task Force. There are governments that mandated security and defense authorities to handle the threat of disinformation, such as Australia's Election Integrity Assurance Taskforce, which comprises several entities such as the Home Affairs Department, Australian Security Intelligence Organization, and the Australian Federal Police. Their focus is to defend against election meddling [136] [192].

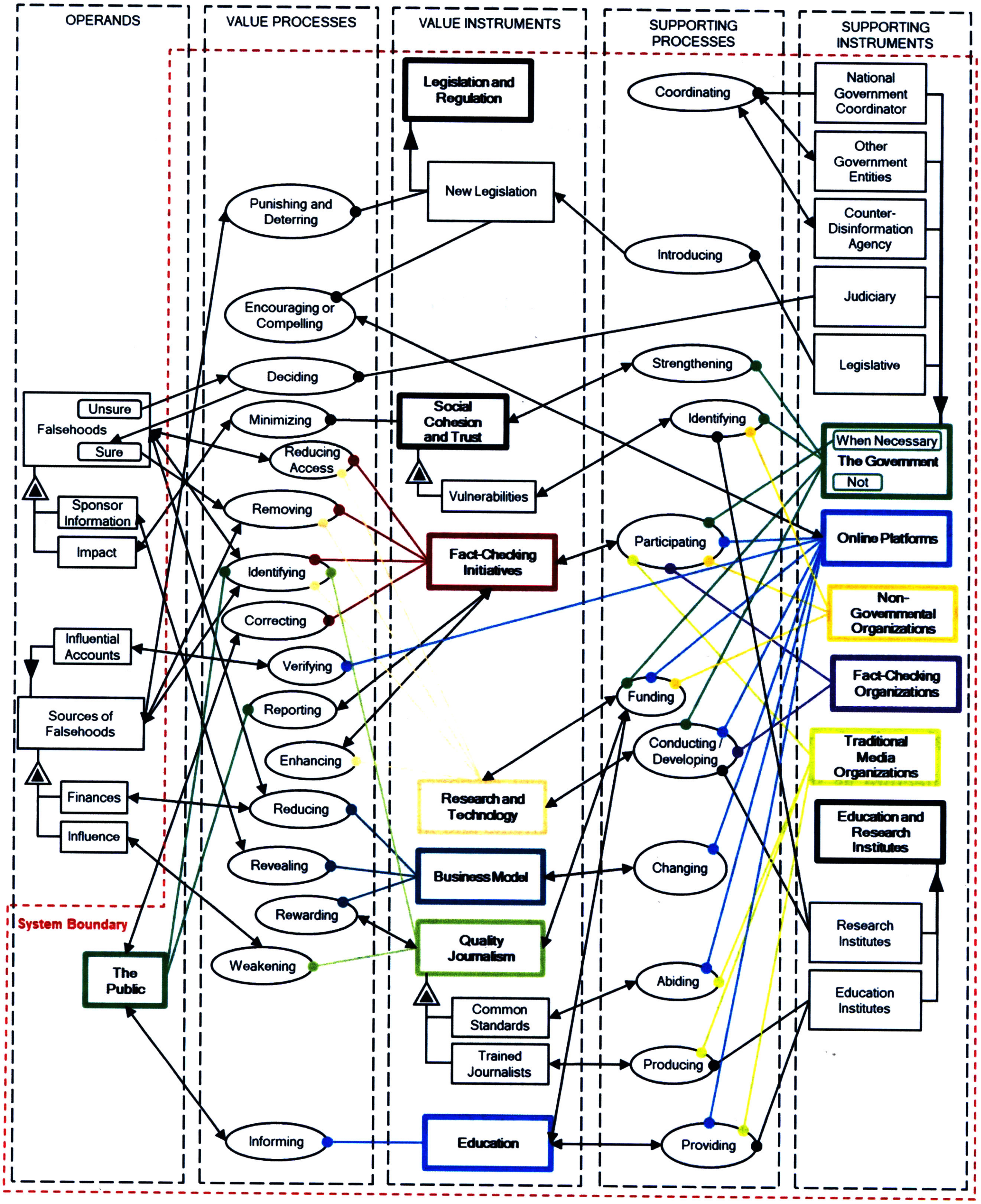
National Government Coordinator. The government could appoint a coordinator for countering disinformation [37]. This coordinator would likely have to be someone with sufficient authority, such as a member of the cabinet, in order to pull resources from various departments and agencies.

Proposed Option: Separate Agency AND National Government Coordinator. To effectively combat the threat of disinformation, dedicated resources would probably be needed. Hence the establishment of a separate agency with its own manpower and resources is the proposed option. It would also send a signal to everyone that the government is serious about tackling the issue. However, it is likely that security and intelligence agencies would continue to maintain their own dedicated units that combat disinformation, due to the sensitivities of their work. Hence it would seem necessary that a national government coordinator, with sufficient authority and clout, be appointed as well. This coordinator would have to coordinate the activities between the separate agency and other government units that combat disinformation. The objective would be to optimize resources and prevent the duplication of efforts wherever possible. To help pay for the expanded work of the government, a UK parliamentary committee has suggested imposing a levy on online platforms, similar to how the banking sector pays for the upkeep of the Financial Conduct Authority [124].

5.3 Proposed Architecture

Based on the key countermeasures from Chapter 4 and the architectural decisions from the previous section, the system architecture is depicted in Figure 4 using Object-Process Diagram (OPD) representation. The legend for OPD is available in Appendix B. The key entities (or stakeholders) and the categories of countermeasures, are in bold boxes and color-coded for easier reference. It is apparent that an effective response requires a whole-of-society approach [145], with collaboration between various stakeholders for each countermeasure. While not represented in this architecture as it is beyond the national-level response, there could be bilateral, regional, and international partnerships with other governments and international bodies. For example, in terms of bilateral cooperation, Russia signed an agreement with Spain in November 2018 to set up a joint cybersecurity group with the objective of preventing disinformation from affecting relations between the two nations [140]. The EU's East StratCom Taskforce is an example of regional cooperation, and was set up in 2015 to counter Russia's disinformation campaigns. The task force serves as a regional mechanism to facilitate collaboration with a wide network of government officials, experts, journalists, and think tanks [134]. At the international level, twenty countries – including France, Britain, India, South Africa, and Canada – have signed an agreement to prevent the spread of online falsehoods [193].

Figure 4: Proposed system architecture.



Chapter 6: Case Study and Validation

This chapter aims to qualitatively validate the proposed system and ascertain its applicability in a real scenario. The annexation of Crimea by Russia on 17 March 2014 is a prime example of how disinformation can be a powerful tool in the hands of adversaries. The spread of propaganda and the manipulation of facts were one of the key factors that helped Russia to succeed in this respect. While the occupation was accomplished using physical military forces, the actual invasion began "in the minds" of the Crimeans [194]. Hence this case study seems appropriate to help validate the applicability of the system in an actual scenario.

6.1 Approach

Firstly, the key events and disinformation campaigns conducted by Russia are compiled to provide a timeline of the events that eventually led to the annexation of Crimea in March 2014. The time period of 2000 to March 2014 is selected – 2000 marked the beginning of Russia's seeding and exploitation of Ukraine's vulnerabilities, while March 2014 was the culmination of those efforts [195]. This aims to provide an appreciation of how the Crimeans were subjected to long-term disinformation efforts that exploited "slow burn" issues and fault lines in society.

Secondly, the key themes of the disinformation campaigns are extracted and aggregated to summarize the types of disinformation that were spread, including the objectives and means of dissemination. The aim here is to determine the key themes or issues that need to be handled by the proposed system.

Finally, validation of the proposed system and its applicability is qualitatively discussed through three main approaches: (a) illustrate the lack of effective countermeasures employed by Ukraine during that time period, (b) how the proposed system, and particularly the countermeasures from Chapter 4 (as they are the value instruments[6] of the system), could have prevented or mitigated the impact of Russia's disinformation campaigns, and (c) list similar countermeasures that the

[6] These are the objects of the system that are essential for the system to deliver value.

Ukrainian government and other entities have since introduced to combat Russian disinformation in Ukraine, to lend further credibility that the proposed measures are applicable in the real world.

6.2 Chronology of Events

The key events and disinformation campaigns conducted by Russia that led to the annexation of Crimea are listed in Table 4.

Table 4: Chronology of key disinformation campaigns leading to the annexation of Crimea.

Time Period	Russian Disinformation
2000 to Mar 2014	Russia financed Ukrainian oligarchs with the aim of corrupting, privatizing, and monopolizing strategic Ukrainian industries, media, infrastructure, and hence politics. Consequently, multiple key Ukrainian societal functions could be easily influenced by Russian disinformation campaigns [195].
2004	During the period of the 2004 Ukrainian presidential election and "Orange Revolution", Russian information campaign portrayed Ukrainian fascists as preparing to invade and massacre local Russian speakers [195].
Nov 2013 to Feb 2014	From late 2013, tensions began to flare in eastern Ukraine as pro-Russian supporters rallied against pro-Ukrainian "Euromaidan" protesters. Ukrainian authorities believed that Russian state media, such as Russia 24, NTV, Channel One (ORT), and Russia-1, were painting a negative picture of Ukraine and Europe, creating divisions between the pro-Ukraine and pro-Kremlin camps. These channels were widely popular in Crimea at that time [52] [194]. Although many Ukrainians were protesting against corruption and oppressive policies of the existing regime, a vastly different narrative was painted to the Crimeans [194]. Russian information warfare portrayed the Euromaidan revolution as "fascist". There was also a barrage of anti-Ukrainian, "anti-fascist" and anti-Maidan propaganda that inflamed passions and stoked tensions [196]. The aim was to prevent the integration of Ukraine with the EU [52].
Feb 2014 to Mar 2014	From 22 February 2014, Russia's military spy agency, the GRU, launched a covert operation with the objective of influencing key decision-makers and the wider public to support the impending Russian military action, which was eventually launched on 27 February. This culminated in the seizure of the Crimean parliament building by armed men [197]. Multiple fake accounts were created on social media platforms and used to call Ukrainian demonstrators "Nazis" and "fascists", as well as post messages that bolstered Moscow's claim of radical Ukrainians inciting violence against Russians in the region [197]. Disinformation spread on social media also aimed

Time Period	Russian Disinformation
	to paint the new government in Kiev as illegitimate and determined to oppress Russian-speaking Ukrainians [195]. Following the seizure of the Crimean parliament building, social media was used to encourage Crimeans to support secession from Ukraine. Advertisements were also purchased on Facebook to increase the groups' popularity, which received nearly 200,000 views on 27 February [197].
Early-Mar 2014	Russia initially denied military intervention in Crimea, claiming "little green men" were groups of local militia illegitimately seizing Russian weapons and uniform for spontaneous self-defense against Euromaidan protesters in Crimea [195]. This sophisticated operation of half-truths and deliberate distortions were aimed at "clouding the picture of what was really going on". Although initially denying their presence in Crimea, the Russian story eventually changed and admitted to sending troops for the protection of ethnic Russians from Ukrainian "fascists" [198]. Russia's intervention in the Crimea was welcomed by many local residents because of their long-standing pro-Russian and pro-Soviet sympathies [196], as well as by the fear generated by Russian disinformation that Ukraine's ethnic Russians were being oppressed and attacked. For example, a video on Russian television supposedly showed the Ukrainian military firing on civilians, when the men filmed were actually Russian militia [52]. Such narratives led people in Crimea to believe their lives and freedoms were in danger from their fellow citizens in Kyiv. As the result, when the Russian military came offering protection, many gladly accepted [194].
Mid-Mar 2014	Crimea's pro-Russian government conducted a referendum and allegedly manipulated the results, stating that 96.7% were in favor of Crimea joining Russia [195], with a turnout of 83%. As no international observers were allowed, the result could not be verified. However, the website of the President of Russia's Council on Civil Society and Human Rights, Putin's own council, estimated that only about 55% voted for annexation, with a turnout of about 40% [199].

6.3 Disinformation Themes

This section highlights the key themes of the disinformation campaigns from Table 4, in terms of objectives and dissemination means.

6.3.1 Objectives

Influence Government Policy by Targeting Key Institutions and Personnel. By financing Ukrainian oligarchs, Russia intended to render key institutions related to the country's industry, media, and infrastructure, susceptible to Russian disinformation campaigns. The key objective was to find means to influence government policies. Russia's military spy agency, the GRU, also allegedly launched disinformation operations to influence key decision makers to support Russian military action.

Fuel Discontent within Society by Exploiting Fault Lines. Since 2004, Russia regularly painted pro-Ukrainian citizens as "fascist" and sought to create divisions and stoke tensions between pro-Ukrainian and pro-Russian camps. False stories were also spread to give the illusion that ethnic Russians were systematically being targeted, to gain wider public support of Russian military intervention. By exploiting these "slow burn" issues, many Crimeans welcomed Russian military intervention because of their long-standing pro-Russian and pro-Soviet sympathies.

Mask Russian Intervention and Increase Legitimacy for its Actions. Disinformation was spread to paint the new government in Kiev as illegitimate and determined to oppress Russian-speaking Ukrainians. To hide the fact that it was infringing on the sovereignty of another nation, Russia initially denied military intervention in Crimea and claimed that the "little green men" were groups of local militia. The story eventually changed and Russia claimed that troops were sent to Crimea to protect ethnic Russians from Ukrainian "fascists". The referendum results on the secession of Crimea were also allegedly manipulated to create the impression that there was significant support from the people, and that Russia was merely abiding by the people's will.

6.3.2 Dissemination Means

Inauthentic Accounts on Social Media. Social media was used to encourage Crimeans to support secession from the Ukraine. Multiple fake accounts were created on social media platforms to label Ukrainian demonstrators as "fascist" and bolster Moscow's claims of violence against ethnic Russians in the region.

Advertisements on Social Media. Advertisements were purchased on Facebook to increase the popularity of inauthentic social media groups created by Russia, which served to spread disinformation and encourage secession from Ukraine.

Russian State Media. Russian state media, such as Russia 24, NTV, Channel One (ORT), and Russia-1, were used to paint negative pictures of Ukraine and Europe. These channels were widely popular in Crimea at that time. The aim was to prevent the integration of Ukraine with the EU.

6.4 Validation of System Applicability

This section primarily discusses how the proposed system could have prevented or mitigated the effects of Russia's disinformation campaigns from Section 6.3. The lack of effective countermeasures employed by Ukraine during that time period, and examples of countermeasures that have since been introduced, are also discussed. These serve to validate the applicability of the system in a real scenario. The discussions are organized around the seven categories of countermeasures from Chapter 4, as they are the value instruments of the system.

6.4.1 Fact-Checking

Lack of Fact-Checking Prior to Annexation. Local fact-checking initiatives were only implemented about two weeks prior to the annexation of Crimea, and hence were perhaps too late to effectively mitigate the effects of the disinformation campaigns. For example, the fact-checking website, StopFake, was only launched on 2 March 2014 to "verify and refute disinformation and propaganda about events in Ukraine being circulated in the media". This fact-checking initiative is completely independent from the Ukrainian government [200], quite similar to the proposed option for the architectural decision in Section 5.2.4 on the "degree of government involvement in fact-checking initiatives".

Potential Effectiveness of Fact-Checking to Prevent Annexation. As shown in the proposed architecture (Figure 4), key functions for fact-checking involve "identifying", "reducing access", or "removing" the falsehoods and/or the sources of falsehoods, as well as "correcting" the public by broadcasting the facts. This would have mitigated the effects of Russian disinformation – such as the simmering of tensions between ethnic Russians and the rest of society, which was one of the key issues that led to the successful annexation. Think tanks or NGOs could also have helped to watchdog and scrutinize key institutions and politicians, identifying and calling out those sympathetic to or were influenced by Russian disinformation. This could have prevented Ukrainian government policies from being influenced by foreign powers. Russian military intervention and results of the referendum could also have been more closely scrutinized and facts highlighted to the rest of the world.

Fact-Checking Initiatives Implemented Since Annexation. To reduce access to sources of disinformation, the Ukrainian government started banning Russian channels from 2014 [201], and in 2017, blocked Russian social media networks and search engine. A year later, 192 websites with alleged pro-Russian sympathies were also blocked [202]. Although Facebook launched its fact-checking program in December 2016, it seems that there are still no Ukrainian fact-checking organizations participating in the initiative [127] [203]. Hence this is certainly one area that Ukraine needs to look into.

6.4.2 Research and Technology (R&T)

Lack of R&T Prior to Annexation. In 2014, Russian troll factories were responsible for promoting falsehoods, which led to a "tsunami" of fake news that prevented the Ukrainian government from monitoring everything [204]. It is evident that sophisticated and advanced tools to aid in fact-checking and identification of inauthentic accounts were lacking at that point in time.

Potential Effectiveness of R&T to Prevent Annexation. As shown in the proposed architecture (Figure 4), a key function of R&T is "enhancing" the fact-checking initiatives through automation and advanced technologies, to more effectively and efficiently identify and remove falsehoods and their sources. For example, automated fact-checking tools, as well as advanced techniques such as machine learning to identify and remove inauthentic accounts could have aided the Ukrainian government significantly by reducing both the sources of falsehoods and falsehoods themselves, thus preventing the spread of disinformation.

R&T Initiatives Implemented Since Annexation. The Ukrainian government was among the first countries to approach Facebook and Twitter for help but were brushed-off. The social media sites have since realized their errors and have taken down state-tied accounts in Russia [205]. Facebook engineers have also since built automated tools to tackle the large-scale problem of fake accounts [204]. In March 2019, Facebook removed more than 1,900 pages, groups, and accounts that were linked to Russia. Some of them were engaged in "coordinated inauthentic behavior", which promoted content including the ongoing conflict in eastern Ukraine. Facebook also mentioned the need to continue "building better technology" to uncover such behavior [206].

6.4.3 Business Model

Business Model Prior to Annexation. During that period, advertisements could still be purchased on Facebook to spread disinformation and encourage secession from Ukraine. It is thus evident that online platforms were focused on earning revenue through advertisements and did not have adequate safeguards in place to prevent their misuse.

Potential Effectiveness of Business Model Initiatives to Prevent Annexation. As shown in the proposed architecture (Figure 4), key functions related to a "changing" business model are "revealing" sponsor information and "reducing" falsehoods. Proposed countermeasures such as increasing transparency of sponsored content and even restricting the sale of political advertisements, could have curbed the spread of disinformation on social media that aimed to support secession from the Ukraine. This could have reduced the level of unrest and Crimean support for independence.

Business Model Initiatives Implemented Since Annexation. In March 2019, Facebook introduced transparency requirements for advertisements related to politics and elections in Ukraine, similar to that in other countries such as US, UK, Brazil, Israel, and India. This would inform users of the sponsor of political advertisements [207]. Twitter has taken it a step further and, since November 2019, has prohibited the promotion of political content on its platform – defined as "content that references a candidate, political party, elected or appointed government official, election, referendum, ballot measure, legislation, regulation, directive, or judicial outcome" [208]. If such a policy is implemented on all online platforms, disinformation such as those aimed at garnering support for Crimea's referendum on independence, would be prohibited.

6.4.4 Quality Journalism

Lack of Quality Journalism Prior to Annexation. Based on a survey conducted in 2014, television was the dominant news medium in Crimea, with 95.7% watching TV for news at least weekly. Unfortunately, TV channels owned by the Russian state were also the most important sources of news and information for Crimeans [209]. Thus it seems that Crimeans were heavily exposed to Russian propaganda and lacked quality sources of news during that period. Ukrainian authorities were initially poor at getting quality information out to their own citizens and the international community, thus forcing Ukrainian civil society to step up [198]. Consequently, the Ukraine Crisis Media Center was launched in March 2014 by Ukrainian experts with backgrounds in international relations, communications, and public relations, with the aim of providing accurate and up-to-date information on the events in Ukraine [210].

Potential Effectiveness of Quality Journalism to Prevent Annexation. As shown in the proposed architecture (Figure 4), key functions associated with quality journalism include "identifying" and "weakening" the influence of sources of falsehoods. Russian state media could have been identified and "blacklisted" if deemed to be a source of falsehoods. This could have prevented Russian state media from becoming an important source of information for Crimeans. Tightening of journalistic controls to enhance quality journalism and improve transparency of information sources, could also have helped to provide the public with more reliable and trustworthy sources of information, thus preventing them from turning to less reliable sources.

Quality Journalism Initiatives Implemented Since Annexation. As a means to tighten journalistic controls, the Ukrainian government expelled dozens of Russian journalists and revoked the accreditation of more than 100 media outlets between 2014 and 2015 [201]. Ukraine also created a Ministry of Information Policy in December 2014 that is tasked to ensure accurate information is available to Ukrainians and the rest of the world, as well as ensure that disinformation attempts are challenged and corrected [198]. This is quite similar to the proposed option of the architectural decision in Section 5.2.6 on "structure of national agency". It was also announced that the Ministry would create a broadcaster to provide news about Ukraine internationally. This platform was eventually launched in October 2015 [198]. It seems that the government's objective was to provide a reliable source of quality information.

6.4.5 Education

Lack of Education Prior to Annexation. The International Research & Exchange Board (IREX) media literacy program in Ukraine only commenced in October 2015, and the head of IREX's Media and Information Literacy Initiatives admitted that they were "ten years too late" [211]. Prior to the commencement of the program, a survey in 14 regions across Ukraine showed that more than 50% of the respondents were unable to distinguish fake news stories and blindly trusted information from the media [212]. It was also found that only 23% of Ukrainians cross-checked news sources [213]. Hence it is evident that media literacy was lacking during the critical period prior to the annexation of Crimea.

Potential Effectiveness of Education to Prevent Annexation. As shown in the proposed architecture (Figure 4), the key function of education is "informing" the public. A more informed and discerning Crimean public would have been better positioned to identify falsehoods spread by Russia. This would have mitigated the impact of Russia's disinformation campaigns, especially those exploiting fault lines and fueling discontent within society over a significant period of time. Disinformation about ethnic Russians being targeted would also not have been so easily accepted by the public, potentially reducing the level of support for Russian military intervention and Crimean independence. In addition, key government personnel should have been trained to identify and resist disinformation operations, thus preventing government policy from being influenced by foreign powers.

Education Initiatives Implemented Since Annexation. From October 2015 to March 2016, the IREX conducted training for 15,000 people of all ages and professional backgrounds to equip Ukrainian citizens to identify disinformation and demand better quality information [201]. In 2018, students in four cities across Ukraine received training conducted by IREX to help identify falsehoods. The aim is for the program to be conducted in approximately 650 schools across Ukraine by 2021 [211].

6.4.6 Social Cohesion and Trust

Lack of Social Cohesion and Trust Prior to Annexation. It can be said that the Ukrainian government failed to build and maintain the trust of its people, and particularly those in Crimea.

One of the first acts of the new Ukrainian government, which was installed following the 2014 revolution, was to annul a bill that allowed the Russian language to be the second official language in regions with Russian-speaking population [214]. This could be considered a "misstep" by the government, and caused it to lose any remaining influence and trust in Russia-friendly regions. The move infuriated ethnic Russians and served as evidence that the protestors were indeed radical fascists and that the new government was intent on pressing for a nationalistic agenda [215].

Potential Effectiveness of Social Cohesion and Trust to Prevent Annexation. As shown in the proposed architecture (Figure 4), the key function of social cohesion and trust is "minimizing" the impact of falsehoods, and to do so, "identifying" vulnerabilities is critical. A cohesive society that has trust between its communities and in the government can serve as an effective countermeasure against Russian disinformation. It is this lack of social cohesion and trust that was exploited by disinformation campaigns to great effect – pitting pro-Ukraine and pro-Kremlin camps against each other, and instilling fear in ethnic Russians that they were being targeted. These vulnerabilities in society should have been identified early and regularly through studies conducted by the government, think tanks, or academia; policies and ground-up initiatives could have been introduced to reinforce social cohesion and trust. Key institutions and personnel vulnerable to Russian influence should also have been identified and addressed. In addition, if the Crimeans had trust in the government, they would not have easily believed that their own government was determined to oppress Russian-speaking Ukrainians.

Social Cohesion and Trust Initiatives Implemented Since Annexation. Despite Ukraine's ever-changing political landscape, including the recent reforms in key sectors, Ukrainian authorities have not been prioritizing social integration issues such as the protection of internally displaced persons (IDPs), which is currently one of the largest in the world [216]. Hence, it would seem that social integration is still not a priority of the Ukrainian government at the moment. In April 2019, the Ukrainian parliament approved a law that would require all citizens to know the Ukrainian language and make it mandatory for civil servants, soldiers, doctors, and teachers [217]. Proponents of the law argue that it would strengthen national identity [218], and hence possibly social cohesion, while opponents have argued that it would instead further divide Russian and non-Russian speaking citizens [219]. Hence it is evident that the issue of social cohesion and trust is one area that needs to be further addressed by the Ukrainian government.

6.4.7 Legislation and Regulation

Lack of Legislation and Regulation Prior to Annexation. To date, the only means of fighting defamation in Ukraine is by filing a law suit, which can take years to be resolved. This is ineffective during elections or disinformation campaigns launched by Russian that aim to damage the reputation of public figures. Hence, there is a need for new legislation to prevent "outright lies" from causing damage long before they can be refuted in court [220]. This is similar to the proposed option of the architectural decision in Section 5.2.1 on "legislation".

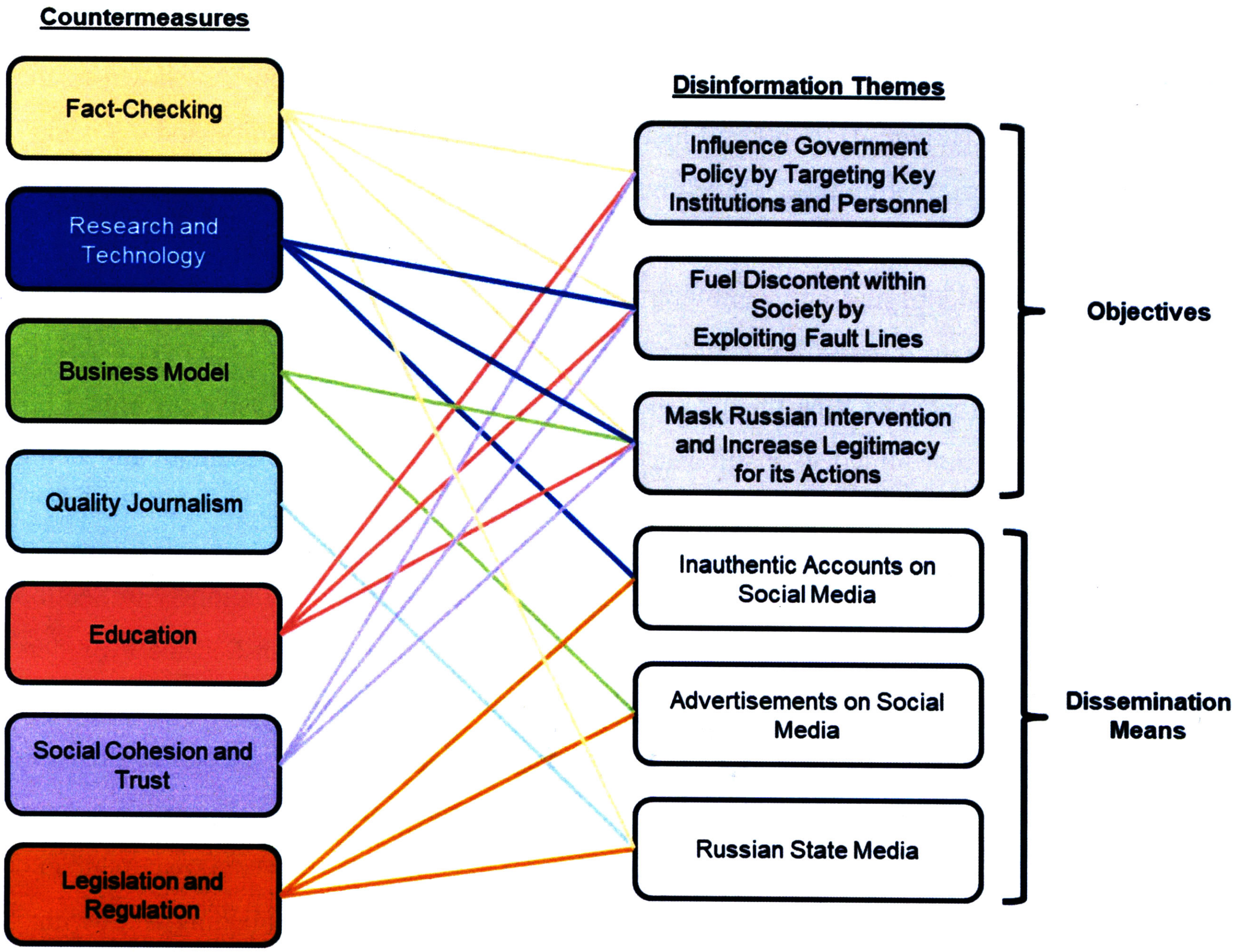
Potential Effectiveness of Legislation and Regulation to Prevent Annexation. As shown in the proposed architecture (Figure 4), two key functions of legislation are "punishing and deterring" those responsible for creating and spreading disinformation, as well as "encouraging or compelling" online platforms to be part of the solution. For example, online platforms should be actively involved in identifying and removing inauthentic accounts, as well as preventing the misuse of advertisements. Legislation could also have been used to restrict or block access to sources of falsehoods, such as Russian state media. In particular, this could have prevented the spread of falsehoods that aimed to instill fear that ethnic Russians were systematically being targeted, as well as to sow discord and fuel discontent within society.

Legislation and Regulation Implemented Since Annexation. As mentioned previously in Section 6.4.1, sanctions have been imposed against Russian companies, and Internet Service Providers (ISPs) were ordered to block access to Russian social media platforms, search engine, and websites as part of a national security decree [221]. In September 2015, a law was introduced that provides new rules on ownership of television stations in Ukraine, and establishes financial disclosure requirements for the owners [222]. The law aims to ensure transparency of media ownership and prevents certain entities or individuals, such as those from Russia, from establishing and owning broadcast companies and program service providers [222] [223]. The Ukrainian parliament is also considering two draft legislations – one to block sources of information that are considered a threat to national security [221] [224]; and another to "criminalize the dissemination of false information" both online and in the media, including severe penalties on outlets that disseminate election-related falsehoods [224].

6.5 Summary

A summary of how the disinformation themes are primarily addressed by the various countermeasures, is depicted in Figure 5. This qualitatively validates the applicability of the system in a real scenario, as all the disinformation themes from the case study have been addressed by the countermeasures proposed.

Figure 5: Mapping of countermeasures to disinformation themes.


Countermeasures
Fact-Checking
Research and Technology
Business Model
Quality Journalism
Education
Social Cohesion and Trust
Legislation and Regulation
Disinformation Themes
Influence Government Policy by Targeting Key Institutions and Personnel
Fuel Discontent within Society by Exploiting Fault Lines
Mask Russian Intervention and Increase Legitimacy for its Actions
Inauthentic Accounts on Social Media
Advertisements on Social Media
Russian State Media
Objectives
Dissemination Means

Chapter 7: Conclusion

This chapter aims to summarize the findings from the research, as well as identify limitations of the current work and possible future research on this topic.

7.1 Summary of Research Findings

Chapter 1 highlights that technological advances and social changes in the Internet era has made disinformation a more potent problem than ever before. Although disinformation campaigns can have a significant impact on the population, the manpower and costs required to carry them out are disproportionately low. It is thus imperative that nations are adequately prepared to deal with this asymmetric and dangerous threat using a suite of countermeasures, as there is no one silver bullet to tackle this problem. A systems thinking methodology is thus selected to develop an integrated approach to deal with this complex issue at the national-level, with the ARIES framework serving as a guide for the entire thought process.

An overview of disinformation is discussed in Chapter 2. Disinformation agents may be foreign or local, state-sponsored or merely civilians. They may seek various objectives such as influencing government policy, discrediting the government, achieving an election outcome, or fueling discontent within society. In terms of impact, disinformation could have short-term or long-term effects, threaten national security, or cause harm to the democratic system and even individuals. Several key challenges in combatting disinformation were also discussed. This include our very own cognitive tendencies, the weakness of truth compared to falsehoods, and the further and faster reach of falsehoods.

The collective group of organizations that need to be involved at the national-level is discussed in Chapter 3. These key stakeholders include the public, the government, online platforms, traditional media organizations, fact-checking organizations, educational and research institutes, and other non-governmental organizations; all of which have an important role to play in preventing the spread of disinformation. Through stakeholder analyses, a key observation is that although online platforms are important to the system as they are a key source of falsehoods, there is little incentive for them to be committed to the success of the system. In fact, their business models seem to facilitate the spread of disinformation. Hence, it is critical that something is done about this.

Chapter 4 provides an overview of possible countermeasures that are being proposed or have been adopted by countries worldwide. The various countermeasures are grouped, aggregated, and abstracted into seven main categories – fact-checking, research and technology, business model, quality journalism, education, social cohesion and trust, and legislation and regulation. These countermeasures include both short and long-term efforts, as well as the involvement of multiple stakeholders. It is argued that the introduction of legislation and other forms of regulation appear necessary to provide the "incentive" for online platforms to implement or support many of the proposed countermeasures.

The proposed architecture is developed in Chapter 5 by building on the proposed countermeasures and identifying key architectural decisions that each government would need to consider when developing its customized solution. Six architectural decisions are identified – type of legislation, entity that decides what is true or false, party that would be held accountable, degree of government involvement in fact-checking initiatives, extent of anonymity allowed on the Internet, and the structure of the national agency. After the pros and cons of each option is discussed, a preferred option that is likely to be applicable to most nations is then proposed for each architectural decision. Finally, the system architecture is developed and represented using OPD.

In Chapter 6, a case study on the annexation of Crimea by Russia is used to qualitatively validate the proposed system and ascertain its applicability in a real scenario. Key themes of Russian disinformation campaigns conducted from 2000 to Mar 2014 are first identified: (a) the main objectives appear to be influencing government policy by targeting key institutions and personnel, fueling discontent within society by exploiting fault lines, and masking Russian intervention and increasing legitimacy for its actions; (b) social media and Russian state media appear to be the primary means of spreading disinformation. It is argued that all the disinformation themes from the case study can be adequately addressed by the proposed countermeasures, thus qualitatively validating the applicability of the system in a real scenario.

7.2 Limitations

The author has identified three key limitations of the current work, and are discussed in this section.

Not All Individual Countermeasures Validated. Under each category of the countermeasures in Chapter 4, there are multiple individual countermeasures that have been proposed or implemented around the world. It is not yet known if every single one of these countermeasures are indeed effective, and further studies may eventually conclude otherwise. For example, while tagging of falsehoods have been shown to be beneficial in some studies, it was also highlighted that a potential downside is the implied truth effect – whereby false headlines that fail to get tagged are considered validated and thus are seen as more accurate. Consequently, the list of individual countermeasures under each category may need to be amended if new studies deem any of them ineffective.

Proposed Architecture May Not Be Suitable for All Nations. As the author has intentionally decided to avoid developing a response for a specific country, the proposed architecture should be considered as a baseline architecture that needs to be further customized for each nation's specific needs and circumstances. Hence it is critical that any architect, who is tasked to develop a specific nation's response against disinformation, takes into consideration and prioritizes the needs of all major stakeholders before selecting the preferred option for each architectural decision in Section 5.2. The architect could also opt to develop a set of evaluation criteria based on what is deemed to be important to the specific nation, and use it to perform a more detailed assessment on the various architectural options.

Lack of Quantitative Validation or Additional Case Studies. With regard to the validation process, two areas of improvement were identified. Firstly, validation was done qualitatively without any quantitative evidence on whether the system would indeed have adequately combatted Russia's disinformation campaigns. A model could be developed to simulate the impact of disinformation and the effectiveness of the proposed countermeasures, in order to quantitatively validate the applicability of the system. This would require more in-depth research into the effectiveness of each countermeasure, either by conducting new experiments or using results from documented studies. On-the-ground surveys may also be needed to understand the psyche of Crimeans and how "deep" the fault lines in society actually were, as this may impact the efficacy of some of the proposed countermeasures, particular those related to enhancing social cohesion and trust. Secondly, due to the time accorded to the thesis, a single case study is used to validate the applicability of the proposed system. Time permitting, additional case studies could be included for a more rigorous qualitative analysis.

7.3 Future Research

In this section, three possible areas for future research are discussed for consideration.

System Dynamics Modelling. One possible tool for modelling the impact of disinformation would be system dynamics, which is a method to enhance learning about complex systems – dynamic complexity, sources of policy resistance, and to design more effective policies. For example, system dynamics has been successfully used to model the dynamics of infectious diseases, and to study the effectiveness of immunization or quarantine to prevent an epidemic [225]. It would seem possible to draw parallels between epidemics of infectious diseases and how disinformation can go "viral". One can also examine how countermeasures such as education can effectively "immunize" a population against disinformation, or how "quarantining" falsehoods by reducing or blocking access to them can be effective.

Follow-up by Security and Intelligence Community. For this thesis, discussions on classified countermeasures and work conducted by intelligence and security agencies have been limited, as such information is not readily available. For example, offensive countermeasures to respond to disinformation operations have been largely omitted. It would be useful for the security and intelligence community to follow-up on this research and include their own specific countermeasures to see how it complements the proposed system. This would provide a more complete response against a potential foreign adversary.

Implementation Details. Most of the current discussions have been done at a more strategic-level, with tactical-level or implementation details largely omitted. It would be useful to consider such finer details, including the potential cost and timeline for implementation associated with each individual countermeasure. This would enable the architect and decision-makers to weigh the cost-effectiveness of the various countermeasures, to decide which should be implemented based on the available budget and time of the specific nation. Cost-effectiveness can potentially be defined as the ratio of the effectiveness or value of the countermeasure, to the effort (i.e. in terms of cost and timeline) needed for it to be implemented.

References

[1] C. Ireton and J. Posetti, "Journalism, 'Fake News' & Disinformation," United Nations Educational, Scientific and Cultural Organization (UNESCO), 2018. [Online]. Available: https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0.pdf. [Accessed 23 October 2019].

[2] Select Committee, "Report of the Select Committee on Deliberate Online Falsehoods - Causes, Consequences and Countermeasures," Parliament of Singapore, 19 September 2018. [Online]. Available: https://sprs.parl.gov.sg/selectcommittee/selectcommittee/download?id=1&type=subReport. [Accessed 5 October 2019].

[3] E. Lucas and P. Pomeranzev, "Winning the Information War," The Center for European Policy Analysis, August 2016. [Online]. Available: https://cepa.ecms.pl/files/?id_plik=2715. [Accessed 21 October 2019].

[4] C. Wardle, "Misinformation Has Created a New World Disorder," Scientific American, 1 September 2019. [Online]. Available: https://www.scientificamerican.com/article/misinformation-has-created-a-new-world-disorder/. [Accessed 7 October 2019].

[5] M. Choy and M. Chong, "Seeing Through Misinformation: A Framework for Identifying Fake Online News," SSON Analytics and Singapore Management University, March 2018. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1804/1804.03508.pdf. [Accessed 23 October 2019].

[6] A. White, "Ethics in the News," Ethical Journalism Network, 2017. [Online]. Available: https://ethicaljournalismnetwork.org/wp-content/uploads/2017/01/ejn-ethics-in-the-news.pdf. [Accessed 23 October 2019].

[7] K. Born, "The future of truth: Can philanthropy help mitigate misinformation?," The William and Flora Hewlett Foundation, 8 June 2017. [Online]. Available: https://hewlett.org/future-truth-can-philanthropy-help-mitigate-misinformation/. [Accessed 21 October 2019].

[8] W. Ahmad, "Dealing with Fake News: Policy and Technical Measures," Massachusetts Institute of Technology (MIT) Internet Policy Research Initiative (IPRI), 20 April 2018. [Online]. Available: https://internetpolicy.mit.edu/wp-content/uploads/2018/04/Fake-news-recommendations-Wajeeha-MITs-IPRI.pdf. [Accessed 23 October 2019].

[9] E. Crawley, B. Cameron and D. Selva, System Architecture - Strategy and Product Development for Complex Systems, Pearson, 2016.

[10] D. Nightingale and D. Rhodes, Architecting the Future Enterprise, MIT Press, 2015.

[11] J. Haines, "Russia's Use of Disinformation in the Ukraine Conflict," Foreign Policy Research Institute, 17 February 2015. [Online]. Available: https://www.fpri.org/article/2015/02/russias-use-of-disinformation-in-the-ukraine-conflict/. [Accessed 21 October 2019].

[12] C. Wardle, "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making," Council of Europe Report, 27 September 2017. [Online]. Available:

https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c. [Accessed 7 October 2019].

[13] F. Filloux, "You can't sell news for what it costs to make," Medium, 6 August 2017. [Online]. Available: https://medium.com/the-walkley-magazine/you-cant-sell-news-for-what-it-costs-to-make-7a4def964ffa. [Accessed 7 October 2019].

[14] Facebook Business, "Find Your Audience," Facebook, 2019. [Online]. Available: https://www.facebook.com/business/ads/ad-targeting. [Accessed 8 October 2019].

[15] Twitter Business, "Ad Targeting," Twitter, 2019. [Online]. Available: https://business.twitter.com/en/targeting.html. [Accessed 8 October 2019].

[16] Google Ads Help, "Targeting your ads," Google, 2019. [Online]. Available: https://support.google.com/google-ads/answer/1704368?hl=en. [Accessed 8 October 2019].

[17] A. Pasick, "Facebook says it can sway elections after all—for a price," Quartz, 1 March 2017. [Online]. Available: https://qz.com/922436/facebook-says-it-can-sway-elections-after-all-for-a-price/. [Accessed 8 October 2019].

[18] Facebook Business, "Toomey for Senate," Facebook, 2019. [Online]. Available: https://www.facebook.com/business/success/toomey-for-senate. [Accessed 8 October 2019].

[19] K. Conger and D. Cameron, "Here Are 14 Russian Ads That Ran on Facebook During The 2016 Election," Gizmodo, 1 November 2017. [Online]. Available: https://gizmodo.com/here-are-14-russian-ads-that-ran-on-facebook-during-the-1820052443. [Accessed 8 October 2019].

[20] D. Cameron, "Facebook Now Estimates 126 Million Americans Viewed Russian-Bought Political Propaganda," Gizmodo, 30 October 2017. [Online]. Available: https://gizmodo.com/facebook-now-estimates-126-million-americans-viewed-rus-1819992713. [Accessed 8 October 2019].

[21] C. Timberg, E. Dwoskin, A. Entous and K. Demirjian, "Russian ads, now publicly released, show sophistication of influence campaign," The Washington Post, 1 November 2017. [Online]. Available: https://www.washingtonpost.com/business/technology/russian-ads-now-publicly-released-show-sophistication-of-influence-campaign/2017/11/01/d26aead2-bf1b-11e7-8444-a0d4f04b89eb_story.html. [Accessed 23 October 2019].

[22] S. Bradshaw and P. Howard, "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation," Computational Propaganda Research Project, University of Oxford, 20 July 2018. [Online]. Available: http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf. [Accessed 7 October 2019].

[23] M. Czuperski and B. Nimmo, "#ElectionWatch: Final Hours Fake News Hype in Germany," Digital Forensic Research Lab, 23 September 2017. [Online]. Available: https://medium.com/dfrlab/electionwatch-final-hours-fake-news-hype-in-germany-cc9b8157cfb8. [Accessed 8 October 2019].

[24] C. Wardle, "Fake News. It's Complicated.," First Draft, 16 February 2017. [Online]. Available: https://firstdraftnews.org/fake-news-complicated/. [Accessed 23 October 2019].

[25] Subcommittee on Crime and Terrorism, "Sean Edgett's Answers to Questions for the Record," Senate Committee on the Judiciary, 19 January 2018. [Online]. Available:

https://www.judiciary.senate.gov/imo/media/doc/Edgett%20Responses.pdf. [Accessed 8 October 2019].

[26] B. Nimmo, "Why Bot Makers Dream of Electric Sheep," Digital Forensic Research Lab, 27 June 2017. [Online]. Available: https://www.atlanticcouncil.org/blogs/new-atlanticist/why-bot-makers-dream-of-electric-sheep/. [Accessed 8 October 2019].

[27] I. Lapowsky, "Parkland Conspiracies Overwhelm the Internet's Broken Trending Tools," WIRED, 21 February 2018. [Online]. Available: https://www.wired.com/story/youtube-facebook-trending-tools-parkland-conspiracy/. [Accessed 8 October 2019].

[28] J. Herrman, "The Making of a No. 1 YouTube Conspiracy Video After the Parkland Tragedy," The New York Times, 21 February 2018. [Online]. Available: https://www.nytimes.com/2018/02/21/business/media/youtube-conspiracy-video-parkland.html. [Accessed 8 October 2019].

[29] C. Silverman and J. Singer-Vine, "The True Story Behind The Biggest Fake News Hit Of The Election," BuzzFeed News, 16 December 2016. [Online]. Available: https://www.buzzfeednews.com/article/craigsilverman/the-strangest-fake-news-empire. [Accessed 9 October 2019].

[30] W. Soeriaatmadja, "Man who uploaded controversial video of ex-Jakarta governor Ahok sentenced to jail," The Straits Times, 14 November 2017. [Online]. Available: https://www.straitstimes.com/asia/se-asia/man-who-uploaded-controversial-ahok-video-sentenced-to-jail. [Accessed 9 October 2019].

[31] E. Maulia, "Fake news charges emotionally driven Jakarta election," Nikkei Asian Review, 13 February 2017. [Online]. Available: https://asia.nikkei.com/Politics/Indonesian-fake-news-spread-ahead-of-Jakarta-election2. [Accessed 9 October 2019].

[32] BBC, "Mass prayer rally in Jakarta against governor 'Ahok'," BBC, 2 December 2016. [Online]. Available: https://www.bbc.com/news/world-asia-38178764. [Accessed 9 October 2019].

[33] R. Khalaf, "If you thought fake news was a problem, wait for ‘deepfakes’," Financial Times, 25 July 2018. [Online]. Available: https://www.ft.com/content/8e63b372-8f19-11e8-b639-7680cedcc421. [Accessed 9 October 2019].

[34] C. Soon and S. Goh, "What Lies Beneath the Truth: A Literature Review on Fake News, False Information and More," Institute of Policy Studies, National University of Singapore, 30 June 2017. [Online]. Available: https://lkyspp.nus.edu.sg/docs/default-source/ips/report_what-lies-beneath-the-truth_a-literature-review-on-fake-news-false-information-and-more_300617.pdf. [Accessed 19 October 2019].

[35] A. Entous, E. Nakashima and G. Jaffe, "Kremlin trolls burned across the Internet as Washington debated options," The Washington Post, 25 December 2017. [Online]. Available: https://www.washingtonpost.com/world/national-security/kremlin-trolls-burned-across-the-internet-as-washington-debated-options/2017/12/23/e7b9dc92-e403-11e7-ab50-621fe0588340_story.html. [Accessed 29 October 2019].

[36] B. Nimmo, "Written evidence submitted to UK Digital, Culture, Media and Sport Committee on “Fake News”," UK Parliament, April 2017. [Online]. Available: http://data.parliament.uk/WrittenEvidence/CommitteeEvidence.svc/EvidenceDocument/Culture,%20Media%20and%20Sport/Fake%20News/written/68987.html. [Accessed 5 October 2019].

[37] J. Janda, "Full-Scale Democratic Response to Hostile Disinformation Operations," European Values, 20 June 2016. [Online]. Available: https://www.europeanvalues.net/wp-content/uploads/2016/06/Full-Scale-Democratic-Response-to-Hostile-Disinformation-Operations-1.pdf. [Accessed 5 October 2019].

[38] P. Pomerantsev and M. Weiss, "The Menace of Unreality: How the Kremlin Weaponizes Information, Culture and Money," Institute of Modern Russia, 22 November 2014. [Online]. Available: https://imrussia.org/media/pdf/Research/Michael_Weiss_and_Peter_Pomerantsev__The_Menace_of_Unreality.pdf. [Accessed 6 October 2019].

[39] National Intelligence Council, "Assessing Russian Activities and Intentions in Recent US Elections," Intelligence Community Assessment, 6 January 2017. [Online]. Available: https://www.dni.gov/files/documents/ICA_2017_01.pdf. [Accessed 5 October 2019].

[40] A. Chen, "The Agency," The New York Times, 2 June 2015. [Online]. Available: https://www.nytimes.com/2015/06/07/magazine/the-agency.html?_r=0. [Accessed 5 October 2019].

[41] A. Prokop, "The new Mueller indictments tell us a lot about Russian trolls," Vox, 16 February 2018. [Online]. Available: https://www.vox.com/2018/2/16/17020966/russia-indictments-mueller-internet-research-agency. [Accessed 5 October 2019].

[42] D. Broniatowski, A. Jamison, S. Qi, L. AlKulaib, T. Chen, A. Benton, S. Quinn and M. Dredze, "Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate," *American Journal of Public Health,* vol. 108, no. 10, pp. 1378-1384, 2018.

[43] BBC, "Russia trolls 'spreading vaccination misinformation' to create discord," BBC, 24 August 2018. [Online]. Available: https://www.bbc.com/news/world-us-canada-45294192. [Accessed 6 October 2019].

[44] Lord Justice Leveson, "Leveson Inquiry - Report into the culture, practices and ethics of the press," The Leveson Inquiry, 29 November 2012. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/270941/0780_ii.pdf. [Accessed 6 October 2019].

[45] D. Arkin and B. Popken, "How the internet's conspiracy theorists turned Parkland students into 'crisis actors'," NBC News, 21 February 2018. [Online]. Available: https://www.nbcnews.com/news/us-news/how-internet-s-conspiracy-theorists-turned-parkland-students-crisis-actors-n849921. [Accessed 7 October 2019].

[46] S. Banjo and N. Lung, "How Fake News and Rumors Are Stoking Division in Hong Kong," Bloomberg, 11 November 2019. [Online]. Available: https://www.bloomberg.com/news/articles/2019-11-11/how-fake-news-is-stoking-violence-and-anger-in-hong-kong. [Accessed 14 November 2019].

[47] M. Novak, "This Video of 'Muslims Celebrating the Paris Terror Attack' Is Totally Fake," Gizmodo, 4 April 2017. [Online]. Available: https://gizmodo.com/this-video-of-muslims-celebrating-the-paris-terror-atta-1794523785. [Accessed 6 October 2019].

[48] J. Stahl, "Purveyor of Fake News Says He Targeted Trump Supporters, Influenced Election," Slate, 17 November 2016. [Online]. Available: https://slate.com/news-and-politics/2016/11/purveyor-of-fake-news-says-he-targeted-trump-supporters-influenced-election.html. [Accessed 7 October 2019].

[49] S. French, "This person makes $10,000 a month writing fake news," MarketWatch, 18 November 2016. [Online]. Available: https://www.marketwatch.com/story/this-person-makes-10000-a-month-writing-fake-news-2016-11-17. [Accessed 7 October 2019].

[50] J. Griffiths, "How Chinese internet trolls go after Beijing's critics overseas," CNN, 18 April 2019. [Online]. Available: https://www.cnn.com/2019/04/18/tech/china-uyghurs-internet-trolls-facebook-intl/index.html. [Accessed 7 October 2019].

[51] Y. Yang, "China's Communist party raises army of nationalist trolls," Financial Times, 29 December 2017. [Online]. Available: https://www.ft.com/content/9ef9f592-e2bd-11e7-97e2-916d4fbac0da. [Accessed 7 October 2019].

[52] D. Paulo and D. Heng, "How fake news fanned the flames of war in Ukraine," Channel News Asia, 22 December 2018. [Online]. Available: https://www.channelnewsasia.com/news/cnainsider/how-fake-news-sparked-war-ukraine-russia-crimea-select-committee-11055154. [Accessed 7 October 2019].

[53] K. Hjelmgaard, "There is meddling in Germany's election — not by Russia, but by U.S. right wing," USA Today, 25 September 2017. [Online]. Available: https://www.usatoday.com/story/news/world/2017/09/20/meddling-germany-election-not-russia-but-u-s-right-wing/676142001/. [Accessed 7 October 2019].

[54] J. Robertson, M. Riley and A. Willis, "How to Hack an Election," Bloomberg Businessweek, 31 March 2016. [Online]. Available: https://www.bloomberg.com/features/2016-how-to-hack-an-election/. [Accessed 7 October 2019].

[55] K. Rawlinson, "Finsbury Park-accused trawled far-right groups online, court told," The Guardian, 23 January 2018. [Online]. Available: https://www.theguardian.com/uk-news/2018/jan/23/finsbury-park-accused-wanted-to-kill-all-muslims-court-told. [Accessed 7 October 2019].

[56] E. J. Kirby, "The city getting rich from fake news," BBC, 5 December 2016. [Online]. Available: https://www.bbc.com/news/magazine-38168281. [Accessed 7 October 2019].

[57] N. Salleh, "Muis condemns ISIS video featuring Singaporean," The Straits Times, 28 September 2017. [Online]. Available: https://www.straitstimes.com/singapore/muis-condemns-isis-video-featuring-sporean. [Accessed 7 October 2019].

[58] Majlis Ugama Islam Singapura (MUIS), "Media Statement - MUIS Statement on ISIS Video," MUIS, 27 September 2017. [Online]. Available: https://www.muis.gov.sg/Media/Media-Releases/27-Sep-17-Media-Statement-Muis-Statement-on-ISIS-Video. [Accessed 7 October 2019].

[59] P. Foster, "'Bogus' AP tweet about explosion at the White House wipes billions off US markets," The Telegraph, 23 April 2013. [Online]. Available: https://www.telegraph.co.uk/finance/markets/10013768/Bogus-AP-tweet-about-explosion-at-the-White-House-wipes-billions-off-US-markets.html. [Accessed 9 October 2019].

[60] The World Staff, "In Myanmar, fake news spread on Facebook stokes ethnic violence," Public Radio International, 1 November 2017. [Online]. Available: https://www.pri.org/stories/2017-11-01/myanmar-fake-news-spread-facebook-stokes-ethnic-violence. [Accessed 15 October 2019].

[61] BBC, "Why is there communal violence in Myanmar?," BBC, 3 July 2014. [Online]. Available: https://www.bbc.com/news/world-asia-18395788. [Accessed 15 October 2019].

[62] S. Frenkel and D. Wakabayashi, "After Florida School Shooting, Russian 'Bot' Army Pounced," The New York Times, 19 February 2018. [Online]. Available: https://www.nytimes.com/2018/02/19/technology/russian-bots-school-shooting.html. [Accessed 16 October 2019].

[63] The Observers, "Fake images spark flare-up of violence in West Bengal," France 24, 13 July 2017. [Online]. Available: https://observers.france24.com/en/20170713-fake-images-causes-flare-violence-west-bengal. [Accessed 16 October 2019].

[64] S. Jawed, "The vicious cycle of fake images in Basirhat riots," Alt News, 7 July 2017. [Online]. Available: https://www.altnews.in/vicious-cycle-fake-images-basirhat-riots/. [Accessed 16 October 2019].

[65] B. Zadrozny, "Fire at 'pizzagate' shop reignites conspiracy theorists who find a home on Facebook," NBC News, 1 February 2019. [Online]. Available: https://www.nbcnews.com/tech/social-media/fire-pizzagate-shop-reignites-conspiracy-theorists-who-find-home-facebook-n965956. [Accessed 16 October 2019].

[66] C. Allbright, "A Russian Facebook page organized a protest in Texas. A different Russian page launched the counterprotest.," The Texas Tribune, 1 November 2017. [Online]. Available: https://www.texastribune.org/2017/11/01/russian-facebook-page-organized-protest-texas-different-russian-page-l/. [Accessed 22 October 2019].

[67] Associated Press, "Chinese panic-buy salt over Japan nuclear threat," The Guardian, 17 March 2011. [Online]. Available: https://www.theguardian.com/world/2011/mar/17/chinese-panic-buy-salt-japan. [Accessed 16 October 2019].

[68] J. Kirtley, "Getting to the Truth: Fake News, Libel Laws, and "Enemies of the American People"," *Human Rights Magazine, American Bar Association,* vol. 43, no. 4, pp. 6-9.

[69] S. Baron and R. Crootof, "Fighting Fake News Workshop Report," Information Society Project, Yale Law School, 2017. [Online]. Available: https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf. [Accessed 16 October 2019].

[70] L. K. Einstein and D. M. Glick, "Do I Think BLS Data are BS? The Consequences of Conspiracy Theories," *Political Behavior,* vol. 37, no. 3, pp. 679-701, 2015.

[71] S. Lewandowsky, U. Ecker and J. Cook, "Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era," *Journal of Applied Research in Memory and Cognition,* vol. 6, no. 4, pp. 353-369, 2017.

[72] A. Applebaum, "The Dutch just showed the world how Russia influences Western European elections," The Washington Post, 8 April 2016. [Online]. Available: https://www.washingtonpost.com/opinions/russias-influence-in-western-elections/2016/04/08/b427602a-fcf1-11e5-886f-a037dba38301_story.html. [Accessed 17 October 2019].

[73] Committee on Foreign Relations, "Putin's Asymmetric Assault on Democracy in Russia and Europe: Implications for U.S. National Security," United States Senate, 10 January 2018. [Online]. Available: https://www.foreign.senate.gov/imo/media/doc/FinalRR.pdf. [Accessed 17 October 2019].

[74] S. Ott, "How a selfie with Merkel changed Syrian refugee's life," Al Jazeera, 21 Feburary 2017. [Online]. Available: https://www.aljazeera.com/indepth/features/2017/02/selfie-merkel-changed-syrian-refugee-life-170218115515785.html. [Accessed 17 October 2019].

[75] B. Nimmo, "#ElectionWatch: Scottish Vote, Pro-Kremlin Trolls," Digital Forensic Research Lab, 13 December 2017. [Online]. Available: https://medium.com/dfrlab/electionwatch-scottish-vote-pro-kremlin-trolls-f3cca45045bb. [Accessed 17 October 2019].

[76] A. Blake, "A new study suggests fake news might have won Donald Trump the 2016 election," The Washington Post, 3 April 2018. [Online]. Available: https://www.washingtonpost.com/news/the-fix/wp/2018/04/03/a-new-study-suggests-fake-news-might-have-won-donald-trump-the-2016-election/. [Accessed 17 October 2019].

[77] Minister for Law, "Deliberate Online Falsehoods: Challenges and Implications," Ministry of Communications and Information and the Ministry of Law, 5 January 2018. [Online]. Available: https://www.nas.gov.sg/archivesonline/government_records/Flipviewer/grid_publish/6/6797717d-f25b-11e7-bafc-001a4a5ba61b-06012018Misc.10of2018/web/html5/index.html?launchlogo=tablet/GovernmentRecords_brandingLogo_.png&pn=1. [Accessed 17 November 2019].

[78] P. Howard, B. Kollanyi, S. Bradshaw and L.-M. Neudert, "Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States?," The Computational Propaganda Project, Oxford Internet Institute, Oxford University, 28 September 2017. [Online]. Available: http://blogs.oii.ox.ac.uk/comprop/wp-content/uploads/sites/93/2017/09/Polarizing-Content-and-Swing-States.pdf. [Accessed 17 October 2019].

[79] P. Howard and B. Kollanyi, "Social media companies must respond to the sinister reality behind fake news," The Guardian, 30 September 2017. [Online]. Available: https://www.theguardian.com/media/2017/sep/30/social-media-companies-fake-news-us-election. [Accessed 17 October 2019].

[80] B. Y. Seow, "Singaporean bore the brunt of xenophobic comments when his photo was misused online," The Straits Times, 28 March 2018. [Online]. Available: https://www.straitstimes.com/politics/singaporean-bore-the-brunt-of-xenophobic-comments-when-his-photo-was-misused-online. [Accessed 17 October 2019].

[81] B. Nyhan and J. Reifler, "When Corrections Fail: The Persistence of Politcal Misperceptions," *Political Behavior,* vol. 32, no. 2, pp. 303-330, 2010.

[82] G. Pennycook, T. Cannon and D. Rand, "Prior Exposure Increases Perceived Accuracy of Fake News," *Journal of Experimental Psychology: General,* vol. 147, no. 12, pp. 1865-1880, 2018.

[83] D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs and C. Mattsson, "Combating Fake News: An Agenda for Research and Action," Harvard University and Northeastern University, 2 May 2017. [Online]. Available: https://shorensteincenter.org/combating-fake-news-agenda-for-research/. [Accessed 19 October 2019].

[84] K. Muller and C. Schwarz, "Fanning the Flames of Hate: Social Media and Hate Crime," University of Warwick, 19 February 2018. [Online]. Available: https://warwick.ac.uk/fac/soc/economics/staff/crschwarz/fanning-flames-hate.pdf. [Accessed 19 October 2019].

[85] S. Hegelich and M. Shahrezaye, "Disruptions to political opinion," Konrad Adenauer Stiftung, July 2017. [Online]. Available: https://www.kas.de/documents/252038/253252/7_dokument_dok_pdf_49188_2.pdf/d3741812-3abb-c52a-fe75-658665e274d7?version=1.0&t=1539649011639. [Accessed 19 October 2019].

[86] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, E. Stanley and W. Quattrociocchi, "The spreading of misinformation online," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 113, no. 3, pp. 554-559, 2016.

[87] Y. Halberstam and B. Knight, "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter," *Journal of Public Economics,* vol. 143, pp. 73-88, 2016.

[88] B. Martens, L. Aguiar, E. Gomez-Herrera and F. Mueller-Langer, "The digital transformation of news media and the rise of disinformation and fake news," European Commission, April 2018. [Online]. Available: https://ec.europa.eu/jrc/sites/jrcsh/files/jrc111529.pdf. [Accessed 23 OCtober 2019].

[89] S. Vosoughi, D. Roy and S. Aral, "The spread of true and false news online," *Science,* vol. 359, no. 6380, pp. 1146-1151, 2018.

[90] T. Lee, "The top 20 fake news stories outperformed real news at the end of the 2016 campaign," Vox, 16 November 2016. [Online]. Available: https://www.vox.com/new-money/2016/11/16/13659840/facebook-fake-news-chart. [Accessed 23 October 2019].

[91] P. Howard, G. Bolsover, B. Kollanyi, S. Bradshaw and L.-M. Neudert, "Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter?," The Computational Propaganda Project, University of Oxford, 26 March 2017. [Online]. Available: http://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2017/03/What-Were-Michigan-Voters-Sharing-Over-Twitter-v2.pdf. [Accessed 6 November 2019].

[92] H. Schellmann, "The dangerous new technology that will make us question our basic idea of reality," New York University, 5 December 2017. [Online]. Available: https://qz.com/1145657/the-dangerous-new-technology-that-will-make-us-question-our-basic-idea-of-reality/. [Accessed 20 October 2019].

[93] B. Nimmo, "#ElectionWatch: Fake Photos in Catalonia?," Digital Forensic Research Lab, 23 October 2017. [Online]. Available: https://medium.com/dfrlab/electionwatch-fake-photos-in-catalonia-fe3f045df171. [Accessed 20 October 2019].

[94] C. Paul and M. Matthews, "The Russian "Firehose of Falsehood" Propaganda Model," RAND Corporation, 2016. [Online]. Available: https://www.rand.org/pubs/perspectives/PE198.html. [Accessed 15 November 2019].

[95] A. Guess, B. Nyhan and J. Reifler, "Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign," 9 January 2018. [Online]. Available: https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf. [Accessed 20 October 2019].

[96] BBC, "Zuckerberg: Facebook is in 'arms race' with Russia," BBC, 11 April 2018. [Online]. Available: https://www.bbc.com/news/world-us-canada-43719784. [Accessed 9 October 2019].

[97] BBC, "Facebook bans pages aimed at US election interference," BBC, 31 July 2018. [Online]. Available: https://www.bbc.com/news/technology-45018516. [Accessed 9 October 2019].

[98] Facebook Newsroom, "Removing Bad Actors on Facebook," Facebook, 31 July 2018. [Online]. Available: https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/. [Accessed 9 October 2019].

[99] N. Newman, R. Fletcher, A. Kalogeropoulos, D. Levy and R. Nielsen, "Reuters Institute Digital News Report 2018," Reuters Institute for the Study of Journalism, 2018. [Online]. Available: http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf. [Accessed 25 October 2019].

[100] R. Epstein and R. Robertson, "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 112, no. 33, pp. 4512-4521, 2015.

[101] R. Waters, "Facebook and Google help showcase Las Vegas fake news," Financial Times, 3 October 2017. [Online]. Available: https://www.ft.com/content/030184c2-a7f1-11e7-ab55-27219df83c97. [Accessed 26 October 2019].

[102] S. Chakrabarti, "Hard Questions: What Effect Does Social Media Have on Democracy?," Facebook, 22 January 2018. [Online]. Available: https://newsroom.fb.com/news/2018/01/effect-social-media-democracy/. [Accessed 6 November 2019].

[103] V. Doshi, "India's millions of new Internet users are falling for fake news - sometimes with deadly consequences," The Washington Post, 1 October 2017. [Online]. Available: https://www.washingtonpost.com/world/asia_pacific/indias-millions-of-new-internet-users-are-falling-for-fake-news--sometimes-with-deadly-consequences/2017/10/01/f078eaee-9f7f-11e7-8ed4-a750b67c552b_story.html. [Accessed 22 October 2019].

[104] REACH, "Findings of Poll on Attitudes towards Fake News," REACH, 26 March 2018. [Online]. Available: https://www.reach.gov.sg/~/media/2018/press-release/media-release-on-findings-of-fake-news-poll-26-mar-2018.pdf. [Accessed 29 October 2019].

[105] E. Dwoskin, A. Entous and C. Timberg, "Google uncovers Russian-bought ads on YouTube, Gmail and other platforms," The Washington Post, 9 October 2017. [Online]. Available: https://www.washingtonpost.com/news/the-switch/wp/2017/10/09/google-uncovers-russian-bought-ads-on-youtube-gmail-and-other-platforms/. [Accessed 6 November 2019].

[106] M. Fisher, J. Cox and P. Hermann, "Pizzagate: From rumor, to hashtag, to gunfire in D.C.," The Washington Post, 6 December 2016. [Online]. Available: https://www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-in-dc/2016/12/06/4c7def50-bbd4-11e6-94ac-3d324840106c_story.html. [Accessed 22 October 2019].

[107] Edelman, "2019 Edelman Trust Barometer Global Report," Edelman, 2019. [Online]. Available: https://www.edelman.com/sites/g/files/aatuss191/files/2019-02/2019_Edelman_Trust_Barometer_Global_Report.pdf. [Accessed 31 October 2019].

[108] E. Nekmat and C. Soon, "Silver lining in the battle against fake news," The Straits Times, 2 November 2017. [Online]. Available: https://www.straitstimes.com/opinion/silver-lining-in-the-battle-against-fake-news. [Accessed 26 October 2019].

[109] K. Fridkin, P. Kenney and A. Wintersieck, "Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising," *Political Communication,* vol. 32, no. 1, pp. 127-151, 2015.

[110] E. Culliford, "Facebook fact-checker says company must share more data to fight misinformation," Reuters, 30 July 2019. [Online]. Available: https://www.reuters.com/article/us-usa-facebook-factcheck/facebook-fact-checker-says-company-must-share-more-data-to-fight-misinformation-idUSKCN1UQ00P. [Accessed 24 October 2019].

[111] J. Burns, "Fake news: Universities offer tips on how to spot it," BBC, 9 November 2017. [Online]. Available: https://www.bbc.com/news/education-41902914. [Accessed 22 October 2019].

[112] Kremlin Watch, "About Kremlin Watch," European Values Think Tank, 2019. [Online]. Available: https://www.kremlinwatch.eu/#about-us. [Accessed 26 October 2019].

[113] A. Mosseri, "Working to Stop Misinformation and False News," Facebook for Media, 7 April 2017. [Online]. Available: https://www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news. [Accessed 26 October 2019].

[114] K. Roose, "Here Come the Fake Videos, Too," The New York Times, 4 March 2018. [Online]. Available: https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html. [Accessed 19 October 2019].

[115] E. Ferrara, O. Varol, C. Davis, F. Menczer and A. Flammini, "The Rise of Social Bots," *Communications of the ACM,* vol. 59, no. 7, pp. 96-104, 2016.

[116] Victoria Derbyshire programme, "The fake video where Johnson and Corbyn endorse each other," BBC, 12 November 2019. [Online]. Available: https://www.bbc.com/news/av/technology-50381728/the-fake-video-where-johnson-and-corbyn-endorse-each-other. [Accessed 12 November 2019].

[117] Federal Ministry of Justice & Consumer Protection, "Draft Act improving law enforcement on social networks," Germany's Federal Ministry of Justice and Consumer Protection, 27 March 2017. [Online]. Available: https://perma.cc/BAE2-KAJX. [Accessed 6 November 2019].

[118] D. Manzi, "Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News," *Fordham Law Review,* vol. 87, no. 6, pp. 2623-2651, 2019.

[119] D. Lazer, M. Baum, Y. Benkler, A. Berinsky, K. Greenhill, F. Menczer, M. Metzger, B. Nyham, G. Pennycook, D. Rothschild, M. Schudson, S. Sloman, C. Sunstein, E. Thorson, D. Watts and J. Zittrain, "The Science of Fake News," *Science Magazine,* vol. 359, no. 6380, pp. 1094-1096, 2018.

[120] High Level Expert Group (HLEG), "Final report of the High Level Expert Group on Fake News and Online Disinformation," European Commission, 12 March 2018. [Online].

Available: https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation. [Accessed 27 October 2019].

[121] M. Warner, "Potential Policy Proposals for Regulation of Social Media and Technology Firms," United States Senator, 23 July 2018. [Online]. Available: https://www.warner.senate.gov/public/_cache/files/d/3/d32c2f17-cc76-4e11-8aa9-897eb3c90d16/65A7C5D983F899DAAE5AA21F57BAD944.social-media-regulation-proposals.pdf. [Accessed 27 October 2019].

[122] European Commission, "Tackling online disinformation: a European Approach," European Commission, 26 April 2018. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236. [Accessed 27 October 2019].

[123] P. Roudik, "Initiatives to Counter Fake News: Comparative Summary," Library of Congress, April 2019. [Online]. Available: https://www.loc.gov/law/help/fake-news/compsum.php. [Accessed 27 October 2019].

[124] The Digital, Culture, Media and Sport Committee, "Disinformation and 'fake news': Interim Report," UK House of Commons, 24 July 2018. [Online]. Available: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf. [Accessed 27 October 2019].

[125] W. Ghonim and J. Rashbass, "It's time to end the secrecy and opacity of social media," The Washington Post, 31 October 2017. [Online]. Available: https://www.washingtonpost.com/news/democracy-post/wp/2017/10/31/its-time-to-end-the-secrecy-and-opacity-of-social-media/?noredirect=on. [Accessed 27 October 2019].

[126] J. Snelling, "Top 10 sites to help students check their facts," International Society for Technology in Education (ITSE), 1 February 2019. [Online]. Available: https://www.iste.org/explore/Digital-and-media-literacy/Top-10-sites-to-help-students-check-their-facts. [Accessed 27 October 2019].

[127] Facebook Business, "Fact-checking on Facebook: What publishers should know," Facebook, 2019. [Online]. Available: https://www.facebook.com/help/publisher/182222309230722. [Accessed 27 October 2019].

[128] N. Gleicher, "Removing More Coordinated Inauthentic Behavior From Iran and Russia," Facebook Newsroom, 21 October 2019. [Online]. Available: https://newsroom.fb.com/news/2019/10/removing-more-coordinated-inauthentic-behavior-from-iran-and-russia/. [Accessed 27 October 2019].

[129] N. Newman, R. Fletcher, A. Kalogeropoulos and R. Nielsen, "Reuters Institute Digital News Report 2019," Reuters Institute for the Study of Journalism, 2019. [Online]. Available: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_0.pdf. [Accessed 27 October 2019].

[130] M. Isaac, "Mark Zuckerberg's Call to Regulate Facebook, Explained," The New York Times, 30 March 2019. [Online]. Available: https://www.nytimes.com/2019/03/30/technology/mark-zuckerberg-facebook-regulation-explained.html. [Accessed 27 October 2019].

[131] M. Zuckerberg, "Mark Zuckerberg: The Internet needs new rules. Let's start in these four areas.," The Washington Post, 30 March 2019. [Online]. Available: https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-

rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html. [Accessed 27 October 2019].

[132] M. Posner, "Dealing With Disinformation: Facebook And YouTube Need To Take Down Provably False "News"," Forbes, 14 March 2019. [Online]. Available: https://www.forbes.com/sites/michaelposner/2019/03/14/dealing-with-disinformation-facebook-and-youtube-need-to-take-down-provably-false-news/#493e080b19e7. [Accessed 15 November 2019].

[133] C. Silverman, "Lies, Damn Lies and Viral Content," Tow Center for Digital Journalism, Columbia University, 5 September 2017. [Online]. Available: https://academiccommons.columbia.edu/doi/10.7916/D8Q81RHH. [Accessed 1 November 2019].

[134] G. Haciyakupoglu, J. Yang, V. S. Suguna, D. Leong and F. Muhammad, "Countering Fake News," S. Rajaratnam School of International Studies, Nanyang Technological University, March 2018. [Online]. Available: https://www.rsis.edu.sg/wp-content/uploads/2018/03/PR180416_Countering-Fake-News.pdf. [Accessed 11 November 2019].

[135] P. Dizikes, "Want to squelch fake news? Let the readers take charge," MIT News, 28 January 2019. [Online]. Available: http://news.mit.edu/2019/reader-crowdsource-fake-news-0128. [Accessed 15 November 2019].

[136] S. Bradshaw, L.-M. Neudert and P. Howard, "Government Responses to Malicious Use of Social Media," Oxford Internet Institute, University of Oxford, January 2019. [Online]. Available: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Nato-Report.pdf. [Accessed 12 November 2019].

[137] A. Mosseri, "Addressing Hoaxes and Fake News," Facebook, 15 December 2016. [Online]. Available: https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/. [Accessed 15 November 2019].

[138] B. Nyhan, "Why the Fact-Checking at Facebook Needs to Be Checked," The New York Times, 23 October 2017. [Online]. Available: https://www.nytimes.com/2017/10/23/upshot/why-the-fact-checking-at-facebook-needs-to-be-checked.html. [Accessed 15 November 2019].

[139] G. Pennycook, A. Bear, E. Collins and D. Rand, "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings," *Management Science,* Forthcoming.

[140] D. Funke and D. Flamini, "A guide to anti-misinformation actions around the world," Poynter, 2019. [Online]. Available: https://www.poynter.org/ifcn/anti-misinformation-actions/. [Accessed 15 November 2019].

[141] A. Luhn, "Ukraine blocks popular social networks as part of sanctions on Russia," The Guardian, 16 May 2017. [Online]. Available: https://www.theguardian.com/world/2017/may/16/ukraine-blocks-popular-russian-websites-kremlin-role-war. [Accessed 18 November 2019].

[142] E. Ullrich, S. Lewandowsky, B. Swire and D. Chang, "Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction," *Psychonomic Bulletin & Review,* vol. 18, no. 3, pp. 570-578, 2011.

[143] J. Swaine, "Twitter admits far more Russian bots posted on election than it had disclosed," The Guardian, 19 January 2018. [Online]. Available: https://www.theguardian.com/technology/2018/jan/19/twitter-admits-far-more-russian-bots-posted-on-election-than-it-had-disclosed. [Accessed 6 November 2019].

[144] S. Larson, "Facebook modifies the way it alerts users to fake news," CNN, 21 December 2017. [Online]. Available: https://money.cnn.com/2017/12/21/technology/facebook-fake-news-related-articles/index.html. [Accessed 20 November 2019].

[145] J. Weedon, W. Nuland and A. Stamos, "Information Operations and Facebook," Facebook, 27 April 2017. [Online]. Available: https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf. [Accessed 4 November 2019].

[146] D. West, "How to combat fake news and disinformation," Brookings, 18 December 2017. [Online]. Available: https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/. [Accessed 16 November 2019].

[147] T. Lyons, "Increasing Our Efforts to Fight False News," Facebook, 21 June 2018. [Online]. Available: https://newsroom.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news/. [Accessed 5 November 2019].

[148] M. Viviani and G. Pasi, "Credibility in Social Media: Opinions, News, and Health Information - A Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 7, 2017.

[149] K. Cheng, "DSO engineers building system that detects fake news," Today, 21 April 2017. [Online]. Available: https://www.todayonline.com/singapore/dso-engineers-building-system-detects-fake-news. [Accessed 15 November 2019].

[150] S. Levin, "Facebook promised to tackle fake news. But the evidence shows it's not working," The Guardian, 16 May 2017. [Online]. Available: https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working. [Accessed 15 November 2019].

[151] C. Silverman, "Facebook Is About To Bring The Hammer Down On Overseas Fake News Operators," BuzzFeed News, 21 June 2018. [Online]. Available: https://www.buzzfeednews.com/article/craigsilverman/facebook-is-now-trying-to-predict-whether-a-page-is-likely. [Accessed 5 November 2019].

[152] Facebook Security, "Disrupting a major spam operation," Facebook, 14 April 2017. [Online]. Available: https://www.facebook.com/notes/facebook-security/disrupting-a-major-spam-operation/10154327278540766/. [Accessed 4 November 2019].

[153] T. Wheeler, "How to Monitor Fake News," The New York Times, 20 February 2018. [Online]. Available: https://www.nytimes.com/2018/02/20/opinion/monitor-fake-news.html. [Accessed 14 November 2019].

[154] S. Woolley and P. Howard, "Computational Propaganda Worldwide: Executive Summary," Computational Propaganda Research Project, University of Oxford, 19 June 2017. [Online]. Available: http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf. [Accessed 11 November 2019].

[155] A. Hern, "Google acts against fake news on search engine," The Guardian, 25 April 2017. [Online]. Available: https://www.theguardian.com/technology/2017/apr/25/google-launches-major-offensive-against-fake-news. [Accessed 13 November 2019].

[156] O. Solon, "Facebook says likely Russia-based group paid for political ads during US election," The Guardian, 6 September 2017. [Online]. Available: https://www.theguardian.com/technology/2017/sep/06/facebook-political-ads-russia-us-election-trump-clinton. [Accessed 6 November 2019].

[157] J. Love, J. Menn and D. Ingram, "In Mexico, fake news creators up their game ahead of election," Reuters, 28 June 2018. [Online]. Available: https://www.reuters.com/article/us-mexico-facebook/in-mexico-fake-news-creators-up-their-game-ahead-of-election-idUSKBN1JO2VG. [Accessed 5 November 2019].

[158] AFP, "Twitter to ban political ads worldwide on its platform," Channel News Asia, 31 October 2019. [Online]. Available: https://www.channelnewsasia.com/news/business/twitter-ban-political-ads-worldwide-platform-jack-dorsey-12049264. [Accessed 5 November 2019].

[159] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach," The Guardian, 17 March 2018. [Online]. Available: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. [Accessed 14 November 2019].

[160] Channel News Asia, "Twitter admits user email addresses, phone numbers may have been used for ads," Channel News Asia, 9 October 2019. [Online]. Available: https://www.channelnewsasia.com/news/business/twitter-admits-user-email-addresses-phone-numbers-data-11983256. [Accessed 9 October 2019].

[161] BBC, "Licence Fee and Funding," BBC, 2019. [Online]. Available: https://www.bbc.com/aboutthebbc/governance/licencefee. [Accessed 31 October 2019].

[162] L. Teodoro, "Combating 'Fake News'," BusinessWorld, 6 June 2019. [Online]. Available: https://www.bworldonline.com/combating-fake-news/. [Accessed 15 November 2019].

[163] A. Bhattacharya, "Here's a handy cheat sheet of false and misleading "news" sites," Quartz, 17 November 2016. [Online]. Available: https://qz.com/839160/heres-a-handy-cheat-sheet-of-false-and-misleading-news-sites/. [Accessed 15 November 2019].

[164] B. Donald, "Stanford researchers find students have trouble judging the credibility of information online," Stanford Graduate School of Education, 22 November 2016. [Online]. Available: https://ed.stanford.edu/news/stanford-researchers-find-students-have-trouble-judging-credibility-information-online. [Accessed 11 November 2019].

[165] K. Ng, "Laws tackling fake news to be introduced next year: Shanmugam," Today, 18 June 2017. [Online]. Available: https://www.todayonline.com/singapore/new-laws-tackle-fake-news-be-introduced-next-year-shanmugam. [Accessed 14 November 2019].

[166] Center for Business and Human Rights, "Harmful Content: The Role of Internet Platform Companies In Fighting Terrorist Incitement and Politically Motivated Disinformation," NYU Stern, 3 November 2017. [Online]. Available: https://static1.squarespace.com/static/547df270e4b0ba184dfc490e/t/59fb31bc0d929735d3d01d95/1509634510957/Final.Harmful+Content.+The+Role+of+Internet+Platform+Companies+in+Fighting+Terrorist+Incitement+and+Politically+Motivated+Propaganda.pdf. [Accessed 15 November 2019].

[167] J. Compton, B. Jackson and J. Dimmock, "Persuading Others to Avoid Persuasion: Inoculation Theory and Resistant Health Attitudes," *Frontiers in Psychology,* vol. 7, p. 122, 2016.

[168] S. van der Linden, A. Leiserowitz, S. Rosenthal and E. Maibach, "Inoculating the Public against Misinformation about Climate Change," *Global Challenges,* vol. 1, no. 2, 2017.

[169] University of Cambridge, "Fake news 'vaccine' works: 'Pre-bunking' game reduces susceptibility to disinformation," ScienceDaily, 24 June 2019. [Online]. Available: https://www.sciencedaily.com/releases/2019/06/190624204800.htm. [Accessed 11 November 2019].

[170] J. Davies, "UK pubs enlist bots to fight against filter bubbles ahead of the UK election," Digiday, 30 May 2017. [Online]. Available: https://digiday.com/media/uk-pubs-enlist-bots-fight-filter-bubbles-ahead-uk-election/. [Accessed 14 November 2019].

[171] A. Hess, "How to Escape Your Political Bubble for a Clearer View," The New York Times, 3 March 2017. [Online]. Available: https://www.nytimes.com/2017/03/03/arts/the-battle-over-your-political-bubble.html. [Accessed 14 November 2019].

[172] M. Richter, "The Kremlin's Platform for 'Useful Idiots' in the West: An Overview of RT's Editorial Strategy and Evidence of Impact," European Values Think Tank, 18 September 2017. [Online]. Available: https://www.europeanvalues.net/wp-content/uploads/2017/09/Overview-of-RTs-Editorial-Strategy-and-Evidence-of-Impact-1.pdf. [Accessed 5 November 2019].

[173] N. Feldman, "Fake News May Not Be Protected Speech," Bloomberg, 23 November 2016. [Online]. Available: https://www.bloomberg.com/opinion/articles/2016-11-23/fake-news-may-not-be-protected-speech. [Accessed 5 November 2019].

[174] Home Affairs Committee, "Hate crime: abuse, hate and extremism online," UK House of Commons, 25 April 2017. [Online]. Available: https://publications.parliament.uk/pa/cm201617/cmselect/cmhaff/609/60904.htm#_idTextAnchor014. [Accessed 5 November 2019].

[175] M. Safi, "Sri Lanka accuses Facebook over hate speech after deadly riots," The Guardian, 14 March 2018. [Online]. Available: https://www.theguardian.com/world/2018/mar/14/facebook-accused-by-sri-lanka-of-failing-to-control-hate-speech. [Accessed 5 November 2019].

[176] V. Goel, H. Kumar and S. Frenkel, "In Sri Lanka, Facebook Contends With Shutdown After Mob Violence," The New York Times, 8 March 2018. [Online]. Available: https://www.nytimes.com/2018/03/08/technology/sri-lanka-facebook-shutdown.html. [Accessed 5 November 2019].

[177] E. Culliford, "Facebook takes down false ad from PAC on Republican Graham," Reuters, 26 October 2019. [Online]. Available: https://www.reuters.com/article/us-usa-election-facebook/facebook-takes-down-false-ad-from-pac-on-republican-graham-idUSKBN1X50IZ. [Accessed 26 October 2019].

[178] R. Bort, "What Alexandria Ocasio-Cortez Exposed About Mark Zuckerberg and Facebook Heading Into 2020," RollingStone, 24 October 2019. [Online]. Available: https://www.rollingstone.com/politics/politics-news/alexandria-ocasio-cortez-mark-zuckerberg-facebook-hearing-903057/. [Accessed 5 November 2019].

[179] Google Transparency Project, "How to Sow Discord Using Google and $100 (or 6,800 rubles)," Campaign for Accountability, 4 September 2018. [Online]. Available: https://www.googletransparencyproject.org/articles/how-sow-discord-using-google-and-100-or-6800-rubles. [Accessed 6 November 2019].

[180] L. Southern, "What to know about Germany's fake-news crackdown," Digiday, 21 March 2017. [Online]. Available: https://digiday.com/media/germanys-proposed-hate-crime-law/. [Accessed 14 November 2019].

[181] C. Radsch, "Proposed German legislation threatens broad internet censorship," Committee to Protect Journalists, 20 April 2017. [Online]. Available: https://cpj.org/blog/2017/04/proposed-german-legislation-threatens-broad-intern.php. [Accessed 23 October 2019].

[182] Human Rights Watch, "Philippines: Reject Sweeping 'Fake News' Bill," Human Rights Watch, 25 July 2019. [Online]. Available: https://www.hrw.org/news/2019/07/25/philippines-reject-sweeping-fake-news-bill#. [Accessed 15 November 2019].

[183] J. Gesley, "Initiatives to Counter Fake News: Germany," Library of Congress, April 2019. [Online]. Available: https://www.loc.gov/law/help/fake-news/germany.php. [Accessed 4 November 2019].

[184] J. Malsin and A. El Fekki, "Egypt Passes Law to Regulate Media as President Sisi Consolidates Power," The Wall Street Journal, 16 July 2018. [Online]. Available: https://www.wsj.com/articles/egypt-passes-law-to-regulate-media-as-president-sisi-consolidates-power-1531769232. [Accessed 23 November 2019].

[185] P. Lee, "Factually website clarifies 'widespread' falsehoods," The Straits Times, 2 March 2017. [Online]. Available: https://www.straitstimes.com/singapore/factually-website-clarifies-widespread-falsehoods. [Accessed 15 November 2019].

[186] J. Kaos, "Government launches 'Tidak Pasti, Jangan Kongsi' to stop spread of false information," The Star Online, 15 March 2017. [Online]. Available: https://www.thestar.com.my/news/nation/2017/03/15/website-to-debunk-fake-news-government-launches-tidak-pasti-jangan-kongsi-to-stop-spread-of-false-in/. [Accessed 15 November 2019].

[187] J. Jackson, "Fact-checkers are weapons in the post-truth wars, but they're not all on one side," The Guardian, 15 February 2017. [Online]. Available: https://www.theguardian.com/media/2017/feb/15/fact-checkers-are-weapons-in-the-post-truth-wars-but-theyre-not-all-on-one-side. [Accessed 15 November 2019].

[188] Z. Liao, "An economic analysis on internet regulation in China and proposals to policy and law makers," *International Journal of Technology Policy and Law,* vol. 2, no. 2/3/4, pp. 242-256, 2016.

[189] BBC, "Government announces anti-fake news unit," BBC, 23 January 2018. [Online]. Available: https://www.bbc.com/news/uk-politics-42791218. [Accessed 22 October 2019].

[190] Agence France-Presse, "Indonesia to set up agency to combat fake news," The Straits Times, 6 January 2017. [Online]. Available: https://www.straitstimes.com/asia/se-asia/indonesia-to-set-up-agency-to-combat-fake-news-0. [Accessed 22 October 2019].

[191] T. Sipahutar and K. Salna, "Inside the Government-Run War Room Fighting Indonesian Fake News," Bloomberg, 24 October 2018. [Online]. Available: https://www.bloomberg.com/news/articles/2018-10-24/inside-the-government-run-war-room-fighting-indonesian-fake-news. [Accessed 23 November 2019].

[192] The Australian Financial Review, "New taskforce to defend against election meddling," The Australian Financial Review, 9 June 2018. [Online]. Available:

https://www.afr.com/politics/federal/new-taskforce-to-defend-against-election-meddling-20180609-h116c6. [Accessed 21 November 2019].

[193] AFP, "Countries at UN commit to fighting fake news," Channel News Asia, 27 September 2019. [Online]. Available: https://www.channelnewsasia.com/news/world/countries-at-un-commit-to-fighting-fake-news-11947714. [Accessed 15 November 2019].

[194] J. Summers, "Countering Disinformation: Russia's Infowar in Ukraine," The Henry M. Jackson School of International Studies, University of Washington, 25 October 2017. [Online]. Available: https://jsis.washington.edu/news/russia-disinformation-ukraine/#_ftnref7. [Accessed 25 November 2019].

[195] R. Tan, "Applying systems thinking towards countering hybrid warfare," Thesis: S.M. in Engineering and Management, Massachusetts Institute of Technology, System Design and Management Program, 2019, 4 January 2019. [Online]. Available: https://dspace.mit.edu/handle/1721.1/121799. [Accessed 25 November 2019].

[196] T. Kuzio and P. D'Anieri, "Annexation and Hybrid Warfare in Crimea and Eastern Ukraine," E-International Relations, 25 June 2018. [Online]. Available: https://www.e-ir.info/2018/06/25/annexation-and-hybrid-warfare-in-crimea-and-eastern-ukraine/. [Accessed 25 November 2019].

[197] E. Nakashima, "Inside a Russian disinformation campaign in Ukraine in 2014," The Washington Post, 25 December 2017. [Online]. Available: https://www.washingtonpost.com/world/national-security/inside-a-russian-disinformation-campaign-in-ukraine-in-2014/2017/12/25/f55b0408-e71d-11e7-ab50-621fe0588340_story.html. [Accessed 25 November 2019].

[198] M. Dyczok, "Ukraine's Media during Revolution, Annexation, War and Economic Crisis," E-International Relations, 20 April 2016. [Online]. Available: https://www.e-ir.info/2016/04/20/ukraines-media-during-revolution-annexation-war-and-economic-crisis/. [Accessed 2 December 2019].

[199] P. R. Gregory, "Putin's 'Human Rights Council' Accidentally Posts Real Crimean Election Results," Forbes, 5 May 2014. [Online]. Available: https://www.forbes.com/sites/paulroderickgregory/2014/05/05/putins-human-rights-council-accidentally-posts-real-crimean-election-results-only-15-voted-for-annexation/#1e49f445f172. [Accessed 25 November 2019].

[200] StopFake.org, "About Us," StopFake.org, 2019. [Online]. Available: https://www.stopfake.org/en/about-us/. [Accessed 28 November 2019].

[201] T. Susman-Peña and K. Vogt, "Ukrainians' self-defense against disinformation: What we learned from Learn to Discern," International Research & Exchanges Board (IREX), 12 June 2017. [Online]. Available: https://www.irex.org/insight/ukrainians-self-defense-against-disinformation-what-we-learned-learn-discern. [Accessed 28 November 2019].

[202] N. Jankowicz, "Ukraine's Election Is an All-Out Disinformation Battle," The Atlantic, 17 April 2019. [Online]. Available: https://www.theatlantic.com/international/archive/2019/04/russia-disinformation-ukraine-election/587179/. [Accessed 2 December 2019].

[203] D. Funke, "In the past year, Facebook has quadrupled its fact-checking partners," Poynter, 29 April 2019. [Online]. Available: https://www.poynter.org/fact-checking/2019/in-the-

past-year-facebook-has-quadrupled-its-fact-checking-partners/. [Accessed 1 December 2019].

[204] D. Priest, J. Jacoby and A. Bourg, "Russian disinformation on Facebook targeted Ukraine well before the 2016 U.S. election," The Washington Post, 28 October 2018. [Online]. Available: https://www.washingtonpost.com/business/economy/russian-disinformation-on-facebook-targeted-ukraine-well-before-the-2016-us-election/2018/10/28/cc38079a-d8aa-11e8-a10f-b51546b10756_story.html. [Accessed 1 December 2019].

[205] G. Cain, "Ukraine's War on Russian Disinformation Is a Lesson for America," The New Republic, 29 March 2019. [Online]. Available: https://newrepublic.com/article/153415/ukraines-war-russian-disinformation-lesson-america. [Accessed 1 December 2019].

[206] N. Gleicher, "Removing Coordinated Inauthentic Behavior from Iran, Russia, Macedonia and Kosovo," Facebook, 26 March 2019. [Online]. Available: https://about.fb.com/news/2019/03/cib-iran-russia-macedonia-kosovo/. [Accessed 2 December 2019].

[207] B. Talant, "Facebook rolls out new political ads policy for Ukraine two weeks before the vote," Kyiv Post, 15 March 2019. [Online]. Available: https://www.kyivpost.com/ukraine-politics/facebook-rolls-out-new-political-ads-policy-for-ukraine-two-weeks-before-the-vote.html?cn-reloaded=1. [Accessed 2 December 2019].

[208] Twitter, "Political Content," Twitter, 2019. [Online]. Available: https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content.html. [Accessed 2 December 2019].

[209] Gallup, "Contemporary Media Use in Ukraine," Broadcasting Board of Governors, June 2014. [Online]. Available: https://www.bbg.gov/wp-content/media/2014/06/Ukraine-research-brief.pdf. [Accessed 2 December 2019].

[210] Ukraine Crisis Media Center, "About Press Center," Ukraine Crisis Media Center, 2019. [Online]. Available: http://uacrisis.org/about. [Accessed 28 November 2019].

[211] S. Ingber, "Students In Ukraine Learn How To Spot Fake Stories, Propaganda And Hate Speech," National Public Radio (NPR), 22 March 2019. [Online]. Available: https://www.npr.org/2019/03/22/705809811/students-in-ukraine-learn-how-to-spot-fake-stories-propaganda-and-hate-speech. [Accessed 2 December 2019].

[212] M. Haigh, T. Haigh and T. Matychak, "Information Literacy vs. Fake News: The Case of Ukraine," *Open Information Science,* vol. 3, pp. 154-165, 2019.

[213] E. Murrock, J. Amulya, M. Druckman and T. Liubyva, "Winning the war on state-sponsored propaganda," International Research and Exchanges Board (IREX), 2018. [Online]. Available: https://www.irex.org/sites/default/files/node/resource/impact-study-media-literacy-ukraine.pdf. [Accessed 2 December 2019].

[214] I. Traynor and S. Walker, "Western nations scramble to contain fallout from Ukraine crisis," The Guardian, 23 February 2014. [Online]. Available: https://www.theguardian.com/world/2014/feb/23/ukraine-crisis-western-nations-eu-russia. [Accessed 11 November 2019].

[215] S. Ayres, "Is it too late for Kiev to woo Russian-speaking Ukraine?," The Christian Science Monitor, 28 February 2014. [Online]. Available:

https://www.csmonitor.com/World/Europe/2014/0228/Is-it-too-late-for-Kiev-to-woo-Russian-speaking-Ukraine. [Accessed 28 November 2019].

[216] M. Jaroszewicz, "Years After Crimea's Annexation, Integration of Ukraine's Internally Displaced Population Remains Uneven," Migration Policy Institute, 19 September 2019. [Online]. Available: https://www.migrationpolicy.org/article/fyears-after-crimea-annexation-integration-ukraine-internally-displaced-population. [Accessed 28 November 2019].

[217] P. Polityuk, "Ukraine passes language law, irritating president-elect and Russia," Reuters, 25 April 2019. [Online]. Available: https://www.reuters.com/article/us-ukraine-parliament-language/ukraine-passes-language-law-irritating-president-elect-and-russia-idUSKCN1S111N. [Accessed 2 December 2019].

[218] A. Roth, "Ukraine Adopts Language Law Opposed by Kremlin," The Guardian, 25 April 2019. [Online]. Available: https://www.theguardian.com/world/2019/apr/25/ukraine-adopts-law-enforcing-use-of-ukrainian-in-public-life. [Accessed 2 December 2019].

[219] R. Huba, "Why Ukraine's new language law will have long-term consequences," openDemocracy, 28 May 2019. [Online]. Available: https://www.opendemocracy.net/en/odr/ukraine-language-law-en/. [Accessed 2 December 2019].

[220] H. Coynash, "Draft law in Ukraine to criminalize 'fake news' condemned as censorship & attempt to silence media," Kharkiv Human Rights Protection Group, 18 March 2019. [Online]. Available: http://khpg.org/en/index.php?id=1552747635. [Accessed 2 December 2019].

[221] Freedom on the Net 2018, "Ukraine," Freedom House, 2018. [Online]. Available: https://freedomhouse.org/report/freedom-net/2018/ukraine. [Accessed 2 December 2019].

[222] P. Roudik, "Ukraine: New Law on TV Ownership," Global Legal Monitor, Library of Congress, 7 October 2015. [Online]. Available: https://www.loc.gov/law/foreign-news/article/ukraine-new-law-on-tv-ownership/. [Accessed 3 December 2019].

[223] J. Adams, "OSCE Representative welcomes law on transparency of media ownership in Ukraine as it comes into force," Organization for Security and Co-operation in Europe, 1 October 2015. [Online]. Available: https://www.osce.org/fom/187956. [Accessed 3 December 2019].

[224] A. Lykholat and N. Buyon, "In Ukraine, Free Elections Must Go Hand-in-Hand with Internet Freedom," Freedom House, 27 March 2019. [Online]. Available: https://freedomhouse.org/blog/ukraine-free-elections-must-go-hand-hand-internet-freedom. [Accessed 2 December 2019].

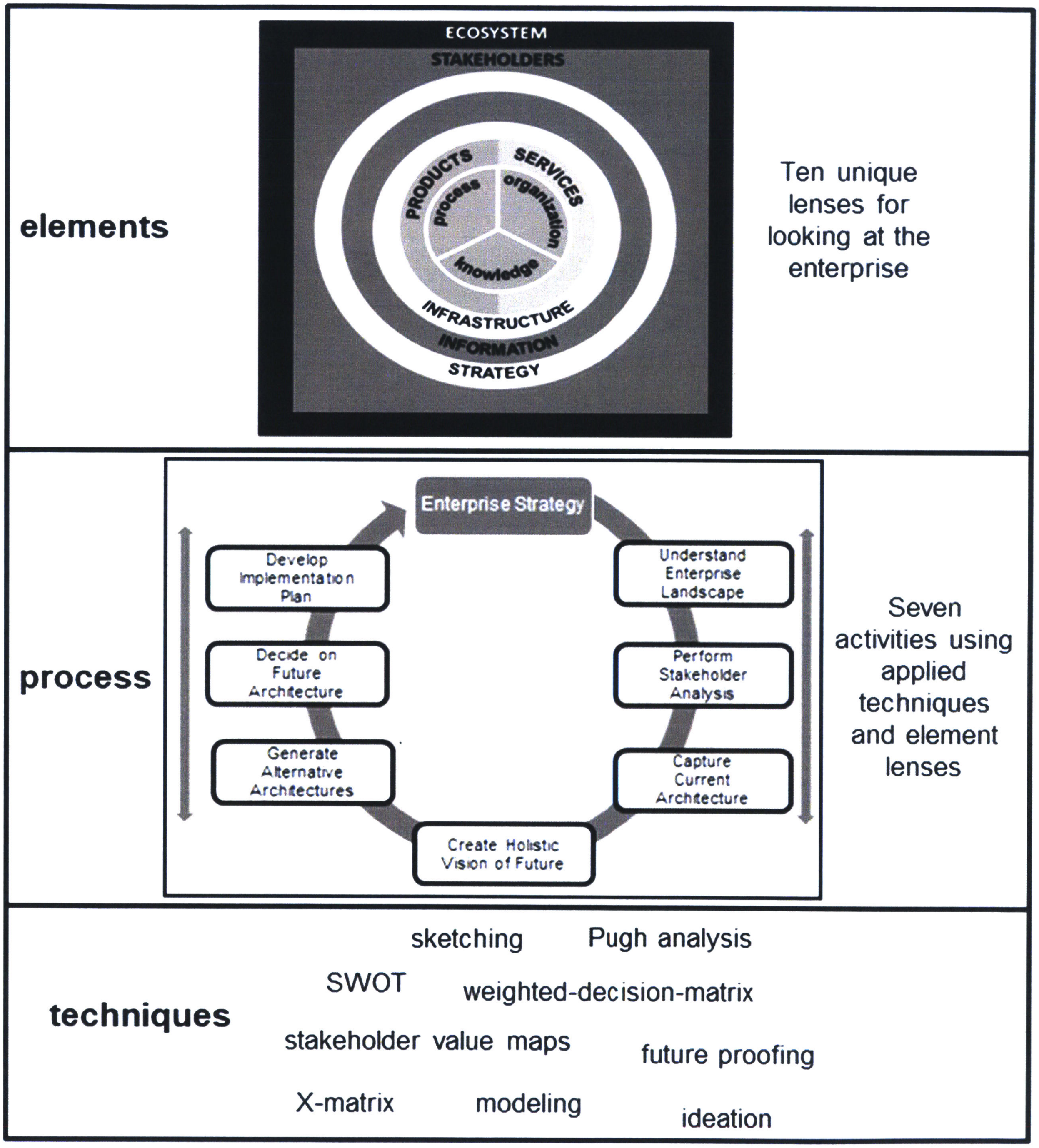
[225] J. Sterman, Systems Thinking and Modeling for a Complex World, Chennai: McGraw Hill Education, 2010.

This page is intentionally left blank

Appendix A

The ARIES Framework and its Application

This appendix provides a brief description of the ARIES framework and discusses its application to this thesis. The framework is as shown:


elements
ECOSYSTEM
STAKEHOLDERS
PRODUCTS
SERVICES
process
organization
knowledge
INFRASTRUCTURE
INFORMATION
STRATEGY
Ten unique lenses for looking at the enterprise
process
Enterprise Strategy
Understand Enterprise Landscape
Perform Stakeholder Analysis
Capture Current Architecture
Create Holistic Vision of Future
Generate Alternative Architectures
Decide on Future Architecture
Develop Implementation Plan
Seven activities using applied techniques and element lenses
techniques
sketching
Pugh analysis
SWOT
weighted-decision-matrix
stakeholder value maps
future proofing
X-matrix
modeling
ideation

Overview of Framework. The ARIES framework was jointly developed in the Massachusetts Institute of Technology by Dr. Deborah J. Nightingale and Dr. Donna H. Rhodes. The framework draws on the fundamental theory and practice of multiple fields, including strategic management, stakeholder theory, systems architecting, innovation, scenario analysis, decision science, enterprise theory, and systems science. It aims to provide a holistic approach to the selection of a new architecture for a future enterprise. As the proposed system involves a collective group of organizations, each tasked with distinct roles and responsibilities, the author finds value in using this framework to guide the entire thought process and analyze how this national-level enterprise should function. The framework consists of (1) the *enterprise element model*, using ten unique elements for viewing the enterprise; (2) the *architecting process model*, with seven activities; and (3) selected *techniques and templates*.

Enterprise Element Model. The first two elements are the *ecosystem* and *stakeholders*. An analysis of the ecosystem and the external factors, which would influence the enterprise's ability to combat the spread of disinformation, is conducted in Section 3.2; while stakeholder analyses to better understand the individuals or groups that contribute to, benefit from, and/or are affected by the enterprise, is conducted in Section 3.4. The remaining eight elements are termed *view elements*, which comprises strategy, information, infrastructure, products, services, process, organization, and knowledge. These perspectives aim to provide a better understanding of the enterprise and is used in Section 3.3 to analyze the current architecture and state of affairs.

Architecting Process Model. The process consists of seven activities to develop the architecture.

- The first activity is to *understand the enterprise landscape*, which includes both the external and internal landscapes. These are analyzed in Sections 3.2 and 3.3 respectively. The external landscape includes context factors that characterize the external environment, such as market, economic, and regulatory factors. The internal landscape includes early insights regarding enablers and barriers to transformation, as well as the capabilities that currently exists within the enterprise.

- The second activity is to *perform stakeholder analysis*, which is conducted in Section 3.4, to understand the individuals and groups that influence or are influenced by the enterprise. Gaps between current value delivery and anticipated future needs are also identified.

- The third activity is to *capture the current architecture*, and to examine how the enterprise is currently structured and how it operates. To do so, the various organizations that form the national-level enterprise are discussed in Section 3.1, while the enterprise element model is used in Section 3.3 to capture the current architecture and state of affairs.

- The fourth activity is to *create a holistic vision of the future*. Vignettes are used in Section 5.1 to describe the envisioned future – a point in time when the system has become a reality – through the eyes of key stakeholders.

- The fifth and sixth activities are to *generate alternative architectures* and to *decide on the future architecture*, respectively. These are performed in Section 5.2, where key architectural decisions are identified with various options for each. A preferred option for each decision is then proposed after discussing the pros and cons of each option.

- The seventh and final activity is to develop the implementation plan, including resources, timelines, and roles and responsibilities. This activity has not been performed in this thesis as the proposed architecture is not for a specific nation. Hence, discussions have been kept to a more strategic level. This activity would be useful when customizing a solution for a specific nation.

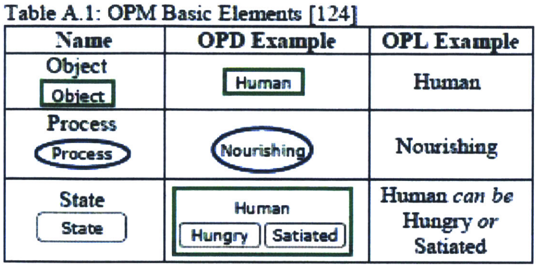
Techniques and Templates. Besides the use of techniques from the ARIES framework (e.g. for the stakeholder analyses in Section 3.4), other systems thinking techniques such as system decomposition and mapping of form to function are utilized in Chapter 4 and Section 5.3 to develop the proposed architecture.

This page is intentionally left blank

Appendix B

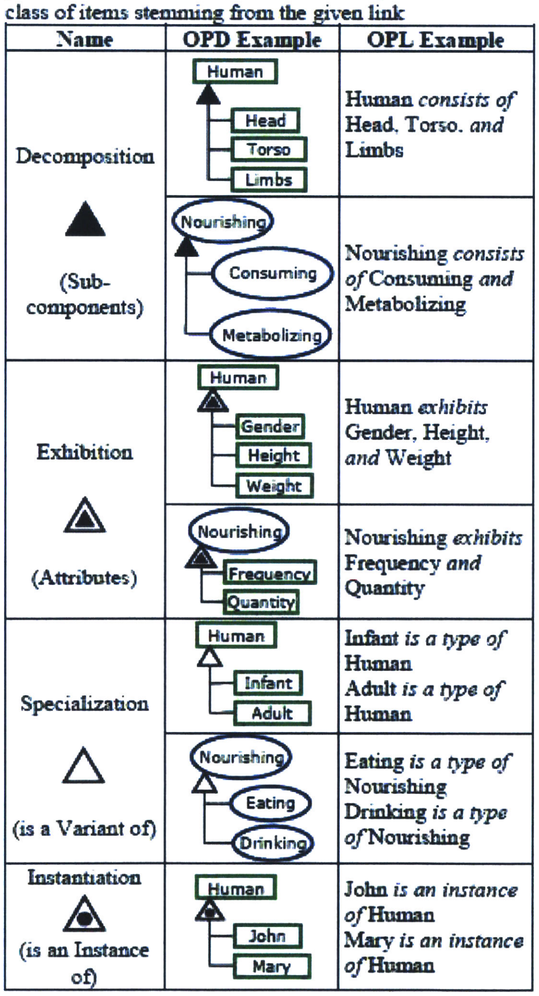
Legend for Object-Process Diagram (OPD)

Table A.1: OPM Basic Elements [124]



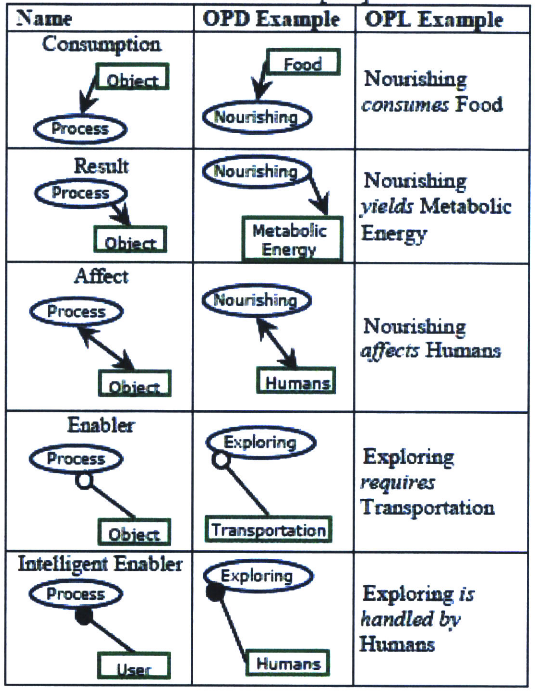
Name	OPD Example	OPL Example
Object Object	Human	Human
Process Process	Nourishing	Nourishing
State State	Human Hungry Satiated	Human *can be* Hungry *or* Satiated

Table A.2: OPM Structural Links [124]. Terms in parentheses can be considered as descriptions of the class of items stemming from the given link



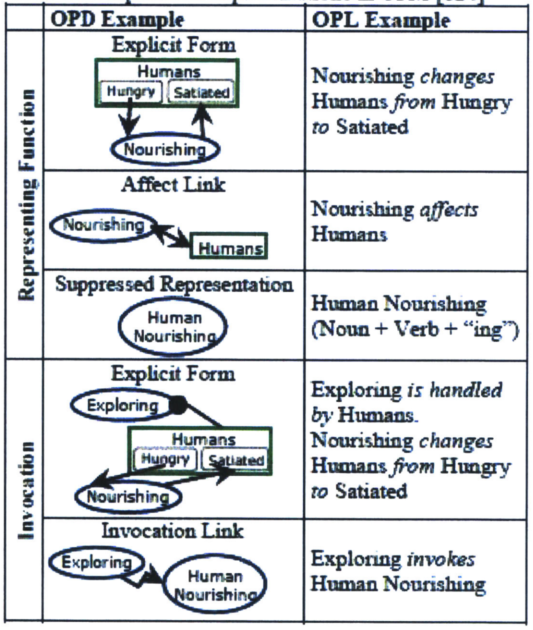
Name	OPD Example	OPL Example
Decomposition ▲ (Sub-components)	Human Head Torso Limbs	Human *consists of* Head, Torso, *and* Limbs
	Nourishing Consuming Metabolizing	Nourishing *consists of* Consuming *and* Metabolizing
Exhibition ◮ (Attributes)	Human Gender Height Weight	Human *exhibits* Gender, Height, *and* Weight
	Nourishing Frequency Quantity	Nourishing *exhibits* Frequency *and* Quantity
Specialization △ (is a Variant of)	Human Infant Adult	Infant *is a type of* Human Adult *is a type of* Human
	Nourishing Eating Drinking	Eating *is a type of* Nourishing Drinking *is a type of* Nourishing
Instantiation ◬ (is an Instance of)	Human John Mary	John *is an instance of* Human Mary *is an instance of* Human

Table A.3: OPM Procedural Links [124]



Name	OPD Example	OPL Example
Consumption Object Process	Food Nourishing	Nourishing *consumes* Food
Result Process Object	Nourishing Metabolic Energy	Nourishing *yields* Metabolic Energy
Affect Process Object	Nourishing Humans	Nourishing *affects* Humans
Enabler Process Object	Exploring Transportation	Exploring *requires* Transportation
Intelligent Enabler Process User	Exploring Humans	Exploring *is handled by* Humans

Table A.4: Equivalent Representations in OPM [124]



	OPD Example	OPL Example
Representing Function	Explicit Form Humans Hungry Satiated Nourishing	Nourishing *changes* Humans *from* Hungry *to* Satiated
	Affect Link Nourishing Humans	Nourishing *affects* Humans
	Suppressed Representation Human Nourishing	Human Nourishing (Noun + Verb + "ing")
Invocation	Explicit Form Exploring Humans Hungry Satiated Nourishing	Exploring *is handled by* Humans. Nourishing *changes* Humans *from* Hungry *to* Satiated
	Invocation Link Exploring Human Nourishing	Exploring *invokes* Human Nourishing

Credits: Sydney Do, MIT Aeroastro Ph.D. '16

This page is intentionally left blank