INTEL WATCH

# Building Digital Resilience

## A Counter-Electoral Disinformation Guide for Sub-Saharan African Countries

## EXECUTIVE SUMMARY

Electoral disinformation has sown disorder across Sub-Saharan Africa, damaging electoral integrity, social cohesion, and public trust in democratic norms and institutions.

This guide offers a practical and context-specific roadmap to understanding and combating electoral disinformation in the region. While grounded in regional and international best practice, it dismantles the often "Global North"-lensed assumptions about information disorders. And instead, foregrounds the lived realities, vulnerabilities, and technological patterns specific to the region and its countries.

Part One of the guide unravels the **WHAT, WHO, WHY, AND HOW** of electoral disinformation:

- **WHAT**: The Conceptual Frameworks of Disinformation: What is disinformation? What are the different types of disinformation? What are its theoretical frameworks? Guidance is provided to aid in distinguishing disinformation from other information disorders, as well as from political propaganda.
- **HOW and WHY**: Cultural and Social Drivers: How do local contexts and specific social, political, and cultural contexts of individual countries shape susceptibility to disinformation?
- **HOW and WHY**: Psychological Drivers: Why does disinformation resonate so deeply and undermine trust in institutions, media, and truth?
- **HOW and WHY**: Technical Drivers: How are social media platforms and recommender algorithms and other design choices exploited by malign actors to amplify falsehoods at scale?
- **WHO**: Actors and Tactics: Who is spreading disinformation, and what strategies, both covert and overt, are they using to sway public opinion or disrupt democratic processes?

Part Two outlines the **HOW TO**, a practical roadmap for countering electoral disinformation. It includes:

- Establishing Electoral Disinformation Response Teams: Practical guidance on the essential roles and skills needed to monitor and combat disinformation, along with advice on building partnerships with organisations already working in the information integrity space in Sub-Saharan Africa.
- Digital Ecosystem Statistics: Guidelines for mapping national-level vulnerabilities, including gaps in digital access, weaknesses in the media landscape, and disinformation risks. Guidance is also provided on how to factor in the unique social and cultural conditions of each country as part of this process.
- Using Social Media Analytics (SMA) to Monitor Misinformation: A proactive and reactive strategy that covers (1) collecting and analysing social media data, (2) a recommended SMA workplan covering the lead-up, during, and after an election is held.

A substantial annex is included to reinforce the guide's key learnings. It features:

- A comprehensive glossary of terms commonly used in the counter-disinformation field. In addition, the footnotes were specifically chosen to include text that explains the various concepts and issues extensively for those requiring deeper explanation
- Case studies of electoral disinformation monitoring projects from various African countries.
- Recommended books and articles offering deeper insights into the theoretical and technical foundations of disinformation and associated topics.
- A curated selection of similar electoral disinformation guides developed by international organisations.

## INTRODUCTION

Disinformation continues to be a destructive societal force across the globe, with far-reaching impacts that have eroded and undermined the foundations of civic life. From public healthcare and political participation to education and gender rights, no area of civic life remains unaffected.

It is to democracy and its institutions that disinformation has perhaps been the most harmful.

For the second year in a row, the World Economic Forum's (WEF) 2025 Global Risks Report[1] identified, along with social polarisation, the spread of false information as one of the top global risks. This, drawn from their annual Global Risks Perception Survey, a collation of insights from over 900 experts across academia, business, government, international organisations, and civil society. Ahead of the 2024 "super election year," within the context of approximately 4 billion people in 70 countries going to the polls, these experts had pegged misinformation and disinformation as the number one global risk, presenting a serious threat that could risk "political unrest, violence and terrorism, and a longer-term erosion of democratic processes."[2]

The prevalence of electoral disinformation across the African continent must not be underestimated. It is grim, best described by Tessa Knight, an Africa researcher with the Atlantic Council's Digital Forensic Research Lab (DFRLab), who warned:[3]

> *Every time I have set out to search for coordinated disinformation in advance of an election or around conflicts, I have found it. I have not investigated an online space in*

[1] World Economic Forum. (2025) *Global Risks Report 2025* https://www.weforum.org/publications/global-risks-report-2025/

[2] World Economic Forum. (2024). *Global Risks Report 2024* https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

[3] Africa Center for Strategic Studies. (2022, January). *Mapping disinformation in Africa* https://africacenter.org/spotlight/mapping-disinformation-in-africa/

> *Africa and not found disinformation. I think a lot of people are not aware of the scale of disinformation that is happening in Africa and how much it is distorting information networks.*

Despite growing regulatory and multisectoral efforts, disinformation remains a seductive tool for intentionally undermining democracy. Across the world electoral disinformation is being deliberately deployed by malign actors, despite full knowledge of its devastating impacts. Its appeal lies in its ability to bypass traditional checks, distort public discourse, and manipulate citizens at scale, turning democratic processes into contested battlegrounds and paving the way for authoritarian consolidation.

Disinformation is, of course, not a new phenomenon. It has long been used as a tool of political manipulation and social control for as long as humans have been able to communicate. What has changed is the speed, scale, and reach enabled by digital technology and, in Africa, fuelled by the continent's digital boom, with approximately 570 million internet users in 2022, projected to reach 1.1 billion by 2029.[4] In high-usage countries like South Africa, Nigeria, and Kenya, where social media is a primary news source, public concern about false information is among the highest globally.[5]

With this - arguably the most urgent crisis at the intersection of technology and democracy - what is to be done in Sub-Saharan Africa? How do we improve our strategies to mitigate the devastating impacts of electoral disinformation?

Unfortunately, there is no universally applicable approach. There is no one-size-fits-all approach for all countries and contexts. It is, or at least should be, a deeply contextual effort. Strategies must be tailored to the specific sociopolitical, cultural, and economic realities of each country, accounting for the wide differences in internet access, digital infrastructure,

---

[4] Statistica. (2024, February). *Internet Penetration Rate Africa 2022, Compared to the Global Rate.* https://www.statista.com/statistics/1176654/internet-penetration-rate-africa-compared-to-global-average/

[5] Pew Research Center. (2025, April 24*). Widespread Global Public Concern About Made-Up News*. https://www.pewresearch.org/global/2025/04/24/widespread-global-public-concern-about-made-up-news/

media freedom, and digital rights, all of which shape how information disorders spread across regions.

This guide aims to provide a strong starting point for those seeking to do the meaningful and urgent work of countering electoral disinformation. It serves as a back-pocket reference, providing readers with the essential concepts and tools needed to do this work. It does not attempt to explain every nuance in a field of knowledge that draws on numerous disciplines, including data science, behavioural science, political science, communications, and others.

The guide must be considered a living document. With its domain as digital technology, disinformation is both a rapidly and constantly evolving threat. It must therefore be updated over time to keep pace with the shifting dynamics of emerging technology, information disorder tactics, platform changes, and geopolitical contexts.

It is hoped that the guide will empower the Sub-Saharan community to build the capacity to monitor and respond to disinformation in a way that is informed, locally grounded, sensitive to each country's unique socio-political and economic context, and, importantly, effective. The Sub-Saharan region must build its own electoral disinformation institutional knowledge to protect electoral integrity and build digital and information resilience based on solutions suited for each country.

## PART 1: the WHAT, WHO, HOW, and WHY

### Decoding the Actors, Motives, and Mechanisms behind Sub-Saharan Africa's Electoral Disinformation Landscape

**Disinformation: a conceptual framework**

This section provides a concise conceptual overview of disinformation and related topics. It is not intended as a complete theoretical treatment of the topic. Instead, it offers a working knowledge of key concepts and patterns essential for navigating the Sub-Saharan African digital information landscape during election cycles.

Readers are strongly encouraged to consult the accompanying glossary (Annexure 1) while reading this section. Many of the definitions used in the counter-disinformation field are relatively new and constantly evolving, reflecting a field that has developed and continues to develop its own lexicon while attempting to keep pace with emerging technology and the ever-evolving tactics of disinformation merchants. For example, the phrase "fake news" initially referred to all information disorders but now specifically refers to falsified news reports. "`Disinformation" is now the preferred term. It is sometimes referred to as "misinformation and disinformation," but this is not strictly necessary unless conceptual delineation is required. The phrase "disinformation," on its own, captures the scope of the issue.

What is disinformation? The phrase is often used as a catch-all term; however, in practice, it refers to a specific, deliberate act. In the context of elections, disinformation is false or misleading information intentionally created and disseminated to manipulate public opinion, suppress voter turnout, discredit opponents, and/or distort democratic processes and outcomes. Misinformation, on the other hand, is disinformation shared on by those unaware of its intent to mislead, thus unintentionally amplifying its reach.

## Disinformation vs Information Disorders

Disinformation and misinformation are only two parts of a broader field known as "information disorders," the umbrella term that includes any content, whether false, misleading, or malicious, that distorts truth and undermines public discourse.

Information disorders encompass not only misinformation and disinformation but also malinformation, propaganda, conspiracy theories, clickbait, satire or parody presented as fact, and other forms of falsified and manipulated content.

| Type | Definition | Intent | Example |
|---|---|---|---|
| **Disinformation** | False or misleading information created or shared deliberately to deceive. | Intentional harm or manipulation | A fake news story, shared by political operatives, claims that a candidate rigged the election. |
| **Misinformation** | False or misleading information shared without intent to cause harm. | Unintentional | A friend shares a fake COVID-19 cure, believing it to be true. |
| **Malinformation** | Genuine information used in a misleading or harmful context. | Intentional, often malicious | Leaking private emails to discredit someone outside their original context. |
| **Propaganda** | Information used to promote a political cause or ideology. | Persuasion, control, influence | State media broadcasts only positive news about the ruling party. |
| **Satire/Parody** | Humorous or exaggerated content that mimics real news for entertainment. | Not intended to mislead | A satirical website publishes a fake headline about a politician's antics. |
| **Conspiracy Theories** | An unfounded explanation attributing major events to secret plots. | Mistrust, ideological framing | Claims that elections are controlled by a global elite. |
| **Hoax** | A fabricated story or event intended to deceive or trick the public. | Deception for attention or disruption | A viral tweet falsely announcing the death of a public figure. |
| **Clickbait** | Sensationalised or misleading headlines or visuals designed to attract attention and maximise clicks, often at the expense of accuracy or context. | Primarily to drive traffic or engagement | A headline claiming, "You won't believe what this politician just said!" that links to unrelated or exaggerated content. |

*Table 1: Information Disorders*

It must be clarified, for conceptual clarity, that disinformation is distinguished by its core feature: the intent to deceive.

Unlike satire, memes, or parodies, which are often legitimate forms of social, political, or artistic expression, disinformation is crafted and shared to deceive and manipulate, not entertain. For example, satire, while conveying inaccurate information, seeks to provoke thought or entertain, not mislead. Disinformation, on the other hand, does not have this motive; it seeks to manipulate belief or behaviour through deception.

That said, those forms of content can become disinformation when repurposed or shared in bad faith with the intent to mislead. These grey areas demonstrate the importance of contextual, intent-aware analysis, rather than simplistic classifications. This is a murky and often contested space that demands nuanced, context-aware analysis.

The table below offers a practical checklist to help determine whether content qualifies as disinformation or another form of information disorder.

| Type Of Content | Definition | Potential Legitimate Use | When It Becomes Disinformation | Key Questions to Ask |
|---|---|---|---|---|
| **Disinformation** | False or misleading information deliberately created or shared to cause harm or manipulate. | Rarely legitimate | When the intention is to deceive, manipulate opinions, or damage trust, such as during elections. | Who created it? What is their intent? Is there evidence of deliberate |
| **Misinformation** | False or misleading information shared without intent to cause harm. | Common in everyday sharing | When bad actors use it to launder disinformation or share it repeatedly despite being | Is the person sharing it aware that it is false? Has it been corrected before? |
| **Satire/Parody** | Humorous or exaggerated content is not meant to be taken literally. | Legitimate artistic, political, and social critique. | When presented out of context or repurposed to deceive, the information can be misleading. | Is it labelled as satire? Could audiences mistake it as factual? |
| **Memes** | Image and text formats conveying ideas, often humour or emotion. | Expression of opinion, humour, identity, or critique. | When deliberately crafted to mislead or incite, particularly using manipulated visuals. | Is the meme misleading? Who is circulating it and why? |
| **Deepfakes** | AI-generated synthetic media typically imitating real people or events. | Satire, parody, and memes | When used to impersonate, deceive, or manipulate (e.g., fake political statements). | Does it clearly disclose its synthetic nature? Could it impact public trust or |
| **Malinformation** | Factual information used out of context to cause harm. | Common in everyday sharing | When presented out of context or repurposed to mislead, deceive and cause harm. | Is it contextualised fairly? Is harm the goal? Is it framed to mislead? Is it in the public interest? |

*Table 2. Distinguishing Disinformation from other Information Disorders*

**Disinformation vs Political Propaganda**

It is also crucial to distinguish between political communication and disinformation, as the two often collide in a grey area. Not all politically motivated content that manipulates opinion is inherently disinformational. For instance, political propaganda may incorporate skewed, ambiguous interpretations of the truth or divisive or emotionally charged messaging, but it may not necessarily contain outright falsehoods, and therefore, it may not be strictly classified as disinformation. It may be selective, misleading, or manipulative, but not factually incorrect.

In electoral contexts, this distinction is crucial: political actors frequently employ strategic messaging and agenda setting that, although ethically questionable, may fall within the bounds of legitimate expression.

Guidance is provided below as a helpful checklist for assessing whether a piece of political communication is propaganda or disinformation.

| Criteria | Political Propaganda | Disinformation |
|---|---|---|
| **Intent** | Persuade, influence opinion, promote ideology or political agenda | Deceive, mislead, and manipulate public perception |
| **Truthfulness** | May be biased, selective, or emotionally charged, but not necessarily false | Intentionally false, misleading, or manipulated content |
| **Legality** | Often legal, protected as free speech (depending on jurisdiction) | Often illegal, unethical, or violates platform/community standards |
| **Harm Level** | Can polarise or reinforce narratives, but not inherently harmful | Undermines trust, electoral integrity, and democratic participation |
| **Tactics** | Slogans, selective facts, repetition, and emotional appeal | Disinformation content types include deepfakes, manipulated media, impostor content, and fabricated information. |

*Table 3: Distinguishing Disinformation from Political Propaganda*

Understanding these nuances and complexities is essential for those monitoring elections and digital discourse. Interventions must be grounded in technically correct content analysis. It is essential to carefully categorise content as either disinformation, another form of information disorder, or, while dicey, legitimate political, artistic, or literary expression.

**The Different Mediums of Disinformation**

Disinformation takes many forms and operates across multiple domains. To understand its full impact, it is important to distinguish between how content is falsified and where it's spread, i.e., the mediums or techniques used to deliver disinformation and which issue areas or topics it is being deployed in.

The following are the common mediums of disinformation, focusing on how the content is created.

1. Fabricated Content: Entirely false content, created to mislead or deceive.
2. Manipulated Content: Real images or information altered to distort meaning (e.g., misleading edits or headlines).
3. Imposter Content: When false sources mimic trusted brands or institutions (e.g., fake news websites styled as credible outlets).
4. Misleading Content: Misuse of factual information to present a false narrative (e.g., opinion framed as objective fact).
5. False Context: Accurate content placed in a misleading or false setting (e.g., recycled images used for new, unrelated events).
6. False Connection: When headlines, visuals, or captions are unrelated to the actual content.
7. Sponsored Content: Paid advertisements presented as genuine news without disclosure.
8. Deepfakes and Synthetic Media: Videos, audio, or images designed to depict people or events falsely, typically mimicking real people to deceive or manipulate perception. Often AI-generated or created with Photoshop and similar editing tools.

**Electoral Disinformation and Its Sub-categories**

Disinformation can also be delineated according to specific domains, or topics where it can exploit existing tensions and amplify social divides. Domains typically include health, climate

change, migration, race, gender, and ethnicity, as well as other socially polarising issues. These issues are particularly vulnerable to exploitation because they tap into identity, fear, and morality, making them fertile ground for manipulation and polarisation.

Electoral disinformation is such a domain, which can further be divided into distinct subcategories.

For the purposes of this guide, we focus on five key subcategories of electoral disinformation that have appeared repeatedly in disinformation campaigns not only across the Sub-Saharan Africa region but also across the globe. These are:

- Violent Extremism Disinformation,
- Election Denialism Disinformation,
- Voter Suppression, Disinformation,
- Identity-based Microtargeted Disinformation, and
- Gendered Disinformation

At the core of each of these subcategories is the exploitation of differences to create enmity, polarisation, political tribalism, and chaos because, in disinformation, division is not a by-product. It is the point. The "Psychological Drivers" section of the guide covers these emotions and the psychological manipulation aspect of disinformation in more detail.

Violent Extremism Disinformation is strategically leveraged by violent extremist actors to disseminate hateful narratives, incite violence, and recruit followers. It has been used to justify security crackdowns, delegitimise opposition movements, or incite communal violence and political unrest.

**Violent Extremism Disinformation Case Study: South Africa's July 2021 Unrest**[6]

In July 2021, South Africa experienced its worst civil unrest since the end of Apartheid. Triggered by the arrest of former President Jacob Zuma, the violence was fuelled by coordinated online disinformation campaigns. False claims circulated across Twitter, Facebook, and especially WhatsApp, alleging everything from a planned "civil war" to imminent food shortages and racial targeting. Hash tags like #FreeJacobZuma and #ShutdownSA were used to mobilise support and incite violence, often with misleading or fabricated videos of supposed attacks.

Disinformation exacerbated South Africa's racial tensions, inflamed political grievances, and amplified calls to mobilise and destroy public and private property. It blurred the line between protest and organised sabotage, resulting in more than 350 deaths, massive infrastructure damage in the billions of rands, and severe economic losses. The unrest occurred just months before the 2021 local government elections, disrupting voter registration processes and deepening distrust in state institutions.

Voter Suppression Disinformation seeks to discourage turnout by disseminating false information about voting dates, eligibility, or safety at polling stations. It often targets marginalised groups or opposition parties. Tactics include spreading fake polling place details, false voting deadlines, or fabricated eligibility requirements. It leverages platform manipulation, narrative seeding, astroturfing, fake news, deepfakes, manufactured amplification, and other disinformation tactics.

**Voter Suppression Disinformation Case Study: Nigeria's 2023 Elections** [7]

In the lead-up to Nigeria's February 2023 presidential election, sophisticated disinformation campaigns strategically targeted voter turnout in opposition strongholds and among young urban voters. Messages circulated widely on WhatsApp, Facebook, and Twitter, warning about imminent violence at specific polling stations and announcing

[6] Petla, V. (2023, November). Information disorders and civil unrest: An analysis of the July 2021 unrest in South Africa. *University of Johannesburg*. https://orcid.org/0000-0003-0320-6154

[7] Abba, A. I., Aluko, G. A., Chioma, A., Iruke, C., Ogide, V., Raji, A., Olatunji, A., Onoboh, H., Tijani, M., & Tola-Winjobi, F. (2023, June 8). Distorting Nigeria's Elections? How Disinformation Was Deployed in 2023. *IssueLab.* https://www.issuelab.org/resource/distorting-nigeria-s-elections-how-disinformation-was-deployed-in-2023.html

fabricated changes to voting dates and locations. Some messages explicitly told women, youth, or urban residents to "stay home" for their safety or hinted that voting would be futile due to fraud.

These narratives spread faster than official corrections, leading to confusion and fear. Civil society monitoring groups noted that political operatives and influencers were behind much of this disinformation. The coordinated effort suppressed turnout in key constituencies, particularly among demographics that threatened incumbent advantage.

Election Denialism Disinformation campaigns falsely claim elections are fraudulent or illegitimate, aiming to undermine trust in democratic outcomes. A great example is the U.S. "stolen election" narratives following the election of Joe Biden in 2020. Election denialism disinformation is typically spread post-election through conspiracy theories or narrative hijacking.

**Election Denialism Case Study: Kenya's 2022 Elections [8]**

Following Kenya's 2022 presidential election, opposition candidate Raila Odinga rejected the declared victory of William Ruto, alleging manipulation and a lack of transparency. Although the country's Supreme Court upheld the results and international observers deemed the process credible, a wave of denialist disinformation flooded social media, particularly Twitter, Facebook, and WhatsApp, with claims of vote rigging, manipulated results, and conspiracy theories targeting the Independent Electoral and Boundaries Commission (IEBC) and its officials. Viral hashtags and altered and fabricated statistical analyses were circulated by partisan influencers and anonymous accounts, framing the election outcome as illegitimate. This persistent narrative, despite its legal validation by the Supreme Court, sparked public doubt, polarised communities, and undermined confidence in both the IEBC and the judiciary, illustrating the powerful role of disinformation in weakening democratic institutions.

Identity-focused Microtargeted Disinformation targets specific groups of people based on their race, ethnicity, gender, religion, location, or social class by using customised messages that manipulate their particular fears or issues. It can convert to offline violence.

[8] Agbele, F. (2023) Disinformation and Misinformation During Kenya's 2022 Election: Implications for Voter Confidence in the Electoral Process. *Megatrends Policy Brief* 14, 30.03.2023, 9 Seiten. http://dx.doi.org/10.18449/2023MTA-PB14

**Identity-Based Microtargeted Disinformation Case Study: Ethiopia 2021-2022 Electoral Context**[9]

Identity-based disinformation was microtargeted during the Tigray conflict, with coordinated campaigns portraying Tigrayans as violent, disloyal, or inherently traitorous. These narratives were customised for different ethnic audiences using local languages and cultural references and disseminated through Facebook, Twitter, and Telegram. The content included fabricated atrocity stories, doctored images, and historical distortions, often pushed through hashtag brigading, fake accounts, and coordinated amplification. The impact was both digital and real-world: the disinformation fuelled ethnic hostility, legitimised exclusionary narratives, and contributed to offline violence. It undermined democratic discourse by framing an entire ethnic group as politically illegitimate, justifying both civic marginalisation and military aggression. The case illustrates how microtargeted disinformation can weaponise identity to destabilise multiethnic societies and fragment the information environment.

Here, it is worth exploring microtargeted disinformation and its exploitation of digital identities. In today's increasingly data-driven world, personal identities have become digitised and can be bought and sold. Social media platforms, in particular, collect vast amounts of behavioural, demographic, and psychographic data. This is not just what users share explicitly but also what platforms infer through engagement patterns, device use, and location tracking. This inferred identity data and psychographic profiling data - often bundled into advertising profiles - is then sold to data brokers or accessed through platform APIs, leaks, hacking, or third-party apps. It is this data that allows digitally microtargeted political digital messaging, both legitimate and covert.

Microtargeting itself is not inherently problematic. Political parties have long microtargeted their political messaging before the internet and continue to do so in the digital age. Ethical and legal questions arise when digital identity data is sourced unethically and weaponised for harmful purposes, such as was the case with the Cambridge Analytica scandal. Cambridge Analytica harvested vast amounts of psychographic profiling data from Facebook - without users' consent - by exploiting a quiz app that accessed not only the data of participants but also their friends. This psychographic data was then used to build detailed voter profiles and

[9] Amnesty International. (2022, December). *Kenya: Meta Sued for 1.6 billion USD for Fuelling Ethiopia Ethnic Violence*. https://www.amnesty.org/en/latest/news/2022/12/kenya-meta-sued-for-1-6-billion-usd-for-fueling-ethiopia-ethnic-violence/

microtarget political messaging in elections such as Nigeria (2015) and Kenya (2017), with the goal of manipulating voter behaviour through fear-based and personalised disinformation.

The growing accessibility of GenAI tools has raised concerns that Large Language Models (LLMs) could be exploited to supercharge political microtargeting ethically, unethically, and illegally.[10] These concerns stem from the potential for GenAI to facilitate highly scalable systems of psychological manipulation, targeting individuals based on their unique susceptibilities.

However, GenAI, as it currently stands, does not introduce any fundamentally new microtargeting techniques; it simply lowers the barrier to crafting microtargeted content based on digital identities. The more immediate threat lies in the use of GenAI to produce convincing synthetic media, rather than in dramatically new advances in the process of microtargeting.

Simchon, Edwards, and Lewandowsky explain:[11]

> *In one sense, this is a democratisation of capability, as anyone may create targeted content. However, the benefits of targeting would be expected to accrue to those actors who are best placed to deliver politically targeted content at scale to very large populations. And there is nothing to stop those actors from devising content that is untruthful or manipulative or both, creating the spectre of "gaslighting" populations by exploiting individual vulnerabilities.*

This issue highlights the fearmongering about GenAI technology that has arisen since the introduction of easy-to-use and widely accessible platforms. Take the 2024 "super election"

[10] Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, 3(2), pgae035. https://doi.org/10.1093/pnasnexus/pgae035
[11] Simchon, A., Edwards, M., & Lewandowsky, S. (2024).

year. Swift, large-scale AI-powered disinformation campaigns were predicted, with voters across the globe vulnerable to "unprecedented disinformation."[12]

However, the anticipated "armageddon" of AI-generated disinformation did not come to pass.[13] While GenAI tools were indeed used to generate deepfakes, traditional misinformation tactics remained. Where present, most deepfakes served to reinforce existing disinformation campaigns or were created as satire, education, or political commentary.[14]

What both the "supercharged disinformation" and "super microtargeting" panics suggest is that the continued focus should be on the structural and psychological conditions that make disinformation persuasive and powerful in the first place, i.e., dealing with the root causes of why disinformation and misinformation is believable and how to mitigate this.

This is not to say that GenAI will never be a problem, nor is it worth preparing for or worrying about. It is quite the opposite. This matter requires careful consideration and emphasises the importance of maintaining focus on the foundational issues that affect the believability of false information. As GenAI technology evolves and more elections take place over the coming years, the nature and scale of its impact may shift. We need to remain prepared in order to effectively mitigate this risk.

This, while bearing in mind that much of the predictive discourse on GenAI-powered disinformation relies on broad generalisations that assume uniform global effects. In reality, the impact of GenAI is deeply context-dependent on adoption rates, internet penetration, and affordability. Ignoring these contextual differences undermines the richness and nuance of

---

[12] de Groot, J. (2024). Electoral integrity is at stake in super election year 2024. *Atlantisch Perspectief*, 48(1). https://www.atlcom.nl/artikel-atlantisch-perspectief/electoral-integrity-is-at-stake-in-super-election-year-2024/

[13] Schneier, B., & Sanders, N. (2024, December). The apocalypse that wasn't: AI was everywhere in 2024's elections, but deepfakes and misinformation were only part of the picture. *Harvard Kennedy School Ash Centre*. https://ash.harvard.edu/articles/the-apocalypse-that-wasnt-ai-was-everywhere-in-2024s-elections-but-deepfakes-and-misinformation-were-only-part-of-the-picture/

[14] Chow, A. R. (2024, October 30). AI's underwhelming impact on the 2024 elections. *Time*. https://time.com/7131271/ai-2024-elections/

the conversations around proactive mitigation. It is therefore essential to root all analyses and mitigatory strategies of GenAI-generated disinformation within the specific, digitally informed realities of each Sub-Saharan African country.

Gendered Disinformation A subset of Technology-Facilitated Gender-Based Violence (TFGBV), Gendered Disinformation refers to false or misleading content specifically designed to target women and gender non-conforming individuals, particularly those in public, political, or activist roles. It invokes gendered narratives about their work, character, sexuality, and appearance. Gendered disinformation fundamentally aims to discredit and discourage women from participating in public life. Common techniques include doxxing, malinformation, deepfakes of a sexual nature, and coordinated harassment campaigns often involving threats of violence and rape. The goal is not only reputational harm but also to deter political participation and limit visibility in public discourse.

**Gendered Disinformation Case Study: Zimbabwean Elections**[15]

In a five-year study, the International Foundation for Election Systems (IFES) assessed gendered disinformation and online violence against women in Zimbabwe's electoral landscape. The study revealed that over 60% of abusive online political discourse targeted women, even though women comprised only about one-third of parliamentary representation. Using a real-time social media sentiment analysis tool, IFES tracked viral hate speech, smear messaging, and sexist threats, often sexualised or character-assassinating content, aimed at suppressing women's political participation. These narratives were spread through platforms such as Twitter, Facebook, and other online forums, reinforcing patriarchal norms and silencing female candidates and political figures. This pervasive digital abuse not only damaged individual reputations and caused emotional harm but also severely eroded trust in democratic institutions and deterred many women from political engagement.

[15] IFES, (2018). *Violence Against Women in Elections in Zimbabwe: An IFES Assessment* https://www.ifes.org/sites/default/files/migrate/vawie_in_zimbabwe_july_2018.pdf

**The cultural and social drivers behind disinformation in sub-Saharan Africa**

Disinformation is a deeply context-bound phenomenon. The socio-political embedding and the regional context in which it circulates are critical to understanding its impact.[16] Without this explicit contextualisation, efforts to combat will not be effective, as it misdiagnoses the motivations behind the sharing of misinformation, in particular.

Put simply: Disinformation doesn't exist in a vacuum; it works differently depending on where it spreads, who spreads it, and what local fears or beliefs it taps into. This requires careful attention to the specific sources, contextual drivers, and psychological mechanisms that shape how disinformation is created and received in each setting.

Therefore, disinformation must always be analysed within the media and political ecosystems of each Sub-Saharan country in which it is produced, distributed, and consumed. Vast differences exist often exist in all Sub-Saharan countries. As an example, South Africa has an internet penetration rate of near 80 percent,[17] while the DRC has 30.6 percent.[18] For the DRC, this represents a data collection challenge for the disinformation researcher and likely means that false information is less likely to be spread on social media and more on messaging platforms like WhatsApp, requiring methods that will allow the collection of that data ethically.

In the same way internet penetration influences the manifestation of disinformation, so do the unique cultural and social aspects of each country, particularly around how information is consumed and re-shared.

---

[16] Hameleers, M., (2023). Disinformation as a context-bound phenomenon: Toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory*, 33(1), 1–10 (2023). https://doi.org/10.1093/ct/qtac021

[17] DataReportal. (2025, March). *Digital 2025: South Africa*. https://datareportal.com/reports/digital-2025-south-africa

[18] DataReportal. (2025, March). *Digital 2025: DRC* https://datareportal.com/reports/digital-2025-democratic-republic-of-the-congo

Unfortunately, much of the research on the cultural and social drivers behind the spread of disinformation have long relied on frameworks developed in and for "Global North" realities. While these frameworks are valuable, they can sometimes obscure the local complexities and specificities that shape how disinformation manifests and spreads in countries that do not share those realities.[19]

Critical Disinformation Studies provides an invaluable framework for analysing disinformation with its call to consider the specific social, political, historical, and cultural contexts in Sub-Saharan Africa.

Rather than treating misinformation as the root problem, Critical Disinformation Studies argues that disinformation is often a symptom of deeper issues like political disenfranchisement, economic inequality, systemic mistrust, or media capture. This requires researchers to examine the underlying conditions that make societies vulnerable to manipulation in the first place.
Kuo and Marwick[20] suggest that the analyses of disinformation are more effective if they are:

- Grounded in history, society, culture, and politics
- Centred on analyses of how social differentiation, such as race, gender, and class, shape the dynamics of disinformation.
- Foregrounded with questions of how institutional power and economic, social, cultural, and technological structures shape disinformation, and
- Maintain explicit commitments to justice and equality

Kuo and Marwick further, and importantly, suggest that countering disinformation and misinformation must go beyond individual-focused solutions like fact-checking or media literacy and acknowledge that digital spaces are often deeply connected to lived, offline

[19] Mare, A., Mabweazara, H., & Moyo, D. (2019). Fake News' and Cyber-Propaganda in Sub-Saharan Africa: Recentering the Research Agenda. *"African Journalism Studies*. https://doi.org/10.1080/23743670.2020.1788295

[20] Kuo, R., & Marwick, A. (2021). Critical disinformation studies: History, power, and politics. *Harvard Kennedy School Misinformation Review*, 2(4). https://doi.org/10.37016/mr-2020-76

realities. In this view, community-led and grassroots efforts that combine political education with demands for systemic change offer more sustainable pathways for resilience.

Applying this framework to sub-Saharan Africa would, rightly, require shifting away from imported, surface-level interventions and towards a decolonised, power-aware, and locally rooted approach.

Due to the vast diversity across Sub-Saharan Africa, it is not possible to generalise what a "locally aware" or "grassroots-driven" approach should look like for the region as a whole. Mitigation and combating strategies must be developed at the country level, informed by national contexts, languages, media systems, and political dynamics.

However, there are also shared regional experiences, including overlapping histories of colonialism, structural inequality, and platform neglect that create points of correlation and opportunity for cross-border learning and collaboration. These are worth exploring.

The proliferation of false information in Sub-Saharan Africa cannot be separated from the realities of increasingly under-resourced newsrooms navigating rapidly evolving communication technologies.[21] Reputable news outlets are increasingly using paywalls to maintain their financial viability, often restricting access to trustworthy information to those who can afford it. As a result, many individuals, especially those with limited resources, are left to rely on freely available but often unverified content circulating on social media and messaging platforms. This has resulted in studies finding on information disorders in Sub-Saharan Africa:[22]

> *Thus, not everyone shares fake news with the intention to cause harm. In some cases, the sharing of fake news is influenced by ignorance and a genuine desire to inform friends, relatives, and family members. Because of limited access to verified*

[21] Mare, A., Mabweazara, H., & Moyo, D. (2019).
[22] Mare, A., Mabweazara, H., & Moyo, D. (2019).

*information, especially in rural and peri-urban areas, citizens are likely to share fake news without having the luxury to cross-check and verify its authenticity.*

A central cultural dimension is also highlighted in a study conducted across six Sub-Saharan African countries - Ghana, Kenya, Zimbabwe, South Africa, and Zambia highlight the importance of factoring in the cultural principles that shape interactions online.[23] African communities, particularly those in the Sub-Saharan region, are influenced by the longstanding cultural principles of Ubuntu - solidarity, interconnectedness, and mutual dependence - which may shape digital communication practices. Due to this cultural inclination, the sharing of misinformation may sometimes be motivated by a genuine desire to share information or warn or protect others and not spread harm.

Another important factor to consider is the broader context of digital repression and authoritarianism and its close links to the fight against disinformation.[24] In some countries where media freedom is limited, authorities frequently suppress freedom of expression, digital rights, and dissent under the guise of combating misinformation. Internet shutdowns, content filtering, surveillance, and restrictive legislation have thus become standard tools of control.

Press freedom also plays a crucial role in countering the prevalence of disinformation. When there is a free and vibrant press, journalists and media organisations can investigate, fact-check, and report on issues accurately and independently. They serve as important gatekeepers, providing reliable information to the public and exposing false narratives. In countries where press freedom is limited or suppressed, there is a higher risk of disinformation thriving. Lack of media independence and pluralism can lead to the

[23] Madrid-Morales, D., Wasserman, H., Gondwe, G., Ndlovu, K., Sikanku, E., Tully, M., Umejei, E., & Uzuegbunam, C. (2021). Motivations for sharing misinformation: A comparative study in six sub-Saharan African countries. *International Journal of Communication*, 15, 1200-1219. https://ijoc.org/index.php/ijoc/article/view/14801

[24] Kimumwe, P. (2022). "Digital Authoritarianism Hurting Democratic Participation in Africa." *Collaboration on International ICT Policy for East and Southern Africa (CIPESA*). https://cipesa.org/2022/06/digital-authoritarianism-hurting-democratic-participation-in-africa/

dissemination of biased or manipulated information by those in power or vested interests. Journalists may face censorship, intimidation, or even legal repercussions for reporting on sensitive or critical topics, which creates an environment conducive to the spread of disinformation.

As a result, disinformation countermeasures must go hand-in-hand with advocating for the protection and realisation of human rights, digital freedoms, and the right to access to information.

The following are suggested interventions that prioritise local agency, challenge the power of global platforms, and connect disinformation to material and historical realities rather than solely to online behaviour.

| **Socio-Cultural Factors Interventions** | |
|---|---|
| **Root Contextual Analysis in Local Realities:** | Action: Ground monitoring efforts in country-specific histories, cultures, politics, and social dynamics.<br>Why: Africa's diverse contexts (e.g., colonial legacies, cultural solidarity) shape disinformation's spread, and misaligned analyses risk ineffective responses to electoral disinformation.<br>To achieve this, train analysts to examine local narratives and power dynamics by utilising tools like social media and digital forensics to monitor sociocultural influences. |
| **Conduct Country-Specific Digital Ecosystem Studies** | Action: Collate digital ecosystem data to map each country's digital landscape, assessing internet access, media freedom, and disinformation vulnerabilities.<br>Why: Diverse sociopolitical environments (e.g., varying internet penetration) drive electoral disinformation, requiring localised insights to counter platform manipulation.<br>How: Identify gaps, such as linguistic disparities in moderation, and use data from the digital ecosystem to inform targeted interventions, which may include platform audits and fact-checking in local languages. |
| **Recognise Non-Malicious Motivations for Sharing** | Action: Acknowledge that disinformation is often shared out of civic duty, solidarity, or a desire to warn others, especially in low-trust settings with limited access to verified information.<br>Why: Cultural values like interconnectedness fuel grassroots amplification, amplifying false content without intent to harm.<br>How: Monitor social media for community-driven narratives, using social media digital forensics to distinguish non-malicious sharing from coordinated disinformation-for-hire. |
| **Address Structural Inequalities** | Action: Incorporate analyses of social differentiation (e.g., race, gender, class) and structural barriers like digital divides and paywalled news into disinformation monitoring.<br>Economic and social inequalities lead individuals to rely on unverified content, which subsequently amplifies electoral disinformation and socio-cultural influences.<br>How: Partner with NGOs to map access barriers and develop inclusive strategies, such as open-access news platforms, to reduce grassroots amplification. |
| **Regulatory Responses to Safeguard Against Authoritarian Misuse of Counter-Disinformation Measures** | Action: Advocate for policies and laws protecting freedom of expression and monitor government actions like surveillance, internet shutdowns, or restrictive laws justified as anti-disinformation efforts.<br>Why: Such measures often suppress dissent while enabling pro-government electoral disinformation, undermining digital democracy. |

| | |
|---|---|
| | How: Collaborate with civil society to expose state-sponsored disinformation using social media and digital forensics, and lobby for transparent digital policies. |
| **Support Independent Media and Equitable Access Action** | Action: Support independent newsrooms and promote affordable, trustworthy news sources to counter reliance on unverified social media content.<br>Why: Under-resourced newsrooms and paywalls limit access to verified information, fuelling electoral disinformation in rural and peri-urban areas.<br>How: Support local media, create open-access platforms, and partner with fact-checking organisations, such as AfricaCheck, to verify content related to elections. |
| **Culturally Relevant Digital Literacy** | Action: Develop community-based digital literacy programs tailored to local languages and cultural contexts, equipping citizens to critically evaluate information.<br>Why: Low media literacy and cultural distrust in local sources drive grassroots amplification of electoral disinformation.<br>How: Work with community leaders to deliver workshops via platforms like WhatsApp, focusing on prebunking and critical thinking to counter cognitive biases. |

*Table 4: The Social and Cultural Factors behind Disinformation*

**The psychological drivers behind disinformation**

This section provides the knowledge base for understanding of the cognitive, emotional, and behavioural mechanics behind the believability of disinformation and how to use those same emotional triggers to counter it based on the lessons of the Behavioural and Cognitive Sciences.

This is not just an academic exercise; it is essential for crafting counter-disinformation messaging that works. If disinformation exploits emotional triggers and cognitive shortcuts to achieve virality and believability, then responses to counter it must speak to those same psychological drivers. Presenting fact-checked information is not enough; if it does not resonate emotionally, it will not stick.

This means adopting a tone, format, and delivery that captures attention, resonates emotionally, and appeals to senses of identity and belonging, all without compromising factual integrity.

***Emotionally Charged Messaging:***

Disinformation is more prevalent, influential, and persistent on topics that are politically charged than neutral or non-divisive ones.[25] That is because politics is inherently intertwined with human emotions and closely connected to the formation of political attitudes and public opinion. Emotions affect both the types of information individuals seek and the ways it is processed. This can lead to selectively seeking information that confirms beliefs, a misinterpretation of that information, or challenging claims that contradict the beliefs.[26]

This is knowledge political parties, and their communicators are aware of, and political messaging is crafted to evoke the core emotions of fear, anger, outrage, loyalty, and hope. When emotionally charged, messaging becomes more persuasive, more shareable, and more likely to bypass critical thinking. With disinformation messaging, these emotional triggers are strategically designed to manipulate, making false content spread faster and resonate more deeply with audiences. As Raffio and Blumenthal explain:[27]

> *The power of misinformation lies not in its factual content but in the emotional response it elicits. Content that provokes outrage, anger, or a sense of injustice can spread rapidly and influence beliefs, regardless of its veracity. Misinformation that triggers fear is one of the most powerful rhetorical devices and has been used in propaganda for centuries.*

There are five strategies commonly observed in political disinformation to evoke the desired emotional reaction to make it not only shareable but also believable. These are[28]

- Emotionally charged language that evokes fear, anger, and other negative emotions

[25] Zhou, Y., & Shen, L. (2024). Processing of misinformation as motivational and cognitive biases. *Frontiers in Psychology*, 15, Article 1430953. https://doi.org/10.3389/fpsyg.2024.1430953

[26] Webster, S. W., & Albertson, B. (2022). Emotion and politics: Noncognitive psychological biases in public opinion. Annual Review of Political Science, 25, 401–418. https://doi.org/10.1146/annurev-polisci-051120-105353

[27] Raffio, N., & Blumenthal, A. (2024, October). Expert explains how misinformation thrives on emotional triggers—and why traditional fact-checking often misses the mark. *Phys.org*. https://phys.org/news/2024-10-expert-misinformation-emotional-triggers-traditional.html

[28] Zhou, Y., & Shen, L. (2024).

- The presentation of an incoherent or mutually exclusive argument
- The framing of issues in false dichotomies
- The scapegoating of individuals or groups to reduce the complexity of a problem
- The resort to ad hominem attacks that target the speaker rather than their arguments

This strategy has been evident in South Africa, where political disinformation typically has a highly exaggerated emotional tone to provoke anger, and that is why, in some instances, it is converted from online hostilities into offline violence.[29] Persistent dichotomising and scapegoating disinformational portrayals of primarily Black foreign nationals as criminals responsible for poverty, crime, and unemployment, paired with violent rhetoric, has and continue to lead to xenophobic violence offline.[30]

The principle is this: political disinformation spreads because it feels true. It taps into core emotions. Therefore, responses to combat it cannot be clinical or emotionally flat. It cannot appeal to reason. It must appeal to emotion. It must equally feel true.

Below are some guidelines on crafting emotionally resonant messaging to combat disinformation.

| Principle | What It Means | Say This (Effective) | Not This (Ineffective) |
|---|---|---|---|
| Evoke relatable emotions | Tap into pride, frustration, or shared pain to make truth emotionally resonant. | "We all want leaders who tell the truth. Lies about the election aren't just politics; they're an insult to our intelligence." | "This post contains misinformation. Electoral processes were followed correctly." |
| Anchor in identity | Tie the message to the audience's values, such as patriotism, faith, justice, or civic duty. | "As proud citizens, we have the right to question, but let's not let false claims tear apart the country we've built together." | "You are being manipulated by fake news. Please check the facts." |
| Use narrative, not just data | Share real people's experiences instead of citing rules or reports. | "I was at the polling station. I saw the process. We disagreed on the outcome, but I saw integrity in action." | "According to the electoral commission report, no irregularities occurred." |

[29] Gagiano, M., & Marivate, V. (2023). Emotionally driven fake news in South Africa. In *EPiC Series in Computing: Vol. 93. Proceedings of Society 5.0 Conference 2023* (pp. 56-67). https://easychair.org/publications/paper/gzdS

[30] Fokou, G., Yamo, A., Kone, S., Koffi, A. J. d'Arc, & Davids, Y. D. (2022). Xenophobic violence in South Africa, online disinformation and offline consequences. *African Identities*, 22(4), 943–962. https://repository.hsrc.ac.za/handle/20.500.11910/19661

| | | | |
|---|---|---|---|
| Visual and sensory cues | Use images, slogans, and tone that grab attention and trigger emotional reaction. | | |

Table 5: Crafting Emotionally Resonant Messaging

***Cognitive Biases as Behind Disinformation Believability***

Subjective cognitive biases shaped by emotion, identity, and memory often override logical reasoning, distorting judgements and choices. Once processed and embedded in cognitive biases, information becomes resistant to correction. The brain favours consistency over accuracy.

Disinformation exploits these biases deliberately, making it not only persuasive but also difficult to debunk or dislodge once believed.

Social media platforms further exacerbate this problem by curating the information users see based on their personal preferences. By catering to personal preferences and building echo chambers and filter bubbles, algorithms confirm and reinforce existing subjective opinions and cognitive biases.[31]

At play is also social media's information overload constantly forcing the brain to make quick judgements, reducing its ability to process content critically. Filtering out false information becomes mentally taxing. How algorithms and other design choices by social media platforms influence the spread and believability of disinformation is covered in the next section, "The Technical Drivers Behind Disinformation."

A study found five cognitive biases frequently present when processing false political information. These are:[32]

| Bias | Description | Effect In Disinformation |
|---|---|---|

[31] Boonprakong, N., Tag, B., & Dingler, T. (2023). Designing technologies to support critical thinking in an age of misinformation. *IEEE Pervasive Computing*, 22(3), 8–17. https://doi.org/10.1109/MPRV.2023.3275514

[32] French, A.M., Storey, V.C. and Wallace, L. (2025) The impact of cognitive biases on the believability of fake news. *European Journal of Information Systems* 34, no. 1, 72-93. https://scispace.com/pdf/the-impact-of-cognitive-biases-on-the-believability-of-fake-2sgxrbjx10.pdf

| | | |
|---|---|---|
| Herd Mentality | Individuals often adopt the behaviours, beliefs, or actions of a group without critical evaluation, simply because they see that everyone else is doing it. | Can fuel viral spread as people share content to conform or signal belonging. |
| Framing Bias | The way information is presented, or "framed," can influence how it is perceived and interpreted. | Disinformation often exploits framing to manipulate emotional responses. |
| Overconfidence Bias | The tendency to overestimate one's knowledge, abilities, or accuracy of judgement. | Can lead individuals to confidently spread false information, believing they are correct even in the absence of evidence. |
| Anchoring Bias | The tendency to rely too heavily on the first piece of information received (the "anchor") when making decisions or forming judgements. | Initial exposure to a false claim can shape how subsequent information is interpreted, even if corrections are presented later. |

*Table 6: The Cognitive Biases Influencing Disinformation Belief*

Is it possible to correct false information once deeply embedded in cognitive biases?

Inoculation theory, rooted in behavioural science, explains how to strengthen beliefs or attitudes by exposing individuals to a weakened form of misinformation in advance. This process, known as psychological immunisation or prebunking, helps build resistance to future manipulation or persuasive falsehoods. "Misinformation in and of itself is not inherently dangerous if nobody believes it. If everyone simply scrolled past it and gave it no attention, the problem would be much easier to contain."[33]

Unlike debunking or fact-checking and its response to misinformation after it spreads, prebunking aims to build resilience before exposure to it.

In addition, debunking, while useful, comes with several other challenges:[34]

- Establishing what counts as factual information is epistemologically difficult, particularly in the context of politics.
- Fact-checks are unlikely to reach everyone who was exposed to the initial misinformation

[33] Traberg, C. S., Harjani, T., Basol, M., et al. (2023). Prebunking against misinformation in the modern digital age. In T. D. Purnat, T. Nguyen, & S. Briand (Eds.), *Managing infodemics in the 21st century: Addressing new public health challenges in the information ecosystem* (Chapter 8). Springer. https://doi.org/10.1007/978-3-031-27789-4_8

[34] Roozenbeek, J., Maertens, R., McClanahan, W. P., & van der Linden, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(31), eabo6254. https://doi.org/10.1126/sciadv.abo6254

- Getting people to believe fact-checks is challenging.
- Correcting misinformation does not always nullify its effects entirely, a phenomenon known as the "continued influence effect."

This is not to imply that debunking is pointless. The two must be used in tandem. Prebunking builds resistance before misinformation is encountered, while debunking corrects false beliefs after exposure.

Below are best practices for crafting prebunking messaging that restores truth and strengthen democratic discourse to be used in tandem with messaging crafted to be emotionally resonant.[35]

| Prebunking Principle | Best Practice |
|---|---|
| **Focus Your Prebunk** | Choose either a specific false claim or a broader manipulation tactic to address. |
| **Lead with Truth** | Start and end with accurate information—use a "truth sandwich" format. |
| **Provide a Clear Warning** | Alert the audience that they may encounter misinformation and why it's misleading. |
| **Explain the Manipulation** | Clearly outline the misinformation technique (e.g., scapegoating, false experts). |
| **Offer Simple Counters** | Keep rebuttals concise, focused, and easy to remember. |
| **Be Transparent** | Share what is known and unknown to build credibility and trust. |
| **Keep Language Simple** | Avoid jargon; make messages accessible and straightforward. |
| **Make It Shareable** | Use mobile-friendly, visually engaging formats suitable for social sharing. |
| **Tailor to the Platform & Audience** | Adjust tone, style, and format for the platform and cultural context. |
| **Use Familiar Contexts** | Frame examples around well-known or neutral topics to build recognition and trust. |

*Table 7: Prebunking Best Practices*

[35] First Draft. (2020). *A Guide to Prebunking: A Promising Way to Inoculate Against Misinformation.* https://firstdraftnews.org/articles/a-guide-to-prebunking-a-promising-way-to-inoculate-against-misinformation/

**The technical drivers behind disinformation**

There is no better and more accessible explanation of how social media platforms are designed in ways that facilitate the spread of disinformation than the Netflix documentary The Social Dilemma,[36] which, through interviews with former tech insiders and experts, illustrates how popular platforms like Facebook, Twitter, and YouTube are built to maximise user engagement, not truth.

It explains how recommender algorithms, data-driven profiling, and engagement-based ranking systems work together to create echo chambers, amplify outrage, and prioritise emotionally charged or misleading content. The film makes it clear that these systems are not neutral; they are optimised for attention, often at the expense of accuracy, democratic health, and human wellbeing. While simplified, its depiction provides an important starting point for understanding the technical dynamics behind the spread of disinformation. It is strongly recommended viewing!

Rather than overwhelming readers with the vast and rather technical literature on how disinformation spreads through digital systems, a distilled overview of the key concepts, curated as a summary that captures the essentials, is provided below. The annexure also includes a glossary and further reading.

| Design Decision | Impact on Disinformation |
|---|---|
| Engagement-Driven Algorithms[37] | Social media platforms like Facebook, Twitter/X, Instagram, and YouTube use engagement-driven algorithms that prioritise content with more likes, shares, and views. This often boosts sensational, emotional, or divisive content. While people spread misinformation, algorithms exacerbate the problem by promoting content that attracts attention, rather than what is accurate. |
| Reward Structure for Sharing[38] | Social media platforms' reward structures, such as likes, retweets, and comments, encourage habitual sharing. Users who post frequently for social validation or visibility are up to six times more likely to share |

[36] The Social Dilemma (2020) https://thesocialdilemma.com/the-film/

[37] Ferrara, E., Chang, H., Chen, E., Muric, G., & Patel, J. (2020). Characterizing social media manipulation in the 2020 U.S. presidential election. *First Monday*, 25(11). https://doi.org/10.5210/fm.v25i11.11431

[38] Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

| | |
|---|---|
| | false information than those who post occasionally. This happens because habitual sharing shifts focus from accuracy to engagement, as platforms reward attention over truth through constant feedback loops. |
| Echo Chamber Design[39] | Social media algorithms amplify echo chambers by prioritising content that aligns with users' beliefs and creating polarised communities were misinformation spreads easily. A study of 100 million posts revealed that recommendation systems cluster users into like-minded groups, limiting exposure to diverse or corrective views. While humans seek out confirming content, algorithms reinforce this by curating feeds that keep users engaged and trapped in echo chambers. |
| Language Disparity in Content Moderation [40] | Content moderation in non-English-speaking regions, especially in the Global South, such as those in Sub-Saharan Africa, is weak due to underinvestment and colonial biases in Natural Language Processing (NLP) systems. Although 75% of internet users come from non-English-speaking countries, online platforms prioritise English, which results in inadequate moderation tools for other languages. This enables misinformation to spread unchecked. Humans post culturally specific falsehoods, but platforms worsen the problem by failing to detect them. Malign actors often exploit this by using word camouflaging to evade detection. |
| Anonymity and Lack of Source Verification[41] | Anonymity and lack of source verification on platforms like Facebook and Twitter enable disinformation campaigns by allowing actors to operate without accountability. In 81 countries, state and non-state actors exploited anonymous accounts to spread propaganda, as seen with Russia's Internet Research Agency reaching 126 million users during the 2016 U.S. election. Humans manipulate narratives using unverified identities, while weak platform verification systems allow misinformation to spread unchecked. |
| Forwarding and Sharing Mechanisms[42] | Features like retweeting and one-tap forwarding allow misinformation to spread rapidly with little scrutiny. On WhatsApp, users could once forward messages to 250 groups, fuelling the viral spread of false content. A study of 2.6 million messages from Brazilian groups revealed that misinformation spread more quickly and frequently than accurate content. Trust within groups and impulsive sharing by users amplified this, while WhatsApp's design lacked safeguards to prevent the spread of harmful content. |
| Limited Transparency in Algorithmic Decisions[43] | Platforms like Facebook and Twitter use opaque algorithms that prioritise content based on engagement, often amplifying sensational or misleading posts. Users and regulators lack insight into how these feeds are curated, making it difficult to understand or challenge the spread of viral misinformation. Humans fuel this by engaging with provocative content, but hidden algorithms worsen the problem by boosting it without transparency. |

[39] Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

[40] Shahid, F., & Vashistha, A. (2024). Colonialism in content moderation research: The struggles of scholars in the Majority World. *Center for Democracy and Technology*. https://cdt.org/insights/colonialism-in-content-moderation-research-the-struggles-of-scholars-in-the-majority-world/

[41] Howard, P. N., & Kollanyi, B. (2016). Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. *Oxford Internet Institute*. https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2018/07/ct2018.pdf

[42] Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., & Benevenuto, F. (2019). Analyzing textual (mis)information shared in WhatsApp groups. *arXiv preprint* https://arxiv.org/abs/1909.08740

[43] Gillespie, T. (2018). Custodians of The Internet: Platforms, Content Moderation, and The Hidden Decisions That Shape Social Media. *Yale University Press.*

| | |
|---|---|
| Monetisation Incentives for Viral Content[44] | Social media platforms' business models rely on ad revenue tied to user engagement. Divisive or misleading content often generates higher engagement, incentivising platforms to tolerate it, as it drives clicks and ad revenue. For example, 44% of the top 50 Facebook posts about mail-in voting contained misinformation, yet their virality benefited the platform's engagement metrics. |
| Inadequate Fact-Checking Integration[45] | Fact-checking on social media platforms is often delayed, inconsistent, and less visible than the original false content. Platforms prioritise speed over verification, allowing misinformation to spread widely before corrections appear. False stories can outpace fact-checks by a factor of 10. While users share misinformation impulsively, weak and reactive fact-checking systems let falsehoods dominate. |
| Design for Speed Over Accuracy[46] | Platforms prioritise speed through features like live streaming and instant posting, enabling misinformation to spread faster than it can be moderated. The 2019 Christchurch shooting was livestreamed on Facebook, reaching thousands before takedown, with copies shared widely. This speed-first design, driven by engagement and competition, outpaces verification. Humans post sensational content in real time, but platforms' rapid dissemination tools let it spread unchecked. |

*Table 8: The Technical Drivers Behind Disinformation*

The design decisions, from engagement-driven algorithms, reward structures, echo chambers, inadequate moderation, anonymity, easy sharing, opaque algorithms, monetisation incentives, weak fact-checking, and speed-over-accuracy architecture, create fertile ground for misinformation to flourish.

However, as Vosoughi et al.[47] incisively note, "False information spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it."

[44] Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., & Benevenuto, F. (2019). (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. *The World Wide Web Conference*, 818–828. Association for Computing Machinery. https://www.researchgate.net/publication/330825332_MisInformation_Dissemination_in_WhatsApp_Gathering_Analyzing_and_Countermeasures

[45] Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe*. https://rm.coe.int/information-disorder-report-november-2017/1680764666

[46] Donovan, J., & Boyd, d. (2021). Stop the presses? Moving from strategic silence to strategic amplification in a networked media ecosystem. *American Behavioral Scientist*, 65(2), 238–249. https://www.researchgate.net/publication/356909505_Stop_the_Presses_Moving_from_Strategic_Silence_to_Strategic_Amplification_in_a_Networked_Media_Ecosystem

[47] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. https://ide.mit.edu/wp-content/uploads/2018/12/2017-IDE-Research-Brief-False-News.pdf

This sobering reality cautions against scapegoating algorithms alone; humans, driven by emotion and cognitive bias, eagerly propagate falsehoods, while platforms amplify these impulses through flawed design. The blame is shared: humans fuel the fire, and platforms provide the kindling.

We must demand transparency, robust moderation, and designs that prioritise truth over virality on social media platforms to combat disinformation while also encouraging users to critically reflect on their sharing. Only through this dual accountability can we dismantle the machinery of misinformation and reclaim a digital space rooted in trust and truth. Advocacy, therefore, must also form an essential part of fighting disinformation.

**The actors and tactics behind disinformation in sub-Saharan Africa**

This section examines the key actors fuelling electoral disinformation in the Sub-Saharan region and the tactics they employ to manipulate narratives, suppress dissent, and distort democratic processes. There remains a lack of comprehensive research in this area. Much of what is known is based on limited case studies or episodic reporting, often focused on high-profile incidents or external actors. This is exacerbated by the prevalence of closed information networks, such as encrypted messaging apps, offline peer-to-peer sharing, and local language forums, which make detection and analysis far more difficult.

Data access is another issue. Social media platforms seldom share data with misinformation researchers, leaving scientists with incomplete or biased datasets that may not reflect public sentiment. Due to limited API access, many researchers rely on data scraping, which usually violates website terms of service, though this is often justified because few alternatives exist for studying algorithmic impact. In addition to advocacy for better design decisions, advocacy for better data access is required. As expressed by misinformation researchers at Harvard University:[48]

[48] Pasquetto, I., Swire-Thompson, B., Amazeen, M.A., Benevenuto, F., Brashier, N.M., Bond, R.M., Bozarth, L.C.,

> *Misinformation thrives on social media because of emotion. False, emotional content is clicked on, diffuses widely and rapidly through social networks, and is often believed, particularly when it fits with one's political worldview. Yet, the degree to which emotion influences exposure to, engagement with, and belief in misinformation on social media remains shrouded by insufficient data from prominent platforms. What is needed is a more comprehensive picture of the emotional nature of misinformation in social media environments. Open data from social media platforms would help address critical, unanswered questions. Open data from social media platforms would also facilitate an understanding of how different emotions like anger and fear uniquely amplify misinformation and deepen misperceptions. More transparency and open data practices from social media platforms would illuminate the processes and mechanisms through which emotional misinformation is encountered, spread, and believed.*

Despite these limitations, existing investigations do offer valuable insights into the tactics, narratives, and occasional actors behind disinformation efforts. However, these accounts are fragmentary and should not be treated as authoritative but rather as instructive. Without such work, policymaking and civil society responses risk being reactive, incomplete, or misdirected. Addressing this gap is essential for developing effective, context-specific strategies to strengthen information integrity and democratic resilience in Africa.

***Tactics:***

In 2022, the Collaboration on International ICT Policy for East and Southern Africa (CIPESA) conducted a study of disinformation in five African countries - Cameroon, Ethiopia, Kenya,

Budak, C., Ecker, U.K.H., Fazio, L.K., Ferrara, E., Flanagin, A.J., Flammini, A., Freelon, D., Grinberg, N., Hertwig, R., Jamieson, K.H., Joseph, K., Jones, J.J., Yang, K.C. (2020). Tackling misinformation: What researchers could do with social media. Harvard Kennedy School (HKS) *Misinformation Review*, 1(8). https://par.nsf.gov/servlets/purl/10220319

Nigeria, and Uganda found some commonalities.[49] Although the study is now three years old, and disinformation tactics are constantly evolving, it remains beneficial and instructive.

CIPESA found that the most common disinformation tactics in the studied countries were:

- Astroturfing
- Mass brigading
- Mass sharing
- The Use of Fake and Pseudonymous Social Media Accounts

Astroturfing is a deceptive tactic used to fabricate the appearance of grassroots support for a cause, individual, or campaign.[50] It is coordinated inauthentic behaviour where hidden actors, ranging from political operatives and commercial disinformation firms to foreign governments, centrally coordinate these campaigns, preventing them from emerging organically. astroturfing is commonly employed in disinformation efforts to manipulate public opinion, create false consensus, or suggest legitimacy where none exists.

Uncovering the identities of individuals behind astroturfing campaigns is a complex process that typically requires the combined use of Open-Source Intelligence (OSINT) and digital forensics. While OSINT involves collecting and analysing publicly available information, such as posts, metadata, images, and account behaviours, digital forensics goes further, enabling investigators to verify content authenticity, trace digital footprints, and preserve evidence in a manner admissible for legal or regulatory proceedings.

Together, these methods allow investigators to move beyond surface-level content and begin attributing campaigns to specific individuals or networks. However, successful attribution

[49] Collaboration on International ICT Policy for East and Southern Africa (CIPESA). (2022). *Disinformation Pathways and Effects: Case Studies from Five African Countries – Cameroon, Ethiopia, Kenya, Nigeria and Uganda*. https://cipesa.org/wp-content/files/documents/Disinformation_Pathways_and_Effects_Africa_Presentation.pdf

[50] Chan, J. (2022). Online astroturfing: A problem beyond disinformation. Philosophy & Social Criticism, 50(3), 507-528. https://doi.org/10.1177/01914537221108467

remains rare, given the sophistication of disinformation actors, the anonymity afforded by social media platforms, and the technical expertise required to conduct such investigations. In the few instances where attribution has been possible, it has involved collaboration between digital forensic experts, investigative journalists, and cybersecurity analysts.

The following case study is one of the few instances where investigators were able to uncover the identities of those orchestrating a coordinated astroturfing campaign, providing invaluable knowledge about how such operations function and who profits from them.

**Case Study: Astroturfing in Nigeria's 2023 Election**[51]

In the lead-up to Nigeria's 2023 general elections, a BBC Africa Eye investigation exposed a sprawling and well-financed astroturfing operation designed to manipulate public opinion and distort the online political landscape. Political operatives working for major parties covertly hired popular influencers to push orchestrated political narratives while pretending to be ordinary citizens expressing genuine views. The investigation uncovered that influencers were paid up to ₦20 million (approximately USD 45,000) to promote false, polarising, and politically charged content across platforms such as Twitter, Facebook, and WhatsApp. These influencers coordinated the release of identical posts and hashtags, simulating mass support for specific parties or candidates, thereby creating a false sense of grassroots consensus. This campaign bore all the hallmarks of political astroturfing: centralised planning, fake grassroots energy, concealed sponsorship, and deliberate deception. The intent was not merely persuasion but manipulation, targeting ethnic, religious, and regional divisions to discredit opponents and shape voter perceptions. Through undercover reporting and open-source analysis, journalists identified digital consultants and social media managers openly selling influence packages to political clients.

Mass Brigading involves a group of users banding together to discredit another user expressing a different opinion.[52] The aim is to drown out the opposing view. It may be an instance of the Bandwagon Effect, a cognitive bias where individuals adopt beliefs or behaviours because they appear popular. The main effects of brigading are to harass and silence. Malign actors may use narrative hijacking, astroturfing, and sockpuppetry to mass brigade. Mass brigading, similarly, requires the use of OSINT and digital forensics.

**Case Study: Mass Brigading Against the #LindaKatiba Movement in Kenya**[53]

[51] Nwonwu, C., Tukur, F., & Oyedepo, Y. (2023, January 18). Nigeria elections 2023: How influencers are secretly paid by political parties. BBC News. https://www.bbc.com/news/world-africa-63719505

[52] Andrews, P.C.S., (2021). Social Media Futures: What is Brigading. *Tony Blair Institute for Global Change*. https://institute.global/insights/tech-and-digitalisation/social-media-futures-what-brigading

[53] Onyango, E. (2021, September 13). Kenyan influencers paid to take 'guerrilla warfare' online. *BBC News*. https://www.bbc.com/news/world-africa-58474936

During Kenya's 2021–2022 political period, civil society groups under the #LindaKatiba banner (which opposed proposed constitutional amendments) became frequent targets of coordinated mass brigading on Twitter. Activists, legal scholars, and opposition-aligned figures who publicly supported the campaign were repeatedly attacked by large numbers of accounts posting identical or similar responses in an effort to discredit them and drown out their messages.

Mozilla Foundation researchers uncovered that many of these brigading efforts were not spontaneous but driven by WhatsApp-coordinated influencer groups. These influencers received daily talk points, targeted accounts, and hashtags from anonymous organisers, often accompanied by mobile money payments. Posts by #LindaKatiba supporters were systematically dogpiled with negative replies, insults, misinformation, and counter-hashtags, designed to delegitimise the movement in the public eye.

What made this a textbook example of mass brigading was not just the volume but the tactical orchestration: the same influencers who promoted pro-government hashtags during the day were mobilised to attack dissident voices at night. The result was a chilling effect; some individuals reduced their engagement or deleted posts and deactivated their accounts, fearing reputational damage or the prospect of sustained trolling.

The Use of Fake and Pseudonym Accounts is a hallmark of disinformation campaigns and features in nearly every documented case. This is enabled by a core design flaw in most social media platforms: the ability to create accounts without meaningful identity verification. While anonymity can serve legitimate purposes, such as protecting whistleblowers and vulnerable activists, it has also been exploited on a large scale by malicious actors.

Anonymous accounts used in disinformation campaigns typically take the form of bots, sockpuppets, or coordinated account clusters managed by bot farms, content farms, or covert commercial disinformation firms. Increasingly, operators favour human-managed sockpuppet accounts over bots, making them harder to detect using automated bot-detection tools. This shift presents a growing challenge to platform integrity. It has led some to call for stricter identity verification mechanisms, although such proposals also raise concerns about privacy and free expression.[54]

Case Study: Anonymous Accounts and Disinformation in Ethiopia's Tigray Conflict[55]

[54] Media.com. (2024, April). *Study Highlights Overwhelming Support for Identity Verification to Combat Misinformation on Social Media Platforms*. https://finance.yahoo.com/news/study-highlights-overwhelming-support-identity-131800599.html

[55] Knight, T., (2021). Ethiopian diaspora organises social-media campaigns amid information scarcity. *African Digital Democracy Observatory (ADDO).* https://disinfo.africa/ethiopian-diaspora-groups-organize-click-to-tweet-tigray-campaigns-amid-information-scarcity-f7532e7b0b5b

During the 2020–2021 Tigray conflict, anonymous and fake social media accounts—primarily on Twitter and Facebook—were used to spread disinformation and amplify pro-government narratives. These accounts shared manipulated images, false reports, and ethnic hate speech targeted at Tigrayans while attacking international media and NGOs. Many used stolen profile pictures, vague bios, and coordinated hashtags (e.g., #NoMore) to disguise state-linked messaging as grassroots activism. In 2021, Meta removed a network of accounts tied to the Ethiopian government for coordinated inauthentic behaviour.

Disinformation for Hire: While not covered in the CIPESA report, a common tactic used in electoral disinformation in Sub-Saharan Africa is Disinformation-For-Hire, a shadow industry where influencers, including anonymous influencers and micro-influencers, are covertly recruited to amplify false or misleading narratives for political, financial, or social gain.

Disinformation has become an industrialised and is a booming industry.[56] In 2021, the New York Times found 65 companies across 48 countries to be meddling in elections to promote falsehoods on behalf of clients who often consist of governments, politicians, and political parties offering plausible deniability.[57] The names of the majority of the firms were withheld, highlighting how opaque this industry remains.

Disinformation-for- Hire providers are hired to run social media campaigns that spread false or misleading information, often focusing on elections, legislation, or political issues. These firms may be paid to promote or attack specific people, groups, or narratives, using tactics like coordinated impersonation, harassment, hashtag hijacking, as well as standard marketing or social media management practices.[58]

Since the 2017 Bell Pottinger scandal in South Africa, many firms have simply gone underground and are rarely uncovered. Instead, they tend to surface only through whistleblower leaks, long-term investigative reporting, or the painstaking work of research

[56] Lewandowski, A. (2021, December 8). Disinformation-for-hire: The pollution of news ecosystems and erosion of public trust. *Centre for International Media Assistance*. https://www.cima.ned.org/blog/disinformation-for-hire-the-pollution-of-news-ecosystems-and-erosion-of-public-trust/

[57] Thompson, S., & Frenkel, S. (2023, February 15). Disinformation for hire, a shadow industry, is quietly booming. *The New York Times*. https://www.nytimes.com/2021/07/25/world/europe/disinformation-social-media.html

[58] Institute for Strategic Dialogue (ISD). (2023). *Commercial Disinformation*. https://www.isdglobal.org/explainers/commercial-disinformation-product-service/

coalitions. Revelations such as those involving Cambridge Analytica, Team Jorge, and others emerged only after years of evidence gathering, confidential source protection, and cross-border collaboration. Such investigations are resource-intensive and require multidisciplinary teams that blend OSINT, digital forensics, data science, legal analysis, cybersecurity expertise, and investigative journalism.

These efforts are not only technically complex but also expensive, time-consuming, and often dangerous, especially when they target powerful political or corporate interests. In under-resourced media environments, common across much of Sub-Saharan Africa, sustained scrutiny of commercial disinformation actors exceedingly rare. As a result, many of these firms continue to operate in the shadows, adapting quickly to new detection methods and exploiting the global demand for covert influence operations.

**Case Study: Disinformation-for-Hire in South Africa**[59]

In the lead-up to South Africa's 2024 general elections, investigative reporting revealed a concerning trend: political parties and interest groups covertly hiring nano- and micro-influencers to manufacture online support and suppress dissent. These influencers, typically ordinary social media users with a few thousand followers, are perceived as more authentic and are thus more effective at swaying public opinion.

The investigation documented how influencers were paid modest sums (R50–R250) to post coordinated, undisclosed political content in favour of figures such as Deputy President Paul Mashatile, Action SA, and embattled ANC leaders. These campaigns, often appearing spontaneous, mimic organic political enthusiasm while concealing their transactional nature. The influencers themselves acknowledged the deceptive nature of these gigs, which they failed to disclose as sponsored content.

**The actors:**

***Domestic***

Identifying the actors behind disinformation in Africa remains a complex and opaque undertaking. Disinformation campaigns are, by design, elusive, a so-called "dark art." Therefore, the examples below likely only begin to scratch the surface. Intensive investigation is required across countries to reveal the full depth and breadth of the problem.

59 Davis, R., (2025, May 6). Why the rise of political nano-influencers should concern us. *Daily Maverick.* https://www.dailymaverick.co.za/article/2025-05-06-why-the-rise-of-political-nano-influencers-should-concern-us/

In 2024, the Africa Centre for Strategic Studies (ACSS) found 50 documented disinformation campaigns across the continent, with roughly 40 percent domestically driven. Their report remains the most comprehensive public information of disinformation activity across the African continent. However, its findings should be approached with measured caution. ACSS, while an academic institution, operates within the U.S. Department of Defence, an affiliation that may shape the scope or framing of its analysis. Still, the report offers a valuable, if partial, window into the scale and complexity of the problem.

Other reports provide some insight into the domestic actors behind disinformation campaigns.

A 2025 study[60] of electoral disinformation in the Democratic Republic of Congo (DRC), Kenya, Senegal, and South Africa found that while a diverse set of actors drove the spread of electoral disinformation, each playing a distinct role in shaping the information landscape, political figures and campaign strategists have been central to the dissemination of misleading narratives, often leveraging digital militias and paid influencers to amplify their reach.

DRFLab's Tessa Knight echoed this sentiment to the Africa Centre for Strategic Studies when discussing the investigation of actors behind disinformation on the African continent.

> *There were a variety of actors behind the examples of African disinformation I investigated. No two cases were the same, but most of these examples were ultimately connected to domestic governments or political parties.*

A book[61] investigating disinformation operations in ten African countries - Zimbabwe, Mozambique, Ethiopia, Egypt, the Democratic Republic of Congo, Cameroon, Uganda, Angola, Kenya, and Nigeria - highlights how electoral disinformation in those countries typically

[60] Madrid-Morales, D., Wasserman, H., Davies-Laubscher, N., Sow, F., (2025, May). Tackling Disinformation in Four African Elections. *Centre for Information Integrity in Africa*. https://ciia.africa/tackling-disinformation-in-four-african-elections/

[61] Roberts, T., & Hamandishe Karekwaivanane, G. (2024). *Digital Disinformation in Africa: Hashtag Politics, Power and Propaganda*. Zed Books. https://doi.org/10.5040/9781350319240.

involved governments coordinating disinformation campaigns to divert the opposition from participating in digital democracy and to close online civic space in order to promote their own power interests.

The growing, "homegrown" nature of electoral disinformation requires a critical shift from focusing solely on external threats to prioritising domestic accountability within the contexts where disinformation campaigns are deployed. Effective responses must be grounded in local political and media realities, including stronger regulation of political messaging, greater transparency in the use of influencers, and sustained investment in digital literacy.

Acknowledging the domestic roots of disinformation is key to designing context-specific, resilient countermeasures.

***FIMI***

Foreign Information Manipulation and Interference (FIMI) refers to covert or deceptive attempts by foreign actors, including state or state-linked entities, to disrupt electoral integrity, sow division, or shift geopolitical alignments.

FIMI campaigns may not always originate from outside the continent but often piggyback on existing domestic tensions and utilise local proxies to obfuscate their origin. For example, South African researchers[62] have suggested adopting the term (F)IMI - Foreign (and) Information Manipulation and Interference - to reflect the country's unique dynamics better. Unlike traditional definitions of FIMI that emphasise foreign actors as the primary drivers of manipulation, this framing recognises that much of the interference in South Africa originates from domestic actors. These local players often build and sustain the infrastructure through which influence is exerted. In contrast, foreign entities typically exploit or amplify these existing systems and narratives rather than establishing new ones. This perspective challenges

[62] Van Damme, P., Findlay, K., Cornelissen, A. (2024, December). Generative AI and its influence on South Africa's 2024 elections. *German Council for Foreign Relations*. https://dgap.org/en/research/publications/generative-ai-and-its-influence-south-africas-2024-elections

the binary distinction between foreign and domestic interference, highlighting instead their intertwined and mutually reinforcing roles.

Nevertheless, "traditional" FIMI remains prevalent across the continent and region. The above-mentioned ACSS found that FIMI and FIMI-adjacent disinformation campaigns documented in 2022 had nearly quadrupled to 189 across 39 African countries, a figure likely underreported.[63] Of these, 60 percent were foreign state-sponsored, with Russia, China, the United Arab Emirates (UAE), Saudi Arabia, and Qatar as the primary sponsors.

***The ACSS's findings were chilling:***

- Disinformation campaigns have directly driven deadly violence, promoted and validated military coups, cowed civil society members into silence, and served as smokescreens for corruption and exploitation.
- One or more coordinated disinformation campaigns have directly targeted 39 African countries.
- Twenty countries have experienced three or more campaigns, up from just seven in 2022.
- Countries experiencing conflict face a median of five disinformation campaigns, far more than stable states, indicating disinformation is both a driver and amplifier of instability.
- Russia leads in external influence, accounting for 80 campaigns across 22 African countries, representing over 40% of all foreign-backed disinformation on the continent.
- Disinformation campaigns often aim to erode trust in elections, discredit opposition movements, and normalise authoritarian governance.
- Twice as many disinformation campaigns target African countries without presidential term limits as those with term limits.

[63] Africa Centre for Strategic Studies. (2024, March) *Mapping a Surge of Disinformation in Africa.* https://africacenter.org/spotlight/mapping-a-surge-of-disinformation-in-africa/

- Foreign disinformation actors frequently collaborate with local influencers, PR firms, and media proxies to boost credibility and local resonance.
- Disinformation is cheap, scalable, and effective. Many campaigns rely on low-cost tactics like WhatsApp forwards, fake pages, meme warfare, and social bots, allowing for broad reach with minimal resources.
- Platforms remain unaccountable and continue to underinvest in moderation, especially in African languages and local contexts, making the region vulnerable to unchecked manipulation.
- The spread of disinformation is contributing to erosion in democratic accountability, polarisation, and a growing crisis of trust in both media and electoral systems.

The latter points emphasise the fact that disinformation is not created in a vacuum and is worth bearing in mind.[64]

> *[FIMI disinformation] resonates more if it builds on existing grievances within society. Inequality, intolerance, distrust (towards both fellow citizens and institutions), and discrimination can pave the way for societal grievances that create a fertile ground for FIMI operations. In polarised political environments, individuals have more difficulty distinguishing between false and accurate information. Exploiting societal divisions and an inclination to embrace information that supports biases against certain groups is thus key for disinformation to gain traction and find a receptive audience. In countries where opposed groups are in conflict, malign actors can thus more easily spread false or distorted narratives that exploit societal tensions.*

The ACSS and other cross-country reports do not give an extensive rundown of the states and actors involved in spreading FIMI in Sub-Saharan Africa. However, below are notable commercial firms and governments that have been linked to disinformation operations in Sub-Saharan Africa, based on credible investigations, repeated study, and reporting.

---

[64] Terren, L., Van Aelst, P., & Van Damme, T. (2023, November 24). The last line of defence: Measuring resilience to foreign information manipulation and interference in West Africa. *European Union of Security Studies.* https://www.iss.europa.eu/publications/briefs/last-line-defence-measuring-resilience-foreign-information-manipulation-and

## Foreign Governments & State-Linked Actors

| Actor | Activity | Countries Targeted | Example |
|---|---|---|---|
| Russia (incl. Wagner Group / Prigozhin networks) | Russia is the single largest and most well-known sponsor of Africa-wide disinformation campaigns. Its FIMI operations typically involve a coordinated mix of disinformation, electoral interference, and support for extraconstitutional claims to power. These tactics are often mutually reinforcing; disinformation campaigns are frequently deployed alongside efforts to influence elections, either to entrench Moscow-aligned regimes, justify unconstitutional term extensions, or legitimise military coups. Russia's interference thus systematically undermines democratic norms while bolstering authoritarian allies across the continent. [65] | 19 countries, including Central African Republic, Mali, Burkina Faso, Sudan, Madagascar, South Africa. | Central African Republic (CAR) journalist Ephrem Yalike exposed the inner workings of Russian disinformation campaigns in CAR, operations in which he was directly involved. These efforts were coordinated through Africa Politology, a covert organisation tied to the "Prigozhin galaxy," which recruited local journalists to shape pro-Russian narratives and manipulate public opinion. One of the key figures orchestrating these campaigns was Mikhaïl Mikhaïlovitch Prudnikov, a Wagner Group associate who had previously conducted similar influence operations in Sudan before turning his efforts to the Central African Republic.[66] |
| China (via state media and proxies) | China is the second most prolific documented Africa-wide sponsor of disinformation, with five known multi-regional campaigns. Beijing's strategy in Africa focuses on narrative control rather than overt disinformation. Through state media partnerships, journalist training programmes, and digital diplomacy, China promotes a consistent narrative that portrays it as a benevolent development partner and model for governance, | Multiple across Africa, notably Kenya, Nigeria, Zimbabwe | China's state-run broadcaster CGTN Africa, headquartered in Nairobi, serves as a key hub for shaping pro-China narratives across the continent. It produces content that promotes China's development model, Belt and Road projects, and diplomatic ties, while avoiding or downplaying sensitive issues like debt, surveillance technology, or human rights abuses. Through content-sharing agreements, Chinese state media outlets like |

[65] Africa Centre for Strategic Studies. (2024). *Tracking Russian Influence on Derail Democracy in Africa.* https://africacenter.org/spotlight/russia-interference-undermine-democracy-africa/

[66] Peruchon, L. (2024, November 21). In the Central African Republic, a former propagandist lifts the veil on the inner workings of Russian disinformation. *Forbidden Stories*. https://forbiddenstories.org/in-the-central-african-republic-a-former-propagandist-lifts-the-veil-on-the-inner-workings-of-russian-disinformation/

| | | | |
|---|---|---|---|
| | while discrediting Western actors as neocolonial or destabilising forces.[67] | | Xinhua and CGTN provide free or subsidised news content to African broadcasters and newspapers. These arrangements often lead to the uncritical republishing of Chinese narratives, giving China outsized influence over how global events and China's role in Africa are framed.[68] |
| Iran (IRGC-linked) | The Islamic Revolutionary Guard Corps (IRGC), a powerful branch of Iran's military, plays a central role in spreading state-aligned disinformation both domestically and abroad. Through front media organisations, fake social media accounts, and proxy networks, the IRGC amplifies narratives that promote Iranian geopolitical interests, attack rivals (especially the U.S., Israel, and Saudi Arabia), and fuel sectarian or anti-Western sentiment. In parts of Africa, this includes targeting Shia communities with propaganda that frames Iran as a defender of oppressed Muslims while discrediting opposing factions and governments.[69] | Nigeria (especially targeting Shia-Sunni divides) | Former members of the Islamic Movement in Nigeria (IMN) ran a covert Facebook operation under the hashtag #ZakzakyLifeMatters, promoting pro-IMN narratives and political messaging. The pages and groups amplified content supporting the movement's leader, Sheikh Ibrahim Zakzaky, while targeting critics and government institutions. The campaign used emotionally resonant messaging and coordinated posts to generate engagement and shape public perception. Investigators traced the operation back to IMN affiliates, revealing an organised attempt to leverage Facebook as a mobilising and persuasive tool.[70] |

Table 9: FIMI operations in Africa

[67] Cissé, D., & Pihl, M. (2025, March 25). China's narrative warfare in Africa: Influence and mechanisms. *China Observers in Central and Eastern Europe*. https://chinaobservers.eu/chinas-narrative-warfare-in-africa-influence-and-mechanisms/

[68] Freedom House. (2022). *Beijing's Media Influence: Kenya*. https://freedomhouse.org/country/kenya/beijings-global-media-influence/2022

[69] Hassaniyan, A. (2022, November 1). What a longstanding Iranian disinformation tactics target protest. *The Washington Institute for Near East Policy*. https://www.washingtoninstitute.org/policy-analysis/how-longstanding-iranian-disinformation-tactics-target-protests

[70] Grossman, S., Gallagher, S., Johnson-Kanu, A., & Wilson, N. (2020, October 8). #ZakzakyLifeMatters: An investigation into a Facebook operation linked to the Islamic Movement in Nigeria. *Stanford Internet Observatory.* https://stacks.stanford.edu/file/druid:vk551rc5348/facebook-NG-202009.pdf

## Commercial Disinformation/Disinformation for Hire

| Organisation | Activity | Countries Operational | Example |
|---|---|---|---|
| Team Jorge | The Israel-based Team Jorge used hacking, fake avatars, and bot networks to run covert disinformation campaigns—spreading false narratives, infiltrating messaging groups, and manipulating elections in dozens of countries, including in Africa. Their tactics included media forgeries, social engineering, and seeding fake stories with real journalists.[71] Still operational. | Multiple African elections, including in Kenya and Nigeria (alleged) | In the lead-up to Kenya's 2022 presidential election, Team Jorge allegedly hacked the emails and messaging accounts of senior campaign aides to candidate Raila Odinga. The operation aimed to gather sensitive political intelligence that could be exploited to benefit rival interests. Investigators identified 25 attempts to access Gmail and Telegram accounts, including that of Odinga's chief legal adviser. Though it's unclear who hired Team Jorge, the hacks form part of a broader pattern of covert digital manipulation linked to elections across Africa and beyond.[72] |
| Cambridge Analytica | Cambridge Analytica used psychographic profiling and microtargeted political ads, often based on improperly harvested Facebook data, to influence voter behaviour, including in Kenya and Nigeria. Its tactics included dark ads, emotionally charged disinformation, and narrative manipulation tailored to specific audiences. Although the firm shut down in 2018, its methods live on through rebranded entities and similar influence-for-hire operations.[73] | Nigeria and Kenya. | In Nigeria's 2015 presidential election, Cambridge Analytica was hired by allies of then-President Goodluck Jonathan to run a covert campaign against his opponent, Muhammadu Buhari. The firm used stolen personal data, including medical records, and created fear-based messaging, including a doctored video aimed at portraying Buhari as sympathetic to Islamic extremism. Their goal was to suppress voter turnout and discredit Buhari among key demographics. The campaign was conducted secretly from London, without formal disclosure or local oversight.[74] |

[71] The Guardian. (2023, February). *Revealed: The Hacking and Disinformation Team Meddling in Elections*. https://www.theguardian.com/world/2023/feb/15/revealed-disinformation-team-jorge-claim-meddling-elections-tal-hanan

[72] The Guardian. (2023, February). *Political Aides Hacked by 'Team Jorge' in Run-Up to Kenyan Election.* https://www.theguardian.com/world/2023/feb/15/political-aides-hacked-by-team-jorge-in-run-up-to-kenyan-election

[73] The Guardian. (2018, March 17). *Revealed: 50 million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach.* https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[74] The Guardian. 2023, February 16) *Dark Arts Of Politics: How 'Team Jorge' And Cambridge Analytica Meddled In Nigerian Election* https://www.theguardian.com/world/2023/feb/16/team-jorge-and-cambridge-analytica-meddled-in-nigeria-election-emails-reveal

| | | | |
|---|---|---|---|
| Internet Research Agency (IRA) and Association for Free Research and International Cooperation (AFRIC) | Both were linked to Yevgeny Prigozhin, a close ally of Vladimir Putin and the financier behind several Russian influence operations, including the Wagner Group.[75]<br><br>The IRA posed as a private company but operated as a covert arm of the Kremlin's information warfare strategy and ran large-scale influence campaigns using troll farms, fake social media personas, and coordinated messaging to sow division, distort public debate, and interfere in elections globally. Though not officially part of the Russian government, it functions as a state proxy, offering the Kremlin plausible deniability while advancing its geopolitical objectives.<br>AFRIC posed as an election observation group and was used to legitimise authoritarian regimes, manipulate perceptions of electoral integrity, and spread disinformation through official-looking reports. | CAR, Zimbabwe, Mozambique, Madagascar, Sudan, DRC | In Zimbabwe's 2018 general election, AFRIC sent so-called election observers who falsely endorsed the election as free and fair, despite widespread reports of irregularities, voter intimidation, and a violent crackdown during the delayed results announcement. AFRIC's role was to legitimise the election outcome, reinforce pro-government narratives, and counter criticism from credible local and international observers. Its presence formed part of a broader Russian strategy to support authoritarian regimes and gain strategic influence in Southern Africa.[76] |
| Archimedes Group | The Archimedes Group is a Tel Aviv-based private intelligence agency that has run political campaigns using social media since 2017. It has been exposed for coordinated disinformation campaigns using fake social media accounts, manipulated images, deceptive ads, and impersonated local political actors to spread propaganda and influence elections, primarily through Facebook. | Nigeria, Senegal, Togo, Angola, Niger and Tunisia | In 2019, Facebook shut down 265 fake accounts run by Archimedes engaging in coordinated inauthentic behaviour around elections in Africa, Latin America, and Southeast Asian countries. The accounts had been posting on behalf of certain political candidates, smearing their opponents, and presenting themselves as local news organisations peddling supposedly leaked information.[77] |

[75] Rampe, W. (2023, May 23). What is Russia's Wagner Group doing in Africa? *Council on Foreign Relations*. https://www.cfr.org/in-brief/what-russias-wagner-group-doing-africa

[76] Shekhovtsov, A. (2020, November). Fake election observation as Russia's tool of election interference: The case of AFRIC. *European Platform for Democratic Elections*. https://epde.org/?news=fake-election-observation-as-russias-tool-of-election-interference-the-case-of-afric-2599

[77] PBS News (2019, May). *Facebook Busts Israeli Company's Campaign to Disrupt Elections.* https://www.pbs.org/newshour/world/facebook-busts-israeli-companys-campaign-to-disrupt-elections

| | | | |
|---|---|---|---|
| Bell Pottinger | The now-defunct UK-based Bell Pottinger used narrative manipulation, fake social media accounts, astroturfing, information laundering, and racially divisive messaging to shape public opinion, discredit critics, and protect client interests. | South Africa, Egypt, Nigeria, Guinea, and the Democratic Republic of Congo. | Hired by the Gupta family, close associates of South Africa's former President, Jacob Zuma, Bell Pottinger orchestrated a disinformation campaign around the term "white monopoly capital" to deflect attention from state capture allegations, inflame racial tensions, and smear journalists and civil society—ultimately leading to the firm's collapse in 2017.[78] |
| UReputation | UReputation is a Tunisia-based digital communications and reputation management firm founded by Lotfi Bel Hadj. While it presents itself as offering online PR and lobbying services, investigations have shown that the company has engaged in covert influence operations across Francophone Africa. These operations include spreading disinformation, running fake social media networks, and manipulating online narratives to support political interests, particularly during elections.[79] | Madagascar, Tunisia, Côte d'Ivoire, Central African Republic | In 2020, Facebook removed 446 pages, and 96 groups linked to UReputation for attempting to manipulate political discourse in Madagascar during its electoral cycle.[80] |

Table 10: Commercial FIMI Operations in Africa

[78] DW. (2017, May). *PR Firm Inflamed Racial Discord in South Africa*. https://www.dw.com/en/bell-pottinger-in-south-africa-pr-firm-inflamed-racial-discord/a-40362110

[79] DRFLab. (2020). *Operation Carthage: How A Tunisian Company Conducted Influence Operations In African Presidential Elections* https://www.atlanticcouncil.org/in-depth-research-reports/operation-carthage-how-a-tunisian-company-conducted-influence-operations-in-african-presidential-elections/

[80] Business & Human Rights Resource Centre. (2020). *Facebook Statement on UReputation*. https://www.business-humanrights.org/en/latest-news/facebook-statement-on-ureputation-may-2020-coordinated-inauthentic-behavior-report/

## PART 2: The HOW to

### Countering the Chaos: Strategies to Combat Electoral Disinformation

The picture is grim. The scale, speed, and sophistication of disinformation across Sub-Saharan Africa is deepening distrust, undermining democratic institutions, and enabling authoritarian creep. However, even in the face of this growing storm, there is cause for resolve. Across the region, civil society, journalists, researchers, and communities are beginning to chip away at the machinery of manipulation, one strategy, one intervention, and one informed citizen at a time.

Part Two of this guide outlines practical steps for doing just that: how to build systems, teams, and tools that can disrupt disinformation, restore trust, and help reclaim our information space.

**Setting up a disinformation response team**

Combating disinformation in electoral contexts demands a multidisciplinary approach, requiring expertise in data science, communications, political science, behavioural psychology, gender studies, and others to effectively counter false narratives and their societal impacts.

A Disinformation Response Team can be established in two ways: developing an in-house team with these specialised skills or forming coalitions with established organisations in the field. The latter is preferable, as it leverages existing expertise, fosters knowledge-sharing, and enhances collective impact in addressing the human-driven and platform-amplified spread of disinformation.

### In-house team

| Skill | Description | Why Needed |
|---|---|---|
| Digital Forensics and Open-Source Intelligence (OSINT) | Expertise in monitoring disinformation sources through platforms such as TweetDeck or CrowdTangle, scrutinising digital traces like metadata and bot detection, and pinpointing instances of coordinated inauthentic behaviour (CIB). | To pinpoint domestic actors (e.g., political operatives, influencers) spreading false narratives, |
| Behavioural Science and Messaging | Expertise in crafting and pre- and debunking messages using behavioural science principles (e.g., accuracy nudges, trusted voices) to counter cognitive biases like confirmation bias and emotional reasoning. | To shift user behaviour and reduce misinformation sharing |
| Media Literacy and Community Engagement | Ability to design and deliver community-based media literacy programs, such as workshops, using trusted local voices to teach source verification and critical thinking. | To empower voters in low-internet-access regions to counter disinformation. |
| Data Analysis and Social Media Monitoring | Proficiency in analysing social media trends and engagement metrics using tools like Python or platform analytics to identify and assess viral falsehoods. | To track disinformation in real time. |
| Cultural and Contextual Knowledge | A deep understanding of local political, social, and cultural dynamics is necessary to tailor responses to specific disinformation campaigns, especially those related to domestic rivalries or ethnic tensions. | To address region-specific falsehoods |
| Communication and Public Relations | Expertise in crafting clear, accessible messages for diverse audiences (e.g., via radio, SMS, social media) and coordinating with media to amplify corrections. | To ensure corrections reach the same audience as disinformation. |
| Policy and Legal Awareness | Knowledge of local and international disinformation laws (e.g., Nigeria's Cybercrimes Act) to navigate regulatory constraints and advocate for rights-respecting policies. | We must adhere to legal frameworks and strive for platform accountability. |
| Gender Rights Specialist | Expertise in analysing how TFGBV targets women and marginalised groups (e.g., based on gender, ethnicity, or sexual orientation) to identify gendered disinformation patterns. | To address TFGBV's disproportionate impact on women in public roles (e.g., journalists, activists) in African elections |

*Table 11: Disinformation Response Team*

### Prospective Partner Organisations for Coalitions: OSINT, digital forensics, disinformation investigations

#### *Code for Africa*

It is recognised as Africa's largest civic tech and data journalism initiative, with operations spanning over 20 countries. The initiative also houses the iLAB, a digital forensic and investigation lab that tracks influence operations, platform manipulation, and election-related disinformation.

***DFRLab***

While global, the Digital Forensic Research Lab has partnered with African organisations to train local researchers and map disinformation trends in elections across Nigeria, Kenya, and Ghana.

***Institute for Strategic Dialogue (ISD)***

A global think tank with growing work across Africa, ISD uses open-source investigation, social media monitoring, and digital forensics to track disinformation, extremism, and influence operations. In Africa, ISD has partnered with civil society to expose cross-platform disinformation campaigns, assess electoral risks, and build capacity for digital investigations in high-stakes contexts.

***Centre for Information Resilience (CIR)***

A UK-based nonprofit working globally to expose disinformation, human rights abuses, and influence operations through open-source intelligence (OSINT), social media analysis, and digital forensics. In Africa, CIR has collaborated with local partners to monitor electoral disinformation, map cross-platform manipulation, and build investigative capacity for civil society and independent media.

***Centre for Information Integrity in Africa (CIIA)***

Research and capacity-building centre focused on combating disinformation and safeguarding information ecosystems, based at the University of Stellenbosch, South Africa

***Centre for Analytics and Behavioural Change (CABC)***

A nonprofit based at the University of Cape Town Graduate School of Business, aiming to use advanced technology in the social media space to deliver social justice outcomes. The CABC is composed of specialised skills and technologies devoted to (a) identifying, analysing, reporting on, and countering harmful social media narratives; (b) building rigorous academic knowledge to help organisations around the world counter misinformation; and (c) using the power of social media to address social prejudice at scale.

**DEEPFAKES & AI POLICY**

***WITNESS***

Equips civil society, journalists, and human rights defenders with the tools to detect, verify, and respond to synthetic media. Leads global work on ethical approaches to deepfakes and has supported African partners with resources and training on media authentication and response strategies during elections and crises.

**DIGITAL RIGHTS & POLICY ADVOCACY**

***Paradigm Initiative***

Pan-African digital rights group advancing internet freedom, data protection, and online civic participation, with offices in Nigeria, Senegal, Zambia, and beyond.

***Collaboration on International ICT Policy for East and Southern Africa (CIPESA)***

Based in Uganda, CIPESA research and advocates for inclusive, rights-based ICT policies, having produced critical work on disinformation laws and platform regulation.

***Research ICT Africa (RIA)***

RIA, located in South Africa, carries out research that involves different fields and aims to benefit the public by studying the digital economy and society in Africa. They focus on digital governance, policy, and regulation to improve how people access, use, and apply digital technologies for social and economic growth in Africa.

***Access Now***

A global group with an Africa-specific programme. Provides support to civil society on internet shutdowns, content regulation, and surveillance threats, including those related to elections.

**FACT-CHECKING**

***Africa Check***

Africa's leading independent fact-checking organisation has offices in South Africa, Nigeria, Kenya, and Senegal. Works to verify public claims, debunk misinformation, and build a culture of evidence-based discourse.

***PesaCheck***

East Africa's largest fact-checking initiative, verifying public statements, media content, and viral claims in countries like Kenya, Tanzania, and Uganda.

***Dubawa***

A verification and media literacy platform led by the Premium Times Centre for Investigative Journalism (PTCIJ) in Nigeria and Ghana, aimed at fighting mis/disinformation through fact-checking and research.

***FactSpace West Africa***

A Ghana-based organisation combats misinformation in West Africa by implementing initiatives for fact-checking, training, and media monitoring.

***Media Monitoring Africa (MMA)***

This South African NGO focuses on media ethics, digital literacy, and children's digital rights. Runs programmes like Real411 to track digital harms, including disinformation and hate speech.

**COLLECTING DIGITAL ECOSYSTEM DATA**

In order to contextualise the work of the Disinformation Response Team, it is essential to gather statistics about the digital and social media ecosystems of each country. Understanding

the demographics, locations, and quantities of users is essential for grounding the work of the team in local realities.

The type of statistical data would include:

***Digital Access and Connectivity:***

Understanding who is online, how they access the internet, and where gaps exist is crucial.

- Internet penetration rate (overall and by region/gender/and age)
- Mobile phone ownership (smartphone vs. feature phone)
- Data affordability (cost per GB)
- Urban vs. rural access differentials
- Access to electricity/digital infrastructure

Sources: ITU, GSMA, World Bank, national telecom regulators

***Platform and Media Use***

Identifying which platforms and media channels dominate public discourse.

- Top social media platforms by user base (e.g., Facebook, WhatsApp, TikTok, X)
- Messaging app usage rates
- TV, radio, and print media consumption stats
- Preferred news sources (local vs. foreign; formal vs. informal)
- Language use on digital platforms

Sources: DataReportal, Afrobarometer, Pew Research, platform transparency reports

***Trust and Information Behaviours***

Assessing how people engage with information, who they trust, and why.

- Levels of trust in media, government, and social media
- Digital literacy rates
- Perceived threat of disinformation
- Who do people turn to for "truth" (religious leaders, family, influencers, etc.)?

Sources: Afrobarometer, national surveys, media research organisations

***Regulatory and Political Environment***

Contextualising the digital space within governance and law.

- Existence of cybercrime or anti-disinformation laws
- Press freedom rankings
- Censorship or surveillance practices
- History of internet shutdowns
- Regulatory independence of media/ICT authorities

Sources: Freedom House, CIPESA, Reporters Without Borders, Access Now

***Disinformation Landscape***

Mapping the risks, actors, and patterns of manipulation.

- Known disinformation incidents (local and foreign led)
- Actors involved (e.g., political parties, influencers, foreign states)
- Platforms used for disinformation
- Election-related disinformation trends
- Existence of fact-checking or monitoring organisations

Sources: Centre For Strategic Studies, fact-checkers (e.g. Africa Check, PesaCheck), civil society reports.

# USING SOCIAL MEDIA ANALYTICS TO GENERATE DATA ABOUT DISINFORMATION AND COMBAT IT

## What is SMA?

The gathering and analysis of social media data is crucial for understanding the how, why, and when of disinformation on social media.

Social Media Analytics (SMA) serves as a critical information-gathering tool that not only aids in providing evidence-based data for crafting combating strategies but also may offer guidance regarding how to implement these strategies effectively.

SMA is a highly complex process with different approaches for specific websites and various collection methods. As an original and passive data source, social media data poses several analytical challenges, such as identifying target information from unstructured and unprompted data, quantifying highly qualitative visual and textual messages, and ensuring representative data for an examined topic.

To standardise the process, Stieglitz et al.[81] proposed a four-stage process. This is the most widely accepted method of SMA, particularly in the political context.

[81] Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social Media Analytics – Challenges in Topic Discovery, Data Collection, And Data Preparation. *International Journal of Information Management, 39*, 156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

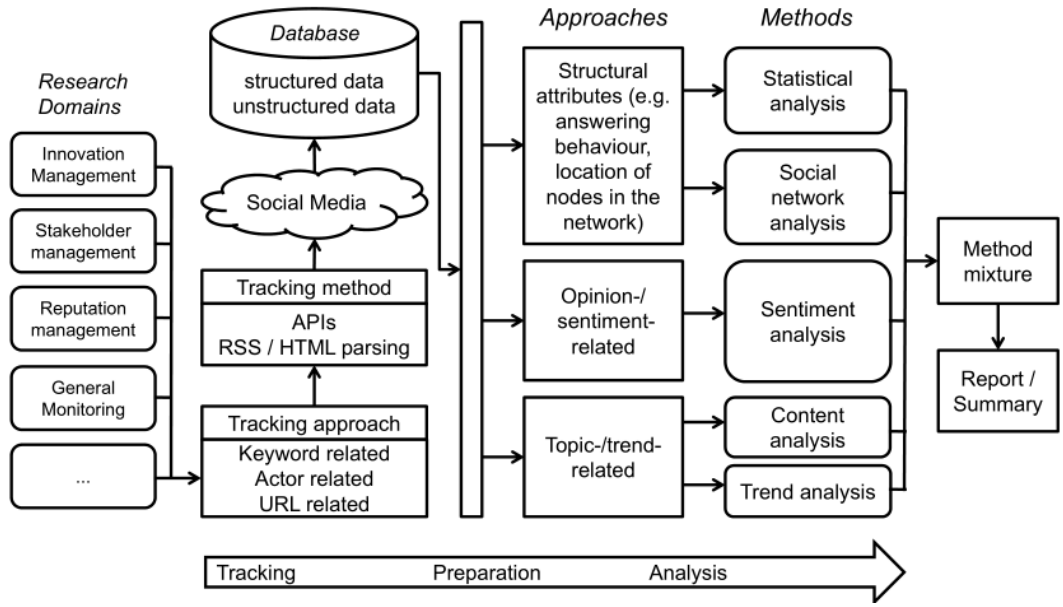


Figure 1 - Stieglitz et al., The Social Media Analytics Framework

**The four steps are:**

- **Discovery**: identification of content and corresponding keywords, hashtags, images, etc., used when discussing a specific topic(s) that will provide input into framing the analysis objectives and the primary hypotheses to be tested.
- **Tracking**: determination of the data sources, collection approach, and output of data on the discovered topic.
- **Preparation**: the approaches (e.g., natural language processing, topic categorisations to prepare the source data for subsequent analysis.
- **Analysis**: conducting various analysis methods/techniques on the prepared data set that aim to answer the questions outlined in the discovery phase

Conducting this work is the forte of data science, the multidisciplinary grouping of mathematics, statistics, probability, computing, and data visualisation. Data scientists build, clean, and structure datasets to analyse and extract meaning. However, commercial tools like Brandwatch, Talkwalker, Meltwater, and others can generate SMA if provided the correct information about what kind of data is required.

SMA can also generate different types of insights: descriptive, diagnostic, predictive, and prescriptive.

| Type | Description | Benefits | Limitations |
|---|---|---|---|
| Descriptive Analytics | What happened? | Can help identify patterns and trends related to the spread of false information and the tactics used to manipulate online narratives. | Provides a summary of historical data, which may limit the depth of insights gained. It is, therefore, reactive, as it is focused on past events and trends rather than offering proactive insights or predictions for future actions or outcomes. |
| Diagnostic Analytics | Why did this happen? | Diagnostic analytics can provide insights into the root causes and factors contributing to the spread of false information. Examining metrics and engagement patterns helps identify sources, influencers, and strategies employed. | While it provides insights into what factors contributed to success or failure, it relies on historical data, which may not provide real-time insights into ongoing social media activities. It is reactive. |
| Predictive Analytics | What might happen in the future? | Using statistical algorithms and machine learning techniques, it identifies the likelihood of future outcomes based on historical data. By analysing historical data and monitoring online activities, predictive models can identify emerging trends, anticipate future misinformation campaigns, and highlight potential targets. | Relies on historical data and not real-time trends to forecast future trends and behaviours. Factors such as evolving user preferences, changing social media algorithms, or unforeseen events can impact the reliability of predictions. Proactive. |
| Prescriptive Analytics | What should we do next? | Offers possible recommendations and strategies. By leveraging insights from descriptive, diagnostic, and predictive analytics, prescriptive analytics can guide the development of targeted interventions, countermeasures, and policies to mitigate the impact of disinformation and promote digital freedom. | A highly complex form of analytics that requires large datasets and the application of data science modelling. While it cannot replace human analysis, research, or research from relevant academic disciplines, it provides valuable information to support data-driven, forward-looking decision-making. Proactive. |

*Table 12: SMA Types*

SMA does have its limitations. It draws from data that users post to social media, thus providing insights based on public social media posts. These insights may represent the views of only a fraction of the population.

## An SMA Strategy for Electoral Disinformation: A Phased Electoral Approach

| PRE-ELECTION PHASE<br>Date: 3–6 months before election day<br>Goal: Establish baselines, detect slow-burn narratives, and anticipate disinformation tactics | |
|---|---|
| Task | Rationale |
| Historical Data Collection and Analysis | • Examine past election cycles and recent political discourse to identify narrative trends (e.g., scapegoating, election denialism).<br>• Use social media digital forensics to investigate existing disinformation networks<br>• Community Mapping and Data Visualisations |
| Platform Prioritisation | Based on findings from the digital ecosystem analysis, the prioritised platforms for data collection include Facebook (e.g., for rural messaging), TikTok (for youth narratives), and WhatsApp (for encrypted messages and disinformation). |
| Sentiment & Narrative Tracking | Use social media listening to identify:<br>• Emergent conspiracy theories<br>• Identity-based disinformation<br>• Voter suppression rhetoric<br>• Influence Operations |
| Network Mapping & Influence Tracking | Begin tracing high-engagement accounts (including anonymous influencers or suspected influencers-for-hire) that shape political narratives. Track platform manipulation tactics such as astroturfing, keyword squatting, or sock puppetry. |
| Weekly Reports (All Four Analytics Types) | • Descriptive: Track volume of disinformation mentions by theme<br>• Diagnostic: Identify tactics and actors behind surges<br>• Predictive: Model potential flashpoints (e.g., party registration deadlines)<br>• Prescriptive: Recommend prebunking strategies or platform escalation |
| Public-Facing Pre- and Debunking | • Identity Narratives for Pre- and Debunking |

| ELECTION WEEK PHASE<br>Date: 7–10 days before and after election day<br>Goal: Detect and respond to disinformation in real time; coordinate platform escalation and media rebuttals | |
|---|---|
| Task | Rationale |
| Monitoring & listening | rack misinformation related to:<br>• Voting logistics (e.g., fake polling station info)<br>• Voter intimidation or violence<br>• Election rigging claims<br>• Deepfakes, or synthetic media impersonations |
| Rapid Response Workflow | • Flag high-risk content for fact-checkers or journalists |

| | |
|---|---|
| | • Escalate platform violations (e.g., coordinated inauthentic behaviour)<br>• Trigger pre-approved debunking countermeasures |
| Daily SMA Reports | • Descriptive: Real-time dashboards on engagement spikes<br>• Diagnostic: Identifying whether surges are organic or manufactured<br>• Predictive: Likelihood of narrative migration across platforms<br>• Prescriptive: Content recommendations for rebuttal or prebunking |
| Public-Facing Pre- & Debunking | Partner with fact-checkers and independent media to:<br>• Debunk voting station and similar narratives<br>• Counter Election Suppression disinformation |

| POST-ELECTION PHASE<br>Date: 1–4 weeks after election day<br>Goal: Debrief, debunk post-election narratives, and begin long-term resilience work | |
|---|---|
| Task | Rationale |
| Disinformation Debrief | Analyse the dominant post-election narratives:<br>• Allegations of fraud or election denialism<br>• Narrative laundering of fringe claims<br>• Targeted attacks on the electoral commission or media |
| Longitudinal Analysis | Compare pre-, during-, and post-election data to:<br>• As the effectiveness of countermeasures<br>• Evaluate amplification patterns<br>• Identify cross-platform spillover (e.g., fringe-to-mainstream narrative migration) |
| Public-Facing Debunking | Partner with fact-checkers and independent media to:<br>• Debunk post-election hoaxes<br>• Counter scapegoating or conspiracy theory framing<br>• Address election denialism targeting marginalised groups |
| Final Report | • Descriptive: Summary of disinformation campaign timelines<br>• Diagnostic: Which actors, narratives, and platforms dominated<br>• Predictive: Long-term risks to electoral integrity or democratic trust<br>• Prescriptive: Recommendations for legal, media, and regulatory follow-up |

## III. Post-Election Phase (1–4 weeks after election)

*Goal: Debrief, debunk post-election narratives, and begin long-term resilience work*

**Key Activities:**

1. Disinformation Debrief: Analyse the dominant post-election narratives:

- Allegations of fraud or election denialism
- Narrative laundering of fringe claims
- Targeted attacks on the electoral commission or media

2. Longitudinal Analysis: Compare pre-, during-, and post-election data to:
    - As the effectiveness of countermeasures
    - Evaluate amplification patterns
    - Identify cross-platform spillover (e.g., fringe-to-mainstream narrative migration)

3. Public-Facing Debunking: Partner with fact-checkers and independent media to:
    - Debunk post-election hoaxes
    - Counter scapegoating or conspiracy theory framing
    - Address voter suppression disinformation targeting marginalised groups

4. Final SMA Report

    - Descriptive: Summary of disinformation campaign timelines
    - Diagnostic: Which actors, narratives, and platforms dominated
    - Predictive: Long-term risks to electoral integrity or democratic trust
    - Prescriptive: Recommendations for legal, media, and regulatory follow-up

# ANNEXURES

## Annexure 1: Glossary of terms

***Algorithmic Bias***

Systematic errors in algorithms that lead to unfair or skewed outcomes, such as prioritising certain content or demographics. Disinformation, or algorithmic bias, can amplify misleading narratives and marginalise accurate information, often reinforcing stereotypes or polarising content. *See also: Recommender Algorithm*.

***Algorithmic Transparency***

The degree to which the operations, criteria, and decision-making processes of algorithms, particularly recommender algorithms, are openly disclosed and understandable to users and researchers. Lack of transparency on social media platforms fuels disinformation by obscuring how content is prioritised or amplified, hindering efforts to detect manipulation or bias. *See also: Recommender Algorithm, Algorithmic Bias, Data Access.*

***Algorithms***

A fixed series of steps that a computer performs to solve a problem or complete a task. On social media platforms, algorithms compile, and present content based on users' engagement history and predicted interests, often influencing the spread of disinformation. *See also: Recommender Algorithm, Algorithmic Bias.*

***Amplification***

The process of increasing the reach or visibility of content, either organically (through shares, likes, and comments) or artificially (via bots, sock puppets, or astroturfing). Amplification can also occur independently of algorithms or through coordinated efforts to manipulate platform rankings. *See also: Manufactured Amplification*.

***Anonymous Influencers***

High-reach social media accounts that conceal the identity of the person or group behind them. They shape narratives, influence public opinion, and often spread disinformation or polarising content while evading accountability. Frequently hired for coordinated messaging. *See also: Influencers-for-Hire*

***Application Programming Interface (API)***

A set of protocols and tools that allow software to extract social media data from platforms hosting user-generated content. APIs are used in disinformation research to monitor, detect, and analyse information risks. *See also: Data Access*.

***Artificial Intelligence (AI)***

Computer systems performing tasks that typically require human intelligence, such as learning or pattern recognition. In disinformation, AI generates convincing fake content (e.g., deepfakes, text, images) or aids in detecting manipulation campaigns.

***Astroturfing***

A deceptive tactic that creates the illusion of grassroots support for a cause, individual, or campaign, orchestrated by hidden actors such as political operatives or PR firms. This tactic is commonly used in disinformation campaigns to manipulate public opinion or create a false sense of legitimacy. *See also: Coordinated Inauthentic Behaviour*

***Automated Reporting***

The creation of social media monitoring reports that collate analysis and information at relevant intervals without manual user intervention.

***Automation***

Software tools designed to complete tasks with minimal human direction. In disinformation, automation amplifies misleading narratives through the use of bots or coordinated campaigns. *See also: Bots, Botnet*

***Bandwagon Effect***

A cognitive bias where individuals adopt beliefs or behaviours because they appear popular. Disinformation campaigns exploit this by creating the illusion of widespread support through likes, shares, or trending status. *See also: Cognitive Biases.*

***Behavioural Science***

Behavioural science is an interdisciplinary field that draws from psychology, cognitive science, economics, and neuroscience to study how people make decisions and behave in real-world contexts. In the context of disinformation, behavioural science helps explain how cognitive biases, emotional triggers, and social influences shape how individuals interpret, believe, and act upon false or misleading information.

***Bots***

Social media accounts operated by computer programs to generate content. Bots after often used to amplify misleading narratives, hijack trending lists, or create the illusion of public support. *See also: Botnet*

***Botnet***

A single operator coordinates a network of bots, often numbering in the tens of thousands, to amplify disinformation or manipulate online discourse. Also known as a **bot farm**. *See also: Bots.*

***Clickbait***

Headlines designed to entice clicks, often leading to misleading or low-value content. Clickbait uses provocative or deceptive headlines/images to lure users, frequently spreading disinformation. *See also: Deceptive Design.*

***Cognitive Biases***

Unconscious thinking patterns that influence how people interpret information. Disinformation exploits these unconscious thinking patterns to increase the likelihood of individuals accepting or sharing false narratives. *See also: Bandwagon Effect, Confirmation Bias, Behavioural Science, and Inoculation Theory.*

***Commercial Disinformation***

Marketing, communications, or PR firms offer disinformation services to manipulate elections, legislation, or political issues for profit. Services include impersonation, harassment, and hashtag hijacking, alongside legitimate marketing. The firms are typically not linked to a particular state but provide services to diverse clients, including governments, politicians, and businesses. *See also: Disinformation-for-Hire, (Covert) Social Media Management*

***Community***

A network of users connected by shared interests, beliefs, or behaviours, often formed around hashtags, influencers, or private groups. *See also: Community Mapping.*

***Community Mapping***

The process of identifying and analysing social media data to detect online communities to understand their structure, key actors, and content spread. Used to trace disinformation origins and influence patterns and typically presented as Data Visualisations. *See also: Community, Data Visualisations.*

***Confirmation Bias***

The tendency to interpret information in ways that confirm existing beliefs, making individuals more susceptible to disinformation (e.g., accepting positive narratives about favoured candidates while dismissing negative ones). *See also: Cognitive Biases*

***Conspiracy Theory***

Narratives exploiting grievances about power structures, proposing that a small group controls global events. These narratives erode trust in institutions by using speculative reasoning. *See also: Information Disorders.*

***Content***

Any material created and shared by users on social media platforms, including text, images, video, audio, and data, whether posted publicly or privately.

***Content Farms***

Entities or networks that mass-produce low-quality, misleading, or sensational content to exploit recommender algorithms, generate ad revenue, or spread disinformation. Often using AI tools or coordinated accounts, content farms flood platforms with clickbait or false narratives, overwhelming information ecosystems.

***Content Moderation***

The process of detecting and addressing content that violates platform terms of use utilises both automation and human review. Actions include demonetisation, downgrading, or removal. Disinformation persists due to inconsistent moderation, particularly in non-English languages. *See also: Linguistic Disparity in Moderation.*

***Content Removal***

A moderation decision to delete content violating a platform's Terms of Service. Enforcement varies across languages and regions, raising concerns about transparency and consistency. *See also: Content Moderation.*

***Coordinated Inauthentic Behaviour (CIB)***

Networks of accounts secretly work together to sway online narratives by employing strategies such as identical posts and coordinated timing. Central to many influence operations. See also: Astroturfing, Influence Operations. Linguistic Disparity in Moderation.

***(Covert) Social Media Management***

The clandestine coordination of social media accounts, particularly vast networks of influencers like anonymous influencers and micro-influencers, to amplify disinformation campaigns typically by Commercial Disinformation operators. Unlike legitimate social media management, which focuses on transparent marketing or brand promotion, covert operations are conducted by entities tied to commercial disinformation, using deceptive tactics to manipulate narratives, boost engagement, or create the illusion of organic support. *See also: Commercial Disinformation, Disinformation-for-Hire, Influencers-for-Hire, and Anonymous Influencers*.

***Dark Web***

Parts of the internet that are not indexed or searchable frequently host illegal content. In disinformation, it may be used to coordinate campaigns or share manipulated content and malinformation. See also: Malinformation

***Data Access***

The ability to retrieve digital information from platforms, often via APIs or scraping, is crucial for disinformation research. Platform policies are increasingly restricting this crucial tool for disinformation research. *See also: API, Web Scraping*

***Data Mining***

The process of discovering patterns in large social media datasets to detect coordinated behaviours, influential accounts, or the spread of narratives allows data collection (e.g., scraping) to transform raw data into insights. See also: Social Media Digital Forensics, Social Media Analytics

***Data Science***

An interdisciplinary field that uses scientific methods, algorithms, and systems to extract insights and patterns from structured and unstructured data. In the context of disinformation, data science enables the detection, analysis, and prediction of information manipulation by tracking digital behaviours, mapping networks, and identifying anomalies in content dissemination.

***Data Visualisations***

Graphical representations (e.g., charts, network graphs) of data to reveal patterns or trends. In disinformation research, data visualisations simplify complex information ecosystems (e.g., using Gephi for network analysis and community mapping). See also: Community Mapping

***Debunking***

Exposing and correcting false claims through fact-checking, investigations, or exposés. Aims to counter disinformation and misinformation. *See also: Fact-checking, Prebunking*.

***Deceptive Design/Dark Design***

User interface features intentionally designed to trick or manipulate users into actions they might not otherwise take, such as sharing data, clicking misleading content, or engaging with disinformation. Also known as Dark Design. *See also: Clickbait.*

***Deepfakes***

Synthetic multimedia content that convincingly mimics real people or events, typically created to deceive. Now increasingly produced using accessible, user-friendly Generative AI platforms like ChatGPT, Grok, and others, deepfakes enable malicious actors to craft realistic fake videos, audio, or images for disinformation campaigns, such as impersonating public figures or spreading false narratives. *See also: Synthetic Media, Generative AI*

***Demonetisation***

A moderation action removing a creator's access to platform revenue without deleting their account. See also: Content Moderation.

***Deplatforming***

A moderation decision to temporarily or permanently ban a user from a platform. See also: Content Moderation.

**Digital Democracy**

The use of digital technologies to support democratic participation, transparency, and accountability. It encompasses online civic engagement, access to information, and the ability to express political views in safe and inclusive digital spaces without fear of retribution. Digital democracy depends on the integrity of the online information environment. When polluted by disinformation, misinformation, hate speech, or manipulation, these spaces can become tools for exclusion, polarisation, democratic erosion, and undermining electoral integrity. See also: Electoral Integrity, Digital Resilience.

***Digital Literacy***

The capacity to evaluate and interact with digital content critically, identify information disorders, and safeguard online privacy is essential. also: Media Literacy, Digital Resilience, Information Disorders

***Digital Resilience***

The capacity of individuals or societies to withstand and adapt to digital threats like disinformation, surveillance, censorship, and others.

***Digital Rights***

Online freedoms and protections include privacy, freedom of expression, and access to information. Encompasses safeguards against censorship, surveillance, and other online harms.

***Disinformation***

False information deliberately created or spread to cause harm, often for political, financial, or social motives. See also: Misinformation, Malinformation, Information Disorders.

***Disinformation-for-Hire***

A shadow industry where influencers, including anonymous influencers and micro-influencers, are covertly recruited to amplify false or misleading narratives for political, financial, or social gain. *See also: Commercial Disinformation, Influencers-for-Hire, Anonymous Influencers, (Covert) Social Media Management*

***Doxing/Doxxing***

The act of disclosing a person's private or personally identifiable details, like their home address, phone number, or workplace, without their consent, frequently aims to intimidate, harass, or silence them. Coordinated disinformation or harassment campaigns frequently employ this tactic. *See also: Malinformation.*

***Echo Chamber***

An online environment where users are primarily exposed to information, opinions, or beliefs that reinforce their own, while opposing views are filtered out. Echo chambers amplify confirmation biases and can intensify polarisation, making individuals more vulnerable to disinformation. *See also: Filter Bubble.*

***Election Denialism Disinformation***

Disinformation campaigns that falsely claim elections are fraudulent or illegitimate, aiming to undermine trust in democratic outcomes. Often propagated post-election through narrative hijacking and astroturfing, these efforts leverage feedback loops and computational propaganda to polarise voters and destabilise institutions. *See also: Electoral Disinformation*

***Electoral Disinformation***

False or misleading information deliberately spread to influence elections, undermine electoral integrity, or manipulate voter behaviour, threatening digital democracy. It includes tactics like voter suppression, disinformation, election denialism, microtargeted disinformation, fake news, deepfakes, or narrative hijacking to sow distrust, polarise voters, or discredit candidates.

***Electoral Integrity***

The degree to which electoral processes are free, fair, and credible, supported by transparent systems, impartial institutions, and an informed electorate. In digital spaces, electoral integrity relies on the integrity of digital democracy in the online information environment: voters must be able to access accurate, trustworthy information without being misled by disinformation, misinformation, hate speech, or foreign and domestic manipulation. A compromised digital environment can distort public perception, suppress participation, and undermine trust in electoral outcomes. *See also: Digital Democracy, Electoral Disinformation, Voter Suppression Disinformation, Election Denialism, Electoral Integrity*

***Engagement Rate***

The measurable interaction (e.g., likes, shares, comments) users have with online content, often manipulated in disinformation campaigns

***Fact-Checking***

Verifying the accuracy of public statements or reports.

***Fake Followers***

Fake social media accounts are purchased to inflate the perceived popularity of a specific account, which creates false credibility.

***Fake News***

Disinformation formatted to resemble authentic news, often as falsified articles or websites.

***Filter Bubble***

An algorithm-driven information environment curates content based on user behaviour, which passively isolates users from diverse perspectives. Unlike echo chambers, filter bubbles are primarily driven by platform algorithms. *See also: Echo Chamber, Recommender Algorithm*.

***Follow Train***

A coordinated tactic where users mutually follow each other to artificially boost metrics or visibility. Can be exploited to build fake credibility for disinformation accounts.

***Foreign Information Manipulation and Interference (FIMI)***

The deliberate actions undertaken by a foreign government or entity to exert influence over another country's decision-making, policies, or public opinion, often in a way that benefits the foreign actor's interests. These operations can employ covert and overt methods, including disinformation campaigns, cyberattacks, and financial inducements, and are often designed to undermine democratic institutions, manipulate public discourse, or advance a foreign government's strategic objectives. Also known as **Foreign Influence Operations**.

***Freedom of Speech***

The right to express opinions without censorship or penalty. Often misused to resist content moderation, though disinformation can suppress targeted groups' speech.

***Fringe Networks***

Niche platforms (e.g., Parler, 4chan, Truth Social) attracting non-mainstream users with controversial voices. Often hosts disinformation or extremist content.

***Gendered Disinformation***

Gendered disinformation refers to false or misleading content specifically designed to target women and gender non-conforming individuals, particularly those in public, political, or activist roles. It blends traditional disinformation tactics with gender-based abuse to silence, discredit, or intimidate its targets. Common techniques include the spread of misogynistic narratives, the sexualisation and manipulation of images or videos (including deepfakes), the reinforcement of harmful gender stereotypes, and coordinated harassment campaigns involving threats of violence, doxxing, or cyberattacks. The goal is not only reputational harm but also to deter political participation and limit visibility in public discourse.

***Generative AI***

AI that creates new content (e.g., text, images, audio), used in disinformation to produce deepfakes or tailored propaganda. Examples include ChatGPT, DALL-E, and Grok. *See also: Deepfakes, Large Language Models, Synthetic Media.*

***Geolocation Analysis***

Using digital tools, often satellite imaging, to identify the geographic context of content (e.g., images, videos). Helps verify or debunk disinformation tied to specific locations. *See also: Image Recognition*

***Hate Speech***

Communication attacking individuals or groups based on protected attributes (e.g., race, gender). Often overlaps with disinformation to incite violence or silence voices. *See also: Digital Rights, Freedom of Speech, Identity-Based Disinformation.*

***Hoax***

A fabricated piece of disinformation tied to a specific event, designed to mislead or provoke. See also: Information Disorders.

***Identity-Based Microtargeted Disinformation***

A form of disinformation that strategically targets specific demographic or social identity groups, such as those defined by race, ethnicity, gender, religion, geography, or class, through tailored content designed to exploit their unique fears, grievances, or social vulnerabilities. This tactic leverages audience segmentation tools, inferred identities, and increasingly, GenAI technologies to personalise false or misleading narratives at scale, often aiming to inflame identity-based divisions, sow distrust, suppress voter turnout, or incite hostility towards other groups. It is especially potent in electoral contexts where the "us vs. them" dynamics are weaponised for political gain. *See also: Inferred Identities, Electoral Disinformation, Microtargeting, Voter Suppression Disinformation*

***Image Recognition***

Machine learning processes analyse images, which are used in disinformation research to detect manipulations or verify authenticity (e.g., via Google Reverse Image Search). *See also: Geolocation Analysis*

***Inferred Identities***

A set of personal or social characteristics, such as race, gender, religion, location, or political affiliation, is deduced by algorithms or platforms based on a user's online behaviour, interactions, and data patterns rather than voluntarily disclosed information. These algorithmically derived profiles are often used in microtargeting, including disinformation campaigns, to craft tailored content that exploits perceived vulnerabilities or group affiliations.

***Information Disorders***

An umbrella term for various forms of harmful or misleading content that distort truth and undermine public discourse. Includes disinformation, misinformation, malinformation, propaganda, conspiracy theories, clickbait, satire or parody shared as fact, hoaxes, trolling, imposter content, synthetic media, and others.

***Information Integrity***

An ecosystem where accurate, reliable information is consistent and accessible and where freedom of expression is protected. *See also: Information Disorders.*

***Information Overload***

A state where excessive information volume overwhelms critical processing, enabling disinformation to spread unnoticed, especially during crises. *See also: Information Vacuum.*

***Information Vacuum***

A lack of timely, accurate information, allowing disinformation, rumours, or conspiracies to fill the gap, common during crises or elections. *See also: Information Overload.*

***Influence Operations***

Coordinated efforts to manipulate public opinion or behaviour using deceptive tactics like disinformation or fake accounts. Common in elections or conflicts. *See also: Coordinated Inauthentic Behaviour, Foreign Information Manipulation and Interference, Platform Manipulation.*

***Influencers***

Individuals with large social media followings who shape opinions or behaviour. *See also: Anonymous Influencers, Influencers-for-Hire.*

***Influencer-for-Hire***

An influencer paid to amplify specific narratives or discredit opponents, often covertly. *See also: Anonymous Influencers, Disinformation-for-Hire, (Covert) Social Media Management.*

***Inoculation Theory***

The strategy, grounded in behavioural science, aims to foster psychological resilience against misinformation by presenting individuals with a toned-down version of deceptive content, accompanied by refutations or counterarguments. This “prebunking” approach equips individuals to recognise and reject future attempts at manipulation. *See also: Prebunking, Debunking, Behavioural Science*.

***Internet Shutdowns***

Intentional interruptions of internet connectivity aim to regulate the flow of information, frequently intensifying misinformation by restricting the availability of precise information. *See also: Information Vacuum.*

***Labelling***

Content moderation practice of applying informational labels to posts or accounts to provide context (e.g., marking disinformation or sensitive content). *See also: Content Moderation.*

***Large Language Models (LLMs)***

AI trained on vast text data to generate human-like language, used to create disinformation or analyse content. Examples include ChatGPT, Grok, and LLaMA. *See also: Natural Language Processing.*

***Linguistic Disparity in Moderation***

The inconsistent or inadequate moderation of harmful content, such as disinformation or hate speech, in non-English languages due to limited automated systems or human oversight. Malicious actors exploit this gap by using native languages or word camouflage to bypass detection, particularly in electoral disinformation campaigns targeting diverse linguistic communities. *See also: Word Camouflaging*

***Malign Actors***

Individuals, groups, or commercial entities intentionally spreading disinformation or manipulating information ecosystems.

***Malinformation***

Truthful information shared to cause harm, often by revealing private data or using facts out of context (e.g., doxxing). *See also: Disinformation, Misinformation*.

***Manufactured Amplification***

Deliberate boosting of content visibility through deceptive means (e.g., bots, sockpuppets) to distort perceived popularity or credibility.

***Mass Brigading***

A coordinated online tactic where large groups of users target a specific individual, post, or viewpoint, often by flooding it with replies, quote posts, or negative comments. The goal is to discredit, silence, or intimidate the target, distort public perception, and create the illusion of consensus. *See also: Astroturfing*

***Media Literacy***

Competencies to critically engage with media, assess source credibility and truthfulness. *See also: Digital Literacy.*

***Meme***

Humorous and shareable content that can be manipulated to oversimplify or distort facts for political impact in disinformation campaigns.

***Microtargeting***

The practice of sending highly tailored content or ads to small, specific groups based on personal characteristics and beliefs. *See also: Microtargeted Disinformatio*n.

***Misinformation***

False information spread without intent to mislead and is often believed to be true by sharers. *See also: Disinformation, Malinformation.*

***Narrative Hijacking***

The intentional appropriation of existing narratives, events, or trending topics aims to spread disinformation or redirect public discourse. Often used to exploit crises or popular hashtags, this tactic amplifies false narratives by blending them with legitimate conversations, particularly in electoral disinformation campaigns. Typically occurs during election denialism campaigns. *See also: Narrative Seeding, Narrative Warfare.*

***Narrative Laundering***

Legitimising false information by republishing it through increasingly credible sources. *See also: Narrative Seeding, Narrative Hijacking*

***Narrative Seeding***

The act of introducing disinformation into online spaces, such as forums and comments, encourages its organic spread among unsuspecting users or through media. *See also: Narrative Laundering, Narrative Hijacking.*

***Narrative Warfare***

Strategic manipulation of narratives to influence public perception or behaviour in disinformation campaigns. *See also: Narrative Hijacking, Narrative Laundering, Narrative Seeding*

***Natural Language Processing (NLP)***

A field of AI that enables computers to understand, interpret, and generate human language. In disinformation research, NLP is used to analyse large volumes of social media content, detect harmful narratives, identify emotional tones, and automate the recognition of misinformation and disinformation patterns. *See also: Large Language Models.*

***Online Violent Extremism***

Using digital platforms to promote or incite ideologically motivated violence, often through echo chambers and algorithmic reinforcement.

***Open-Source Intelligence (OSINT)***

The practice of collecting, analysing, and interpreting publicly available information from digital, print, and broadcast sources to generate actionable insights. In the context of disinformation, OSINT leverages social media platforms, news outlets, websites, forums, and multimedia content to detect coordinated manipulation, trace the origins of false narratives, and identify threat actors. OSINT is a foundational method in digital investigations, election monitoring, and media forensics, prized for its transparency, verifiability, and ethical alignment when conducted responsibly.

***Parody***

Humorous imitation of a person or style, often mistaken for real content in disinformation contexts.

***Platform Manipulation***

The deliberate exploitation of social media platform features, algorithms, or affordances to amplify disinformation, distort narratives, or influence public perception. Tactics include gaming trending algorithms, coordinating fake accounts, or leveraging (covert) social media management to create artificial engagement or visibility.

***Post-Truth***

A situation where emotions and beliefs shape public opinion over facts, enabling disinformation to thrive.

***Prebunking***

Anticipating and countering disinformation before it spreads, using past fact-checks to prepare responses. See also: Inoculation Theory, Debunking.

***Propaganda***

A form of strategic political communication aimed at influencing public opinion or behaviour in support of a political, ideological, or institutional agenda. Propaganda typically involves the selective use of facts, emotional appeals, repetition, and symbolic messaging to persuade and mobilise audiences. While not always false or harmful, propaganda can become problematic when it distorts reality, suppresses dissent, or is used to legitimise authoritarianism. In electoral contexts, it is distinct from disinformation, although the two may intersect.

***Psychographic Profiling Data***

Information that categorises individuals based on psychological attributes such as values, beliefs, interests, attitudes, lifestyles, and personality traits. This data is often derived from online behaviour, including social media activity, likes, shares, and browsing habits, and is used to predict and influence decision-making, especially in targeted advertising and political microtargeting campaigns. *See also: Inferred Identities, Microtargeting*

***Recommender Algorithm***

An automated system used by social media platforms to select, rank, and present content based on user behaviour, interests, and engagement signals. Powered by machine learning, this algorithm prioritises attention-grabbing content, often amplifying disinformation by creating feedback loops that reinforce cognitive biases, such as confirmation bias, and entrench user beliefs. *See also: Cognitive Biases, Confirmation Bias, Feedback Loop, Filter Bubble, Algorithmic Bias, Platform Manipulation*.

***Regulatory Responses***

Government or institutional policies and laws aimed at combating disinformation, hate speech, or platform manipulation. Examples include content moderation mandates, transparency requirements for algorithms, or penalties for spreading false information. These efforts aim to enhance information integrity but face challenges in enforcement and striking a balance with free speech.

***Satire***

Humorous content using irony or exaggeration, often mistaken for real news in disinformation contexts. *See also: Information Disorders*

***Shadow Banning***

Limiting a user's content visibility without their knowledge, often for content moderation purposes. *See also: Content Moderation.*

***Social Media Data***

Publicly available content and metadata from platforms used to detect disinformation trends or campaigns. *See also: API, Data Scraping*

***Social Media Digital Forensics***

The specialised process of collecting, preserving, and analysing social media data to uncover evidence of harmful activities, such as disinformation, cyberbullying, or hate speech, often perpetrated by anonymous accounts. Techniques include metadata analysis, linguistic profiling, network mapping, and reverse image searches to trace origins, identify hidden

networks, or attribute content to malicious actors despite anonymity. This field is critical for exposing coordinated manipulation and ensuring admissible evidence for legal or public accountability. *See also: Open-Source Intelligence, Algorithmic Transparency, Data Access*

***Social Media Metrics***

The analysis of social media data to provide a quantitative measurement of a topic. For example, analysing the conversation volume on a specific topic and comparing that against other topics.

***Social Media Monitoring***

The real-time tracking and recording of social media activity, such as mentions, hashtags, or keywords, to observe engagement, flag incidents, and identify disinformation as it spreads. *See also: Social Media Listening.*

***Social Media Listening***

The process of tracking and analysing online conversations to understand public sentiment, detect emerging trends, and uncover disinformation patterns. Unlike social media monitoring, which focuses on observing and recording activity, social listening interprets meaning and context. *See also: Social Media Monitoring.*

***Social Media Metrics***

Quantitative analysis of social media data (e.g., conversation volume) to measure engagement or trends.

***Social Media Platforms***

Online services that enable users to create, share, and engage with content and communities. This includes public-facing platforms (such as Facebook, X, TikTok) as well as private or encrypted messaging apps (such as WhatsApp, Messenger, Telegram), which are increasingly used to spread disinformation beyond public view.

***Sock Puppet***

A human-managed account using a false identity to deceive or amplify disinformation, unlike automated bots. *See also: Sock Puppetry*

***Sock Puppetry***

The coordinated use of multiple fake online identities, often managed by a single actor, to manipulate discussions, amplify narratives, or create false consensus. Unlike bots, sockpuppets are typically human-operated and used in disinformation to deceive audiences or evade detection. *See also: Sock Puppet*

***Synthetic Media***

AI-generated or manipulated content (e.g., deepfakes, fabricated text) used to create convincing disinformation. See also: Deepfakes, Generative AI.

***Targeted Harassment***

Coordinated online attacks to threaten or silence individuals, often overlapping with disinformation or hate speech.

***Technology-Facilitated Gender-Based Violence (TFGBV)***

Gender-based harm via digital platforms, including harassment, doxxing, or gendered disinformation targeting women or gender-diverse individuals. See also: Gendered Disinformation.

***Trolling***

Inflammatory online behaviour to provoke negative reactions, often used in disinformation to distract or polarise. See also: Troll Farm.

***Troll Farm***

A group engaging in coordinated trolling or bot-like narrative promotion, also called a troll army. See also: Trolling,

***User-Generated Content (UGC)***

Any form of content created and voluntarily shared by individual users on digital platforms, rather than by the platforms themselves, professional media, or paid content producers. UGC is typically unpaid and reflects the personal views, creativity, or experiences of the user. It stands in contrast to coordinated content produced by content farms, bot farms, or commercial disinformation operators. *See also: Content Farms, Commercial Disinformation, Disinformation-For-Hire, (Covert) Social Media Management*

***Violent Extremism Disinformation***

Content that falsely associates individuals, groups, or political movements with terrorism or extremist ideologies incites fear, justifies repression, or undermines opposition. Often used to frame opponents as threats to national security or public order, especially during elections, this form of disinformation can escalate tensions and contribute to real-world violence. *See also: Electoral Disinformation.*

***Voter Suppression Disinformation***

False or misleading information spread to discourage or prevent voters from participating in elections, often targeting marginalised groups to undermine electoral integrity and digital democracy. Tactics include spreading fake polling place details, false voting deadlines, or fabricated eligibility requirements, as well as leveraging platform manipulation or using anonymous influencers. *See also: Electoral Disinformation.*

***Web Scraping***

Extracting data from websites without APIs, used in disinformation research but may violate platform terms.

***Word Camouflage***

Deliberate word alterations, like the use of slang, misspellings, symbols, or non-English terms, aim to circumvent content moderation systems. Often employed in disinformation or hate campaigns, especially in native languages, to spread harmful narratives while avoiding automated filters, contributing to linguistic disparity in moderation. *See also: Linguistic Disparity in Content Moderation*

**ANNEXURE 2: Further reading**

**Disinformation monitoring guides**

- OSCE. (2021). Guidelines for Observation of Election Campaigns on Social Networks. https://www.osce.org/files/f/documents/4/1/500581_0.pdf.
- UNESCO. (2022) Counteracting Electoral Disinformation: Practical Guide for Organizations and Electoral Bodies. https://unesdoc.unesco.org/ark:/48223/pf0000380594_eng.
- Carnegie Endowment for International Peace (2024). Countering Disinformation Effectively: An Evidence-Based Policy Guide. Countering Disinformation Effectively: An Evidence-Based Policy Guide | Carnegie Endowment for International Peace.
- UN Office on Genocide Prevention. (2024). A Comprehensive Methodology for Monitoring Social Media. A Comprehensive Methodology for Monitoring social media.
- The Ghana Center for Democratic Development. (2021) Dealing With Disinformation And Misinformation During Elections: A Toolkit To Guide WAEON Members https://www.waeon.org/assets/downloadables/WAEON-toolkit-ENGLISH-full-WB-1-1.pdf.

**Case studies:**

In addition to those referenced in the guide, these additional case studies provide contextual information of the manifestations of electoral disinformation in a number of elections of African countries. This a sample of useful studies.

***Côte d'Ivoire***

- van Baalen S., Gbala A., (2023) Patterns of Electoral Violence During Côte D'Ivoire's Third-Term Crisis, *African Affairs*, Volume 122, Issue 488, July 2023, Pages 447–460, https://doi.org/10.1093/afraf/adad020

  *This study examines the dynamics of electoral violence during Côte d'Ivoire's 2020 third-term crisis, highlighting how political actors used identity-based narratives and*

*disinformation to incite unrest. The study maps how online mobilisation, hate speech, and contested legitimacy claims contribute to offline violence, offering insights into the relationship between political communication and electoral instability in fragile democracies.*

***Malawi***

- Kanyang'wa, M., & Lotshwao, K. (2023). Understanding misinformation and disinformation in elections: Lessons from the Malawi 2019 presidential elections. *African Journal of Democracy and Electoral Reform*, 3(2). https://hdl.handle.net/10520/ejc-aa_ajder_v3_n2_a4

  *The study examines the dissemination and influence of misinformation and disinformation in Malawi's 2019 presidential elections. The study highlights how false narratives circulated via social media and traditional platforms shaped voter perceptions, delegitimised institutions, and fuelled post-election protests. It offers lessons on electoral communication vulnerabilities in fragile democracies and underscores the need for stronger counter-disinformation strategies.*

***Tanzania***

- Ishengoma, D. J., & Mutinta, G. (2024). Tanzanian journalists in countering fake news: disinformation and misinformation. *Communicare: Journal for Communication Studies in Africa*, 43(2), 50–64. https://doi.org/10.36615/gz3ezj29

  *Examines Tanzanian journalists' abilities to combat fake news by assessing their awareness, challenges, and strategies. The study explores strategies used by media practitioners, including fact-checking and source verification, while highlighting institutional knowledge barriers. It provides insight into the challenges of sustaining media integrity and public trust in low-resource and politically sensitive contexts.*

***South Africa***

- Van Damme, P., Findlay, K., Cornelissen, A. (2024, December). Generative AI and its influence on South Africa's 2024 elections. *German Council for Foreign Relations*.

https://dgap.org/en/research/publications/generative-ai-and-its-influence-south-africas-2024-elections

*Examines how generative AI influenced South Africa's 2024 elections through deepfakes, disinformation, and microtargeting, with a focus on both domestic manipulation and Foreign Information Manipulation and Interference (FIMI). Highlights regulatory gaps and platform failures and proposes safeguards to protect electoral integrity in the era of synthetic media.*

***Kenya***

- CIPESA. (2022, June). Disinformation in Kenya's Political Sphere. https://cipesa.org/wp-content/files/Disinformation-in-Kenyas-Political-Sphere-Actors-Pathways-and-Effects.pdf

*Examines disinformation in Kenya's 2022 election, detailing actors (e.g., anonymous influencers, disinformation-for-hire), pathways (e.g., deepfakes, WhatsApp messages), and effects on electoral integrity. It highlights strategies like fact-checking, digital literacy, and government engagement to counter grassroots amplification and platform manipulation, offering lessons for protecting digital democracy in African elections*

***Nigeria***

- Abba, A. I., Aluko, G. A., Chioma, A., Iruke, C., Ogide, V., Raji, A., Olatunji, A., Onoboh, H., Tijani, M., & Tola-Winjobi, F. (2023, June 8). Distorting Nigeria's Elections? How Disinformation Was Deployed in 2023. *IssueLab*. https://www.issuelab.org/resource/distorting-nigeria-s-elections-how-disinformation-was-deployed-in-2023.html

*Analyses how disinformation was deployed during Nigeria's 2023 general elections to manipulate public perception, suppress voter turnout, and erode trust in democratic institutions. It outlines tactics such as ethnically charged narratives, impersonation of official sources, fear-based messaging, and coordinated influencer campaigns. The*

*authors emphasise the role of social media in amplifying these efforts and call for stronger monitoring and civic education to counter future threats.*

***Democratic Republic of Congo, Kenya, Senegal, and South Africa***

- Centre for Information Integrity in Africa (2025, May). Tackling Disinformation in Four African Elections. https://ciia.africa/tackling-disinformation-in-four-african-elections/

*Highlights cross-country trends in disinformation tactics, such as coordinated influence operations, narrative hijacking via WhatsApp, and deepfake deployment, and outlines tailored countermeasures, including community-led fact-checking, digital literacy programmes, and platform partnerships. Offering practical, context-specific lessons, the report aims to support election observers and civil society in safeguarding electoral integrity through proactive, locally anchored responses.*

***Ghana***

- Media Foundation for West Africa. (2025, April). The State of Mis/Disinformation, Polarisation, and State Threat to Ghana. https://mfwa.org/wp-content/uploads/2025/04/The-State-of-MisDisinformation-Polarisation-and-State-Threat-to-Ghana -NEWLY-EDITED.pdf

*It examines disinformation in Ghana's 2024 election, highlighting politically motivated falsehoods, the role of pro-government media, rising online polarisation, and the growing use of state-aligned influencers to attack journalists and civil society actors.*

***West Africa***

- Atlantic Council's Digital Forensic Research Lab. (2023) Weaponised: How Disinformation Became a New Threat to West Africa. https://www.atlanticcouncil.org/in-depth-research-reports/report/disinformation-west-africa

*Investigates disinformation campaigns across West Africa, exposing the tactics of state-aligned actors, foreign influence operations, and disinformation-for-hire firms.*

*The study highlights the coordinated attacks on opposition figures, civic activists, and election observers in Nigeria, Mali, Côte d'Ivoire, and other regions.*

***Sahel Region***

- Sahel Research Group. (2024). Final Report on OPEN Study Day: Russian, Chinese, and Other Misinformation and Disinformation *Efforts in the Sahel Region* (Vol. 9, No. 7). https://sahelresearch.africa.ufl.edu/wp-content/uploads/sites/170/Open-Study-2024_-Final-Report.pdf

*This study delves into the ways in which Russian and Chinese influence operations take advantage of instability in the Sahel region, employing coordinated disinformation strategies that target civil society, elections, and anti-French sentiment. The study records the manipulation of narratives through local media, Telegram channels, and influencers who support the regime.*

***Zimbabwe***

- Chibuwe, A. (2024). Fake News as Political Communication: On Fake News, Digital Media, and the Struggle for Hegemony in Post-Mugabe Zimbabwe. *Political Research Quarterly*, 77(4), 1181–1195. https://doi.org/10.1177/10659129241262230

*Analyses how fake news functions as a tool of political communication in Zimbabwe's post-Mugabe era, particularly during elections. The study underscores the ways in which both ruling elites and opposition actors utilise digital media to challenge legitimacy and influence public perception within a highly polarised information landscape.*

***Zambia***

- The Carter Centre. (2022). Analysing Zambia's 2021 General Elections. https://www.cartercenter.org/resources/pdfs/news/peace_publications/election_reports/zambia-final-report-2021.pdf

*Documents the surge in social media disinformation during Zambia's August 2021 elections, especially via state-aligned Facebook pages masquerading as impartial news sources. Reports include false narratives used to sway votes toward the ruling party, weak fact-checking infrastructure, and the government's role in amplifying manipulative content. The report underscores the importance of civil society's grassroots rebuttals, the necessity for verified channels, and the importance of digital literacy.*

***Liberia***

- Mozilla Foundation. (2024, February 27). The 2023 Election in Liberia: Platform Interventions in a Low-Trust Election Environment. The 2023 Election in Liberia - Mozilla Foundation.

*The study delves into the involvement of social media platforms in the 2023 election in Liberia, where a climate of institutional mistrust fostered the growth of disinformation. Highlights include limited platform moderation, a lack of local language support, and minimal transparency regarding content enforcement. Recommends better platform accountability and civil society coordination to strengthen electoral integrity.*

***DRC***

- Internews. (2024, May). Social Media and Misinformation in the Electoral Context of the DRC. https://internews.org/wp-content/uploads/2024/08/social-media-and-misdisinformation-in-electoral-context_DR_Congo_English.pdf

*The article highlights patterns of narrative manipulation, platform misuse, and low media literacy as key vulnerabilities, while calling for enhanced monitoring, digital literacy interventions, and context-specific regulatory frameworks.*

***The regional dynamics of misinformation in sub-Saharan Africa***

- Gondwe, G. (2023). This paper discusses misinformation in countries with limited technological literacy, focussing on how individuals in sub-Saharan Africa engage with

fake news. *Paper presented at the Institute for Rebooting social media, Harvard University.* https://www.semanticscholar.org/paper/Misinformation-in-countries-with-limited-How-in-Gondwe/037dffb06e8f46d64dcb0bfb850bc64ed067d56d

*Examines how low technological literacy in sub-Saharan Africa fuels grassroots amplification of misinformation, offering insights into socio-cultural influences and strategies to enhance media literacy for digital democracy.*

- Gondwe, G. (2024). Artificial intelligence, journalism, and the Ubuntu robot in sub-Saharan Africa: Towards a normative framework. *Digital Journalism*, 1-19. https://www.tandfonline.com/doi/full/10.1080/21670811.2024.2311258

  *The article explores AI's role in African journalism, including risks of deepfakes and electoral disinformation, and proposes ethical frameworks to protect electoral integrity in digital ecosystems*.

- Madrid-Morales, D., Wasserman, H., Gondwe, G., Ndlovu, K., Sikanku, E., Tully, M., Umejei, E., & Uzuegbunam, C. (2021). Motivations for sharing misinformation: A comparative study in six sub-Saharan African countries. *International Journal of Communication*, 15, 1200-1219. https://ijoc.org/index.php/ijoc/article/view/14801

  *Investigates why individuals share misinformation in Sub-Saharan African countries, highlighting cognitive biases and grassroots amplification, with implications for countering electoral disinformation*.

- Mare, A., Mabweazara, H. M., & Moyo, D. (2019). 'Fake news' and cyber-*propaganda in sub-Saharan Africa: Recentering the research agenda. African* Journalism Studies, 40(4), 1-12. https://www.tandfonline.com/doi/full/10.1080/23743670.2020.1788295

  *Reframes disinformation research in Sub-Saharan Africa, focusing on computational propaganda and socio-cultural influences, offering insights into local dynamics of electoral disinformation.*

- Okolo, C. T. (2024). African Democracy in the Era of Generative Disinformation: Challenges and Countermeasures against AI-Generated Propaganda. *arXiv*. https://arxiv.org/abs/2407.07695

  *Provides a continent-wide overview of how generative AI technologies, such as deepfakes and synthetic narratives, are being used to disrupt democratic processes in Africa. The article provides case examples and proposes countermeasures that are based on AI governance, platform regulation, and civil society resilience.*

- Albrecht, E., Fournier-Tombs, E., & Brubaker, R. (2024). Disinformation and Peacebuilding in Sub-Saharan Africa: Security Implications of AI-Altered Information Environments. *United Nations University*. https://collections.unu.edu/view/UNU:9419#viewAttachments

  *Investigates how AI-driven disinformation undermines peacebuilding across Sub-Saharan Africa. Highlights the security risks associated with manipulated narratives in conflict and post-conflict contexts and calls for coordinated responses that link digital governance, peacebuilding, and AI regulation frameworks.*

- Timcke, S., Orembo, L., Hlomani, H., & Schültken, T. (2023). The Materials of Misinformation on the African Continent: Mid-Year Report. Research. *ICT Africa.* https://idrc.sun.ac.za/wp-content/uploads/ria_information_disorders_2023_mid_year_report.pdf

  *Analyses the underlying political economies that enable misinformation across Africa, focusing on infrastructure gaps, platform governance failures, and media capture. Highlights how context-specific disinformation thrives in environments marked by structural inequality and weak regulatory oversight.*

- Cunliffe-Jones, P. et al. (2021). The State of Media Literacy in Sub-Saharan Africa 2020 and a Theory of Misinformation Literacy, pp. 5–96, in Misinformation Policy in Sub-Saharan Africa: *Fulcrum.* https://www.fulcrum.org/downloads/794081189?locale=en

  *Presents an in-depth analysis of media literacy initiatives across Sub-Saharan Africa and proposes a theory of "misinformation literacy" tailored to the region's specific political, social, and technological contexts. Highlights the limitations of existing legal approaches and emphasises the importance of culturally grounded, community-led media literacy as a counter to disinformation.*

- Norman, I. D. (2024). Foreign Election Interference in Africa's De-Democratization Culture. *European Journal of Law and Political Science*, 3(3), 23–33. https://doi.org/10.24018/ejpolitics.2024.3.3.130

  *Explores how foreign actors have actively contributed to the erosion of democratic norms and practices in African countries through election interference. Examines the intersection of external manipulation, ranging from cyber operations and disinformation campaigns to economic coercion, with local vulnerabilities such as weak institutions, elite complicity, and limited digital literacy. Argues that foreign interference is not just opportunistic but deeply embedded in the continent's evolving political culture, where democracy is increasingly treated as a procedural façade rather than a substantive norm.*

**Books**

- Harari, Y. N. (2024). Nexus: A brief history of information networks from the Stone Age to AI. Random House.

  *Explores the evolution of information networks, including modern disinformation challenges, relevant to understanding platform manipulation and AI-driven threats.*

- McIntyre, L. (2023). On disinformation: How to fight for truth and protect democracy. MIT Press.

  *Offers strategies to combat disinformation, focusing on fact-checking and media literacy, which are essential for protecting digital democracy.*

- Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.

  *Examines how algorithms perpetuate bias, particularly in the context of recommender algorithms and platform manipulation, with implications for addressing socio-cultural influences in African digital spaces.*

- Van der Linden, S. (2023). Foolproof: Why misinformation infects our minds and how to build immunity. W.W. Norton & Company.

  *Explains psychological drivers of misinformation and proposes prebunking strategies.*

- Wasserman, H., & Madrid-Morales, D. (Eds.). (2022). Disinformation in the Global South. Wiley-Blackwell.

  *The study focusses on disinformation in developing regions, including Africa, and addresses grassroots amplification and linguistic disparity in moderation, along with strategies to protect digital democracy.*

- Woolley, S. (2023). Manufacturing consensus: Understanding propaganda in the era of automation and anonymity. Yale University Press.

  *The system analyses automated propaganda, offering insights into countering electoral disinformation.*

- Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books.

  *Explores how data-driven platforms enable microtargeted disinformation, relevant to addressing platform manipulation and algorithmic transparency.*