# Verification, monitoring and responsible reporting in an age of information disorder

## A guide for practitioners in Southeast Asia

By: Dr. Anne Kruger (APAC Director)
Esther Chan (APAC Bureau Editor)
Stevie Zhang (Associate Editor)
First Draft APAC bureau, UNESCO IPDC grant recipient

# Acknowledgments

# Table of contents

# 1. Introduction

**The landscape of mis- and disinformation circulating in online spaces and efforts to slow this down vary across Southeast Asia, with different political systems, socio-economic profiles, languages, cultural and religious backgrounds across the region.**

As Southeast Asia experts[1] note, this is also a region of "autocratic and illiberal democratic regimes whose information environment has long been marked by censorship and low press freedom". The recent wave of health misinformation that spread globally, spurred by the Covid-19 pandemic, was further compounded in countries such as Myanmar which experienced a "confluence of crises"[2] after the coup. During this time, barriers to vaccine access and the global spread of online rumours exposed less developed countries to the dangers of "information disorder" in terms of health and safety[3].

That is not to say that information disorder was new to the region, which remains one of the most dangerous places for journalists to work in[4]. Studies about disinformation in Southeast Asia are still lagging, despite ample reason for deeper analysis, especially as disinformation research gained prominence over the past years[5]. As we explore below, governments in the region have taken unprecedented steps in legislating against mis- and disinformation, with researchers noting that some of these laws are cracking down on political dissent instead. Meanwhile, fact-checking initiatives continue to spring from both governments and grassroots efforts.

In 2018, UNESCO launched "Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training"[6]. As there is no one specific disinformation landscape[7] and systems

---

1 Email quote from Dr Aim Sinpeng from the University of Sydney with approval

2 Son, J. & Paing, K.T. (2021, July 16). COVID-19 Disaster Adds to Confluence of Crises in Myanmar. *Reporting ASEAN*. https://www.reportingasean.net/covid-19-disaster-adds-to-confluence-of-crises-in-myanmar/

3 Carson, A. & Fallon, L. (2021, January). *Fighting Fake News: A Study of Online Misinformation Regulation in the Asia Pacific*. La Trobe University. https://www.latrobe.edu.au/__data/assets/pdf_file/0019/1203553/carson-fake-news.pdf

4 Email quote from Dr Aim Sinpeng from the University of Sydney with approval

5 Xia, Y. (2021, August 22). Disinformation after Trump. *Media, Culture & Society* 43(7). https://doi.org/10.1177/01634437211040684

6 UNESCO. (2018). *Journalism, fake news & disinformation: handbook for journalism education and training*. https://en.unesco.org/fightfakenews

7 Haciyakupoglu, G. (2019, February 12). Southeast Asia's Battle Against Disinformation. *The Diplomat*. https://thediplomat.com/2019/02/southeast-asias-battle-against-disinformation/

and governments differ across Southeast Asia, First Draft[8] was approached in 2021 by UNESCO to create a short booklet based on our 'information disorder' approach, to draw out some concepts from the Handbook applicable to the region. This booklet expands upon and supplements the handbook by briefly examining the recent legal developments in Southeast Asia, situating them in their respective social and political contexts, alongside examples of mis- and disinformation encountered in our research. These examples help illustrate the types of information shared in Southeast Asian social networks as well as the motivations for sharing, better grounding concepts on how best to fight the spread of disinformation, such as inoculation theory.

Journalists and media professionals in the sub-region are only too aware of the real-life consequences that stem from information disorder in the cybersphere[9]. The aim of the booklet is to support Southeast Asia's journalists, academics, students and civil society in their education and skills development in the fight against mis- and disinformation. As UNESCO Director General Audrey Azoulay and Es'hailSat Chairperson and Working Group co-chair Dr Hessa al-Jabar noted in the UNESCO 2020 Broadband report, "UNESCO defends the essential role journalists play in our societies – by encouraging public debate, they help build citizen awareness."[10] This booklet recognises the essential role of independent (those who can report without fear of interference or influence) and professional journalists to monitor for misinformation, debunk and pre-empt problematic areas to provide quality information and promote healthy civil discourse. Yet it also acknowledges the realities and differences in laws and practices in the region.

This booklet acknowledges that every method of

combating disinformation must be situated in the specific social, political, and economic contexts of where it will be deployed — hence the need for this region-specific booklet. This guide begins with an overview of developments in legal frameworks and fact checking activities in the region, which uses a thematic approach to place into a Southeast Asian context the opening chapters from the 2018 handbook "Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training"[11].

First Draft's definitions and a conceptual framework of information disorder are then outlined. This is based on the 2018 handbook's Module Two: "Thinking about 'Information Disorder' by Claire Wardle & Hossein Derakshanin", which provides a scaffold for examples from the Southeast Asia region. This booklet therefore situates the spectrum of information disorder and provides activities that will help readers apply the concepts in their own context and communities.

The examples draw on desktop research and email interviews with experts as well as first-hand experience from the First Draft Asia-Pacific staff, who have conducted in-person training programs and verification research throughout the region prior to the pandemic, including in Cambodia, Myanmar, Vietnam, Thailand, Singapore, Malaysia, the Philippines and Indonesia; with new and additional online training and collaborations during the pandemic with Brunei, Malaysia, Singapore and Laos. The case studies and short overview of laws and fact-checking is indicative only of themes, and examples supporting the definitions and issues is by no means exhaustive, due to the diverse and ever-changing "disinformation landscape"[12] of Southeast Asia. It also takes into account safety considerations of those in our networks living in the region.

The guidebook further contextualises fact-checking from Module Five and verification from Module Six of the 2018 handbook in Southeast Asia by outlining practical steps to monitoring — including search and setting up workflows — allowing readers to sharpen their skills in verification as applied to their own region.

Then, journalists and civil society can learn about the

---

8  First Draft's mission has moved to the Information Futures Lab, at Brown University. The Lab builds on the legacy of media verification, journalism education and innovation pioneered by the nonprofit First Draft, with the ongoing work moving to the Lab after six years of groundbreaking work.

9  Towhid, A.B. (2021, July). INFORMATION DISORDER IN THE INFORMATION AGE: ACTORS, TACTICS, AND IMPACTS IN SOUTH AND SOUTHEAST ASIA. *Bangladesh Institute of International and Strategic Studies* Papers 28. http://dx.doi.org/10.13140/RG.2.2.18056.34568

10  Broadband Commission for Sustainable Development. (2020, September). B*alancing Act: Countering Digital Disinformation While Respecting Freedom of Expression.*  https://www.broadbandcommission.org/wp-content/uploads/2021/02/WGFoEDisinfo_Report2020.pdf

11  UNESCO, 2018.  *Journalism, fake news & disinformation*.

12  Haciyakupoglu, 2019. Southeast Asia's Battle Against Disinformation.

psychology of misinformation to better understand why we're vulnerable, why misinformation can be so difficult to correct, and how better to prevent it from spreading and influencing people. This is followed with details on how to report responsibly and with maximum effect.

As an extension of the 2018 Handbook (and indeed other training resources), this booklet proposes and demonstrates a possible new workflow for journalists, academics, students and civil society invested in the fight against mis- and disinformation, particularly ahead of an event likely to draw disinformation campaigns, such as an election. Before the event, an in-depth understanding of the misinformation landscape of the place is crucial. This can be achieved by analysing the platforms and communication methods local residents use, legislation in place to curb disinformation, if any, and any existing fact-checking initiative already working to combat information disorder. Trusted information providers may then use these concepts to implement prebunking — preparing the public for any potential narratives that may circulate during the event.  This proactive approach is essential to build further impact to slow down and prevent mis- and disinformation from spreading, compared to the usual reactive measures organisations typically deploy during or after the event.

As noted above, the aim of this guide is to support journalists, students, communications professionals,or anyone who may be relied upon for quality information in an era of information disorder. We were invited to include key lessons, activities and resources based on First Draft's definitions and foundational research and explain how and why false information spreads. This guide can be used to develop and build specialist skills into workflows and learn how to slow down the spread of mis - and disinformation. It can be used in creating educational curriculum for practical application.
We wish all stakeholders well in these important endeavours.

Sincerely,

Dr Anne Kruger, Esther Chan and Stevie Zhang

Editorial team, October 2022

# 1.1 Law 'laboratory'

Getting the balance right in the fight against misinformation is a huge challenge and a complex terrain to navigate for governments, stakeholders and policy makers. The pandemic has shown the urgent need for all stakeholders to address information disorder for the betterment of society, as the region continues to grapple with an 'infodemic'. Even well before the pandemic, the region saw the urgent need for stakeholders to work together to address the effects of online misinformation: a prime case study being how misinformation resulted in violent offline action in 2014, as Myanmar went from limited news to smartphones almost overnight.

Research has described Southeast Asia as the "world's most vibrant laboratory of anti-fake news laws since 2018,"[13] however, the nature of this "laboratory" is often highly sensitive. Concerns about the ramifications and implementation of these laws have been a particularly weighty issue in Southeast Asia. Malaysia's parliament enacted the 'Anti-Fake News Act' in April 2018, in the midst of reports about the 1MDB corruption scandal; however this was repealed in December 2019. Meanwhile, Singapore's 'Protection from Online Falsehoods and Manipulation Act' — known as POFMA — captured global attention throughout its development and was passed by parliament in May 2019. Researchers note that "since the Act entered into force in October 2019, the government has used it on several occasions to order official corrections to be posted next to news or social media posts."[14] In addition, "all actions taken under POFMA are documented in the POFMA Office's media centre". Critics noted new laws by Indonesia and Singapore to fight and slow down misinformation were "alleged to have been misused by governments to crack down on political dissent and suppress freedom of expression and the media."[15]

The background information chapter in UNESCO's 2018 Handbook noted:

*While some governments are seeking to tackle the problems through regulation, it is unlikely that this can resolve the matter at scale. Moreover, it carries major risks of abuse, where legitimate freedom of expression and authentic journalism can become subject to new censorship by a 'Ministry of Truth'.*

This is further nuanced in Southeast Asia, where the implementation of laws needs to be understood against political landscapes they exist in. Take, for example, Vietnam, where a cybersecurity law was introduced in 2019 and followed up by national guidelines on social media behaviour in June 2021. Vietnam's government and the ruling Communist Party own most of the country's press, and any privately owned news outlets are tightly controlled by the state (however, most media organisations are financially independent, with little state subsidy). In Indonesia, while "unfounded political propaganda was an integral part of the 32-year authoritarian rule of President Suharto" which ended in 1998, "the post-Suharto era of democracy" has been plagued by disinformation filling new space for political competition and "exploiting the religious and ethnic fault lines that seem to consume the country."[16]

As the further readings below show, the regions' "anti-fake laboratory" has been criticised as too restrictive, used as an excuse to enhance political or government controls, and target journalists — all of which result in the erosion of trust in media as reliable sources of information. For example, Myanmar's military recently proposed a cybersecurity bill that included sections to limit misinformation, but Human Rights Watch said the provisions "would allow the authorities to order the removal of any content critical of individual military leaders or others linked to the junta". In Cambodia, the Ministry of Information's "Fake News Monitoring Committee" reported nearly 2,000 cases of "fake news" in 2021, amid calls from press freedom groups for more transparency over its methodology.

Indonesia's primary piece of legislation that governs

13 Schuldt, L. (2021). Official Truths in a War on Fake News: Governmental Fact-Checking in Malaysia, Singapore, and Thailand. *Journal of Current Southeast Asian Affairs* 40(2) 340-371. https://doi.org/10.1177/18681034211008908

14 *Ibid*.

15 Carson & Fallon, 2021. *Fighting Fake News.*

16 Kaur, K., Nair, S., Kwok, Y., Kajimoto, M., Chua, Y.T., Labiste, M.D., Soon, C., Jo, H., Lin, L., Le, T.T., Kruger, A. (2018, October 10). Information Disorder in Asia and the Pacific: Overview of Misinformation Ecosystem in Australia, India, Indonesia, Japan, the Philippines, Singapore, South Korea, Taiwan, and Vietnam. http://dx.doi.org/10.2139/ssrn.3134581

digital content and hate speech is the 2008 'Electronic Information and Transaction Law', created to address commentary online.[17] In 2018, the government launched the National Cyber and Encryption Agency (BSSN) focused on hoaxes. This agency reports directly to the president — a system which has been criticised by researchers that it "creates the potential for arrests that are politically motivated and based largely around whether the content is anti-government, leading to a perception that the government does not tolerate criticism and is using the security forces as an instrument of political repression."[18]

In 2022, Thailand's approval of new anti-fake news regulations was the culmination of the country's long running efforts to regulate online misinformation, beginning with the country's Computer Crime Act of 2007. Researchers note this is set against recent history where the "2014-2019 period of military rule consolidated and expanded... bureaucratic-legal infrastructure... the junta packed once independent and civilian entities such as the National Broadcasting and Telecommunications Commission (NBTC) with military loyalists."[19] The same research notes that the government's Anti-Fake News Centre (AFNC) fact-check system, founded in 2019, "often cites official sources as true news despite these merely being an official interpretation of political events".[20]

On the other hand, there is the Philippines, where local Facebook pages coordinated in 2016 to promote a hyper-partisan agenda[21]. Then-President Rodrigo Duterte resisted passing "fake news laws" during a 2017 senate review; instead, as CNN Philippines reported, Duterte "said lawmakers should focus on increasing penalties for libel and slander." Researchers in the Philippines noted in 2018 that "President Rodrigo Duterte has emerged as a, if not the, major source or creator of state-level misinformation, disinformation and mal-information delivered in conventional platforms

such as press conferences, interviews and speeches, including the State of the Nation Address."[22] In June 2020, Rappler co-founder Maria Ressa made international headlines once more after she was found guilty of 'cyberlibel' in a series of legal investigations she faced in the Philippines. The decision was upheld by the Court of Appeals in July 2022.

The 'law laboratory' phase was further challenged by complications in addressing the Covid-19 pandemic. The sharing of false news is a criminal offence in Cambodia where arrests were made in relation to Covid-19[23]. During the pandemic, Malaysia introduced emergency "fake news" laws amid opposition, as reported by Reuters, that "the new law was too sweeping and could be used to instil fear among the public about criticising the government." While laws against fighting Covid misinformation might be well intentioned, arrests in Thailand (see media reports by the Bangkok post below), gave way to fears that people could be arrested for sharing "fake virus news" accidentally, despite having no ill intent.

Indeed, the pandemic raised a nuanced point about the definitions of misinformation (see chapter 2.1 below), in that people were often unwittingly sharing information that they didn't realise was false, but did so as they thought they were protecting their loved ones. This is supported by findings from a research team led by Dr Phansasiri Kularb at Chulalongkorn University which "examined messages related to Covid-19 that were flagged by fact-checkers — the Anti-Fake News Centre (AFNC), Cofact, AFP Fact Check and Sure And Share", (see "Gone Viral" by the Bangkok Post in the further readings below). The study noted, "A closer look at messages labelled as misleading or fabricated revealed that they usually contain warnings for people to protect themselves more than an intent to stir up public panic or attack the government."

While Facebook is seen as 'the internet' in many areas of Southeast Asia, researcher Ross Tapsell noted in 2021 that "the biggest shift caused by fear of crackdown is the rise of discourse via closed

17  Tapsell, R. (2019, September 20). Indonesia's Policing of Hoax News Increasingly Politicised. *ISEAS Perspective* 2019(75). https://www.iseas.edu.sg/images/pdf/ISEAS_Perspective_2019_75.pdf

18  *Ibid*.

19  Sombatpoonsiri, J. (2022, April 7). Labelling Fake News: The Politics of Regulating Disinformation in Thailand. ISEAS Perspective 2022(34). https://www.iseas.edu.sg/wp-content/uploads/2022/03/ISEAS_Perspective_2022_34.pdf

20  *Ibid*

21  Kaur et al, 2018. Information Disorder in Asia and the Pacific.

22  *Ibid*.

23  UNDP. (2021, September 9). *Combatting Disinformation: A Preliminary Assessment of the Information Landscape in Cambodia*. https://www.undp.org/cambodia/publications/combatting-disinformation-preliminary-assessment-information-landscape-cambodia

messenger groups, especially WhatsApp in Indonesia and Malaysia, LINE in Thailand, and Facebook Messenger in the Philippines." The reliance on closed app groups has brought about ongoing challenges in how to slow the spread of mis- and disinformation, particularly on encrypted services, and deserves ongoing research.

Additionally, the growing popularity of voice messaging in Southeast Asia may be a convenient means of communicating, but offers new space for the sharing of mis- and disinformation. This is increasingly popular between friends and family in Cambodia, although "the precise source of the information shared is often unknown, with some senders claiming the voice messages are that of government officials or authorities; however, this cannot always be verified."[24]

# FURTHER READING

→ "Myanmar: Scrap Draconian Cybersecurity Bill" Human Rights Watch, July 15, 2022.

→ "Southeast Asia's Battle Against Disinformation" The Diplomat, February 12, 2019, by Gulizar Haciyakupoglu.

→ "Two held for sharing fake news on virus", Bangkok Post, 30 January 2020.

→ "Pair nabbed for spreading fake virus news", Bangkok Post, 4 March 2020.

→ "Govt adopts multi-level approach to fake news", Bangkok Post, 2 February 2022.

→ "Information Disorder in Asia and the Pacific: Overview of Misinformation Ecosystem in Australia, India, Indonesia, Japan, the Philippines, Singapore, South Korea, Taiwan, and Vietnam" see SSRN, October 10, 2018, by Kanchan Kaur et al.

→ "Fighting Fake News: a study of online misinformation regulation in the Asia Pacific" La Trobe University, January 2021, by Andrea Carlson, with Liam Fallon.

→ "Official Truths in a War on Fake News: Governmental Fact-Checking in Malaysia, Singapore, and Thailand" SAGE Journals, May 2021, by Lasse Schuldt.

→ "Malaysia defends coronavirus fake news law amid outcry" *Reuters*, 13 March, 2021.

→ "Propaganda war: Weaponizing the internet", Rappler, October 3, 2016.

→ "Gone Viral: Covid-19 and fake news in Thailand", Bangkok Post, 2021.

→ 'Hoax killed my father': Indonesia's other pandemic Finding different ways to address hoaxes and misinformation" August 45, 2021, Unicef Indonesia by Jenny Lei Ravelo.

---

24 *Ibid.*

# 1.2 Fact checking developments

The number of fact checking organisations more than doubled in Asia from 2019-2020[25], and for many years, Southeast Asia has paved the way with some of the most innovative and creative approaches. In Indonesia, Mafindo began in 2015 as a small community-driven Facebook group: 'Forum Anti Fitnah Hasut dan Hoax' (Anti-Hoax Slander and Sedition Forum). Its network of fact-checkers now includes over 85,000 members in its online forum who participate in crowdsourced "hoax busting". CekFakta, a fact-checking consortium comprising International Fact-Checking Network (IFCN) certified fact-checkers and a number of local publishers, was launched in Indonesia in 2018 by 22 media companies. They also collaborate with the Google News Initiative and experts such as the Indonesian Cyber Media Association (AMSI), the Indonesian Anti-Slander Society (Mafindo), the Alliance of Independent Journalists (AJI), and Internews to tackle mis- and disinformation unique to their country.

Other innovative programs include Thailand's collaborative Cofact project. Launched in 2020 and modelled on a project of the same name in Taiwan, its website describes it as "the integration between the use of "civic tech" and "journalism", where the editorial team works with volunteers in verifying news or content from credible sources and providing the space for the public to share and to exchange factual information and personal opinions".

There has also been strong growth in government-run fact checking outlets in Southeast Asia, including Singapore (Factually since 2012); Malaysia (sebenarnya.my — a government portal launched in 2017, as well as MyCheck — a fact-checking unit launched by the government-owned news agency Bernama in March 2020); Thailand (Anti-Fake News

Center since November 2019 as mentioned in the 'law laboratory' section above) and Vietnam (Anti-Fake News Center launched in January 2021).

Many of these government fact checking units have helped to debunk Covid misinformation throughout the pandemic and other recent crises, which has drawn some acknowledgement by independent fact-checkers as beneficial on those topics. On the other hand, as noted in the 'law laboratory' section above, government involvement in fact check initiatives brings to bear questions about influence and independence, in particular on more political topics. Researchers note an imbalance exists with the high number of government institutions and agencies with the power to prosecute compared with independent fact-checking organisations[26]. A 2021 article in the Journal of Current Southeast Asian Affairs into government-led fact checking in Malaysia, Singapore and Thailand argued that the correction practices mainly existed to perpetuate an "omnipresent fake news threat" and could pave the way for restrictive laws[27].

Other initiatives include hybrid as well government collaborations with fact checking units in Indonesia (StopHoax.id launched in 2018 by the Communications Ministry, however the IP address is no longer managed by the Ministry, in its place is Covid19.go.id at the time of writing); and Thailand's SureAndShare centre initiative run by the state-owned national broadcast MCOT (since 2015). The project also collaborates with universities: for example, Thammasat University's "Fake News Fighter Project: "A Fact-Checking Mechanism In Television Newsrooms and Digital Media" trains students in a newsroom simulation setting.

In Cambodia, the United States Agency for International Development (USAID) supported the Women's Media Centre in Cambodia to develop and launch Cambodia's first fact-checking Khmer language website in 2020, however another fact-checking project launched in July 2021 was closed very shortly after it began.

25  Stencel, M. & Luther, J. (2020, October 13). Fact-checking count tops 300 for the first time. Duke Reporters' Lab. https://reporterslab.org/fact-checking-count-tops-300-for-the-first-time/ and Stencel, M. (2019, June 11). Number of fact-checking outlets surges to 188 in more than 60 countries. Duke Reporters' Lab https://reporterslab.org/number-of-fact-checking-outlets-surges-to-188-in-more-than-60-countries/

26  Sinpeng, A. & Tapsell, R. (2020). From Grassroots Activism to Disinformation: Social Media Trends in Southeast Asia. In From Grassroots Activism to Disinformation: Social Media in Southeast Asia (p 1-18). ISEAS – Yusof Ishak Institute. https://bookshop.iseas.edu.sg/publication/2454

27  Kaur et al, 2018. Information Disorder in Asia and the Pacific

The linguistic diversity of Southeast Asia also adds to the challenge of fact checking. Resources, expertise in disinformation and language proficiency are key to getting corrective messages across groups that speak different languages or dialects. In other words, there is no 'one size fits all'.

## FURTHER READING

→ "Faster Facts: The rapid expansion of fact checking", News In Asia JNI Institute & Hong Kong University, September 2021, by Masato Kajimoto.

→ "Fact-checking count tops 300 for the first time" Duke Reporters' Lab, October 13, 2020, by Mark Stencel & Joel Luther.

→ "Welcome to 'year of fact-checking'" Bangkok Post, 3 April 2021.

→ "New innovation "CoFact" launched to fight fake news" ThaiHealth, 28 Aug 2020.

→ "Fake News Fighter Project, A Fact-Checking Mechanism In Television Newsrooms and Digital Media" Thammasat University, 9 March 2022.

→ "'What You Share Defines You': Indonesia Has World's Biggest Fact-Checking Network" Jakarta Globe, 16 January 2021.

# 2. Information Disorder

**Imposter websites designed to look like professional outlets, sock puppet accounts posting outrage memes, click farms, online conspiracy communities busy trying to fool reporters into covering rumours or hoaxes — the term 'fake news' doesn't begin to cover all of this. Most of this content isn't even fake: it's often genuine, used out of context and weaponized by people who know that falsehoods based on a grain of truth are more likely to be believed and shared. And most of this can't be described as 'news' — it's rumours, it's memes, it's manipulated videos and old photos re-shared as new.**

First Draft advocates using the terms that are most appropriate for the type of content; whether that's propaganda, lies, conspiracies, rumours, hoaxes, hyperpartisan content, falsehoods or manipulated media. We also prefer to use the terms disinformation, misinformation or malinformation. Collectively, we call it 'information disorder'[28]. This has been adopted globally, including in Southeast Asia, and serves as a foundation

for which to understand the phases of information, which includes the original creation of a message, how it is then produced and turned into a media production; how it is distributed; and then a recirculation of these messages being further picked up, reproduced and distributed — and is therefore subject to change from its original intention or provenance. This part of the cycle is potentially limitless.

One reason why many people may fall for mis- or disinformation is that it is often anchored in a 'kernel of truth'; however this 'kernel' is often twisted or taken out of context. That makes messages spread faster because it's harder to determine what part of the information is false. Some prominent historical cases of harm caused by vaccine hesitancy due to this 'grain of truth' come from Asia, including the Philippines' controversy around the Dengvaxia vaccine.

28  Wardle, C. & Derakhshan, H. (2017, September 27). *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.  https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

# 2.1 Definitions of misinformation, disinformation and malinformation

The term 'fake news' does not cover the complexity of the phenomenon, and has been weaponized by politicians to discredit seemingly critical media stories or questions. Whilst this is most famously attributed to former US President Donald Trump, the term is prevalent throughout Asia. Researcher Janjira Sombatpoonsiri suggested this stubborn prevalence of the "fake news" term is somewhat due to convenience for official use, "because maintaining a clear-cut dichotomy between what is considered as true and false information makes it convenient to associate information diverging from that of official sources with falsehood that yields harmful effects on the public at large."[29] To summarise these terms broadly:

MISINFORMATION is false information spread mistakenly, even if people genuinely believe it, or are trying to be helpful. The sharing of misinformation is driven by socio-psychological factors. Online, people perform their identities. They want to feel connected to their "tribe", whether that means members of the same political party, parents that don't vaccinate their children, activists who are concerned about climate change, or those who belong to a certain religion, race or ethnic group. During the pandemic, there was an influx of family members forwarding online messages to their loved ones with 'Covid cures' which they genuinely believed would be helpful, but were actually harmful, and in some cases with deadly consequences.

For example, misinformation promoting the unproven use of herbal medicine as a cure for Covid-19 was prominent throughout the pandemic, especially before the vaccines were available. Thailand's Anti-Fake News Center found misinformation circulating online alleging that a concoction of ginger, galangal, lemongrass, pepper, garlic and shallot would make an effective Covid treatment. The Department of Thai

Traditional and Complementary Medicine dismissed the claim as "false information", according to the Center in a July 2022 report. Misinformation about alternative treatments is often reported when there is a data deficit around a subject, i.e. when the amount of credible information isn't enough or available to answer all the questions and address all the concerns. This was the case during numerous points of the Covid-19 pandemic, especially when there was a lack of data about vaccines' effectiveness and possible adverse reactions.

Elsewhere, WhatsApp messages circulating in Brunei in September 2021 that authorities planned to tighten Covid-19 measures had been dismissed as false by the country's Ministry of Health. Bruneian participants in First Draft's verification training said they too had received the viral misinformation via WhatsApp groups, including those that belong to their families, and it had triggered panic buying in supermarkets[30].

In another example, First Draft saw the circulation of rumours speculating about the role of the Chinese government in the 2021 February 1st Myanmar coup. This was underpinned by existing questions surrounding the relationship between the two countries, as China's expansion of its Belt and Road Initiative into Myanmar led to concerns about a "debt trap" and China later blocked attempts by the UN to condemn the military junta. Claims about alleged Chinese support for the military coup overwhelmed social media, and after the military junta announced plans to introduce sweeping new cybersecurity laws, speculation emerged online that Chinese IT equipment and technicians were being flown in to help implement an internet firewall. These fears were exacerbated when a technical glitch led mobile devices, smart TVs and apps to display random Chinese characters, and people in Myanmar began reporting that they were receiving automated messages from telecommunicators operators in Chinese, rather than Burmese, language. Investigations at the time by the since-closed HK Citizen News found that the gibberish Chinese-language messages were likely the result of an

---

29  Sombatpoonsiri, 2022. Labelling Fake News, *ISEAS Perspective*

30  The Star. (2021, September 27). "Refrain from spreading false information", Brunei health minister tells people as Covid-19 cases keep rising in country. *The Star*. https://www.thestar.com.my/aseanplus/aseanplus-news/2021/09/27/refrain-from-spreading-false-information-brunei-health-minister-tells-people-as-covid-19-cases-keep-rising-in-country

encoding issue on the telecom providers' side, possibly a [side effect of the new internet restrictions](). But due to the fearful political environment and the opaqueness of Chinese-Myanmar relations, these technical errors became just one one node in a large complex network of conspiracy theories that ran rampant online after the coup.

**DISINFORMATION** is content that is intentionally false and designed to cause harm. It is mainly motivated by three distinct factors: to make money; to have political influence, either foreign or domestic; or to cause trouble or mischief for the sake of it. When disinformation is shared it often turns into misinformation. Wardle and Derakhshan[31] further specify the motivations as:

» Financial: Profiting from information disorder through advertising;

» Political: Discrediting a political candidate in an election and other attempts to influence public opinion;

» Socio-psychological: Connecting with a certain group online or off; and, seeking prestige or reinforcement.

For example, the Indonesian-language text inserted on-screen in [a video]() viewed over 700,000 times since November 2021, says "Timor-Leste returns to Indonesia". This is false; [an earlier version of the video](), uploaded by Indonesian broadcaster Kompas TV in August that year, was titled "Hundreds of East Timorese Deported from Indonesia". The description of the video says in part that more than 300 people from Timor-Leste were to be deported after entering Indonesia illegally. While ties have strengthened in recent years, the historically fraught relationship between the two countries can still facilitate politically-driven disinformation.

Disinformation can be used not only to discredit political opponents but also to fuel hate speech and racism. In 2018, Facebook took down hundreds of pages followed by millions of people in Myanmar for "[coordinated inauthentic behavior]()". An investigation by the [New York Times]() revealed shocking content: "The Facebook posts were not from everyday internet users. Instead, they were from Myanmar military personnel who turned the social network into a tool for ethnic cleansing, according to former military officials,
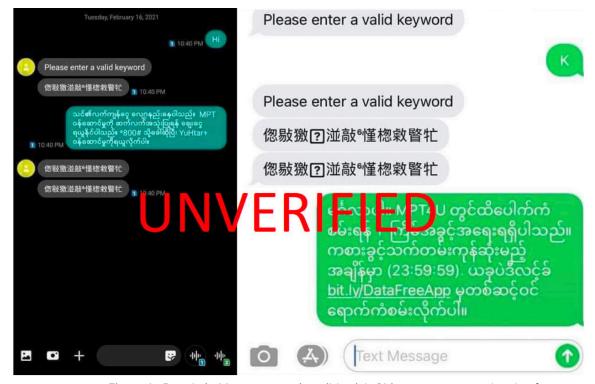


Figure 1: People in Myanmar receive gibberish Chinese messages shortly after coup

---

31 Wardle & Derakhshan, 2017. *Information Disorder*, Council of Europe.

researchers and civilian officials in the country." The report drew together accounts by people who spoke on the condition of "anonymity because they feared for their safety."  In a statement, Facebook thanked the New York Times for the report and noted: "As part of our ongoing investigations into this type of behaviour in Myanmar, we discovered that these seemingly independent news, entertainment, beauty and lifestyle Pages were linked to the Myanmar military." Note, lawyers representing Rohingya people around the world later launched legal actions[32].

As many of our further readings and references note, elections are a particularly fraught time for the creation of disinformation campaigns in Southeast Asia. Myanmar was hit again by disinformation during the 2020 elections.

**MALINFORMATION** is a term that describes genuine information shared with an intent to cause harm. An example can include hacked emails that leak certain details to the public to damage reputations. Revenge porn is also an example of malinformation. In the Philippines, one of "the most visible and the most visceral" motivations behind malinformation would be politics[33]. As a result, researchers note[34], "some victims of disinformation and malinformation have filed libel suits against propagators"[35].

---

32  Rohingya Facebook Claim. https://www.rohingyafacebookclaim.com/

33  Kaur et al, 2018. Information Disorder in Asia and the Pacific.

34  *Ibid*.

35  Ballaran, J. (2017, September 22). Trillanes files raps vs Mocha Uson over 'fake news'. Inquirer.net. https://newsinfo.inquirer.net/932625/trillanes-files-raps-vs-mocha-uson-at-ombudsman-over-fake-news

## 2.1.1 Tips on deciphering intent

The intent behind the publication of a social media post and/or the purported endorsement for a particular online narrative may not always be apparent. Here are some tips to help you decipher the intent:

» Check the publisher's account: Do their bio, cultural background, location and profession match how they describe themselves? Does it make sense that this person would be interested in the particular subject?

» Profile picture: Does the user's photo strike you as odd, perhaps the eyes, ears or teeth appear warped? Artificially-generated images, or "deep fakes", can sometimes create glitches around the eyes, ears and teeth;

» Followership and interests: The users, pages and groups a person follows can give you an idea about their interests and preferences and can help you decide whether they would be likely invested in the issue at hand;

» Language: What language(s) does the user speak? Are they writing or speaking in the language they say they have proficiency in with any grammatical or syntactic mistakes? You may want to consult a native speaker of the language to make sure the nuances in how a language is used colloquially are not missed;

» Digital footprint: Check the user's digital footprint across different social media platforms, forums and other online spaces. This is a good way to form a more comprehensive picture about the user's background, beliefs and interests.

## ACTIVITY
### WHY WE DON'T SAY F*** NEWS

→ Watch a video on why we don't say f*** news here: http://bit.ly/sms-apac-2

→ Here's a reflection on the video from a journalist in Malaysia: "People tend to share misleading news without checking the credibility of the news source. This is common especially with the older generation as they tend to share previous news articles or misleading information concerning a current issue. The reason why this is common with the older folks is because they are technologically [less savvy], they believe their friends more, or they can't differentiate between 'fake news' or credible news. For example, my parents will receive news articles from their friends via Whatsapp and will relate the news with a current happening or issue. When I cross-checked the article, it is often taken from an article published years before or the event did not happen in Malaysia but happened in a different country" – *by Shahirah Hamid, Journalist, Malaysia.*

# 2.2 The 7 types spectrum

Within these three overarching types in information disorder, we also refer frequently to seven categories, as we find that it helps people to understand the complexity of this ecosystem. In "Fake News. It's Complicated"[36], First Draft founder Claire Wardle first outlined the seven major types of mis- and disinformation (Figure 2 ) as a way of moving the conversation away from a reliance on the term 'fake news' and illustrated why the use of more precise language is advised. The 'deceptive seven' still acts as a useful way of thinking about different examples.

These include satire or parody (content that isn't intended to cause harm, but has the potential to fool, although has increasingly been weaponised as we explain below); False connection (headlines, images or captions that over-sell the content such as clickbait); Misleading content (information that frames an issue or a person in a misleading way); False context (genuine content that is shared out of its original context); Imposter content (content that impersonates or falsely claims to be from a genuine source); Manipulated content (genuine information or imagery that is manipulated or edited to deceive): and fabricated content (new content that is 100% false, made to deceive and do harm).

As the diagram shows, we consider this to be a spectrum based on the level of manipulation required to create the disinformation, with satire at one end, through clickbait content, misleading content, genuine content reframed with a false context, imposter content when an organisation's logo or influential name is linked to false information, to manipulated or edited content, and finally fabricated content that is 100 percent false.

**HIGH HARM**

**FABRICATED CONTENT**
New content that is 100% false, designed to decieve and do harm.

**MANIPULATED CONTENT**
When genuine information or imagery is manipulated to decieve.

**IMPOSTER CONTENT**
When genuine sources are impersonated.

**FALSE CONTEXT**
When genuine content is shared with false contextual information.

**MISLEADING CONTENT**
Misleading use of information to frame an issue or individual.

**FALSE CONNECTION**
When headlines, visuals or captions don't support the content.

**LOW HARM**

**SATIRE OR PARODY**
No intention to cause harm but has potential to fool.

---

36  Wardle, C. (2017, February 16). Fake news. It's complicated. *First Draft*. https://firstdraftnews.org/articles/fake-news-complicated/

## Satire

Making fun of figures of authority or reflecting on events with humour is a reflection on life and what it is to be human. The problem is that if satirical content gets shared outside of its original context or if people don't realise it is satire or parody, they may mistakenly believe it to be true. This becomes a form of misinformation. Additionally, satire has also been used as a tool for journalists or the public to challenge some ideas or powerful authority. In Southeast Asia, the risks and possible ramifications of criticising governments have been outlined in the above 'law laboratory' section.

Increasingly, what is labelled as 'satire' is often used as an excuse to spread hateful, polarising and divisive content. Satire and parody can be used as a tactic to strategically bypass fact-checkers and to distribute rumours and conspiracies. Agents of disinformation can dismiss any criticism by stating that it was never meant to be taken seriously.

In Indonesia, an image circulated that has been described in Indonesian-language posts on social media as showing "halal beer" even though alcohol is banned in Islam. But the image was originally published in a German satirical website. Here, the language barrier adds to the challenge for social media users to trace back to the source.

In Malaysia, a parody Twitter account pretending to be state news agency Bernama, or Berita Nasional Malaysia (Malaysian National News), had managed to build a followership of more than 30,000 users within a year. It has since been suspended by the platform in December 2020 for violating its rules.

Dr Masato Kajimoto, founder of Hong Kong-based fact checking project Annie Lab, observed that in Asia, compared to genuine parody or when jokes are exploited by agents or disinformation, sometimes it is the case when satire in another language is being mistranslated or misinterpreted. For instance, Annie Lab found a Chinese-language Weibo post taking a parody video produced by US show host Jimmy Kimmel's team seriously. The parody video was made using a Fox News clip, which was edited to include the voice of then US President Donald Trump and frames of furniture being thrown out of the window of the White House.

This lesson can also be applied to Southeast Asia given the linguistic diversity as noted above. For instance in Vietnam, some mainstream media outlets have translated satire as serious news. In one example, a local news site translated about the release of the three US citizens by North Korea in 2018.  But in doing so, the news site quoted satirical site The Onion that the three US citizens were taken to a correctional prison in the US state of South Carolina to ease the shock as they returned to normal life after a long period in North Korea (meant by The Onion as satire).

## False connection

False connection is when headlines, visuals or captions don't support the actual content supplied. This can include "clickbait" headlines when the content of a story or video doesn't match up to the headline. It's often not as exciting as the headline or the title made it out to be and can damage an audience's relationship with media outlets or news providers.

A meme that shows a photo of New Zealand Prime Minister Jacinda Ardern next to a quote in support of current president of the Philippines BBM, or Ferdinand "Bongbong" Marcos Jr, has been shared by pro-Marcos Jr accounts. Ardern's photo has been falsely connected to the quote of unknown origin; there's no record of her making the statement publicly, while keyword searches on Marcos also returned no results on the websites of the New Zealand government or Parliament.
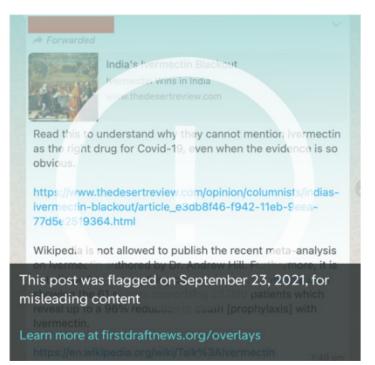
## Misleading content

Misleading content can include the framing of statistics, claims or quotes in a way that is problematic. Often the most effective mis and disinformation has a kernel of truth, which people relate to, but it has been reworked, or re-framed.

A [photo](#) of Burmese mixed martial arts fighter Aung La Nsang holding a gun was shared on Facebook in January 2022 alongside a claim that he joined the Spring Revolution. The Revolution is how local people refer to the anti-coup protests that broke out in early 2021. The photo was actually posted by the fighter in [July 2019 on Facebook](#), where he said he was visiting Las Vegas.

An [article](#) circulating in WhatsApp groups in Singapore in September 2021 argued that Ivermectin is an effective Covid-19 treatment, but the use of it had been rejected in parts of India. The opinion article attributed the low Covid death rate in Uttar Pradesh at the time to the antiparasitic drug that was included in the medicine kits distributed to millions of people. It failed to take into account other reasons that could have contributed to a low death rate, such as effective restriction measures.

Meanwhile, AFP FactCheck [reported](#) in May 2021 that an advisory circulating on Facebook purportedly shows an announcement by a drug store in the Philippines that antiparasitic drug Ivermectin had been approved by the country's Food and Drug Administration. Images of the notice were shared alongside a claim that the medicine was a substitute for Covid-19 vaccination. This is misleading as the advisory actually states that the drug had been approved as an antinematodal agent to treat parasitic roundworms and has nothing to do with Covid.



## False context

False context is genuine content shared with a false context. In breaking news events, it is common to see images and videos being shared as if they are from the current time, but are in some cases actually many years old and may even be from a different location.

A picture went viral on social media in Brunei in January 2021 and caused panic as it claimed to show cracks at a newly-built flyover in Telanai. The Jabatan Kerja Raya (JKR), or Public Works Department, released a [statement on social media](#) explaining it was taken in a neighbouring country. Malaysian newspaper The Borneo Post featured the same photo in a [January 18, 2021](#), article about a bridge in the state of Sabah.

A screenshot of the January 19, 2021, statement on JKR's Instagram can be seen below:

In Myanmar, photos that purport to show a group of captured people have been liked hundreds of times after being shared in a Facebook post in July 2022. The images are described as showing members of the People's Defence Force being captured in the Taung Phyu village. However they surfaced online at least a year prior, as featured in media reports published in July 2021 by the news agency Myanmar Now and by BBC Burmese.

In an example from Thailand, an image of a submarine has been viewed tens of thousands of times on Facebook alongside a Thai-language claim that the numbers inscribed on the boat are the winning lottery numbers in Thailand, meaning the draw is a "scam". However, AFP FactCheck found the photo shows a submarine that belongs to the Singapore navy, and the numbers were added to it later.

## Imposter content

People rely on heuristics, or the mental shortcuts that help us judge the credibility of information and decide if something is from a source they trust or know. If a logo is familiar or well known, for example the trademark of a brand, a media outlet or government organisation, it can give instant credibility — so this logo can be used to fool people into believing information is from a trusted source.

For example, in early 2020, a false claim that an alkaline diet can help prevent Covid spread widely in Southeast Asia on various platforms and messaging apps, such as WhatsApp in Indonesia and Viber in the Philippines. On Facebook, a purported image of a flyer was shared in over 500 groups in Indonesia alone with nearly 40 million followers combined between April and June 2020, according to University of the Philippines Associate Professor Yvonne Chua. However the flyer credits a fictional "Journal of anti-virology and antiviral research".

According to AFP FactCheck, a Facebook page donning a Philippine Department of Social Welfare and Development logo was shared on Facebook ahead of the 2022 Philippine election, calling on voters to get

in touch in exchange for a handout of 10,000 pesos (around AUD263 or USD190). However the department said it was not associated with neither the post nor the Facebook page.
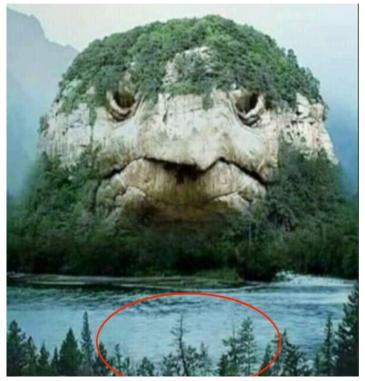
## Manipulated content

Manipulated content is when genuine information or images are manipulated to deceive or mislead. Online editing software has made it very easy for example to splice two different photos together and create a new one. Often a reverse image search will help people to figure this out.

For example, supporters for Philippine President Ferdinand Marcos Jr. targeted independent presidential candidate Leni Robredo with false and misleading information in the 2022 presidential election (N.B. Robredo was with the Liberal Party previously, and still chairs the party but did not run as the standard bearer). Manipulated content included a photo that was described as showing a large turnout at a Leni Robredo rally in the Philippine city of Pasay on April 23, 2022. However, the photo has been doctored. While at least tens of thousands of people turned up to support the then liberal candidate and vice president, the original photo was taken the day before the rally and did not show the purported supporters.

In another example, a photo shared in a May 2020 Facebook post includes a doctored image of Burmese actor and film director Aung Min Naing, or Ko Pauk. In the image, Ko Pauk is seen holding a placard in support of Aung San Suu Kyi, whose National League for Democracy won the election in November that year. However the original image was posted on Ko Pauk's verified Facebook page in March 2020 with a message calling on people to stay indoors in the midst of the pandemic.

Another example purports to show a photo of a turtle head-shaped mountain in Cambodia has been doctored. The original image, which shows the Pilot Mountain in North Carolina (similar uploaded here on Wikipedia) has been edited to include parts of a photo of the Wild Goose Island at Montana's Glacier National

Park. The following is a screenshot comparison of the doctored photo (top) and images that show the exact location (middle) and where the lower part of the photo was taken (bottom):



## Fabricated content

Fabricated content is new content that is 100 percent false, created to deceive or mislead. This could be anything from false information on a meme, to the other extreme of deep fakes (discussed in further detail below).

Fabricated text, audio and visuals are common and don't necessarily require advanced skills or technology to achieve. For example, in early 2022, a claim circulating on Chinese-language social media platforms cited Singaporean newspaper Lianhe Zaobao as saying that the World Health Organisation had declared the pandemic would be over within 2022, and that nucleic acid tests would no longer be proof of Covid infection. However, Lianhe Zaobao said in an April 2022 statement that they had never published the claim.

In Thailand, an image that purports to show a circular is featured in Thai-language social media posts that say the document details how an Islamic textbook in Thailand offers advice not in accordance with traditional Islamic teachings. The textbook is accused of telling its readers not to pay homage to their parents, teachers or monks, among other things. Thailand's Anti-Fake News Center reported in June 2022 that the Ministry of Education had confirmed the content has been fabricated and is not found in said textbook.

An example in Vietnam showed an orchestrated effort to deceive the public for donation. In August 2021, when Ho Chi Minh City was at the height of the pandemic, a Facebook account posted a story of "Dr. Khoa," a medical doctor who removed respirators from his coronavirus-stricken parents to save a mother who gave birth to twins. According to the story, Khoa also delivered the twins. It was spread widely on Vietnamese social media. Ho Chi Minh City's Law newspaper explained the case and the police's findings that the post was created by a group of people who attempted to fraudulently call for donations, and the "Dr Khoa" avatar used was stolen from a Singaporean doctor.

## FURTHER READING

→ "Don't Blame Bat Soup for the Coronavirus", Foreign Policy, January 29, 2020, by James Palmer.

→ "The unproven lab leak theory, Wuhan lab and virus origin: Reporting best practices" First Draft News, June 11, 2021, by Stevie Zhang.

→ "Disinformation, stigma and Chinese diaspora: policy guidance for Australia", First Draft News, August 21, 2021, by Esther Chan and Stevie Zhang.

## ACTIVITY

→ Watch a video on the 7 most common types of information disorder here: http://bit.ly/sms-apac-1

→ Watch a video on the need for 'emotional skepticism': http://bit.ly/sms-apac-6

→ Read more about skepticism in our Psychology of Misinformation series: http://bit.ly/sms-apac-7

# 2.3 Tools and techniques

The methods and tools used by agents of disinformation have become more sophisticated[37]. As Joan Donovan and Brian Friedberg reported in *Data&Society*, these include:

» Sockpuppet Accounts: the anonymous figures — bot, human, or hybrid — pretending to be something they are not.

» Imposter Content: using trusted logos, branding or names as a shortcut for credibility.

» Source Hacking: manipulating the news media and influential figures through lies and deception.

» Keyword Squatting: associating a word with a worldview.

» Attentional Denial of Service: information bombardment to overwhelm and confuse. Additionally, see the 'further reading' below in this section about Spyware use in Southeast Asia.

Another way of thinking about information disorder comes from the work of Camille François, former chief innovation officer at Graphika, a company that tracks disinformation. François has written about the "ABCs" of disinformation — which covers actors, behaviour and content:

## A: Manipulative actors

Manipulative actors "engage knowingly and with clear intent in viral deception campaigns."[38]The actors' intent and their campaigns are "covert, designed to obfuscate the identity and intent of the actor orchestrating them".

## B: Deceptive behavior

This is focused on the techniques used by deceptive actors. The goal of these techniques is to give the impression of a greater impact as if there were larger numbers of actors. These techniques range from "automated tools (e.g., bot armies used to amplify the reach and effect of a message) to manual trickery (e.g., paid engagement, troll farms)[39].

## C: Harmful content

Francois noted "entire categories of content can be deemed "harmful" because they belong to the realm of viral deception, eg, health misinformation"[40]. Additional ways Francois noted the intersection of harmful content and disinformation campaigns can manifest include:

» The content of a campaign itself can be manipulated to deceive users and therefore belong to the realm of 'disinformation', and,

» 'Harmful content' can be promoted by deceptive actors or by campaigns leveraging distortive behaviours.[41]

Similar to the definitions provided by First Draft above, the "ABC" approach notes that the entities that engage in disinformation have a diverse set of goals. Some are financially motivated, engaging in disinformation activities for the purpose of turning a profit. Others are politically motivated, spreading disinformation to foster specific viewpoints among a population, to exert influence over political processes, or for the sole purpose of polarising and fracturing societies. There are also those who share disinformation for their own entertainment, which often involves bullying, and they are commonly referred to as "trolls"[42].

---

37 First Draft Training resources. https://firstdraftnews.org/en/education/curriculum-resources/

38 François, C. (2019, September 20). *Actors, Behaviors, Content A Disinformation ABC*. Transatlantic Working Group. https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf

39 *Ibid*, p 4

40 *Ibid*.

41 *Ibid*.

42 *Ibid*.

## ACTIVITY

→ What do you think are the ABCs behind the vaccine misinformation examples in this article?: http://bit.ly/sms-apac-12

→ 'Think like a troll' game: An important step in learning how to operate within our digital environment, is to understand how online trolls or agents of disinformation think. If we know how they work, how they create content with the purpose to mislead, how they play on people's emotions to ensure false information spreads then we can start to anticipate and "inoculate" our audiences against this type of information.

If you need any ideas or some practice these are two excellent games you can play to put yourself in the shoes of agents of disinformation: "Get Bad News" (developed by Cambridge University Researchers) and "Troll Factory" (developed by Yle, Finish Broadcasting Company). But please note ahead of this activity, it can be quite confronting to see how this type of thinking operates.

## FURTHER READING

→ Journalist Maria Ressa shares her experience of being targeted by gender-based disinformation campaigns: http://bit.ly/sms-apac-14

→ Here is a fascinating investigation by Citizen Lab of spyware against Thai activists: https://citizenlab.ca/2022/07/geckospy-pegasus-spyware-used-against-thailands-pro-democracy-movement/

→ Here's a report from First Draft on the main Covid-19 vaccine narratives and misinformation that circulated on social media in 2020. As we have observed, these themes continued into 2021 http://bit.ly/sms-apac-11

# 2.3.1 Deep fakes in Southeast Asia

Technology companies, governments and journalists are challenged and concerned about how to react to Artificial Intelligence technology that can make real people say and do things they never actually did. Known as deep fakes, genuine pieces of media are manipulated with the intent of deceiving viewers and can undermine trust in news media and governments. While much of the manipulation is done through sophisticated computer software, there are also low quality manipulations, e.g. which slow down video of a politician to make them appear intoxicated, that are known as 'shallowfakes' and "cheap fakes".

Director of WITNESS Sam Gregory wrote in Wired:

> "Statements like 'It's a deepfake' or 'It's been manipulated' have often been used to disparage a leaked video of a compromising situation or to attack one of the few sources of civilian power in authoritarian regimes: the credibility of smartphone footage of state violence. This builds on histories of state-sponsored deception. In Myanmar, the army and authorities have repeatedly both shared fake images themselves and challenged the veracity and integrity of real evidence of human rights violations."

Similarly, journalists and audiences need to be aware of the risks of "the liar's dividend", where it can become easier for those in a position of power to claim something they don't want seen in public is a "deep fake".

Journalists in Southeast Asia should consider ahead of time what scenarios and disinformation actors are most likely to use deep fakes (which are still more technologically sophisticated and expensive to produce) versus other cheaper/easier shallow fakes (technology such as mis-captioning a photo, etc). They should also consider when and why would they expect this — for example, during political campaigns. Deep fakes of public figures will be based on prior footage, so both being prepared with a 'library' of existing

footage or where to find it should be combined with familiarity in using detector tools (reverse video/image search) can help to show if something is different if a deep fake emerges.

Similarly, journalists and audiences need to be aware of the risks of "the liar's dividend", where it can become easier for those in a position of power to claim something they don't want seen in public is a "deep fake".

Journalists in Southeast Asia should consider ahead of time what scenarios and disinformation actors are most likely to use deep fakes (which are still more technologically sophisticated and expensive to produce) versus other cheaper/easier shallow fakes (technology such as mis-captioning a photo, etc). They should also consider when and why would they expect this — for example, during political campaigns. Deep fakes of public figures will be based on prior footage, so both being prepared with a 'library' of existing footage or where to find it should be combined with familiarity in using detector tools (reverse video/image search) can help to show if something is different if a deep fake emerges.

Deep fakes and other forms of synthetic media certainly have the potential to cause harm, but an inclination to find signs of deception can lead to premature conclusions. Following the 2021 coup in Myanmar, a video of former chief minister Phyo Min Thein was viewed over a million times and many online users believed it was deep fake, pointing out, among other things, that his lips appear to be out of sync with his words. However, as Sam Gregory told the news website The Irrawaddy, "There are some confusing signs in the video — for example, digital artefacts around his mouth as it moves — but experts indicate that these are as likely to be from compression as from digital fakery."

Having a plan to prepare and respond to these threats may require multi-stakeholder coordination between actors like technology companies, media forensics experts, and disinformation experts. WITNESS addressed this in a workshop for Southeast Asia, with participants who attended from Cambodia, Indonesia, Myanmar, Singapore, Thailand, and Malaysia. Detailed reports can be found in the link below.

# FURTHER READING

→ "Deepfakes prepare now perspectives from South and Southeast Asia. WITNESS Media Lab, (Report on the 1st Southeast Asia expert meeting on deepfakes preparedness and solutions).

→ "Authoritarian Regimes Could Exploit Cries of 'Deepfake'", Wired February 14, 2021, by Sam Gregory.

→ "Deepfakes and cheapfakes" Data & Society, September 18, 2019, by Britt Paris and Joan Donovan.

→ Here's a video on the real problem with deepfakes: http://bit.ly/sms-apac35

# 3. Monitoring and verification

**A systematic approach to monitoring social media content gives journalists the best chance at discovering notable content and data. This section helps journalists and fact checkers to debunk mis- and disinformation that people are actually seeing, understand where they might be receiving it from, but also help to get ahead of it and slow it down.**

## 3.1 Discovery

Boolean queries are strings of words that allow you to cut through the usual social media chatter and use a multifaceted, specific search to find more precise snippets of information. This is possible with 'operators', which allow you to combine multiple keywords (note: using keywords from the target population's language/ script opens up much more content). There are three operators for basic searches: AND, OR, and NOT.

AND allows you to narrow your search to only retrieve results that combine two or more terms. Taking the earlier example from the Philippines, if you were looking for narratives shared by the supporters of Leni Robredo, who ran as an independent candidate in the 2022 Philippine presidential election, you might want to mix pro-Robredo hashtags with a specific issue. For example you could search for (#MarcosDuwag OR #LetLeniLead OR #SolidSnort) AND (Covid OR pandemic) .

OR allows you to broaden your search to retrieve results connecting two or more similar terms. This can be good for misspellings and typos.

In the case of the Philippine election, you could search for #LeniDuwag OR #LetLeniLeave OR #SolidNorth. This will retrieve all results containing either of the hashtags supporting Ferdinand Marcos Jr.

NOT allows you to exclude terms you don't want to appear in your search results. For example, you might want to exclude mentions of Ferdinand Marcos Sr, former ruler and father of the current president,  and so you could write "Ferdinand Marcos Jr" NOT ("Marcos Sr" OR Macoy OR Ferdie).

As a next step and as seen in some of the examples above, grouping sections together is most useful for long and complicated searches. At this point it is handy to write these out first in a separate document such as notes, and copy and paste this over into the search bar when ready.  It helps to give a better visual view of the operators.

## Key points

» Operators (AND, OR, NOT) must be written in capitals, or they won't work.

» If you're searching for phrases (terms made up of multiple words) then you have to put them in quotation marks (e.g. "Notre Dame").

» You won't be able to find information that has been made private by a user.

» OR statements expand your results, AND statements limit them.

» Group similar words in ( ).

» Phrases need quotations.

» Capitals don't matter (apart from the operators in all-caps as noted above).

» Think about language people actually use, including misspellings, slang, swear words, nicknames.

## ACTIVITY

→ As noted above, a systematic approach to monitoring social media content gives journalists the best chance at discovering notable content and data. Keep a spreadsheet of the keywords that you are using, and update it daily as you track changes in keywords and phrases such as new uses of slang or spelling/mis spellings. If you are in a collaborative environment it is good to have a central spreadsheet that colleagues can see at the start of a shift and help to update.

→ A second spreadsheet should also order the different social media accounts and topics you are following.  This list is useful if you are entering in information for the first time into tools such as CrowdTangle (login version). It may get quite long but is a good way to keep track of everything.

→ From here a spreadsheet can be used to combine key words and narratives of particular issues or projects you are monitoring.

## 3.2 Google Trends

Google Trends uses real-time search data to show you what people are actively interested in at any given moment. It allows you to see the topics people are searching for, practically in real time — you can access the "trending searches" of the day, or you can look at data surrounding specific topics or terms (non-realtime data from 2004 until up to 36 hours before your searches is also available). Journalists can use this information to explore potential story ideas, or use Google Trends data in stories to illustrate the general level of interest in a person, issue or event in a specific period of time and country or territory. The data is [anonymised, categorised, and aggregated](#) — meaning search queries made for the same topic are grouped together.

You can filter the data through various aspects, such as the topic, time of searches, location, and language. Values shown in the sections "interest over time" and "interest by region" are in relation to the highest search volume in the specified time and region. Google Trends will also show you related topics, which are aggregated and categorised, and related queries, which are exactly search queries inputted into Google. "Breakout" denotes that the topic or query has experienced a sudden increase in interest. [Booleans also work on Google Trends](#) and can help refine your searches.

For example, we can use Google Trends to find search data related to the 2022 Philippine presidential election, worldwide over the past 90 days. The data shows that interest in English-language searches about the election was relatively low, but spiked on May 9, 2022, when the election occurred. Through Google Trends, we can see that there was a high level of interest worldwide in [Robin Padilla](#), a Filipino film actor who received the highest number of votes in the Senate election. Whether you are a journalist based in or outside of the Philippines, you can use this data to form a story — for example, about Padilla's image overseas, or what voters think of him.

How do booleans work in Google Trends? Since both the Philippine presidential election and Australia's

federal election were held in May 2022, the boolean query "election -AU" conducted in the Philippines can help eliminate results related to the election in Australia.
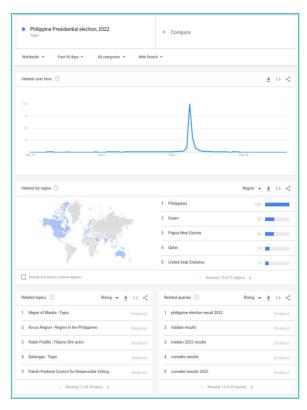


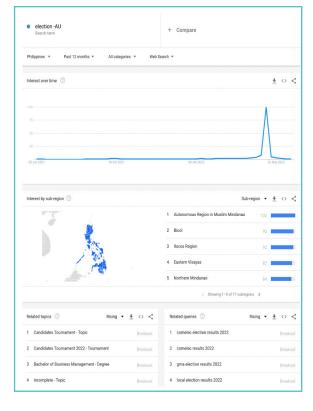Figure 6: Google Trends search for 2022 Philippine Presidential election conducted on June 14, 2022



Figure 7: Google trends search for elections not related to Australia conducted on June 16, 2022

# 3.3 Data deficits

Data deficits are topics where the demand for credible information outstrips supply. For example, during the initial rollout of the Covid-19 vaccines it was difficult to find information about how any of the vaccines affected menstrual cycles — there simply wasn't very much research on it. The vacuum created by the lack of credible information allowed misinformation to flourish.

Data deficits are easy to predict if you know what to look for. Here are six characteristics of topics that might produce them:

» Novel, limited expertise

» Technically complex

» Aligns with pre-existing narratives

» Political dimension

» Emotionally charged

» Legitimate questioning

To find possible data deficits begin with a search on Google or YouTube on a particular topic and take note of the results. Fill gaps where quality information is lacking.

Take note of questions people are asking in social media conversations.

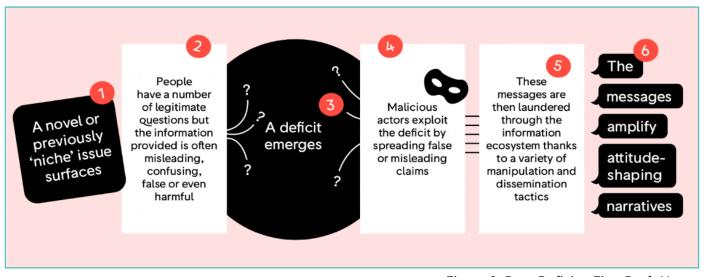Tip: You can include "?" as a "keyword" in your Boolean queries to find posts that include questions.



Figure 8: Data Deficits, First Draft News

# 3.4 Verification

There is no 'one size fits all' when it comes to verification, rather it is a matter of collecting and corroborating numerous clues. The good news is, as First Draft reported, "whether you are looking at an eyewitness video, a manipulated photo, a sock puppet account or a meme, the basic checks you have to do are the same":

1. Provenance: Are you looking at the original account, article or piece of content?

2. Source: Who created the account or article, or captured the original piece of content?

3. Date: When was it created?

4. Location: Where was the account established, website created or piece of content captured?

5. Motivation: Why was the account established, website created or the piece of content captured?

## VERIFICATION: THE FIVE PILLARS OF VISUAL VERIFICATION

**Provenance**
Where did the content originate? Is this the first version?

**Source**
Who created the original content? Remember that the person who posted the content isn't always the original source.

**Date**
Remember that the upload time can be different to the time something was captured.

**Location**
If the content is geotagged, does the location make sense? Have you independently verified the location?

**Motivation**
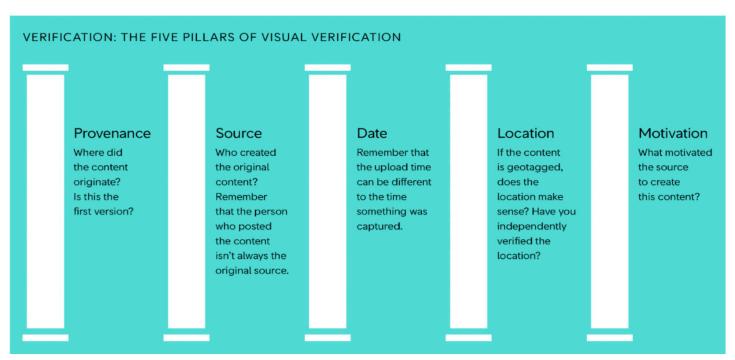What motivated the source to create this content?

Figure 9: The Five Pillars of Visual Verification, First Draft News

# RESOURCES

→ Here's a video on how to quickly verify social profiles on your phone: http://bit.ly/sms-apac34

→ There are a number of online tools that can be used for verification such as different reverse image search options.  It's important not to be overwhelmed by the options, but to have a general idea of what is available and to practice one or two at first.  Here's an excellent resource by OSINT Essentials: https://www.osintessentials.com/. Gradually try out different tools depending on where you are in the five pillars of verification. Here's few to get you started with reverse image search:

RevEye extension: best reverse image search tool
InVID: video verification tool

→ Apply for a CrowdTangle login version.  If you are unable to gain access, CrowdTangle has curated public dashboards for countries including in Asia. Public Live Displays are organized by region and country and show content from local media, regional World Health Organization pages, government agencies, and local politicians as well as social media discussion from Facebook, Instagram, and Reddit. Here's a handy guide: https://www.youtube.com/watch?v=BZs9bO8crjE

# ACTIVITY

## THE OBSERVATION CHALLENGE

→ Put your geolocation skills to the test with this challenge https://firstdraftnews.org/articulate/obsc/story_html5.html

# 4. Fighting mis- and disinformation

Ecker's[44] research into misinformation on climate change exposed argumentation techniques that used mis- and disinformation over scientific climate consensus to falsely encourage climate denialism.

In research findings published in 2022, Carol Soon, Nandhini Bala Krishnan and Shawn Goh detailed the assessment of a "misinformation-immunity gap" between different groups of interviewees in Singapore. They found that those with stronger immunity against mis- and disinformation are generally more capable to differentiate between official, verified sources from information shared by people they know, and they tend to be more accustomed to the practice of verification. The researchers concluded that in order to narrow the "immunity gap", media literacy, awareness of fact-checking publications and resources, as well as knowledge of verification are key.

## 4.1 Inoculation Theory

Similar to the age-old advice from a doctor that 'prevention is preferable to cure', so too a proactive approach to preventing misinformation is recommended as much as possible. As the Australian scholar John Cook explained, psychological research known as inoculation theory[43] "borrows from the logic of vaccines: A little bit of something bad helps you resist a full-blown case." As First Draft reported in 2021: "Much like vaccines train your immune response against a virus, knowing more about misinformation can help you dismiss it when you see it."

Inoculation theory teaches people the generic tactics and techniques of mis and disinformation. For example, John Cook, Stephan Lewandowsky and Ullrich K H

## 4.2 Prebunking

Prebunking is a form of inoculation. It is important for journalists to be aware of possible problematic narratives and get ahead of the curve with quality reports ready to fill any voids (see data deficits, above). This is because audiences may be urgently searching for information on a new or niche issue, or in breaking news.

### Prebunking tips

» Figure out what information people need. Don't assume that your questions are the same as your audience's.

» Choose your example carefully. If you can explain a tactic that is being used to manipulate, it will help build more general resilience to misinformation.

---

43  McGuire, W.J. & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *The Journal of Abnormal and Social Psychology* 62(2) 327-337. https://doi.org/10.1037/h0042026

44  Cook, J., Lewandowsky, S. & Ecker, U.K.H. (2017, May 5). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS One* 12(5). https://doi.org/10.1371/journal.pone.0175799

» Wrap it in truth. Lead with the facts or a clear warning of how information is being manipulated. Remember that it is common for people simply to read a headline and move on. Use your space and words efficiently.

» Warn your audience. Before you repeat the falsehood, warn your audience. This increases your audience's mental resistance to misinformation.

» Explain why something isn't correct. This helps increase belief and arms your audience with counter-arguments they can use to debunk the misinformation the next time they see it.

» Explain how you know what you know (and what we don't know yet). Explanation and transparency help build trust.

» Make them shareable. If you want your prebunk to go far, design it to be passed on. Think about which social media platforms your audience is getting its information from and how you can use those platforms to their  fullest potential.

## ACTIVITY
## TABLE YOUR PREBUNKS

→ As with the systematic approach to booleans and spreadsheets it is helpful to create a table to prebunk issues.

→ For example, the First Draft APAC team trained a group of journalists in Myanmar ahead of the November 2020 general election. In the seminar, we discussed four main categories of mis- and disinformation our research found circulating in the country ahead of the election, including that about the election, targeting then leader Aung San Suu Kyi and the Rohingya people, as well as health misinformation related to the Covid-19 pandemic.

→ Apart from sharing research findings with local journalists, the team also highlighted ways for them to protect their personal safety online and offline, avoid amplifying falsehoods and some reporting considerations that can help stop the spread of misinformation.

→ Prebunking can be made more effective with a collaborative approach involving journalists, researchers, academics and more. Some of the attendants found our presentation helpful and offered to translate some of First Draft's existing materials from English to Burmese for local newsrooms. We also stayed in touch with our trainees via messaging apps through the election and beyond, keeping each other informed about the latest misinformation in the country.

# 5. Responsible reporting and responding

**Sometimes rumours, misleading articles or fabricated visuals are confined to niche communities – addressing the content might actually help to spread it further or 'amplify' information that we don't want to give any more public attention.  Before rushing to report or fact check, consider:**

» Who is my audience? Are they likely to have seen a particular piece of mis- or dis-information already? If not, what are the consequences of bringing it to the attention of a wider audience?

» When should we publish stories about mis- and dis-information? How much traffic should a piece of mis- and dis-information have before we address it? Is it spreading across social media networks or into the mainstream? In other words, what is the "tipping point".

» These are not always easy questions to answer, and as we saw in the pandemic, can depend on the risk of real world harms.

The tipping point is when content moves out of a community, crosses platforms and starts moving at speed. This point will vary according to factors such as the size of the relevant publications and platforms, and their audiences.

If misinformation hasn't reached the tipping point, reporting on it might give it extra oxygen and inadvertently help it spread.

After the tipping point, it is important to push out debunks as widely as possible.

THE TIPPING POINT

Figure 10: The Tipping Point, First Draft News

# 5.1 Psychology of debunks

Additionally, research from the fields of psychology and communication can be used to maximize the positive impact of our corrections. Psychological studies have shown that unless a debunk is done appropriately—or when people don't read the explanation, just a low quality headline or first line in a Facebook or WhatsApp post — our brains are more likely to remember the falsehood. For that reason, it is best to avoid repeating the falsehood in the debunk or headline. If you spot a falsehood about vitamins being a cure rather than a vaccine, think about how to craft this without repeating the memorable false claim. If possible, provide an alternative explanation for the evidence that people use to make their claim. For example the debunk headline or social media post should read: False claims are circulating that certain supplements prevent measles. These are being pushed by people trying to make money from selling these supplements. Additional context is always helpful to help your audience understand the background against which the false or misleading claim is made.

# 5.2 Responsible and effective debunking tips

The importance of taking a proactive stance to information disorder was outlined in the prebunking section. Correcting misinformation is always trickier than it seems. Whether you are a journalist, professional fact checker, academic or communications specialist, here are the best practices.

Avoid asking open-ended questions on social media. If you do, have a moderator ready to offer fact checks in the comments.

» Acknowledge legitimate concerns. Even the most outlandish conspiracy theories tend to be rooted in some kind of deeper apprehension.

» Start from a place of empathy. Treat your audience with dignity. Most people are just trying to make important decisions for themselves and loved ones.

» Focus on the facts. Especially with digital content, remember that attention spans are short. Try to communicate reliable information concisely and warn audiences before repeating falsehoods.

» Provide an alternative explanation. Avoid simple retractions ("this is false"). Clear explanations help the truth stick and prepare people to look out for similar misinformation.

» Keep it simple, but not too simple. Detail is important for earning trust, but too much can be overwhelming. Remember that misinformation is usually short, simple and visually compelling.

# RESOURCES

→ Here's an article that shows how people turn to misinformation when they are searching for answers and there is nothing else out there: http://bit.ly/sms-apac38

# ACTIVITY

## CRAFT YOUR PREBUNKS

→ Similarly to the example above, there were many false cure claims in the pandemic. Rather than listing the claims on the left then the corrections on the right - how else could you craft these debunks?

→ With more attention focused during campaigns, elections are the perfect opportunity for agents of disinformation to target, and can also be rife with misinformation and bias. Think ahead about how this might play out in your country. To prepare for the 2022 Australian federal election, First Draft's APAC team published a misinformation playbook that shows how international narratives and tactics often get adopted domestically. How would you adopt this to your country?

→ Try this chatbot — it can help you figure out what to say to a friend who is unsure about the Covid-19 vaccine: http://bit.ly/sms-apac37

# 5.3 We are all participants

Finally, it is time to stop thinking in terms of the 'top-down' broadcast model. First Draft reported in 2020 that we need to stop thinking that audiences are passive online:

> In trying to explain the influence of false and misleading information online, researchers and commentators frequently focus on the recommendation algorithms that emphasize emotionally resonant material over facts. Algorithms may indeed lead some toward conspiracy theories, but the dominant yet deeply flawed assumption that internet users are passive consumers of content needs to be discarded once and for all. We are all participants, often engaging in amateur research to satisfy our curiosity about the world.

In other words we need to stop thinking of disinformation as something carried out only by influential bad actors to passive audiences, as false and misleading information circulates and develops to create a form of collective sense-making online. For example in First Draft's research into the "Great Reset" conspiracy theory, we found the content underpinning the theory had spread widely across social media in many local and interest-based spaces, garnering considerable opportunities to persuade the curious.

# 6. A final word

**A thoughtful communication strategy will consider not just how to counter misinformation by debunking it, but how to proactively build trust with your audience. When it comes to building trust and identifying trusted sources, think about who the influencers are in your community. This is not only government authority figures or online influencers, but could be respected community leaders. Where possible develop connections with these leaders to find out what their communities' needs are, and how best to reach them.**

As noted in the introduction, the aim of this guide was to support journalists, students and communications professionals (or anyone who may be relied upon for quality information) in an era of information disorder. The case studies and review of laws and fact-checking is indicative only of themes, and examples supporting the definitions and issues is by no means exhaustive due to the diverse and ever-changing "disinformation landscape" of Southeast Asia. We wish readers well as they use this guide to develop and build specialist skills into workflows and learn how to slow down the spread of misinformation in their own communities.

## Debunking rumors

- What we think people need to know
- Appealing to source credibility
- Citing experts
- Simple fact checks

## Building trust

- What people are worried about
- Addressing concerns
- Identifying trusted sources
- Developing critical thinking skills

Figure 11: Debunking versus building trust

# 7. Appendix

**List of Figures**

Figure 1: People in Myanmar receive gibberish Chinese messages shortly after coup

Figure 2: The Deceptive Seven, First Draft News

Figure 3: Misleading content circulating in WhatsApp groups in Singapore

Figure 4: The Jabatan Kerja Raya statement on Instagram

Figure 5:  An example of manipulated images spliced together

Figure 6: Google Trends search conducted on June 14, 2022

Figure 7: Google trends search conducted on June 16, 2022

Figure 8: Data Deficits, First Draft News

Figure 9: The Five Pillars of Visual Verification, First Draft News

Figure 10: The Tipping Point, First Draft News

Figure 11: Debunking versus building trust