

Digital Disinformation Landscape and Online Disinformation-Countering Tools

Draft Research Report – May 2022

Research Lead:

Dr. Umar Ruhi

Core Research Team Members:

Kimberley Nault

Ola El-Telbany

Noora Kassab

Other Research Team Members:

Idris Alghazali

Katianna Beaudoin

Nikolas Belanger

Bradley Conlin

Reyhaneh Kalantari

Mackenzie Kathmann

Hind Mebtoul

Pulkit Mogra

Karim Sultan

Lacey Theoret

Note:

This draft report is an interim work product. It is an amalgamation of research work related to literature reviews and empirical investigations conducted over the course of the project. This information will be used to produce final products in the form of short media articles, academic articles and conference papers, and research seminars.

Please cite this work as follows: Ruhi, U., Nault, K., El-Telbany, O., Kassab, N. (2022, May). *Digital Disinformation Landscape and Online Disinformation-Countering Tools* (Working Paper), Telfer School of Management, University of Ottawa, Ottawa, ON.

*This research project is supported through the
Canadian Heritage's Digital Citizen Contribution Program*

Funded by the Government of Canada
Financé par le gouvernement du Canada

| **Canada**

Abstract

The spread of digital disinformation is threatening social cohesion and democracies globally. Malicious agents and corrupt individuals have waged an information war in an already saturated ecosystem by creating and spreading false information and manipulated digital media at unprecedented levels since the boom of the internet. As a result, scholars and researchers across many disciplines have taken an interest in strategizing against the dissemination of harmful media online. To begin, the array of terms describing the current ecosystem of “information disorders” has prompted the conglomeration of the growing taxonomy currently circulating in research communities. Inspired by the lack of a unified framework to describe the current infodemic, the literature review seeks to analyze the various interpretations and categorizations of the terms, as well as the enablers that promote the spread of digital disinformation. Sociotechnical, psychological, emotional, and technological enablers that promote the dissemination of digital disinformation and hinder its resolve are identified and discussed. Considering the severity of disinformation in current times and the need to eliminate its spread, we also consider current technology, as well as take an empirical look at the tools currently available. Moreover, we examine the usability and functionality of select tools to identify key-success factors that would promote their use and uptake amongst users. Finally, this review provides a discussion of recommendations from the literature, targeted for civil society, government, and the technology sector. These suggestions focus on fixing the underlying structural and institutional issues that have allowed for the widespread dissemination of disinformation and are integral to the future success of eradicating disinformation.

Table of Contents

Abstract..... ii

1. Introduction 8

2. Challenges in the Typology of Misinformation and Disinformation 9

3. Terminology and Characteristics 10

 3.1 Hypernyms 10

 3.2 Misinformation/Disinformation Taxonomy Ecosystem 12

 3.3 Misinformation: Unclear Intention, Uncertain Truth..... 13

 3.4 Disinformation: Certain Intention to Harm, Certain Falsification..... 14

 3.5 Malinformation: Certain Intention to Harm, Certain Truth..... 17

 3.6 Mixed Category Terms 18

 3.7 Amplifiers 18

4. Affective Associations of Disinformation..... 19

 4.1 Analysis of Disinformation and Emotion 20

 4.1.1 Interconnection between Emotions and Disinformation..... 21

 4.1.2 Role of Positive and Negative Emotions 22

 4.1.3 Association of Depression and Disinformation Beliefs 23

5. Sociotechnical and Psychological Enablers 24

 5.1 Cognitive Bias..... 24

 5.2 Confirmation Bias..... 25

 5.3 Motivated Reasoning 25

6. Sociological Enablers..... 26

 6.1 Social Belonging..... 26

 6.2 Journalism and Media Mistrust..... 26

 6.3 Institutional Mistrust..... 27

7. Technological Enablers 28

7.1 Social Media	28
7.2 Platform Amplification	29
7.3 Types of Content.....	29
7.4 Information Vacuums	30
7.5 Information Overload	30
7.6 Targeted Ads.....	30
8. Disinformation Drivers	31
8.1 Commercialization and Financial Incentives.....	31
8.2 Political Incentives.....	32
9. Technology and Tools.....	32
9.1 Technical Approaches - Machine Learning.....	32
9.1.2 Textual Analysis.....	33
9.1.3 Sentiment Analysis.....	33
9.1.4 Metadata Analysis	34
9.1.5 Image and Video Analysis	34
9.1.6 Forged Images.....	35
9.1.7 Deepfakes.....	35
9.2 Industry Approaches -Social Media Platforms.....	35
9.2.1 Meta.....	35
9.2.2 Youtube	37
9.2.3 TikTok.....	38
9.2.4 Warning Styles.....	38
9.2.5 Investments in Media Literacy.....	39
9.3 Online Tools.....	39
9.3.1 Education and Training	39

9.3.2 Human Fact-Checking	40
9.3.3 Credibility Scoring Tools	40
9.3.4 Bot Detection and Tracking	40
9.4 Functionality of Existing Tools	41
10. Empirical Review	43
10.1 Research Design.....	43
10.2 Methods	43
10.3 Inspection Methods.....	44
10.4 Empirical Methods.....	44
10.5 Participants.....	45
10.6 Data Collection and Analysis	45
10.7 Scenarios and Digital Disinformation Authenticating Tools.....	47
11. Results	47
11.1 Heuristic Evaluation	47
11.2 Moderated Usability Evaluation	49
Task 1: Locating Tools to Authenticate Digital Disinformation	50
Task 2: Article Credibility	51
Task 3: MediaBias/FactCheck Website	58
Task 4: isitcredible Website	61
Task 5: The Newsroom (Beta) Extension	64
Task 6: Newstrition Extension	66
Task 7: MediaBias/FactCheck Extension	69
11.3 Conclusions.....	71
11.4 Recommendations.....	75
Establish Credibility.....	75

Tool Layout and Design	75
Search Capability	75
Information and Presentation	76
12. Final Recommendations	78
12.1 Recommendations for Civil Society	78
12.1.1 Media Literacy	78
12.1.2 Building Community Resilience	79
12.1.3 Communication and Gatekeeping	80
12.2 Recommendations for the Government	80
12.2.1 Funding and Legislative Support	80
12.2.2 Rebuilding Trust.....	81
12.3 Policy Recommendations	81
12.3.1 Policy Mandate on the Use of Artificial Intelligence.....	81
12.3.2 Towards "Susceptibility by Design"	82
12.4 Recommendations for (Big) Tech.....	82
12.4.1 Cross-Disciplinary Recommendations for Big Tech’s Considerations	83
12.4.2 Detection	84
12.4.3 Protection	85
12.4.4 Social Media Campaigns and Gaming.....	86
12.4.5 Fact-Checking Tool Interface and Design	86
12.4.6 Use of Blockchain Technology	87
13. Future Research: Digital Disinformation Issues and Behaviors Survey	88
14. References	89
15. Appendices	103
Appendix A.....	103

Appendix B.....	104
Appendix C.....	105
Appendix D. Heuristic Evaluations – MediaBias/FactCheck Website.....	112
Appendix E. Heuristic Evaluations – isthiscredible Website	114
Appendix F. Heuristic Evaluations – The Newsroom Beta Extension.....	116
Appendix G. Heuristic Evaluations – Newstrition Extension	118
Appendix H. Heuristic Evaluations – MediaBias/FactCheck Extension.....	120
Appendix I. Online Survey – Digital Disinformation Issues and Behaviours.....	122

1. Introduction

The interconnectedness of individuals, communities, and societies through the web has allowed for the spread of information quickly on a global scale. Correspondingly, the discourse surrounding false information has rapidly evolved as misinformation and disinformation is amplified in the online sphere in mass. Despite the spread of false information not being a new phenomenon as seen historically (Hamilton et al., 2021; Posetti & Matthews, 2018), the scale of the problem poses a public health and security risk when considering the speed of the spread in the age of the internet (Gradoń et al., 2021; Mehta et al., 2021). Just last year, the Center for Countering Digital Hate identified “The Disinformation Dozen” - a mere 12 individuals that were responsible for 65% of the Covid-19 anti-vaccine misinformation online, and 73% of the related misinformation on Facebook (Centre for Countering Digital Hate, 2021).

The spread of false information has been likened to the spread of disease, known as an infodemic (Gradoń et al., 2021; Hamilton et al., 2021; Rubin, 2019). The World Health Organization (WHO) defines an infodemic as a “volume of information that grows exponentially due to a specific occurrence, consequently resulting in unverifiable information spreading like a virus” (2020). Like the mutation of a virus, language is organic and evolves as well, resulting in new terminology used to describe false information, misinformation, and disinformation. In the first phase of this research, this literature review seeks to classify the various terms and taxonomy describing several aspects of false information through the lens’ of the following disciplines: communication, information science, and computer science/data science, to formulate an interdisciplinary taxonomy of disinformation. Various academic researchers and industry research groups have recently called for such type of research (Rubin, 2019; Kapantai et al., 2021; Hewlett Foundation, 2018).

Digital disinformation occurs in articles, photographs, and even videos (known as deepfakes) and spread easily and quickly online due to many factors. Researchers across multiple disciplines have attempted to isolate different enablers in order to devise solutions that would allow for elimination. The dissemination of online disinformation is enabled by many sociological, technological, and psychological factors. Historically, division was sown by corrupt politicians to gain power, targeting marginalized communities, fear mongering, and fueling mistrust for the benefit of the elite. With the dawn of the internet and social media, the spread of disinformation is able to spread to further corners of the world and exploit cultural and language barriers. It is driven by financial and political incentives of a few who wreak havoc on many. Disinformation spreaders capitalize on humans’ natural cognitive shortcomings from being too trusting to not being trusting enough. Human emotions, psyche, and socioeconomic status all play a role in the belief in disinformation. Cognitive biases enable disseminators to slowly radicalize and further divide individuals, separating them from the majority, unbeknownst to the victims. Disinformation succeeds because it takes advantage of intentional deceit, but also because of fallible minds.

Unhelpful to the situation, is big tech’s algorithms and natural platform amplification, as well as their tracking and targeting. While this has enabled the commercialization of many aspects of the internet, it has also become an incentive to draw the attention of users through lies and sensational headlines. Put simply, more user engagements = more money for the author and publisher. The commercialisation of the news media industry has had especially devastating impacts due to the subsequent loss of credibility and trust in well-known publications. Biased headlines, wrong information, and unverifiable information spread by major news outlets has resulted in a decreased reliability of news outlets for many citizens.

To identify and counter the different types of disinformation, various online tools have been developed in recent years by private sector companies, non-profit organizations, and civil society organizations. These tools range from websites powered by human fact-checkers to verify the accuracy of information, to bot and spam detection tools that can identify automated bot activity on social media, to automated artificial intelligence applications that can generate credibility ratings for online content. The purpose of these tools is to help information consumers navigate today's challenging information environment by separating reliable information sources from false or misleading information. Nonetheless, such disinformation-countering tools has revealed a dearth in academic and industry research on the efficacy of these tools, both in terms of their functionality and usability, and there have been various calls for this type of research from both academic as well as practitioner communities.

Accordingly, in the second phase of this research project, various online disinformation-countering tools were examined by mapping their functionality against the taxonomy of disinformation that was developed in the previous phase. This enabled us to determine any gaps in tools to counter disinformation.

Consequently, in the third phase of this research, several tools were selected, and a user-experience evaluation was performed through a combination of heuristic usability evaluation and remote usability testing with representative end-users. This helped in validating the overall efficacy of these tools in meeting the contextual, technical, and cognitive needs of people who would potentially use these tools as part of their daily information consumption activities. Additionally, it also identified the gaps in the functionality and usability of current online disinformation countering tools and provide recommendations for improving the design and development of these tools. Finally, this research also highlights the key success factors for the widescale adoption of disinformation-countering tools, and ultimately help foster wider institutional and societal resilience to disinformation. This also paved the path to some future research recommendations.

2. Challenges in the Typology of Misinformation and Disinformation

The paradoxes of defining the terms used within the literature surrounding mis-/disinformation are akin to the root causes behind the spread of misinformation: an overuse and amplification of some terms such as “fake news;” a lack of uniform understanding and framework for use of the terms; and, the emergence of new terms (Kapantai et al., 2021; Wardle & Derakhshan, 2017). Considering the nature of the infodemic, a multidisciplinary strategy is fundamental in combating its effects (Kapantai et al., 2021; Mehta et al., 2021; Rubin, 2019; Tandoc et al., 2018).

Accordingly, different weights towards the effects and usage of the terms are emphasized by different researchers and disciplines. Some literature provided narrow definitions while others opted for more comprehensive options, such as definitions that examined the nature and intent of the spread itself. Furthermore, considering the broad scope of research, novel words were referred to in the text by some authors but seldom others. This augmented the quantity of terms to define, further exacerbating the challenge of creating a comprehensive typology and consolidating the disparate definitions. While authors acknowledge the oversaturation of terms to describe disinformation, terms have been studied separately to establish strategies for detection by understanding patterns in language syntax and spread (Ghanem et al., 2020; Giachanou et al., 2022). While these findings represent significant advancements in the effort to combat digital disinformation, they have little impact on citizens when it comes to identifying false information (Carmi et al., 2020; Hinsley & Holton, 2021).

3. Terminology and Characteristics

Upon review of the literature, several themes and classifications emerged surrounding the typology of mis-/dis-information. Some authors analyzed the specific terminology to shape their research (Ghanem et al., 2020; Shu et al., 2017; Tandoc et al., 2018) while others conveyed a more blasé interpretation, paying little attention to the nuances of intention and impact for each type of misinformation. The first theme of typology groups hypernyms that have been used to describe or explain the spread of mis-/dis- information and the general ecosystem of information. These terms are used to describe the over-arching crisis of mis-/dis- information. Beyond hypernyms, many authors relied on factors of verifiability, facticity, or even level of author involvement (Gradoń et al., 2021; Kapantai et al., 2021; Wardle, 2018; Zannettou et al., 2019). Relying on factors such as on the facticity and intention of the disseminator to deceive or cause harm allowed for secondary and tertiary groupings of the terms based on intention of the disseminator (Kumar & Shah, 2018; Tandoc et al., 2018; Wardle & Derakhshan, 2017). These are analyzed through the terms misinformation, disinformation, and sometimes mal-information in this literature. Finally, a section is presented on terms that don't belong to either of the preliminary groupings.

Considering the breadth of the subject, the variations in the usage of terms are plentiful. Recent literature reveals the lack of a unified framework that is shared across the disciplines (Allcott & Gentzkow, 2017; Kapantai et al., 2021; Kumar & Shah, 2018; Wardle, 2018; Zhou & Zafarani, 2018). The nature of the disciplines draws each to uniquely identify with specific encompassing terms. In the context surrounding these terms, researchers have relied on the overarching terms to introduce the concept of mis-/dis-information, as well as underscore the dangers of it. In 2019, the WHO named the spread of misinformation as a top 10 global threat, solidifying the impending doom of disinformation in the digital era. Beginning with hypernyms this review will define then synthesize the literature surrounding the taxonomy of digital disinformation.

3.1 Hypernyms

The era of misinformation has plagued communities with deception in combination with “societal mega trends, such as decline in social capital, growing economic inequality, increased polarization, increased polarization, declining trust in science, and an increasingly fractioned media landscape,”(Lewandowsky et al., 2017, p.1) causing people to rely on opinions rather than facts and discredit experts (Lewandowsky et al., 2017; Westerlund, 2019; Zannettou et al., 2019).Throughout the literature, the above-mentioned terms were used in a broadscale manner to describe the consequences of the age of misinformation. Compared to other terms, they were distinctly used more frequently in some disciplines compared to others.

Infodemic: A union of “information” and “pandemic,” *infodemic* first appeared in 2003 after the SARS breakout. David Rothkopf who coined the term describes it as “a few facts, mixed with fear, speculation and rumor, amplified and relayed swiftly worldwide by modern information technologies”(Hamilton et al., 2021, p.1) Recent literature refers to the infodemic in light of Covid-19 and proliferation of digital disinformation (N. K. Agarwal & Alsaeedi, 2020; Gradoń et al., 2021; Hamilton et al., 2021; Kumari et al., 2022).

Disinfodemic: A union of the terms “disinformation” and “infodemic,” *disinfodemic* refers specifically to the uncontained spread of disinformation especially with regards to dangerous and severe consequences of false and fabricated information, particularly around Covid-19. Authors

commonly cite Posetti and Bontcheva as found in UNESCO policy brief authored (Giachanou et al., 2021; Horne, 2021; Levak, 2020; Santos-D'Amorim & Miranda, 2021).

Information Disorder: Wardle and Derakhshan (2017; 2018) use the term “information disorder” to conceptualise their framework describing the *types*, *phases*, and *elements* of misinformation and disinformation in a detailed manner. Types in the information ecosystem encompass misinformation, disinformation, and mal-information. Phases include the creation, production, and distribution of false information. Finally, elements include agent, message, and interpreter of the message.

Information pollution: A hypernym used to describe “the presence and spread of unwanted messages in human society, in large enough quantities to cause significant negative effects on human and social activity” (Bran et al, 2021, p.1). Other authors refer to term within the context of contamination and the resulting harm (Akers et al., 2018; Kapantai et al., 2021; Meel & Vishwakarma, 2020; Wardle & Derakhshan, 2017).

Post-truth (era): As a consequence to the copious amount of false information available, the term post-truth emerged as descriptor for the devastating impact that resulted from it. The spread of digital mis-/dis-information has created a social ecosystem of distrust in science and granted anyone to sway public opinion through an appeal to emotion or personal belief rather than objective fact (Horne, 2021; Mayorga et al., 2020; Molina et al., 2021; Oxford Languages, 2016; Paschalides et al., 2021).

From our full-text review, *infodemic* appears mostly in literature involving data science and information science. Various researchers have relied on the term because of its relation to biology, information mapping, Natural Language Processing (NLP), and machine learning (N. K. Agarwal & Alsaedi, 2020; Giachanou et al., 2022; Gradoń et al., 2021; Hamilton et al., 2021; Yesmin & Ahmed, 2022). Infodemic accurately describes a contagion model for understanding the spread of false information and misinformation through an epidemiological lens. Researchers in these disciplines dissect the spread of information in order to compare it to the viral load of misinformation, its transmissibility, susceptibility, acceleration of spread, and even its effect on governmental and public responses (Fallis, 2015; Gradoń et al., 2021; Hamilton et al., 2021; Yesmin & Ahmed, 2022). It has been popularized by the WHO and mainstream media, especially with relevant literature surrounding COVID-19.

Similarly, information pollution is used in some literature with regards to the spread of harmful information (Akers et al., 2018; Bran et al., 2021; Kapantai et al., 2021; Meel & Vishwakarma, 2020; Wardle & Derakhshan, 2017). Disinfodemic is used less frequently in the text compared to other hypernym, likely due to how recent it is.

Contrastingly, the term *information disorder* has mostly been cited in the communications discipline; however, information scientists have also used it as a blanket term for the consequences of misinformation, disinformation, and mal-information (Herrero-Diz & López-Rufino, 2021). Literature that uses the term *information disorder* has consistently cited the work of Wardle and Derakhshan (2017) for the European Council on policy making. Finally, *post-truth* has been used heavily in the communications discipline. In some instances, post-truth is used to describe the era in a neutral manner on a time horizon; other times researchers make a point to deliver on the malicious political intent of disinformation. Hypernyms are used most often in literature that studies the harmful effects of misinformation and disinformation on a social level.

3.2 Misinformation/Disinformation Taxonomy Ecosystem

The ecosystem of terms is ridden with a variety of older and newer terms. Wardle and Derakshan (2017) create a valuable framework for the information pollution ecosystem, helping to dissect the message, creator and disseminator. However, their framework is criticized for adding new terms that are seldom used by others in an already vast taxonomical set (Bran et al., 2021). The elected classification seeks to explore the unifying features of the terminology to better understand the ecosystem of which these terms proliferate. Furthermore, special attention was paid to the frequency of which these terms occurred in the text in order to identify significant patterns and connections in the literature. Appendix A and Appendix B provide additional tables and graphs used when developing the analysis. Appendix C presents a table, complete with definitions and examples for each of the terms. It is important to note the subjectivity of the matter. For example, while some individuals might view government propaganda as protective and patriotic within their borders, when it arises from foreign actors it becomes a threat. This alteration in understanding would carry propaganda from misinformation to disinformation.

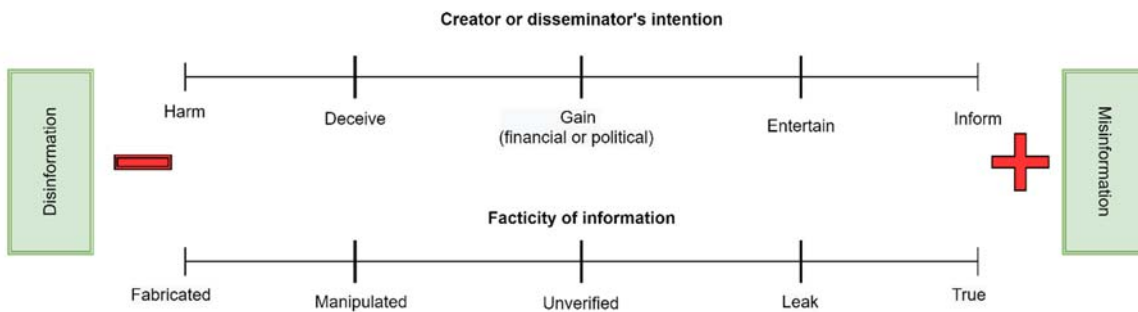


Figure 1. Author's conception of a scale to measure intention and facticity of term used to describe message.

To further illustrate the methodology behind the classification, the above figure was incepted during the literature review process. Each of the terms could be measured on a scale based of the intention of the disseminator and the level of facticity of the message itself. This figure denotes the extreme ends, marking intention as the discerning feature. In addition to the scale, the frequency (Appendix B), interpretation, and understanding of the terms is depicted in Figure 2. Multi-layer Venn diagram classifying misinformation and disinformation ecosystem typology (subjective). (below).

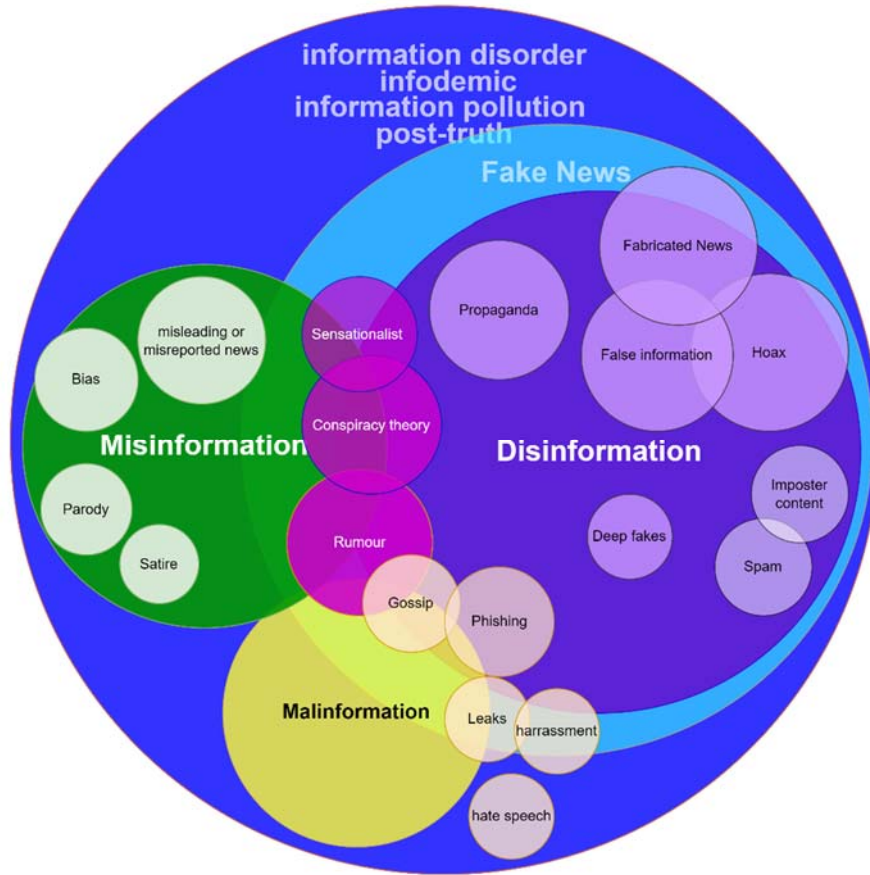


Figure 2. Multi-layer Venn diagram classifying misinformation and disinformation ecosystem typology (subjective).

3.3 Misinformation: Unclear Intention, Uncertain Truth

Misinformation was one of the most commonly occurring terms in the literature. Much of the literature references misinformation using it synonymously with disinformation (Flores-Saviaga & Savage, 2019; Froehlich, 2017; Pierre, 2020). This is likely based on its debut in Dictionary.com’s 2018 Word of the year, defining it as information that is spread regardless of intention to mislead and may contain inaccurate or incomplete information (Ha et al., 2021). However, in most of the recent literature accessed for this review, it is defined as false information that is spread without the intent to harm, usually because of carelessness or cognitive biases (Kumar & Shah, 2018; Mayorga et al., 2020; Meel & Vishwakarma, 2020; Wardle & Derakhshan, 2017). Molina et al (2021) also argue that the definition of misinformation has changed over the years however, it is established that misinformation is now interpreted as misleading and highlights the importance of identifying intention when labeling or defining misinformation. As such, other terms that were determined to be well intentioned (towards the positive scale on the above-mentioned framework), were included in this section.

Parody and satire are used interchangeably in communication and information science. Literature refers to parody and satire misinformation that is spread without the intent to harm but with a potential to fool the reader or viewer. Much of the literature tends to group the two together without a clear distinction of the terms: mockery in the case of parody, and irony / dry humour in the case of satire (Hinsley & Holton, 2021; Horne, 2021; Mayorga et al., 2020; Molina et al., 2021; Santos-D'amorim & Miranda, 2021). However, it is worth mentioning that Tandoc et al (2019) in their research note that while satire relies on humour and exaggeration to mock issues, parody uses “non-factual information to inject humour”. It is arguable that false facts used in parodies would qualify them as disinformation; however, the nature of satire and parodies requires an assumed understanding between the author and the reader for the use of humour to deliver the news piece (Shu et al., 2017; Tandoc et al., 2018). While this is relevant, it neglects the cultural cues that could be missed by consumers of different cultures (Rubin, 2019).

Bias appears in the literature often within the context of cognitive bias and media “one-sided” bias in journalism. Although biased news/information is not neutral, it is not considered to be intentionally harmful and instead, a cognitive bias based on experience and social upbringing, or pure ignorance, that shapes the author’s writing (Kapantai et al., 2021; Santos-D'amorim & Miranda, 2021; Zannettou et al., 2019).

When misinformation is discussed in the literature, it is often used to describe social consequences and polarization (Marwick & Lewis, 2017; Wardle & Derakhshan, 2017) rather than intentional deception. A spreader of misinformation might consider themselves informing others by sharing information they think is helpful (Chadwick & Stanyer, 2022; Rubin, 2019). Some authors have also acknowledged the humanity of reporters, noting that they are as susceptible to deception and bias as everyone else (Marwick & Lewis, 2017; Rubin, 2019). Thus, misleading news and misleading reporting can also be included under misinformation as they are a consequence of bias. This is due to their use of “sensationalist” words and headlines that capture the reader’s attention (Meel & Vishwakarma, 2020). Though their motive is primarily economic, attention-grabbing headlines as a tactic for website traffic is debatable in terms of information ethics (Au et al., 2021; Froehlich, 2017; Meel & Vishwakarma, 2020). It is important to note that the intention to mislead by itself would be considered as disinformation (Ha et al., 2021; Meel & Vishwakarma, 2020; Wardle & Derakhshan, 2017).

3.4 Disinformation: Certain Intention to Harm, Certain Falsification

Based on the review, the literature reveals a huge interest in disinformation which is commonly agreed on to be known false or incorrect information, spread with deliberate intention to harm or deceive (Fallis, 2015; Kapantai et al., 2021; Kumar & Shah, 2018; Mayorga et al., 2020; Meel & Vishwakarma, 2020; Wardle & Derakhshan, 2017). Much of the literature that describes disinformation emphasizes the level of deception intended by the author. This could be further divided by the intent to mislead or the intentional fabrication of facts. Kumar and Shah (2018) have offered terms to describe certain types as opinion-based *such as, conspiracy theories or fake news* or fact-based *such as false news or hoax*; however, these are not common categories. Communications researchers expand the definition of disinformation to include fabricated or manipulated content in addition to gossip and rumours (Carmi et al., 2020; Hinsley & Holton, 2021; Wardle & Derakhshan, 2017). Across the disciplines, disinformation is consistently defined within the context of the intent behind it (Hinsley & Holton, 2021; Kapantai et al., 2021; Shu et al., 2017;

Zannettou et al., 2019). For this reason, this literature review categorizes terms describing types of disinformation based on them being intentionally fabricated and harmful. This would be the far left of the framework in Figure 1.

The underlying theme is media, whether news, videos, or pictures, that were deliberately created with the intent of causing harm regardless of whether it is to people, organizations, or countries. Researchers attribute the intent to harm for reasons of economic, political, and financial gain (Fallis, 2015; Gradoń et al., 2021; Molina et al., 2021; Rubin, 2019; Shu et al., 2017). Studies on deception, the role of trust in people's inherent nature, and information ethics, highlight the dark and ugly intentions of disinformation propagators and those who take advantage of vulnerable users (Chadwick et al., 2022; Froehlick, 2017; Waldman, 2022).

False information: Put simply, false information is known wrong information. It is a direct, neutral and well understood term as it is commonly used *within* the definitions of other terms, with little regard to the nuances of the intent behind the spread and intention to deceive in most literature (Chadwick & Stanyer, 2022; Molina et al., 2021; Wardle & Derakhshan, 2017; Zannettou et al., 2019).

Fake News: Google Trends marked February 2017 as the peak month for “fake news” searches, but the term existed has long before former U.S. President Donald Trump popularized it in his 2016 election (Quandt et al., 2019). Citing the work of Wardle and Derakhshan (2017), many researchers in the communication sector have taken a stance against the use of “fake news” due to its over-generalization to several types of false information as well as its misappropriation by politicians. “Fake news” is used as an umbrella term to describe a range of information types; and is often used by politicians to dismiss the news, undermine its authenticity, silence free speech, and evade any accountability or news they disagree with (Ansar & Goswami, 2021; Horne, 2021; Kapantai et al., 2021; Wardle & Derakhshan, 2017). Other researchers, most notably in data science and information science, simply define fake news as false information spread online with the intent to harm for financial or political gain (Choraś et al., 2021; Meel & Vishwakarma, 2020; Zhang & Ghorbani, 2020). Communications authors underscore the detrimental effects of the use of the term “fake news” and opt for more descriptive terms in their research.

Fabricated News: Entirely false or fictional stories created with the intention of causing harm (Ansar & Goswami, 2021; Kapantai et al., 2021; Zannettou et al., 2019). Fabricated news is found in the same format as legitimate news articles (Ansar & Goswami, 2021), making it one of the most dangerous types of false information (Zannettou et al., 2019). The difficulty in identifying fabricated news and severity of harm that it could cause is that it is often shared by organizations that are seemingly objective (Tandoc et al., 2018). In the context of the articles analyzed, fabricated news across the disciplines referred specifically to false news articles with a deceptive motive such as social confusion, political, or economic gain. It is a preferred alternative to “fake news” considering the overuse of the latter term (Allcott & Gentzkow, 2017; Ansar & Goswami, 2021; Zubiaga et al., 2018). It is most often distinguished in the communication and information science disciplines and is synonymous with “fake news” in the field of data science.

Deep fakes: Rapid technological innovation in machine learning and artificial intelligence have allowed for a type of hyper-real, fabricated multimedia content to emerge. Reference to deep fakes typically imply that they are created synthetically with artificial intelligence and result in fake videos that are hard to distinguish from real (S. Agarwal et al., 2019; Vaccari & Chadwick, 2020; Westerlund, 2019). So much so that they could result in the “implantation of false memories and

manipulate individuals towards nefarious actions” thus having dire cognitive and social impacts and influences (Liv and Greenbaum, 2020 as cited in Bastick, 2021).

Propaganda: Propaganda is defined as deceptive information spread with the intention of influencing behavior in targeted communities, for political reasons or to push an agenda, usually by government institutions (Fallis, 2015; Gradoń et al., 2021; Hamilton et al., 2021; Meel & Vishwakarma, 2020; Shu et al., 2017; Zannettou et al., 2019). The motive behind propaganda is highly agreed to be highly politically motivated and intended to cause harm. The term is frequently used across all disciplines and in political context.

Sensationalist news: Although not a frequently used term, sensationalist news refers to information that is intended to evoke a reaction and thus qualifies as intentionally manipulative. Molina et al describe this type of content as polarizing and meant to align with particular ideologies that are based on inciting fear (2021). *Sensationalist* appears in text that references fake news or clickbait as content that is meant to increase engagements from social media users through emotional and attention-grabbing headlines (Akers et al., 2018; Horne, 2021; Molina et al., 2021; Santos-D’amorim & Miranda, 2021).

Hoax: A large-scale, intentionally deceptive fabrication or propagation of a conspiracy theory that is presented as a true or legitimate story (Chew & Eysenbach, 2010; Molina et al., 2021; Rubin, 2019; Santos-D’amorim & Miranda, 2021). A hoax is often so deceptive in nature it transpires any fun or scam and causes harm to its victims by convincing them of a false of “paranoia-filled story” (Ghanem et al., 2020; Kumar & Shah, 2018; Rubin, 2019). Data science articles find similarities with communication researchers and note that a hoax is intentionally circulated by the author with the aim to deceive; and, the accuracy of the narrative being quite low (Choraś et al., 2021; Ghanem et al., 2020). Contrastingly, information scientists state that hoaxes are often utilised in news stories to present false claims as legitimate and that in the research community they are known as half-truths, implying that they are not fully fabricated as other article types suggest (Zannettou et al., 2019).

Clickbait: Clickbait is a term coined wherein media sources hyperbolize their fabricated or true stories to attract attention to the intended webpage through emotionally charged titles or by withholding information in the title (Ghanem et al., 2020; Kapantai et al., 2021; Meel & Vishwakarma, 2020; Rubin, 2019; Shu et al., 2017). Content creators use their article or video titles as well as the thumbnail picture to emphasise the most attractive part of the story to the reader and in clickbait, the content is rarely as enticing as the original imagery lead the consumer to believe (Kapantai et al., 2021; Rubin, 2019). Although clickbait refers directly to content found on the internet, this phenomenon in print articles and was known as “yellow journalism” (Chew & Eysenbach, 2010; Molina et al., 2021; Zannettou et al., 2019). Wardle and Derakshan file clickbait under their term “false connection” where the headline and content do not match. Overall, the understanding and usage of the term “Clickbait” is common between data scientists, communication journalists and information scientists with no notable discrepancies between their definitions.

Dark Patterns: application and website interface features designed to intentionally mislead users into providing consent, purchasing items or subscribing to news (Carmi et al., 2020; Waldman, 2020). This term is seldom discussed in the literature whether as a term or enabler. However, considering the impact of dark patterns on privacy, security, and tracking of habits, it is an

important aspect of the information sphere deserving of attention and regulation (Carmi et al., 2020; Waldman, 2020).

Spam: Spam was not clearly defined in the literature, but researchers agree that it is intentionally deceitful information used to take up space of legitimate information, or propagate false information (Gradoń et al., 2021; Rubin, 2019). Meel and Vishwakarma (2020) also consider *fake reviews* as spam.

The difficulty with disinformation is that the receiver of the message may not be cognizant of the author or disseminator's intention (Ha et al., 2021). As such, it may be shared or engaged with unintentionally by misled users (Bran et al., 2021; Chadwick & Stanyer, 2022). Some authors have chosen to leave out terms that are falsified but well intentioned such as humorous hoaxes, parodies, satire (Shu et al., 2017), deep fakes and defamation ((Kapantai et al., 2021). However, there is a rising trend in the inclusion of satire and parody under disinformation or fake news since they have a strong influence on the media ecosystem (Ha et al., 2021; Kapantai et al., 2021; Tandoc et al., 2018). When considering the influence of satirical news programs on young consumers especially, it becomes increasingly obvious that satire has shaped political discourse (Bran et al., 2021; Tandoc et al., 2018).

3.5 Malinformation: Certain Intention to Harm, Certain Truth

Mal-information is defined as genuine or true information deliberately manipulated or taken out of context with the intent to deceive and cause harm (Hinsley & Holton, 2021; Mayorga et al., 2020; Wardle & Derakhshan, 2017). It is primarily used in the communication discipline and seldomly referenced in data science. Wardle & Derakhshan (2017) include harassment, leaks, and hate speech in their definition and distinguish mal-information from other types by virtue of true information being exploited in the public sphere.

Phishing is referred to in the context of digital security and threat response since it has to do with theft of personal and identity information (Akers et al., 2018; Figueira & Oliveira, 2017; Wardle & Derakhshan, 2017) and in the context of spam and bot detection in data science (Kim et al., 2018; Shu et al., 2017). Others consider phishing as being done by social spammers (Bondielli & Marcelloni, 2019), or grouped with clickbait such as (Meel & Vishwakarma, 2020). Interestingly, Santos-D'Amorim & Miranda (2021) categorize phishing as malinformation because it contains personal and confidential information.

Leaks: Based on Wardle and Derakhshan's (2017) definition of true information released with the intent to harm and classified as malinformation. Context of leaks in the literature refers to leaks with connection to foreign malicious actors (Allcott & Gentzkow, 2017; Bastos & Farkas, 2019; Tandoc et al., 2018; Zhang & Ghorbani, 2020).

Harassment: Unwanted advances that subject a person or group to stalking or online attacks. Wardle and Derakhshan group harassment with malinformation. Some authors group harassment with spamming or trolling or doxing (Froehlich, 2017; McLane, 2021; Wardle & Derakhshan, 2017).

Hate speech: Factually inaccurate (Golebiewski & Boyd, 2018) or out of context information, used to cause harm (Wardle & Derakhshan, 2017). In the literature, hate speech is a type of offensive

speech or defamation arising from disinformation campaigns that needs to be regulated due to its offensive racial and sexual nature (Kapantai et al., 2021; Marwick & Lewis, 2017; Mayorga et al., 2020; Rubin, 2019).

There is a clear pattern in the literature of malinformation being mentioned in connection with Wardle & Derakshan (2017) such as Hinsley and Holton (2020) and Mayorga et al (2020), but items under this category are not categorized similarly across the literature. For example, Wardle and Derakshan list hate speech under malinformation, where it is “out of context” but it is difficult to imagine any hate speech to be out of context when it directly assaults a person or group. Very clear legislation has been implemented around hate speech to this end (Marwick & Lewis, 2017; Rubin, 2019). Furthermore, the above-mentioned terms are traditionally used to explain the consequences of a polluted information ecosystem and not a direct type of information (Carmi et al., 2020; Marwick & Lewis, 2017).

3.6 Mixed Category Terms

Conspiracy Theory: Conspiracy theories are not based on factual information and are often used to explain important events and aim to implicate governments or powerful individuals in illegal acts (Kapantai et al., 2021; Meel & Vishwakarma, 2020; Santos-D’amorim & Miranda, 2021; Zannettou et al., 2019). This understanding of a conspiracy gives it similar characteristics to a rumour however many researchers argue that conspiracies are unverified information and should not be categorized with “fake news” as intentionally misleading since by definition, they difficult to verify as true or false (Bondielli & Marcelloni, 2019; Choraś et al., 2021; Kapantai et al., 2021; Molina et al., 2021). Taking into account the questionable intention of a conspiracy theory propagator, conspiracy theories could be classified as both disinformation or misinformation once confirmed. However, given their long-standing nature, this is hard to establish at first revelation especially since research has found that evidence presented to refute claims of conspiracy often further proves the original theory (EAVI, 2018 as cited in(Kapantai et al., 2021). Data and information scientists as well as communication journalists are in alignment with their definitions of conspiracy theories and lack of unified agreement on intention.

Rumour: Rumours are referred to as anecdotes or stories that, although may be widely known, are ambiguous in nature and are not confirmed to be true (Ansar & Goswami, 2021; Kapantai et al., 2021; Kumar & Shah, 2018; Shu et al., 2017). They are often propagated in social networks and media sources (Peterson and Gist, 1951 as cited in Kapantai et al., 2021). Data Science articles as well as communication understand rumours to be ultimately false whereas information science articles make a point to question whether or not there is truth to rumours, noting that they are unverified pieces of information which are not necessarily false (Ansar & Goswami, 2021; Bondielli & Marcelloni, 2019; Meel & Vishwakarma, 2020). They also note, which data scientists allude to, that rumours originate with one source and spread quickly to other information mediums (Yesmin & Ahmed, 2022).

3.7 Amplifiers

The terms in this category are used interchangeably within the literature, often in connection with confirmation biases and technological enablers of misinformation spread. Some argue that humans as a species have always lived in echo chambers, as these are simply their closest networks and communities (Wardle & Derakshan, 2017). Technology companies, particularly

those providing forms of social communications, such as Meta, Twitter, and Youtube have simply replicated a naturally occurring phenomenon on their platforms to create online communities.

Echo Chambers: Consistently documented by data and information scientists as a phenomenon wherein consumers are repeatedly exposed to information that favours a perspective they already have, particularly on social media (Ansar & Goswami, 2021; Chen et al., 2022; Meel & Vishwakarma, 2020; Wardle & Derakhshan, 2017; Zhang & Ghorbani, 2020). This extensive creation of a confirmation bias can exacerbate ideological polarisation and intolerances as well as the quick spread of false information, further homogenizing online information ecosystems (Kumari et al., 2022; Santos-D'Amorim & Miranda, 2021; Shu et al., 2017). Echo chambers were seldomly mentioned within the selected communications literature.

Filter bubbles: Similar to echo chambers, filter bubbles are algorithmically created, narrowed down online ecosystems created on the basis of interests and search results that isolate users from ideas or beliefs that are contrary to their ideologies (Agarwal & Alsaedi, 2020; Bastick, 2021; Horne, 2021). Communication journalists understand this to be true and expand on their thought by noting that filter bubbles are main causes of concern in the dissemination of misinformation. They are present where members of a community have a desire to take a clear stance on a partisan divide (Mayorga et al., 2020). Information Scientists share this definition of filter bubbles and add that they create a personal ecosystem of information, which allows not only for a strong confirmation bias, but spreading of misinformed narratives as well (Gleick, 2011 as cited in Santos-D'Amorim & Miranda, 2021).

4. Affective Associations of Disinformation

Disinformation is a critical problem over the internet, especially on social media and messaging platforms, due to its potency to influence the cognitive patterns of the people. Due to the affordability and freemium model on messaging apps and social media websites, access to technology has not been a concern anymore. With roughly 2.91 billion monthly active users reported during the fourth quarter of 2021, Facebook has emerged as one of the most used online social networks worldwide, followed by WhatsApp, with approximately 2 billion active users monthly. (*Facebook MAU Worldwide 2021*, n.d.; *WhatsApp - Statistics & Facts | Statista*, n.d.)

With the convenience of communication and information dissemination to a wider audience, social media and messaging apps have become a part of daily life. Social media has not only been limited to communications but also has turned into a major source of news dissemination (Groshek & Bronda, n.d.). According to a Pew Research Centre survey conducted in 2021, the report states that approx. 48 % of the United States population gets their news from social media (Walker & Matsa, 2021). However, despite being a great source of information, over the last two decades, the internet has also emerged as a common channel that has been responsible for the propagation of false information, turning into a "Web of Deception"(Kumar & Geethakumari, 2014a).

Influenced by a fusion of both accurate information and disinformation, the internet has left users prone to emotional fluctuations. They tend to publish their views, often driven by different emotions, on social media platforms (Zhang et al., 2017). Eventually, social media algorithms tend to replicate the set of information that the users are interested through various social networks and groups with homogenous public opinions (Pennycook & Rand, 2019a). Often the credibility of the information sources remains undefined, resulting in the spread of disinformation. House of Commons, in their report, defined disinformation as "the deliberate creation and sharing of false

and/or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm or for political, personal or financial gain” (House of Commons Digital, Culture, Media and Sport Committee, 2019). The spread of rumours and misleading information, especially on social media platforms, not only causes public concern but also poses a significant threat to the user’s physical and psychological health but also brings challenges to the governance and social order (Lazer et al., 2018).

Often, when disinformation is spreading amongst a mass, the primary intention of the information is to mislead audiences (Bradshaw & Howard, 2019). However, on a wider scale, the intended effects of disinformation include increasing group polarization, political divisions, reducing trust, and undermining civil society (Howard et al., 2018). Experimental studies have shown that disinformation’s potency to mislead can affect users’ psychological health (Buchanan, 2020) and can result in changes in behavioural and attitude patterns (Zerback et al., 2021). Inferences of the same were noticed when a mob attacked telecommunications masts in response to disinformation based on 5G telecommunication technology, causing a novel coronavirus (Parveen & Waterson, 2020). Such computational propaganda has been described as a pervasive and ubiquitous part of user’s life, used as a tool for spreading disinformation through social media (StratCom | NATO Strategic Communications Centre of Excellence Riga, Latvia, n.d.).

Indistinguishability between real and intentionally misleading information remains a challenging task due to the presentational design of disinformation (Karami et al., 2021). Though researchers have identified characteristics of posts and user profiles that spread disinformation, factors that instigate users to spread disinformation on social media are yet to present a conclusionary evidence. Past research has reflected that users are more prone to propagating disinformation in situations that make one emotionally overwhelmed and when the topics are of personal importance (Pennycook & Rand, 2019a). Emotional pressurizing conditions such as anxiety leaves people more vulnerable towards spreading unproven claims as disinformation can often serve to justify feelings and relaxes one from emotional tension (Zeller, 1948). Also, the content of disinformation possesses a higher likelihood of stimulating emotions such as anxiety, fear, surprise, and disgust compared to the emotions evoked by true information, such as sadness and joy (Vosoughi et al., 2018).

4.1 Analysis of Disinformation and Emotion

Emotions play an essential role in spreading disinformation through social media. Because of its capability to meet the psychological needs of the users. Related studies have concluded that there exists a close connection between the spread of disinformation and emotions (Oh et al., 2013). Negative emotions such as anxiety play a pivotal role in spreading disinformation (Oh et al., 2013). Studies conducted to study disinformation on social media during public crises to identify emotions conducted sentiment analysis and found that public sentiments change along with the change in content, font, and linguistics of disinformation. The more conflicting the content, attractive and exaggerated use of language, the more intense the sentiments (Leng et al., 2020).

Hence, the disinformation on social media affects the user’s psychology and directly impacts the stability of social order and security. Therefore, investigating what sets of emotions drive one to consume and share disinformation can help us understand how human emotions have a deep-rooted connection to spreading disinformation. Also, it can be beneficial to provide suggestions for emotion management for users of social media platforms.

Based on a comprehensive literature review, this section discusses the role of emotions and the spread of disinformation in three spheres

- a. Interconnection between emotions and disinformation.
- b. Role of positive and negative emotions in spreading disinformation.
- c. Association of depression with disinformation beliefs.

The section analyses how emotion plays an instrumental role in a human's ability to perceive fake news as accurate. Also, it discusses if positive and negative emotions such as happiness, depression, and anxiety can promote the tendency to spread disinformation amongst users.

4.1.1 Interconnection between Emotions and Disinformation

Several studies have found that people who critically engage in reasoning over the accuracy of the subject matter are less likely to fall for disinformation. However, emotions play a pivotal role in establishing belief over disinformation that can influence people's decision-making power. Research has shown how short-term moods that result in positive and negative emotions influence people's ability to detect deception in the information offered (Lane & DePaulo, 1999). While experiencing positive and negative emotions, humans are more likely to fall for disinformation as the ability to distinguish between the truth and deception falls considerably (Martel et al., 2020). Koch et al. have analyzed how a person's mood can act as a means for acceptance or rejection to a particular set of information. His research found that positive moods promote while negative moods reduce the tendency of an individual to rely on information while making judgments (Koch & Forgas, 2012).

Different negative moods, such as anger, affect cognitive processes and can extend to the extent that the physiological concomitants of emotional experiences can affect how information is processed while a person undergoes a particular set of negative emotions (Bodenhausen et al., 1994). Emotions such as anger and sadness can have completely different effects such as anger can stimulate reliance upon heuristic cues, whereas sadness can decrease reliance on the heuristic cues (Bodenhausen et al., 1994). Further studies on the relationship between emotion and gullibility have found that a negative mood state leads to an increase in skepticism, while a positive mood tends to increase gullibility, decreasing one's ability to detect deception (Forgas and East 2008; Forgas 2019). From a theoretical perspective, the classical reasoning theory states that when analytical thinking and reasoning are applied to a presented set of facts, this can aid in uncovering the truth from the contended subject matter (Bago et al., 2020). Therefore, individuals engaged in reasoning over the set of news are likely to identify the disinformation since disinformation often thrives when individuals are unable to apply critical reasoning and analytical thinking. Therefore the application of classical reasoning advocates analytical thinking while addressing the news rather than relying on intuition or gut feeling. (Evans, 2003; Stanovich & West, 2007). The theory has been tested on fake news, where engagement in analytical thinking with suspected beliefs patterns such as superstitious beliefs has resulted in a skeptic thought pattern on the research subjects (Pennycook et al., 2012). Moreover, analytical thinking has resulted in disassociation in fake news sources due to a break in trust patterns (Bago et al., 2020).

Therefore, when applied to disinformation and fake news perception, the classical reasoning theory states that heightened reliance on emotions lacks reasoning and analytical thinking (Pennycook & Rand, 2019b). That, in turn, accounts for failure in identifying disinformation and

fake news. Hence, over-reliance on emotions and intuition is the reason why persons fall into a category of susceptible to disinformation. This leads us to analyze its implications on disinformation that spreads across social media platforms and other web-based sources. Since emotional reasoning is often a sub-categorization of non-deliberative thinking, disinformation that spreads through social media platforms can often invoke emotional reasoning through the content. In turn, this factor leads to contributes to the virality of disinformation. Research further substantiates the current stance where a sentiment analysis of articles was conducted that were identified as a source of disinformation. It was revealed that the articles that intend to spread disinformation comprised of increased use of negative emotional language (Zollo et al., 2015; (Ecker et al., 2022; Horne & Adali, 2017; Martel et al., 2020; Paschen, 2020). Therefore, stimulating content can stimulate a biased thought pattern based on emotional reasoning rather than a critical thought pattern.

Hence, when presented with disinformation, one's negative emotions create a vigilance effect and accelerate one's skeptical processing when analyzing this set of information (references). This improves the assessment of the information, but it also promotes one's ability to detect lies.

4.1.2 Role of Positive and Negative Emotions

This part identifies how acceptance or non-acceptance of disinformation is influenced by the mood a person is experiencing, i.e. positive or negative mood. (Forgas, 1995). Conspiracy theories, alternative therapies, pyramid schemes, and miracle diets have been some topics that have continued to thrive since past decades, primarily through web sources. The previous part on, Interconnection between emotions and disinformation, has reflected on how emotions play a role while addressing the choices of disinformation's acceptance or non-acceptance. The human propensity to accept and believe information from other persons is essential for cultural evolution and social integration. However, the same can also result in the growth of gullibility and irrational beliefs (Gilbert, 1991; Harari, 2014).

Dr. Joseph Forgas, in his research, has reported that negative mood effects, such as anxiety, sadness, and disgust, when compared with positive mood effects, such as happiness and excitement, have resulted in reducing truth biases, confirmation biases, and improving the individual's ability to detect deception (Forgas, 2019). Specifically, negative moods promote a straightforward and data-driven processing style, leaving individuals more attentive towards their decisions and critical thinkers, while positive moods produce more theory-driven, creative, and heuristic thinking (Forgas 2013). Clinical work on "depressive realism has previously substantiated the same stance on negative mood." The studies have reflected that negative mood produces a better accuracy in the perception of reality (Alloy & Abramson, 1988). Theoretically, the claims are in harmony with Bless & Fiedler's (2006) assimilative/ accommodative processing dichotomy (Bless & Fiedler, 2006). The model suggests that a positive mood promotes while a negative mood reduces the tendency to rely on internal cues while making truth judgments (Bless & Fiedler, 2006). These results support our analysis that negative mood decreases the importance of internal fluency cues due to accommodative processing, associated with better focus on external stimulus features (Bless & Fiedler, 2006). In a study conducted, participants were made to watch mood-inducing films that induced happy and sad moods (Forgas, 1998). After that, participants watched the interrogation of either truthful or deceptive targets in denying a theft. It was found that negative mood resulted in the identification of more guilty judgments overall, but at the same time, it also improved the accurate detection of deception (Forgas, 1998).

However, it has been noticed that not every negative emotion has laid down a similar effect of increasing the accuracy of the brain over detecting deception. In their work, Michael Greenstein and Nancy Franklin have examined anger's impact on disinformation acceptance (Greenstein & Franklin, 2020). Anger is an emotion that has been characterized under the category of negative emotions that encompasses high arousal of emotions and induces critical reasoning (Carver & Harmon-Jones, 2009). It is an emotion that features approach-oriented characteristics, and it is adapted to guide behaviour with the influence of time pressure and situations. Studies suggest that anger enhances goal-relevant information processing that impacts a person's attention and memory (Bodenhausen et al., 1994). However, it has also been shown to increase reliance on cognitive processes, such as increased belief in stereotypes (Bodenhausen et al., 1994).

Though anger tends to reduce skepticism, which increases susceptibility to post-event disinformation, it also impairs an individual's ability to dismiss errors, affecting the accuracy of decision-making (Bodenhausen et al., 1994). Therefore, while affecting an individual's accuracy and memory to retain & process the information, anger increases the frequency of errors in judgment, promoting biased, intuitive, motivated reasoning (MacKuen et al., 2010). This tends to become a perfect platform for online disinformative content to stimulate anger amongst the population. Hence public anger can act as a vital tool amongst content creators to be stimulated through the online disinformative content, resulting in a viral sharing of the misinformative content amongst the population. Since it decreases the cogitative ability of critical reasoning amongst the individuals, it poses a severe problem of unintended sharing of disinformation. However, studies have shown that. However, studies have shown that anger plays a differential role when compared with other negative emotions, such as disgust and sadness, that improve the accuracy of judgment. (Alloy & Abramson, 1988).

Also, apart from anger as an exceptional negative emotion, positive mood feelings in totality have been evaluated as stimulating factors toward belief in disinformation (Bless & Fiedler, 2006; Fiedler, 2001). Bless and Fiedler's accommodative processing model suggests that positive emotions signal a benign environment, where people “impose internal structures on the external world,” while negative emotions signal a problematic situation, where people “modify internal structures in accordance with external constraints” (Bless & Fiedler, 2006, p. 66). Several experiments have substantiated that positive mood increases while negative mood decreases the tendency to rely on perceived beliefs. (Bodenhausen, Sheppar, & Kramer, 1994; Forgas, 1998; 2011; Forgas, & Denson, 2008). Affect-cognition research indicates that negative emotions create a cautious, pessimistic and critical interpersonal thinking pattern because of the selective priming and effect of negative information in memory. At the same time, a positive mood, on the other hand, creates a confident and optimistic thinking pattern that reduces levels of suspiciousness (Forgas, 1999, 2002). Hence, positive emotions may not always be desirable, primarily when we evaluate them in terms of their effect on deceptive situations. People in a good mood have been prone to errors (Forgas et al., 2005) and are less effective in detecting deception (Bless & Fiedler, 2006), while negative emotions offer opposite results.

4.1.3 Association of Depression and Disinformation Beliefs

During the past 24 months, vaccine disinformation has been one of the most pressing issues globally that have led to a division amongst people based on different beliefs, which has led to threatening the effectiveness of measures aimed toward bringing an end to the pandemic. Similarly, other prevailing narratives, such as 5G cellular technology being a reason for the growth of the virus, are other examples of disinformation (Vincent, 2020). While some researchers have

maintained that disinformation and conspiracy theories are mundane (digital) artifacts or fringe phenomena and lay down a negligible impact on the real world, the above-stated examples during the COVID-19 pandemic demonstrate the opposite results (Vincent, 2020). Studies have indicated that exposure to traditional media sources was negatively associated with depression, and exposure to digital media was positively associated (De Coninck et al., 2021)

In numerous studies, it has been reported that a continuous sense of uncertainty is related to increased levels of psychological distress (Barzilay et al., 2020; Salari et al., 2020). During the COVID-19 pandemic, one-quarter of the adult population in the US reported moderate or greater depressive symptoms (Qiu et al., 2020). Cognitive models of depression posit that depressed individuals exhibit negative biases during information processing, including memory perception, interpretation and attention (Mathews & MacLeod 2005). Study data published in JAMA Network Open reported that the population with depressive symptoms had a 2.2 times more likelihood of falling into the prey of disinformation and endorsing the same further (Perlis et al., 2022).

On theoretical grounds, Beck's cognitive theory of depression, negative thoughts generated by dysfunctional beliefs are one of the primary causes of chronic depression (Beck, 1967). Hence more negative thoughts an individual experiences, the more depressed one becomes. Also, cognitive theoretical models suggest that negative schemata catalyze the need for an excess of information based on any threatful or stressful situation to gain control over the situation (De Coninck et al., 2021). However, the researchers have stressed that the same practice has resulted in a backfire due to the wide circulation of disinformation and conspiracy theories available on web-based sources (Taylor et al., 2020). Hence the feeling of depression can act as a stimulator toward increasing beliefs in the disinformation and conspiracy theories. Further, research suggested that one of the primary motivations amongst the patients with depression for being associated with disinformation is to gain relief from the discomforting situation (De Coninck et al., 2021). Therefore disinformation and fabricated reports can exacerbate symptoms of depression amongst the population (Zhou et al., 2020).

5. Sociotechnical and Psychological Enablers

The traditional news and media spread was a systematic, unidirectional process in which news professionals wrote, edited, and published information for consumers (Turk, 2018). Contrastingly, Web 2.0 has transformed this process into an iterative one, where news can be altered and consumed in many forms (Akers et al., 2018; Turk, 2018). The alterations, false amplification, and false impressions of support are reinforced and spread by spambots and misguided individuals at incredibly fast rates (Akers et al., 2018; Au et al., 2021; Chew & Eysenbach, 2010). Additionally, the increased interconnectedness of civil society online, combined with the ease of sharing one's thoughts and opinions creates an environment of information overload making it hard for humans to identify and differentiate real news from fake (Gradoń et al., 2021; Mayorga et al., 2020; Rubin, 2019). This review seeks to analyze the most prominent enablers of misinformation and disinformation in the literature.

5.1 Cognitive Bias

Cognitive biases weighed heavily in the research as a strong human-centered enabler for the spread of mis- and dis- information. The areas of focus for enablers tended to be more about individual cognition and less about the influence of social upbringing (Carmi et al., 2020). Cognitive biases refer to deviations from rational judgement (Au, 2021). Researchers agree that susceptibility

to fake news is partially due to being prone to deception, humans' inclination to trust, or *truth-bias*; as well as cognitive biases (Hinsley & Holton, 2021; Kumar & Shah, 2018; Rubin, 2019). These could also include information avoidance or overgeneralizations in understanding (Au et al., 2021; Froehlich, 2017). These biases are further reinforced when the spreaders of misinformation are within our trusted networks of family and friends who are inherently believed to be more credible and reliable (Horne, 2021; Mehta et al., 2021). As like-minded people continue to share similar information within the same social sphere, this creates an echo-chamber effect, further upholding previously held biases and beliefs (Horne, 2021; Kapantai et al., 2021; Kumar & Shah, 2018; Mayorga et al., 2020). At this point, those receiving news online through social media (a whopping 67%) become incredibly susceptible to repeated exposure to news that aligns with their belief, further polarizing them (Buchanan, 2020; Rubin, 2019; Wardle & Derakhshan, 2017). While many researchers observed the effects of cognitive biases on the consumption and dissemination of digital mis-/dis-information, few noted behaviour from the lens of social belonging.

5.2 Confirmation Bias

A prominent theme in the literature when it came to believing and spreading misinformation and disinformation was confirmation bias (Chadwick & Stanyer, 2022; Cooke, 2017; Hinsley & Holton, 2021; Horne, 2021; Mayorga et al., 2020). Confirmation biases are often shaped by one's social identity, partisanship, and values (Digital Democracy Project, 2020). Literature suggests that individuals' confirmation bias can have effects on their judgments about source credibility, affect level of information acceptance, make them overconfident in their own assessments (Hinsley & Holton, 2021; Horne, 2021; Mayorga et al., 2020). Additionally, trusting misinformation even when contradicting evidence exists, is associated with lower critical thinking skills and increased conspiracy theory ideation and skepticism of media credibility (Hinsley & Holton, 2021; Mayorga et al., 2020; Pierre, 2020). Several studies have noted the increasing hold on one's own beliefs despite being presented with belief-correcting information (Cook et al., 2017; Rubin, 2019).

5.3 Motivated Reasoning

Motivated reasoning is considered an enabler due to its effects on news consumers' susceptibility in believing misinformation, particularly in its effects on politics (Allcott & Gentzkow, 2017; Chadwick & Stanyer, 2022; Walter et al., 2020); as well as a lack of motivation in seeking information that counters their belief (Kahne & Bowyer, 2017; Kunda et al., 1990; Mayorga et al., 2020; Walter et al., 2020). Kahan (2017) found that motivated reasoning was linked to ideological polarization through a study on social issues such as climate change and gun violence. In the context of misinformation, Pennycook and Rand (2019) use this theory to further analyze the effect of motivated reasoning vs classical reasoning (defined as a lack of analytic reasoning) on susceptibility to fake news. Pennycook and Rand concluded that when it comes to believing misinformation, it was due to a lack of reasoning rather than motivated reasoning. In terms of motivated reasoning when it comes to fact checking, Walter et al (2020) found that motivation to fact-check information that aligned with individual's beliefs was higher than fact-checking information that did not align with their beliefs.

6. Sociological Enablers

6.1 Social Belonging

Humans are social beings by nature and story telling was one way to pass on information, knowledge, and entertainment. Polletta and Callahan (2017) observe the significance of storytelling as a mechanism for social belonging in the Trump era. From sociologist Arlie Hochschild's perspective, part of Trump's appeal to his followers was the deep promise he cast with the narrative to "Make America Great Again"(Polletta & Callahan, 2017). Troubling economic prospects due to automation and outsourcing were happening at the same time as the influx of immigrants and refugees. "[Trump] refused to empathize with immigrants and poor people, while casting white working men as the victims of liberal's misplaced sympathies (Polletta & Callahan, 2017). This created a relatable story to these folks who belonged to social groups where their understandings were shaped and reinforced by online media, stories told by friends, and experiences of others, regardless of how factual these stories were. Polletta and Callahan argue that the root cause of the dissemination of fake news is not only because of biased media outlets, but of how stories from multiple sources within our social group are integrated and perceived.

Social identity theory is briefly referenced in some of the literature review as it provides an explanation for the belief and dissemination of fake news, regardless of facticity (Chen et al., 2021; Levak, 2020; Mayorga et al., 2020; Shu et al., 2017; Zhou & Zafarani, 2018). Social identity theory leverages one's need for belonging, social acceptance, has the ability to create a collective identity, foster relationships, and build bonds among individuals (Levak, 2020; Mayorga et al., 2020; Polletta & Callahan, 2017; Shu et al., 2017). Some sociologists note that people's sense of personal experiences encompass experiences that are not their own and instead are based on the experiences of people in their social group in order to reinforce the group's values (Prins et al, 2013, Bonilla-Silva, 2004, as cited in Polletta & Callahan, 2017). This is congruent with the theory of "identity-protective cognition" that focuses on how people process information in ways that helps them "signify their loyalty to important affinity groups" (Kahan, 2013 as cited in Chadwick & Stanyer, 2022). The exchange of stories and rumors is thus argued to be an important element in solidifying membership to a group and increasing one's trust and value in that group, regardless of the quality of information exchanged (Buchanan, 2020; Chadwick & Stanyer, 2022; Polletta & Callahan, 2017; Tandoc et al., 2018). One's sense of security and belonging to a group shapes how news and information is perceived and processed, and how closely individuals adhere to social rules, as well as how strongly we believe in narratives that align with our own (Abdalla et al., 2021; Polletta & Callahan, 2017; Tandoc et al., 2018). The role of social belonging and its significance to the dissemination of fake news was overlooked in most of the literature sought out in this review.

6.2 Journalism and Media Mistrust

Much of the literature touched on the role of journalists in the spread of digital mis/dis-information.(Bastos & Farkas, 2019; Lewandowsky et al., 2017; Marwick & Lewis, 2017; Wardle & Derakhshan, 2017; Zannettou et al., 2019) The literature acknowledges that much of the information spread by journalists could be categorized as misinformation, that is, information that is unintentionally biased or wrong, some even referring to it as *misreporting* or *yellow-journalism* (Molina et al., 2021; Shu et al., 2017; Tandoc et al., 2018). Some authors are also clear to distinguish "journalists" from "critics" or "commentators", citing that political satire, which is sometimes misinterpreted based on cultural differences, language differences or political

partisanship, should not be confused for disinformation (Ansar & Goswami, 2021; Bondielli & Marcelloni, 2019; Molina et al., 2021). This is especially important to highlight for teenage demographics who may rely on satirical news programs to stay informed and simultaneously entertained (Mayorga et al., 2020).

Another aspect of the distrust in media is due to the general public's inability to differentiate between opinions authored by commentators from real, objective news (Molina, 2021; Polletta and Callahan, 2017). Polletta and Callahan (2017) mention the role of government in the deregulation of media outlets, monopolization of media ownership, as well as the elimination of the "balanced programming" requirement under the Fairness Doctrine in 1987 that allowed for the emergence of hybrid news and entertainment programming. This gave networks an incentive to produce "provocative and deliberately questionable content" in order to increase their popularity as much as possible. The expectation of constant news and updates from an urgency-oriented society exasperates the need for authors to come up with frequent, new, engaging content for profit, making "publish first, correct if necessary" the new motto (Meel and Vishwakarma, 2020). These financially incentivised news outlets have been criticized by journalists as they reduce credibility of journalists and erode public trust in mainstream media.

In a report by Data & Society, authors Marwick and Lewis highlight the significance of biased reporting by *The New York Times* during the Bush administration to support the invasion of Iraq, citing the devastating impact this has had on public trust in the media. Additionally, the competitive environment created through the rise of "citizen journalism" has drastically increased the number of content creators online, blurring the lines between legitimate journalism that is governed by laws and regulations versus commentary uploaded by anyone with internet access (Giachanou et al., 2021; Kumar & Shah, 2018; Kumari et al., 2022; Molina et al., 2021; Tandoc et al., 2018). The seeming "realness" of fabricated news has also polluted online news feeds making it hard to identify the false news from the real news (Gradon et al., 2021; Meel and Vishwakarma, 2020)

Some of the literature that analyzes propaganda as a form of mis-/dis-information criticizes the role of government institutions in media mistrust. The denouncement of journalism as part of some disinformation campaigns illustrates this example (Akers et al, 2018; Quandt et al, 2019). Quandt et al (2019) underscore the historical role of the press during times of war to push "strategic control over the morale of the people...[and] spread demoralizing messages to enemy nations in the form of propaganda. These strategies continue to be implemented today which has created a general element of journalist skepticism from the public as well as a questioning of journalist's motivations (Hinsley & Holton, 2021; Meel & Vishwakarma, 2020; Quandt et al., 2019).

6.3 Institutional Mistrust

The full-text review found that many researchers have attributed early cases of disinformation to political motives, using information warfare and propaganda, to sway political and ideological opinion or cause confusion (Bastos & Farkas, 2019; Bran et al., 2021; Hamilton et al., 2021; McLane, 2021; Posetti & Matthews, 2018). The deliberate spread of false information by political groups and politicians has resulted in the erosion of public trust in democratic institutions (Bran et al., 2021; Kahne & Bowyer, 2017; Wardle & Derakhshan, 2017; Zannettou et al., 2019). Trust is further undermined by "political cynicism": a belief that politicians are motivated by self-interest rather than integrity and public well-being (Balmas, 2014; Pierre, 2020). On a broader scale, the general sentiment of skepticism towards any information arising from societal or

governmental bodies could be labeled as “informational disbelief” (Brandtzaeg et al., 2018). The impact of trust in governments on personhood and social welfare especially is seldom mentioned in the literature surveyed. For example, knowing that trust in government is correlated with better health and less trust-related barrier issues is a far-reaching impact beyond the information epidemic (Melki et al., 2021).

Institutional distrust could also be explained in light of other social and economic factors such as democratic institutions’ failure to provide public services (Wardle & Derakhshan, 2017) and structural and systemic mistreatment of marginalized communities (Pierre, 2020; Songs et al., 2019; Wang et al., 2021). Frequent and chronic violations of trust interfere with the public’s ability to engage in politics and develop a positive perception of the legitimacy of democratic institutions (Kahne & Bowyer, 2017). Some authors have also noted the effect of political partisanship on trust citing that: individuals are more likely to trust the political party closest to them, especially at times of social tension thus making them more vulnerable to fake news (Levak, 2020; Lewandowsky et al., 2017; Tandoc et al., 2018). This year’s Edelman Trust Barometer highlights a global collapse of trust in democracies noting that less than 50% of the surveyed democracies’ populations trust institutions within their country (*Edelman Trust Barometer*, 2022).

7. Technological Enablers

7.1 Social Media

The role of social media accounts have been extensively studied in relation to their role in recent elections and political events, especially in its role in the dissemination of false information online (Allcott & Gentzkow, 2017; Lazer et al., 2018; Vosoughi et al., 2018; Wu et al., 2019). Recent scholarship has begun to understand the role of highly networked organization that uses “covertly funded think tanks, astroturf lobbyists, alternative media [such as] documentaries, podcasts, and sponsored social media accounts” to influence public discourse on key topics such as immigration, healthcare, and racial inequality (Chadwick & Stanyer, 2022, p. 32). While social media platforms were first considered to be an important transformation in the pursuit of news media decentralization, it has been overshadowed by its exploitation through ill intentioned political actors and even used as method of censorship and state control (Bastos & Farkas, 2019; Bradshaw & Howard, n.d.; Chadwick & Stanyer, 2022; Foster, 2022).

Researchers cite that almost 60% of the incoming traffic for false information was linked to Facebook and Twitter (Kumar & Shah, 2018). The platforms which social media sites are built on exasperate the spread of misinformation through their financially incentivized business models that reward higher reactions and engagements from their followers (Carmi et al., 2020; Choraś et al., 2021; Gradoń et al., 2021; Rubin, 2019). These engagements are incredibly easy to purchase and help boost the visibility and perceived authenticity of a post (Choraś et al., 2021; Domenico et al., 2021; Kumar & Shah, 2018). Through provocative, sensationalist, and emotional headlines, authors can attract readers and begin the spread of false information extremely quickly. Social media therefore accelerates the speed of spread based on amplification algorithms.

Vosoughi et al (2018) found that that on Twitter, fake news travels six times faster than truthful news and is 70% more likely to be retweeted regardless of the news topic. Kumar and Shah (2018) cite that there’s an average delay of 12 hours between the start of the spread and its debunking, and that even though 99% of hoaxes are detected within an hour, 1% of really well-written hoaxes can go undetected for over a year. Researchers frequently used theories of network

density, social influence and propagation theory, and the size of social circle all to frame the spread of digital mis-/dis- information (Chen et al., 2022; Gradoń et al., 2021). Researchers agree that online social networks play a vital role in some aspects of user susceptibility such as those influenced by the number of likes a post has, the influence of the person sharing it, their social proximity, and how often has it been repeated within their ecosystem (Gradoń et al., 2021; Horne, 2021; Kapantai et al., 2021; Kumar & Shah, 2018). Social media amplification thus plays into consumers' cognitive biases to influence and alter behavior.

Literature that explains the diffusion of information through social networks as well as literature that studies right-wing ideology explains the high inter-connectedness between the two (Davey et al., 2020; Marwick & Lewis, n.d.; Shu et al., 2017; Vosoughi et al., 2018; Wu et al., 2019). Hashtags, memes, Discord channels, and Reddit groups all have been demonstrated to have a role in the way users connect online to share information, exchange ideas, and stay up to date with relevant material. Due to the low barrier to entry, ease of information presentation, and echo chambers, social media had become increasingly democratized and easy to access, making it an ideal medium for the spread of false information (Akers et al., 2019; Domenico et al., 2021). With the added amplification of misinformation by influencers (both knowingly and unknowingly), false information is amplified at unprecedented rates to vulnerable users (Carmi et al., 2020; Domenico et al., 2021; Foster, 2022).

7.2 Platform Amplification

On one hand, the technological features of social media can drive amplification in local networks, repeating stories to users that are similar in tone and context (Humphrecht, 2019; Sample et al., 2020). On the other hand, human behaviour such as cognitive biases can cause individuals to share news without verifying it, fall for sensational headlines, and engage with groups that reinforce their beliefs. (Humphrecht, 2019). Researchers have studied the dissemination of false information across the web to different conclusions: some of the literature underscores the role of human behaviour in the amplification of mis-/dis-information (Pantazi et al., 2021; Pennycook & Rand, 2020) while others highlight the structural differences of how false information spreads across the internet (Juul & Ugander, 2021; Vosoughi et al., 2018). False information tends to be more emotionally driven with negative associations, and thus more triggering, contributing to its amplification by users and bots (Ghanem et al., 2020; Vosoughi et al., 2018).

7.3 Types of Content

The prevalence of different types of false content shares the commonality of being deceptive and confusing, aimed at undermining trust in public discourse and distracting citizens from critical issues (Vaccari & Chadwick, 2020). Content could be divided and studied under the various names that encompass false information such as propaganda, satire, hoaxes, etc., to understand the vectors of why each variation would succeed in influencing human behaviour (Hinsley, 2021; Tandoc et al., 2018; Wardle & Derakshan, 2017). Content with the highest engagement tends to use elements of persuasion, deception, and is aimed at influencing attitudes and behaviours (Balmas, n.d.; Chadwick & Stanyer, 2022; Foster, 2022). Scholars have also noted the emotional aspects of fake content depending on its type and why it may be more successful at spreading (Ghanem et al., 2020; Vosoughi et al., 2018). Studies in psycholinguistics as well as data and information science have tried to understand the patterns that allow for the spread of specific

content based on elements that contribute to its virality and infectiousness (Domenico et al., 2021; Juul & Ugander, 2021).

7.4 Information Vacuums

Emerging research in the spread of misinformation and disinformation identify data voids (alternatively known as information vacuums) as areas of exploitation by malicious actors with ideological, political, or economic agendas (Flores-Saviaga & Savage, 2019; Golebiewski & Boyd, 2018). Data voids are defined as search terms where there is limited or non-existent verifiable data available (Flores-Saviaga & Savage, 2019; Golebiewski & Boyd, 2018; Pierre, 2020). Information vacuums are filled with misinformation through differing strategies and often it is too late to correct. Flores-Saviaga & Savage (2019) note that data voids could emerge as a result under terms or searches that people do not usually search for; this is especially problematic in languages other than English where sufficient, credible information does not exist. Furthermore, it becomes difficult for a person to identify that there is harmful or false data when no other information exists to refute it (Flores-Saviaga & Savage, 2019).

Mare et al. (2019) and Moyo (2009) attribute the misinformation that emerges from data voids to citizen journalism; that is, citizen engagement in the production and dissemination of information due to a lack of sufficient news from accredited sources. Innes et al. (2021) explain one such example of the Manchester Arena terrorist attack where the Greater Manchester Police department communicated quickly via twitter that there was an ongoing investigation in the City Centre. However, hours later there was a failure to provide much more information, creating an information vacuum susceptible to the influence of other malicious actors. Yang & Lee (2020) in their examination of news coverage during the 2015 Middle East Respiratory Syndrome (MERS) outbreak in South Korea uncover the critical role of mass media in bridging communication gaps between authorities and the public during time of crises, as well as how the timely release of information and content can influence public perception and emotion.

7.5 Information Overload

The sheer volume of content that is created and spread on a daily basis has created an ecosystem of information pollution that interferes with individuals' decisions making and information processing abilities (Bran et al., 2021; Gradoń et al., 2021; Wardle & Derakhshan, 2017). Information overload enables the spread of misinformation by overwhelming cognitive thinking processes to the extent where individuals will overgeneralize knowledge and fail to think critically due to the distress this causes (Au et al., 2021; Bran et al., 2021; Cooke, 2017). This further results in believing information based on it being convincing enough without effort to verify or authenticate the source and accuracy of information (Au et al., 2021; Cooke, 2017; Gradoń et al., 2021; Rubin, 2019). Despite information overload being acknowledged in some of the literature, measurable factors were scant.

7.6 Targeted Ads

Advances in online marketing and advertising have allowed for user-specific targeting through algorithms that seek to further manipulate and polarize users (Akers et al., 2019; Domenico et al., 2021; Shu et al., 2017). Targeted advertising is one such method where private data is used to create online identities that further drive users towards polarization (Carmi et al., 2020; Foster, 2022). Researchers have criticized the algorithmic curation of content online that pushes users

further towards extreme content, particularly because users do not have control over the viewing experiences they have online (Akers et al., 2019; Carmi et al., 2020). This allows online sites to attract users and thus be able to promote products to them through targeted advertising (Akers et al., 2019; Domenico et al., 2021). The methodology of online advertising where click-through rates generate ad revenue incentivises the creators of disinformation and even news media to continue to attract users through the spread of harmful or biased information (Akers et al., 2019; Allcott & Gentzkow, 2017; Domenico et al., 2021). Click-bait is a term frequently used to explain the types of headlines used in digital media to reel users into sites that are riddled with advertising (Shu et al., 2017). Optimization of advertising utilized sophisticated and hidden methods of tracking such as embedded social media widgets, website analytics, and browser history tracking to correlate a user's website visits with purchases and interests (Akers et al., 2019; Domenico et al., 2021).

8. Disinformation Drivers

8.1 Commercialization and Financial Incentives

The potential for financial gain as a result of engagements and viewership was highlighted as a motive for spreading fake news (Allcott & Gentzkow, 2017; Au et al., 2021; Carmi et al., 2020; Figueira & Oliveira, 2017; Hinsley & Holton, 2021). Companies such as react365(dot)com are subsets of larger companies like Media Vibes SNC who own hundreds of URLs devoted to generating fake news (Figueira & Oliveira, 2017). Generating web traffic to fake news sites can earn profit through advertising by employing catchy headlines in today's attention economy (Allcott & Gentzkow, 2017; Carmi et al., 2020; Figueira & Oliveira, 2017; Pierre, 2020). Other financial incentives fall on mass media where journalists are expected to produce content that would generate profit for news media companies (Marwick & Lewis, 2017; Tandoc et al., 2018).

Commercialization, excessive product personalization, and consumer hyper-engagement in digital technologies has led to industry wide competition for consumer's attention via filter bubbles, echo chambers, and bias (Foster, 2022; Pace & Markannen, 2021; Petratos, 2021). While much of the literature focuses on political motivators that lead to the creation and distribution of misinformation, few have studied the economically motivated reasons (Domenico et al., 2021). In the context of commercialism, misinformation can be used for anything from generating ad revenue through catchy click-bait, destroying a company's brand through the spread of disinformation, or using cyberattacks to exploit system vulnerabilities for data theft through phishing and spam (Domenico et al., 2021; Humprecht, 2019; Petratos, 2021). The literature also highlights the shift in attitudes of enterprises towards prioritizing profits over civic responsibility (Marwick & Lewis, n.d.); while at the same time not incentivizing solutions (Sample et al., 2020).

The media economy is fueled by user-engagement driving big-tech, marketers, and even journalists to be a driving force behind the quickest creation and dissemination of misinformation (Rubin, 2019; Sample et al., 2020). Researchers explain how constant user engagement helps 'fuel recommendation algorithms to eventually influence our actions and behaviour, making users easy prey for advertisers and propagandists' (Foster, 2022, p. 4). Using the science of human deception and persuasion, disinformation is created trick consumers into purchasing products, decreasing credibility of competitors, and even manipulate stock market prices (Domenico et al., 2021; Foster, 2022; Stewart, n.d.). Furthermore, in the interest of big data and customization, the tech industry is creating a norm where individuals have to give up their privacy and data in exchange for convenience (Carmi et al., 2020; Sample et al., 2020).

8.2 Political Incentives

Political motivation has received notable attention from researchers across many disciplines (Allcott & Gentzkow, 2017; Balmas, n.d.; Freeze et al., 2021; Gorwa & Guilbeault, 2020; Vargo et al., 2018; Vosoughi et al., 2018). Many researchers agree that the intention behind political disinformation is to present events to fit a certain narrative that drives partisanship and sows divide between individuals, communities, and nations for the benefit of a few (Allcott & Gentzkow, 2017; Humprecht, 2019; Marwick & Lewis, n.d.; Vosoughi et al., 2018; Wardle, 2018). Propaganda and information warfare have historically been studied in the context of foreign operations (Allcott & Gentzkow, 2017; Bastos & Farkas, 2019; Shu et al., 2017); however, in recent examples researchers have underscored the rise of neoliberalism and populism as drivers of false information (Rogenhofer & Panievsky, 2020). These researchers argue that political mis-/disinformation could be explained as a tool in the playbook of populist leaders seeking to create narratives of national identity unity based on the ethnoreligious supremacy of one group, and the inferiority of another (Bastos & Farkas, 2019; Bradshaw & Howard, n.d.; Chadwick & Stanyer, 2022; Rogenhofer & Panievsky, 2020).

The success of polarizing political content is argued by some researchers as dependent on societal norms that influence cognitive biases and allow for more accepting attitudes towards information that aligns with their beliefs (Chadwick & Stanyer, 2022; Humprecht, 2019; Moravec et al., 2018). This is further confirmed by studies that aim to understand the types of political misinformation that is successful in differing democracies: for example, Humprecht (2019) examined Western democracies (US/UK/Germany/Austria), and found that in English-speaking countries, “disinformation attacks political actors, whereas in German-speaking countries, immigrants are most frequently targeted” (Humprecht, 2019, p. 12). Therefore, the type of disinformation that flourishes depends on societal knowledge and information environments. Moreover, a society’s level of knowledge plays a role in how they react when confronted with disinformation; those who are better informed being more critical and less affected (Foster, 2022; Humprecht, 2019; Marwick & Lewis, n.d.). Based on these findings, attention to deception and disinformation must go beyond individual variables, and connections to broader, systemic and institutional factors must be considered (Carmi et al., 2020; Chadwick & Stanyer, 2022).

9. Technology and Tools

To address the problem of digital disinformation, there are multiple approaches that can be taken. This section will explore at a high level, the various approaches to address disinformation, as well as a brief view of current initiatives undertaken by academia, social media platforms, organizations. Lastly, we will review digital tools created to verify information to assess their efficacy in addressing the various forms of disinformation.

9.1 Technical Approaches - Machine Learning

With the increasing use of technology in all areas of life, the domain of digital disinformation is no different. In as much as technology can be an enabler to the spread of disinformation, it can equally be a means of addressing it. One common approach being applied is Machine Learning (ML), which is a “technique that improves system performance by learning from

experience via computational methods [...] the main task of machine learning is to develop learning algorithms that build models from data [that can predict outcomes]” (Zhou, 2021, p.2).

There are several approaches to ML that are better equipped to perform certain functions relating to detection. Algorithms commonly discussed include Support Vector Machines (SVM), Long Short-Term Memory (LSTM), Neural Networks, and Naïve Bayes (Al Shariah, & Saudagar, 2019; Alenezi & Alqenaei, 2021; Anjali, Reshma & Lekshmy, 2019; Bharti & Jindal, 2020; Choudhary & Arora, 2021; De Oliveira, Medeiros & Mattos, 2020; Giachanou et al., 2021; Kaur, Sawinder, et al., 2020; Kim & Ko, 2021; Tan et al., 2020; Vinolin & Sucharitha, 2021). While algorithms often produce good results on their own, an effective approach is to combine methods and techniques to produce more promising results. A lot of the literature that we reviewed dealt with this subject and built upon existing experiments to further improve disinformation detection tools. The main ways that they did this was by combining multiple tools, to tackle the problem from multiple avenues (Aslam et al., 2021; Bharti & Jindal, 2020; Kim & Ko, 2021; Kolluri et al., 2021).

9.1.2 Textual Analysis

One common approach is to assess textual evidence to determine if it contains forms of disinformation (Bharti & Jindal, 2020; Choudhary & Arora, 2021; Giachanou et al., 2021; Yang et al., 2021). One such example is the research of Kim and Ko (2021) who determined that there were many internal and external characteristics to consider when trying to determine if a social media post contained misinformation. They found that trying to analyse external features can become costly, so instead they decided to analyse the text, rather than the metadata. They approached the problem of online misinformation by creating sentence relevancy scores before they ran the text through the ML algorithm. They used graphing techniques to assess the relevancy of each sentence in the post. In doing so, they had the algorithm focus on the most useful sentences to achieve a more targeted result. Their model achieves a better performance than baseline and comparison models.

Beyond topic analysis, a text-based author analysis can be performed. Based on their research, Aslam et al. (2021) determined that CNN and Deep Learning were the two best ML algorithms for detecting misinformation, as such they combined the two techniques to see whether they could achieve a more accurate result when used together. They used the bi-LTSM-GRU algorithm to analyse statement features, and the Deep Learning Dense Model to analyse other features, like speaker job title, subject, venue etc. Their approach outperformed other models, using the same dataset, by a large margin demonstrating that combining techniques is an effective approach to analysing social media posts for misinformation.

As will be later discussed, an approach commonly being used by major social media platforms is a human-machine approach. To assess its efficacy, researchers simulated this method by creating an open-source website called CoVerifi, where they used a two-step authentication process to tag posts as either real or fake. This website uses an existing ML algorithm capable of detecting whether a post was created by a human or a machine. Then, users are also asked to vote on whether the content appears credible. The interface then displays a user credibility rating and a ML credibility score. By incorporating two types of authentications this website provides multiple credibility metrics in one location (Kolluri et al., 2021).

9.1.3 Sentiment Analysis

Another avenue to determine whether content was likely misinformation was rather than relying on the words that are used, the emotional connection between the words could be explored.

Giachanou et al.'s (2021) research investigates the relationship between the intense emotional response people have to disinformation and the likelihood that the information presented is false. They hypothesized that disinformation would be written in a way that would be extremely emotional, as compared with genuine information. As such, their model, Emocred, assesses the emotional language in a post to help determine the veracity of the statement. It outperformed Deep Learning approaches to detecting misinformation, suggesting that emotional language is indicative of misinformation. In addition to these findings, a combined Neural Network and Deep Learning Model also assessed sentiment-based evidence, by determining whether there was a correlation between intense emotional language in the writing and misinformation. Combining multiple secondary features, including syntax, grammar, readability, character count, word count, word density, number of nouns, adjectives, verbs, pronouns, and more, the approach outperformed other models tested using the same dataset. Thus indicating that these outside features, including sentiment, were useful in detecting disinformation (Choudhary & Arora, 2021).

9.1.4 Metadata Analysis

To assist in locating potential sources of disinformation, such as coordinated inauthentic behaviours, metadata can be explored to determine the originating location of the author or information being shared. To explain, we use the example of one study where the posts on the popular Chinese micro-blogging site Sina Weibo were used to train a classifier to automatically detect rumours. It was determined that by adding information about client-based features, such as whether the post was made from a tablet or a phone, and location-based features, whether the post is about something that occurred within China or outside, they improved the algorithm's ability to detect misinformation and rumours (Yang et al, 2021).

Considering the substantial number of metadata stored on each social media post, these findings suggest that more researchers should consider reviewing data outside the normal scope of the text, and how these factors affect the likelihood that a statement is true or false. For example, Bharti and Jindal (2020) made extensive use of the available attributes on Twitter. They analysed length, Tweet URLs, number of hashtags used, tweet-count and user specific features, whether the account was confirmed or not, number of followers, account activation days, and status number. By incorporating all these features, they were able to create a highly accurate model for detecting statement credibility. These outcomes suggest that metadata is an important tool for detecting misinformation and should be considered by ML approaches to the problem of misinformation. It is important that researchers consider more than the text itself, as the metadata can provide insightful clues into the validity of information.

9.1.5 Image and Video Analysis

One aspect of countering disinformation that is at risk of being overlooked is the fact that disinformation takes many forms. A lot of the research around countering disinformation using ML centers on detecting cues in text. However, two forms that are worth considering are forged photographs and deepfakes, or forged videos. Forged images and videos pose harm to society and governments, just as much or more than written words (AlShariah & Khader, 2019). This is because many consumers may not realize that these mediums can be altered convincingly. As a result, it is important to consider and combat the different mediums in which disinformation can spread. This section will consider those two types of disinformation and discuss ML tools to counter them.

9.1.6 Forged Images

To deceive others, some engage in forging images to present an event that did not actually occur. While this is a common problem, with fake images shared around social media platforms, AlShariah & Khader (2019) found that the research around detecting fake images is lacking. As a potential solution, they tested three different algorithms, CNN, Alexnet Network, and Transfer Learning using Alexnet, to determine which was best for detecting altered images. They found that Alexnet performed the best, with an accuracy rating of 93.4, thereby providing a potential means of detecting forged images in the future. In addition to the Alexnet Network, Vinolin and Sucharitha (2021), set out to create a new algorithm to detect fake images. Rather than testing existing algorithms with this new problem, or dataset, they created a new forgery detection approach called “Taylor-adaptive rag-bull rider (RR) optimization algorithm-based deep convolutional neural network”, or Taylor-RNet for short. Their method is specially adapted to fight a common type of forged image called image splicing. Image splicing is when you photoshop someone's face on to someone else's body. As you can imagine, in a political context for example, this can be highly problematic. As such, their method used light coefficients to detect whether the image had been altered. They achieved an accuracy rating of 96.921. Their research indicates that by targeting specific types of forgery algorithms can accurately detect forged images.

9.1.7 Deepfakes

Due to the widely shared deepfake videos of world presidents, deepfakes have been highlighted as an area of concern (Kaur et al., 2020). Deepfakes are AI generated forged videos or images created using Neural Networks to create convincing face swaps. Again, it is easy to see how this kind of alteration could be harmful to society and very convincing to a public who does not fully understand the technology that makes these videos possible. As such, Kaur et al. (2020) developed a dataset of Deepfake videos that they created using the DeepFaceLab tool downloaded from GitHub.com. They then tested a tool that they created, which exploits the unique characteristics of deepfake videos, being image rescaling, shear mapping, zooming augmentation, and horizontal flipping, to detect whether these videos are likely forged. Their model achieved an accuracy rating of 98.21, thus demonstrating that ML tailored to address specific types of disinformation can achieve remarkable results.

9.2 Industry Approaches -Social Media Platforms

In addition to the work researchers have undertaken to explore solutions to digital disinformation, social media corporations are also heavily engaging in research and methods to address disinformation. Particularly due to the large public backlash after stories of disinformation being spread during election campaigns and during the early days of the covid-19 pandemic, companies are scrambling to minimize the spread on their platforms. Albeit a brief look at current initiatives, this section explores several of the larger platforms, their policies and approaches undertaken.

9.2.1 Meta

Meta, who includes Facebook, Instagram, and WhatsApp has often been criticized in the media for their role in allowing for disinformation to spread, particularly on Facebook. In recent years they have taken a more proactive approach to addressing disinformation and preventing its spread.

To identify disinformation, Meta follows a machine-human approach where technology is used to identify potential disinformation using similarity and keyword detection (Meta, 2021), and human users can also provide feedback through built in mechanisms. Human fact checkers are then tasked with making final decisions. For this, they rely on third-party fact checkers to verify content in posts (Meta, n.d.). Fact-checkers must be certified through the International Fact-Checking Network (IFCN) and must adhere to an established Code of Principles. With a focus on viral information, as of June 1, 2021, Meta works with more than 80 IFCN-certified organizations, addressing content in 60 languages (Meta, 2021).

The first step in Meta's approach is to completely remove prohibited content addressed in their Community Standards and Ads policies. This includes any content related to "...hate speech, fake accounts, and terrorist content..." (Meta, 2021, paragraph 5). The second step to problematic content is to reduce accessibility (Meta, 2021) by removing it from the 'explore' and 'hashtag' pages, as well as reducing visibility in news feeds and stories (Instagram, n.d.). Whereby they balance between free speech and reducing access to fact-checker identified misinformation. Accounts that have repeatedly posted false information may be prevented from being tagged in posts, have their visibility reduced or disabled (Instagram, n.d.). Lastly, Meta provides warnings, in which users are informed of the fact-checkers findings and have the opportunity to either continue accessing the flagged content or go back. In addition, anyone who previously shared the content is informed of its rating, and AI is used to label duplicate content (Meta, 2021).

Specific to election periods and political campaigns, Meta takes a more specific approach, with three identified categories of needs. The first, is to fight foreign interference, in which policies were updated and politicians accounts are protected via Facebook Protect. One integral role in fighting foreign interference is to address coordinated inauthentic behaviour, which is when "groups of pages or people work together to mislead others about who they are or what they are doing" (Gleicher, 2018). Groups are identified and taken down primarily based on their behaviours, such as representing themselves as operating from one part of the world, when they are actually operating from another. Those engaging in groups of coordinated inauthentic behaviour, can be foreign governments, as well as non-state actors such as domestic groups, and groups formed for financial gain (Gleicher, 2018). The second need is to increase transparency, in which owners of a page or account go through a process to confirm their identity, state-controlled media is labeled, and candidate ad spending is visible to all. Lastly, to reduce misinformation within the elections, fact-checking labels are used, paid ads urging people not to vote are banned and media literacy projects for middle and high school were established (Rosen et al., 2019).

Potentially due to their criticism in the media, Meta has provided the most transparency and detailed information on the methods being used. Once content has been labelled as disinformation by the independent fact checkers, an image matching model, SimSearchNet++ is used to search for similar content. Using self-supervised learning, it can find the same or similar content, i.e., the same terms but with different imagery or different variations of the terms. When similar content is identified, it is labelled with the same warnings as the original content (Meta AI, 2020).

In anticipation of an expansion of deepfake videos being created for the purpose of disinforming society, Meta is in the process of creating a deepfake detection model using neural networks to identify videos that have been altered (Meta AI, 2020).

The approaches taken by Meta are generally a positive step in the right direction, however, one limitation is that the same approaches are not extended to politicians, as Meta argues that it

may limit the “voice” of less known politicians or advocacy groups (Constine, 2019). Thus, some politicians and groups may still retain a green pass for spreading disinformation.

Twitter

Twitter’s approach to disinformation is to consider the propensity for harm for three types of misleading content. The first, misleading information, is defined by Twitter as “statements or assertions that have been confirmed to be false or mislead by subject-matter experts, such as public health authorities” (Roth & Pickles, 2020, paragraph 6). The second type, disputed claims, are defined as “statements or assertions in which the accuracy, truthfulness, or credibility of the claim is contested or unknown” (Roth & Pickles, 2020, paragraph 6). The third type, unverified claims are, “information (which could be true or false that is unconfirmed at the time it is shared” (Roth & Pickles, 2020, paragraph 6). Depending on the type of misleading content and the propensity for harm, i.e., moderate, or severe, content posted on Twitter could be either have no action for unverified claims, to labels for moderate misleading information and disputed claims, to warning labels for severe risk due to disputed claims and complete removal for severe misleading information (Roth & Pickles, 2020). If an account has advanced or continuously shares harmful misleading narratives that violate the synthetic and manipulated media policy, Twitter may temporarily reduce the visibility of the account or lock or suspend the account.

Considering Twitter has become a breaking news source, particularly in times of crisis, with individuals posting updates as they happen. Given the potential for disinformation to be spread during this time, Twitter unveiled that they would pay particular attention during this time, making every effort to consult journalists, conflict monitoring groups, humanitarian organizations and independent investigators to establish the credibility of information being posted. Similar to their counterparts, Twitter will then apply forms of warnings and limit the opportunity to share tweets found to contain inaccuracies (Roth, 2022).

To help identify disinformation, while still in the pilot stage, Twitter has also established a community-driven pilot program, ‘Birdwatch.’ Where users can identify misleading content within tweets and post a note providing context to support their claims (Twitter, n.d.-a). Currently the notes are only accessible in a separate website, and not visible within Twitter itself. Once the content has been rated as misleading and the notes have been rated as helpful by the birdwatch community members, it will be shown on a blurred Tweet (Twitter, n.d.-a). To rank the notes, matrix factorization is used. This approach was first established during a 2006 Netflix recommender system competition; however, the approach was modified to calculate a global score, rather than as a recommender system (Twitter, n.d.-b). Twitter is testing certain new ways in some areas around the globe to enable its users to actively participate in misleading info reporting, including Misleading Info Reporting Flow to allow users to report tweets containing disinformation. During important events (e.g. COVID-19 pandemic, elections), Twitter proactively feature informative messages or updates to counter misleading narratives that emerge (Roth, 2022).

9.2.2 Youtube

In their policies, Youtube approaches disinformation from the perspective of the amount of harm it may cause others. Based on this, Youtube has established community guidelines wherein the types of content and links to external sources that are not allowed are detailed, which includes, “...promoting harmful remedies or treatments, certain types of technically manipulated content, or content interfering with democratic processes (Youtube, n.d.-c). Those found to be in violation will

have their content removed, after three ‘strikes,’ the channel violating the terms will be deleted (Youtube, n.d.-a). Videos that have been determined to be close to, but not in direct limitation of the prohibited content will not be included in the options for recommended videos. In addition, content that has been fact-checked will also include a form of informative context to provide more accurate information (Youtube, n.d.-b) . If users consider a video to be in violation of Youtube’s community guidelines, they may report it for Youtube’s review. In addition, channel owners may delete any violating material posted in comments made to their posted material (Youtube, n.d.-d). While Youtube has not provided detailed technical information on their approaches to reviewing content, however, it is known that in addition to human verification, they are using machine learning to automate the review of uploaded videos and the removal of those that are detected to contain prohibited content (Youtube, n.d.-e). Their algorithms also flag content for human verification, in which the results from their decisions are used to train the models for improved future detection (Youtube, n.d.-b).

9.2.3 TikTok

One of the most famous social media platforms today, TikTok has followed a similar approach as the other platforms in their implementation of community guidelines, which details actions and content prohibited. In addition to the prohibition of content that misleads people on topics related to health, elections and civic matters, and general disinformation campaigns, they have also clearly outlined the ban on “synthetic or manipulated content” referring to shallow and deepfakes (Pappas, 2020). Further, coordinated inauthentic behaviour has been explicitly called out to prevent “...exert influence and (to) sway public opinion while misleading individuals, our community or the larger public about the account’s identity, location, or purpose” (TikTok, n.d., section 9). Similar to Youtube, other than stating that they use technology and human verification, TikTok does not provide details related to the technical approaches to addressing disinformation.

9.2.4 Warning Styles

Contextual warnings placed near the information that is being warned against. Can be some form of an indicator, “congenial styling,” or a pop up, which users particularly have become accustomed to ignore (Kaiser, Mayer, et al., 2021). This type of warning has proven to be ineffective, with many users either not seeing or choosing to ignore the warnings (Kaiser, Wei, et al., 2021). An alternative approach is through the use of interstitial warnings, in which a users browsing would be interrupted with messaging describing why the warning is occurring and the associated risk in proceeding to access the flagged content (Kaiser, Wei, et al., 2021). Research has revealed that interstitial warnings are a much more effective approach, in that users are more likely to notice the interstitial style warnings and then choose to either not review the flagged content or performed a search to verify the content flagged (Kaiser, Wei, et al., 2021).

Social media companies have tested forms of interstitial warnings, using full page warnings (O’Sullivan, 2021), and facebook, Instagram have blurred out images and videos of content that has been deemed either fully or partially false by independent fact-checkers. Users are given an option to either see why it was labeled as false information, in which they are taken to the report provided by the fact checker, or to proceed to view the flagged content (Rosen et al., 2019). Similarly, if users try to share content that has been found to be false, they will receive any advisory message explaining why it is false and informing users that if they chose to share the content anyways, it will include the label that the information is false (Rosen et al., 2019). A similar approach is being followed by Twitter, who places warning messages on content that it considers to be harmful or misleading. Users then have the option to be directed to a Twitter page that

provides additional reasoning as to why it was flagged, or to an external, trusted source (Roth & Pickles, 2020).

As a limitation, embedded tweets and users viewing tweets without being logged into a Twitter account will not receive warning messaging (Roth & Pickles, 2020).

9.2.5 Investments in Media Literacy

Signalling the importance of moving beyond fact checking and warnings, many of the social media platforms have indicated that they are engaging in some forms of investment in digital and media literacy. Meta has invested in multiple media literacy programs for general digital literacy, to middle and high school curriculum, first time voters and seniors (Meta, 2020). The literacy programs focus on teaching fact-checking skills, how to find reliable information, as well as integrity in creating and sharing information (Meta, 2020). Along similar lines, Google and Youtube have invested in organizations and individual media literacy programs for children, parents, and educators (Youtube, n.d.-b). To show their support, Twitter supports ‘media literacy week,’ (Gamboa & Cartes, 2017) which takes place yearly in October, is a time when schools across the globe focus on engaging in forms of educational methods to educate students on safely engaging with digital media (UNESCO, n.d.).

9.3 Online Tools

9.3.1 Education and Training

According to the RAND Corporation, who have developed a database of tools that combat disinformation, tools classified as education and training are interactive games, courses, or activities aimed to teach individuals new skills in combatting disinformation (RAND Corporation, 2019). The assumption for these programs would be that they would aim to shift individuals from a heuristic processing approach, which essentially means they are automatically processing information, to more systematic processing approach, which focuses on the attempt to thoroughly understand information through careful attention, deep thinking, and intensive reasoning, (Chaiken & Ledgerwood, 2012).

Education and training evidently have great potential in aiding in the fight against disinformation as they can attack the problem on more of a root level. The more individuals in our society who are equipped to identify disinformation, the more it will hypothetically get reported. The other main benefit of education and training as it relates to fighting disinformation is the ease of implementation for youth in primary, secondary, or even post-secondary education. Finding relevant information online in a scholarly context is something that most students learn about at some point, and it would be an easy transition to a course or training that helps them identify disinformation in other contexts such as health or politics. Overall, education and training can be a strong solution in combatting the characteristics of velocity and variety of disinformation. Velocity due to the increase in informed individuals who will not re-post disinformation, and variety as more people will be informed in identifying disinformation in multiple contexts and with tools other than just the headline.

9.3.2 Human Fact-Checking

Building off education and training, these tools are generally websites that have a team of educated journalists or fact-checkers who either fact-check statements, posts, or articles that website users submit to them, or aim to designate information that is trending heavily at that time as fact or fiction (Zhang & Ghorbani, 2020).

Because of the general nature of news being very mixed and with a variety of different statements of information, sometimes a binary true or false system cannot provide an accurate answer regarding the validity of something you may read online (Zhang & Ghorbani, 2020). Thus, the main capability comes with the fact that there are actual human experts performing the proper and relevant research to determine whether a claim is factual. Due to having a human element, these sites are best equipped to deal with the characteristics of the variety and the veracity of disinformation. For example, an article headline may be true, but the content may be riddled with incorrect statistics. A machine learning algorithm may classify this article as disinformation, but a human could identify the fact versus fiction within the article. Another capability of these fact-checking websites is that they hold a much lower level of bias than any individual. In an American-focused study, Grady et al. (2021) found that individuals were less likely to validate information that aligned with their supported political platform, and vice-versa. Even when individuals were forced to fact-check something aligning with their political affiliation, they were skeptical of the 'false' claim.

9.3.3 Credibility Scoring Tools

Credibility scoring tools are very similar to fact checking websites; however, they are focused on the source of the disinformation, and not the claim itself. While most human based fact-checking tools are websites, these tools can not only be websites, but many are plug-ins, browser extensions, or applications. Some use human scorers, others use machine learning, some use crowdsourcing, while most use a combination of these three approaches.

Credibility scoring tools are interesting when contrasted with fact-checking websites, as they deal with their shortcomings but also have some of their own. The large benefit here is that it solves the issue of time latency in a sense. Because they are evaluating the whole source of the information, it's almost as if they can take a bulk approach to identifying disinformation; As opposed to having to evaluate each individual claim, the user can know if the website tends to spread disinformation and find a new source. Due to their automation, these types of tools combat the volume and velocity characteristics of disinformation.

9.3.4 Bot Detection and Tracking

These two tools are very similar and therefore have been grouped together. Bot detection consists of tools intended to find automated accounts on social media (which tend to be key spreaders of disinformation) while disinformation tracking consists of tools that track the spread of a specific piece of disinformation or report on disinformation statistics on an application or website (RAND Corporation, 2019).

Almost entirely contrasting credibility scoring tools, many bot detection and disinformation tracking tools set their focus on social media sites such as twitter, where disinformation spread is plentiful. Another major capability that separates these tools from others

mentioned is that their use of machine learning and/or artificial intelligence means that users get almost instant feedback into their searches. Therefore, these tools perform well in the volume and velocity of disinformation due to their automated nature.

9.4 Functionality of Existing Tools

As seen, there are numerous types of disinformation, as well as numerous tools that have been created to address the varying forms. While theoretical discussions provide a reassuring nod that the various forms of disinformation are being considered and addressed, however, to ensure that they are adequately being addressed should also be studied.

In considering this aspect, we performed an empirical analysis of some existing tools to determine whether all forms of digital disinformation are being addressed. To do so, we selected tools from RAND Corporation’s list of tools (RAND Corporation, 2019). To focus on tools that would immediately verify information for users, we selected tools that provided human fact-checking, credibility scoring, or bot detection.

Table 1. Types of Disinformation Addressed by Current Tools

Tool Name	Falsified/Fake News	Sensationalist News/Clickbait/Tabloid	Biased News	Pseudoscience	Satire/Parody/Irony	Hoax/Rumor	Imposter Content/Bots	Spam	Phishing	Conspiracy	Deep Fakes
Bot Sentinel							1				
Botometer							1				
BotSlayer	1						1				
Captain Fact	1										
ClaimBuster	1		1								
Climate Feedback	1										
CrossCheck	1	1	1							1	
Digital Polarization Initiative	1										
Dirt Protocol	1										
Disinformation Index	1										
Duke Videofactchecking Tool	1										
Emergent.Info						1				1	
Factcheck.org	1		1								
Factchecking	1		1								
Forensically Image Verification Tool											1
Glorious Contextubot	1										
Hamilton 2.0	1		1								
Hoaxy (Observatory on Social Media)	1						1				
Iffy Quotient	1		1								

KnowNews	1		1			1				1	
Lead Stories FactChecker	1					1				1	
MediaBias Ratings	1		1		1					1	
Misinformation Detector	1										
NewsCheck Trust Index	1		1								
Newstrition	1			1	1						
OpenSources	1	1	1	1	1					1	
Our.News	1		1		1						
PolitiFact	1										
Polygraph (BBG)	1										
Public Editor	1		1								
Reveal Image Verification Assistant											1
Reveal Image Verification Assistant	1										
Share the Facts Widget	1										
Snopes	1	1									
The Factual	1									1	
Trive Verify	1										
Trusted Times	1		1								
Verification Tool	1										
Video Verification Plugin (InVid)	1										
YouTube Data Viewer - Citizen Evidence Lab	1										
Total	35	3	13	2	4	3	4	0	0	7	2

As can be seen in *Table 1*, almost all of the tools are created to address falsified, or fake information. This is encouraging in that individuals' have multiple tools they can consult to verify the credibility of information they may have come across. However, there are multiple areas in which it may not be possible to verify certain types of disinformation. For example, *Table 1* shows that as of the time that the RAND Corporation published their list, there were no tools identified to address known spam and phishing attempts. Such tools may assist recipients in verifying whether the email or other form of information received is actually spam or a phishing attempt. Especially those with limited technical experience may not be able to recognize the difference between a legitimate email or a nefarious one.

In addition, we consider the low number of tools available to address sensationalist news, pseudoscience, hoaxes, and conspiracy theories. While the number of tools are low, given that falsified information is often used within these types of content, it could be possible that the tools created to address falsified information may address some sensationalist, pseudoscience and conspiracy theories, thus, additional research should be performed to ensure that these types of disinformation are adequately being addressed.

Lastly, it can be seen that there are only four tools dedicated to unveiling imposters/bots. With the increasing amount of cyber bullying and disinformation being spread, further resources should be allocated to clarifying online identities. This would help to know whether the profiles a user is engaging with is a legitimate source, and whether they may be a source of negative influence or bias.

10. Empirical Review

10.1 Research Design

When creating products or services for public benefit, the goal of a developer is for their product or service to be used, and the extent to which someone enjoys, or benefits in using a product or service impacts the likelihood of ongoing use. Therefore, a main requirement is to create an appealing environment for users, wherein particular attention is given to positively engage their feelings and emotions (Chen et al., 2018). In addition to making a product or service enjoyable to use, developers must ensure that the product works, fulfills users' need, and it is intuitive, meaning, easy to understand or manipulate.

In doing so, they assess the user experience (UX), which considers the users' behaviours and sentiments while using a product or service (Law et al., 2009). Specifically, as per the ISO Standard definition, user experience can include all emotions, beliefs, benefits, physical and psychological responses, behaviours, and performances before, during, and after using it, thus considering "...that user experience is influenced by the system structure, the user and context of use" (ISO 9241-210, 2019, p.43).

To assess the user experience and the extent to which the product or service is enjoyable to use, fulfills users' needs and is intuitive, a usability evaluation should be undertaken to assess users' experiences and identify any areas of weakness. We note that while usability is typically considered a subset of UX, it accounts for the "holistic experience" (Rosenzweig, 2015, pg. 7) of the user. Therefore, considering that "evaluating usability and creating usable objects are all part of the development of a successful UX" (Rosenzweig, 2015a, pg. 116), usability and UX should be considered in unison. To assess usability, five essential areas should be considered. They include the extent to which the product or service is easy to learn, how efficient it is, the occurrences of errors, the extent to which it is memorable, and the user is satisfied while engaged with it (Nielsen, 1993).

As an overall design to usability research, multiple approaches can be taken, including a focus on qualitative or quantitative forms of data collection. Alternatively, a mixed methods approach can be used whereby qualitative and quantitative data can be analyzed together, in triangulation, which may add to the reliability of the findings (Jacko, 2012; Lazar et al., 2017). Moreover, "this combination can provide rich data that can identify the big picture issues, patterns, and more detailed findings for specific issues" (Rosenzweig, 2015, pg. 145). Thus, as per the benefits of triangulating data from both quantitative and qualitative approaches, a mixed method design will be used within the evaluation.

10.2 Methods

Usability evaluations can be performed via multiple methods of assessment, particularly when a mixed method research design is undertaken. Opportunities range from inspection methods where quick expert overviews are undertaken to promptly identify glaring issues, to more intensive

empirical evaluations wherein users' experiences are explored while engaged with a product or service (Rosenzweig, 2015a).

For our usability evaluation, in line with our mixed method approach, we will incorporate multiple methods of evaluation. Our rationale for using multiple methods of data collection stems from reports that some forms of testing can discover usability problems not discovered through other types of user testing and vice versa (Desurvire, 1994). Thus, to gain a complete view of the usability and overall user experience in using the digital disinformation countering tools, a combination of approaches will be used.

10.3 Inspection Methods

Heuristic Evaluations

Also completed by experts, heuristic evaluations assess the components of the product or service and rate it against best practices to produce a rating and make recommendations (Rosenzweig, 2015a). While not an in-depth form of evaluation, the usability heuristics provide a quick assessment of an interface's usability. To conduct the heuristic evaluations, the research team members individually assessed each tool by using usability heuristics established by an industry expert (Nielsen, 1994).

10.4 Empirical Methods

User-Based Testing

Within the user-based testing phase of our study, we conducted a task-based, moderated, remote usability evaluation to learn more about how users interact with selected tools and any challenges pertaining to their use. A task-based approach was used to assess how a user would interact with the digital disinformation countering tools using a real-life situation (Rosenzweig, 2015b). In doing so, participants were requested to perform tasks representing scenarios they may typically engage in if trying to validate information read or received online.

Moderated usability testing was used as having a moderator present allows for direct observation of user experiences, take notes, and assist if task explanations are needed (Rosenzweig, 2015b). Furthermore, to ensure that users' perspectives are well understood, the moderators may ask clarifying questions. Remote testing was chosen as it facilitated the sessions without the need for participants to leave their homes (Jacko, 2012; Lewis & Sauro, 2021), and enabled the sessions to be easily recorded for data analysis.

Within the evaluations, a think aloud protocol was followed as it required participants to verbalize their thoughts while completing the tasks. Gathering participants' verbal feedback provides their opinions of the digital disinformation countering tools, and clarifies any frustrations or confusions experienced (Lewis & Sauro, 2021; Riihiahho, 2017).

Summative (Survey) Evaluation

In addition to the remote usability evaluation, a survey instrument was utilized to gather data from end-users about their overall perceptions and experience using the tools. The System Usability Scale (SUS) was selected due to its popularity in UX research as a quick approach to usability testing. It provides a comprehensive set of questions to assess participants' opinions on the usability of the item being assessed (Brooke, 1996).

10.5 Participants

As digital disinformation countering tools are created for the general population, we wished to explore whether the experiences were consistent or varied across different user types. Therefore, to establish a representative set of participants, we divided our target population into three categories of technical capability, thus known as ‘beginner,’ ‘intermediate,’ and ‘advanced’ user personas.

To determine a representative number of participants, we referred to the opinion that for qualitative evaluations, five people can find most of the usability issues (Nielsen & Landauer, 1993), however, this number can go up to eight (Rosenzweig, 2015b). As the evaluation has three personas in this research, considering the opinions cited, our aim was to have a minimum of five participants for each persona, for a minimum total of fifteen participants. As researchers have recommended up to nine testers, up to twenty-seven participants would be accepted. For the quantitative, or summative (survey) evaluations, it is recommended to have a minimum of 20 participants to have a statistically representative sample (Rosenzweig, 2015b).

A call for participation was shared with peers and publicly posted on networking and social media platforms to solicit participants. Those interested completed an online form with their name, email, and their self-identified level of technical competence.

To select participants from those who expressed interest, purposive sampling was used. Such a sampling approach is used when participants are required for a specific objective (Tashakkori & Teddlie, 2003, pp. 713). As this research design called for a specified number of participants with varying levels of technical capability, we wanted to ensure that we fulfilled the minimum number required for each persona. Moreover, to seek heterogeneity in the final sample and maximize the generalizability of the results, we wished to select participants of varying age, and education levels (Lewis & Sauro, 2021).

As a compensation, a \$40 participant-selected Amazon or Starbucks gift card was given to each participant.

10.6 Data Collection and Analysis

Heuristic Evaluations

The evaluations were based on an established set of usability heuristics (Nielsen, 1994) and can be found within the completed evaluations in Appendix D to Appendix H. Each team member (n=4) assessed the individual digital disinformation countering tool against the established heuristic and provided a rating and recommendations. The individual findings of each team member were then compared to ascertain the severity of issues and provided useful insights for user-experience problems with the tools. In cases where there were disagreements in the assessed severity of each category, the research team members revisited the tool to reach a mutually agreed upon severity rating. In cases that could not be agreed upon, the mean severity rating was taken as the final rating. Given the recommendation to have three to four evaluators to increase the reliability of the mean severity rating (Nielsen, 1994), our use of four evaluators meets this recommendation.

User-Based Testing

Within the task-based usability evaluation, seven tasks were formulated for participants to complete in a moderated (n=2) usability test to assess users experience in carrying out functions that could be performed in fact checking situations. As it was a remote usability evaluation, the platform used was '[Loop 11](#),' which enables users screen and voice to be recorded while participating in the evaluation. Additional data, including but not limited to the task completion time, heat maps and questionnaires can also be collected via the platform.

A combination of self-guided and scenario-based usability evaluation procedures were used to uncover potential issues within each tool. Scenarios were defined based on the most important and representative task (Dumas & Fox, 2009; Jacko, 2012; Riihiaho, 2017), which is information credibility checking. During the evaluations, after answering several demographic questions, participants were directed to three different online articles and were asked to express their opinion about the credibility of the articles. The same articles were later used to verify the source and/or the article claims with five disinformation countering tools.

Prior to conducting the usability evaluation with the main participants, the two moderators created the evaluation script to be used during the evaluations. A pilot study was then carried out with four participants to identify any deficiencies with the scenarios and questions. Modifications were then made according to the feedback.

To analyze the task-based usability evaluation data, a content analysis approach was utilized. Content analysis involves the analysis of data to identify ideas, patterns or trends, for the purpose of establishing an overall understanding of the phenomenon (Krippendorff, 2012, 2018). Specifically, a deductive, content analysis approach was taken. In line with the deductive approach, our goal in analyzing the usability evaluations was to assess experiences and usability issues within the digital disinformation countering tools, rather than creating a new overarching theory (Silver & Lewins, 2017). To create the code book for deductive content analysis, the moderators watched the video recordings to note specific incidents or comments made by participants, alongside a review of the evaluation transcripts. The notes and comments were then used to categorize common experiences and opinions, to create 'codes,' which were then assembled into a code book. The researchers then applied a 'human-driven' coding approach whereby the data was read and coded according to the pre-defined code labels (Silver & Lewins, 2017). Inductive coding was also used to a limited extent, to identify any experiences or opinions that were not previously included in the code book.

Summative Evaluations

After testing each tool, several open-ended questions and a shortened version of the SUS questionnaire were posed to obtain an understanding of the participants' overall user experience and specific challenges that stood out. As part of the user testing methods available in Loop11, the questionnaire was completed directly in the software. To reduce the length of the usability evaluation and the cognitive burden of completing multiple tasks and questions, only five of the ten SUS questions were included. Questions were selected based on their relevance to the tools being assessed and the objectives of our study.

To calculate the score, Loop11 provides an overall calculation based on the feedback provided by all participants for all tools. Thus, to calculate the scores of each individual tool, a

programmed Excel spreadsheet was used (Excel Spreadsheet for Calculating SUS Scores, 2008) To interpret the score in a form more easily understood, we used adjective based scoring to assess the calculated results (Bangor et al., 2009).

10.7 Scenarios and Digital Disinformation Authenticating

Tools

The articles, shown in *Table 2* for fact checking were selected from diverse forms of disinformation that individuals could be subjected to in their daily lives. Examples include articles discussing pseudoscience, conspiracy theory, and satire.

Table 2. Articles to Authenticate

Articles to Authenticate	
Article	Disinformation Category
Chinese Herb Kills Cancer Cells	Pseudoscience
Biggest Lie in World History: There Never Was A Pandemic.	Conspiracy
Hairdresser Arrested for Making Voodoo Dolls from Customer' Hair	Satire

Based on the results from the scan and review of currently available tools to address digital disinformation, we selected five tools, two websites and three extensions that provide disinformation-countering functionality to perform a formal usability evaluation for each of the selected tools. The selected tools for the usability evaluation include:

Websites

‘MediaBias/FactCheck’: Website for fact and source checking.

‘isthiscredible’: Website for rating news articles.

Extensions

TheNewsroom: browser extension that provides information on the trustworthiness of news.

Newstrition: browser extension that provides information on the source of online information.

MediaBias/FactCheck: browser extension for the MediaBias/FactCheck website.

11. Results

11.1 Heuristic Evaluation

As previously mentioned, the heuristic evaluations were individually completed by four project team members. Results were then compared and discussed to reach a mutual consensus. In cases where an agreement could not be made, the mean of the scores were calculated to provide for the final score.

Table 3 presents a comparison of the scores for the websites included in our study. Scores range from zero, where there are no issues, to four which represents a severe issue. As seen, while some categories do not have vastly different scores, ‘isthiscredible’ clearly presents with less

severity ratings, thus suggesting it may be a more user-friendly tool. Given the potential for an appreciable experience, ‘isthiscredible’ may be a better option for disinformation-countering. The full evaluation, along with issues and recommendations are included in Appendix D (MediaBias/FactCheck Website) and Appendix E.

Table 3. Results of the Heuristic Evaluations - Websites

Results of Heuristic Evaluation - Websites		
Heuristic	MediaBias/FactCheck Website	isthiscredible
Visibility of system status	2	0
Match between system and the real world	4	3
User control and freedom	2	1
Consistency and standards	3	3
Error Prevention	0	0
Recognition rather than recall	3	0
Flexibility and efficiency of use	0	0
Aesthetic and minimalist design	4	0
Help users recognize, diagnose, and recover from errors	0	0
Help and documentation	4	4

The project team members followed the same approach for the browser extensions, with results highlighted in *Table 4*. As with the websites, in many of the heuristic categories, the severity ratings did not deviate greatly. This may suggest that many of the issues experienced are consistent across disinformation-countering tools, and not just isolated incidents. Considering the importance of these tools, and the need for increased use to circumvent the harms of disinformation, this may be a serious hinderance to their uptake. The full evaluation, along with issues and recommendations are included in Appendix F (The Newsroom Beta), Appendix G (Newstrition), and Appendix H (MediaBias/FactCheck Extension).

Table 4. Results of the Heuristic Evaluations - Browser Extensions

Results of Heuristic Evaluation – Browser Extensions			
Heuristic	The Newsroom	Newstrition	MediaBias/FactCheck Extension
Visibility of system status	4	3	3
Match between system and the real world	2	3	2
User control and freedom	0	2	0
Consistency and standards	0	4	2
Error Prevention	2	3	2
Recognition rather than recall	3	3	2
Flexibility and efficiency of use	0	0	0
Aesthetic and minimalist design	0	0	3

Help users recognize, diagnose, and recover from errors	0	3	2
Help and documentation	3	4	1

11.2 Moderated Usability Evaluation

Participants

The final number of participants who took part in the usability evaluation is 18. Of this number, according to self-identification, they were classified into three different technical capability personas. Five were ‘beginners,’ eight were ‘intermediate,’ and five were ‘advanced.’

Table 5. Participant Demographics

ID	Age	Completed Education	Technical Capability	Digital Literacy	Technology Adoption
P3	18-24	High School	Advanced	Producer	Before most of my colleagues/friends/family
P2	25-34	Undergraduate	Beginner	Consumer	When it becomes mainstream
P4	35-44	Graduate	Advanced	Consumer	When it becomes mainstream
P5	18-24	College	Intermediate	Activist	After most of my colleagues/friends
P6	25-34	Graduate	Intermediate	Consumer	When it becomes mainstream
P7	35-44	Graduate	Intermediate	Consumer	After most of my colleagues/friends
P9	18-24	High School	Advanced	Activist	Before most of my colleagues/friends/family
P10	18-24	High School	Intermediate	Consumer	Before most of my colleagues/friends/family
P11	25-34	Undergraduate	Intermediate	Consumer, Producer	When it becomes mainstream
P12	45-54	College	Beginner	Consumer	After most of my colleagues/friends
P14	18-24	Undergraduate	Advanced	Producer	Before most of my colleagues/friends/family
P15	35-44	Undergraduate	Beginner	Consumer	After most of my colleagues/friends
P16	18-24	High School	Intermediate	Producer	After most of my colleagues/friends
P17	25-34	Graduate	Intermediate	Consumer, Producer, Activist	After most of my colleagues/friends
P18	25-34	Graduate	Beginner	Consumer	Before most of my colleagues/friends/family
P21	18-24	College	Beginner	Consumer	After most of my colleagues/friends
P23	25-34	Undergraduate	Beginner	Consumer	Before most of my colleagues/friends/family
P24	18-24	High School	Intermediate	Producer	After most of my colleagues/friends

Task 1: Locating Tools to Authenticate Digital Disinformation

Participants were given ten minutes to search for a disinformation countering tool, such as a website, application or browser extension that can be used to verify information. The purpose of this scenario was to determine how easy it is to locate digital disinformation countering tools, solicit opinions on the capabilities offered, and how easy it is to use them.

Table 6. Digital Disinformation Tool Search Experiences

Experience	Beginner		Intermediate		Advanced	
	Frequency	%	Frequency	%	Frequency	%
Did not find a tool within 10 minutes	3/5	60	4/8	50	0/5	-
Didn't know what to search for	4	80	7	87.5	1	20
Didn't know what to look for in search results	1	20	4	50	-	-
Found a tool	2	40	4	50	5	100
Found a tool, didn't understand how it works	1	20	3	37.5	2	40
Found a tool, could not get any results	1	20	-	-	2	40
Found a tool, understood results	1	20	1	12.5	2	40
Prefer Google to authenticate information	1	20	1	12.5	1	20

Based on the highlighted results of Task 1, in Table 6, participants who self-identified as either ‘beginners’ or ‘intermediate’ users of information technology were more likely to not know what to search for (80% and 87.5% respectively) and thus were not able to find a tool within the ten minutes allocated (60% and 50% respectively). It is noted that half of the ‘intermediate’ participants were able to find a tool, however, only one participant was able to use it and understand the results (12.5%).

Examples of feedback received were *“it wasn't easy to find a tool. Multiple websites came back as the results with lists of tools. I had to read and go through them and then chose one randomly”* [P18], *“I don't know what to look for because I don't have enough baseline knowledge”* [P15]. In addition, *“the tools are difficult to find because it's hard to define what I am looking for...”* [P7], *“the tool I search would depend on the type of news”* [P17].

While it may appear that the ‘advanced’ participants were successful in being able to locate a tool, however, only two out of the five participants (40%) were able to use it and understand the results, thus suggesting that while the advanced participants may be better positioned to perform searches, however, in regard to applied tool use and in understanding output, they do not fare much better than those who identify as either ‘beginner’ or ‘intermediate’ users.

As [P18] had pointed out, the search results were expected to include actual tools to explore, however, 11 out of 18 participants first received results related to articles and blog posts discussing tools and used that to select a tool to explore. On this note, [P16] mentioned that *“when*

trying to find a tool it was difficult to gauge if a tool was accurate since results included business pages that recommended the tools, and it wasn't clear if they had any affiliation or bias. It was also hard to figure out if the tool was really related to what I was looking for.” The topic of affiliation or bias is particularly noted as this was a re-occurring comment by some participants throughout the usability evaluation and will be further discussed.

Key Take-Aways:

Given participant experiences in their search for a digital disinformation countering tool, it is apparent that most of the beginner and intermediate users did not know what to search for to find a tool. While the advanced participants fared better in terms of formulating a search query, when it came to using the tool and understanding the results, their experiences were similar to the beginner and intermediate participants. With some participants commenting on needing to know in which context would the tool be used, it may suggest that participants would feel more confident in the search for a tool if they know exactly what they will verify, rather than searching for a disinformation-countering tool in general.

We noted that 11 out of 18 participants received initial results related to blogs and articles discussing tools, rather than direct results for the tools themselves. This may suggest that the search engine optimization (SEO) of the websites hosting the tools may be less optimal than the websites containing articles. Going through lists after the initial search is an added step in the process, thus creating a potential barrier to quickly finding a suitable tool.

Task 2: Article Credibility

Prior to using the tools, the participants were asked to review three different articles to provide their opinion on whether the website itself looked credible and whether the claims the article was making appeared to be credible. In doing so, the moderators wished to assess the participants' level of awareness and identification of forms of digital disinformation without assistance.

Table 7 provides a summary of each persona’s opinion on whether the articles were considered ‘not credible,’ ‘not sure’ (if the article was credible), and ‘not credible.’ Each article will be further discussed to understand which factors influenced the participants' conclusions.

Table 7. Participant Opinions on Article Credibility

		Beginner			Intermediate			Advanced		
		Not credible	Not sure	Credible	Not credible	Not sure	Credible	Not credible	Not sure	Credible
1	Article	80%	20%	-	87.5%	-	12.5%	100%	-	-
	Website	100%	-	-	100%	-	-	100%	-	-
2	Article	60%	-	40%	100%	-	-	100%	-	-
	Website	40%	20%	40%	65.5%	-	37.5%	80%	-	20%
3	Article	80%	-	20%	50%	37.5%	12.5%	40%	40%	20%
	Website	80%	-	20%	75%	-	25%	100%	-	-

Article 1: Chinese Herb Kills Cancer Cells

As can be seen in Table 8, all participants immediately recognized that the website is not a credible source, with most commenting on the multitude of advertisements and pop-ups on the website: “especially with the saggy jowls, how to heal nerve damage ads, it makes it look like the website isn't legit” [P12]. Particularly, “with ads popping up and asking me to enter my information,

it does not look like a scientific website” [P9]. Moreover, one commented *“it has ads for other things, it looks like it is trying to sell stuff”* [P15]. Comments went even further with [P16] commenting on the store and movie *“the website has articles but also a lot of ads and ways to monetize the website. It looks like they are really focused on how to monetize the website”* [P16]. [P5] had similar thoughts and said, *“it looks like David Wolfe is trying to sell a 10-step style program”*, while *“it looks snake oily”* was [P10]’s immediate instinct.

Table 8. Participant Opinion on Article 1 Website Credibility

Article 1 - Website Credibility			
	Beginner	Intermediate	Advanced
Website Not Credible	100%	100%	100%
Looks like they want to sell something	20%	62.5%	-
Too many ads/Pop ups	60%	50%	80%
Looks like a blog	60%	25%	60%
Website design/layout doesn't look professional	-	25%	40%
Unsure if Credible	-	-	-
Website Credible	-	-	-

In addition to the feedback of too many ads, many participants also highlighted that *“the way the website is set up, it seems more like a blog than an academic source”* [P9], it *“looks like a health blog, it’s nothing formal”* [P2]. [P11] also concluded that *“it looks like a blog because I don’t see contact info for where its coming from”* [P11], insinuating that if a website does not include a physical address, it appears less professional.

Further to providing their opinions on the website itself, participants also provided their opinion on the credibility of the article claims. Results are highlighted in Table 9. As shown, only one participant believed the article claims, which they decided after reviewing a ‘source’ provided at the end of the article. We note that the ‘source’ was not referenced within the text, rather, it was placed at the end of the article. They deemed the source (National Library of Medicine) to be reputable, saying *“reputable organizations are less likely to put out fake news”* [P11], therefore, by association, they considered the claims the article was making to be credible as well.

The participants who immediately doubted the claims commented, *“the topic is rough on common sense. It isn’t something that I can believe in”* [P7]. Further, they thought that the arguments being made were broad, misleading, and without sufficient academic support, as [P14] pointed out *“they link to a study, but the study isn’t easily accessible since they might just be trying to link to something. If they use “studies say” but do not really give much more information, I am skeptical.”* [P16] also mentioned, *“it kind of cites a journal but does not have a lot of quantifiable statistics or anything that would solidify their argument”* and while [P17] was reviewing the article’s sources, they commented *“it’s a government site which is good...there is nothing about cancer. So, if this herb can kill cancer, then the government page should say something about that, but it doesn’t even say anything about cancer.”*

Table 9. Participant Opinion on Article 1 Article Content Credibility

Article 1 - Article Credibility			
	Beginner	Intermediate	Advanced
Article Credible	-	12.5%	-
Initially thought it was credible	-	12.5%	-
Use of references and/or figures	-	-	-
Author name available	-	-	-
Article Not Credible	80%	87.5%	100%
Not enough content/supporting arguments	60%	25%	20%
Arguments too broad/misleading	60%	50%	20%
Other related articles not credible, makes the article appear not credible	40%	-	-
No direct links (citations) to review their references	-	-	20%
Supporting photos/references not credible looking	20%	25%	20%
Author name/credentials missing	40%	37.5%	20%
If this topic discussed on an academic website/mainstream news, it could then be considered credible.	20%	12.5%	40%
Unsure if Article Content Credible	20%	-	-

An interesting observation is these concerns were more frequently mentioned by the beginner and intermediate participants, rather than the advanced participants. However, the advanced participants more frequently mentioned that they would believe this claim if they were to read about it from an established hospital, an academic source or in the mainstream news media. *“If it was something as big as this, I would want to see something from large hospitals or universities”* [P24] and *“since it’s a popular topic, it should be everywhere if it would be an actual thing. Their claims are substantial”* [P21]. To summarize many participants' sentiment upon reading the article, [P17] put it well when they said, *“I think it’s a personal opinion, it doesn’t mean its correct.”*

Article 2: Covid is a Lie

The article with claims that ‘covid is a lie’ comes from Global Research – The Centre for Research on Globalization, who write opinion pieces on world events, from what is represented as a ‘research’ perspective.

While reviewing the first article, most participants quickly deemed it not credible, however, opinions on the second article were not as conclusive. Multiple participants either initially considered the website credible and later decided against its credibility, or they maintained its credibility throughout the article review. As demonstrated in *Table 10*, the beginner and advanced groups show the most indecisiveness where an equal number of participants thought the website was credible versus not credible. While 80% of advanced participants ultimately deemed the

website not credible, the same participants had initially thought it to be credible when first reviewing the website.

Reasons that lead them to consider its credibility include that *“the website looks more professional, the way it is set up”* [P9], and *“off the bat, this website seems more academic because of the lack of ads everywhere and having people being quoted”* [P14] and, *“the website looks more like what you would see for professional resources, national publications, PubMed”* [P21]. Thus, *“the website seems more trustable”* [P4] as *“the design gives off element of trust”* [P3]. To explain how it translates to trust, we quote [P16] who said *“if the layout is nice and has professional branding and consistency, people would be more willing to think it is accurate. Whether or not it is actually accurate or not is different but professional sites seem more trustworthy”* (we note that [P16] did not consider the website professional). The name of the source also appears to reflect on the credibility, with [P5] commenting, *“Global Research - fits the part,”* this sentiment was also shared by [P6] and [P15].

Another element deeming it more trustworthy is the presence of an author’s name and numbers, *“it actually has an author. It looks more convincing than the previous one. There is more information. I can see numbers. Typically, we associate numbers with science, fact, and reality”* [P24]. Along similar lines, [P3] stated, *“if information is nicely presented, it would seem more credible.”* However, they acknowledged that this is a downside as should something be nicely presented then they might believe it, regardless of its accuracy.

Those who maintained its credibility did so due to *“...how many shares it has says that it is credible. I can trust this source because it is shared by many people”* [P18]. Moreover, [P11] stated that *“the website looks credible because the author claims to be an award-winning professor at the University of Ottawa, and I would doubt that an award-winning professor would jeopardize his career unless he was sure about what he was doing.”* [P23] also recognized that it was a professor who published the article, and when they also realized it was possible to download the article as a PDF, they said *“I would definitely trust the source.”*

Table 10. Participant Opinion on Article 2 Website Credibility

	Beginner	Intermediate	Advanced
Website Credible	40%	37.5%	20%
Discussing scientific topics	-	-	-
Professional Domain name	20%	37.5%	-
Website design/layout looks professional	20%	25%	60%
Ability to download/share content /leave comments	40%	-	
Initially thought to be Credible	-	12.5%	80%
Website Not Credible	40%	62.5%	80%
Looks like they want to sell something	-	37.5%	-
Too many ads/Pop ups	-	-	-
Looks like a blog	-	-	-
Website design/layout doesn't look professional	-	25%	20%
Unsure if Website is Credible	20%	-	20%

Those who considered the website not credible found the opposite of their peers in thinking that the website did not have a professional appearance and said *“...can tell that it is a propaganda website based on 9/11 having its own tab”* [P2]. Based on a review of feedback, it appears that

participants concluded that the website was not credible after deciding that the article claims were not credible. Therefore, it appears that the participants were judging credibility not just based on layout and design, but also that of the claims, which then reflected on the website credibility.

Table 11. Participant Opinion on Article 2 Article Content Credibility

	Beginner	Intermediate	Advanced
Article Credible	40%	-	-
Use of references and/or figures	40%	-	-
Author name available	40%	-	-
Article Not Credible	60%	100%	100%
Not enough content/supporting arguments	20%	25%	60%
Arguments too broad/misleading	60%	62.5%	-
Appears to be an opinion, not factual	-	62.5%	40%
Other related articles not credible, makes the article appear not credible	-	-	-
No direct links (citations) to review their references	20%	12.5%	20%
Supporting photos/references not credible looking	-	-	40%
Author name/credentials missing	-	-	-
If this topic discussed on an academic website/mainstream news, it could then be considered credible.	-	12.5%	-
Unsure if Article Content Credible	-	-	-

For those who found the article to not be credible, the biggest complaint that beginner and intermediate participants had was related to the arguments being too broad or misleading (*Table 11*), with particular complaints of the language being too extreme. Specifically, with comments such as, “*the claim itself echoes propaganda. Most news sites, if they are legitimate, it doesn't go off propaganda and sensational sentences. The ‘diabolical agenda’ is too aggressive*” [P5]. Similar sentiments were expressed by [P16] who said, “*this website seems to use more pathos and buzzwords for the impact of Covid-19...but ‘economic-warfare’ and ‘war against humanity’ is extreme wording...I guess before they bring in facts, they are bringing in a certain perspective but without tangible facts or statistics to back it up...it just seems like they are writing an opinion article and they are angry...it's definitely biased - they don't separate the facts from opinions...it's very extreme and a click bait because they don't elaborate more on their claims.*”

While reviewing this article, multiple participants commented on the issue of confirmation bias, with [P17] providing the most thorough opinion on the possibility, “*there is confirmation bias, it depends on your outlook. If you already believe in misinformation then you will believe what you are reading but I already know that the news is a lie, so I will think it's wrong.*” Such sentiment was shared by [P6] and [P11] who both stated that their experience tells them that covid is not a lie, thereby leading them to believe that the article claims are not credible. To such effect, [P14] commented “*I do think that people's biases will affect how much they are willing to accept the information*” and [P16] said “*the claim they are making is catered towards people who are looking for evidence or opinion stating that it is a lie.*” While [P3] acknowledged the issue of filter bubbles, they “*think it would be ridiculous if anyone would believe this news.*”

Article 3: Hairdresser Makes Voodoo Dolls

The last article, which is taken from a satirical ‘news’ website, was chosen to assess whether participants can distinguish between fictional satire and factual news stories.

While *Table 12* shows that the majority of participants believed the website itself to not be credible, a small number of beginner and intermediate participants thought that the website could either have some credibility or were unsure. Some participants noted that the domain name ‘world news daily report’ was a usual news name [P18], [P2] and the ability to share the story enhanced its credibility [P18]. This was corroborated by [P24] who thought that it “...looks like a normal news story.” [P5] mentioned “*the website looks credible, but it isn't something I would completely believe.*”

Table 12. Participant Opinion on Article 3 Website Credibility

	Beginner	Intermediate	Advanced
Website Credible	20%	25%	-
Professional Domain name	40%	-	-
Website design/layout looks professional	20%	12.5%	20%
Website Not Credible	80%	75%	100%
Looks like they want to sell something	-	12.5%	-
Too many ads/Pop ups	40%	50%	40%
Looks like a gossip magazine	40%	37.5%	20%
Website design/layout doesn't look professional	-	-	40%
Unsure if Website is Credible	-	-	-

Reasons for thinking the website is not credible were similar to what was expressed while reviewing article one. Many spoke of the advertisements, “*initial thoughts are that the page is a bunch of ads for unrelated articles*” [P16] and “*so many ads, I don't like to read it. It kills the credibility*” [P17]. With one participant recognizing that they “*have seen some of the click bait ads on other websites. I am always curious but don't want to click on it*” [P12], thus signaling experiences with advertisements across websites is not uncommon, however, the participant did not trust it enough to click on it.

Further to the presence of ads, multiple participants reflected on the type of website it appears to be, commenting “*this looks like a Hollywood star kind of website*” [P7]. Other terms used to describe the website include “*gossip magazine*” [P2], “*sensational news*” [P15], “*tabloid*” [P11], for “*entertainment*” [P10], and “*like the daily mail*” [P3], with ‘the daily mail’ being a self-described “tabloid-format” news source. Interestingly, no one suggested that it could be satire.

To re-address [P5]’s belief in the article’s claims, as shown in *Table 13*, many participants (40% of beginners, 50% of intermediate, and 80% of advanced participants) either initially thought the claims could be true or maintained the possibility throughout their time reviewing the article. Many commented along the lines of “*this story could be true due to stuff that happens in America,*” [P2] and the world, “*there is so much crazy stuff happening in the world that I wouldn't immediately think it was fake*” [P21], or just people in general, “*the story could be credible, people do weird things*” [P15]. In addition to “*the claim isn't too crazy*” [P5], [P11] suggested “*...its New Orleans and they have a history of black magic.*” While many acknowledged the claims could be true, [P9] summarized some participants opinion that “*the article could be true, a one-off thing that could have happened. But it's just a one-off news story so it does not have as much weight, nor will it*

really matter to most people.” Thus, in addition to those who thought the article could be true, there were also many who acknowledged that it could just be entertainment and wouldn’t fully believe it, for example, we further [P5]’s thoughts on the website and the article claims, *“the website looks credible but isn’t something I would completely believe. I wouldn’t bring it up in conversation or take it to heart. It’s something I would read but wouldn’t necessarily believe it”* [P5]. As well as [P16], who stated, *“the claims the story is making - it’s such an obscure topic, it could go either way. I may read it and think it’s interesting but wouldn’t take it to heart - like it is real news. The charges seem extreme. I wouldn’t internalize it and think it’s true”* [P16]. [P3] particularly acknowledged that the website has weird news but *“I might still want to read it word for word because it is so interesting. She would be more likely to send this to a friend due to shock factor but not actually believe it”* [P3] and [P10] mentioned, *“if a friend sent me this, I would just take it as a laugh. I would not put much thought into it. It’s more for entertainment purposes.”*

Table 13. Participant Opinion on Article 2 Article Content Credibility

	Beginner	Intermediate	Advanced
Article Credible	20%	12.5%	20%
Use of References and/or Figures	20%	25%	-
Article Not Credible	80%	50%	40%
Initially thought to be Credible	40%	50%	80%
Arguments too Broad or Misleading	-	25%	40%
Appears to be an Opinion, Not Factual	20%	12.5%	-
Related Articles Lack Credibility,	60%	25%	40%
Reflects Poorly on Article Considered			
Photos/References Lack Credibility	40%	25%	-
Author Name/Credentials Missing	-	12.5%	-
Unsure if Article Content Credible	-	37.5%	40%

One participant mentioned that *“it’s hard to confirm if its real or not since it’s just a story, it’s not science”* [P24], thereby suggesting that articles with scientific claims may be easier to debunk rather than news stories. Considering this, alongside the claims that some people may read (satirical) news and share it with others without internalizing it or taking it seriously, may suggest that articles making scientific claims may be given more importance to verify claims, over news stories with just ‘entertainment value.’

What ultimately lead participants to decide that the article claims were not credible, was mainly due to the extreme headlines of the other news articles on the website, to elaborate, [P9] commented *“the title leads me to believe that someone could have done this. But with the other news articles around it, it doesn’t seem like they are all true.”* Moreover, the claims being made within the article itself sounded too extreme, *“it seems like a heavy fine for the charges...”* [P21], and *“1-million-dollar fine is too much for voodoo”* [P12]. In addition to the claims, 40% of beginners and 25% of intermediate participants thought that the supporting references and images appeared to be less credible. [P7] was surprised that a photo was even provided, given that *“...in the past even for real articles, I was not able to find a photo.”* Commenting on the photos that were provided, [P12] pointed out, *“the photo is hazy, so it looks like it has been taken from somewhere else.”*

Key Take-Aways:

It was encouraging to see many participants recognize that the article claims, and the websites were not credible, with some distinctly searching for specific clues to assess its credibility. We highlight some aspects that led participants to think the website or article claims were uncredible:

- Unprofessionally designed website;
- The presence of excessive advertisements;
- Lack of an author name or contact information;
- The lack of scientific evidence for science-based claims;
- Extreme, sensational wording;
- Extreme claims;
- Mixing opinions with fact;
- Questionable references or images.

We also highlight what led participants to think a website or article claims were credible:

- Professionally designed website;
- Inclusion of quotes and scientific referencing;
- Professional, suitable domain name;
- Ability to share the article;
- Ability to download as a PDF (for ‘academic’ articles);
- An award-winning author.

Digital Disinformation Countering Tools – Usability Evaluation

Once participants reviewed and provided their opinions on the three articles selected, they were asked to utilize the identified websites and browser extensions to verify either the author, the website, or the article claims.

Task 3: MediaBias/FactCheck Website

As per information presented on the MediaBias/FactCheck website, it has the capability to both check the author or source, as well as fact check. As participants were verbalizing their thoughts on the website while attempting to verify the article claims or source, we collected their feedback and thoughts on the website, the interface, the search process, and the results obtained.

Table 14. MediaBias/FactCheck Website - Participant Opinions

Interface/Website	Beginner	Intermediate	Advanced
Website/extension overwhelming/too much content	80%	50%	40%
Too many advertisements	40%	37.5%	80%
Website/extension looks trustworthy	20%	12.5%	-
Website/extension doesn't look trustworthy	40%	50%	60%
Thought had to create an account	20%	-	-

As can be seen in the results highlighted in *Table 14*, each persona commented more on certain issues than the others. To elaborate, 80% of the beginners were overwhelmed by the content that was included on the website, with [P2] complaining that that “*the layout is distracting. The news ticker that has changing headlines diverts my attention,*” followed by the presence of too many advertisements, to which [P12] questioned if she will get a virus by clicking on something. These complaints possibly affected their opinion of the website not appearing to be trustworthy, or

credible itself, as [P12] pointed out alongside her other complaints, “*it does not seem wholly credible either.*”

In turn, the intermediate participants were slightly more focused on the lack of trustworthiness the website had, followed by the overwhelming content and ads, with [P24] commenting that the “*website is crowded*” [P24] and [P16] elaborating, “*the website gives information that I was looking for, but it had a lot of ads and pop ups, so it doesn't seem very legitimate.*” [P7] also commented “*I know people need to make money to eat but it's a fact check website; they shouldn't have all these ads.*”

The advanced participants were more focused on the advertisements and were more likely to think the website was not trustworthy, with [P3] immediately commenting that the “*website looks sketchy*” and “*I should have left my ad blocker on*” as soon as the website opened. Overall comments on the website are that “*the way the website is set up could be better, it's not as user friendly as it could be but it's not bad. It was pretty easy to use but it could be better*” [P9] and “*a website that checks facts should have a more professional website*” [P4].

The search process is where we attempted to assess users' interaction with the website to verify either the source of the claim, or the claim itself. As the focus of a usability study, we wished to determine how easy it was to interact with the website and to carry out the tasks for which the website was being visited. Table 10 provides an overview of user experiences during this process.

Table 15. ‘MediaBias/FactCheck’ Website - Search Experiences

Search Process	Beginner	Intermediate	Advanced
Immediately found where to perform a search	60%	87.5%	80%
Did not know how to perform a search	40%	37.5%	20%
Took multiple searches to figure out how the search works	40%	37.5%	80%
Tried to (incorrectly) search using the article URL	40%	37.5%	80%
Tried to (correctly) search using the article URL	-	-	-
Thought they could do a "google style" topic search query	80%	100%	20%

While most participants were able to quickly determine where to perform the search, as Table 15 shows, beginners had the most challenging time to find the search box. [P23] commented, “*the website isn't clear how I should confirm if the information is accurate or not,*” while [P12] was distracted by the pop ups and content on the website and had to be prompted to look for a standard search box. Further, [P18] was overwhelmed by the sections of the website to visit, therefore, was not sure if they should select one to start their search. Due to the fact that there are two different places where a search can be performed on the same page, [P3] wasn't sure if she was searching in the right place. It was “*difficult to see where to actually search.*”

Once the search process was started, there was a mix between searching using the URL or a ‘google-style’ search query. Of the beginner and intermediate participants, 80% and 100% respectively, at one point during the search process thought that they could perform a ‘google-style’ search, and 80% of the advanced participants tried to search using the URL of the article. MediaBias/FactCheck website did indicate that a search could be performed by entering in the URL of the article to be verified, therefore, it could be understood that this information was understood by participants, however, this type of search did not return any results, despite the website containing information on the website being verified.

As *Table 16* demonstrates, the advanced participants fared much better in the results received, followed by intermediate participants. Considering results that were directly for the author, website or article claims, 60% of the advanced participants were able to receive results, in comparison to only 37.5% of intermediate participants and a mere 20% of the beginners. Given that it took 80% of advanced participants multiple search attempts (*Table 16*) before they obtained results, it could be suggested that while it may have taken multiple searches, the advanced participants may have better search skills.

The same 60% of advanced participants who received direct results understood the results received, whereas fewer (25%) intermediate participants understood the results, compared to the 37.5% that received directly relevant results. As we don't have any comments from participants on not understanding terminology used within the results, based on [P24]'s comment, "*because of the ads it's hard to find the information you are looking for,*" we can understand that the reason for not understanding results may stem from the placement of advertisements and the confusion it creates.

Another finding related to the types of searches that can be performed, the website represents itself as being able to fact check, however, it appears that searches can only be performed for media sources, rather than searching and findings answers for a specific topic. Thus, when participants were performing the "google-style" search queries, they included terms related to the topic, i.e., 'Chinese herb kills cancer,' or 'covid is a lie,' and so forth, rather than the name of the website or article author. To this note, [P5] stated that if they wanted to search for facts, i.e., is the pandemic real, then he would want to find results indicating if it was real, why, and so forth. [P5] expressed that "*maybe the website provides this but was not able to find it.*"

Table 16. MediaBias/FactCheck Website - Results Received

Results	Beginner	Intermediate	Advanced
Content Received			
Result directly related to website/author	20%	37.5%	60%
Results related to the topic	40%	62.5%	80%
Didn't receive related results	20%	62.5%	20%
Understanding of Results			
Understood results	20%	25%	60%
Didn't understand results - general	-	62.5%	20%
Didn't understand terminology	-	-	-
Had to infer understanding based on reviewing topic-related results	40%	12.5%	40%
Thinks if there are no results then the topic is fake	-	25%	20%
Ended task without finding an "answer"	40%	75%	20%
Felt that not enough information was provided	-	-	-
Interaction with Results			
Read initial result information only	-	-	20%
'Read more' for additional information	60%	25%	60%
Looked for/read the "About Us or How it works" section of the website	-	-	-
Commented that a google search is easier	20%	12.5%	-
Ads distracted from understanding/reading results	20%	-	-

Surprisingly, intermediate participants had the highest occurrence (62.5%) of not receiving results related to the topic being searched, with the same number also not understanding the results

received. Such is likely a direct contributor to the high percentage of participants who ended the task without validating the article source or claims (75%). Another possible explanation is that MediaBias/FactCheck only provides brief information in the initial search results, with more details provided if you click ‘read more,’ however, only 25% clicked to read the additional information. We clarify that for the intermediate participants, 62.5% both received results related to the topic and also did not receive related results. This is due to the task requiring participants to verify more than one article, in which they received results related to the topic for one article but not the other. In fact, while verifying the Chinese herb claims, participants received results deeming alternative therapies to be pseudoscience, however, similar search formats did not return any results for the remaining articles. This demonstrates that 62.5% of participants did not discover the search format that would provide direct results.

As to why some participants did not follow up to receive additional information, [P2] mentioned that they initially didn't feel the need to ‘read more’ as they felt informed enough by initial results. Once they did, they appreciated the additional information on the website owner and funder. This was also shared by [P14] who said, *“I Like that they have exactly what is wrong with it,”* and that the content was broken down under headers and some items bolded for easier reading. One negative effect of all the advertisements and overwhelming content is that when [P12] selected ‘read more,’ they focused more on the ads, not realizing there was more information, and ended the task without reading the clarifying information. Regarding the rating scales provided within the additional information, differing sentiments were expressed, with some commenting that they like the rating as *“it's funny”* [P15], while [P21] found *“the rating scales don't look professional.”*

Another interesting finding is that using the topic-related search results, both beginner and advanced participants were more likely to use their own knowledge and reasoning to infer whether an article was credible. To explain, when authenticating the first article, [P12] searched the name of the herb ‘lei gong teng,’ questionable source was provided as a rating to news sources, but it did not refer to the herb directly. Based on the results, they considered that it insinuates that the article claims are not credible. Moreover, even with exact results for the source being searched, [P16] mentioned that they would consider the information provided but would still use her own personal judgement alongside the information.

Key Take-Aways:

Considering feedback made during the task completion, a common complaint of MediaBias/FactCheck was that the website was difficult to navigate. With an extensive number of advertisements and cluttered text, several participants were left unsure as to where they should perform a search. The search process presented further challenges, with participants facing issues in formulating the right search query. Despite the website instructions, URL searches did not provide results, nor did it appear to be able to fact check. Rather, searches could only be performed using the website name. Those who received results related to the source they were validating did appreciate the information provided, including the website owner and funder as it helped to verify information.

Task 4: isitcredible Website

The second website for verifying digital disinformation was, “isthiscredible.” The website has a simple interface, with one central location to perform either URL, or topic searches, located in the middle of the page. Additional links are provided at the top of the page, i.e., ‘how it works.’

Table 17 provides an overview of initial impressions of the website. As some participants may have not verbalized their opinions of the website, the results may not reflect all opinions. However, between the beginner and intermediate participants, it was clear that upon seeing the website, they preferred ‘isthiscredible’ interface, in comparison to ‘MediaBias/FactCheck.’ Of which, [P2] mentioned, *“the website itself is very easy and straightforward to use and easier on the eyes. I like the layout as the other one was crowded and had writing everywhere.”* Similarly, [P15] said *“this website seems easier to use because there is nothing else to look at,”* followed by [P12] who said, *“I really like the simplicity and the lack of ads of this site. I like things simple; I don't like things complicated. If it's complicated, it's not going to be legit.”*

Table 17. Opinions of 'isthiscredible' Website

Interface/Website	Beginner	Intermediate	Advanced
Likes the layout of the website/extension	100%	87.5%	40%
Website/extension looks trustworthy	20%	-	-
Website/extension doesn't look trustworthy			20%

While all participants immediately saw where to enter in the information for their search, Table 18 shows that not all participants noticed that you could search using the article’s URL, in fact, while [P15] was confident in saying *“it's pretty obvious as to what you need to do,”* they appeared to not notice the option to search via URL, rather, they carried out a ‘google-style’ topic search, without a positive result. Moreover, participants who performed a topic search did not receive results directly related to the claims being made within the articles, thus suggesting that the ‘isthiscredible’ platform may only be useful for ‘viral’ widely spread topics.

Table 18. 'isthiscredible' Website Search Experiences

Search Process	Beginner	Intermediate	Advanced
Immediately found where to perform a search	100%	100%	100%
Didn't understand how to perform a search	-	25%	40%
Took multiple searches to figure out how the search works	-	12.5%	-
Tried to (incorrectly) search using the article URL	-	-	-
Tried to (correctly) search using the article URL	80%	62.5%	80%
Thought they could do a "google style" topic search query	40%	50%	40%

When performing a search using the URL, the result is provided via a red box that highlights the search bar and with ‘questionable news,’ written in small, red font. Given the format in which the results were returned, a small number (20% of beginners, 37.5 of intermediate, and 20% of advanced participants) did not notice the result. The moderator noted that it appeared that the participants assumed it was an error message and did not focus to read the small print. This was confirmed by [P9] who said, *“the red text here makes me think it cannot find anything but then it says, questionable news.”*

Of those that recognized the result, for some it was not clear what ‘questionable news’ meant. As [P16] questioned, *“it doesn't really do anything. Questionable, in what sense? it doesn't really do anything for me,”* along similar lines, [P14] commented *“it is telling me that it's questionable, but I don't know why. Is it about the source, website, article?”* After receiving this result, most of the advanced users (80%) went to the ‘how it works’ section of the website to see if there were definitions provided or how to find additional information. Upon viewing the content provided, [P9] liked that it identifies how the rating is done, that there is both an algorithm and

human fact checking. To which they said, “it makes me feel more confident in using it,” however, [P10] pointed out that, “it mentioned a rating, it would be nice to see that rating. It’s lacking information.” Thus, we note that the ‘how it works’ section did not provide any definitions or additional information to further explain what ‘questionable news’ meant.

As highlighted in *Table 19*, based on the lack of explanation, on the lower end, 37.5% of intermediate and on the higher end, 60% of advanced participants felt that not enough information was provided. Regarding the lack of information, [P2] commented, “this is so unsatisfying, I need to know why,” and [P16] elaborated, “it should give more details as to why it is not credible for people to know what to look out for in the future.” In fact, the lack of information lead at least one participant to not feel any more confident in knowing whether the article or author are credible or not, “I’m getting my answer that I am looking for but the way it is searching for it, it isn’t giving me any information on why it is telling me this, so it doesn’t really make me feel any more confident than before” [P14].

Table 19. 'isthiscredible' Website Results Received

Results	Beginner	Intermediate	Advanced
Content Received			
Result directly related to website/author	60%	62.5%	80%
Results related to the topic	-	-	-
Didn't receive related results	40%	50%	20%
Understanding of Results			
Understood results	60%	50%	60%
Didn't understand results - general	20%	25%	-
Didn't see the result that was given	20%	37.5%	20%
Didn't understand terminology	-	-	-
Had to infer understanding based on reviewing topic-related results	-	-	-
Thinks if there are no results then the topic is fake	20%	25%	-
Ended task without finding an "answer"	40%	25%	20%
Felt that not enough information was provided	40%	37.5%	60%
Interaction with Results			
Read initial result information only	-	-	-
'Read more' for additional information	-	-	-
Looked for/read the "About Us or How it works" section of the website	40%	37.5%	80%
Commented that a google search is easier	-	12.5	20%
Ads distracted from understanding/reading results	-	-	-

Despite most participants liking the website and layout (*Table 17*), the lack of information was a deterrent in fully embracing this website as a better option for fact checking, with [P2] stating, “I preferred the other website though as it did provide more information.”

Key Take-Aways:

While participants appreciated the layout and the ease of performing a URL search, those who performed a topic search received irrelevant results. Furthermore, some participants searched ‘www.globalresearch.ca,’ in which they received a message that it was ‘not able to analyze.’ When they searched ‘worldnewsdailyreport.com,’ they received results unrelated to the website.

Once search results were received, not all liked the method of providing feedback, particularly, the lack of information.

Task 5: The Newsroom (Beta) Extension

Once participants finished using the websites to validate the articles, they were asked to validate the same articles using browser extensions. Before navigating to the websites to view the articles, participants were instructed to install the extension and had the opportunity to review any content provided by its creator. Once it was installed and participants were ready to move forward, they were instructed to navigate to the articles to assess how the extension would interact with the website and/or authenticate information found on the website.

The first chrome extension, ‘The Newsroom’ is a beta version, thus, in the process of being tested. How ‘The Newsroom’ works is that an icon appears on the top left side of the screen when a user scrolls through a website (see Figure 3). Once the user hovers over the icon, a box appears with a source and political bias score, a summary of the source, website history and ownership.

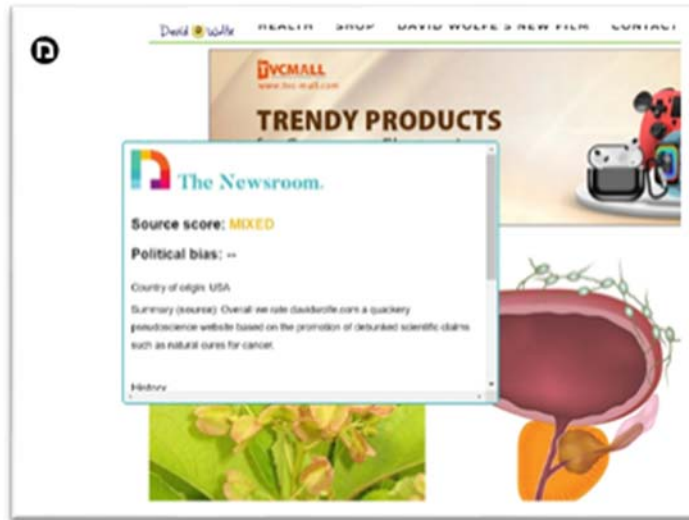


Figure 3. The Newsroom Browser Extension

One major issue experienced by most, and particularly all beginner participants, was how to open the extension or have it appear on their screen (Table 20). As they didn’t know what to expect, 60% of them went to the settings to see how the extension works and how to get it to appear. It was only by chance that when they went back to the website and scrolled through to see if there were any changes that they noticed the icon appear, “*yea, it only takes a bit of scrolling and then it shows up.*” Once it appeared, participants enjoyed the ease of which it opened, “*oh, it's that simple enough*” [P15].

Table 20. Participant opinions of 'The Newsroom'

The Newsroom (Beta) Extension			
Interface/Website	Beginner	Intermediate	Advanced
Likes the layout of the website/extension	60%	50%	40%
Thought had to create an account	-	-	-

Couldn't immediately figure out how to open the website/extension	100%	87.5%	80%
Couldn't close the website/extension	40%	25%	80%
Pinned the Extension	20%	-	20%

Once the icon appeared, there were mixed reviews of the extension layout (*Table 20*) and the information provided (*Table 21*). Some participants, such as [P23] commented, *“that’s actually brilliant. I like how they did the ownership, history, and the site score.”* [P10] also remarked *“I like getting information about the entire website in the summary, it gives an idea of what will be covered in the article.”* Some participants also appreciated that *“it does present the information in a nice and clean way”* [P16].

While the results shown in *Table 16* demonstrate that most participants understood the information being presented on the website source, the beginner participants (60%) did not understand the terminology and the rating scales being used. [P18] commented *“the information in the extension was easy to read but the terminology wasn’t clear.”* [P18] further elaborated, *“they use different terms, so I need to read to understand what they mean by mixed. Does ‘low’ mean low accuracy or low risk? I don’t know if, for native speakers, they get the terms directly, but since I am not a native speaker, it confuses me.”* Considering the confusion for participants, [P23] recommended, *“my suggestion is to provide more information or reasons for why they rated it as a low or mixed. It would be more professional and accurate to do so.”* Moreover, in addition to providing more explanations and information, some participants had mixed feelings regarding the terminology being used. Although [P12] *“appreciated the quiriness and dry humor”* in terms being used, such as ‘tin foil hat’ and ‘quackery,’ others, such as [P14] thought it showed bias, *“I think calling a website quackery, makes this fact checker seem less credible because that shows me bias on its own.”* In addition to the aversion to seeing such terms, we refer to [P23] who didn’t understand what satire meant, and again refer to [P18] who commented about the terminology being used, that she is *“... not a native speaker, it confuses me.”*

Table 21. 'The Newsroom' Information Provided

The Newsroom (Beta) Extension			
Information Provided	Beginner	Intermediate	Advanced
Understanding of Results			
Understood information	60%	50%	80%
Didn't see information that was given			
Didn't understand website terminology	60%	37.5%	-
Ended task without finding an "answer"	-	37.5%	20%
Felt that not enough information was provided	40%	25%	40%
Interaction with Results			
'Read more' for additional information	40%	50%	40%
Looked for/read the "About Us or How it works" section of the website	60%	62.5%	60%

To address some of the confusion on what the rating scales and terms mean, we note that there was an almost equal division between participants who accessed the additional information, compared to those who only read the initial information (as shown in *Table 21*). If all participants had read the additional information, it may have clarified some of the confusion. Nevertheless, considering the feedback of [P2], *“if I just used the extension, without going to ‘more information’ then it would not be useful. The text provided in the box was not enough. Their grading system*

wasn't clear, and I wouldn't go out of my way to find it." Thus, considering the feedback, the initial ratings should be clear enough without the need to visit additional websites. However, should users desire to receive further information about the source and reasons for the rating, the ability to receive additional information should be clearly indicated. At least one participant almost did not realize that it was possible to do so and said, "...how am I supposed to know that I can click on the history/ownership information that does not seem clickable?" [P14]. While there was only one participant who expressed this frustration, it may explain why others did not follow the links for additional reading.

Lastly, we raise the issue of the icon window not closing when participants have finished reviewing the content provided. As shown in Table 16, it was not a major issue for all participants, however, 80% of advanced participants were not able to close it. As one participant commented, "if you move your mouse away it should go away" [P21], however, to do so was not clear, which obstructed participants from being able to continue reading from the website.

Key Take-Aways:

Overall, participants appreciated 'The Newsroom' more than the websites due to the ratings immediately being provided without the need to formulate a successful search query. Thus, despite the uncertainty of how to open the extension, once they figured it out, the usability of the tool was much better than their website experiences.

Given that TheNewsroom sources their data from the MediaBias/FactCheck website, they use the same terminology. It was interesting that participants commented more on the terms used while using The Newsroom, than while using MediaBias/FactCheck, however, it may be due to the limited information being presented, in which case the participants could focus on what was being presented more than on the MediaBias/FactCheck website where the information and content is cluttered and overwhelming.

Task 6: Newstrition Extension

The second of the browser extensions is 'Newstrition,' which, as seen in Figure 4, provides a warning banner across the screen of 'problematic' sources, an extension window that can be opened for additional information, as well as a website that can be visited for further information.



Figure 4. Newstrition Banner and Extension Window

While some of the participants immediately noticed the orange banner and appreciated the immediate feedback, in comparison to the beginner participants, of whom 100% saw the banner, surprisingly only 50% of the immediate and 40% of advanced participants noticed it. Regarding the extension, [P2] mentioned, "the banner was useful because it gives you

information directly. If I was asked to find authentic sources and the banner tells me it's not authentic then I would accept it as is and move on." [P11] shared a similar opinion and said, "I like this, it just pops in your face and if you want more information, you can click to go to the website." Despite the positive comments, one participant did not like the banner and said, "I don't like that it's at the top of the website. It's intrusive for me" [P10]. While it was just one comment, it is something that designers should consider when creating digital disinformation countering tools.

Table 22. 'Newstrition' - Extension Experiences

Newstrition Extension			
Interface/Website	Beginner	Intermediate	Advanced
Thought had to create an account	40%	25%	80%
Couldn't immediately figure out how to open the website/extension	20%	-	-
Noticed the orange banner	100%	50%	40%
Opened the extension (window)	80%	75%	80%
Relied on the orange banner only	20%	25%	20%
'Read more' for additional information	100%	75%	100%

Another interesting finding is that despite having the orange banner, only a small number of participants relied on the banner alone, with a large number opening the extension window. Once they were there, however, many participants were disappointed in the amount of information available, "interesting, they don't have any information on this one, but they tell you it's problematic." "I like how straightforward this is but doesn't provide enough information" [P2]. [P21] elaborated on the lack of information, "this provides nothing. It seemed really cool at the start because there were so many tabs that I thought it would provide something, but I can't even - it said learn why but I can't. Outside of saying it is problematic it doesn't say why. When I open the website there is even less information."

In addition to seeking and reviewing information provided in the extension window, all but two of the intermediate participants clicked on 'learn why' for additional information. This takes users to the Newstrition website, which, as per [P21]'s comment, does not have any further information. [P14]'s comments elaborate on seeing the orange banner but then seeking more information, "there is not enough information, I would have liked more. I like that it was very clear, and I don't have to do anything, but I would have like to see more information. Where the ratings came from." In fact, most of the intermediate (87.5%) and all advanced participants complained of a lack of information. One participant mentioned the need for supporting evidence from an academic perspective, "if I was just doing this for personal reasons, I would be satisfied with the problematic rating but if I was doing it for school, my teachers would expect more information" [P2].

Table 23. 'Newstrition' - Extension Opinions

Newstrition Extension			
Information Provided	Beginner	Intermediate	Advanced
Understanding of Results			
Understood information	100%	100%	100%
Didn't understand website terminology	20%	-	-
Ended task without finding an "answer"	-	-	-

Felt that not enough information was provided	40%	87.5%	100%
Interaction with Results			
Looked for/read the "About Us or How it works" section of the website	40%	25%	20%

To further discuss [P21]’s complaint of wanting to know where the ratings came from, this was a point of grievance for multiple participants, with [P10] mentioning, *“I don’t understand what goes into the aggregate ratings, this isn’t anything helpful,”* and [P14], *“I wish I knew how they did the ratings. Maybe you have to log in to find more information.”* In addition, some participants worried about who is doing the rating, *“they should make it more clear how the ratings are done and if its general users that could provide their rating. People have biases, and if just anyone can change the rating, then wouldn’t be so confident in the response it is giving”* [P9]. [P21] also had similar concerns, stating, *“its community based so the problem is that you don’t know who is rating it and what their stance is.”*

To receive in-depth information, the advanced participants particularly noticed that there was an option to create an account and log in, to which they said, *“sign(ing) in for more information or subscribing to receive information seems like a bit much”* [P16] and *“it’s community based so the problem is that you don’t know who is rating it and what their stance is”* [P21].

To summarize many participants opinion of ‘Newstrition,’ we quote [P5] who said after opening the website *“well this one itself looks a bit sketchy - I don’t like the sign up, it looks like it’s trying to sell me something. I also don’t like the ads. It doesn’t look legitimate. If I opened this, I wouldn’t think it was legitimate and would go back to the website.”*

Key Take-Aways:

Based on participants' experience with the browser extension, they all understood the information conveyed and there were no major issues with the terminology being used. While a good number of participants noticed the orange banner and liked the immediate feedback, it was not noticed by all, thus suggesting that unless users are provided with clear explanations including visuals when they install the extension, it may not be suitable for all users. Moreover, considering the number of participants who clicked on the extension with the apparent anticipation that something will open may suggest that past experiences using browser extensions provide for an additional resource appearing. Developers should account for this expectation when creating new tools.

Similar to what has been experienced with other tools used within this study, while participants like the initial feedback, in most circumstances they wished to have the ability to click on a link and receive additional information on the author, the website, and/or on the claims being made. Although not all participants commented on the issue of how the ratings are performed, i.e., a community rating approach and individual biases, the participants who raised this concern weighed in with heavy importance. Rightly so, as those believing in conspiracy theories and/or wishing to engage in forms of disinformation may “hijack” the opportunity to rate something inauthentic as authentic to disinform society. Therefore, those providing community-based ratings should also provide a form of monitoring to prevent this from occurring.

In the search for an explanation as to why the publisher was problematic, those who commented on the requirement to sign up to receive additional information were not interested in doing so. Thus, it is our recommendation that users are not required to sign up to access resources to authenticate information, as it may pose as a barrier and further enable disinformation to spread.

Task 7: MediaBias/FactCheck Extension

While the participants reviewed the MediaBias/FactCheck website at the beginning of the usability evaluation, we then asked them to evaluate the extension at the end of the session to see if their opinions would be maintained.

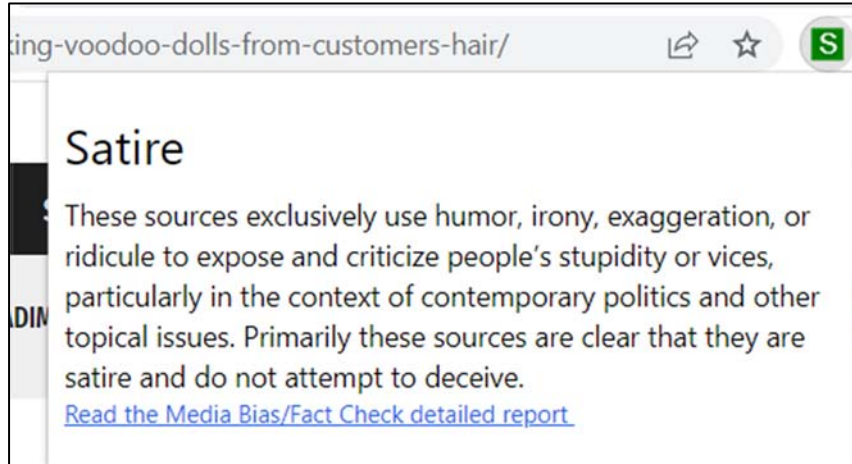


Figure 5. 'MediaBias/FactCheck' Browser Extension

One action that the researchers found interesting was that 80% of the beginners and 60% of the advanced participants immediately pinned the extension (Table 24), which was not noticed while they were interacting with the first two extensions. Once the extension was pinned, if it was a rated website, users would see that the extension icon changes colour and letter based on the rating. Based on this, pinning the extension plays a major role in the user experience and could be the reason for pinning this extension, over the others whose icon did not change. One participant commented, "I don't think I would have had the same experience if I had not pinned the extension to the tool bar. I wouldn't have seen the ratings and would have to manually check each time" [P15]. Considering this, [P11] mentioned, "people just have to know to pin it otherwise they wouldn't see the rating." Regarding the colour and rating change, [P2] and [P18] both commented that they liked it, however, [P11] questioned "what the reasoning behind which colours are used. Since green was used for satire, if I am in a hurry and I see green then I might assume that it means that it is ok."

Table 24. "MediaBias/FactCheck" - Extension Experiences

MediaBias/FactCheck Extension			
Interface/Website	Beginner	Intermediate	Advanced
Couldn't immediately figure out how to open the website/extension	-	12.5%	-
Appreciated changing colour/code	80%	37.5%	60%
Pinned the Extension	80%	37.5%	60%

Once participants open the extension, a small text box appears with the overall rating, i.e., 'conspiracy,' 'pseudoscience,' and so forth, then provides a short paragraph on the definition of the rating, and a link to go to the website for more information. This method of providing information received mixed reviews. While most of the participants "liked step by step, incremental information

provided” [P11], because “...if you click on it, you can get a bit more information and if you still want more you can click to go to the website” [P2].

However, multiple participants commented on how the information was being presented within the extension, “I don't like the way they are presenting the information, it's not easy to understand. If they show the main theme and they can provide a link to the details” [P10], particularly, “everyone has a short attention span so it's better to see it and click for more information rather than a lot of text” [P21], and “the extension is ok, but people don't like to read paragraphs” [P7]. Some provided suggestions for improvement, such as, “if you put the information using 2-3 words, that would be perfect” [P18], and “bullet points should be used instead of paragraphs” [P21]. On that note, we mention one participant who only reviewed the extension and did not click for more details, who said, “it doesn't make her feel better informed since it mostly just provides definitions rather than explaining why the website has that rating” [P16]. Thus, the information provided within the extension itself should be short and to the point of explaining why the website has such a rating.

Table 25. 'MediaBias/FactCheck' - Extension Opinions

MediaBias/FactCheck Extension			
Information Provided	Beginner	Intermediate	Advanced
Understanding of Results			
Understood information	100%	87.5%	100%
Didn't understand website terminology	20%	12.5%	-
Ended task without finding an "answer"	20%	-	-
Felt that not enough information was provided	-	12.5%	-
Interaction with Results			
'Read more' for additional information	100%	50%	80%
Looked for/read the "About Us or How it works" section of the website	-	-	-

While using the extension, although participants didn't enjoy reading the paragraph of text, all of the beginners and 80% of the advanced participants went to the website for additional information (Table 25). Almost all participants understood the information being provided, with only one complaint of not enough being provided (Table 25). We highlight this finding in comparison to the website use alone, where only 20% of beginners and 60% of advanced participants understood the results, with up to 40% of the same participants having to infer an understanding of whether the topic or source was credible or not, based on results related to the topic. Moreover, 75% of the intermediate participants ended the task in the MediaBias/FactCheck website without finding an answer that verified the articles' information or source. Therefore, in comparison to using the website, it appears that by using the extension, participants were better able to access the required information.

On that note, we quote some participants who addressed our goal of identifying whether the extension provided a better user experience, “the extension was easy to use, much better than the website” [P4], and “now that I am going through the website from the extension, I really like how its working and the information it provides” (i.e., the failed fact checks) [P14]. Lastly, “I appreciate the website more after using it with extension as opposed to just the website itself. MediaBias has potential to be better but needs improvement” [P7].

Similar to comments made when reviewing ‘Newstrition,’ [P9] said, “*I would like to see how biased is MediaBiasFactCheck and how they do the categorization because human bias is something else to take into consideration,*” and [P11], who did not previously address this concern while reviewing ‘Newstrition’ noted that “*...it would be nice to know who verifies the information.*”

Key Take-Aways and Recommendations:

Based on the comments made by participants, it appears that the MediaBias/FactCheck browser extension presents a much-improved user experience. Particularly since participants did not have to learn what search methods are supported by the website, it eliminated one major usability issue that participants had with the website. We also note that the participants understood the results and did not have to deduce whether a source or the article claims were credible based on related results, thus eliminating another barrier to successful use of the platform.

To further improve users’ experience with the browser extension, participants’ complaints on the visual presentation of information should be considered. Only essential information related to fact or source checking should be provided in simple terms, rather than definitions of the rating. Additional definitions and context may be placed on the website, should users choose to receive additional information. Moreover, to improve users’ confidence in the ratings, the website should clearly and transparently provide information on how their fact checking is performed and who is responsible.

11.3 Conclusions

Once the participants completed the tasks, they were asked to provide a SUS rating. In addition to their thoughts, opinions, and a count of the frequency of which an experience (whether good or bad) was occurring, this rating would provide a quantitative score for each tool assessed.

As can be seen in *Table 26*, the lowest SUS score across all personas is the MediaBias/FactCheck website, while the same extension was rated the highest for both beginners and advanced users. The intermediate participants rated isthiscredible as the highest. We note that the highest rates of standard deviation are between the same tools, with the MediaBias/FactCheck having the highest rate at 5.30, followed by isthiscredible at 4.25 points. This accounts for the difference in which the different personas rated each tool, and a clear appreciation for one tool over the other.

Despite the clear tool preference, we consider the meaning ascribed to the ratings received. On this account, we refer to established guidelines for interpreting the scores. Such guidelines confer that anything rated below 60 is considered to receive a ‘F’ grade. More specifically, anything between 25-39 is poor, whereas any score between 39-52 is rated as ‘ok,’ however still not acceptable (Bangor et al., 2009). Based on the rating scale, and while the MediaBias/FactCheck extension received the highest rating at 44 points, it still falls within an unacceptable rating from a usability perspective.

Table 26. Final SUS Scores

	SUS Scores				
	MediaBias/Fact Check Website	isthiscredible	TheNewsroom	Newstrition	MediaBias/FactCheck Extension
Beginner	27.00	31.00	35.50	35.00	38.00
Intermediate	22.50	35.00	31.25	31.56	33.44

Advanced	23.50	39.50	35.00	35.50	44.00
Average	24.33	35.17	33.92	34.02	38.48
STDDV	2.36	4.25	2.32	2.14	5.30

In addition to the SUS scores, at the end of the session, participants were asked to rate which website or extension they found to be most up to date. The reason for this question is that the researchers wanted to ascertain which tool participants thought would provide them with the most current information to assist them in the verification process.

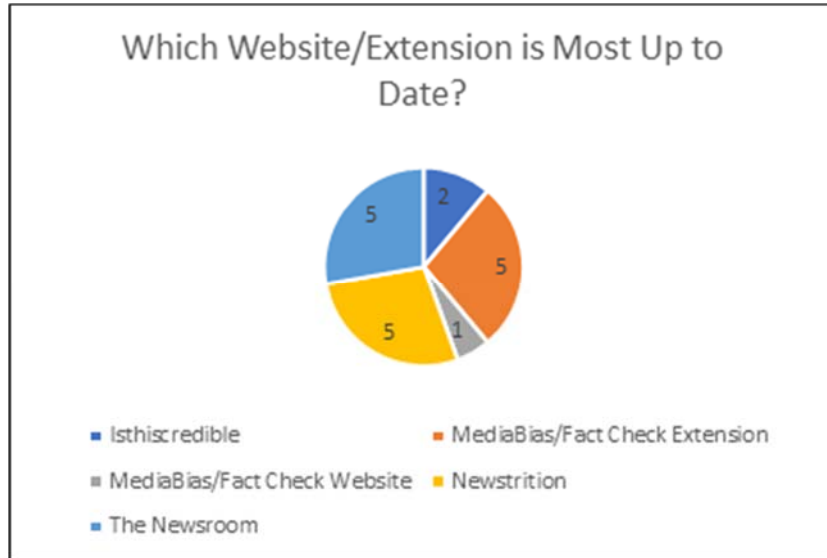


Figure 6. Which Tool was Most Up to Date?

As shown in Figure 6, surprisingly, despite the MediaBias/FactCheck Extension receiving the highest SUS score, it was tied with Newstrition and The Newsroom as the tool considered most up to date. Also unexpected is that while ‘isthisccredible’ had the second highest SUS score, it had the second lowest ranking for being up to date.

Participants were also asked which tool they would be more likely to use, post evaluation. The reason for this question was to determine which tool the users preferred and would more likely use in their daily lives.

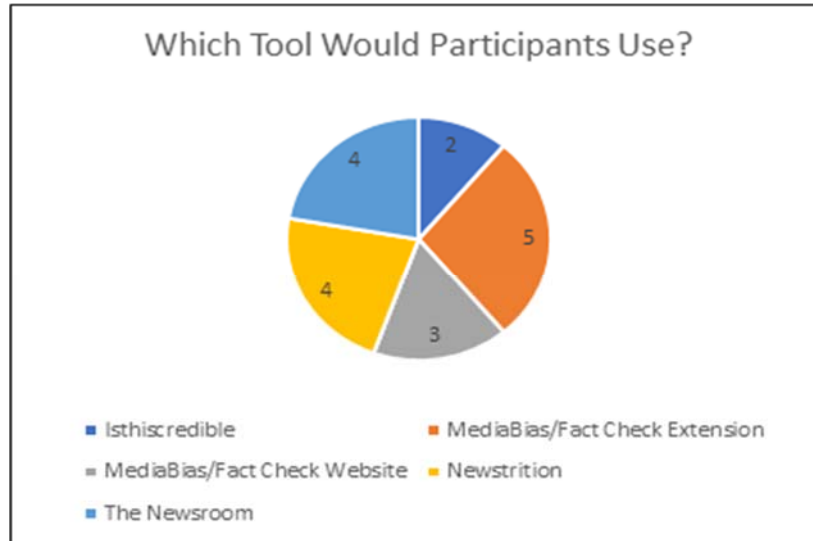


Figure 7. Which Tool Would Participants Use (Post Evaluation)

In line with the results received regarding the tools that were most up to date, the same three tools were selected, with the MediaBias/FactCheck Extension only having one more selection than the others (Figure 7). Albeit only an incremental lead, this aligns with the SUS score. Although it is surprising that The Newsroom and Newstrition would be included in the list of tools considered most up to date and most likely to be used post-evaluation, their similar SUS scores and lower standard deviation could explain why it would be preferred over ‘isthisccredible,’ which had a higher standard deviation, signaling a larger difference of opinion.

Further elaborating on ‘isthisccredible,’ despite it having a nice and easy to use interface, which participants initially appreciated more than other tools, it had the lowest rating for future use and second lowest rating in terms of being considered up to date. Therefore, while a good interface is an important aspect of the user experience, other aspects that influence its usability, such as how efficient it is and the rate of satisfaction while using it will influence whether users want to continue using it. In the case of ‘isthisccredible,’ only users that recognized that a URL search could be performed received a result that helped verify the article, leaving it to be inefficient for those who did not receive a relevant result. Moreover, the format in which the results were provided, as well as the lack of supporting evidence led to dissatisfaction with the website.

This finding can be further compared to participants' opinions of the MediaBias/FactCheck website. Although many initially had challenges performing a search and receiving relevant results (in addition to the overall displeasure with the excessive advertisements and cluttered information), once they used the browser extension and were directly brought to relevant information on the source of the articles, they appreciated the website more than while completing the initial task. Thus, the elimination of opportunity for error increased the efficiency, while then led to users being satisfied.

Based on our review of currently available tools, the participants experiences, feedback, SUS scores, and multiple-choice question selections, we conclude several insights and recommendations for future digital disinformation countering tools.

Websites

The usability of current websites is plagued by multiple issues. From an interface perspective, except for *isthiscredible*, many of the websites were poorly designed, which impacted the ability to navigate the tools, where and how to perform searches, read content provided, or know where to find more information. Considering the search process, while participants appeared to expect a 'google-style' search capability, based on results received, the websites seem to only accept specific forms of search queries to return relevant results. This increased the errors experienced and impacted their ability to use and verify information using the tools. In some tools, the advanced participants fared better than beginner and intermediate participants, however, this was not always the case. At times, the beginners had better experiences than the intermediates, with many cases where their experiences aligned with advanced participants, thus suggesting the severity of the issues, as they plagued all personas. Despite some differences found in persona profiles and considering that the beginners often had similar experiences as the advanced participants, it cannot be stated that one group of user profiles may fare better. Accordingly, and particularly as the tools are created for the general population, they should be designed to accommodate all users.

In addition to the poor design and search capabilities, a lack of information was another common issue experienced. While *MediaBias/FactCheck* did provide extensive information, the form in which it was presented took away from the readability and likelihood of participants receiving the information presented. On the other extreme, *isthiscredible* provided no information to support the rating provided, which greatly diminished the experience using the tool, and as we have seen, the likelihood of continued use.

Browser Extensions

Though the browser extensions provided a much-improved user experience, many participants complained about how information was being presented. Initially, participants appeared to prefer the Newsroom, however, the rating scales and terminology used was confusing for some users, particularly those who speak English as a second language. *Newstrition* presented multiple 'tabs' under which it was expected that information would be found, however, both the extension and the website that was supposed to provide additional information were lacking. While it appeared that more information could be found by creating an account and signing in, participants saw that as a barrier and did not wish to do so. The *MediaBias/FactCheck* extension was brief and to the point, which the participants noted that if it was pinned, they could immediately see the rating and appreciated. However, to this point, they noted that the colours used were confusing as green was used for a 'satire' rating. Whereas green usually indicates 'good' or 'go,' a satire-type website should be read with caution, thus, a yellow or orange colour may be more appropriate. Considering these aspects, the efficiency of the tools could be markedly improved by reconsidering design, eliminating unnecessary sections and content that does not provide immediate value. Moreover, at a minimum, the terminology used should be basic terms that varying language profiles would understand. A better approach would be to use a numerical rating scale and standardized colours to reduce the cognitive load and opportunity for error in interpreting the rating given.

Tool Availability

One observation is that locating a tool can be a challenge, with many participants not being sure of what to search for and not being successful in locating a tool. To which we note that the researchers' search for tools found that searching for browser extensions in the chrome extension

menu did not return results, rather, it can only be done via the web store. When searching for tools using “fact checker,” “misinformation,” “disinformation,” a minimal number of extensions were found.

11.4 Recommendations

The aim of our research is to encourage the public to counter the plague of disinformation, whereby the disinformation countering tools may assist. Thus, given the findings of our study, we have summarized several aspects that should be considered while developing digital disinformation countering tools. Such recommendations are given with the goal of avoiding any of the usability issues experienced by participants during the evaluations, as well as to improve the overall user experience while using the tools.

Establish Credibility

To encourage users to trust and turn to digital disinformation countering tools, the most important task is to present the tools as credible and trustworthy sources. While it is understood that developers require funding to create and maintain the tools, the placement of extensive advertisements conjures less credibility. Accordingly, advertisements as a source of funding should be avoided or kept to a minimum, if necessary.

To further establish credibility, tool developers should clearly present how they are performing the fact or source ‘checks,’ including whether it is human, or machine (learning) based, or a combination of methods. Moreover, any data sources, decision-based criteria, or rating scales they follow in the verification process should be communicated. They should further provide information on who is involved in the process, what their roles are, past experience, educational backgrounds, and what their affiliations are, if any. Such information would promote transparency and reduce worry of biased ratings, thus, solidifying trust in the tool. Such may then lead to a potential increased use in digital disinformation countering tools.

As at least one of our browser extensions (Newstrition) appeared to solicit community feedback to rate sources, some of our participants worried about the bias that it may introduce into the ratings being provided. Thus, should fact-checking tools choose to follow such an approach, it must be made clear when results are provided and address how they mitigate the risk of falling prey to conspiracy theorists and others who wish to spread disinformation. To eliminate this risk completely, we conjure that it is not advisable to allow anonymous, community-based voting on the credibility of information or sources.

Tool Layout and Design

Should a website or browser extension be difficult to understand and perform the necessary functions, users will be less likely to want to use them. Given that many participants had difficulty either navigating the tool, were initially unsure how to get it to work, or in some cases, unable to get it to work, increased attention must be made to the design and layout of the tool. Modern approaches to graphic design and user interfaces should be applied to ensure that there is adequate white space, easy to read fonts, colours, and an inclusion of only required information.

Search Capability

Performing searches on disinformation countering websites featured a major usability challenge for participants, thus, further to the design of a website that is easy to understand, the

process of performing searches should be given importance in tool development. When users can first quickly and clearly identify where to perform their search, the process is greatly simplified. In addition, the formats in which searches can be performed should be clearly outlined for users to understand how to perform a search in a manner that will most likely return results directly related to their topic. Considering that many of the participants attempted to perform ‘google-style’ search queries, importance should be given to the optimization of the tool’s search capabilities, to accommodate differing forms of search queries. We also note that the search instructions given should be matched with the website’s search capabilities.

Alternatively, as the usability of the browser extensions was much better than the websites, thus providing a better user experience, the promotion of browser extensions could reduce errors experienced in the search process. As the extensions directly provide ratings for each website visited, users are not left with the challenge of figuring out an appropriate query to receive a direct result. Moreover, they do not have to try to infer whether or not an article or source is credible based on indirect results received from website search queries.

Information and Presentation

When the tool presents the results of the verification, attention should be given to how and what information is presented. As many participants complained of large paragraphs of text being too much to read, only the initial rating should be provided with an option to navigate to another page or section to read additional information, should the user wish to do so. We note participants varied regarding whether they wished to receive additional information or not. Some were satisfied to see the overall rating and move on, whereas others wanted to know why such a rating was given. Thus, both options should be made available to users.

Along with providing users an option to view an initial, or overall rating, along with the ability to receive additional information, we recommend that developers do not require users to create accounts and login to view information. As witnessed within our study, this requirement in the Newstrition extension was not well received by some participants. Those who did notice and commented on it mentioned that they would not likely sign up for such a service. Therefore, to encourage accessibility, we recommend developers allow for open access to information, without the need to create an account.

To further discuss the rating scale, as some participants complained of not having a consistent rating scale across tools, in which case, it took additional time to search the definitions of the ratings to see what each one signified. Moreover, the terminology used should be simplified for users of differing educational and language levels to understand. Terms such as “quackery” and “tin-foil hat” may not be understood by those who English is not their first language or may not have previously come across such terms. While it may be difficult to coordinate across all disinformation countering platforms, those involved with a centralized body, such as the International Fact-Checking Network, may be better positioned to do so.

Should this recommendation not be possible, developers should strive to make their ratings more general, such as in the form of a percentage or the use of a colour rating scale to visually receive initial feedback. Examples include the use of green to indicate a source is good, yellow to signify caution, orange a more serious caution, and red to warn users of serious disinformation. In doing so, tools will not have to allocate space for terminology definitions, rather they can focus on the rating and reasoning directly. As some participants indicated that it may be possible for an untrustworthy website to feature credible information, and vice versa, they therefore requested that tools should clarify whether it is the source itself, or an article topic that is being verified and rated.

Along similar lines of the information that should be included in disinformation countering tools to establish their credibility, similar types of information should be collected and presented on the source authors, website, and so forth. Some participants appreciated knowing who the website owner was, and who funded the organization, as it helped to provide context of who they are affiliated with, or any biases they may have. As previously mentioned, providing this information may raise awareness of platforms or authors to directly avoid in the future, without the need for countering tools.

Lastly, as one participant particularly mentioned, *“for people who don't know what conspiracies are or who believe conspiracies already, then it won't really be useful for them.”* [P7]. To that note, we comment that one participant in the study admitted to believing in some conspiracy theories and it was observed that they were particularly questioning the results that some of the tools provided, as well as the tools themselves. Therefore, our final recommendation is that in addition to creating websites and browser extensions that can be used to counter disinformation, further efforts must be made in educating the public on the dangers of disinformation and the use of critical thinking skills. This will assist in preventing individuals from reaching the stage where disinformation countering tools may be considered a conspiracy in itself, and thus avoided.

Table 27. Summary of Recommendations

SUMMARIZED RECOMMENDATIONS: DISINFORMATION-COUNTERING TOOLS
<p>1. Establish Credibility</p> <ol style="list-style-type: none"> 1. Advertisements should be avoided 2. Clearly present how they are performing the fact or source ‘checks,’ 3. Data sources, decision-based criteria, or rating scales should be communicated. 4. Provide information on who is involved in the process, what are their roles, their past experience, educational backgrounds, and/or what their affiliations are. <p>2. Easy to Understand Layout and Design</p> <ol style="list-style-type: none"> 1. Attention must be made to the design and layout of the tool. 2. Modern approaches to graphic design and user interfaces should be applied to ensure that there is adequate white space, easy to read fonts, and colours. 3. Include necessary information only. <p>3. Search Capability</p> <ol style="list-style-type: none"> 1. The process of performing searches should be given importance. 2. The format(s) in which searches can be performed should be clearly outlined. 3. Optimize the tool’s search capabilities to accommodate differing forms of search queries. 4. Search instructions should match the website’s search capabilities. 5. Promotion of browser extensions. <p>4. Information and Presentation</p> <ol style="list-style-type: none"> 1. Avoid large paragraphs of text. 2. Provide an initial rating with an option to read additional information. 3. Allow for open access to information. Do not require users to create accounts to view information. 4. Have a consistent rating scale across tools. 5. Simple terminology. 6. Use percentages and/or a colour rating scale to visually provide initial feedback. 7. Ratings should clarify whether it is for the source itself, or an article topic that is being verified and rated. 8. Establish tool credibility: Provide information on the source authors, website, and so forth.

12. Final Recommendations

Recommendations for the combat of digital disinformation can generally be divided into two categories, proactive and reactive (McLane, 2021; Shu et al., 2017) or detection and protection (Walker et al., 2019). For the purpose of this literature review, they are divided into recommendations that specifically target the aforementioned psychological, social, and technological enablers based on the full-text synthesis. Researchers have called on the internationalization of efforts for infodemic management and a holistic multidisciplinary approach to combat the creation and spread of mis-/dis-information (Agarwal & Alsaeedi, 2020; Ansar & Goswami, 2021; Gradoń et al., 2021; Zannettou et al., 2019). This idea is also referred to as the whole society approach to building resilient citizens and its goal is to assist in the comprehension of the threat environment and the identification of malicious intent (Ivan et al., 2021). This would be a multifaceted network that would empower citizens through “knowledge of propaganda and disinformation tactics and modus operandi, critical thinking, fact checking, source validation etc.” (Ivan et al., 2021, p. 9).

12.1 Recommendations for Civil Society

12.1.1 Media Literacy

One of the most prominent recommendations in the literature is media literacy. This is often addressed to the public sector and educational institutions. The goal is to extend digital literacy to groups that are most likely to be susceptible to mis-/dis-information. Based on some participants' inability to distinguish between a credible and uncredible website and article claims, further resources should be made available to educate the public on how to critically assess and identify sources of disinformation. While many of the criteria for deeming an article or website to be credible are requirements of credible sources, all the same criteria can be present in uncredible sources. Therefore, it is imperative to educate society on all factors to consider.

Many researchers believe the best way to combat fake news is through individual skills in digital literacy (Agarwal & Alsaeedi, 2020; Bran et al., 2021; Froehlich, 2017; Hwang et al., 2021). The main principles of the recommended models include using intelligence services to educate citizens about detection tools and active disinformation campaigns, creating a transparent ethical code of values for those intelligence services, and creating an “early warning” system to detect emerging operations (Ivan et al., 2021). Due to the covert nature and history of intelligence services, it is important to develop a transparent code of ethics to enhance credibility and foster trust with the public and stakeholders (Ivan et al., 2021). A far-reaching network like this proposed model would require the incorporation of many facets of society including libraries, educational institutes, think tanks, social welfare, and civil society organizations (Ivan et al., 2021). There is a need for not only education on identifying misinformation by oneself but also combining that education into looking at different automated tools to try and find a happy medium between speed and accuracy. Further focus on the actions one can take once they have confidently identified something as misinformation or disinformation. For example, one can report a user on social media if they feel confident enough, but for websites and other sources, perhaps a publicly funded website that has users submit sources they find less than credible for revision would be useful.

There are several ways that the public sector could increase media literacy. One notable idea is improving human detection skills through inoculation (Cook et al., 2017; Gradoń et al., 2021; Kumar & Shah, 2018). Allowing people to be repeatedly exposed to holes in arguments and differing perspectives, would enable media consumers to utilize their critical thinking skills more often when faced with a piece of “news” (Mayorga et al., 2020). Another idea, although similar, suggest that the public should take it upon themselves to occasionally explore opposing views (Au et al., 2021; Innes et al., 2021). This type of intrinsic criticism would ensure that the heuristic rule be adhered to, meaning that there would never be an opportunity to develop or believe an extreme viewpoint about an opposing view. This idea would in turn decrease the amount of disinformation as well with contribute to the depletion of extremism in news coming from a lack of understanding other points of view (Au et al., 2021; Innes et al., 2021).

In the public sector, researchers believe that information professionals have an opportunity to influence the decrease in the spread of mis-/dis-information (Cooke, 2017; Froehlich, 2017; Herrero-Diz & López-Rufino, 2021). Notably, librarians have a duty to become competent, fight personal biases, stay knowledgeable on current events and avoid violating information ethics or being ignorant to the issue of mis-/dis-information at hand (Froehlich, 2017). Gatekeeping of information must be avoided in an effort to promote individual and collective information literacy (N. K. Agarwal & Alsaeedi, 2020; Froehlich, 2017). Group media literacy would also neutralize echo chambers wherein it is already understood that people surround themselves with like-minded individuals because it requires less cognitive work and increases belonging (Ha et al., 2021; Wardle & Derakhshan, 2017). Having knowledge holders and leaders encourage the public to seek perspectives would counteract the artificial systems that social media platforms have created to capitalize on basic human behaviour to ensure more time spent and participation on their sites (Wardle & Derakhshan, 2017).

Digital literacy is another major contributor to the discrepancy in recognizing real and fake media. If people are not exposed to fake or hard news, they are less likely to be able to distinguish between fake and real news (Buchanan, 2020). Given the importance of digital literacy in younger generations to utilize in work and on a personal level, this phenomenon more acutely impacts older community members who have had less exposure to relevant technology. Without a widely accepted process for measuring digital literacy, it will be difficult to know how to educate populations at risk of mis-/dis-information (Buchanan, 2020). Digital media literacy is at the foundation of many anti-disinformation initiatives as exemplified in “News Hero”; a Facebook game developed by NATO Strategic Communications Centre of Excellence (<https://www.stratcomcoe.org/news-hero>) to use gamification as a method to enhance media literacy.

12.1.2 Building Community Resilience

Methods that combat misinformation and disinformation require different strategies when the factors of intention and facticity are considered. Disinformation is a systemic problem that is rooted in malicious intent and a desire to cause harm. At its core, disinformation thrives in societal divide and hate. In order to combat this, community-oriented strategies would be best deployed to unite online communities. Furthermore, efforts to understand and deradicalize polarized online communities would be fundamental in addressing the infodemic. Digital and media literacy should be tailored to individual users and demographics in order to be most effective (Carmi et al., 2020; Wang et al., 2021).

Some researchers would argue that ideological polarization is one of the most dangerous effects of fake news (Bran et al., 2021; Golebiewski & Boyd, 2018; Ha et al., 2021; Kumar & Shah, 2018). The closed system, or echo chamber, in which communication and news is repeated with limited opposing beliefs, excluding voices of moderation and dividing communities, creates a homogenous community which threatens democracy and social stability (Allcott & Gentzkow, 2017; Au et al., 2021; Shu et al., 2017). To counteract these effects, authors underscore the need for progressive public discussions and exposure to different ideological and political perspectives (Au et al., 2021; Carmi et al., 2020; Innes et al., 2021; Tandoc et al., 2018). This will create a precedent for building community resilience.

Raising awareness and improving societal resilience is imperative to creating a type of herd-immunity against the dissemination of fake news. The infodemic presented by mis-/dis-information acts as a virus that spreads quickly through digital media. Similarly to pandemics, a strong defence against mis-/dis-information is like-minded people who share a community-centred approach (Carmi et al., 2020; Choraś et al., 2021). The stronger the structure of a community, the more likely the members will be to support each other from understood on-coming threats (Johnson, 2022). The process known as inoculation will allow community members to recognize the threat of mis-/dis-information and protect community members from it (Carmi et al., 2020; Wang et al., 2021).

12.1.3 Communication and Gatekeeping

The lack of good communication from authorities and experts cannot be understated, especially at critical times (Innes et al., 2021; Mare et al., 2019; Melki et al., 2021; Moyo, 2009). Effective communication remains crucial for mitigating internal, external, or global threats to citizens (Innes et al., 2021; Mehta et al., 2021; Yang & Lee, 2020). Researchers have pointed out the decline of quality gatekeepers from traditional media outlets coinciding with the rise of social media and the internet (Au et al., 2021; Melki et al., 2021; Yang & Lee, 2020). The irony of quality checkers or gatekeepers being used by scientists and experts to disseminate information while anonymous internet users can peddle false information to the public without any verification whatsoever is the epitome of the information crisis (Melki et al., 2021). For this reason, gatekeeping is now expected to be the responsibility of public figures and people with more influential power because they have the potential to do more damage than a regular citizen when sharing information on social media (Au et al., 2021; Mehta et al., 2021). Researchers emphasize the need to close the communication gap and infiltrating online echo chambers through better communication between experts and the public in order to mitigate the effects of information pollution (Carmi et al., 2020; Choraś et al., 2021; Innes et al., 2021; Meel & Vishwakarma, 2020).

12.2 Recommendations for the Government

12.2.1 Funding and Legislative Support

There are reports that more and more citizens are relying on business to solve societal problems as they view their governments as incompetent (Edelman Trust Barometer, 2022). While this is great for business, it is a dangerous rhetoric for social welfare. Enhancing collaboration with online platforms, social networks and the tech industry is essential in addressing the current crisis; the government needs to have an active role in the private sector (Carmi et al., 2020; Choraś et al., 2021; Meel & Vishwakarma, 2020). Literature recommends implementing policy and laws that would discourage users from sharing or posting harmful content (Choraś et al., 2021; Meel &

Vishwakarma, 2020). Additionally, since most well-established fact-checking are based in the United States, increased funding and promotion for a Canadian based fact checking tool is needed. This tool would greatly benefit from combining the human element with automated tools in a hybrid approach. For example, sending the claim through a machine learning algorithm first to give the human expert more context in their classification decision Furthermore, legislation should regulate platforms from exploiting citizens and jeopardizing human rights (Carmi et al., 2020; Rubin, 2019) Governments must also update laws to be in line with the digital age and age of information, prioritizing their citizens and protecting them from private interests (Rubin, 2019).

An important feature of government action is to rally international efforts to combat mis-/dis-information through a global NGO or regulating body that would lead efforts in combat. This would require stronger cooperation and joint threat responses (Choraś et al., 2021; Lazer et al., 2017; Meel & Vishwakarma, 2020). This is especially important when analyzing mis/dis information through the lens of informational warfare since they tend to target societies and cultures across geographical areas, timezones, and subject matter (Songs et al., 2019). Local and national efforts to combat misinformation should include funding and research targeted towards understanding and connecting to local community leaders (Carmi et al., 2020; Lazer et al., 2017; Wang et al., 2021). Additionally, funding in the form of subsidies should be directed towards journalists and local media that would empower media outlets to rely more on fact-checking outlets, as well as encouraging local reporting through universities (Lazer et al., 2017).

12.2.2 Rebuilding Trust

Governments are the original disseminators of misinformation (Hamilton et al., 2021; Lazer et al., 2017; Marwick & Lewis, 2017). Over the years, this has deteriorated citizen's trust, especially for individuals from marginalized communities who have grown to feel cynical and alienated from society (Carmi et al., 2020; Songs et al., 2019; Wang et al., 2021). One general observation from the literature surveyed within the disciplines of data science, information science, and communications, is that the importance of government trust was not emphasized enough. Further research in the health domain allowed for the extraction of meaningful steps in rebuilding trust between authorities and communities.

The focus of these articles was vaccine uptake considering the on-going pandemic and the devastating effects that had on communities and nations (Melki et al., 2021; Songs et al., 2019; Wang et al., 2021). Recommendations begin at the grassroots level, calling on governments to partner with local institutions, experts, and leaders, especially when serving marginalized communities (Carmi et al., 2020; Melki et al., 2021; Songs et al., 2019; Wang et al., 2021). Furthermore, transparency, public engagement in policy making, and forums are fundamental for the long-term success of such initiatives (Carmi et al., 2020; Wang et al., 2021).

12.3 Policy Recommendations

12.3.1 Policy Mandate on the Use of Artificial Intelligence

AI is a critical tool currently being used and is under the development stage by social media giants such as Facebook to tackle disinformation. Technology has proven to be beneficial for scaling the work of human experts and, at the same time, has proven to reduce the action rate per potentially problematic post or comment intended to harm people. For example, Facebook has deployed SimSearchNet, an improved image matching model trained using self-supervised learning modes to match variations of an image with a very high degree of precision

(Ramasubramanian, 2020). It has been deployed as part of an end-to-end image matching system that runs on images uploaded on Facebook and Instagram (Ramasubramanian, 2020). This technology, in turn, can identify the variations in images and posts that have been re-designed to be circulated as disinformation (Ramasubramanian, 2020). However, these actions have been taken from a corporation's point of view, and legislations are yet to mandate the use of such technologies for the public social platform. Legislatures need to develop and draft a range of policies and mandate the use of AI-based products for dealing with disinformation on social media platforms. These may include warnings, the appointment of an ombudsman, and the use of plugins by third-party fact-checkers to keep a check on viral content. This, in turn, shall help reduce distribution and remove disinformation that may result in imminent harm.

12.3.2 Towards "Susceptibility by Design"

Technological companies must focus on innovations towards combating fake news prevalent on the media platform. Former FCC Commissioner Tom Wheeler stated that "public interest algorithms" can aid towards identifying and publicizing fake news posts and hence can prove to be an effective, valuable tool for protecting consumers (Wheeler, 2017). Algorithms have proven to be an effective vehicle for shaping up the web-based software and websites that are pre-engrained to detect disinformation posted on the website (Wheeler, 2017). It automates hoax detection and, at the same time, helps educate the consumers without censoring the content. This, in turn, proves to be effective in three manners; firstly, it creates a safe space for the online consumers on the content they read and informs them duly if the content has virality and its susceptibility to being a piece of fake news (Wheeler, 2017). Secondly, it also educates a user over the prevalence of fake news and data, which acts as a literacy program for the users (Wheeler, 2017). Lastly, it maintains the thin line between freedom of expression and freedom to raise opinions on the platform, but at the same time, it duly checks the effectiveness and trustworthiness of the content being posted on the platform (Wheeler, 2017). One such example of the effective use of the same can be reflected where several media platforms in the US, such as Wikipedia, have installed a "disputed news" tag on their websites that warn readers about controversial content (West, 2017). This could vary from information being outrightly false to where most parties have disagreed on its factualness. This way, it warns readers on possible threats and inaccuracies in the information being posted (West, 2017). Since it is a crowdsourced platform, Wikipedia has dealt with the problem by adding tags on the materials it has identified as "disputed news." Therefore, integrating such algorithms for media platforms must be mandated by insurance companies and governmental agencies first to protect the users from disinformation and secondly to provide a safe harbour to the media platform where it provides a space to the users for posting the desired results content. Hence, a public policy recommendation for the technological organization can yield results in favour of the entire user population. Further, this would enable private technological organizations and government to work on this issue together for carving out better results.

12.4 Recommendations for (Big) Tech

Even though researchers establish that the spread of mis-/dis-information is largely due to political and financial incentives, many place the onus of combating it on individuals rather than try to dissuade the perpetrators (Carmi et al., 2020). While this is effective for combating the effects on an individual level, it does not address factors that that an individual has no control over, such as online platforms' *dark patterns* (Carmi et al, 2020). Carmi et al. (2020) explain that dark patterns include involuntary and intrusive default settings, and misleading wording that give a false illusion of control (p.8). Wardle and Derakshan (2017) argue that tech companies have simply taken

advantage of human's social nature, other authors describe these dark patterns as a feature and result of the private sector's creative pursuits – in other words, the cost of having an online presence (Waldman, 2020). However, much of the literature argues of the dangers of echo chambers and filter bubbles (Chen et al., 2021; Ha et al., 2021; Zhang & Ghorbani, 2020) as well as other features that reduce human autonomy and infringe on privacy (Carmi et al., 2020; Waldman, 2020). The effects of echo chambers, human or digital, are well-researched in their effects on ideological polarization and radicalization (Au et al., 2021; Innes et al., 2021; Pierre, 2020).

Despite the well documented exploitation of citizens by big Tech, some authors consider the effect of social media algorithms on the spread of fake news to be overestimated by (Au et al., 2021; Vosoughi et al., 2018) thus they may consider more emphasis or responsibility on the citizen rather than the social media platform. However, other authors emphasize the need for a community-based approach to fighting the infodemic (Abdalla et al., 2021; Carmi et al., 2020; Mclane, 2021; Songs et al., 2019). Rubin (2019) argues that consequences of misinformation could be resolved through a framework of education, autodetection, and regulation. Limitations in the literature on recommendations lacked research on the effectiveness of recommendations. With regards to information voids and information overload, there was insufficient data available to dissect (Bran et al., 2021). Further areas for exploration should examine open-source data tools, blockchain technology, strategies for the development of digital literacy programs, and individual security (Bran et al., 2021).

12.4.1 Cross-Disciplinary Recommendations for Big Tech's Considerations

Using the framework of an “infodemic,” authors (Hamilton et al., 2021; Meel & Vishwakarma, 2020; Rubin, 2019) propose tackling the spread of misinformation through structured networks by addressing various stages of the spread with questions such as:

1. Duration: How long does it take before the item is removed/debunked?
2. Opportunities: How visible is the news item? What is the size of the audience?
3. Transmission probability: how likely are people likely to share the item?
4. Susceptibility: how likely are people to believe the item?

By analyzing the spread of rumors spread on online networks in the same light of viral disease spread, they propose the infiltration of echo chambers through algorithms that identify the most influential nodes and planting an anti-rumour propagator (Choraś et al., 2021; Meel & Vishwakarma, 2020). When considering the effect of repeated exposure on belief, it becomes especially important to reduce the repetition of messages in an individual's feed (Buchanan, 2020; Lazer et al., 2017). To expand further, some of the tactics include the spread of anti-rumours and the potential to block rumours from influential users (Au et al., 2021; Meel & Vishwakarma, 2020; Zannettou et al., 2019). Researchers note the level of human intervention needed in fact-checking and detection methods, whether through independent methods or crowdsourcing (Figueira & Oliveira, 2017; Shu et al., 2017; Zannettou et al., 2019). Credibility-based rankings can be done with user support, however, some authors note lack of papers that detect fake news using social aspect features such as user popularity and influence (Shu et al., 2017; Zannettou et al., 2019). The emphasis of user influence ties in with gatekeeping efforts since it is agreed that a small user's account impact should not be weighed against the impact of that of a social influence or public figure (Zannettou et al., 2019).

Some of the current efforts involve Google disabling ad revenue to disinformation spreaders, Facebook working with fact checkers, improving ranking system, and removing trolls and fake accounts (N. K. Agarwal & Alsaedi, 2020; Wardle & Derakhshan, 2017). However, much of the discussion in the literature revolves around social media tactics to combat misinformation (Au et al., 2021; Meel & Vishwakarma, 2020; Shu et al., 2017; Zannettou et al., 2019). Social media strategies could be divided into linguistic based (lexical or syntax based), visual based (images and videos such as those used by propaganda), social aspect based (number of followers or percent of verified users), network based (user engagements/friendship or co-occurrence network), or post based (stance, topic, credibility) (Shu et al., 2017; Zannettou et al., 2019).

Researchers believe that online platforms should be able to actively close down fake accounts, identify bots and label them, as well as collaborate with national regulators and independent fact-checkers (Choraś et al., 2021; Meel & Vishwakarma, 2020). Social media platforms should enable community signals, user flags, and expert opinions to assist with detection methods (Meel & Vishwakarma, 2020; Zannettou et al., 2019). However, some authors note a lack of information regarding early detection of information that has already been disseminated, especially because some of the data sets studies cannot be generalized to all types of misinformation (Au et al., 2021; Guo et al., 2016; Zannettou et al., 2019). Tech companies could also work with researchers and scientists who could help create profiles based on personality for media literacy or counter-messaging efforts (Buchanan, 2020; Carmi et al., 2020). In the sphere of detection for example, Ghanem et al. (2020) found that different types of fake news have different emotional patterns, establishing that emotions play a role in deceiving the reader, or even evoking an engagement with the message. It is worth investigating and tracking how users on social media engage with their daily feeds when scrolling through social media or examining a specific news article (Ghanem et al., 2020). Drawing on works by Ansar and Goswami (2021), emotional lexicon can further be used to categorize message content and cross referenced with predicted social response to help with early detection methods.

12.4.2 Detection

The first step in detecting disinformation involves educating the targets of the operations. Simply identifying posts as fake news has proven not to be sufficient to dissuade individuals from believing disinformation as there is a tendency to believe what they can read and see in images (Baptist & Gluck, 2021). There is a need to educate society to increase awareness of the tactics and types of manipulation that are being used against the public in order to better equip them to identify these types of disinformation campaigns and can help temper emotional responses (Baptist & Gluck, 2021; Ivan et al., 2021). TrollHunter2020 has been designed with the purpose of early detection of trending disinformation campaigns and their development on Twitter (Jachim et al., 2021). Trolling strategies have evolved and as such the detection systems must evolve as well. This system will assist in the detection of present or emerging trends indicated by associated accounts or through state sponsorship (Jachim et al., 2021). TrollHunter2020 is an impressive advancement because it only requires a dataset of a few dozen to a couple of hundred tweets (Jachim et al., 2021). There are ethical concerns and possibilities for abuse when targeting the TrollHunter2020 system when there is a loose definition of what is considered to be “alternative, distorted or trolling narratives” (Jachim et al., 2021, p. 9). Additional improvements and the implementation of safeguards to prevent malign abuse should be developed to control the distinction between what is determined to be false narratives and what is simply trending news. There are two forecasted results of the application of TrollHunter2020 to Twitter. The first is that this system will force trolls to more peripheral fringe social media to avoid the system, or the inverse will occur, and it will draw trolls from fringe social media to exploit the limitations of the system when they are learned (Jachim

et al., 2021). Further research is needed to overcome the limitations of this model to improve the detection of bot run coordinated accounts.

Another recommendation suggested that includes studying both conspiracy theories and real conspiracies through the application of automated machine learning techniques to assess the structure of the narratives (Moffitt et al., 2021). The theory suggests that identifying the differences between the narratives would enhance and assist in the detection of new disinformation campaigns. The BERT-based model (Bidirectional Encoder Representations from Transformers) is suggested as a useful model for distinguishing between the two types of narratives (Moffitt et al., 2021). Some of the promising qualities of the BERT-based model include its scalability, and the efficiency and effectiveness in categorizing theories. This suggests that additional investigation is required into existing conspiracies and conspiracy theories to further understand their unique features to properly identify new narratives and counteract their attacks (Moffitt et al., 2021). Zannettou (2019) focuses on machine learning and detection combat strategies for each type of fake news type (rumour, hoax, etc). While this is effective for detection, a study on people's perception of the fake news type and self-efficacy, demonstrate that the type of fake news has null impact on their confidence in identifying it, with the exception of parody and satire (Hinsley & Holton, 2021). This underscores the importance of well-trained machine learning techniques as well as the value of understanding human psyche in the development of detection tools since for example, overwhelming individuals with the different terms of fake news may not be effective when creating a media literacy strategy.

There is also a recommendation to further develop a method of identifying disinformation campaigns with coordinated accounts without previously identifying any part of the accounts involved in the campaign. The AMDN-HAGE (Attentive Mixture Density Network with Hidden Account Group Estimation) is a model to, "capture inherent characteristics of coordination which is, accounts that coordinate must strongly influence each other's activities, and collectively appear anomalous from normal accounts" (Sharma et al., 2021, p. 1). Two of the focal points of the model will be identifying strong hidden influence, and highly concerted activities to identify coordinated accounts. This recommended method of using coordinated accounts to identify disinformation campaigns would be independent of platform specific features, linguistics, and metadata. This will allow the model to be generalized to act as a universal model across languages and countries (Sharma et al., 2021). While coordinated bot accounts can be identified and tracked, disguised human driven accounts are much more difficult to identify and requires the cooperation and coordination of social media companies (Farkas & Bastos, 2018). This indicates that identifying linked accounts should only be one of the focus areas for improving the detection of disinformation campaigns and additional research into the detection of disinformation campaigns is needed.

12.4.3 Protection

There is a call for the development of combining misinformation with information security to create MisinfoSec WG (Working Group) (Walker et al., 2019). The combination of these two facets will facilitate the creation a structured framework for understanding targeted misinformation attacks (Walker et al., 2019). There is even discussion that the definition of cybersecurity should be broadened to include mis/disinformation operations (Walker et al., 2019). There are many areas of overlap between the two topics and there are already systems in place for InfoSec that would provide a starting point for adaption to misinformation detection requirements (Walker et al., 2019). An example of this overlap can be seen in Advanced Persistent Threats (ATP) which are a long duration campaign perpetrated by an attacker that is often a nation state, which is comparable to a misinformation campaign (Walker et al., 2019). It is recommended that a list of best practices and threats be compiled in regard to the disinformation environment and institutional requirements to create a framework to apply tested and applicable security principles (Walker et al., 2019). By creating this framework, it will assist in identifying gaps and future requirements including

planning defenses, planning countermoves, and creating assessment tools (Walker et al., 2019). Further mapping of frameworks and testing are required before implementation is feasible (Walker et al., 2019).

12.4.4 Social Media Campaigns and Gaming

Social Media campaigns and gaming to tackle the problem of online disinformation, social media developed an online game called "Go Viral!" (gusmanson.nl, n.d.). The game aimed to spread awareness amongst Facebook users on how disinformation spreads and affects their decision-making capacity. Researchers found that such games and infographics can train people to spot untrustworthy information and develop critical thinking abilities amongst the users to identify disinformation (van der Linden et al., 2020).

The game was designed with the idea of building attitudinal resistance against disinformation through pre-emptively 'prebunking' of disinformation. Prebunking is based on a key component of inoculation theory (Eagly and Chaiken, 1993). It is based upon psychological inoculating of the mind, similar to the biological immunization process (McGuire, 1964). Research studies have provided cross-cultural evidence on the effectiveness of prebunking interventions in order to reduce susceptibility to disinformation (Basol et al., 2021). Go Viral! has been found to positively impacts players' ability on identifying disinformation (Basol et al., 2021). Research reflected that similar use of prebunking games and training improved players' confidence for detecting disinformative content and reduced willingness to share disinformation (Basol et al., 2021). Hence prebunking can be utilized further to mitigate the spread of disinformation as it can both minimize the spread as well as train the users for dealing with disinformation.

Prebunking, though, has been utilized by social media organizations, the concept should not be limited to only such organizations. Prebunking should also be utilized in the form of training as a part of academic curriculums, corporate trainings and general public training through government awareness camps. This would help train the population time and again, which would not only increase awareness about disinformation but also train the minds from falling into the loop of disinformation.

12.4.5 Fact-Checking Tool Interface and Design

Throughout this evaluation, it became evident that the tools oftentimes had no clear display of the process by which they analyse information and achieve the data or information presented. More transparency is needed, as well as better product marketing considering none of the participants were aware that such tools exist online, before the evaluation began. Additionally, while we recognize that there are multiple terms and synonyms used to refer to forms of disinformation, the creators of such tools should include this when building the SEO parameters of their sites. This would increase the chances of their tools being found and used. Tools should also be easy to identify visually and have working widgets. The need to set up an account was found to be hugely inhibiting to users positive perception of tools suggesting that future countering tools should not require the creation of a membership based account for access. Search capabilities should also be improved in order to facilitate the user's search on the website. Alternatively, a query format structure could be advertised on the site for ease of use. Instructions on how to use extensions, widgets, and the site are considered highly valuable to users and would promote further use of the tools.

Users appreciate visual dispersion charts and color coding when using misinformation countering tools, suggesting that the visual aspect and ease of comprehension are key factors in the success of these tools. Furthermore, an explanation behind the detection algorithm helps reassure

users of its credibility. Despite the humour found by some users in some sites' determination of facticity using terms such as "quackery" and "tin foil hat level", others found these indicators potentially dismissive and belittling to those who may have aligned views. This would create a barrier for use by some individuals. Less biased language would be easier to understand, overcome cultural barriers, and not be perceived negatively.

12.4.6 Use of Blockchain Technology

Blockchain is a disruptive technology, revolutionary in its nature and disintermediation capabilities. Yet, it is perhaps one of the most misunderstood buzzwords of the decade, with a fervent demand for implementation in business use cases which may not truly benefit from it. A succinct yet complete definition of blockchain is: "a decentralized, sequential database containing cryptographically linked blocks of digitally signed asset transactions, governed by a consensus model" (Sultan, Ruhi, & Lakhani, 2018). Blockchain to combat disinformation can be used to certify the source, as opposed to the content. The proposed goal with blockchain is to identify the source, and to apply one or more metrics against that source to decide whether to trust this author in their efforts. *Table 28* below explains how various metrics could be used to target and combat online disinformation:

Table 28. Metrics to Target and Combat Online Disinformation

Metric	Use case:
Proof of Authorship	A known entity (exposed public key) publishes an article. They first run the article through a smart contract, in which they provide a hash of the meta-data. A 3rd party can recreate this hash using the same metadata attributes and see who wrote the article. The content can not be validated against the author; but the author can be validated. At this point, reputation can be relied upon
Reputation	A smart contract can be used to provide an author's name and a validity score from 0 to 1 inclusive. The interested reader can run the author's name and retrieve a validity score. When reputation services are coupled with Proof of Authorship, the service becomes even more reliable.
Certification	An article is published. The author certifies the document on the date of publication. The document's secure hash is calculated and sent to a smart contract, with the document's title. The secure hash is replicable. A reader can replicate the secure hash and then query the blockchain for it (via the smart contract) thus determining if the document is certified. This allows the reader to determine if the content is from a trusted source.
Proof of Origin/Proof of Authenticity	Proof of Origin can be determined by using Certification with Proof of Authorship and Reputation Services. An article is published. The agent publishes the public key of the author. The author first Certifies the document on the blockchain. Next, the author uses Proof of Authorship with the documents meta-data to establish authorship rights. Finally, the author can be looked up via reputation service.

Furthermore, this information can be saved to the blockchain, making it available to the next reader. Applying this to any articles of misinformation can be an effective method in combating misinformation globally, and on a fast scale. Blockchain can be used in almost any sector that has digital assets, has a requirement for trust and disintermediation, has a need for cryptocurrency or tokens for payment and incentive, or requires the execution of logical contracts to govern transactions, making it applicable on a global scale (Glaser, 2017; Sultan et al., 2018; Yaga et al., 2018). Deloitte identified core domains of application including cybersecurity, identification, compliance, financial reporting, and digital asset management. Deloitte concluded that "while

blockchain was once classified as a technology experiment, it now represents a true agent of change that is affecting the entire organization” (Deloitte, 2020). The effect of having decentralized copies of each node enforces the characteristic of availability: Blockchain has guaranteed availability (McConaghy et al., 2016; Yaga et al., 2018). This poses an important question of Blockchain. If all nodes are equal, then how are decisions made? Blockchain achieves this through a consensus model, which guarantees agreement between 'n' nodes in a finite time (McConaghy et al., 2016; Yaga et al., 2018). This is perhaps Blockchain's most important achievement. Using public key infrastructure (PKI), Blockchain can form trusted results (Mougayar, 2016). This has led to it being called a trust-machine and is the key to its disruptive potential.

The combination of digital signatures, integrity hashes and archival blocks creates data immutability (McConaghy et al., 2016; Sultan et al., 2018). Previous entries in a blockchain cannot be changed without triggering an instantly noticeable cascade of errors in all subsequent blocks. This underlines the importance of participants maintaining their own nodes; they will be able to detect modification attempts immediately and reject altered blocks (Glaser, 2017; Sultan et al., 2018; Yaga et al., 2018). Due to increased node participation, mature public blockchains are highly resistant to attacks, and thus manage to achieve exceptional reliability (Glaser, 2017; McConaghy et al., 2016; Yaga et al., 2018). Private and Consortium hybrid models remain at risk if one participant controls most of the computing power.

13. Future Research: Digital Disinformation Issues and Behaviors

Survey

Building upon the findings of this research, and to further understand how and why people take corrective actions towards dis/misinformation by engaging actively on social media, we built a survey to further understand the factors that impact people's responses to misinformation or to what could also be referred to as hotly debated and controversial topics. These are topics for which people might have different views and beliefs, and there may be disagreements about the accuracy or reliability of information.

The purpose of this survey is to investigate users' perceptions, attitudes and behaviours towards misinformation and disinformation online, especially within the context of social media use. The findings from this research are expected to identify potential gaps in current individual, technological and organizational practices for the mitigation of online misinformation and disinformation, and to provide recommendations that can help foster societal resilience to these threats. A full copy of the survey questions is available under Appendix I.

Using data collected from the literature review, constructs were established to be studied in this survey. The survey was then built to encompass six main themes and constructs, namely Sociological Environment Perceptions, Information Ecosystem Perceptions, Cognitive Perceptions, Information Seeking Behavior, Affect & Attitude, and Disinformation Countering Responses. The survey was then circulated to friends and family for pilot testing and is now ready to be released to the public for data collection, given the funding availability.

14. References

- “Disinformation” vs. “Disinformation”: Get Informed On The Difference—Dictionary.com. (n.d.). Retrieved February 24, 2022, from <https://www.dictionary.com/e/disinformation-vs-disinformation-get-informed-on-the-difference/>
- Abdalla, S. M., Koya, S. F., Jamieson, M., Verma, M., Haldane, V., Jung, A. S., Singh, S., Nordström, A., Obaid, T., Legido-Quigley, H., & McNab, C. (2021). Investing in trust and community resilience: Lessons from the early months of the first digital pandemic. *The BMJ*, 375. <https://doi.org/10.1136/bmj-2021-067487>
- Agarwal, N. K., & Alsaeedi, F. (2020). Understanding and fighting disinformation and fake news: Towards an information behavior framework. *Proceedings of the Association for Information Science and Technology*, 57(1). <https://doi.org/10.1002/pra2.327>
- Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2019). *Protecting World Leaders Against Deep Fakes*.
- Akers, J., Bansal, G., Cadamuro, G., Chen, C., Chen, Q., Lin, L., Mulcaire, P., Nandakumar, R., Rockett, M., Simko, L., Toman, J., Wu, T., Zeng, E., Zorn, B., & Roesner, F. (2018). *Technology-Enabled Disinformation: Summary, Lessons, and Recommendations*. <http://arxiv.org/abs/1812.09383>
- Akers, J., Bansal, G., Cadamuro, G., Chen, C., Chen, Q., Lin, L., Mulcaire, P., Nandakumar, R., Rockett, M., Simko, L., Toman, J., Wu, T., Zeng, E., Zorn, B., & Roesner, F. (2019). *Technology-Enabled Disinformation: Summary, Lessons, and Recommendations* (arXiv:1812.09383). arXiv. <http://arxiv.org/abs/1812.09383>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. In *Journal of Economic Perspectives* (Vol. 31, Issue 2, pp. 211–236). American Economic Association. <https://doi.org/10.1257/jep.31.2.211>
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Ansar, W., & Goswami, S. (2021). Combating the menace: A survey on characterization and detection of fake news from a data science perspective. *International Journal of Information Management Data Insights*, 1(2), 100052. <https://doi.org/10.1016/j.ijime.2021.100052>
- Au, C. H., Ho, K. K. W., & Chiu, D. K. W. (2021). The Role of Online Misinformation and Fake News in Ideological Polarization: Barriers, Catalysts, and Implications. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10133-9>
- Baccarella, C. V., Wagner, T. F., Kietzmann, J. H., & McCarthy, I. P. (2018). Social media? It’s serious! Understanding the dark side of social media. *European Management Journal*, 36(4), 431–438. <https://doi.org/10.1016/j.emj.2018.07.002>
- Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake News, Fast and Slow: Deliberation Reduces Belief in False (but Not True) News Headlines. *Journal of Experimental Psychology. General*, 149(8), 1608–1613. <https://doi.org/10.1037/xge0000729>
- Balmas, M. (2014). When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism. *Communication Research*, 41(3), 430–454. <https://doi.org/10.1177/0093650212453600>
- Balmas, M. (n.d.). When Fake News Becomes Real. *Communication Research*, 25.
- Bamberger, L. (2019). Philip N. Howard/Bharath Ganesh/Dimitra Liotsiou/John Kelly/Camille François: The IRA, Social Media and Political Polarization in the United States, 2012–2018. Oxford: Oxford University Project on Computational Propaganda, Working Paper, 2018. *SIRIUS: Zeitschrift Für Strategische Analysen*, 3(1), 92–92. <https://doi.org/10.1515/sirius-2019-1017>
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, 4(3), 114–123.

- Barzilay, R., Moore, T. M., Greenberg, D. M., DiDomenico, G. E., Brown, L. A., White, L. K., Gur, R. C., & Gur, R. E. (2020). Resilience, COVID-19-related stress, anxiety and depression during the pandemic in a large population enriched for healthcare providers. *Translational Psychiatry*, 10(1), 1–8. <https://doi.org/10.1038/s41398-020-00982-4>
- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 disinformation. *Big Data & Society*, 8(1), 20539517211013868. <https://doi.org/10.1177/20539517211013868>
- Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*, 116. <https://doi.org/10.1016/j.chb.2020.106633>
- Bastos, M., & Farkas, J. (2019). “Donald Trump Is My President!”: The Internet Research Agency Propaganda Machine. *Social Media and Society*, 5(3), 1–13. <https://doi.org/10.1177/2056305119865466>
- Bastos, M., & Farkas, J. (2019). “Donald Trump Is My President!”: The Internet Research Agency Propaganda Machine. *Social Media + Society*, 5(3), 205630511986546. <https://doi.org/10.1177/2056305119865466>
- Beck, A. T. (1967). *Depression; clinical, experimental, and theoretical aspects*. Hoeber,.
- Bless, H., & Burger, A. M. (2017). Mood and the Regulation of Mental Abstraction. *Current Directions in Psychological Science : A Journal of the American Psychological Society*, 26(2), 159–164. <https://doi.org/10.1177/0963721417690456>
- Bless, H., & Fiedler, K. (2006). Mood and the regulation of information processing and behavior. *Undefined*. <https://www.semanticscholar.org/paper/Mood-and-the-regulation-of-information-processing-Bless-Fiedler/1efeea85eb915ef8a5e481883a36334c770a262d>
- Bless, H., Schwarz, N., & Kemmelmeier, M. (1996). Mood and Stereotyping: Affective States and the Use of General Knowledge Structures. *European Review of Social Psychology*, 7(1), 63–93. <https://doi.org/10.1080/14792779443000102>
- Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology*, 24(1), 45–62. <https://doi.org/10.1002/ejsp.2420240104>
- Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, 497, 38–55. <https://doi.org/10.1016/j.ins.2019.05.035>
- Bradshaw, S., & Howard, P. (2019). The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation. *Copyright, Fair Use, Scholarly Communication, Etc*. <https://digitalcommons.unl.edu/scholcom/207>
- Bradshaw, S., & Howard, P. N. (n.d.). *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. 26.
- Bran, R., Tiru, L., Grosseck, G., Holotescu, C., & Malita, L. (2021). Learning from each other—a bibliometric review of research on information disorders. In *Sustainability (Switzerland)* (Vol. 13, Issue 18). MDPI. <https://doi.org/10.3390/su131810094>
- Brandtzaeg, P. B., Følstad, A., & Chaparro Domínguez, M. Á. (2018). How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services. *Journalism Practice*, 12(9), 1109–1129. <https://doi.org/10.1080/17512786.2017.1363657>
- Brooke, J. (1996). SUS: A “Quick and Dirty” Usability Scale. In *Usability Evaluation In Industry* (pp. 207–212). <https://doi.org/10.1201/9781498710411-35>
- Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLoS ONE*, 15(10 October). <https://doi.org/10.1371/journal.pone.0239666>

- Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLoS ONE*, 15(10), e0239666. <https://doi.org/10.1371/journal.pone.0239666>
- Canadian Opinion Research. (n.d.). Angus Reid Institute. Retrieved February 26, 2022, from <https://angusreid.org/>
- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, 9(2). <https://doi.org/10.14763/2020.2.1481>
- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, 9(2). <https://doi.org/10.14763/2020.2.1481>
- Carver, C. S., & Harmon-Jones, E. (2009). Anger Is an Approach-Related Affect: Evidence and Implications. *Psychological Bulletin*, 135(2), 183–204. <https://doi.org/10.1037/a0013965>
- Centre for Countering Digital Hate. (2021). The Disinformation Dozen. In *Cent*.
- Chadwick, A., & Stanyer, J. (2022). Deception as a Bridging Concept in the Study of Disinformation, Misinformation, and Misperceptions: Toward a Holistic Framework. *Communication Theory*, 32(1), 1–24. <https://doi.org/10.1093/ct/qtab019>
- Chadwick, A., & Stanyer, J. (2022). Deception as a Bridging Concept in the Study of Disinformation, Misinformation, and Misperceptions: Toward a Holistic Framework. *Communication Theory*, 32(1), 1–24. <https://doi.org/10.1093/ct/qtab019>
- Chaiken, S., & Ledgerwood, A. (2011). A theory of heuristic and systematic information processing. *Handbook of theories of social psychology: Volume one*, 246-166.
- Chen, X., Zhou, F., Trajcevski, G., & Bonsangue, M. (2022). Multi-view learning with distinguishable feature fusion for rumor detection. *Knowledge-Based Systems*, 240. <https://doi.org/10.1016/j.knosys.2021.108085>
- Chen, X., Zhou, F., Zhang, F., & Bonsangue, M. (2021). Catch me if you can: A participant-level rumor detection framework via fine-grained user representation learning. *Information Processing and Management*, 58(5). <https://doi.org/10.1016/j.ipm.2021.102678>
- Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PLoS ONE*, 5(11). <https://doi.org/10.1371/journal.pone.0014118>
- Choraś, M., Demestichas, K., Giełczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., Urda, D., & Woźniak, M. (2021). Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. *Applied Soft Computing*, 101. <https://doi.org/10.1016/j.asoc.2020.107050>
- Constine, J. (2019). *Zuckerberg on Chinese censorship: Is that the internet we want?* TechCrunch+. <https://techcrunch.com/2019/10/17/zuckerberg-speech/>
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS ONE*, 12(5). <https://doi.org/10.1371/journal.pone.0175799>
- Cooke, N. A. (2017). Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age. In *Library Quarterly: Information, Community, Policy* (Vol. 87, Issue 3). <http://www.oed.com>.
- Davey, J., Hart, M., & Guerin, C. (2020). *An Online Environmental Scan of Right-wing Extremism in Canada*. Institute for Strategic Dialogue (ISD).
- De Coninck, D., Frissen, T., Matthijs, K., d'Haenens, L., Lits, G., Champagne-Poirier, O., Carignan, M.-E., David, M. D., Pignard-Cheynel, N., Salerno, S., & Généreux, M. (2021). Beliefs in Conspiracy Theories and Disinformation About COVID-19: Comparative Perspectives on the Role of Anxiety, Depression and Exposure to and Trust in Information Sources. *Frontiers in Psychology*, 12, 646394–646394. <https://doi.org/10.3389/fpsyg.2021.646394>

- Deloitte. (2020). *Deloitte's 2020 Global Blockchain survey. From promise to reality*. Retrieved from https://www2.deloitte.com/content/dam/insights/us/articles/6608_2020-global-blockchain-survey/DI_CIR_2020_global_blockchain_survey.pdf
- Desurvire, H. W. (1994). Faster, Cheaper!! Are Usability Inspection Methods as Effective as Empirical Testing? In J. Nielsen (Ed.), *Usability Inspection Methods* (pp. 173–202). John Wiley & Sons, Inc. <https://doi.org/10.5555/189200.189217>
- Digital Democracy Project. (2020). *LESSONS IN RESILIENCE*. www.ppforum.ca
- Domenico, G. D., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124, 329–341. <https://doi.org/10.1016/j.jbusres.2020.11.037>
- Domenico, G. di, Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. In *Journal of Business Research* (Vol. 124, pp. 329–341). Elsevier Inc. <https://doi.org/10.1016/j.jbusres.2020.11.037>
- Dumas, J. S., & Fox, J. E. (2009). Usability Testing: Current Practice and Future Directions. In A. Sears & J. A. Jacko (Eds.), *Human-Computer Interaction* (1st Editio, p. 22). CRC Press.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes* (pp. xxii, 794). Harcourt Brace Jovanovich College Publishers.
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of disinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Edelman Trust Barometer*. (2022). <https://www.edelman.com/trust/2022-trust-barometer>
- Evans, J. S. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10), 454–459. <https://doi.org/10.1016/j.tics.2003.08.012>
- Excel Spreadsheet for Calculating SUS Scores*. (2008). Measuring UX. https://measuringux.com/sus/SUS_Calculation.xlsx
- Facebook MAU worldwide 2021*. (n.d.). Statista. Retrieved April 19, 2022, from <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Fallis, D. (2015). What is disinformation? *Library Trends*, 63(3), 401–426. <https://doi.org/10.1353/lib.2015.0014>
- Fallis, D. (2015). What Is Disinformation? *Library Trends*, 63(3), 401–426. <https://doi.org/10.1353/lib.2015.0014>
- Figueira, Á., & Oliveira, L. (2017). The current state of fake news: Challenges and opportunities. *Procedia Computer Science*, 121, 817–825. <https://doi.org/10.1016/j.procs.2017.11.106>
- Fisher, N. A. K., Karen E. (2013, March 15). *A social diffusion model of disinformation and disinformation for understanding human information behaviour* [Text]. Professor T.D. Wilson. <http://informationr.net/ir/18-1/paper573.html#.YhpG4pPMK3I>
- Flores-Saviaga, C., & Savage, S. (2019). *ANTI-LATINX COMPUTATIONAL PROPAGANDA IN THE UNITED STATES*.
- Forgas, J. P. (1995). Mood and Judgment: The Affect Infusion Model (AIM). *Psychological Bulletin*, 117(1), 39–66. <https://doi.org/10.1037/0033-2909.117.1.39>
- Forgas, J. P. (1998). On Being Happy and Mistaken: Mood Effects on the Fundamental Attribution Error. *Journal of Personality and Social Psychology*, 75(2), 318–331. <https://doi.org/10.1037/0022-3514.75.2.318>
- Forgas, J. P. (2017). Chapter 3 - Mood Effects on Cognition: Affective Influences on the Content and Process of Information Processing and Behavior. In *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 89–122). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-801851-4.00003-3>

- Forgas, J. P. (2019). Happy Believers and Sad Skeptics? Affective Influences on Gullibility. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, 28(3), 306–313. <https://doi.org/10.1177/0963721419834543>
- Forgas, J. P., & Bower, G. H. (1987). Mood Effects on Person-Perception Judgments. *Journal of Personality and Social Psychology*, 53(1), 53–60. <https://doi.org/10.1037/0022-3514.53.1.53>
- Foster, C. L. E. (2022). Truth as social practice in a digital era: Iteration as persuasion. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-021-01306-w>
- Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J. R., Olin, J., Ross, M. Q., & Szafran, J. (2021). Fake Claims of Fake News: Political Misinformation, Warnings, and the Tainted Truth Effect. *Political Behavior*, 43(4), 1433–1465. <https://doi.org/10.1007/s11109-020-09597-3>
- Froehlich, T. J. (2017). A Not-So-Brief Account of Current Information Ethics: The Ethics of Ignorance, Missing Information, Misinformation, Disinformation and Other Forms of Deception or Incompetence. *Textos Universitaris de Bibliotecomia i Documentacio*, 39. <http://temaria.net/simple.php?origen=1575-5886&idioma=en>
- Gamboa, L., & Cartes, P. (2017). *Twitter's Contribution to Media Literacy Week*. Twitter Blog. Investing in U.S. Media Literacy Initiatives
- Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, 38(1), 84–117. <https://doi.org/10.22329/il.v38i1.5068>
- Ghanem, B., Rosso, P., & Rangel, F. (2020). An Emotional Analysis of False Information in Social Media and News Articles. *ACM Transactions on Internet Technology*, 20(2). <https://doi.org/10.1145/3381750>
- Ghanem, B., Rosso, P., & Rangel, F. (2020). An Emotional Analysis of False Information in Social Media and News Articles. *ACM Transactions on Internet Technology*, 20(2), 1–18. <https://doi.org/10.1145/3381750>
- Giachanou, A., Ghanem, B., & Rosso, P. (2021). Detection of conspiracy propagators using psycholinguistic characteristics. *Journal of Information Science*. <https://doi.org/10.1177/0165551520985486>
- Giachanou, A., Ghanem, B., Ríssola, E. A., Rosso, P., Crestani, F., & Oberski, D. (2022). The impact of psycholinguistic patterns in discriminating between fake news spreaders and fact checkers. *Data and Knowledge Engineering*, 138. <https://doi.org/10.1016/j.datak.2021.101960>
- Gilbert, D. T. (1991). How Mental Systems Believe. *The American Psychologist*, 46(2), 107–119. <https://doi.org/10.1037/0003-066X.46.2.107>
- Glaser, F. (2017). Pervasive Decentralisation of Digital Infrastructures: A Framework for Blockchain enabled System and Use Case Analysis. *50th Hawaii International Conference on System Sciences*, 1543–1552. <https://doi.org/10.1145/1235>
- Gleicher, N. (2018). *Coordinated Inauthentic Behavior Explained*. Meta Newsroom. <https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>
- Golebiewski, M., & Boyd, D. (2018). *Data Voids: Where Missing Data Can Easily Be Exploited*. <https://datasociety.net/about/#funding>.
- Gorwa, R., & Guilbeault, D. (2020). Unpacking the Social Media Bot: A Typology to Guide Research and Policy. *Policy & Internet*, 12(2), 225–248. <https://doi.org/10.1002/poi3.184>
- Gradoń, K. T., Hołyst, J. A., Moy, W. R., Sienkiewicz, J., & Suchecki, K. (2021). Countering misinformation: A multidisciplinary approach. *Big Data and Society*, 8(1). <https://doi.org/10.1177/20539517211013848>
- Grady, R. H., Ditto, P. H., & Loftus, E. F. (2021). Nevertheless, partisanship persisted: fake news warnings help briefly, but bias returns with time. *Cognitive research: principles and implications*, 6(1), 1-16.

- Greenstein, M., & Franklin, N. (2020). Anger Increases Susceptibility to Disinformation. *Experimental Psychology*, 67(3), 202–209. <https://doi.org/10.1027/1618-3169/a000489>
- Groshek, J., & Bronda, S. (n.d.). *How social media can distort and misinform when communicating science*. The Conversation. Retrieved April 19, 2022, from <http://theconversation.com/how-social-media-can-distort-and-misinform-when-communicating-science-59044>
- Guo, B., Ding, Y., Yao, L., Liang, Y., & Yu, Z. (2016). The Future of Misinformation Detection: New Perspectives and Trends. *Journal of the ACM*, 2657, 1–9. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Gusmanson.nl. (n.d.). *Play GO VIRAL! | Stop Covid-19 disinformation spreading*. Go Viral! Retrieved April 19, 2022, from <https://www.goviralgame.com/books/go-viral/>
- Ha, L., Andreu Perez, L., & Ray, R. (2021). Mapping Recent Development in Scholarship on Fake News and Misinformation, 2008 to 2017: Disciplinary Contribution, Topics, and Impact. *American Behavioral Scientist*, 65(2), 290–315. <https://doi.org/10.1177/0002764219869402>
- Hamilton, K., Božić, B., & Longo, L. (2021). *Interrupting the Propaganda Supply Chain*. <https://www.tanbih.org/prta>
- Harari, Y. N. (2018). *Sapiens: A brief history of humankind* (First Harper Perennial edition.). Harper Perennial.
- Herrero-Diz, P., & López-Rufino, C. (2021). Libraries fight disinformation: An analysis of online practices to help users' generations in spotting fake news. *Societies*, 11(4). <https://doi.org/10.3390/soc11040133>
- Hinsley, A. (2021). *Fake News Cues: Examining the Impact of Content, Source, and Typology of News Cues on People's Confidence in Identifying Mis- and Disinformation*. 20.
- Hinsley, A., & Holton, A. (2021). Fake News Cues: Examining the Impact of Content, Source, and Typology of News Cues on People's Confidence in Identifying Mis-and Disinformation. In *International Journal of Communication* (Vol. 15). <http://ijoc.org>.
- Horne, B. D., & Adali, S. (2017). This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. *ArXiv:1703.09398 [Cs]*. <http://arxiv.org/abs/1703.09398>
- Horne, C. L. (2021). Internet governance in the “post-truth era”: Analyzing key topics in “fake news” discussions at IGF. *Telecommunications Policy*, 45(6). <https://doi.org/10.1016/j.telpol.2021.102150>
- Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, Social Media and Political Polarization in the United States, 2012-2018*. Project on Computational Propaganda.
- Humphrecht, E. (2019). Where ‘fake news’ flourishes: A comparison across four Western democracies. *Information, Communication & Society*, 22(13), 1973–1988. <https://doi.org/10.1080/1369118X.2018.1474241>
- Hwang, Y., Ryu, J. Y., & Jeong, S. H. (2021). Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 188–193. <https://doi.org/10.1089/cyber.2020.0174>
- Innes, M., Dobрева, D., & Innes, H. (2021). Disinformation and digital influencing after terrorism: spoofing, truthing and social proofing. *Contemporary Social Science*, 16(2), 241–255. <https://doi.org/10.1080/21582041.2019.1569714>
- Instagram. (n.d.). *Reducing the Spread of False Information on Instagram*. Help Center. Retrieved July 14, 2022, from Reducing the Spread of False Information on Instagram %7C Instagram Help Center
- Jacko, J. A. (Ed.). (2012). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications* (3rd Editio). CRC Press. <https://doi.org/10.1201/b11963>

- Johnson, S. S. (2022). Knowing Well, Being Well: well-being born of understanding: The Urgent Need for Coordinated and Comprehensive Efforts to Combat Misinformation. *American Journal of Health Promotion*, 36(3), 559–581. <https://doi.org/10.1177/08901171211070957>
- Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, 118(46), e2100786118. <https://doi.org/10.1073/pnas.2100786118>
- Kahne, J., & Bowyer, B. (2017). Educating for Democracy in a Partisan Age: Confronting the Challenges of Motivated Reasoning and Misinformation. *American Educational Research Journal*, 54(1), 3–34. <https://doi.org/10.3102/0002831216679817>
- Kaiser, B., Mayer, J., & Matias, N. (2021). *Warnings That Work: Combating Misinformation Without Deplatforming*. Disinformation. <https://www.lawfareblog.com/warnings-work-combating-misinformation-without-deplatforming>
- Kaiser, B., Wei, J., Lucherini, E., Lee, K., Matias, J. N., & Mayer, J. (2021). Adapting security warnings to counter online disinformation. *Proceedings of the 30th USENIX Security Symposium*, 1163–1180.
- Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media and Society*, 23(5), 1301–1326. <https://doi.org/10.1177/1461444820959296>
- Karami, M., Nazer, T. H., & Liu, H. (2021). Profiling Fake News Spreaders on Social Media through Psychological and Motivational Factors. *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, 225–230. <https://doi.org/10.1145/3465336.3475097>
- Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining, 2018-February*, 324–332. <https://doi.org/10.1145/3159652.3159734>
- Krippendorff, K. (2012). Content Analysis. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 234–238). Sage Publications Inc. <https://doi.org/https://dx.doi.org/10.4135/9781412961288>
- Krippendorff, K. (2018). *Content Analysis: An Introduction to its Methodology* (4th ed.). SAGE Publications Inc.
- Kumar, K. P. K., & Geethakumari, G. (2014). Detecting disinformation in online social networks using cognitive psychology. *Human-Centric Computing and Information Sciences*, 4(1), 1–22. <https://doi.org/10.1186/s13673-014-0014-x>
- Kumar, S., & Shah, N. (2018). *False Information on Web and Social Media: A Survey*. 2657, 1–35. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Kumari, R., Ashok, N., Ghosal, T., & Ekbal, A. (2022). What the fake? Probing misinformation detection standing on the shoulder of novelty and emotion. *Information Processing and Management*, 59(1). <https://doi.org/10.1016/j.ipm.2021.102740>
- Kunda, Z., Dunning, D., Jones, E., Jussim, L., Miller, D., Nisbett, R., Petty, R., & Prentice, D. (1990). *The Case for Motivated Reasoning* (Vol. 108, Issue 3). <http://spider.apa.org/ftdocs/bul/1990/november/bul1083480.html>
- Lane, J. D., & DePaulo, B. M. (1999). Completing Coyne's Cycle: Dysphorics' Ability to Detect Deception. *Journal of Research in Personality*, 33(3), 311–329. <https://doi.org/10.1006/jrpe.1999.2253>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science (New York, N.Y.)*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A.,

- Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., & Mattsson, C. (2017). *Combating Fake News: An Agenda for Research and Action* | Shorenstein Center. Shorenstein Center on Media, Politics, and Public Policy. <https://shorensteincenter.org/combating-fake-news-agenda-for-research/>
- Leng, Y., Zhai, Y., Sun, S., Wu, Y., Selzer, J., Strover, S., Fensel, J., Pentland, A., & Ding, Y. (2020). Analysis of disinformation during the COVID-19 outbreak in China: Cultural, social and political entanglements. *ArXiv:2005.10414 [Cs]*. <http://arxiv.org/abs/2005.10414>
- Levak, T. (2020). Disinformation in the new media system – Characteristics, forms, reasons for its dissemination and potential means of tackling the issue. *Medijska Istrazivanja*, 26(2), 29–58. <https://doi.org/10.22572/mi.26.2.2>
- Lewandowsky, S., Ecker, U., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the “Post-Truth” Era. *Journal of Applied Research in Memory and Cognition*, 353–369.
- Lewis, J. R., & Sauro, J. (2021). Usability and User Experience: Design and Evaluation. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (pp. 972–1015). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119636113.ch38>
- MacKuen, M., Wolak, J., Keele, L., & Marcus, G. E. (2010). Civic Engagements: Resolute Partisanship or Reflective Deliberation. *American Journal of Political Science*, 54(2), 440–458.
- Mare, A., Mabweazara, H. M., & Moyo, D. (2019). “Fake News” and Cyber-Propaganda in Sub-Saharan Africa: Recentering the Research Agenda. In *African Journalism Studies* (Vol. 40, Issue 4, pp. 1–12). Taylor and Francis Inc. <https://doi.org/10.1080/23743670.2020.1788295>
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, 5(1), 47. <https://doi.org/10.1186/s41235-020-00252-3>
- Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online*.
- Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, 1, 167–195. <https://doi.org/10.1146/annurev.clinpsy.1.102803.143916>
- Mayorga, M., Hester, E. B., Helsel, E., Ivanov, B., Sellnow, T. L., Slovic, P., Burns, W. J., & Frakes, D. (2020). Enhancing Public Resistance to “Fake News” A Review of the Problem and Strategic Solutions 1 Defining and Understanding “Fake News.” In *The Handbook of Applied Communication Research* (Vol. 1).
- Mcconaghy, T., Marques, R., Müller, A., De Jonghe, D., Mcconaghy, T., McMullen, G., ... Granzotto, A. (2016). BigchainDB: A Scalable Blockchain Database (DRAFT). *BigchainDB*, 1–65.
- Mclane, L. (2021). *DUAL DISRUPTIONS: OVERCOMING THE EFFECTS OF DISASTERS AND MIS-, DIS-, AND MAL-INFORMATION ON DEMOCRACIES*. NAVAL POSTGRADUATE SCHOOL MONTEREY.
- Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, 153. <https://doi.org/10.1016/j.eswa.2019.112986>
- Mehta, A. M., Liu, B. F., Tyquin, E., & Tam, L. (2021). A process view of crisis misinformation: How public relations professionals detect, manage, and evaluate crisis misinformation. *Public Relations Review*, 47(2). <https://doi.org/10.1016/j.pubrev.2021.102040>
- Melki, J., Tamim, H., Hadid, D., Makki, M., el Amine, J., & Hitti, E. (2021). Mitigating infodemics: The relationship between news exposure and trust and belief in COVID-19 fake news and social media spreading. *PLoS ONE*, 16(6 June). <https://doi.org/10.1371/journal.pone.0252830>

- Meta AI. (2020). *Here's how we're using AI to help detect misinformation*. ML Applications & Integrity. <https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/>
- Meta. (2020). *Investing in U.S. Media Literacy Initiatives*. Meta Journalism Project. <https://www.facebook.com/journalismproject/media-literacy-initiatives>
- Meta. (2021). *How Facebook's third-party fact-checking program works*. Meta for Media. <https://www.facebook.com/formedia/blog/third-party-fact-checking-how-it-works>
- Meta. (n.d.). *Meta's Third-Party Fact-Checking Program*. Meta for Media. Retrieved July 14, 2022, from <https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking>
- Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*, 65(2), 180–212. <https://doi.org/10.1177/0002764219878224>
- Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*, 65(2), 180–212. <https://doi.org/10.1177/0002764219878224>
- Mougayar, W. (2016). *The business blockchain : promise, practice, and application of the next Internet technology*. Wiley.
- Moyo, D. (2009). Citizen journalism and the parallel market of information in zimbabwe's 2008 election. *Journalism Studies*, 10(4), 551–567. <https://doi.org/10.1080/14616700902797291>
- Nielsen, J., & Landauer, J. (1993). A Mathematical Model of Finding the Usability Problems. *CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, 206–213. <https://doi.org/10.1145/169059.169166>
- Nielsen, J. (1994). Heuristic Evaluation. In J Nielsen & R. L. Mack (Eds.), *Usability Inspection Methods* (pp. 25–62). John Wiley & Sons. <https://doi.org/10.5555/189200.189209>
- Nielsen, Jakob. (1993). *Usability Engineering*. Morgan Kaufmann Publishers Inc. <https://doi.org/10.5555/2821575>
- Nielsen, Jakob. (1994). Enhancing the Explanatory Power of Usability Heuristics. *CHI '94: Conference Companion on Human Factors in Computing Systems*, 210. <https://doi.org/10.1145/259963.260333>
- O'Sullivan, D. (2021). *Twitter is now showing a warning*. Twitter. <https://twitter.com/donie/status/1346055145793220611>
- Oh, O., Agrawal, M., & Rao, H. R. (2013). Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises. *MIS Quarterly*, 37(2), 407–426.
- Pandemic fatigue: One-in-three Canadians report struggles with mental health; 23% say they're 'depressed.' (2022, January 24). *Angus Reid Institute*. <https://angusreid.org/pandemic-mental-health-addictions/>
- Pappas, V. (2020). *Combating misinformation and election interference on TikTok*. Newsroom. <https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok>
- Parveen, N., & Waterson, J. (2020, April 4). UK phone masts attacked amid 5G-coronavirus conspiracy theory. *The Guardian*. <https://www.theguardian.com/uk-news/2020/apr/04/uk-phone-masts-attacked-amid-5g-coronavirus-conspiracy-theory>
- Paschalides, D., Christodoulou, C., Orphanou, K., Andreou, R., Kornilakis, A., Pallis, G., Dikaiakos, M. D., & Markatos, E. (2021). Check-It: A plugin for detecting fake news on the web. *Online Social Networks and Media*, 25. <https://doi.org/10.1016/j.osnem.2021.100156>
- Paschen, J. (2020). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product and Brand Management*, 29(2), 223–233. <https://doi.org/10.1108/JPBM-12-2018-2179>

- Pennycook, G., & Rand, D. G. (2019a). Fighting disinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7), 2521–2526. <https://doi.org/10.1073/pnas.1806781116>
- Pennycook, G., & Rand, D. G. (2019b). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Perlis, R. H., Ognyanova, K., Santillana, M., Lin, J., Druckman, J., Lazer, D., Green, J., Simonson, M., Baum, M. A., & Della Volpe, J. (2022). Association of Major Depressive Symptoms With Endorsement of COVID-19 Vaccine Disinformation Among US Adults. *JAMA Network Open*, 5(1), e2145697. <https://doi.org/10.1001/jamanetworkopen.2021.45697>
- Pierre, J. M. (2020). Mistrust and misinformation: A two-component, socio-epistemic model of belief in conspiracy theories. *Journal of Social and Political Psychology*, 8(2), 617–641. <https://doi.org/10.5964/jspp.v8i2.1362>
- Polletta, F., & Callahan, J. (2017). Deep stories, nostalgia narratives, and fake news: Storytelling in the trump era. *American Journal of Cultural Sociology*, 5(3), 392–408. <https://doi.org/10.1057/s41290-017-0037-7>
- Posetti, J., & Matthews, A. (2018). *A short guide to the history of 'fake news' and disinformation* (Vol. 15, Issue 2). Informa UK Limited. <https://www.icfj.org/news/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module>
- Privitera, M. B. (2019). Heuristic analysis, cognitive walkthroughs & expert reviews. In *Applied Human Factors in Medical Device Design*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-816163-0.00010-4>
- Qiu, J., Shen, B., Zhao, M., Wang, Z., Xie, B., & Xu, Y. (2020). A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: Implications and policy recommendations. *General Psychiatry*, 33(2), e100213. <https://doi.org/10.1136/gpsych-2020-100213>
- Quandt, T., Frischlich, L., Boberg, S., & Schatto-Eckrodt, T. (2019). Fake News. In *The International Encyclopedia of Journalism Studies* (pp. 1–6). Wiley. <https://doi.org/10.1002/9781118841570.iejs0128>
- Ramasubramanian, S. (2020, November 23). Facebook introduces new AI systems to detect disinformation. *The Hindu*. <https://www.thehindu.com/sci-tech/technology/internet/facebook-introduces-new-ai-systems-to-detect-disinformation/article33159027.ece>
- Riihiahho, S. (2017). Usability Testing. In K. L. Norman & J. Kirakowski (Eds.), *The Wiley Handbook of Human Computer Interaction* (pp. 257–275). John Wiley & Sons. <https://doi.org/10.1002/9781118976005.ch14>
- Rosen, G., Harbath, K., Gleicher, N., & Leathern, R. (2019). *Helping to Protect the 2020 US Elections*. Meta Newsroom. <https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/#reducing-misinformation>
- Rosenzweig, E. (2015a). Usability Inspection Methods. In *Successful User Experience: Strategies and Roadmaps* (pp. 115–130). Morgan Kaufmann, Elsevier. <https://doi.org/10.1016/b978-0-12-800985-7.00006-5>
- Rosenzweig, E. (2015a). Usability Inspection Methods. In *Successful User Experience: Strategies and Roadmaps* (pp. 115–130). Morgan Kaufmann, Elsevier. <https://doi.org/10.1016/b978-0-12-800985-7.00006-5>
- Rosenzweig, E. (2015b). Usability Testing. In *Successful User Experience: Strategies and Roadmaps* (pp. 131–154). Elsevier Inc. <https://doi.org/10.1016/b978-0-12-800985-7.00007-7>
- Roth, Y. (2022). *Introducing our crisis misinformation policy*. Twitter Blog. https://blog.twitter.com/en_us/topics/company/2022/introducing-our-crisis-misinformation-policy

- Roth, Y., & Pickles, N. (2020). *Updating our approach to misleading information*. Twitter Blog. https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information
- Rubin, V. L. (2019). Disinformation and misinformation triangle: A conceptual model for “fake news” epidemic, causal factors and interventions. *Journal of Documentation*, 75(5), 1013–1034. <https://doi.org/10.1108/JD-12-2018-0209>
- Salari, N., Hosseinian-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulpoor, S., Mohammadi, M., Rasoulpoor, S., & Khaledi-Paveh, B. (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health*, 16(1), 57. <https://doi.org/10.1186/s12992-020-00589-w>
- Santos-D’amorim, K., & Miranda, M. K. F. de O. (2021). Misinformation, disinformation, and malinformation: Clarifying the definitions and examples in disinfodemic times. *Encontros Bibli*, 26. <https://doi.org/10.5007/1518-2924.2021.e76900>
- Sengupta, E., Nagpal, R., Mehrotra, D., & Srivastava, G. (2021). ProBlock: a novel approach for fake news detection. *Cluster Computing*, 24(4), 3779–3795. <https://doi.org/10.1007/S10586-021-03361-W>
- Shrauger, J. S., Mariano, E., & Walter, T. J. (1998). Depressive Symptoms and Accuracy in the Prediction of Future Events. *Personality and Social Psychology Bulletin*, 24(8), 880–892. <https://doi.org/10.1177/0146167298248007>
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective*. <http://www.journalism.org/2016/05/26/news-use-across-Data-Mining-Perspective>
- Silver, C., & Lewins, A. (2017). Qualitative Coding in Software: Principles and Processes. In *Using Software in Qualitative Research: A Step-by-Step Guide* (pp. 158–185). Sage Publications Inc. <https://doi.org/10.4135/9781473906907.n8>
- Songs, W., Krasodomski-Jones, A., Smith, J., Jones, E., Judson, E., & Miller, C. (2019). *Information Operations in the Digital Age*.
- Stanovich, K. E., & West, R. F. (2007). Natural Myside bias is independent of cognitive ability. *Thinking & Reasoning*, 13(3), 225–247. <https://doi.org/10.1080/13546780600780796>
- StratCom | NATO Strategic Communications Centre of Excellence Riga, Latvia. (n.d.). Retrieved April 19, 2022, from <https://stratcomcoe.org/publications/government-responses-to-malicious-use-of-social-media/125>
- Sultan, K., Ruhi, U., & Lakhani, R. (2018). Conceptualizing Blockchains: Characteristics & Applications. In *11th IADIS International Conference on Information Systems* (pp. 49–57). Lisbon, Portugal.
- Talwar, S., Dhir, A., Singh, D., Virk, G. S., & Salo, J. (2020). Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. *Journal of Retailing and Consumer Services*, 57. <https://doi.org/10.1016/j.jretconser.2020.102197>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions. In *Digital Journalism* (Vol. 6, Issue 2, pp. 137–153). Routledge. <https://doi.org/10.1080/21670811.2017.1360143>
- Tang, S., Willnat, L., & Zhang, H. (2021). Fake news, information overload, and the third-person effect in China. *Global Media and China*, 6(4), 492–507. <https://doi.org/10.1177/20594364211047369>
- Tashakkori, A., & Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research*. Sage Publications Inc.
- Taylor, S., Landry, C. A., Paluszczek, M. M., Fergus, T. A., McKay, D., & Asmundson, G. J. G. (2020). COVID stress syndrome: Concept, structure, and correlates. *Depression and Anxiety*, 37(8), 706–714. <https://doi.org/10.1002/da.23071>
- TikTok. (n.d.). *Community Guidelines*. Community Guidelines. Retrieved July 5, 2022, from <https://www.tiktok.com/community-guidelines?lang=en#37>

- Turk, Z. (2018). *Technology as Enabler of Fake News and a Potential Tool to Combat It*.
Twitter. (n.d.-a). *Notes shown on Twitter*. Birdwatch Guide. Retrieved July 14, 2022, from <https://twitter.github.io/birdwatch/notes-on-twitter/>
- Twitter. (n.d.-b). *Ranking Notes*. Birdwatch Guide. Retrieved July 14, 2022, from <https://twitter.github.io/birdwatch/ranking-notes/>
- UNESCO. (n.d.). *Global Media and Information Literacy Week*. Retrieved July 11, 2022, from <https://www.unesco.org/en/media-information-literacy-week>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media and Society*, 6(1). <https://doi.org/10.1177/2056305120903408>
- van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating Against Fake News About COVID-19. *Frontiers in Psychology*, 11. <https://www.frontiersin.org/article/10.3389/fpsyg.2020.566790>
- Vincent, J. (2020, June 3). *Conspiracy theorists say 5G causes novel coronavirus, so now they're harassing and attacking UK telecoms engineers*. The Verge. <https://www.theverge.com/2020/6/3/21276912/5g-conspiracy-theories-coronavirus-uk-telecoms-engineers-attacks-abuse>
- Vosoughi, S., Roy, D., & Aral, S. (2018). *The spread of true and false news online*. <https://www.science.org>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Waldman, A. E. (2020). Cognitive biases, dark patterns, and the ‘privacy paradox.’ *Current Opinion in Psychology*, 31, 105–109. <https://doi.org/10.1016/J.COPSYC.2019.08.025>
- Walker, M., & Matsa, K. E. (2021, September 20). *News Consumption Across Social Media in 2021*. *Pew Research Center's Journalism Project*. <https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/>
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350–375. <https://doi.org/10.1080/10584609.2019.1668894>
- Wang, Z., Maria, M., Morales, R., Husak, K., Kleinman, M., & Parthasarathy, S. (2021). *In Communities We Trust Institutional Failures and Sustained Solutions for Vaccine Hesitancy*.
- Wardle, C. (2018). The Need for Smarter Definitions and Practical, Timely Empirical Research on Information Disorder. *Digital Journalism*, 6(8), 951–963. <https://doi.org/10.1080/21670811.2018.1502047>
- Wardle, C., & Derakhshan, H. (2017). *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making* *Information Disorder Toward an interdisciplinary framework for research and policymaking*. www.coe.int
- West, D. M. (2017, December 18). *How to combat fake news and disinformation*. *Brookings*. <https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/>
- Westerlund, M. (2019). *The Emergence of Deepfake Technology: A Review*.
- WhatsApp—Statistics & Facts | Statista*. (n.d.). Retrieved April 19, 2022, from <https://www.statista.com/topics/2018/whatsapp/#dossierKeyfigures>
- Wheeler, T. (2017, November 1). Using “public interest algorithms” to tackle the problems created by social media algorithms. *Brookings*. <https://www.brookings.edu/blog/techtank/2017/11/01/using-public-interest-algorithms-to-tackle-the-problems-created-by-social-media-algorithms/>
- Yaga, D., Mell, P., Roby, N., & Scarfone, K. (n.d.). *Blockchain Technology Overview*. <https://doi.org/10.6028/NIST.IR.8202>

- Yang, J., & Lee, S. (2020). Framing the MERS information crisis: An analysis on online news media's rumour coverage. *Journal of Contingencies and Crisis Management*, 28(4), 386–398. <https://doi.org/10.1111/1468-5973.12292>
- Yesmin, S., & Ahmed, S. M. Z. (2022). Infodemic surrounding COVID-19: Can LIS students recognize and categorize “problematic information” types on social media? *Digital Library Perspectives*, 38(1), 3–15. <https://doi.org/10.1108/DLP-03-2021-0020>
- Youtube. (n.d.-a). *Community Guidelines strike basics on YouTube*. Youtube Help. Retrieved July 11, 2022, from <https://support.google.com/youtube/answer/2802032>
- Youtube. (n.d.-b). *How does YouTube address misinformation?* Youtube Commitments. Retrieved July 9, 2022, from <https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/#providing-context>
- Youtube. (n.d.-c). *Misinformation policies*. Youtube Help. Retrieved July 14, 2022, from https://support.google.com/youtube/answer/10834785?hl=en&ref_topic=10833358
- Youtube. (n.d.-d). *Report inappropriate videos, channels, and other content on YouTube*. Youtube Help. Retrieved July 9, 2022, from <https://support.google.com/youtube/answer/2802027?hl=en&co=GENIE.Platform%3DDesktop#zippy=%2Creport-a-comment%2Creport-a-video%2Creport-a-playlist%2Creport-a-thumbnail>
- Youtube. (n.d.-e). *YouTube Community Guidelines Enforcement*. Transparency Report Help Center. Retrieved July 8, 2022, from <https://support.google.com/transparencyreport/answer/9209072?hl=en#zippy=%2Cwho-reviews-flags-to-make-decisions-about-removing-or-restricting-content%2Cwho-are-the-human-reviewers-trained%2Cwho-else-is-content-identified-as-violating-youtubes-community-gu>
- Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality*, 11(3). <https://doi.org/10.1145/3309699>
- Zeller, B. (1948). ALLPORT, GORDON W., and LEO POSTMAN. The Psychology of Rumor. Pp. xiv, 247. New York: Henry Holt and Company, 1947. \$2.60. *The ANNALS of the American Academy of Political and Social Science*, 257(1), 240–241. <https://doi.org/10.1177/000271624825700169>
- Zerback, T., Töpfl, F., & Knöpfle, M. (2021). The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them. *New Media & Society*, 23(5), 1080–1098. <https://doi.org/10.1177/1461444820908530>
- Zhang, L., Xu, L., & Zhang, W. (2017). Social media as amplification station: Factors that influence the speed of online public response to health emergencies. *Asian Journal of Communication*, 27(3), 322–338. <https://doi.org/10.1080/01292986.2017.1290124>
- Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing and Management*, 57(2). <https://doi.org/10.1016/j.ipm.2019.03.004>
- Zhou, S.-J., Zhang, L.-G., Wang, L.-L., Guo, Z.-C., Wang, J.-Q., Chen, J.-C., Liu, M., Chen, X., & Chen, J.-X. (2020). Prevalence and socio-demographic correlates of psychological health problems in Chinese adolescents during the outbreak of COVID-19. *European Child & Adolescent Psychiatry*, 29(6), 749–758. <https://doi.org/10.1007/s00787-020-01541-4>
- Zhou, X., & Zafarani, R. (2018). *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*. <https://doi.org/10.1145/3395046>
- Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Emotional Dynamics in the Age of Disinformation. *PloS One*, 10(9), e0138740–e0138740. <https://doi.org/10.1371/journal.pone.0138740>

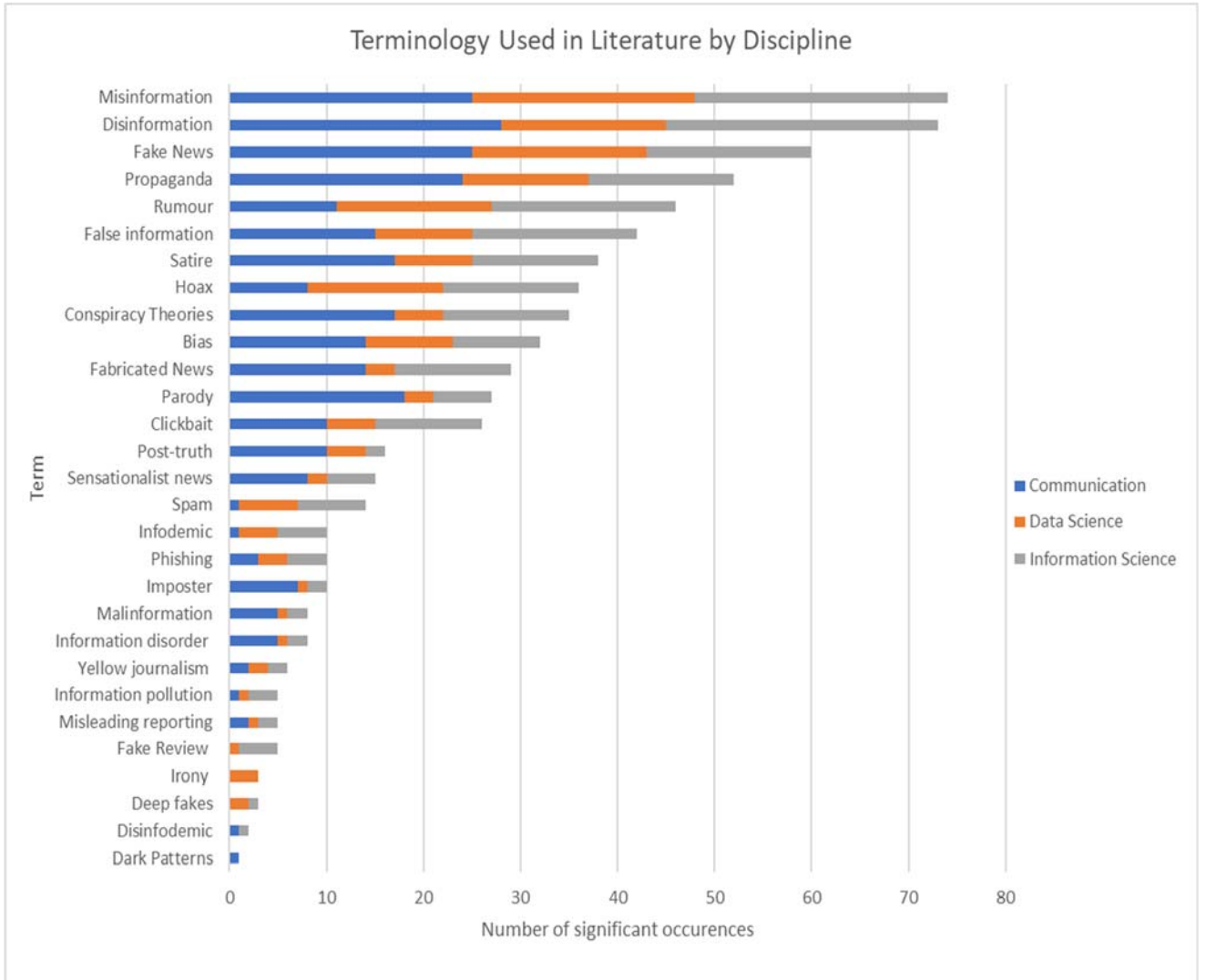
Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. In *ACM Computing Surveys* (Vol. 51, Issue 2). Association for Computing Machinery. <https://doi.org/10.1145/3161603>

15.Appendices


Appendix A


Post- truth era										
	Information Disorder			Intent				Veracity		
	Misinformation	Disinformation	Malinformation	Political gain	Financial gain	Mislead	Fun / Other	Fabricated/False	Unverified	Truth
Fake news	1	1	1	1	1	1	1	1	1	1
false news		1		1				1		
fabricated news		1		1		1		1		
sensationalist news		1		1	1	1	1	1	1	
bias news	1			1	1					1
propaganda		1		1		1				1
parody	1			1			1	1		
satire	1			1			1			1
rumour		1		1		1				1
hoax		1		1		1	1			1
clickbait		1			1	1		1		
imposter content		1		1	1			1		
spam		1			1			1		
phishing		1			1			1		
irony	1						1			1
conspiracy		1		1			1			1
deep fakes		1		1			1	1		
fake reviews		1			1	1		1		
leak			1	1			1			1
harrassment			1	1						1
hate speech			1	1						1
misleading news/reporting										
gossip		1					1	1		1



Appendix B





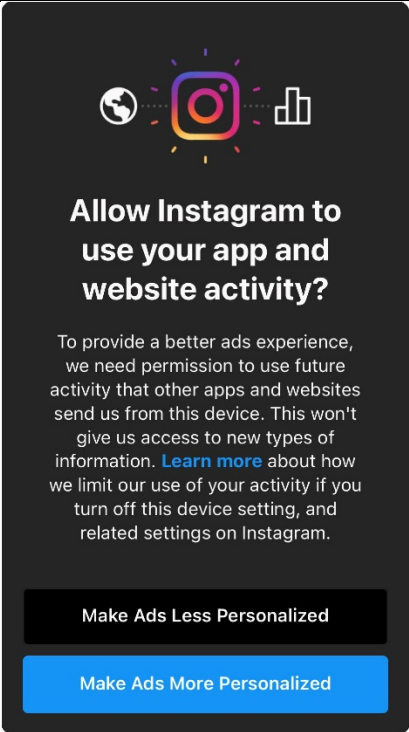

Appendix C


Misinformation		
Term	Definition	Example
Satire	Misinformation that is spread without the intent to harm but with a potential to fool the reader or viewer. Relies on humour and exaggeration to mock issues. (Hinsley & Holton, 2021; Horne, 2021; Mayorga et al., 2020; Molina et al., 2021; Santos-D'amorim & Miranda, 2021)	The Onion – Online “news” source that publishes stories that relate to real world issues but are unrealistic and comical/mocking said issues.
Parody	Misinformation that is spread without the intent to harm but with a potential to fool the reader or viewer. Parody uses non-factual information to inject humour. (Hinsley & Holton, 2021; Horne, 2021; Mayorga et al., 2020; Molina et al., 2021; Santos-D'amorim & Miranda, 2021)	Saturday Night Live – Popular American media that features parody skits of relevant pop culture references. The sketches take on inspiration from real world events and inject non-factual information to fool and entertain the consumer.
Bias	A cognitive bias based on experience and social upbringing, or pure ignorance, that shapes the author’s writing. (Kapantai et al., 2021; Santos-D'amorim & Miranda, 2021; Zannettou et al., 2019)	<p>Researched produced to indicated that a drug was successful in passing clinical trials but was funded by the drug company.</p> <p>Lexchin, J., Bero, L. A., Djulbegovic, B., & Clark, O. (2003). Pharmaceutical industry sponsorship and research outcome and quality: systematic review. <i>BMJ (Clinical research ed.)</i>, 326(7400), 1167–1170. https://doi.org/10.1136/bmj.326.7400.1167</p>
Disinformation		
Fake News	An umbrella term to describe a range of information types. Also defined as false information spread online with the intent to harm for financial or political gain. (Ansar & Goswami, 2021; Horne, 2021; Kapantai et al., 2021; Wardle & Derakhshan, 2017; Choraś et al., 2021; Meel & Vishwakarma, 2020; Zhang & Ghorbani, 2020)	<p>Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement</p> 
False News	Known wrong information. (Chadwick & Stanyer, 2022; Molina et al., 2021; Wardle & Derakhshan, 2017; Zannettou et al., 2019)	<p>This false news story was taken from a satire website and circulated with the intension of political gain from a number of Republican, Donald Trump, supporters during his 2016 presidential run.</p>
Fabricated News	Entirely false or fictional stories created with the intention of causing harm. (Ansar &	



	Goswami, 2021; Kapantai et al., 2021; Zannettou et al., 2019)	https://web.archive.org/web/20161115024211/http://wtoc5news.com/us-election/pope-francis-shocks-world-endorse-donald-trump-for-president-releases-statement/ https://this-person-does-not-exist.com/en
Deep Fake	Deep fakes typically imply that they are created synthetically with artificial intelligence and result in fake videos that are hard to distinguish from real. (S. Agarwal et al., 2019; Vaccari & Chadwick, 2020; Westerlund, 2019)	A website dedicated to generating pictures of AI fabricated humanlike faces that are not real photographs or real people captured.
Propaganda	Deceptive information spread with the intention of influencing behavior in targeted communities, for political reasons or to push an agenda, usually by government institutions. (Fallis, 2015; Gradoń et al., 2021; Hamilton et al., 2021; Meel & Vishwakarma, 2020; Shu et al., 2017; Zannettou et al., 2019)	 https://thebarentsobserver.com/en/security/2022/03/wave-warmongering-gushing-over-russian-youth The “Z” on there are used as a symbol for Russia’s war on Ukraine. The three young women in the video photographed above, are members of the “Yunarmiya”, a military youth movement established by the Russian Minister of Defense. The symbol is used as propaganda targeted towards youth to insinuate a Ukrainian “genocide against the Russians”.
Rumor	Anecdotes or stories that, although may be widely known, are ambiguous in nature and are not confirmed to be true. (Ansar & Goswami, 2021; Kapantai et al., 2021; Kumar & Shah, 2018; Shu et al., 2017)	https://www.the-sun.com/entertainment/4368721/kourtney-kardashian-clues-pregnant-travis-barker-bump-grammys/ Pregnancy is amongst the most prevalent rumors in mainstream media. Although “The Kardashian’s” fans know that Kourtney Kardashian has been undergoing IVF treatments, no pregnancy has been confirmed and yet tabloid articles allege that it is true.
Hoax	A large-scale, intentionally deceptive fabrication or propagation of a conspiracy theory that is presented as a true or legitimate story. (Chew & Eysenbach, 2010; Molina et al., 2021; Rubin, 2019; Santos-D’amorim & Miranda, 2021).	https://www.aljazeera.com/opinions/2022/4/17/britains-trojan-horse-a-hoax-that-still-harms-muslims A letter sent to a British school wherein the sender claimed schools in Britain were being taken over by terrorists and were centering Islamic teachings in their curriculum and it was “extremism and radicalization” in high-performing schools in deprived areas. After extensive observation it became clear this letter was a xenophobic hoax to deteriorate the programs and reputations of the schools and Islamic leaders.
Clickbait	Media sources hyperbolize their fabricated or true stories to attract attention to the intended	https://www.t TMZ.com/2022/04/21/asap-rocky-rihanna-arrest-cops-gun-airport/

	<p>webpage through emotionally charged titles or by withholding information in the title. (Ghanem et al., 2020; Kapantai et al., 2021; Meel & Vishwakarma, 2020; Rubin, 2019; Shu et al., 2017)</p>	<p>popular media source TMZ titled this article “ASAP Rocky, Rihanna Blindsided, Shocked when Cops Arrested Him... Cops Say They Had Their Reasons”. The article was written the day after ASAP Rocky was allegedly arrested by police in LAX. Although no direct quote from Rhianna or credible source, they announce in the title her dismay which would lead a consumer to clicking on this article to understand more about the situation, although not all of the information alluded to could be found in the article.</p>
<p>Sensationalist News</p>	<p>Information that is intended to evoke a reaction and thus qualifies as intentionally manipulative. (Molina et al, 2021)</p>	 <p>Source: Bloomberg News Week – September 29, 2014 Issue: https://www.bloomberg.com/magazine/businessweek/14_40</p>
<p>Spam</p>	<p>Intentionally deceitful information used to take up space of legitimate information, or propagate false information. (Gradoń et al., 2021; Rubin, 2019)</p>	

		<p>Source: https://www.nbcnews.com/news/world/facebook-spam-attack-pushes-diet-miracle-flna24104</p> <p>This NBC News article depicts spam posted to Facebook to advertise a “miracle” weight-loss solution. No actual solution was depicted in any of the messages.</p>
<p>Imposter/ imposter content</p>	<p>Wardle and Derakshan (2018) use the term to describe content that is manipulated or falsely produced under the guise of original authorship and organizations (Wardle, Santos, Levak, Bran, Hinsley). Kapantai et al consider imposter to be psychologically motivated and verifiably false.</p>	 <p>Source: Freedom Forum Institute’s <i>Quick guide to spotting fake news</i>: https://www.freedomforuminstitute.org/first-amendment-center/primers/fake-news-primer/</p>
<p>Phishing</p>	<p>Referred to in the context of digital security and threat response since it has to do with theft of personal and identity information (Wardle, UNESCO, Akers, Figueira) and in the context of spam and bot detection in data science (Kim, Shu), done by social spammers (Bondielli), consequence of Clickbait (Meel and Vishwakarma), categorized as malinformation (Santos). Mostly used in information science</p>	 <p>The image above depicts a phishing attempt wherein a vague e-mail is sent to several randomized or specific e-mail addresses with links that will allow the sender to hack into the receiver’s data and information via their computer.</p>

<p>Dark Patterns</p>	<p>Application and website interface features designed to intentionally mislead users into providing consent, purchasing items or subscribing to news. (Carmi et al., 2020)</p>	 <p>Allow Instagram to use your app and website activity?</p> <p>To provide a better ads experience, we need permission to use future activity that other apps and websites send us from this device. This won't give us access to new types of information. Learn more about how we limit our use of your activity if you turn off this device setting, and related settings on Instagram.</p> <p>Make Ads Less Personalized</p> <p>Make Ads More Personalized</p> <p>An example of UI design feature that deceives users into subscribing to personalized ads and data collection. Source: https://www.vox.com/recode/22351108/dark-patterns-ui-web-design-privacy</p>
<p>Leaks</p>	<p>Based on Wardle and Derakshan’s definition of true information released with the intent to harm, classified as malinformation. Context of leaks in the literature refers to leaks with connection to foreign malicious actors (Tandoc, Allcott, Bastos, Zhang)</p>	 <p>Daily Express @Daily_Express</p> <p>Putin humiliated as hackers expose 600 in Russia's secret service and leak 87,500 emails express.co.uk/news/science/1...</p> <p>3:34 am · 19 Apr 2022 · TweetDeck</p> <p>25 Retweets 3 Quote Tweets 86 Likes</p> <p>Leaks of Russian plans while they continue to invade Ukraine have become increasingly more common and a tool for Ukraine and allies to defend their country.</p>

<p>Harassment</p>	<p>Wardle and Derkashan group harassment with malinformation. Some authors group harassment with spamming or trolling or doxing (McLane, Wooley, Froehlich)</p>	 <p>Pictured here is an Instagram picture posted by Kanye West. During this time he was known to post on Instagram, slanderous pictures and comments in regards to his ex-wife, Kim Kardiashian's alleged new boy-friend, Pete Davidson. West's Instagram page was suspended by Instagram a few days later due to his use of racial slurs in an attack on Trevor Noah which was a violation of their harassment policy.</p>
-------------------	---	---

<p>Hate Speech</p>	<p>Factually inaccurate or out of context information information, used (boyd, Wardle). In the literature, hate speech is a type of offensive speech or defamation arising from disinformation campaigns that needs to be regulated due to its offensive racial and sexual nature (Marwick and Lewis, Mayorga, Rubin, Kapantai).</p>	 <p>This slanderous and racist tweet posted by Randy Hillier targeting someone with an opposing of the Freedom Convoy to Ottawa is depicted here and called out as hate speech by another Twitter user.</p>
<p>Conspiracy Theory</p>	<p>Theories not based on factual information and are often used to explain important events and aim to implicate governments or powerful individuals in illegal acts. (Kapantai et al., 2021; Meel & Vishwakarma, 2020; Santos-D'Amorim & Miranda, 2021; Zannettou et al., 2019)</p>	<p>https://www.theguardian.com/us-news/2020/dec/27/nashville-blast-investigators-examine-5g-paranoia</p>  <p>5G Network Paranoia – Claims regarding 5G network, supported by QAnon, refer to 5G as being a tool used by the deep state to spy and could potentially cause cancer and spread the coronavirus.</p>

Appendix D. Heuristic Evaluations – MediaBias/FactCheck Website

Heuristic Evaluation MediaBias/FactCheck Website				Tool Name:
1	Visibility of System Status	Issues	Recommendation	Severity
	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	No feedback indication given.	Provide a feedback indicator to inform users when the website is processing information.	0 1 2 3 4
2	Match between system and the real world	Issues	Recommendation	Severity
	The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	The information on the website is highly cluttered to the point that you skip over likely useful content. Terminology used for ratings will not be understood by all users.	Reduce the amount of content on the page, structure it better. Revise the terms being used to ensure that even those with limited English proficiency would understand.	0 1 2 3 4
3	User control and freedom	Issues	Recommendation	Severity
	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support, undo and redo.	Once you have clicked on some links, the only option is to use the back button on the browser.	Provide breadcrumbs for users to track where they are and can easily get back to their desired section of the website.	0 1 2 3 4
4	Consistency and standards	Issues	Recommendation	Severity
	Users should not have to wonder whether different words, situations, or actions mean the same thing.	Albeit poorly designed, the website is consistent. Improvements and clarifications are required for the rating standards.	Use different terminology for the rating scales and clarify their meaning.	0 1 2 3 4
5	Error Prevention	Issues	Recommendation	Severity
	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Not an issue.		0 1 2 3 4
6	Recognition rather than recall	Issues	Recommendation	Severity
	Minimize the user's memory load by making objections, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	Since the website is cluttered, it is not clear how or where to perform some tasks.	Restructure the website, make it more intuitive and provide less content all on the same page.	0 1 2 3 4
7	Flexibility and efficiency of use	Issues	Recommendation	Severity
	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	Not a major issue.		0 1 2 3 4
8	Aesthetic and minimalist design	Issues	Recommendation	Severity
	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	Too much information is provided on one page.	Redesign and restructure the website.	0 1 2 3 4
9	Help users recognize, diagnose, and recover from errors	Issues	Recommendation	Severity
	Error messages should be expressed in plain language (no codes), precisely indicate the	Not an issue.		0 1 2 3 4

	problem, and constructively suggest a solution.			
1 0	Help and documentation	Issues	Recommendation	Severity
	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task. List concrete steps to be carried out, and not be too large.	Information on the website, what the ratings mean and so forth is not clearly visible.	Restructure and redesign the website to make it more visible.	0 1 2 3 4
<p>0 - I don't agree that this is a usability problem at all.</p> <p>1 - Cosmetic problem only. Need not be fixed unless extra time is available on project.</p> <p>2 - Minor usability problem. Fixing this should be given low priority.</p> <p>3 - Major usability problem. Important to fix, so should be given high priority.</p> <p>4 - Usability catastrophe. Imperative to fix this before product can be released.</p>				

*This heuristic evaluation template was taken from: <https://uxplanet.org/how-to-conduct-heuristic-evaluation-85548a355dca>

Appendix E. Heuristic Evaluations – isthiscredible Website

Heuristic Evaluation isthiscredible Website				Tool Name:
1	Visibility of System Status	Issues	Recommendation	Severity
	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.			0 1 2 3 4
2	Match between system and the real world	Issues	Recommendation	Severity
	The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	The rating provided for URL searches is not very clear and could be misinterpreted as an error message.	Make the rating more visible, do not highlight the search bar in red.	0 1 2 3 4
3	User control and freedom	Issues	Recommendation	Severity
	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support, undo and redo.	Not a major issue, however, there is no option to cancel a search once it has started.	Provide an option to cancel a search.	0 1 2 3 4
4	Consistency and standards	Issues	Recommendation	Severity
	Users should not have to wonder whether different words, situations, or actions mean the same thing.	There should be more explanation as to what the ratings mean and what information was used to come to such rating.	Increase the amount of information provided.	0 1 2 3 4
5	Error Prevention	Issues	Recommendation	Severity
	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Not an issue.		0 1 2 3 4
6	Recognition rather than recall	Issues	Recommendation	Severity
	Minimize the user's memory load by making objections, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	Not an issue.		0 1 2 3 4
7	Flexibility and efficiency of use	Issues	Recommendation	Severity
	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	Not an issue.		0 1 2 3 4
8	Aesthetic and minimalist design	Issues	Recommendation	Severity
	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	Not an issue.		0 1 2 3 4
9	Help users recognize, diagnose, and recover from errors	Issues	Recommendation	Severity
	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	Not an issue.		0 1 2 3 4
	Help and documentation	Issues	Recommendation	Severity

1 0	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task. List concrete steps to be carried out, and not be too large.	When going to 'how it works,' the information was on the browser extension and not the website. No explanations or definitions are given for the ratings.	Provide an explanation for what the ratings mean.	0 1 2 3 4
<p>0 - I don't agree that this is a usability problem at all.</p> <p>1 - Cosmetic problem only. Need not be fixed unless extra time is available on project.</p> <p>2 - Minor usability problem. Fixing this should be given low priority.</p> <p>3 - Major usability problem. Important to fix, so should be given high priority.</p> <p>4 - Usability catastrophe. Imperative to fix this before product can be released.</p>				

Appendix F. Heuristic Evaluations – The Newsroom Beta Extension

Heuristic Evaluation Newsroom Beta Extension				Tool Name: The
1	Visibility of System Status	Issues	Recommendation	Severity
	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	The browser extension does not always appear. There is no way of knowing if and how to open it yourself.	Have an option to open the tool and ascertain whether it has any information/feedback on the website	0 1 2 3 4
2	Match between system and the real world	Issues	Recommendation	Severity
	The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	It is not clear what the source score and the respective ratings mean.	They should clarify what the source score is and what the rating means.	0 1 2 3 4
3	User control and freedom	Issues	Recommendation	Severity
	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support, undo and redo.	No option to enter information, therefore, not an issue.		0 1 2 3 4
4	Consistency and standards	Issues	Recommendation	Severity
	Users should not have to wonder whether different words, situations, or actions mean the same thing.	The tool is very minimal in terms of its functionality, therefore, not an issue.		0 1 2 3 4
5	Error Prevention	Issues	Recommendation	Severity
	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Not everyone may understand how to get the icon to appear.	There should be an option to have the icon appear once you click on the extension icon.	0 1 2 3 4
6	Recognition rather than recall	Issues	Recommendation	Severity
	Minimize the user's memory load by making objections, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	It is not clear if you can open the tool and/or how to do so. If the icon does not appear, you search the page to see if it is located somewhere else on the page. The colors of the icon also blend in with the page, so you have to remember what the icon looks like in order to find it on a busy page.	Chose more distinctive colours for the icon or make it more obvious that the tool has feedback on the page/article. Also provide an option to open the tool manually to see if there is any feedback.	0 1 2 3 4
7	Flexibility and efficiency of use	Issues	Recommendation	Severity
	Accelerators – unseen by the novice user- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	The tool is very minimal in terms of its functionality, therefore, not an issue.		0 1 2 3 4
8	Aesthetic and minimalist design	Issues	Recommendation	Severity
	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	The tool is very minimal in terms of its functionality, therefore, not an issue.		0 1 2 3 4
	Help users recognize, diagnose, and recover from errors	Issues	Recommendation	Severity

9	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	The tool is very minimal in terms of its functionality, therefore, not an issue.		0 1 2 3 4
10	Help and documentation Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task. List concrete steps to be carried out, and not be too large.	Issues There is no option to open the tool manually and/or get clarification on what some terms mean.	Recommendation Include an option to open the tool manually and have a user guide/FAQ.	Severity 0 1 2 3 4
0 - I don't agree that this is a usability problem at all. 1 - Cosmetic problem only. Need not be fixed unless extra time is available on project. 2 - Minor usability problem. Fixing this should be given low priority. 3 - Major usability problem. Important to fix, so should be given high priority. 4 - Usability catastrophe. Imperative to fix this before product can be released.				

Appendix G. Heuristic Evaluations – Newstrition Extension

Heuristic Evaluation Newstrition Extension				Tool Name:
1	Visibility of System Status	Issues	Recommendation	Severity
	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	It is not always clear if Newstrition is running. I.e. headline at the top of the page, if you don't know to look for it, you might miss/ignore it. Also, the icons do not always come up to provide feedback on the news source/article.	Make the headlines more obvious. Perhaps use a pop-up or something with more contrast to the web site.	0 1 2 3 4
2	Match between system and the real world	Issues	Recommendation	Severity
	The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	"Source score" – it is not clear what is meant by "low" – does that mean the trustworthiness, etc is low or the risk factor is low = a trustworthy site? "tin foil hat and quackery level" – unless the users have a deep base in western education/culture – they may not understand what is meant by this statement.	Use better terminology to clarify what source score means.	0 1 2 3 4
3	User control and freedom	Issues	Recommendation	Severity
	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support, undo and redo.	There is no clear option to close the icon once it is opened.	Include an option to close the window that appeared, without closing the icon/extension.	0 1 2 3 4
4	Consistency and standards	Issues	Recommendation	Severity
	Users should not have to wonder whether different words, situations, or actions mean the same thing.	The Newstrition tool has "summary" "newstrition" "fact check" and "ratings" but it is not clear what is the difference between the options. The tool mentions that it is a fact-check for individual articles and that it does not have information about the publisher, yet the icons that appear in the website have feedback on the news site in general and not for the news articles within the site.	Very inconsistent. Needs to be fixed.	0 1 2 3 4
5	Error Prevention	Issues	Recommendation	Severity
	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Once the icon is opened, it is not clear how to close it.	Include an option to close the window that appeared, without closing the icon/extension.	0 1 2 3 4
6	Recognition rather than recall	Issues	Recommendation	Severity
	Minimize the user's memory load by making objections, actions, and options visible. The user should not have to remember information from one part of the dialogue to	The summary – which contains the overall feedback about the news source is not as visible.	The summary should be brought further up in the results box and made more visible.	0 1 2 3 4

	another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.			
7	Flexibility and efficiency of use	Issues	Recommendation	Severity
	Accelerators – unseen by the novice user may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	No accelerators	If the other issues are addressed, I don't think there would be a need for accelerators.	0 1 2 3 4
8	Aesthetic and minimalist design	Issues	Recommendation	Severity
	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	No issues.		0 1 2 3 4
9	Help users recognize, diagnose, and recover from errors	Issues	Recommendation	Severity
	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	Not much feedback		0 1 2 3 4
10	Help and documentation	Issues	Recommendation	Severity
	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task. List concrete steps to be carried out, and not be too large.	No option for help or tutorial	More intuitive design and help/tutorial.	0 1 2 3 4
0 - I don't agree that this is a usability problem at all. 1 - Cosmetic problem only. Need not be fixed unless extra time is available on project. 2 - Minor usability problem. Fixing this should be given low priority. 3 - Major usability problem. Important to fix, so should be given high priority. 4 - Usability catastrophe. Imperative to fix this before product can be released.				

Appendix H. Heuristic Evaluations – MediaBias/FactCheck Extension

Heuristic Evaluation MediaBias/FactCheck Extension			Tool Name:	
1	Visibility of System Status	Issues	Recommendation	Severity
	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	On some websites, a rating does not appear. There is no feedback indicating if it is processing or if there is an option to refresh the extension itself.	There should be feedback indicating if the extension is processing information or refresh.	0 1 2 3 4
2	Match between system and the real world	Issues	Recommendation	Severity
	The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	The terms for ratings (i.e., pseudo-science, satire, etc.) are established terms, however, not everyone may understand what they mean.	A definition is provided in the extension window that appears, but it may be more helpful to use more general terms that all can understand + use the space within the window to provide reasoning for the rating, specific to the source.	0 1 2 3 4
3	User control and freedom	Issues	Recommendation	Severity
	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support, undo and redo.	Not an issue for browser extensions.		0 1 2 3 4
4	Consistency and standards	Issues	Recommendation	Severity
	Users should not have to wonder whether different words, situations, or actions mean the same thing.	The terminology can be difficult for some to understand (i.e., those who speak English as a second language). In addition, the icon when a rating is not provided is in the form of a red microphone, which looks like it is recording audio.	The terminology used should be simplified and the icon should be changed so users don't worry that the extension is recording their audio.	0 1 2 3 4
5	Error Prevention	Issues	Recommendation	Severity
	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	If the extension is not working, there should be a way to refresh it.	Provide an option to refresh the extension.	0 1 2 3 4
6	Recognition rather than recall	Issues	Recommendation	Severity
	Minimize the user's memory load by making objections, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	There should be recommendations to pin the extension as this greatly enhances the user experience.	Provide a brief overview of the functions the extension performs and recommendations for an improved experience while using it.	0 1 2 3 4
7	Flexibility and efficiency of use	Issues	Recommendation	Severity
	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	Not applicable.		0 1 2 3 4
	Aesthetic and minimalist design	Issues	Recommendation	Severity
	Dialogues should not contain information which is irrelevant or rarely needed. Every	The definition of the rating is provided in a paragraph	It would be more suitable to structure information using	0 1 2 3 4

8	extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	that takes up the whole extension window.	headers, bullet points, numbering, etc. rather than putting it all together in a single paragraph.	
	Help users recognize, diagnose, and recover from errors	Issues	Recommendation	Severity
9	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	No error messages are provided. The main issue that could be experienced is that it does not provide a rating, in which there is no known way to recover from the error.	Provide an option to refresh the extension.	0 1 2 3 4
	Help and documentation	Issues	Recommendation	Severity
10	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task. List concrete steps to be carried out, and not be too large.	Some information is provided when you first install the extension and on the web store page, however, it is overwhelming and cluttered.	Simplify the documentation and provide guidelines for a good user experience (i.e., to pin the extension).	0 1 2 3 4
<p>0 - I don't agree that this is a usability problem at all. 1 - Cosmetic problem only. Need not be fixed unless extra time is available on project. 2 - Minor usability problem. Fixing this should be given low priority. 3 - Major usability problem. Important to fix, so should be given high priority. 4 - Usability catastrophe. Imperative to fix this before product can be released.</p>				

Appendix I. Online Survey – Digital Disinformation Issues and Behaviours

Study Context

The questions in this survey pertain to individual opinions and behaviors related to hotly debated and controversial topics. These are topics for which people might have different views and beliefs, and there may be disagreements about the accuracy or reliability of information.

We will start the survey with some basic demographic questions, and questions about your use of the Internet and social media platforms:

Demographic Questions

1. Gender

What Is Your Gender?

- Male
- Female
- Non-Binary
- Prefer Not to Disclose
- Prefer to Self-Describe

2. Age

Select Your Age Range:

- 18–24
- 25–34
- 35–44
- 45–54
- 55–64
- 65+

3. Race/Ethnicity

What is your cultural background? Choose all that apply:

- African
- European
- East Asian
- South Asian
- South East Asian
- First Nations or Indigenous (please specify) _____
- Hispanic or Latinx
- Middle Eastern
- Other (please specify) _____
- Prefer Not to Answer

4. Income Class

If you were asked to use one of these commonly used income classes according to your country of residence, which would you say you belong in?

- High income
- Upper-middle income
- Middle income
- Low income
- Poor
- Prefer Not to Answer

5. Education

What is Your Highest Level of Education (obtained or in-progress)?

- Less than high school degree
- High school graduate
- Some college but no degree
- Associate degree in college (two years)
- Bachelor's degree in college (four years)
- Master's degree
- Doctoral degree

6. Political View

Some people talk about politics in terms of left, center or right. On a scale from 1 to 5, with 1 indicating left and 5 indicating right, where would you place yourself? (click on the grey slider bar below)

Left	Leaning Left	Center	Leaning Right	Right	Don't Know / Not Sure
	1	2	3	4	5

Political Partisanship	
------------------------	--

Technographic Questions:

7. In terms of your Internet skills, do you consider yourself to be:

- Not very skilled
- Fairly skilled
- Skilled
- Very skilled
- Expert

8. How long have you been using the Internet?

- Less than 6 months
- 6 to 12 months
- 1 to 3 years
- 4 to 6 years
- 7 years or more

9. To what extent do you use the following instant messaging applications?

	Don't Use At All	Use Rarely	Use Sometimes	Use Often	Use All The Time
WhatsApp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Telegram	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Weixin/WeChat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Facebook Messenger	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Signal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
QQ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. To what extent do you use the following social media platforms?

	Don't Use At All	Use Rarely	Use Sometimes	Use Often	Use All The Time
Facebook	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Twitter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instagram	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
LinkedIn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TikTok	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Snapchat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reddit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quora	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Information Seeking Behaviour

11. While consuming or sharing information through social media?

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
I tend to rely on information I have before me rather than looking around for more information.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I pay more attention to catchy headlines or appealing images.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I focus on a few key points.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

12. While discussing or researching a hotly debated or contentious topic, I tend to instinctively:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
Look for evidence that supports my ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Believe information that is shared by close connections	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Seek information that aligns with my way of thinking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Share information I consider to be aligned with my social cause	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

13. I feel confident in:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
Detecting false information on social media	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My skills to check for information authenticity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using tools and techniques to verify information credibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14. Thinking about news (by news we mean information about events and issues that involve more than just your friends and family), how often do you:

	Never	Rarely	Sometimes	Often	Always
Read any newspapers in print?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Listen to news on the radio?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Watch local television news?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Watch national evening network television news?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Get news from a social media sites?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Get news from a website or app?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15. Thinking about when you are interacting with news items on a social media site, how often, if at all, do you:

	Never	Rarely	Sometimes	Often	Always
Click on links to news stories?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Post links to news stories yourself?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Share or repost links to news stories that were originally posted by someone else?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“Like” news stories?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Post my own photos or videos of a news event?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Discuss issues in the news with others on that site?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16. I might share news without verification because:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
I don't have time to check its authenticity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Verifying information is not worth spending time on	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Verifying information is very time consuming	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Information Ecosystem Perceptions

17. Please indicate your level of agreement with the following statements about hotly debated or contentious topics:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
There is lack of quality reporting about such issues.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There is not enough information on such subjects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The pace of information communication about such topics is slow.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Information related to such topics keeps constantly changing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

18. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
The information I get from social media is reliable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can trust information that I get from social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I consider information obtained from social media to be credible.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Sociological Environment Perceptions

19. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
Mainstream media is a reliable source of information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mainstream media often exaggerates and sensationalizes the news	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mainstream media often represents their own bias and interests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

20. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
I have confidence in government institutions to carry out their responsibilities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Government institutions are run by a few big interests looking out for themselves	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most people running the government are crooked	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I generally trust government institutions to do what is right	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Cognitive Perceptions

21. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
I believe other people are very concerned about the spread of misinformation on social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I believe other people are very concerned about the authenticity of news that they receive on social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I believe misinformation misleads other people's understanding of specific topics.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

22. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
I have thought about the possible consequences caused by misinformation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I often think about the issue of misinformation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I often recall instances where I have witnessed misinformation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

23. Please indicate your level of agreement with the following statements

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
Many important things happen in the world which the public is never informed about.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Politicians usually do not tell us the true motives for their decisions.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Government agencies closely monitor all citizens.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Events which superficially seem to lack a connection are often the result of secret activities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There are secret organizations that greatly influence political decisions.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

24. To what extent are you concerned about the negative effects of the following types of online content:
(Note: If you are not familiar with any of the terms listed below, please select the "Not Sure or Don't Know" option)

	Not At All Concerned	Slightly Concerned	Somewhat Concerned	Moderately Concerned	Extremely Concerned	Not Sure or Don't Know
Clickbait	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deepfakes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Satire	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fake News	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Propaganda	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conspiracy Theories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PseudoScience	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sponsored Content	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Affect and Attitude

**25. When you encounter misinformation about a hotly debated or contentious topic, to what extent do you feel:
(click on the grey sliders below)**

	Not At All	A Little	Moderately	Very	Extremely
	1	2	3	4	5
Anxious					
Angry					
Depressed					
Nervous					
Worried					
Outraged					
Annoyed					

26. Please indicate your level of agreement with the following statements:

To verify whether information is accurate:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
It is important to check the original source of information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important to check who first uploaded the information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important to check when the information was first created	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important to check where the information was first created	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Disinformation Countering Responses

27. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
When I detect misinformation on social media, I would report it to the platform.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I detect misinformation on social media, I would place a complaint against its author.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would check the authenticity of the misinformation message before I forward it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>


28. To what extent have you used the following tools or techniques to safeguard yourself against consuming or sharing misinformation online:(Note: If you are not familiar with any of the tools or techniques listed below, please select the "Not Sure or Don't Know" option)

	Never Used	Use Rarely	Use Sometimes	Use Often	Use All the Time	Not Sure or Don't Know
Search engine queries to verify information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fact-Checking Websites such as Snopes or HoaxSlayer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Browser Plugins or Extensions such as NewsGuard or Fake News Alert	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bot Detection Tools such as BotOMeter or BotSentinel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (Please Specify):	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

29. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
Accounts who post misinformation on social media should be removed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would support legislation to prohibit the spread of misinformation on social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would support that misinformation should be blocked/censored by social media platforms.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would sign an online petition demanding the government to contain the spread of misinformation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

30. How satisfied are you with the way social media platforms currently deal with the issue of misinformation? On a scale from 1 to 5, with 1 indicating "Not At All Satisfied" and 5 indicating "Extremely Satisfied", please indicate your level of satisfaction. (click on the grey slider bar)

	Not At All Satisfied	Slightly Satisfied	Moderately Satisfied	Very Satisfied	Extremely Satisfied	Don't Know / Not Sure
	1	2	3	4	5	
Level of Satisfaction						

31. Please indicate your level of agreement with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
It is important that social media users be taught how to analyze online information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important that social media users be taught how to recognize false or misleading information online	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important for social media users to understand how to evaluate information critically	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Open Ended Question

32. (Optional) Please feel free to share any additional thoughts about online misinformation:
