

# BEYOND THE RADAR:

## Emerging Threats, Emerging Solutions



**DISINFO  
RADAR**

 DEMOCRACY  
REPORTING  
INTERNATIONAL





# BEYOND THE RADAR:

## Emerging Threats, Emerging Solutions



DEMOCRACY  
REPORTING  
INTERNATIONAL

## About Democracy Reporting International

DRI is an independent organisation dedicated to promoting democracy worldwide. We believe that people are active participants in public life, not subjects of their governments. Our work centres on analysis, reporting and capacity-building. For this, we are guided by the democratic and human rights obligations enshrined in international law. Headquartered in Berlin, DRI has offices in Lebanon, Libya, Myanmar, Pakistan, Sri Lanka, Tunisia, and Ukraine.

## About Disinfo Radar

As part of the Disinfo Radar project, DRI will examine three core pillars of disinformation:

- Emerging technological tools used to produce disinformation
- New tactics for propagating manipulated content
- Untold stories harnessing these tools and tactics to frame false narratives

For more information on the project click [here](#)

## Acknowledgements

*This report was written by Beatriz Almeida Saab, Digital Democracy Research Associate, Jan Nicola Beyer, Digital Democracy Research Coordinator, and Lena-Maria Böswald, Digital Democracy Programme Officer. ForSet designed the layout of this publication.*

**Date: December 2022**

This paper is part of the Disinfo Radar project funded by the German Federal Foreign Office. Its contents do not necessarily represent the position of the German Federal Foreign Office.



This publication is available under a Creative Commons Attribution Non-Commercial 4.0 International license.

# Table of Content

<b>1. Executive Summary</b> . . . . .	6
<b>2. Glossary</b> . . . . .	9
<b>3. Introduction: Understanding Disinformation in Local Contexts</b> . . . . .	11
<b>4. Local Contexts: Emerging Threats and Solutions to Disinformation</b> . . . . .	12
Brazil: Disinformation and the Brazilian 2022 Elections. . . . .	12
Taiwan: Strong Democracy, Low Media Trust. . . . .	22
Kenya: From Cambridge Analytica to Home-Grown Misinformation . . . . .	27
Disinformation in Scandinavia: Sweden and Finland . . . . .	35
<b>5. Going Beyond the Threats: Lessons Learned</b> . . . . .	41
<b>6. The Disinfo Radar: A Novel Approach to Early Detection</b> . . . . .	44



# Executive Summary

In 2022, there were several highly significant political events that were subject to systematic disinformation campaigns. Russia's full-scale invasion of Ukraine and elections in countries including Brazil and Kenya were all subject to either domestic or international efforts to spread false or deceptive news. In parallel, significant technological innovations, such as text-to-image creation, emerged, revolutionising the nature of synthetic media creation.

Fundamental political changes affect the global digital ecosystem by allowing open space for disinformation to emerge and thrive. Building on the foundational first report, "On the Radar",<sup>1</sup> which took a bird's eye view of these trends, this report examines case studies from various countries, spanning multiple regions. The aim is not only to map emerging disinformation tools, tactics and narratives, but also to provide insights into emerging initiatives by local stakeholders to counter disinformation.

These are the main findings:

## Emerging Threats: Tactical Diversification, Rather Than Technical Revolutions?

The findings of this report are consistent with those of previous research, in that they highlight the prevalence of tactically innovative disinformation efforts over technologically sophisticated ones. This is reflected in the report's case studies, which feature countries from around the world. Disinformation actors use cheapfakes – altered images and videos that do not require deep learning expertise or sophisticated tools – as their weapons of choice. Although disinformation agents continue to have access to the same technological arsenal, the global disinformation ecosystem is changing. Malicious actors recognise the value of capturing traditional institutions of trust, along with the power of imitation. By laundering information through proxies (see case study on Sweden and Finland) or purchasing election polls (see case study on Kenya), they seek new ways to provide an air of legitimacy to their disinformation schemes.

---

<sup>1</sup> Jan Beyer & Lena-Maria Böswald, "[On the Radar: Mapping the Tools, Tactics and Narratives of Tomorrow's Disinformation Environment](#)", Democracy Reporting International, June 2022.

## Emerging Solutions: Putting Knowledge at the Centre Stage

The case studies, however, show not only adaptations on the side of disinformation actors, but also illustrate great innovations on the part of anti-disinformation actors. A couple of trends are apparent from the research and interviews.

There is an increasing effort to form anti-disinformation networks, by bringing together state and commercial actors, as well as civil society organisations (CSOs) (see for example, case studies on Finland, Kenya, and Brazil). At the same time, there is a growing tendency to combine tech and non-tech solutions, such as combining automated debunking with face-to-face pre-bunking and digital-literacy-promotion strategies. Finally, the case studies show the tightrope states often need to walk, balancing the containment of disinformation against the principle of freedom of speech. While many of the solutions reviewed in this study focus on the dissemination of knowledge, either by correcting specific disinformation (debunking) or preparing societies for disinformation efforts (through pre-bunking or enhancing digital literacy), banning malicious agents from the public sphere remains a highly controversial solution.

## Lessons Learned: Anticipating the Road Ahead

The case studies offer lessons that transcend country-specific contexts, providing a broader outlook on the future of (dis)information. Such insights concern both the changing nature of disinformation threats and the sustainability and scalability of the current approaches to fight them:

### a) Threats:

- In their quest for new business models, tech companies have created new opportunities for disinformation actors. For example, as we see in Brazil, “consumer bait”, such as unrestricted data access on messenger apps, has opened new opportunities for disinformation.
- The role of domestic proxies: Foreign influence operations have adopted the disguise of domestic institutions, whether in the shape of influencers, junk news, or polling companies.
- Vulnerable “outsiders”: As states become more resilient, disinformation actors will continue to search for weak links in societies. Immigrants and ethnic minorities may be particularly affected by this, as the report’s case study on Finland illustrates.

This report also provides insights to address these emerging disinformation threats.

**b) Solutions:**

- Investment in long-term payoffs: Specifically, the case studies on Finland and Taiwan presented here show just how important it is to think about disinformation early on, even in relation to early childhood education. Doing so increases societal resilience against false and deceiving information.
- Inclusion of citizens: The case studies show that the public is receptive to anti-disinformation solutions and craves healthier digital ecosystems; they are willing to adopt new solutions – whether tech or non-tech – and become anti-disinformation agents.
- Diverse alliances: The more diverse the body of stakeholders, the more innovative the solutions will be. In particular, our study of the Code for Africa reveals how large anti-disinformation networks are able to develop a wide range of both technical and non-technical solutions to the problem of deceptive narratives.

## **Disinfo Radar: An Innovative Approach to Early Detection**

In addition to using comparative case studies, the report introduces Disinfo Radar. Disinfo Radar is DRI's own initiative to combat disinformation by employing an innovative early detection method. The website-based registry, developed by DRI, identifies disinformation tools and tactics in their early stages of development. For example, by using automated text analysis tools, Disinfo Radar attempts to identify new disinformation technologies at an early stage. By auto-collecting and auto-analysing electronic preprint repositories (e.g., arXiv), industry papers (e.g., syncedreview.com), and policy publications (e.g., IEEE), Disinfo Radar scans the environment for indications of emerging technologies. Hence, rather than providing information about current disinformation threats, Disinfo Radar helps detect the emerging threats of tomorrow

.

# 2

# Glossary

## **Artificial Intelligence**

The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

## **Astroturfing**

Organised activity on the Internet that is intended to create the false impression of a widespread, spontaneously arising grassroots movement in support of or in opposition to something (such as a political policy) but that is, in reality, initiated and controlled by a concealed group or organisation.

## **Bots**

Social media accounts that are operated entirely by computer programs and are designed to generate posts and/or engage with content on a particular platform. Researchers and technologists take different approaches to identifying bots, using algorithms or simpler rules based on the number of posts per day.

## **Cheapfakes**

Content altered by technologically low-level manipulation of audio-visual material (created with easily accessible software).

## **Coordinated Inauthentic Behaviour**

Groups of pages or people working together to mislead others about who they are or what they are doing in the online environment.

## **Deepfakes**

Content manipulated or created by technologically highly sophisticated manipulation of audio-visual media, using AI-driven technology.

## **Debunking**

The act of uncovering half-truths or false information, and communicating about this.

## **Digital Literacy**

Having the skills to live, learn, and work in a society where communication and access to information are increasingly through digital technologies.

## **Disinformation**

False information that is deliberately created or disseminated with the express purpose to cause harm.

## **End-To-End Encryption**

A secure communication process that prevents third parties from accessing data transferred from one endpoint to another.

## Fact-Checking

The process of determining the truthfulness and accuracy of official, published information, such as politicians' statements and news reports.

## Information Manipulation

The strategies employed by a source or producer of information to deceive the receiver or consumer into interpreting that information in an intentionally false way. The user thinks they are receiving the information in a genuine way but, in fact, they are being deceived by its manipulation.

## Misinformation

Incorrect or misleading information, unintentionally presented as fact.

## Pre-Bunking

The act of uncovering tactics and tropes of misleading information before they are encountered in a given context.

## Recommender System

A subclass of machine learning that generally deals with ranking or rating products, content or users.

## Social Media Amplification

Where recommender systems specifically amplify certain content (based on built-in features or for paid promotion), thereby increasing its exposure.

## Sock Puppet

An online account that uses a false identity designed specifically to deceive. Sock puppets are used on social platforms to inflate another account's follower numbers and to spread or amplify content to a mass audience.

The term is considered by some to be synonymous with the term "bot".

## Synthetic Media

Data and media (audio, text, image, or video) artificially produced, manipulated and modified by automated means, especially through the use of AI algorithms.

# 3

# Introduction: Understanding Disinformation in Local Contexts

Times of political instability and technological innovation offer disinformation actors opportunities to expand their arsenals through the adoption of new technologies. In its first report, “On the Radar”, DRI mapped the emerging disinformation tools, tactics and narratives that are likely to shape and reshape tomorrow’s public sphere.<sup>2</sup> As discussed in this report, declining entry barriers to cutting-edge technology, such as complex machine-learning models, multiple avenues for creating synthetic media, and the availability of ever-increasing data sets, could lead to an avalanche of false and deceptive information.

“On the Radar” provided a bird’s-eye view of disinformation tools, tactics, and narratives; in this report, we will zoom in and observe them in a diverse set of local or national contexts. Additionally, by examining innovative means by which both public and private actors combat disinformation, we seek to understand the array of current solutions.

This report will do the following:

- (a)** dissect the emerging disinformation tools, tactics, and narratives in four selected case studies;
- (b)** discuss innovative approaches anti-disinformation agents have created to tackle these arising threats; and, finally
- (c)** present DRI’s own initiative in the fight against disinformation – Disinfo Radar. This online tool facilitates the early detection of emerging disinformation tools and tactics by leveraging various machine-learning models and natural language processing (NLP) tools.

---

<sup>2</sup> Jan Beyer & Lena-Maria Böswald, [“On the Radar: Mapping the Tools, Tactics and Narratives of Tomorrow’s Disinformation Environment”](#), Democracy Reporting International, June 2022.

# 4

# Local Contexts: Emerging Threats and Solutions to Disinformation

## Brazil: Disinformation and the 2022 Elections

Disinformation online has a particularly large impact on Brazilian elections, as 54 per cent of Brazilians use social media platforms as their main source of news.<sup>3</sup> This is a high percentage when compared to the global average of 46 per cent, and to the country with the lowest rate, Germany, where only 18 per cent of its population perceives social media as their main news source.<sup>4</sup> Furthermore, the most popular social media platform in Brazil is WhatsApp, being present on 99 per cent of all mobile phones. A recent report from the Reuters Institute revealed that 53 per cent of Brazilians trust news disseminated on WhatsApp.<sup>5</sup>

Since 2018, Brazil has developed a complex disinformation ecosystem, characterised by powerful actors, mass messaging, and the production and dissemination of false and falsified content. This disinformation ecosystem has been – and continues to be – responsible for an avalanche of false information that confuses the population and adversely impacts institutions. For example, during the first years of the COVID-19 pandemic, 110 million Brazilians received false information about the pandemic, with 6 out of 10 internet users receiving the news through WhatsApp.<sup>6</sup> Disinformation has many harmful consequences, and one of the most serious is the loss of confidence in democratic institutions.<sup>7</sup> Fifty-two per cent of Brazilians have little or no trust in the Brazilian Supreme Court, for example.<sup>8</sup>

Brazilian elections have particular characteristics, such as, since 1996, the systematic

<sup>3</sup> Samuel Tan, "[Global: To What Extent Do People Use Social Media to Catch up on the News?](#)" YouGov, 239 July 2022.

<sup>4</sup> Ibid.

<sup>5</sup> Camila Mont'Alverne, Sumitra Badrinathan, Amy Ross Arguedas, Benjamin Toff, Richard Fletcher, and Rasmus Kleis Nielsen. "[The Trust Gap: How and Why News on Digital Platforms Is Viewed More Sceptically Versus News in General](#)", Reuters Institute for the Study of Journalism, p. 73.

<sup>6</sup> "[O Brasil está sofrendo uma infodemia de Covid-19](#)", Avaaz, 4 May 2020.

<sup>7</sup> Adauto Soares, Amanda Yumi, Beatriz Barbosa, Danilo Doneda, Deborah Dalbart, Diego Machado, Diego Canabarro, et al, "[Internet, Desinformação e Democracia](#)", Comitê Gestor da Internet no Brasil.

<sup>8</sup> Marlen Couto, "[Partidos, igrejas e STF: veja os índices de confiança dos brasileiros nas instituições](#)", O Globo,

use of electronic voting machines, and electoral management by a court (rather than a dedicated commission or by government ministries, as is the case in many other countries).

The main electoral institution in Brazil is the Superior Electoral Court (TSE), the highest structure within the system of electoral management and justice, thus playing a fundamental role in the administration and defence of democracy.<sup>9</sup> The TSE's tasks include dealing with complaints about candidate registration and investigating electoral wrongdoing, which includes some control of social media platforms and the right to block accounts that share disinformation. This approach has been criticised from some quarters as an illegitimate limitation of free speech.<sup>10</sup>

## Disinformation in Brazil: Tools, Tactics and Narratives



### Electronic Voting at the Centre of the Debate

The main disinformation narrative before and after the 2022 Brazilian elections concerned the procedural integrity of the vote. Similar to such narratives in the United States, actors raised unsubstantiated doubts among the public about the voting process. False information concerning the safety of electronic voting machines mobilised political debate on social media. The incumbent, Jair Bolsonaro, and his supporters were central to this narrative, claiming that the electronic voting machines used across Brazil are susceptible to fraud,<sup>11</sup> without substantiating these claims. This forced the judges of the TSE to take further steps in their efforts to publicly guarantee fair elections. Together with approaches already in place, such as auditing a selected number of electronic voting machines,<sup>12</sup> in 2022 – for the first time – the TSE allowed the military, known in Brazil to be great supporters of Bolsonaro, to perform an integrity test to assure the electronic machines were operating correctly.<sup>13</sup>

---

<sup>9</sup> ["Superior Electoral Court"](#), TSE.

<sup>10</sup> Gessica Brandino, ["TSE Age Contra Fake News No Vácuo Do Ministério Público"](#), Folha de São Paulo, 17 November 2022.

<sup>11</sup> Juliana Gragnani and Jake Horton, ["Brazil Election: Do Voting Machines Lead to Fraud?"](#), BBC News, 30 September 2022.

<sup>12</sup> Rosanne D'Agostinno, ["Teste Mostrou Que Não Houve Divergência Entre Votos Dados e Votos Registrados Pelas Urnas, Diz Moraes"](#), G1, 6 October 2022.

<sup>13</sup> Flavia Maia, ["TSE atende Forças Armadas e faz simulação de teste nas urnas com biometria"](#), JOTA Info (blog), 15 September 2022.



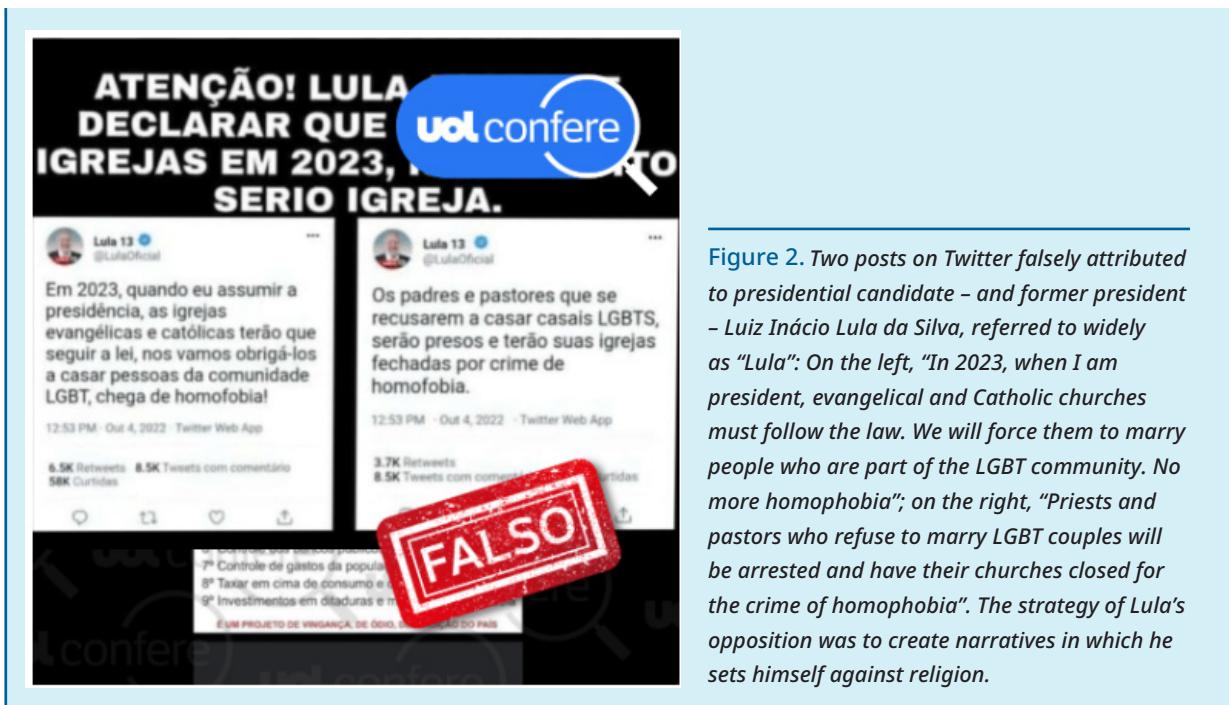
## Chepfakes Dominate the Online Discourse

To spread the electoral fraud narrative, actors in Brazil often use simple manipulation techniques to create false evidence.<sup>14</sup> Such chepfakes are still the dominant means of media manipulation and disinformation campaigns. This technique requires only low technical sophistication and consists of simple content manipulation, such as speeding up, slowing down and cutting or splicing clips. Below are examples of chepfakes from the 2022 election:



**Figure 1.** A post on Instagram with the caption: SCANDAL! Leaks videos and photos of Bolsonaro asking for votes in front of Freemasonry symbols and taking pictures next to the Baphomet, a mystical pagan creature usually associated with Satanism. As evangelical Christians are a significant part of Bolsonaro's support base, one method of attacking him in disinformation campaigns was to relate him to images that are seen as anti-Christian.

<sup>14</sup> Beyer & Böswald. ["On the Radar"](#), op. cit., note 1.



**Figure 2.** Two posts on Twitter falsely attributed to presidential candidate – and former president – Luiz Inácio Lula da Silva, referred to widely as “Lula”: On the left, “In 2023, when I am president, evangelical and Catholic churches must follow the law. We will force them to marry people who are part of the LGBT community. No more homophobia”; on the right, “Priests and pastors who refuse to marry LGBT couples will be arrested and have their churches closed for the crime of homophobia”. The strategy of Lula’s opposition was to create narratives in which he sets himself against religion.

## The Transforming Use of WhatsApp in the Spread of Disinformation

During the 2022 Brazilian elections, disinformation actors continued to use WhatsApp as their main weapon for disseminating false narratives and disinformation.<sup>15</sup> The way disinformation agents used the app has changed, however. During the 2018 elections, fact-checkers observed mass messaging as one of the means to spread deceptive information.<sup>16</sup> After Meta prohibited this tactic, as part of their efforts to protect election integrity,<sup>17</sup> those spreading disinformation quickly updated their tactics. Two tactics stand out here: the YouTube-to-WhatsApp pipeline and the use of WhatsApp’ “Status” function.

Disinformation actors often share snippets of YouTube videos on WhatsApp, allowing users to watch and share the videos without using the data plan on their phone. This is possible because many data provider companies in Brazil offer plans with unlimited access to WhatsApp as an option. WhatsApp disinformation is often sourced from viral videos on YouTube, therefore extending the reach and effect of YouTube’s algorithm

<sup>15</sup> Lais Borges, [“Estudo mostra que uso de fake news cresce no 2º turno; ‘desinformação está mais complexa e sofisticada’, diz pesquisadora”](#), G1, 25 October 2022.

<sup>16</sup> Patricia Campos Mello, [“WhatsApp Admits to Illegal Mass Messaging in Brazil’s 2018”](#), Folha de São Paulo, 9 October 2019.

<sup>17</sup> [“How Meta Is Preparing for Brazil’s 2022 Elections”](#), Meta, 12 August 2022.

on messaging disinformation. During the 2022 elections, YouTube released an updated version of its election guidelines.<sup>18</sup> Over the first two months of the presidential campaign, however, YouTube removed only 4.4 per cent of videos providing mis/disinformation about electoral processes and electronic voting security.<sup>19</sup>

This YouTube-to-WhatsApp pipeline is an effective tactic to spread disinformation; WhatsApp users often share video clips with their contacts. Because open internet access is expensive, and given low levels of media literacy,<sup>20</sup> recipients often do not fact-check the veracity of videos.

Although there is so far no available research on the number of videos shared via WhatsApp during the 2022 election period, a study from 2019 showed that, in most Brazilian WhatsApp groups, one video is uploaded for every 14 text messages.<sup>21</sup> The research also shows that WhatsApp users tend to link to YouTube more than to any other site — 10 times as frequently as they would link to Facebook – thus strengthening the YouTube-to-WhatsApp pipeline.<sup>22</sup> While there are no updated reports on this phenomenon, experts expect a continuation of this trend, with videos being one of the key means of sharing disinformation on WhatsApp.<sup>23</sup>

Another common tactic on WhatsApp uses the “Status” feature, which is visible to all of the user’s contacts. While many people view Status as an overlooked and underutilised tool, it has the capability to reach an audience outside of the user’s typical social media echo chamber. What makes this feature interesting is that, regardless of whether there is any message exchange between contacts, if one shares something on WhatsApp Status, everyone that has that contact saved will have the option to click and visualise the content. Understanding the potential this brings for spreading information to many contacts, people linked to parties and activist groups encouraged voters to use WhatsApp Status to boost the reach of their candidate’s campaigns

This new tactic relies on occupying the spaces that platforms make available and playing with the unsupervised structures, with the specific aim of influencing voters and shaping the information they receive before election day. Despite not yet having reports that show how much of the content shared on WhatsApp status was part of disinformation campaigns, the feature offers great disinformation potential.

---

<sup>18</sup> [“Políticas Contra Desinformação Em Eleições - Ajuda Do YouTube”](#), YouTube.

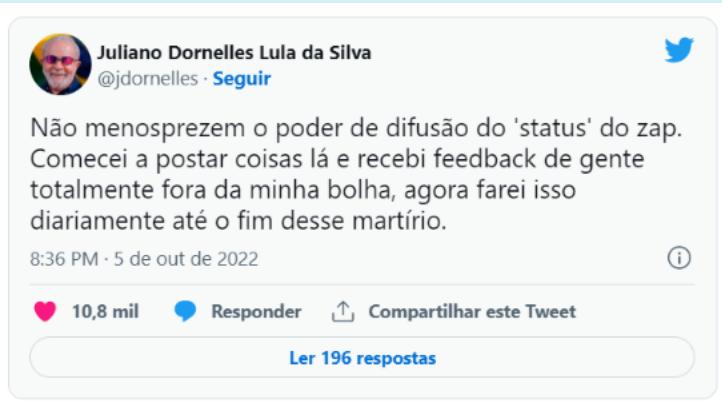
<sup>19</sup> Samuel Lima, [“YouTube remove só 4,4% dos vídeos com desinformação contra urna eletrônica”](#), Timeline: eleições 2022 (blog), 21 June 2022.

<sup>20</sup> [“62% dos brasileiros não sabem reconhecer uma notícia falsa”](#), Veja, 13 February 2022.

<sup>21</sup> Amanda Taub & Max Fisher, [“How YouTube Misinformation Resolved a WhatsApp Mystery in Brazil”](#), The New York Times, 15 August 2019.

<sup>22</sup> Ibid.

<sup>23</sup> Julia Braun, [“Conspiração e apuração paralela: a crescente desinformação no WhatsApp sobre urnas às vésperas da eleição”](#), BBC News Brasil, 1 October 2022.



**Figure 3.** A tweet from a Lula supporter saying "Don't underestimate the dissemination power of WhatsApp status. I started posting stuff there and get feedback from people totally outside my social media circle, now I will do it daily until the end of this election".



## Approaches to Disinformation: The Government Takes Action

One of the initiatives of the TSE to combat the negative impacts of the dissemination of false narratives has been the creation of the “Programme to Combat Disinformation”. In partnership with more than 70 institutions and organisations, the TSE works within the framework of the programme to minimise the impacts of disinformation on democratic processes.<sup>24</sup>

For the 2022 Elections, the TSE intensified its efforts to combat disinformation to fortify the voters' choice as legitimate and to contain interference by deceptive campaigns.

The Programme is organised in three pillars: (1) to inform voters, by disseminating official, reliable, and truthful information; (2) to empower voters, by developing media literacy initiatives, so the public understands the concept of disinformation and the functioning of the electoral process; and (3) to respond, by identifying cases of disinformation and adopting strategies to contain their negative effects.<sup>25</sup>

<sup>24</sup> "O TSE", TSE.

<sup>25</sup> Aline Osorio, Frederico Franco Alvim, Giselly Siqueira, Julia Barcelos, Antonio Vargas, Tainah Rodrigues & Thiago Rondon. ["Programa Permanente de Enfrentamento à Desinformação no Âmbito da Justiça Eleitoral"](#), Tribunal Superior Eleitoral, 2022.

## The Programme to Combat Disinformation (TSE)

Inform	Empower	Respond
A network for the mass dissemination of truthful and official information	Campaigns to educate citizens about disinformation	A permanent coalition for fact-checking
A WhatsApp chatbot to receive electoral enquiries	Mental-health prevention for public servants dealing directly with disinformation	Engagement of digital platforms and their technological resources in confronting structured networks of disinformation and inauthentic behaviour
Access, dissemination and potentialisation of the reach of fact-checking about the electoral process	Campaigns to educate citizens about the electoral process	A channel for denouncing mass messaging, in partnership with WhatsApp
Deepening of electoral transparency	Awareness-raising campaigns on misinformation	Creation of a network for monitoring disinformation practices related to the electoral process
Development of new digital channels to disseminate truthful information	Media literacy actions	Containment of disinformation on Telegram
—	Dialogue with political parties and party federations to make them aware of their responsibility in combating disinformation	Partnership and exchange with the Federal Police and the Electoral Public Prosecutor's Office
—	Support to other public institutions to implement actions to combat disinformation	Creation of a Strategic Committee of Cyber Intelligence
—	—	Review and elaboration of rules that combat the practice of disinformation in the scope of the TSE

The programme included partnerships with fact-checking agencies, social media platforms (Facebook, Instagram, WhatsApp, Google, YouTube, Twitter, and TikTok), telephone companies, research agencies, civil society organisations (CSOs) and media associations. The TSE also created a cyber-intelligence committee, banned accounts that carried out mass messaging in the elections, and created hashtags and a page to debunk false news. From a technical perspective, two of the TSE's main initiatives were particularly innovative:

### ***Virtual Assistant on WhatsApp***

The TSE launched "Tira-Dúvidas Eleitoral no WhatsApp", a chatbot "virtual assistant", originally created in partnership with the messaging app to facilitate voter access to relevant information about the 2020 Municipal Elections that was also used for the 2022 Elections. According to the TSE, this partnership between an electoral authority and WhatsApp is the first of its kind. During the 2020 elections, the chatbot exchanged almost 19 million messages with users.<sup>26</sup>

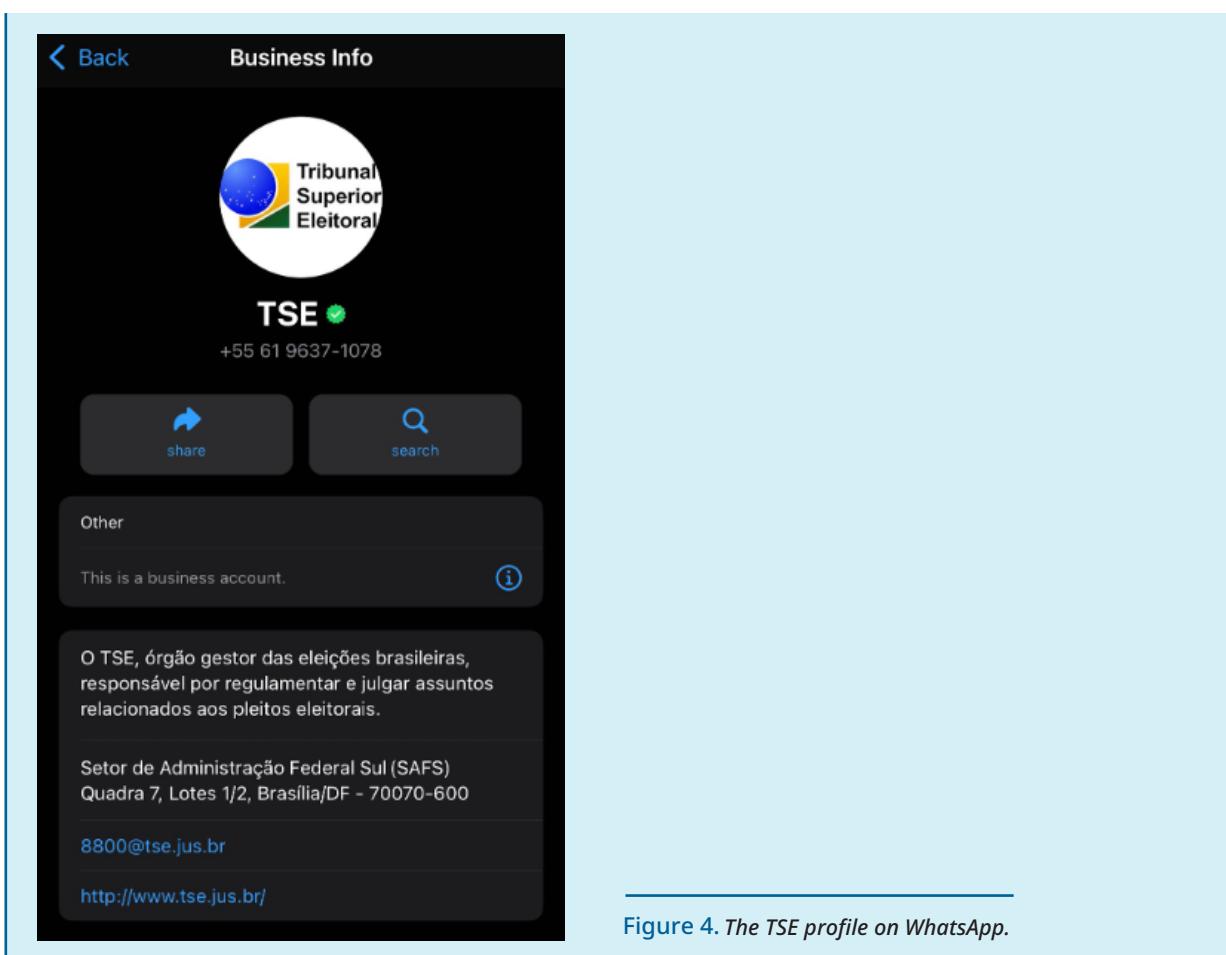


Figure 4. The TSE profile on WhatsApp.

<sup>26</sup> Ibid.

## ***The development of new digital channels to disseminate truthful information and receive reports about electoral crimes***

Over the past two years, the TSE developed two apps to facilitate communication with the public, as well as to digitalise some services. These apps are “E-Título” and “Pardal”.

The e “E-Título” app works as a digital copy of the voter registration card and replaces the paper document. Through the app, it is also possible to update your personal information, receive notifications from the TSE, and report electoral crimes.

“Pardal” is another app available for Brazilian citizens, where they can report irregularities in advertisements and other practices prohibited by the electoral legislation, such as vote-buying and the misuse of public resources for campaigning; the abuse of political power, such as pressuring state workers to vote for a particular candidate; and illegal practices during electoral campaigns, such as disseminating disinformation and mass messaging.<sup>27</sup> This year, during the electoral campaign period (between August and October), the app received around 52 thousand reports from citizens, the highest number since it was launched in 2016.<sup>28</sup>

The TSE developed and operated a digital communication infrastructure (notification centre) with the capacity to send short messages to disseminate quality content via both apps. The notifications provided useful information about the elections, including clarifications about false news. In 2020, the TSE sent over 300 million notifications to around 18 million citizens that have the app installed in their phones.<sup>29</sup> The high number of notifications and users illustrates how citizens adapt to solutions and initiatives proposed.

For example, on election day for the second round of the 2022 elections, complaints were made via the Pardal app about cases where the Federal Police were stopping voters on their way to voting stations.<sup>30</sup> The app allowed for a speedier response from authorities, informing users just one hour later of their rights.

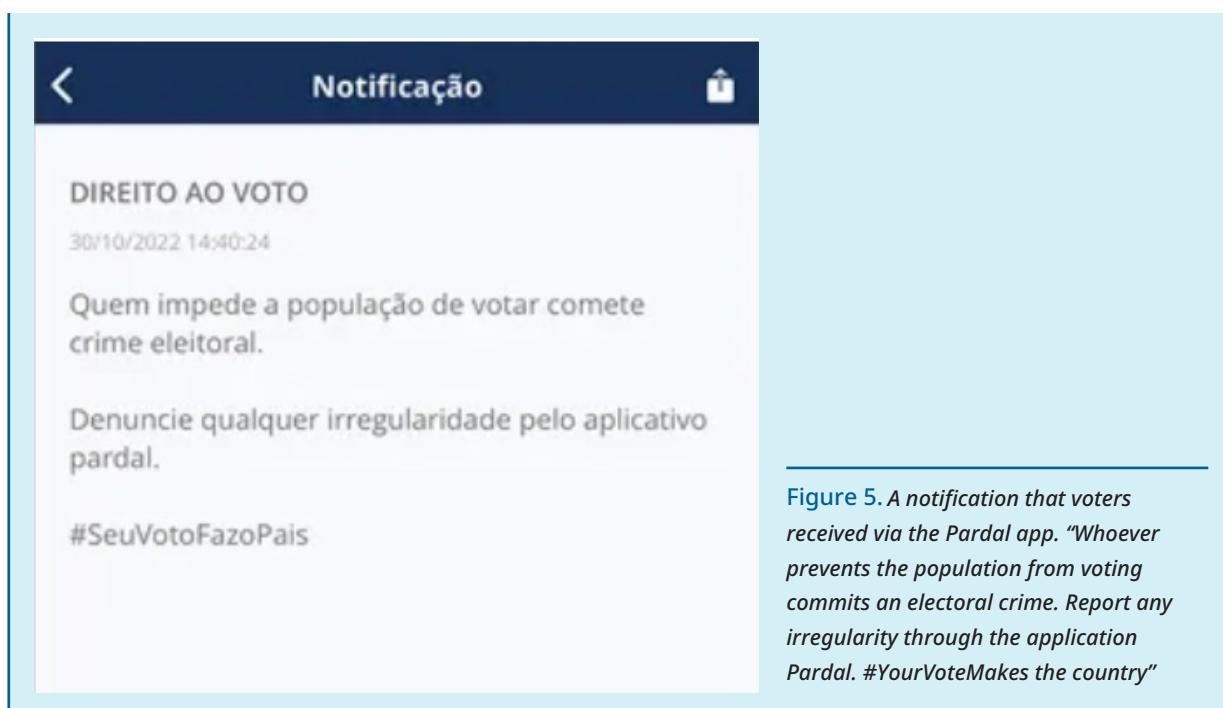
---

<sup>27</sup> [“Eleições 2022: confira o que pode e não pode na propaganda eleitoral”](#), TSE.

<sup>28</sup> [“Eleitor fiscal: aplicativo Pardal bate recorde com mais de 52,9 mil denúncias nas Eleições 2022”](#), TSE.

<sup>29</sup> Aline Osorio, Frederico Franco Alvim, Giselly Siqueira, Julia Barcelos, Antonio Vargas, Tainah Rodrigues & Thiago Rondon. [“Programa Permanente de Enfrentamento à Desinformação no Âmbito da Justiça Eleitoral”](#), Tribunal Superior Eleitoral, 2022.

<sup>30</sup> Isabela Camargo & and Marcio Falcao. [“PRF Descumpre Ordem Do TSE e Para Pelo Menos 610 ônibus de Eleitores Em Blitz; Moraes Intima Diretor-Geral”](#), G1, 30 October 2022.



**Figure 5.** A notification that voters received via the Pardal app. "Whoever prevents the population from voting commits an electoral crime. Report any irregularity through the application Pardal. #YourVoteMakes the country"

Alongside Pardal, the TSE also provided a system of alerts against electoral disinformation available on its website. The channel enabled users to report violations of the terms of use of digital platforms, specifically related to disinformation about electronic voting machines and the work carried out by the TSE. It is too early to judge in detail how impactful these Apps were in 2022. It is clear, however, that the TSE is among the world's most proactive election management bodies in addressing electoral disinformation.

Registrar um alerta

**Tipo de denúncia (obrigatório)\***

Selecionar

- Selecionar
- Desinformação
- Discurso violento ou odioso**
- Disparo em massa (WhatsApp/Telegram)
- Grave perturbação do ambiente democrático
- Indício de comportamento inautêntico
- Vazamento de dados/incidente cibernético
- Violência Política de Gênero

**A mensagem se refere ao processo eleitoral?**

Selecionar

Mensagens que envolvam candidatos e partidos políticos não serão objeto de processamento por meio desse canal

**ENVIAR**

**Figure 6.** The Alert system provided by the TSE. Members of the public can report violations related to: disinformation; violent or hateful speech; mass messaging (WhatsApp/Telegram); serious disruption of the democratic environment; evidence of inauthentic behaviour; data leakage or cyber incidents; and gender-related political violence.

## Taiwan: Strong Democracy, Low Trust in the Media

In March, Taiwan made headlines by topping the Digital Society Project's world rankings of countries receiving the most disinformation online.<sup>31</sup> According to many sources, Taiwan is on the front line of mainland China's disinformation campaigns, with one study suggesting that the Chinese government uses it as a "testing ground" to experiment with propaganda techniques before they are employed elsewhere.<sup>32</sup>

While the Taiwanese enjoy the second-freest democracy in Asia, mainland China has infiltrated a number of print and broadcasting outlets, meaning many Taiwanese have shifted to crowdfunded and open-source foundations for their news.<sup>33</sup>

Although studies suggest that it is difficult to precisely identify the actual impact of mainland China's disinformation campaigns on Taiwanese social media, Beijing's disinformation efforts have generated real-life implications.<sup>34</sup> The most important impact these efforts have had in Taiwan is in lowering overall trust in the government and deepening socio-political divisions.<sup>35</sup> Facing this pressing scenario, Taiwan has developed several approaches to countering disinformation that have been considered innovative, while respecting a pluralistic civil society.<sup>36</sup>

---

31 Maren Sass, "[How Taiwan Is Countering Chinese Disinformation](#)", Deutsche Welle, 25 August 2022.

32 Shih-Shiuan Kao, "[Taiwan's Response to Disinformation: A Model for Coordination to Counter a Complicated Threat](#)" The National Bureau of Asian Research, September 2021.

33 Sass, "[How Taiwan Is Countering Chinese Disinformation](#)", op. cit., note 31.

34 Matthew Becerra, "[The Battle for Reality: Chinese Disinformation in Taiwan](#)", Geopolitical Monitor (blog), 24 August 2022.

35 Ibid.

36 Shih-Shiuan, "[Taiwan's Response to Disinformation](#)", op. cit., note 32.

## Pro-Chinese Disinformation in Taiwan: Tools, Tactics, and Narratives



### Pro-Chinese Narrative: Deligitimisation of Taiwan's Government

According to various reports, mainland China's current disinformation narratives targeting Taiwan aim at dividing and demoralising Taiwanese society. Research shows that "China's disinformation efforts directed at Taiwan generally seek to divide and demoralise Taiwan society, driving up the negative impressions associated with the government of President Tsai [Ing-wen] and creating an image of an incompetent government. It presents an impression of the Tsai administration as being out of touch with the populace and indifferent to the effects of its (purportedly incorrect) policies, driving Taiwan toward disaster, and betraying Taiwan's economic and territorial interests and its true identity".<sup>37</sup> The same study suggests that China's government seeks to convey a dual narrative through social media. The first dimension of this narrative tries to reassure those in Taiwan who believe that mainland China is non-threatening, by conveying positive images of China as an attractive place where Taiwanese businesses can succeed. The second dimension tries to convince target audiences that the re-unification of Taiwan and China is inevitable.<sup>38</sup>



### Creation of Original Content instead of Media Manipulation

According to the report "Detecting Digital Fingerprints: Tracing Chinese Disinformation in Taiwan" from the International Republican Institute (IRI), when it comes to disinformation tools and tactics, some Chinese actors generate original content and disseminate it via online fake accounts. One example was a campaign that ran a month after Tsai's re-election, in 2020. A petition written in broken English appeared on a United States government petition website, WeThePeople, asking the United States to investigate the authenticity of Tsai's Ph.D. Over the following month, a small network of users heavily promoted the petition link on Facebook, Instagram, and Twitter. The posts often used identical messages, and most likely were spread from fake accounts. This disinformation narrative around President Tsai's degree was also perpetuated in the most popular messaging app in Taiwan, LINE.<sup>39</sup>

<sup>37</sup> Scott W. Harold, Nathan Beauchamp-Mustafaga & Jeffrey W. Hornung, "[Chinese Disinformation Efforts on Social Media](#)", RAND Corporation, 2021.

<sup>38</sup> Ibid.

<sup>39</sup> Nick Monaco, Melanie Smith & Amy Studdart, "[Detecting Digital Fingerprints: Tracing Chinese Disinformation in Taiwan](#)", International Republican Institute (IRI), 25 August 2020.



**Figure 7.** Three users post identical messages on a United States government petition website urging Taiwanese users to sign a petition calling on the American government to investigate Taiwanese President Tsai Ing-wen. (Source: [Detecting Digital Fingerprints: Tracing Chinese Disinformation in Taiwan](#)).



## Large-Scale Disinformation with Content Farms

According to these reports, Chinese disinformation actors use tactics such as reposting content originating from within Taiwan, using Taiwanese content farms with links to China to elevate the profile of negative postings about a particular person or issue,<sup>40</sup> and utilising new types of digital platforms, such as news aggregation and discussion platforms, including Reddit.<sup>41</sup>

Reports indicate that China gains access to local social media and pushes messages from the content farms to popular Taiwan social media platforms. To do so, malicious actors create social media profiles and recruit willing participants, who can either fabricate or import deceptive content and spread it on channels such as Facebook and LINE. Increasingly, China has recruited Taiwan-based content producers who will fabricate disinformation locally.<sup>42</sup> Furthermore, research found that Taiwanese internet celebrities are being hired to launch disinformation campaigns.<sup>43</sup>

<sup>40</sup> Harold, Beauchamp-Mustafaga & Hornung, "[Chinese Disinformation Efforts on Social Media](#)", op. cit., note 37

<sup>41</sup> Kenddrick Chan & Mariah Thornton, "[China's Changing Disinformation and Propaganda Targeting Taiwan](#)". The Diplomat, 19 September 2022.

<sup>42</sup> Harold, Beauchamp-Mustafaga & Hornung, "[Chinese Disinformation Efforts on Social Media](#)", op. cit., note 37.

<sup>43</sup> Chung Li-hua & Jake Chung, "[China Using Local 'Agents' to Spread Misinformation Online: Institute](#)", Taipei Times, 4 August 2019.



## Approaches to Disinformation: Parallel Work between Civil Society and Government

As shown above, disinformation from mainland China is a big issue in Taiwan, and the government has taken action to minimise the impacts of influence operations in the society. To tackle the issue, different stakeholders have developed various initiatives to prevent disinformation, to respond more quickly, and to prepare citizens.

### ***Government approach: The Disinformation Coordination Team (DCT)***

According to the report “Taiwan’s Response to Disinformation” from the National Bureau of Asian Research,<sup>44</sup> the Taiwanese government established the Disinformation Coordination Team (DCT) within the executive branch to oversee policies for countering disinformation and to coordinate with other agencies. This team promotes a four-facet framework, focused on:

**Identification:** As the upstream measure for countering disinformation, identification includes efforts to empower citizens to spot disinformation online. The response is closely related to the promotion of media literacy.

**Debunking:** The aim of this facet is to debunk disinformation campaigns more quickly. Before the DCT proposed the new framework, it took governmental agencies around six hours to approve and deliver press releases debunking a given narrative. With the new framework and principles established, the response time fell to just one hour.

**Combat:** This facet of the framework has been the most controversial, and has generated strong opposition from different stakeholders, as the Government aims to combat disinformation with stronger legislation. The resistance is based on stakeholders’ fears that such legislation will limit freedom of expression in the country.

**Punishment:** This facet deals with creating penalties for the harm that has been caused by disinformation. The government introduced legislation with provisions prescribing penalties for spreading disinformation on topics that cause specific harm.

Of the four aspects above, the DCT was able to fully implement the first two. In particular, in their debunking efforts, Taiwanese government agencies have adopted a short, humorous style to deflate rumours and streamline response times.<sup>45</sup>

For identification, the government created early education programmes. To support

---

<sup>44</sup> Shih-Shiuan, [“Taiwan’s Response to Disinformation”](#), op. cit., note 32.

<sup>45</sup> Ibid.

younger generations in dealing with disinformation, the Ministry of Education incorporated media literacy into its teaching guidelines, with the inclusion, in 2019, of the "Elementary and Junior High School Media Literacy Education Base School Programme". This builds on a long tradition: The first media literacy initiative can be traced to 2009, when the government was worried about the influence of traditional media on children's behaviour.<sup>46</sup> Due to this institutional legacy, Taiwan is seen as a world leader in countering online threats.<sup>47</sup>

## ***The civil society approach: Fact-checking and media literacy efforts***

As described above, some elementary and middle schools have incorporated media literacy lessons into their curricula, aiming to help students identify disinformation in their everyday lives. One of the most important CSOs in this area is the Taiwan FactCheck Education Foundation (TFEF).<sup>48</sup>

Among its efforts of fact-checking news in Taiwan, recently, TFEF announced that it will start working on media literacy and hold around 600 workshops between 2022 and 2025. The project is funded by Google, and the workshops will target those who may be disadvantaged by Taiwan's online ecosystem, including the elderly, residents living in remote areas, and recent immigrants.<sup>49</sup>



**Figure 8.** Representatives from the TFEF, Ministry of Education, National Communications Commission, and Google meeting to establish the new partnership. (Source: [Taiwan FactCheck Centre](#)).

<sup>46</sup> ["Media Literacy Education through Broadcasting on Campus"](#), Ministry of Education, Republic of China (Taiwan), 23 April 2009.

<sup>47</sup> Nicola Smith, ["Taiwan: Schoolkids to Be Taught How to Identify Fake News"](#), Time, 7 April 2017.

<sup>48</sup> Shih-Shiuan, ["Taiwan's Response to Disinformation"](#), op. cit., note 32.

<sup>49</sup> ["Taiwan Media Literacy Education Initiatives' launched with Google's US\\$1 million funding"](#), Taiwan Fact-check Center, 29 April 2022.

Taiwan's model to counter disinformation differs from the other case studies cited in this report, as civil society and government often work in parallel, rather than in formal partnerships. The distance between them this creates helps civic actors earn public trust in their independence and integrity, eliminating suspicions that they are part of a propaganda strategy on the part of the current government.<sup>50</sup>

This distance, however, doesn't stop the two stakeholders from complementing each other. Each actor benefits from the other's efforts and programmes to tackle disinformation. Civil society, for example, uses the concept and definition of disinformation provided by the Disinformation Coordination Team, and civil society's demands for the promotion of media literacy have largely been facilitated by guidelines proposed by the government. Similarly, the government uses the efforts of civil society to debunk disinformation, to elaborate on the necessity of combating disinformation, and to justify its own policy agenda.<sup>51</sup>

Hence, while the parallel work between the government and civil society might be seen as two unconnected efforts to tackle the same issue, both stakeholders reinforce each other's efforts and create a broader framework of initiatives to counter disinformation.

---

## Kenya: From Cambridge Analytica to Home-Grown Misinformation

Kenya has a history of domestic and foreign interference in its election processes, including by weaponising social media platforms and messaging apps at election time. During the 2017 Kenyan general elections, politicians and the now-defunct United Kingdom data firm Cambridge Analytica collaborated in using Facebook data to discredit political rivals.<sup>52</sup>

In August 2022, Kenyans went to the polls in another highly contentious presidential race. The general elections were held against the backdrop of two highly contested previous votes – one leading to post-election violence, in 2007, and one being nullified by the Supreme Court, in 2017.

Five years later, despite various efforts taken to mitigate their spread, misinformation, harmful content, and online violence remain key problems in Kenya. The line between false news and fact has become blurrier, with cheapfakes, other manipulated content,

---

<sup>50</sup> Shih-Shiuan, "[Taiwan's Response to Disinformation](#)", op. cit., note 32.

<sup>51</sup> Ibid.

<sup>52</sup> Justina Crabtree, "[Here's How Cambridge Analytica Played a Dominant Role in Kenya's Chaotic 2017 Elections](#)", CNBC, 23 March 201.

and fake profiles polluting the online information environment with unfounded claims. It comes as no surprise that, according to a recent Reuters Institute survey, 75 per cent of Kenyan news consumers find it difficult to differentiate between what is real and what is false information online.<sup>53</sup>

Throughout this year's election cycle, the spread of misinformation and disinformation on social media by political leaders and candidate's supporters sparked mistrust not only in the electoral process, but also in democratic institutions. This was indicative of a rise in coordinated home-grown misinformation campaigns to shift voters' opinions in favour of or against particular candidates.

## Disinformation in Kenya: Tools, Tactics and Narratives



### False Endorsement, Culminating in Fake Election Results

In the run-up to the 9 August vote, social media platforms in the country were swamped with political disinformation.<sup>54</sup> Research by the Mozilla Foundation shows that TikTok acted as a platform for rapid and far-spreading political disinformation before election day.<sup>55</sup> The sample of problematic content contained over 130 videos from 33 accounts, which were viewed collectively over 4 million times.

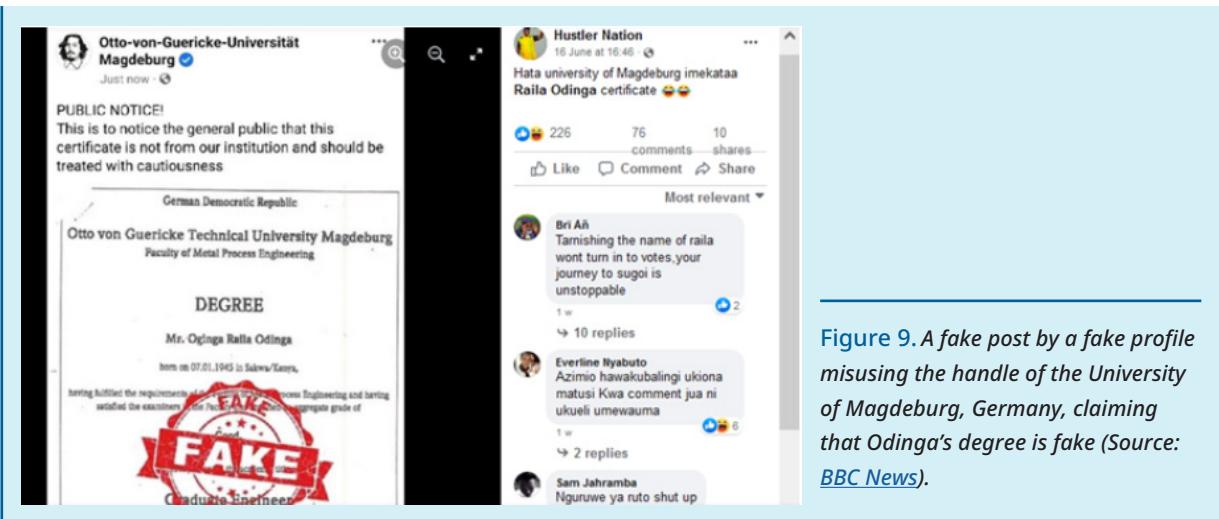
Disinformation campaigns were not only visible on TikTok, but also on Facebook. For example, supporters of both leading candidates, Raila Odinga and William Ruto, sought to cast aspersions on their opponent's educational qualifications. False narratives claimed that Odinga did not study engineering in Germany and that Ruto falsified his university grades. Kenya's electoral laws require candidates for the presidency to have been awarded a degree from a recognised university. These claims were debunked by fact-checkers but, nonetheless, trended on Twitter for days.<sup>56</sup>

<sup>53</sup> ["Reuters Institute Digital News Report 2021"](#), Reuters Institute, 1 September 2021.

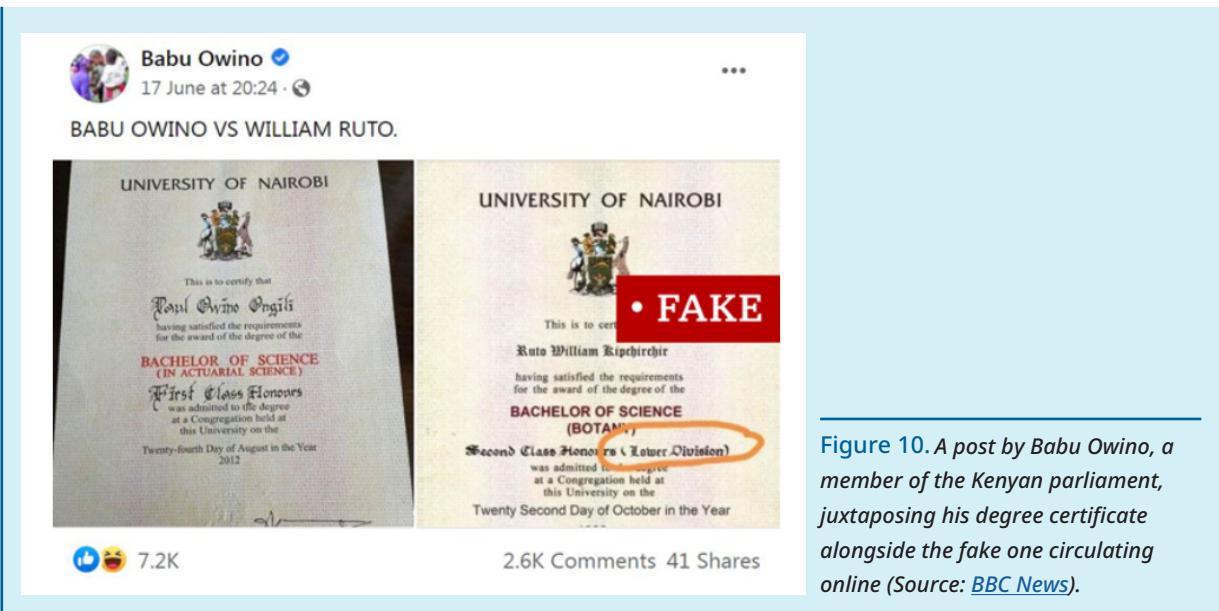
<sup>54</sup> ["Kenya: Tackling Misinformation Is Critical for Electoral Integrity"](#) Article 19, 1 September 2022.

<sup>55</sup> Odanga Madung, ["From Dance App to Political Mercenary: How Disinformation on TikTok Gaslights Political Tensions in Kenya"](#). Mozilla Foundation, 2022.

<sup>56</sup> Peter Mwai, ["Kenya Elections 2022: The Misinformation Circulating over Academic Qualifications"](#), BBC News, 11 July 2022.



**Figure 9.** A fake post by a fake profile misusing the handle of the University of Magdeburg, Germany, claiming that Odinga's degree is fake (Source: BBC News).



**Figure 10.** A post by Babu Owino, a member of the Kenyan parliament, juxtaposing his degree certificate alongside the fake one circulating online (Source: BBC News).

According to the fact-checking organisation AFP Fact Check Africa, campaigners for both frontrunners sought to delegitimise the results by accusing the opposing side of voter fraud and attempting to steal the elections.<sup>57</sup> Electoral results were often shared before the official body (the Independent Electoral and Boundaries Commission – IEBC), published the results.

After the results confirmed Ruto as the new president, his campaign accused the IEBC of mistakenly adding votes to presidential candidate Raila Odinga, questioning the Commission's integrity and competence. A fake version of an article from website of the newspaper Nation.Africa (see below), published a story that the IEBC had flagged 9,200

<sup>57</sup> Mary Kulundu & James Okong'o, ["Election Campaigning Ends in Kenya but Disinformation Battle Drags On"](#), Fact Check, 8 August 2022.

votes that had been “mistakenly added to Raila” in the Kisumu East constituency. The real Nation. Africa site later corrected the claim, given the attention the manipulative use of its identity had garnered.



**Figure 11.** The fake Facebook post about the alleged IEBC misbehaviour (left), with a fake website front page attached and the official Tweet by the website of the newspaper Nation.Africa (right), identifying the claim as false (Source: [PesaCheck](#)).



## Flimsy but Widely Shared Cheapfakes

Similar to the Brazilian case study, in the case of Kenya, cheapfakes – in their many variations – continue to be the dominant means of media manipulation. Here, most of the disinformation campaigns include doctored content targeting one of the frontrunners. One example is a manipulated video of former United States President Barack Obama that was originally posted on TikTok, and then widely shared on Facebook prior to election day. The video falsely suggested that Obama had endorsed presidential candidate William Ruto.

The video works with simple cheapfake editing techniques: a photoshopped picture of Ruto and fake banners across the screen, suggesting that the video is an excerpt from a BBC News Story.



*Figure 12. A doctored video, here including fact-checking labels, allegedly showing former United States President Barack Obama. The banner claims that Obama is announcing his support for candidate William Ruto (Source: [BBC News](#)).*

Another example of how technical disinformation tools were applied in the Kenyan election context can be found in the classic technique of manipulated audio. When Odinga hosted a campaign rally, Dennis Itumbi, a blogger supporting Ruto, tweeted a video of the crowd chanting Ruto's name when asked about their voting intentions. The audio, however, was manipulated.



Dennis Itumbi, HSC ✅  
@OleItumbi

...

ODM: Tarehe 9/8/2022 ni nani (mtachagua)?

#HustlerNation: @WilliamsRuto

• FALSE tena?

ISIOLO Hustler Nation: Ruto. Ruto.Ruto..

Life and its moments....

#Radico

Translate Tweet

*Figure 13. A tweeted video with manipulated audio, here including fact-checking labels, in favour of William Ruto (Source: [BBC News](#)).*

Kenyan disinformation actors, however, also use a variety of other tactics to spread disinformation.



## A Disinformation Industry across Platforms

In the 2022 elections, candidates used a range of different tactics to spread disinformation about their political opponents. These were similar to strategies used in the 2017 elections, but different in scale and scope, seamlessly supporting and feeding into the narratives disseminated across platforms:

### Fabricated opinion polls and press releases

According to AFP Fact Check, fake opinion polls have emerged as a prominent tactic for distorting public opinion, with campaigners in favour or against a particular presidential candidate falsely attributing such polls to legitimate survey companies, such as GeoPoll.<sup>58</sup> Fake press releases were another tactic used to influence voting behaviour.



### Home-grown, paid-for influencer disinformation campaigns

Far-reaching foreign influence campaigns have become a common disinformation practice. In Kenya, however, rather than foreign players, the 2022 election campaign shed light on the rise of domestic paid-for influencers with a wide reach on social

58 Ibid.

media,<sup>59</sup> taking advantage of the lack of enforcement of laws against hate speech and the manipulation of information online. These local influencers often serve as "disinformers", pushing key messages attributed to a particular candidate. With a simple search, LeMonde/AFP found several Facebook pages under the names of the two main presidential candidates.

### Codewords for amplification

While platforms are committed to taking steps to tackle the spread of disinformation, election influencers often rely on codewords to amplify their content and circumvent effective platform regulation, using misleading language or contexts to share information over social media. On TikTok, the Mozilla report reveals, users employed coded language, labelling some ethnic communities as "madoadoa", a Kiswahili word meaning "stain" or "blemish", which was listed as [hate speech](#) by Kenyan authorities ahead of the polls.<sup>60</sup>



## Approaches to Disinformation: Civil Society Taking the Lead

The diversity of tactics identified above calls for a set of diverse solutions if stakeholders in Kenya want to effectively counter disinformation. To combat the negative impacts caused by the dissemination of dis- and misinformation, CSOs in Kenya are actively initiating projects and programmes to counter malicious actors and their information manipulation campaigns. One of the continent's most active organisations implementing effective prevention mechanisms is Code for Africa (CfA). CfA is the largest network of civic technology and data journalism labs in Africa, with a country-specific offshoot in Kenya focusing on fact-checking and debunking disinformation. It is a unique actor in the Kenyan context, as its approaches to fighting disinformation are diverse, use different technologies and resources, and are able to tackle the issue from various angles.

### PesaCheck: Fact-checking and AI-based verification

PesaCheck forms part of CfA's portfolio of forensic verification and fact-checking projects. It examines content posted by public figures marked as potential misinformation on Facebook and other social media platforms. They use a number of social listening and content analysis tools, and collect requests from the public on claims that need checking, using a dedicated [WhatsApp](#) tip line and an [online form](#).

<sup>59</sup> ["In Kenya, Disinformation Factories up Production Ahead of the August Presidential Election"](#), Le Monde. fr., 9 May 2022.

<sup>60</sup> Samuel Kobia, ["Explaining Hatelex, a Lexicon of Hate Speech Terms in Kenya"](#), Nation, 23 April 2022.

**FALSE: Voting and announcement of results have not taken place at the Kenya High Commission in Australia**

Australia is not on IEBC's register of voters residing outside Kenya.

**False**

PESA CHECK

Figure 15. A screenshot of a piece of disinformation PesaCheck debunked.

PesaCheck provides deeper insights into posts the public see in social media feeds. Their approach tracks political promises by politicians, through other initiatives, such as the Wajibisha/PromiseTracker toolkit, unpacks budget and census data, through PesaYetu and TaxClock platforms, and builds machine learning and artificial intelligence tools, such as DebunkBot, to help automate verification.

The DebunkBot is a bot that detects when people share information that has been proven to be misleading. PesaCheck developed the bot to fight the spread of mis- and disinformation on social media, by responding to tweets sharing questionable links.

Replies to @Chris\_1791

Hello @Chris\_1791 , I'm debunkbot. The link you've just shared could lead to false information. Check out this link [checkyourfact.com/2021/09/13/fac...](https://checkyourfact.com/2021/09/13/fac...)



checkyourfact.com  
FACT CHECK: Image Claims A 'New' Study Found Ivermectin Causes Sterility I...  
'Infertility in men is not a known side effect of ivermectin and is not included as a side effect in the labeling.'

10:52 PM · Sep 21, 2021

Figure 16. A tweet from the account DebunkBot replying to a user with a message warning that the link they shared could lead to the spread of false information.

## Kenya elections hate speech/misinformation prevention on Wikipedia

Code for Africa also partnered with the United Nations to fight hate speech and disinformation during Kenya's 2022 elections. The campaign involved Wikipedia in Residence (WiR) fellowships, deployed by CfA, to help fight hate speech, incitement to violence, and disinformation.<sup>61</sup>

WiR mobilised regional Wikipedia editors and contributors to monitor and detect misleading and toxic content on Wikipedia, as the project suggests, given that while English/French-language Wikipedia entries for African content are adequately peer-reviewed, entries in indigenous African languages are riddled with misleading information.

Some of the work CfA did in this project included detecting and correcting poor quality and/or potentially planted disinformation related to the elections, carrying out a systematic review of current content on Kenya elections and democracy; and the uploading of pertinent new information, with a special focus on debunking disinformation with partner initiatives, such as PesaCheck.

---

## Disinformation in Northern Europe: Sweden and Finland

Disinformation in Northern Europe has been intertwined with the general security outlook on the continent. For Finland, with its long land border with Russia, and for Sweden, with its proximity to Russia and its important role in the Baltic sea, Russia's full-fledged invasion of Ukraine triggered fears of Russian disinformation campaigns and sparked renewed efforts to fight disinformation.<sup>62</sup> Following the 2014 Russian annexation of Crimea, both Finland and Sweden realised the danger of Russian disinformation campaigns. Disinformation has not, however, been solely a foreign-source phenomenon. Particularly in Sweden, extreme domestic right-wing actors have been instrumental in spreading false and deceptive information with malign intent.<sup>63</sup>

Despite the increasing threat of foreign influence operations, in particular stemming from Russia, both countries have been successful in fighting disinformation and maintaining confidence in media outlets and high levels of digital literacy. To understand their success, this report analyses the responses by both Finland and Sweden when it comes to fighting disinformation. While both chose state-driven solutions and used information, either in the form of pre-bunking or by promoting digital literacy,

---

<sup>61</sup> ["Kenya Elections Hate Speech/Misinformation Prevention - Meta"](#), Wikimedia.

<sup>62</sup> Lauri Kivinen, ["Pragmatism Defeats Propaganda - Finland's Move to NATO"](#), CEPA, 20 May 2022.

<sup>63</sup> Jack Stubbs & Johan Ahlander, ["Exclusive: Right-Wing Sites Swamp Sweden with 'Junk News' in Tight Election Race"](#), Reuters, 6 September 2018.

as their primary weapon in their fight against disinformation, their approaches have differed slightly, and are rooted in different traditions. The Swedish government opted to focus on disinformation primarily as a defence issue, building on a response developed throughout the Second World War.<sup>64</sup> Finland, on the other hand, has made disinformation a matter of general digital health, attempting to prepare society at large against the influx of deliberately false and misleading narratives.<sup>65</sup> The two approaches represent useful case studies, as they demonstrate two different successful solutions to the problem of disinformation.

## The Swedish and Finnish Disinformation Ecosystems: Tools, Tactics and Narratives



### Anti-Democratic Narratives and the Fear of Muslim Communities

The narratives that have affected both Sweden and Finland have a high degree of congruence. Essentially, they are divided into three categories: a) narratives that question the countries' approach to human rights, b) narratives attempting to instil fears about minorities, and c) narratives that attempt to re-shape the countries' positions towards Russia, particularly during the current process of Finland and Sweden's NATO accession.<sup>66</sup>

A series of disinformation narratives aimed at attacking both nations' socially liberal societal models have claimed the countries are safe havens for "necrophiliacs, paedophiles, coprophagists, and bestialists".<sup>67</sup> Both extreme right-wing actors and Russian-born narratives have attempted to instigate the fear of Islamic communities as breeding grounds for violence and crime.<sup>68</sup>

In Sweden, efforts to undermine social cohesion have not only been aimed at creating fear of Muslims but also at instigating fear of state institutions among Muslims. A particular example of such efforts occurred at the beginning of 2022, in an Internet campaign led by Shuoun Islamiya and other Arabic-language social media accounts, which alleged that Swedish authorities were kidnapping Muslim immigrants' children. The narrative spread rapidly among Arabic-speaking communities in Sweden, and was

---

<sup>64</sup> Miranda Bryant, "[Sweden Returns to Cold War Tactics to Battle Fake News](#)", The Observer, 6 February 2022.

<sup>65</sup> Eliza Mackintosh, "[Finland Is Winning the War on Fake News. Other Nations Want the Blueprint](#)", CNN.

<sup>66</sup> M. Cepurītis, I. Juurvee, A. Keišs, D. Marnot, S. Ruston & B. Carrasco Rodríguez, "[Russia's Footprint in the Nordic-Baltic Information Environment 2019/2020](#)", NATO Strategic Communications Centre of Excellence, 12 November 2020.

<sup>67</sup> Ibid.

<sup>68</sup> Ibid.

further fuelled by established media, such as TRT World and Al Jazeera, reporting the allegations without establishing their veracity.<sup>69</sup>



**Figure 17.** A tweet by the Swedish Ministry of Foreign Affairs warning of an ongoing disinformation campaign.

The Swedish disinformation campaign resonates with experiences in Finland. Ironically, in the Finnish case, it forms part of the success of the country's anti-disinformation efforts (see next section). After struggling to reach Finnish society, due to unusually high resistance to disinformation campaigns, the Kremlin shifted its tactics to target a new population. In Finland, Russian propaganda is now primarily targeted at minority communities, such as African and Middle Eastern immigrants.<sup>70</sup> Thus, disinformation, with the explicit intent of spreading fear of the Finnish authorities, has been spread in various languages of origin.

Finally, a third disinformation narrative has attempted to drive a wedge between the two Nordic countries and NATO, namely by portraying NATO as incompetent and aggressive. Simultaneously, disinformation actors have falsely portrayed Russia's interest in the countries as cooperative, tolerant, and peaceful. Specifically, Russian-backed narratives have stressed the idea that Finland's accession to NATO risks the loss of a special relationship between the two countries.<sup>71</sup>

<sup>69</sup> Elisabeth Braw, ["Americans, Like Swedes, Need Help Telling Fact From Fiction"](#), Foreign Policy, 27 May 2022.

<sup>70</sup> Stanisław Żaryn, ["Stop the Accession. Finland Is a Target"](#), StopFake, 26 August 2022

<sup>71</sup> ["Disinfo: Finland's Membership of NATO Breaks a Treaty with Russia"](#), EUvsDisinfo, 11 May 2022.

## Junk News and Laundered Content

To successfully propagate these narratives, disinformation actors have had to use innovative tactics, given the general public's resilience to disinformation. Two tactics have been particularly prevalent: so-called "junk news", and information and content laundering. In Sweden, there has been a significant proliferation of junk news, meaning news items that purport to be real news but are inaccurate or deceptive in their information content. A study conducted by Oxford University demonstrated that such junk news articles could reach a wide audience through social media amplification. Following a 10-day observation period, covering 275,000 election-related Tweets, the study concluded that roughly one-third of the articles shared during Swedish elections could be classified as "junk".<sup>72</sup> It further concluded that automated Twitter accounts, or "bots", were fundamentally contributing to the spread of junk news. According to the study, bot accounts were 40 per cent more likely to frequent websites hosting right-wing content, such as Samhällsnytt, Nyheter Idag, and Fria Tider. More than 85 per cent of the "junk news" content included references to these pages.

It is often difficult to determine the source of these misleading or inaccurate news items. Research by NATO's Strategic Communications Centre of Excellence indicates that this is at least partly due to strategic efforts by foreign countries to launder content.<sup>73</sup> The practice of content or information laundering (where false or misleading information is legitimised through a network of proxies or intermediaries) has been prevalent both in Sweden and Finland. The research shows that, particularly in Sweden, Russian content laundering attempts have been far-reaching and have been channelled through both Swedish-speaking (Fria Tider, Nyheter Idag) and English-speaking (Svensk Press, The World News, White TV, Offensive) sites to promote a pro-Kremlin narrative.<sup>74</sup> While Finland has demonstrated a higher degree of resilience to content laundering, in particular as domestic news outlets have been quick in identifying and debunking false narratives, Russia has still attempted to operate through Finnish pro-Kremlin media outlets (i.e. MV-Lehti), individual freelancers, and through social media influencers.

---

<sup>72</sup> Stubbs & Ahlander, "[Exclusive: Right-Wing Sites Swamp Sweden with 'Junk News' in Tight Election Race](#)", op. cit., note 63.

<sup>73</sup> Cepurītis et al, "[Russia's Footprint in the Nordic-Baltic Information Environment 2019/2020](#)", op. cit., note 66.

<sup>74</sup> Ibid.



## Approaches to Disinformation: Long-Term Strategies

### Sweden: A focus on psychological defence

While both Finland and Sweden rethought their strategic approaches after 2014, they reached different conclusions. Swedish efforts to counter disinformation culminated in the establishment of the Swedish Psychological Defense Agency (MPF). The Agency was established with the express purpose of maintaining the free flow of knowledge and information in an open society. State-coordinated psychological defence has a long history in Sweden, dating back to the Second World War, and a predecessor to the MPF existed throughout the Cold War.<sup>75</sup> In the current climate of Russian influence operations, the agency is tasked with combatting online deception and disinformation, and ensuring that government authorities communicate effectively with the public, both in times of minimal threat and in times of high alert.

MPF has tried to fight disinformation with various strategies of pre-bunking, as well as by increasing digital literacy. Prior to the 2022 Swedish elections, for example, the agency consulted parties and candidates on how to avoid becoming targets of disinformation campaigns and strategies. The agency has also produced various kinds of information materials and infotainment. As part of recent efforts, it developed a handbook for journalists on how to identify and defuse influence campaigns, as well as how to monitor Russian disinformation related to Sweden's NATO bid.<sup>76</sup> In addition, under the heading "Don't be fooled", the agency has produced a set of short videos aimed at raising awareness of disinformation and its impact. Among the topics discussed are how to identify bots, how to recognise fake news sites, and how to judge the validity of online information.

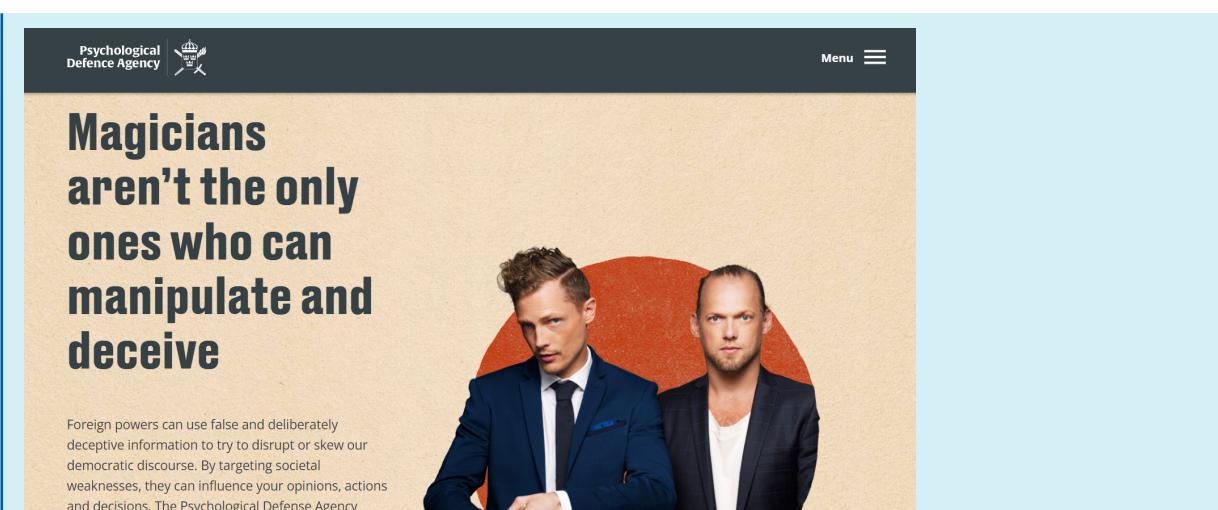


Figure 18. The opening page of the Swedish Psychological Defense Agency's "Don't be fooled" campaign (Source MPF).<sup>77</sup>

<sup>75</sup> Bryant, "[Sweden Returns to Cold War Tactics to Battle Fake News](#)", op. cit., note 44.

<sup>76</sup> Braw, "[Americans, Like Swedes, Need Help Telling Fact From Fiction](#)", op. cit., note 69.

<sup>77</sup> "[Don't Be Fooled](#)", Psychological Defence Agency.

## Finland: Education as a bulwark against disinformation

While the Swedish government established a new agency to combat disinformation, Finland has relied on its education system to maintain its lead in the battle against disinformation. In order to build societal resilience, Finland has employed a bottom-up approach, by making media literacy a cross-departmental priority and a key strategic objective. Since 2016, the Finnish educational curriculum has been revised to include teaching about disinformation in early education.<sup>78</sup> With the assistance of Faktabaari (FactBar), a fact-checking agency, the country has developed digital literacy "toolkits", geared towards both elementary and adolescent students.<sup>79</sup>

As part of the toolkit, students can learn, for example, critical skills for interacting with information sources before re-sharing material on social media. This form of early childhood media literacy education has been a cornerstone in the high resilience to disinformation in Finnish society.<sup>80</sup> It also forms part of a publicly funded security model, in which the government works with private businesses, CSOs, and voluntary organisations to build a network to combat disinformation.



**Figure 19:** A primary school class in Helsinki explains the differences between misinformation, disinformation, and malinformation (Source: *The Guardian*).<sup>81</sup>

<sup>78</sup> "US Experts Gird Finnish Officials for Information War", YLE, 22 January 2016.

<sup>79</sup> Kivinen, "Pragmatism Defeats Propaganda", op. cit., note 62.

<sup>80</sup> Mackintosh, "Finland Is Winning the War on Fake News. Other Nations Want the Blueprint", op. cit., note 65.

<sup>81</sup> Jon Henley, "How Finland Starts Its Fight against Fake News in Primary Schools". *The Guardian*, 29 January 2020.

# Going Beyond the Threats: Lessons Learned

The aim of this section is to understand what the lessons learned are from the various disinformation tools, tactics, and narratives, as well as proposed solutions to tackle such narratives. It is noteworthy that many of the political events that have been discussed in this report are too recent to allow the development of an analysis of their effectiveness and impact in the studied contexts. Furthermore, the intent of this report has been simply to map out different contexts, and not to rank them or make any judgments as to which approaches are better or worse. The aim is to understand what lessons can be learned from a variety of realities and events, and what these tell us about trends in emerging threats and solutions.

## Lessons Learned: Tools, Tactics and Stories

Chepfakes remain the tool of choice for disinformation actors: In most of the disinformation campaigns studied for this report, chepfakes were the stars of the show. This disinformation tool requires only low technical sophistication and consists of manipulating content in simple ways, such as the speeding up, slowing down, cutting, re-stating or re-contextualisation of content.<sup>82</sup> For malicious actors, it is still more advantageous to use simple forms of manipulation than to invest the time and knowledge necessary to use more advanced technologies, which require greater sophistication, such as deepfakes. The previous report highlighted that, with ever-advancing technology, malicious actors will continue to weaponise information and develop increasingly sophisticated tools for disseminating manipulated content; we haven't, however, reached that tipping point yet.<sup>83</sup>

Cross-platform sharing amplifies disinformation campaigns: Even though cross-platform sharing was only observed in the Brazilian context, it is still important to highlight this phenomenon. As "all-in" data plans are the main reason why the YouTube-to-WhatsApp pipeline has such an impact in the context of Brazil, this business model might become

---

<sup>82</sup> Beyer & Böswald. "[On the Radar](#)", op. cit., note 1.

<sup>83</sup> Ibid.

more prevalent in other countries. Disinformation actors often use content produced for one platform, such as YouTube, and amplify its message by resharing it on a different outlet, successfully evading algorithms. This tactic was highlighted in the YouTube-to-WhatsApp pipeline in Brazil. Not only are actors recycling the same content, but they selectively choose snippets of the content to frame and recontextualise the narrative in their favour.

Some foreign influence operations rely strongly on content laundering: The Swedish and Finnish case studies illustrate how content laundering and so-called “junk news” have become popular tactics for propagating manipulated narratives. These tactics are perfidious because, while they originate from abroad, they exploit trust relationships with domestic institutions.

Vulnerable groups are likely disinformation targets: Disinformation actors are likely to target vulnerable groups within societies. In the case of countries with a higher degree of resilience against disinformation, such as Sweden and Finland, disinformation actors have demonstrated the ability to adapt quickly, by selecting new targets to increase their efficacy. By targeting immigrants and developing narratives in their native languages, they have fostered distrust within these communities.

## Lessons Learned: Approaches to Disinformation

Investing in long-term solutions pays off: The case studies presented here illustrate different approaches to systematically combatting disinformation. Some invest in debunking, and others in promoting digital literacy and pre-bunking. The case of Finland shows that longer-term solutions need to involve digital literacy in the education system. Both Finland and Taiwan, in order to build societal resilience, have followed a bottom-up approach, by making media literacy a cross-departmental priority and a key strategic objective.

Citizens are willing to be part of the solution: Even in different geographic contexts, the case studies show that creating new solutions and channels for citizens to be part of the battle against disinformation generates results. In the Brazilian context, for example, the government received a high number of reports of electoral crimes, considered credible by experts from the TSE, through their digital channel created for the elections. Whether tech or non-tech approaches are employed, citizens are willing to adopt these and become anti-disinformation agents.

A multi-stakeholder approach is effective: Throughout most of the case studies, one aspect was a constant: Any solutions must rely on partnerships in order to have a more holistic and effective approach to fighting disinformation. The more diverse the

coalition involved, the more innovative the solutions will be. The Brazilian case study illustrated an expansive approach with multiple initiatives, including by state actors, civil society, and tech platforms. While many of the political events are too recent to make judgements about the effectiveness of these initiatives, they offer good potential to be effective. Furthermore, the study of Code for Africa in Kenya also shows how large anti-disinformation networks are able to develop a wide range of both technical and non-technical solutions to the problem of deceptive narratives.

Emerging disinformation threats call for emerging solutions that target the source causes (not just the symptoms) and embrace prevention (as opposed to recovery). The case studies emphasise how diverse state approaches can be when tackling disinformation in their local contexts. We can see, however, that technical solutions are not yet the go-to approaches for many countries, perhaps because advanced technological tools are not yet the preferred mode of creating disinformation. Yet, when actors do take technical approaches, these usually take place at the stage of debunking, which has limited impact once a successful narrative has reached its audience.<sup>84</sup> For the future, states should consider investing in automation and detection at an early stage, particularly as AI-powered disinformation tools become more prominent.

---

<sup>84</sup> Beyer & Böswald. "On the Radar", op.cit., note 1

# The Disinfo Radar Registry: A Novel Approach to Early Detection

The previous sections examined various initiatives aimed at combating disinformation, either by increasing societal resilience or by exposing false or deceptive information. Disinfo Radar seeks to complement such initiatives with a novel approach. To accomplish this, it shifts focus from messaging to the tools and tactics used to disseminate disinformation narratives. There are various reasons for doing so. First, the development of AI and, in particular, the subtype of machine learning, has changed the technical playing field for disinformation actors, as technological advances and open-access models are rapidly lowering entry barriers.<sup>85</sup> With the erasure of these barriers, the possibility of large-scale disinformation campaigns is growing. Thus, it is imperative to monitor emerging technologies that have the potential to propagate lies and half-truths.

Technology is, however, only one dimension of the changing global disinformation ecosystem. Thus, secondly, Disinfo Radar monitors disinformation tactics. In recent years, we have seen innovative approaches to circumventing fact-checkers and challenging digital forensics, whether they are simulating grass-root movements, cross-sharing between various platforms, or even emulating legitimate news sites.<sup>86</sup>

As developments in disinformation technology and strategies continue at breakneck speed, DRI has introduced an element of automation into the detection and analysis of disinformation tools and tactics. The following presents a system of computer-based tools that DRI has developed to warn of emerging threats.



## Early Detection of Emerging Technology: Finding the Risks

Since many tech innovations, such as synthetic image generation, were originally developed with legitimate goals in mind, their disinformation potential has often been overlooked. Originally, bots, as public or commercial information sources, and deepfakes, as entertainment technologies, served legitimate purposes.<sup>87</sup> In hindsight, their ability to scale up and improve disinformation has become obvious.

---

<sup>85</sup> Beyer & Böswald. "On the Radar", op.cit., note 1.

<sup>86</sup> Ibid.

<sup>87</sup> Tobias Knecht, ["A Brief History of Bots and How They've Shaped the Internet Today"](#), abusix, 4 May 2021.

Disinfo Radar attempts to leverage these insights. By using automated text analysis tools, Disinfo Radar aims to identify new disinformation technologies at an early stage. By auto-collecting and auto-analysing electronic preprint repositories (e.g., arXiv), industry papers (e.g., syncedreview.com), and policy publications (e.g., IEEE), Disinfo Radar scans the environment for indications of emerging technologies. Through a daily updated pipeline, it collects, processes, and subjects texts to state-of-the-art machine learning models, in order to identify technical innovations that could be abused for disinformation purposes.

Once texts have been collected (based on an auto-collection powered by web-scraping), they are assessed using a self-trained classifier (a support vector machine). The classifier serves as a form of pre-selection. In evaluating a text, the classifier determines whether it (a) refers to a particular technology and (b) whether that technology has the potential to mislead or increase mis- and disinformation. This classifier was trained using approximately 1,000 descriptions of diverse AI tools determined by DRI experts to have the greatest disinformation potential. Among these are the latest text-generation models (e.g., GPT-3), and text-to-image or text-to-video generators (e.g., Dall-E 2, Stable Diffusion, or Meta's Make-A-Video).

After passing the pre-selection round, the originality of the texts is re-evaluated. A second machine learning algorithm measures whether a given text is an outlier. What is meant by outliers? Outliers, in this context, are those texts that contain novel textual information. Such novel elements might, *inter alia*, be unheard model names, new approaches to leveraging data, or new forms of synthetic content. It is important to note that outliers can occur for various reasons, including a unique style or vocabulary an author uses. Hence, Disinfo Radar works on the basis of an interplay between automation and expert assessment. Disinfo Radar identifies these outliers by using transformer models (deep learning models that incorporate self-attention mechanisms). As these algorithms assist in clustering texts based on similarity, they can also be leveraged for identifying outliers.

Identifying outliers in the previous steps assists DRI's disinformation experts in their qualitative analysis. Using the registry results, they evaluate the identified technologies and determine the threat potential of each, by conducting additional desk research. When a technology is seen as embodying a potential threat, meaning that it could potentially be used to produce or amplify disinformation, DRI utilises the data obtained from the registry to inform potential stakeholders (see details below).



## Detection of Emerging Tactics

As the second pillar of analysis, Disinfo Radar seeks to detect specific strategies and tactics used by disinformation actors to create, disseminate, and amplify false and misleading information. DRI's extensive experience monitoring social media was a major asset in the design and implementation of Disinfo Radar.

Unlike disinformation tools, which undergo various stages of development and rarely disappear once they have been introduced, disinformation tactics can have significantly shorter lifespans. Often, they are used only for a short period of time before disinformation actors adapt their strategies.

By utilising nearly real-time monitoring of information, Disinfo Radar considers this aspect of disinformation strategies. Using a complex system of cross-references, DRI has identified what it considers to be an epistemic community of disinformation experts. The community comprises over 4,000 knowledge-based experts, including those working at renowned organisations such as the EUDisinfoLab, the Atlantic Council, Bellingcat, and many others.

By analysing the daily discourse that emerges in this community of experts, DRI utilises its skills in social media monitoring to identify new tactical trends that malicious actors employ. To distil these tactics, Disinfo Radar focuses on a set of markers that can serve as signposts for emerging trends.

These are trending topics, as well as the geographic focus of expert discussions. As disinformation strategies are often born out of contentious events (elections, wars, political struggles, etc.), it is paramount to discover early on what topics, regions, and actors disinformation experts focus on. To do so, Disinfo Radar provides various data-analytical tools.

## The Regional Focus Tracker

The Regional Focus Tracker determines in which region disinformation experts are currently most interested. Rather than just identifying global hotspots of interest, it provides the observer the chance to zoom in on specific countries. Providing the user with keywords that occur in conjunction with a given country, the Regional Focus Tracker allows for the identification of country-specific events or developments that might be useful for spotting disinformation.

## Topical Focus Tracker

The Topical Focus Tracker identifies and ranks disinformation experts' most-discussed topics on social media.

Topic Label	Associated Terms
Topic 1	* twitter * tweet * musk * file * moderation * content * account * hate * speech * takeover
Topic 2	* missile * drone * force * air * armed * brigade * tank * airfield * soldier * rocket
Topic 3	* cup * england * football * soccer * world * match * player * team * fan * sport
Topic 4	* minister * prime * foreign * ambassador * affair * meeting * consul * discussed * deputy * thanked
Topic 5	* china * protest * xi * yuan * shanghai * gulf * strict * quarantine * policy * protester
Topic 6	* episode * listen * latest * discus * season * tune * series * finale * dial * sprout
Topic 7	* court * supreme * case * appeal * harper * legislature * judge * argument * oral * ruling
Topic 8	* journalism * investigative * journalist * reporter * correspondent * reporting * intern * covering * career * medium
Topic 9	* feel * fact * thing * truth * wrong * completely * mean * dictionary * doubt * meaning
Topic 10	* runoff * election * walker * senate * turnout * democrat * race * vote * voter * win

Created with Datawrapper

## The Topical Focus Tracker

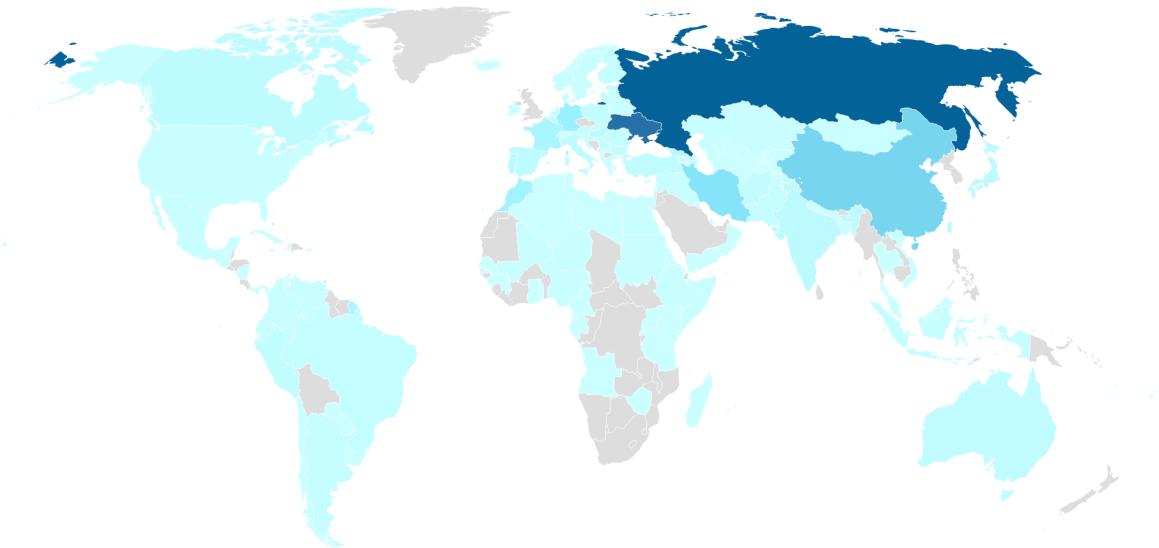
The regional context is just one facet of disinformation eco-systems. Disinformation strategies are not necessarily regionally contained, and can also occur in a global context; they can be linked to specific technological developments, specific events, or the emergence of new actors. Hence, Disinfo Radar does not only take a regional perspective, but also tracks the broader topics that emerge within the disinformation discourse.

## Regional Focus Tracker

The Regional Focus Tracker harnesses social media monitoring to pinpoint disinformation experts' geographical focus in near real-time.

Number of mentions

1 2372



Created with Datawrapper

As with data obtained concerning disinformation tools, data obtained concerning emerging disinformation tactics forms only the basis for additional desk research. Early warnings of arising disinformation tactics, obtained through the previously described data analytical tools, are corroborated by DRI's team of disinformation experts. When tactical innovations are identified, relevant stakeholders are informed.

Whether it is through DRI's rapid response briefs, DRI's own expert group, composed of ten internationally recognised disinformation experts, or DRI's own DisinfoCon conference, we have developed various channels for communicating novel and emerging threats to various interested parties. The stream of data produced by the registry will help to support DRI's continued work on mapping emerging disinformation threats and developing initiatives and studies on how different stakeholders can prepare, detect, and minimise the impacts of advanced disinformation campaigns.

