Strategic Communication and Foresight (SG.STRAT)
Data Team

FOREIGN INFORMATION MANIPULATION & INTERFERENCE

OSINT GUIDELINES

How to Detect and Analyse Identity-Based Disinformation/FIMI

A Practical Guide to Conduct Open Source Investigations

November 2024

European Union
EXTERNAL ACTION

EXECUTIVE SUMMARY

The "OSINT Toolkit to Detect and Analyse IBD-focused FIMI" serves as a comprehensive guide for conducting open-source intelligence (OSINT) investigations, specifically addressing incidents and campaigns of Foreign Information Manipulation and Interference (FIMI) that weaponise identity-based disinformation (IBD). In response to the increasing number of manipulative, coordinated, and intentional attacks against individuals and groups based on their gender, sexual orientation, race, or ethnicity, these guidelines have a two-fold objective. They aim to **equip potential victims with the tools and methodologies to detect, analyse, and gather evidence to counter these threats. At the same time, they wish to raise awareness among OSINT practitioners of recognising and adopting an IBD-specific lens in their investigative activities.**

The guidelines are structured as follows. First, they define IBD and discuss its intersection with FIMI, exploring how the identity dimension is exploited for malign purposes. Then, the guidelines introduce key frameworks that offer a systematic approach to analysing FIMI. The second part outlines best practices for conducting IBD-focused public interest OSINT investigations on FIMI, emphasising the need for thorough threat assessments of the online ecosystem. The section provides practical guidance on navigating the complexities of these campaigns while ensuring responsible investigative methods. **The third section presents a comprehensive and accessible catalogue of OSINT tools to investigate IBD within the context of FIMI.** These tools cover a range of functions, including archiving digital evidence, assessing coordination between actors, verifying the authenticity of content and perpetrators, tracing the source of attacks, and evaluating the overall impact of disinformation operations. Finally, the document concludes with actionable recommendations to strengthen democratic integrity and bolster the global response to IBD-driven FIMI. **These recommendations aim to guide future actions, promote greater collaboration within the OSINT community, and enhance the resilience of targeted groups against this phenomenon.**

Disclaimer: These guidelines have been produced as part of the project on identity-based disinformation/ FIMI, under the EEAS Information Integrity and Countering Foreign Information Manipulation and Interference Division. The content was developed in collaboration with EU DisinfoLab, with contributions from Maria Giovanna Sessa and Raquel Miguel Serrano. The authors thank Amaury Lespingart and Denis Teyssou for their reviews.

Built on existing frameworks, this document highlights a selection of third-party open-source tools and methodologies to incorporate an identity-based perspective in OSINT investigations. Please note that the examples presented are for illustrative purposes only. They represent a limited subset of threat actors' activities and should not be interpreted as conclusive evidence of broader trends in FIMI.

TABLE OF CONTENTS

INTRODUCTION

Long before labelling information operations as Foreign Information Manipulation and Interference (FIMI), open-source intelligence (OSINT) practitioners were actively engaged in detecting and analysing these disinformation campaigns. OSINT's ability to aggregate and analyse data from diverse, publicly available sources makes it an indispensable tool in detecting and countering FIMI and zooming into issue-specific attacks.

The advent of globalisation has brought cultural issues to the forefront of public debate, in conjunction with multicultural societies, the emancipation of women, and the ongoing fight for equal rights for the LGBTIQ+ community. **Perceiving these changes as a threat to traditional values and lifestyles, some groups fueled the rise of identity-based disinformation (IBD), which exploited these social fractures for malign purposes.** In the words of Thomas Rid: "Disinformation operations rely upon tactics that exploit technology, political divisions, and tensions between allies. Political fissures and frictions are a function of the target. The design of the divisive material and the craftsmanship of disinformation are a function of the attacker. The technological substrate and the available media platforms are a function of the operational environment" (Rid, 2021, p. 240)[1].

In this context, the intersection of FIMI and IBD, first explored by the European External Action Service (EEAS) with a specific anti-LGBTIQ+ lens (EEAS, 2023)[2], is marked by the strategic manipulation of identity politics to advance geopolitical agendas. While this interplay is crucial, it is important to note that not all FIMI campaigns focus on IBD, as interference and manipulation cover a broad range of topics beyond identity. **By exacerbating existing divisions and targeting vulnerable communities, these campaigns sow instability and conflict within countries where identity politics play a significant role.**

Paraphrasing the EEAS' definition, the guidelines refer to IBD-focused FIMI as *a non-illegal pattern of behaviour that threatens and causes direct or indirect harm to individuals and groups based on their gender, orientation, race or ethnicity, affects values and disrupts the political process. Such activity is manipulative, intentional, and coordinated, and it is conducted by state or non-state actors, including their proxies inside or outside of their territory*.

Researchers of gendered disinformation have highlighted the need to move beyond a binary conception of gender when addressing the issue and emphasised the critical role of intersectionality, as further discrimination arises from belonging to multiple marginalised groups (Sessa, 2022[3]; Sobieraj, 2019[4]). However, the nature of these vulnerabilities differs significantly. Gender and sexual orientation-related disinformation often targets personal identity and non-conformity to traditional societal norms, while race and ethnicity-focused disinformation frequently engages with legacies of historical injustices and collective national narratives. Therefore, IBD serves as a useful umbrella, but it is essential to recognise the nuances within each category, ensuring that responses are tailored to these diverse experiences rather than applying a one-size-fits-all approach.

In tackling such complex issues, **open-source intelligence proves to be a valuable tool for gathering insights into FIMI operations, including those that focus on IBD**. OSINT enables practitioners across various sectors to access and analyse publicly available data, offering critical transparency and actionable intelligence (Alaphilippe, 2022)[5]. However, it is important to remember that OSINT itself is a neutral method, and its application depends on the context and intent of its users. It can be employed for various purposes beyond IBD or even FIMI, stressing the importance of responsible and ethical use.

Therefore, **the purpose of these guidelines is to mitigate the impact of IBD-focused FIMI threats on democratic integrity**. On the one hand, they wish to equip the OSINT community with awareness of the identity lens when tackling disinformation, manipulation and interference. On the other hand, they hope to offer a toolkit for victims of IBD-focused FIMI to detect incidents and campaigns, so to broaden and strengthen collaboration within the defender community.

EXISTING FRAMEWORKS TO STUDY FIMI

At this stage, the defender community can rely on several frameworks to analyse FIMI, including IBD-related incidents and campaigns.

Frameworks for the analysis of Tactics, Techniques, and Procedures (TTPs)

- **The 'B' in the ABCDE Framework**[6] (Pamment, 2022). The acronym stands for actors, behaviour, content,

degree, and effect. TTPs are the operationalisation of behaviour, which in the case of FIMI is manipulative, intentional, and coordinated.

Example: Concerning the EEAS report on FIMI targeting LGBTIQ+ people, perpetrators deployed various TTPs, including targeting the audience. In detail, they identified social and technical vulnerabilities, existing prejudices, and segmented the audience demographically, geographically, and politically.

- **Standardised frameworks for the analysis of tactics, techniques and procedures (TTPs).** Standardised open source frameworks guide analysts in identifying and describing the manipulative behaviours used at each stage of FIMI operations from planning to execution. These models allow to effectively recognise early warning signs of an FIMI incident, share and compare insights across teams, and disrupt malicious operations more swiftly. Examples of these models include the Online Operations Kill Chain Model(Nimmo and Hutchins, 2023)[7] – whose links include (among others) asset acquisition, coordination and planning, ensuring engagement and longevity) and the DISARM Red Framework (organised into planning, preparation, execution, and assessment).

Example: The deepfake video of Ukrainian First Lady Olena Zelenska allegedly buying a luxury car linked to a Russian disinformation campaign can be coded into the category "develop AI-generated videos" according to the DISARM Red Framework (Mezzofiore, 2024)[8].

- **STIX.** Structured Threat Information eXpression (STIX) is a standardised language using a JSON-based lexicon to express and share threat intelligence information in a consistent and readable format, thus enabling the exchange of cyberthreat information between systems. The Defending Against Disinformation Common Data Model (DAD-CDM) Open Project enhances coordination in counter-disinformation efforts across various domains by establishing a unified data model, building on the OASIS STIX standard with new disinformation-related objects and offering recommendations for its expansion.

Impact assessment

Researchers and experts have long noted the difficulty of measuring the impact of disinformation incidents and campaigns. Technology has enhanced the "seductive illusion of metrics," but "measuring the actual impact of trolling and online influence campaigns is probably impossible" (Rid, 2021, p. 431). Ongoing discussions animate the defender community while some models have been developed.

- **The 'D' and 'E' ABCDE Framework** (Pamment, 2020). The degree of harm and severity of a case, as well as its effect in terms of impact are relevant indicators to assess outreach and influence.

Example: Taking in consideration the EEAS report on FIMI targeting LGBTIQ+ people (EEAS, 2023) once again, attacks were widely disseminated across social media platforms (including private messaging apps), blogs, and news websites, targeting the general public and specific communities to spread disinformation and influence public opinion against LGBTIQ+ rights (i.e., degree). Reported FIMI incidents and campaigns affect trust in democratic institutions and exacerbate social divisions against the LGBTIQ+ community, which may potentially lead to the erosion of legal protection and personal safety (i.e., impact).

- **The Breakout Scale**[9]. Ben Nimmo (2020) proposed to measure the impact of influence operations based on their distribution across communities and platforms.

Example: The gender-based disinformation campaign against top female candidates before the 2021 German federal election falls into "category five" as it entailed multiple platforms, communities, and amplification by mainstream media and public figures (Smirnova et al., 2021).

- **Impact-risk index**[10] (Miguel Serrano, 2022). EU DisinfoLab developed a tool to measure a single hoax's harmful impact and offline risk based on various criteria ranging from media outreach to a call for action.

Example: The doctored nude photo of Annalena Baerbock alleging she used to be a sex worker in her youth gained significant exposure and generated engagement, travelled across platforms, languages, and actors, scoring a high impact-risk (Brady, 2021)[11].

Table 1: SWOT analysis of analysing IBD/FIMI through the existing frameworks

SWOT analysis

These frameworks offer a structured approach to understanding and mitigating the effects of FIMI campaigns, but they also come with challenges and opportunities that the defender community must consider.

STANDARDS FOR IBD-FOCUSED OSINT INVESTIGATIONS

The listed standards focus on suggestions relevant to IBD, though the list is not exhaustive.

- **Berkeley Protocol on Digital Open Source Investigations**[12] (Stover et al., 2022). The protocol suggests ensuring inclusivity and diversity of experts (also integrating the gender perspective) to avoid biases in analysis and language. It mentions the importance of considering differences in using and accessing digital technologies based on identity characteristics in the digital landscape assessment. In addition, it emphasises dedicating resources to the safety and well-being of investigators, who might be at risk due to identity-based characteristics.
- **Guidelines for Public Interest OSINT Investigations**[13] (ObSINT, 2023). The next section will explore these in detail from the perspective of IBD-focused FIMI investigations.
- **OSINT in armed conflict settings**[14] (Millett, 2023). The *vademecum* invites to consider the legal framework (understanding the gaps in international humanitarian law and international human rights law regarding OSINT) and ethical considerations (acknowledging OSINT-related harms, especially concerning vulnerable groups' privacy and data protection).
- **OSINT in cases of sexual violence**[15] (Koenig & Egan, 2021). The paper encourages to evaluate the societal and cultural context and prioritise ethics by respecting dignity, consent, and trauma. They urge to integrate gender and intersectional analyses to address biases, recognise diverse experiences, and be mindful of technology and security-related risks to prevent additional harm during investigations.

HOW TO CARRY OUT AN OSINT INVESTIGATION ON IBD-RELATED FIMI

GUIDELINES FOR IBD-FOCUSED PUBLIC INTEREST OSINT INVESTIGATIONS ON FIMI

Conducting an OSINT investigation for IBD and FIMI requires a systematic approach to ensure diligence, lawfulness, and efficacy. The following steps outline a comprehensive methodology to guide investigators through this process.

Principles

The "Guidelines for Public Interest OSINT Investigations" (ObSINT, 2023) identify five overarching principles: accuracy, community, diversity, accountability, and balance and responsibility.

- **Accuracy:** The investigation has to maintain a gendered, racial, and queer lens to ensure that disinformation campaigns targeting specific groups are accurately identified and addressed. Research processes and the communication of results need to be transparent, replicable, and unbiased. Of course, like in all human-led activities, achieving complete objectivity is impossible, but maintaining rigorous standards and a critical awareness of inherent biases will help mitigate these challenges.

Example: Jankowicz et al. (2021, p. 14) explicitly disclaim that their data collection relied on a series of "colloquial misogynistic slurs in the English language".

- **Community:** Allyship and positive defender community relationships and contributions are vital for a sustainable and inclusive industry. The safety and well-being of OSINT operators, the communities under investigation, and the audience are crucial, especially in the context of disinformation targeting marginalised groups.

Example: The defender community doing OSINT and researching identity- and gender-based disinformation displays good capacity to network and cooperate. To illustrate this, EU DisinfoLab (2022) conducted investigations with other reputable CSOs and wrote about gender-based disinformation with #ShePersisted.

- **Diversity:** Incorporating diverse perspectives and inclusive participation is crucial in OSINT research. It is imperative to maintain an intersectional perspective throughout the investigation and be sensitive towards experiences of discrimination. Teams should include members who are not only skilled in OSINT techniques but also knowledgeable about the unique features of marginalised communities. Understanding the languages and idioms, as well as the norms and subcultures of these groups, helps ensure comprehensive and unbiased investigations, ultimately strengthening the quality and impact of OSINT outputs. Despite the limited resources available to members of the OSINT community, efforts must be made to address this sensitive topic accurately and responsibly. Incorrectly conducted OSINT investigations can be easily weaponised, exacerbating rather than counting harm and misinformation.

Example: A doctored photo of US congresswoman Ilhan Oman was circulated online, showing her without a headscarf (Lead Stories, 2020)[16]. Digitally removing a Muslim woman's hijab equates to nudification. It is a disrespectful act that perpetuates harmful stereotypes, fuelling harassment and discrimination.

- **Accountability:** This is especially important in IBD investigations, where ethical considerations and the real-world effects on targeted communities must be carefully managed to prevent further victimisation and misinformation. Moreover, transparency and openness include acknowledging the drawbacks and limitations of a study.

Example: In July 2016, WikiLeaks published nearly 300,000 emails about President Erdoğan's repressive actions following a failed coup attempt. A significant and dangerous repercussion of this leak was the spreading of numerous women's sensitive personal information, including addresses and contact details. This doxxing led to severe privacy violations and put these women at risk of harassment and violence (Tufekci, 2016)[17].

- **Balance and responsibility:** Investigations must balance technical possibilities with ethical appropriateness, minimising harm to individuals and groups. This involves balancing the right to access information with the right to privacy, ensuring that investigative practices do not inadvertently perpetuate IBD.

Example: The EEAS report on FIMI activities targeting LGBTQ+ communities warns that it covers a limited period and subset of activities that "should not be used to conclude general trends in FIMI" (EEAS, 2023, p. 3).

Scope and objectives

The first step in conducting any investigation is to outline its objectives clearly. Despite a lack of a universal definition, OSINT needs to be guided by public interest. In this regard, researchers may be motivated by the desire to unveil and analyse suspicious activities and identify and stop the actors behind them. While more action is needed to impact the malign actors' capabilities and ultimately deter them, the cost-effectiveness assessment developed by EU DisinfoLab (Miguel Serrano et al., 2024)[18] finds that research and exposure contribute to:

- **Raise situational awareness** of the existence of malign operations and the phenomenon of IBD;
- **Impact the malign actors' capabilities** to produce and distribute the IBD;
- **Advance responses in the defender community** to reduce the distribution of the campaign and risks for the victims being targeted due to their identity;
- **Attribute the campaign** to specific threat actors.

Methodology

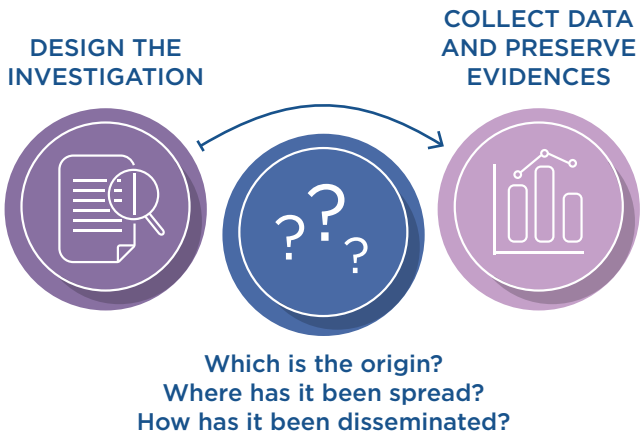


Figure 1: Key steps to conduct an Open Source Investigation

Design

A detailed plan outlining timelines, milestones, resource allocation, and key indicators for identifying and analysing patterns of IBD and FIMI activities should be developed before launching the investigation. Conducting a comprehensive risk assessment is necessary to identify potential legal and ethical issues that could arise. This involves several key components:

- **Awareness of legal constraints.** Familiarising oneself with local, national, and international laws governing digital investigations is essential. This includes ensuring adherence to the General Data Protection Regulation (GDPR) and other relevant legal frameworks.

- **Risk assessment.** Conducting a comprehensive risk assessment is necessary to identify potential legal and ethical issues that could arise during the investigation. This assessment should evaluate the risks to the investigation team, the investigation subjects, and any third parties that might be affected. Developing strategies to mitigate these vulnerabilities is crucial to maintaining the integrity and legality of the investigation.

- **Platform policies.** Familiarity with the platforms' policies being investigated enables compliance with the Terms of Services and understand how these can be exploited for malign purposes. In light of existing national and EU legislations, such as the Digital Services Act (DSA), this awareness favours platform accountability, allowing investigators to identify and mitigate systemic risks.

While FIMI operations span across platforms, the latter lack standard definitions of the phenomena. Most Very Large Online Platforms (VLOPs) refer to "influence operations" and typically link them to state actors. Identity-based attacks are addressed and prohibited under hate speech policies. Mis- and disinformation are usually defined as content determined false and misleading by authoritative third parties, with content removal justified by the risk of imminent physical harm. Although IBD is not explicitly mentioned, it can fall under these categories. However, this leaves out borderline "awful but lawful" content. The table below reports the policies that refer to FIMI or identity-based attacks.

Data collection and preservation

Effective data collection and preservation in OSINT investigations must ensure minimal damage and obtain the best possible evidence. Emphasising the justifiability and reproducibility of actions is crucial to maintaining the

	POLICY REFERRING TO FIMI	POLICY REFERRING TO IDENTITY-BASED ATTACKS
META	Facebook refers to "influence operations", defined as "coordinated efforts to manipulate or corrupt public debate for a strategic goal". Although not explicitly mentioned, it should extend to Instagram, too.	Meta's hate speech policy and Ad Standards prohibit discrimination based on "personal attributes such as race, ethnicity, colour, national origin, religion, age, sex, sexual orientation, gender identity, family status, disability, medical or genetic condition."
YOUTUBE	The platform refers to "coordinated influence operations", for which it does not provide a definition but attributes them to state actors.	YouTube's hate speech policy prohibits discrimination based on "age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigration status, religion, sex/gender, sexual orientation, victims of a major violent event and their kin, veteran status".
TIKTOK	The platform refers to "covert influence operations" as "coordinated, inauthentic behaviour where networks of accounts strategically work together to mislead people or our systems and influence public discussion".	TikTok's hate speech and anti-discrimination ad policy prohibits discrimination based on "race, ethnicity, national origin, religion, tribe, caste, sexual orientation, sex, gender, gender identity, serious disease, disability, and immigration status."
X	Twitter used to speak of "state-linked information operations", but since the platform's rebranding as X, there is no reference to the topic.	X's hateful conduct policy prohibits direct attacks "on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease".

Table 2: How the main VLOPs' policies define FIMI and IBD

integrity and authenticity of potential digital evidence, as indicated by the ISO 27037:2012 "Guidelines for identification, collection, acquisition, and preservation of digital evidence"[19] instruct – hence the following steps.

- **Privacy Impact Assessment (PIA).** Completing a PIA checklist is crucial to ensure compliance with data protection laws and policies, including:
  - The legal basis for the study such as public or legitimate interest and compliance with data protection laws (e.g., Article 6.1 of the GDPR);
  - Data minimisation, pseudonymisation, and anonymisation, implementing adequate security measures, and establishing also appropriate data retention periods;
  - Record all data processing activities in the Treatment Activity Register;
  - Confirm that concerned data subjects have been adequately informed about the data processing activities;
  - Last but not least, it is of the most significant importance, especially when tackling IBD-related matters, to ensure appropriate safeguards for handling special category data – including race or ethnicity, gender, sexual orientation, and minors.

  Moreover, the PIA should be regularly reviewed and updated to reflect changes in the study or data processing activities.

- **Archiving.** The investigation plan should define how to handle data collection and storage securely. Archiving is essential to preserve historical records, hold perpetrators accountable, and ensure access to information. However, ethical considerations are paramount, especially in the sensitive context of IBD. There is a trade-off between immortalising digital content and the right to be forgotten to avoid further victimisation. Therefore, investigations must prioritise public interest, comply with legal standards, and adhere to data minimisation and anonymisation principles.

Finally, major social media platforms have recently tightened their API policies, often citing privacy concerns or commercial interests, limiting the availability of comprehensive, real-time data (European Commission, 2024[20]; Dang, 2023[21]). Such

growing limitations and/or restrictions on platform APIs cause an obstacle to data collection and have significantly hindered access to valuable data. This makes it increasingly difficult for researchers to study online behaviour, disinformation, manipulation and interference threatening information integrity. These restrictions also pose challenges to transparency and accountability, as previously available tools are now out of order for analysts, who are often unable to conduct thorough investigations or validate platform-driven metrics.

ONLINE ECOSYSTEM THREAT ASSESSMENT FOR IBD-RELATED FIMI

Figure 2: Elements to assess IBD/FIMI threats

Who: Targets, perpetrators, and audiences

- **Identify individuals and groups targeted by IBD.** IBD-focused FIMI campaigns target vulnerable groups such as racial or ethnic minorities, LGBTIQ+ individuals, and women, amplifying existing prejudices against them. Three factors trigger identity-based attacks online: non-compliance with traditional gender norms, occupation of previously all-male spaces, and sharpening discrimination from belonging to multiple marginalised groups (Sobieraj, 2019). Moreover, it is essential to recognise these attacks as systematic (as theorised in gender-based disinformation) and coordinated (as mentioned in the definition of FIMI) rather than isolated episodes (Jankowicz et al., 2021)[22].

Example: Foreign interference during the 2016 U.S. elections exploited pre-existing racial tensions to discourage African-American citizens from voting (Zaveri & Fortin, 2019)[23].

- **Profile the perpetrators behind the campaigns.** FIMI actors perpetrating IBD can range from state actors to non-state actors with specific ideological agendas. Attribution can be the most difficult task in any OSINT investigation, as threat actors invest extensive resources to conceal their operations. The next section will explore how OSINT tools can help trace the origins of disinformation campaigns and identify key actors and their connections. A fundamental disclaimer is that endorsement is not attribution; for instance, if a social media account reposts a piece of disinformation, it does not automatically mean this is the source of the content. Furthermore, the lines between FIMI and DIMI (Domestic Information Manipulation and Interference) might be blurred when it comes to IBD, where well-funded transnational networks combining EU and non-EU actors can operate.

Example: A "global network of ultra-conservative organisations" is increasingly operating across Central Eastern Europe (and Europe in general), working to undermine women's reproductive and sexual rights. The movement was founded in Brazil and flourished in the US; Russian oligarch Konstantin Malofeyev was identified as prominent across the network, which also counts on several EU organisations (Janulewicz & Balint, 2021)[24].

- **Analyse the audience consuming the disinformation.** Understanding who is at the receiving hand of the disinformation can provide meaningful insights. Demographic factors and media consumption habits help identify the network contributing to its proliferation.

Example: In the two cases above, threat actors seek to hinder voting and reproductive rights, respectively, evidencing that the audience and the target of the campaign can overlap (Simmons & Martiny, 2024)[25].

What: Assessing threat behaviours

FIMI incidents and campaigns employ adaptive and evolving methods, often shifting in response to exposures, takedowns, and infrastructural disruptions. One way to study these threats is through Tactics, Techniques,

and Procedures (TTPs) outlined in the DISARM Red Framework, which provides valuable insights to explore suspicious behaviours.

Instead of focusing on overly technical details, this analysis takes a descriptive approach to threat behaviours. These include targeting specific audiences by exploiting echo chambers (such as anti-gender forums or pages) and segmenting users based on demographics, beliefs or location (e.g., mobilising a racial divide within a specific community). Disinformation is spread by distorting facts and pushing old and new narratives that prey on social prejudices (misogyny, homophobia, or racism), conspiracy theories – from "gender ideology" to "Eurabia" (Marchlewska & Cichocka, 2020[26]; Brown, 2019[27]) – and current events like elections or world-famous sports competitions).

Both IBD and FIMI rely on all sorts of formats – e.g., visuals, audio, and text – to pollute the information environment with AI-generated and manipulated content, as well as malinformation practices such as doxxing. Malign actors also exploit the architecture of online platforms by creating bots and sockpuppets, pages and groups, and inauthentic and anonymous accounts to target their victims. Common tactics include impersonating authentic sources, delivering misleading ads, and testing platform boundaries with borderline content, especially when platform policies lack explicit prohibitions on gender-based disinformation.

Researchers must enhance IBD reporting, recognising how FIMI attacks majorly affect vulnerable groups based on their gender identity, sexual orientation, and racial or ethnic background. While deepfakes and cheapfakes are widely recognised as tools for manipulation, it is critical to acknowledge that women are disproportionately targeted with deepfake pornography online (Hurst, 2023)[28]. Moreover, special attention should be paid to "awful but lawful" borderline content aimed at identity-based groups, as it can have incendiary consequences.

Where: Mapping online spaces

Threat actors tend to operate across multiple platforms. Mapping the complex online ecosystem of mainstream VLOPs, alternative and fringe platforms with looser content moderation policies, encrypted messaging apps, and crowdfunding platforms (applicable in follow-the-money approaches) allows researchers to reconstruct the amplification loop and grasp the full scope of the operations.

Example: EEAS reported on FIMI incidents targeting LGBTIQ+ (EEAS, 2023) that were distributed through various means, ranging from inauthentic social media accounts and websites to coordinated Telegram posts and publications by state-affiliated outlets controlled by FIMI actors.

When: Timing and events

Incidents targeting individuals based on their ethnicity, gender, and sexual orientation often spike around significant political, social, cultural, or economic events. This includes elections, commemorating dates, emergencies and crises, and international conflicts. Monitoring these periods is key for early detection. Moreover, understanding the correlation between disinformation spikes and specific events helps identify changes in the tactics used. There should be a broad understanding of events, as malign actors can shift their activities in response to initiatives by the defender community (i.e., exposures, takedowns, sanctions). Although the example goes beyond the scope of IBD, this is well evidenced in the Doppelganger operation (Miguel Serrano et al., 2024).

Example: Russian actors spread gender-based disinformation against Moldovan President Maia Sandu in the context of the 2024 elections (#ShePersisted, 2024)[29].

Why: Evaluating objectives of IBD

The motivations behind IBD-fuelled FIMI attacks are multifaceted, potentially causing real-life harm and long-term negative consequences. Malign actors can have a political agenda to influence elections, undermine democratic institutions, and convince candidates to withdraw. Ideological reasons linked to promoting a particular belief system are common – from fuelling Islamophobia to opposing sex educations in schools. Minorities and marginalised communities offer the ideal scapegoat to blame for all evils, exploiting pre-existed social fractures to "divide and rule" by polarising public opinion and shifting attention from more complex economic, political, and social causes and responsibilities.

Example: Drawing on the previous example, the disinformation campaign targeting Moldovan President Sandu wished to undermine her pro-Western leadership, including support for EU membership.

CATALOGUE OF OSINT TOOLS AND METHODOLOGIES FOR IBD-RELATED FIMI INVESTIGATIONS

This section outlines easily accessible tools for addressing IBD-focused FIMI incidents and operations. It covers tools for archiving evidence, assessing coordination, verifying authenticity, identifying sources, and measuring impact, all of which provide valuable insight into the 5Ws (i.e., who, what, where, when, and why) that should guide any investigation. Additional repositories and tools can be found in the Annex.

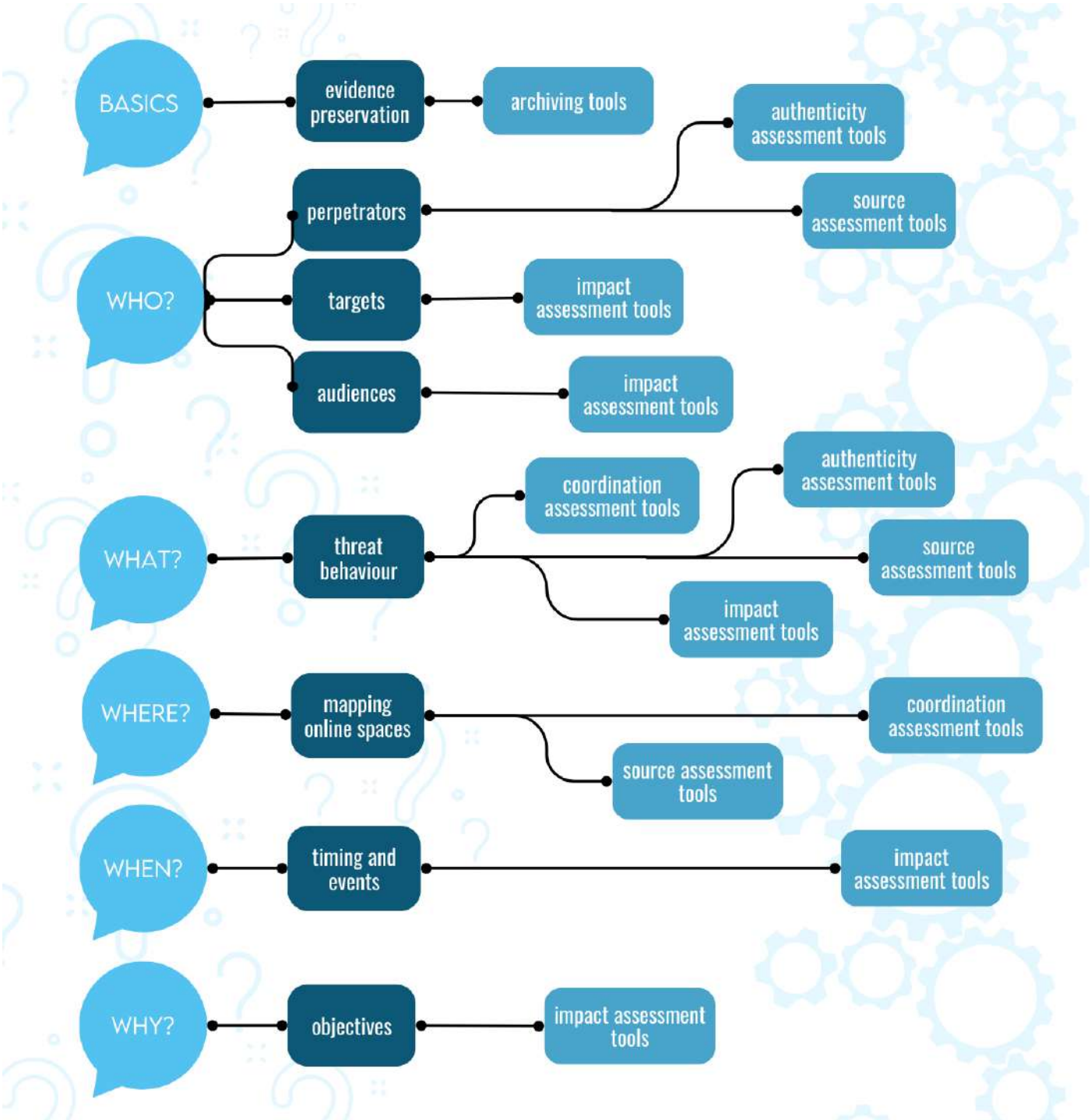


Figure 3: OSINT tools based on online ecosystem threat assessment tree chart

ARCHIVING TOOLS

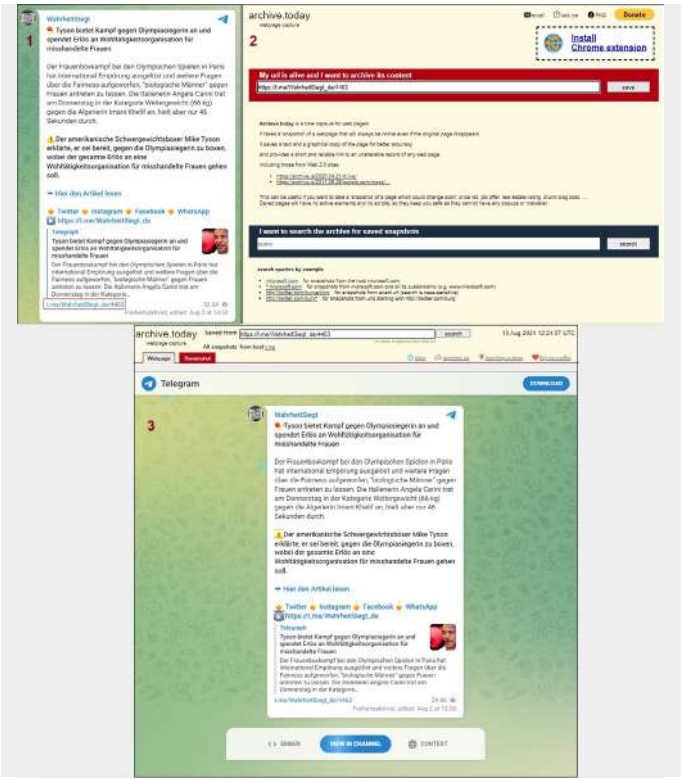
Archiving webpages

A screenshot alone is insufficient for preserving digital evidence, as website source code can be easily altered, leading to manipulated images. Therefore, archiving online content is the only way to retain digital evidence while minimising manipulation risks. An important disclaimer is that there is no one-tool-fits-all for content. For instance, sometimes it is not possible to archive social media links directly, making it necessary to download embedded media (such as videos) or use platform-specific tools. Moreover, to be archivable, the content must be available online. Overall, it may take a few attempts with different tools to succeed, hence the importance of managing frustrations.

Free archiving tools that allows people to create archives and visit archived versions of websites:

- Archive.today;
- Ghost Archive;
- The Wayback Machine, as of mid-October 2024, this Internet Archive service is available in read-only mode after a cyberattack (Warren, 2024)[30].

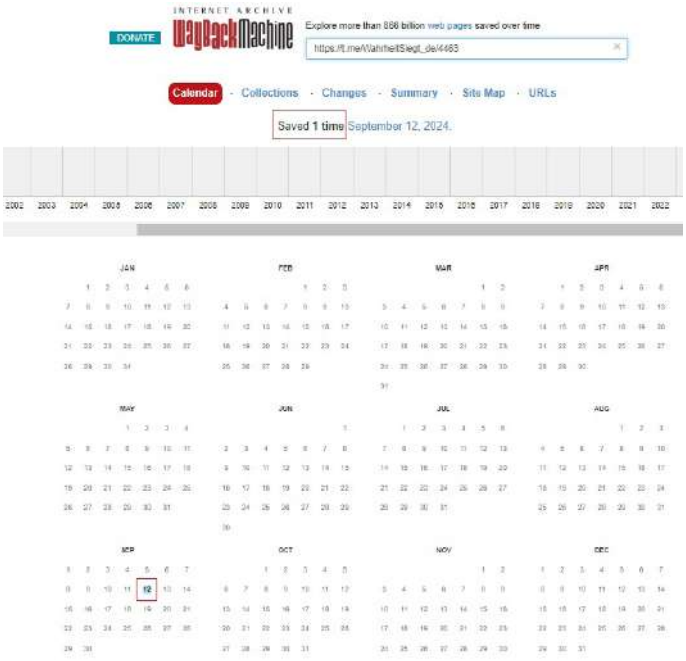
These archiving tools – a more exhaustive list is provided in the Annex – are very straightforward: having the original URL is enough to create an archived version of the desired webpage.



Example: To illustrate the process, users can access "Archive.today" and paste a URL, for example, the hoax debunked by German fact-checker Correctiv during the Paris Olympic Games, questioning female Algerian boxer Imane Khelif's gender, shared, among other sources, on Telegram (Marinov, 2024). Therefore, it will be enough to paste the URL "t.me/WahrheitSiegt_de/4463" on the red box "My url is alive and I want to archive its content". The archived link is available at "archive.is/EZqYq", indicating the archiving date and original link.

Avoiding archiving duplications

To avoid duplications, these tools will suggest the already archived version for a certain webpage if it already exists in the library. In "Archive.today", users can check for it by pasting the original URL into the black box "I want to search the archive for saved snapshots".



Similarly, WayBack Machine compiles a calendar containing all the archived versions of the same webpage at different times. This helps track changes in frequently edited content, such as platform policies against misinformation.

Archiving social media content

This practice presents challenges, particularly with embedded media such as images and videos. The archiving tools described above may not always be effective, making downloading and storing the media offline necessary, detaching it from the post's original context. Useful information can be found in the "Activists' Guide to Archiving Video" (Witness, 2023)[31].

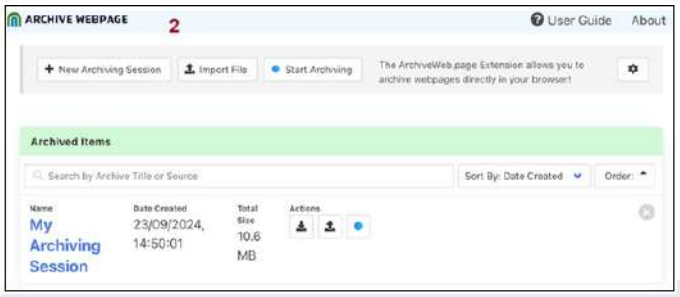
Free tools for archiving content from specific social media platforms:

- The already mentioned Ghost Archive is preferable when archiving social media posts.
- FDOWN.net is a free tool designed to download Facebook videos and save them directly to computers or mobile devices. This can be done by entering the link, which can be retrieved by clicking on the three dots at the top right of the publication and selecting "copy link".
- Getfvid converts videos to MP4 (video) or MP3 (audio) files and enables free downloads, functioning across computers, tablets, and mobile devices.
- SnapSave provides similar downloading capabilities for Facebook and Instagram videos.
- Savefrom.net is recommended for downloading YouTube videos.
- SSSTwitter is a video downloader tool for X.
- Telegram's simplest method to archive media is through the desktop application. Users can right-click on an image or video and select "Save as…" to store it. If the image is attached as a file, it retains its original metadata, as noted by Bellingcat (2022)[32].

Archiving entire websites or sessions

In addition to archiving URLs, users might want to archive entire websites or record an entire session for research purposes. Although it requires more advanced technical knowledge, some options are:

- HTTrack.com to download entire websites for offline use.
- Stone, a "research transparency" app that captures desktop research using screen captures and webcam commentaries.
- WebRecorder, a suite of open-source tools and packages, to capture interactive websites and replay them at a later time as accurately as possible.
- Wget to download entire websites for offline use and resume interrupted downloads.



Example: Users can choose the "ArchiveWeb.page" Chrome extension from WebRecorder,[33] which is also available as a standalone desktop application. Once installed, the recording can be started by clicking on the extension, selecting "Start Archiving," and customising the session name (for the sake of this example, "My Archiving Session"). The Autopilot feature is suggested when archiving websites that feed content at each scroll, such as social media. The tool will start archiving every webpage the user visits. A useful disclaimer is that only loaded content will be saved; therefore, opening images, videos, and embedded links is recommended to archive them successfully. Archived webpages can be accessed via the "View Archived Pages" button, displaying each session's creation date and total size.

COORDINATION ASSESSMENT TOOLS

A key element of a FIMI campaign is the coordination of activities conducted by various agents working together towards a shared goal. In the context of IBD, this systematic organisation of tasks and efforts can serve various purposes, such as silencing the targets or polarising the audience, and rely on various deceptive tactics, like giving the false impression of organic content amplification. Moreover, coordination is a critical component of Coordinated Inauthentic Behaviour (CIB), a practice that social platforms usually ban but can also be applied in FIMI campaigns (Romero Vicente, 2024[34]). Detecting coordination poses significant challenges and requires strong observational and technical skills. The following indicators may assist in identifying this behaviour.

	Description
Temporal indications	▪ Accounts created around the same time ▪ Similar posting timestamps across different accounts ▪ Accounts engaging with each other's posts in a synchronised manner ▪ Sudden spikes in messaging around a certain narrative or events
Content indicators	▪ Narrative alignment, including identical or similar hashtags, images, links, memes, posts, texts or videos ▪ Same or similar content translated and posted in different languages ▪ Accounts focused on a single topic ▪ Consistent posting patterns across various social media platforms
Relational indicators	▪ Same or similar profile images and cover photos ▪ Tightly interconnected clusters of accounts that mostly follow each other
Technical indicators	▪ Multiple accounts sharing the same IP address, analytics IDs, devices, and configurations ▪ Centralised content production
Automation indicators	▪ Presence of bots ▪ Automation practices in content publication

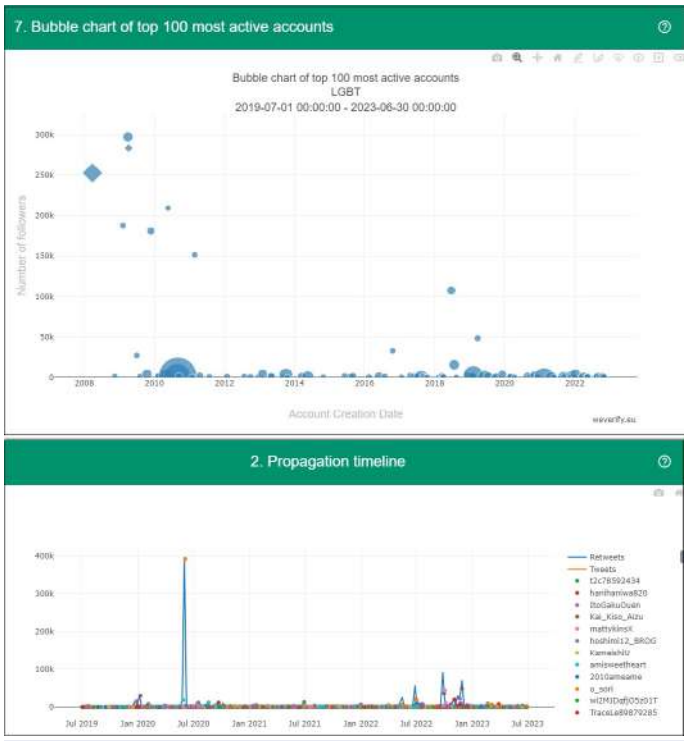
Table 3: Indicators of coordination of IBD-related FIMI activities

Temporal indications

The creation date of an account can be checked manually. This information is typically available in the bio, profile details, or transparency page on Facebook – though availability varies by social network. If the creation date is not explicitly listed, an approximate date can often be determined by the timing of the first publication. However, conducting a manual search can sometimes be tedious. Some tools capture the creation date of several accounts at the same time.

An instrument to grasp temporal insights into accounts:

- The InVID-WeVerify verification plugin is a powerful means to spot disinformation online with multiple functionalities, such as capturing the creation date of several accounts at the same time.



Example: InVID-WeVerify verification plugin's data analysis functionality allows the retrieval of creation dates for the top 100 most active accounts (shown at the top), as well as their posting timelines for content using the term 'LGBT' (shown at the bottom) between July 2019 and June 2023 – the tool cannot fetch any tweet since 1 July 2023 due to API restrictions. The simultaneous or closely timed creation of several accounts, combined with similar posting patterns, often suggests a coordinated campaign. Additional signs of such coordination include sudden surges in messaging around a specific narrative or event, as well as synchronised interactions between accounts, such as spikes in tweets containing "LGBT" in July 2020, when accounts engaged heavily by reposting or liking one another's content.

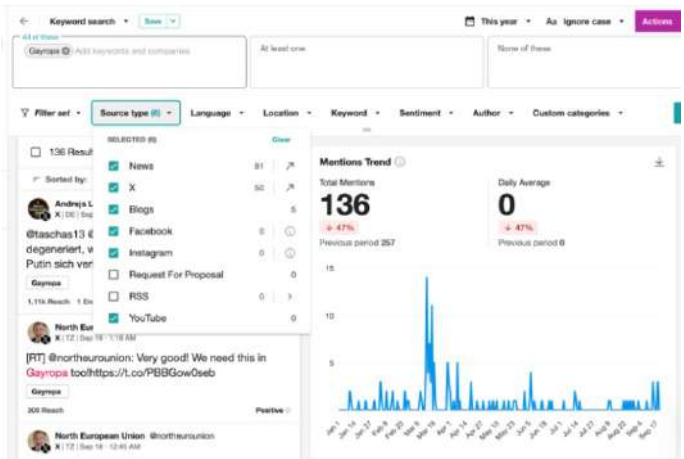
Example: Identical posting patterns can be a sign of coordination. These accounts from the Doppelganger network used the same profile photos and posted the same content on 10 July 2022 at 4:35 pm in French and at 4:49 pm in Italian.

Content indications

Narrative alignment is one of the strongest indicators of coordination, often evidenced by the use of identical or similar text, hashtags, links, images, or memes. This alignment may also appear in comparable posting patterns across multiple social media platforms, including similar post structures. Coordinated content can be translated and shared in different languages, and suspicious behavior may include accounts focusing exclusively on a single topic. Detecting content alignment involves searching for these elements across various social networks and search engines to identify shared publication sources.

Tools to search content:

- The Google Hacking Database (GHDB) – also known as Dorks – offers a set of operators used to filter information within the Google search engine, are particularly helpful for refining these searches.[35] For instance, if the user wants to find tweets containing the conspiracy keyword "gayropa", it will be sufficient to Google the appropriate query, i.e., site:twitter.com "gayropa".
- The Google Custom Search is a customised Google search engine that allows users to conduct searches on different platforms simultaneously, although results will be more or less accurate depending on the platforms' indexation. Similar search operators exist also for X.
- While some free dashboards that could be used to check the spread of content on different platforms are no longer available, a good – but paid – alternative for social media monitoring is Meltwater.
- CooRnet is another social media monitoring tool for detecting coordinated link-sharing behaviour (CLSB), as it generates a network of entities involved in such activities.



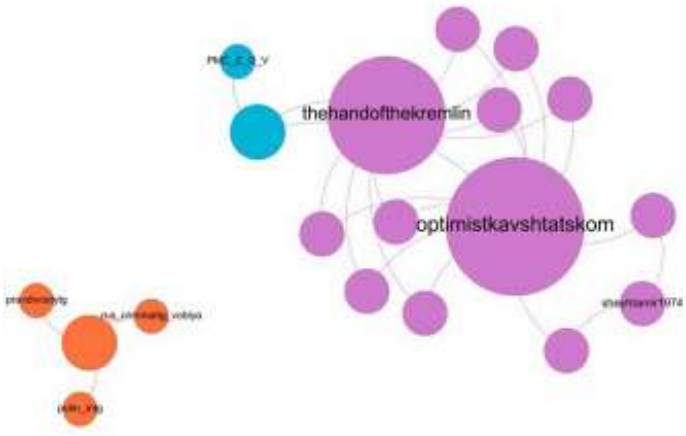
Example: The figure shows how Meltwater can be customised to monitor the spread of the anti-gender term "Gayropa" on different platforms such as news outlets, X, blogs, Facebook, Instagram, and YouTube.

Relational indicators

Other visual indicators, such as the use of the same or similar profile images, avatars, and cover photos, may constitute strong evidence of the connection between accounts. A strong relational indicator among accounts is their interaction, characterised by parameters such as following and engagement. Tightly interconnected clusters of accounts that predominantly follow each other and participate in a campaign provide evidence of coordination.

Tools to extract relational indicators:

- Twiangulate is a tool to search, analytics and mapping of connections between X friends and followers.
- Followerwonk extracts relationships between X accounts, including searching bios, comparing users, analysing, tracking and sorting followers.
- The InVID-WeVerify verification plugin facilitates research on relationships within X, Facebook, and Instagram. It provides with a social network analysis (SNA) feature with two data analysis tools: one for X (functioning with past data until 1 July 2023) and another for Facebook and Instagram (based on CSV files exported from CrowdTangle, but deprecated since 14 August 2024).
- Advanced researchers can also scrape data from other platforms, such as Telegram, download a CSV file, and import the data into Gephi to investigate account relationships further. This approach allows the retrieve of specific accounts involved in a campaign and their connections.



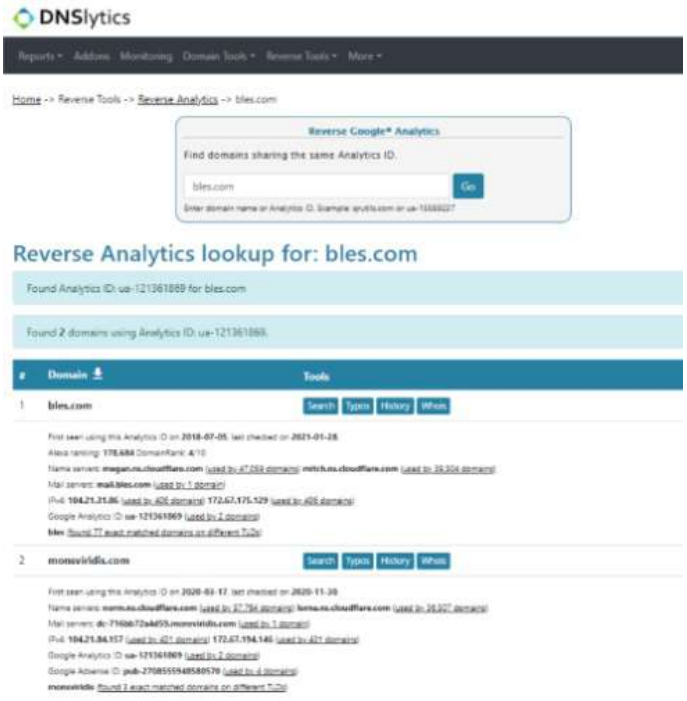
Example: The chart shows clusters of Telegram accounts participating in Operation Overload that were identified using Gephi (Atanasova et al., 2024)[36].

Technical indicators

Using the same IP address, device, or configurations, or engaging in centralised content production, are strong coordination indicators. However, identifying these requires technical expertise and may raise privacy concerns. A simpler method is to examine Google IDs. If multiple pages share the same Analytics or AdSense ID, it suggests they are connected. A reverse search of a Google Analytics ID can reveal all sites using the same ID, indicating they belong to the same entity. However, recent updates in Google Analytics 4 have made thisprocess more complex (Silverman, 2023)[37].

A tool to grasp technical indicators:

- DNSlytics is a web-based monitoring tool offering comprehensive DNS and domain research services.



Example: Using DNSlytics, an Analytics ID search reveals a connection between two sites from EU DisinfoLab's investigation on "Tierra Pura", which share the same ID (Miguel Serrano, 2021)[38].

Automation indicators

Bots can facilitate coordination and streamline processes (further details on bot detection can be found in section 3.3 on authenticity assessment). Social platform dashboards can also be used to schedule post publication. Yet, while automation may suggest coordination, it is not definitive proof of it. For example, a single actor may automate the publication process in an uncoordinated attack.

Although analysing all these factors individually is important, there are platforms that allow for the integration of all available evidence to provide a comprehensive overview of the situation and visualise potential points of coordination, assisting researchers with network analysis.

Tools to help detecting coordination:

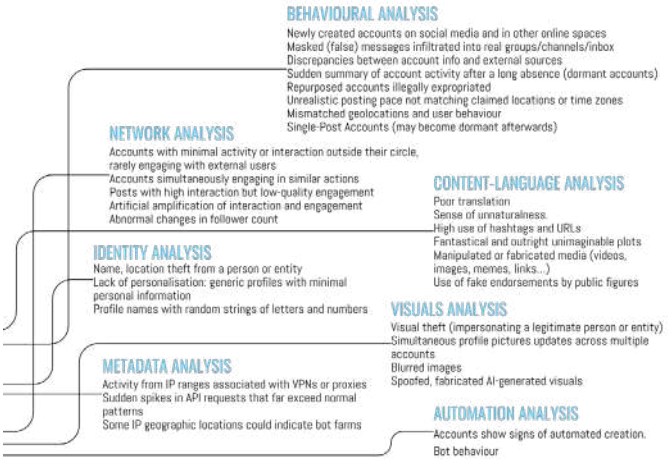
- Maltego is an investigation platform that allows to mine, merge, and map all the available data.
- Cytoscape is an open-source software platform for visualising complex networks and integrating these with attribute data.
- NodeXL is a Microsoft Excel plugin facilitating data collection, storage, analysis, visualisation and reporting.

AUTHENTICITY ASSESSMENT TOOLS

The manipulative nature of FIMI campaigns can take many forms, including false content, inauthentic actors, and deceptive behaviour – e.g., coordination to simulate engagement inorganically (as explained in section 3.2 on coordination assessment). Therefore, checking the authenticity of the shared content and the channels that spread it is crucial to any IBD-focused investigation.

Before delving into the tools, researchers should be aware of fact-checking databases to see if an image, video, document, or quote has been fact-checked:

- The Database of Known Fakes (DBKF) is search tool that enables users to find relevant debunks not only through keyword search but also based on multi-media content.
- Google Fact-Check Tools allows users to search for debunked stories and images, as well as add ClaimReview markup to their own fact-checks.
- EUvsDisinfo database is a searchable, open-source repository of pro-Kremlin disinformation pieces.



Example: As technical tools only provide some of the answers, observation-based indications can offer fundamental leads to identify suspicious activities (Romero Vicente, 2024).

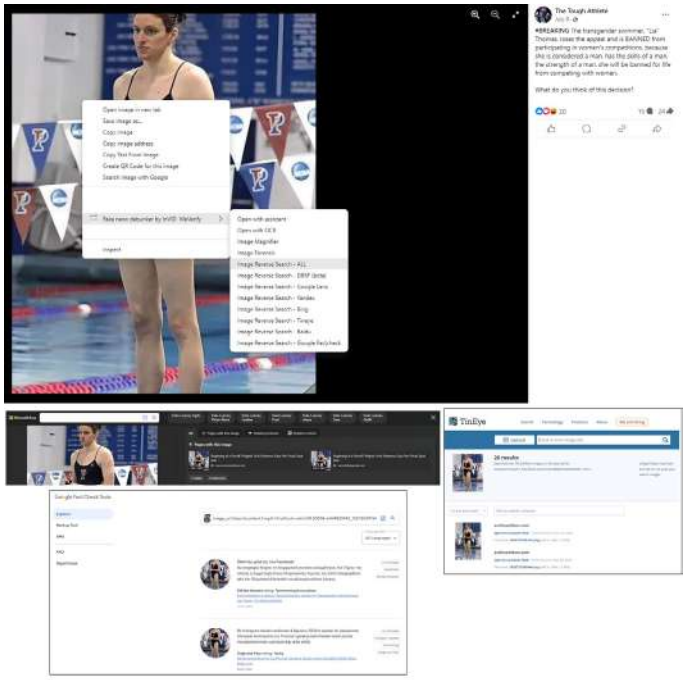
Reverse image searching

One of the easiest ways to verify the authenticity of a photo is to perform a reverse image search to see if the same image has appeared online before and, if so, in what context. Multiple search engines allow this, with similar features and minor variations. Other tools perform this task in different engines simultaneously, saving many steps.

Tools for reverse image searches:

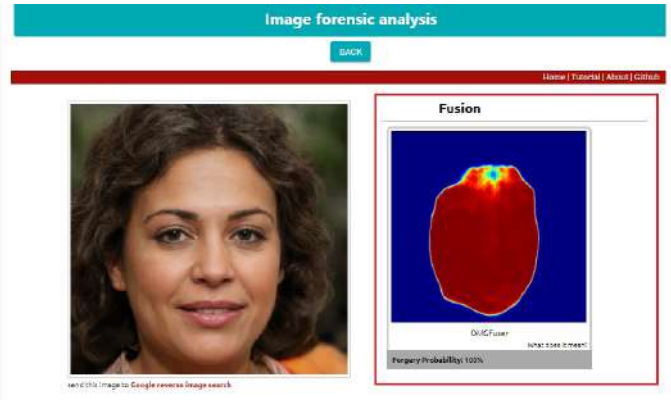
- Bing (Microsoft) is particularly accurate for checking the details of an image.
- Google Reverse Image Search assists researchers in pinpointing the origins of specific images, delving into their context.
- TinEye is especially useful for finding the first time that an image was published.
- Yandex was once considered a useful resource, but its results have recently been noted for declining accuracy. Researchers have raised concerns about its credibility, particularly following its strengthened ties with the Kremlin (Kravets-Meinka, 2024)[39].
- RevEye is a navigator extension that searches on Bing, Google, Yandex, and TinEye simultaneously.
- The InVID-WeVerify verification plugin adds Google Fact-Check, Baidu, and the Database of Known Fakes (DBKF) to the previous list and provides a forensic analysis tool for pictures. This makes the verification plugin the most comprehensive tool for reverse image search available today.

To conduct a reverse image search on any of these search engines, one simply has to copy the image's URL or upload it from a local file. The best solution for simultaneous searches in several search engines is downloading the RevEye or InVID-WeVerify browser extensions.



Example: A previously fact-checked Facebook post (Marinov, 2024) features a doctored photo of trans swimmer Lia Thomas, altered to show visible male genitals. Using the InVID-WeVerify verification plugin, users can right-click the image to access all search options for verification. In the second search, Google Fact-Check Tools is applied, and in the third, TinEye sorts results by the oldest, simplifying the process of finding the original source.

Another recommendation is to verify social media profile pictures to see if they are stolen or fake, as this can be a strong indicator of an inauthentic account or an actor attempting to conceal their true identity.

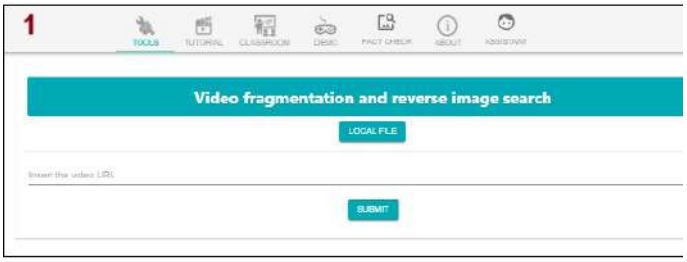


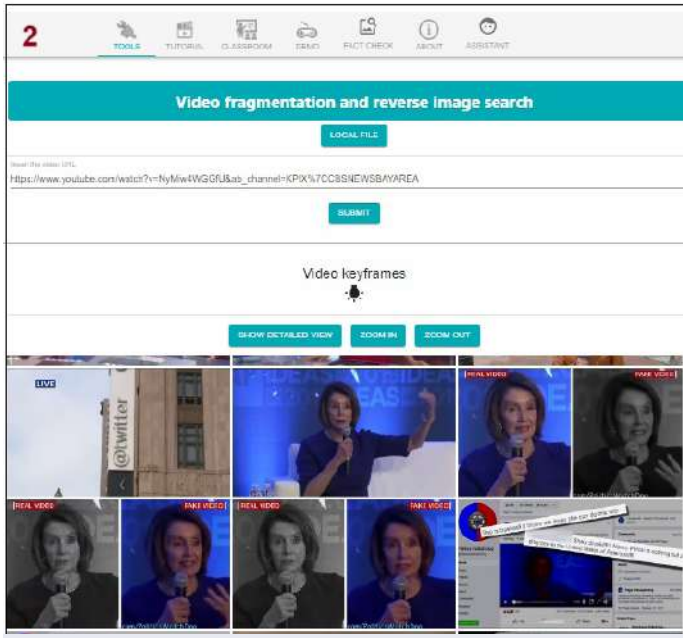
Example: InVID-WeVerify verification plugin's forensic analysis shows that the portrait has a 100% forgery probability, as it was in fact taken from "thispersondoesnotexist.com", a website that generates realistic AI-generated portraits.

Checking video keyframes

Reverse image searches can also be applied to videos to detect potential manipulation:

- The InVID-WeVerify verification plugin video analysis section proves especially valuable, as it breaks down video sequences into individual frames.





Example: The InVID-WeVerify verification plugin fragmentation functionality returns a series of video keyframes that can then be subjected to reverse image searches, enhancing the ability to detect alterations or verify authenticity.

Checking forensics and metadata

A reverse image search is just one way to verify an image's authenticity; forensic image analysis and examining metadata also provide valuable insights. Metadata can reveal important contextual information, such as the date and time when an image was taken, the camera settings, and even the GPS coordinates – also known as the EXIF (i.e., exchangeable image file format).[40]

However, this is not always possible, as some images lose their metadata when uploaded to social networks or when users remove them. Platforms such as Facebook, Instagram, LinkedIn, or X usually hide or delete the metadata of uploaded images, unlike Foursquare, Flickr, Pinterest, and VK. Telegram hides the metadata if the image was uploaded in a compressed version, but the metadata will be visible if the image was uploaded uncompressed. Various instruments support forensic analysis and metadata extraction, although interpreting the indicators returned by these tools requires some expertise.

Forensic analysis tools:

- FotoForensics.
- Forensically.
- InVID-WeVerify verification plugin.

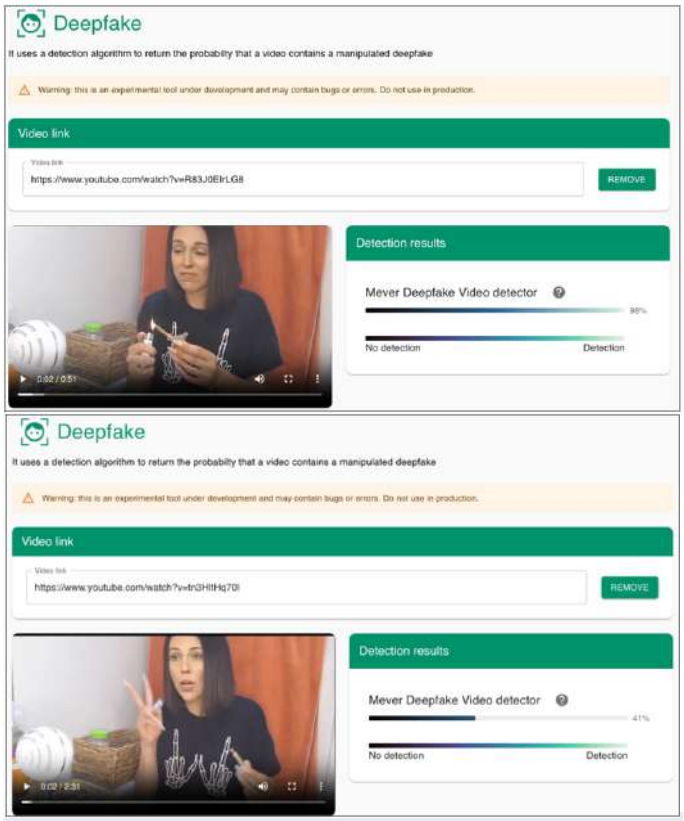
Metadata extraction tools:

- Brandfolder.
- ExifPurge – for removing metadata from photos before uploading them to social media.
- Metadata2Go
- InVID-WeVerify verification plugin.

Checking AI-generated content

Several tools exist to detect AI-generated content:

- Deepware Scanner scans suspicious videos to find out if they have been synthetically manipulated.
- Scribbl detects texts generated by the most popular AI tools, such as ChatGPT, Gemini, and Copilot.
- The InVID-WeVerify verification plugin provides tools for detecting synthetic images, deepfakes, and synthetic audio. All of them are currently in beta, meaning they are still in a testing phase that may have bugs or limited features as development is ongoing. Therefore, the tool is free, but users must apply for access.



Example: The InVID-WeVerify verification plugin provides two crucial features for detecting synthetic images and deepfakes. To illustrate this, two fragments of a YouTube video claiming that former New Zealand Prime Minister Jacinda Ardern was smoking cannabis were tested (Thompson-Fuller, 2021)[41]. On the left, it can be seen that there is a 98% probability that the video is a deepfake, which is true. On the right, another scene from the same video returns a less satisfactory result, highlighting that there is no silver bullet regarding these tools.

Checking account inauthenticity

Another sign of inauthenticity is the possibility that an account or profile is a bot. The main challenge here is not the lack of tools but their accessibility. Many popular instruments to assess the automated nature of X accounts have been entirely or partially deactivated following changes to the platform's API.

Tools to assess bots and automated behaviour:

- BotSentinel is a free platform developed to classify and track inauthentic accounts and trolls. It continues to offer some features, though its access to X's API was revoked in 2022.
- Botometer X calculates the likelihood an account is automated (i.e., a bot). It still functions in a limited capacity, operating in archival mode with pre-calculated results based on data collected before 31 May 2023.

SOURCE ASSESSMENT TOOLS

In any FIMI case, especially when tackling sensitive matters such as IBD, one of the most critical and challenging questions is identifying who is behind a given piece of content, attack, or campaign. Considering the foreign nature of these operations, it is often likely to find clues pointing to foreign actors, such as foreign usernames, traces of third languages and poor translations, or significant activity from locations or timezones that do not align with the user base. Yet, references to the source will not always be this obvious, as threat actors may use proxies (i.e., like-minded local groups that share their objectives), often to target individuals or groups based on their identity.

Getting closer to the source

While determining the exact origin of a campaign can be complex, the goal is to get as close as possible. Some key recommendations are:

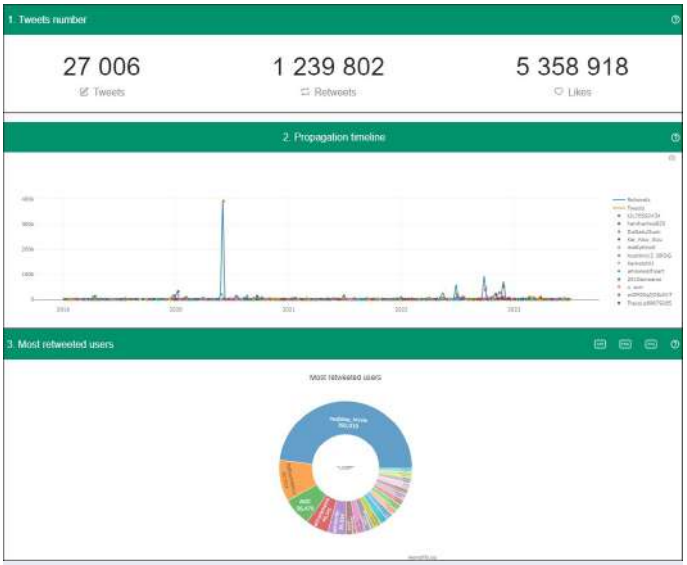
- If multiple posts spread a hoax or attack, researchers should sort them by publication date and focus their research on the earliest one. Additionally, consulting fact-checking databases can help determine the propagation loop.
- Rely on network analysis tools to examine the accounts with the most connections and influence in spreading the message.
- Once key accounts are identified, it is advisable to continue the investigation using available data such as usernames, names, photos, or email addresses.
- Search for the message on different social networks, platforms, or forums to trace possible alternative origins.
- In many instances, technology alone cannot provide a complete solution. Context analysis and visual observation to identify recognisable backgrounds or locations are essential for uncovering who may be behind an attack or disinformation operation.

Network analysis

Performing at least a basic network analysis is crucial to uncovering perpetrators. While social network analysis can be technically complex, several platforms provide a useful starting point.

Tools to support network analysis:

- The InVID-WeVerify verification plugin includes a social network analysis (SNA) tool for X; although it was deprecated in July 2023 due to API access restrictions, it still allows searches up to that date. For Facebook and Instagram, the tool supports SNA based on CSV files exported from CrowdTangle (deprecated in mid-August 2024), enabling research on relationships within those social platforms. Features like the propagation timeline help identify the first posts and associated usernames; examining the most retweeted users highlights the key actors in the distribution network. Additional charts, such as the most active users or the most mentioned accounts (including account creation dates), provide further valuable insights.
- Advanced researchers can also download a CSV and bring the data into other SNA tools, such as Gephi, to explore the relationships between the accounts further.



Example: The InVID-WeVerify verification plugin X data analytics tested for the keyword "LGBT" between July 2019 and June 2023 displays various information including the number of posts, reposts, and likes; the keyword's propagation timeline, and most reposted users.

Investigating malicious actors with publicly available data

After identifying accounts that may be significant in an attack or campaign, various tools can help to investigate further the actors behind the malicious activity. The range of available resources is extensive, and should always be used within legal and ethical frameworks.

Tools to check domain and username use and availability across social media:

- Instant Username Search.
- Namech_k.
- UserSearch.

Tools to investigate other online assets:

- Epieos, for email and phone reverse lookups.
- FreeCarrierLookup.com allows to perform a free phone carrier lookup on any phone number in any country and check which country and service provider a phone number currently is allocated to. It offers additional details such as the phone number line type such as cellular, wireless, landline, VOIP (virtual number), or pager services.
- Hunter, to find and verify professional email addresses.
- Truecaller provides a caller ID and spam blocking software, as well as conducting research on call and SMS harassment.
- WhocalId offers a caller ID feature to identify instant calling and message via phone, Viber and WhatsApp.

Additional tools and tips for other identifiable information:

- IntelTechniques offers a comprehensive suite of OSINT resources for researching individuals.
- Manual searches in search engines or specific social platforms can also yield valuable results.

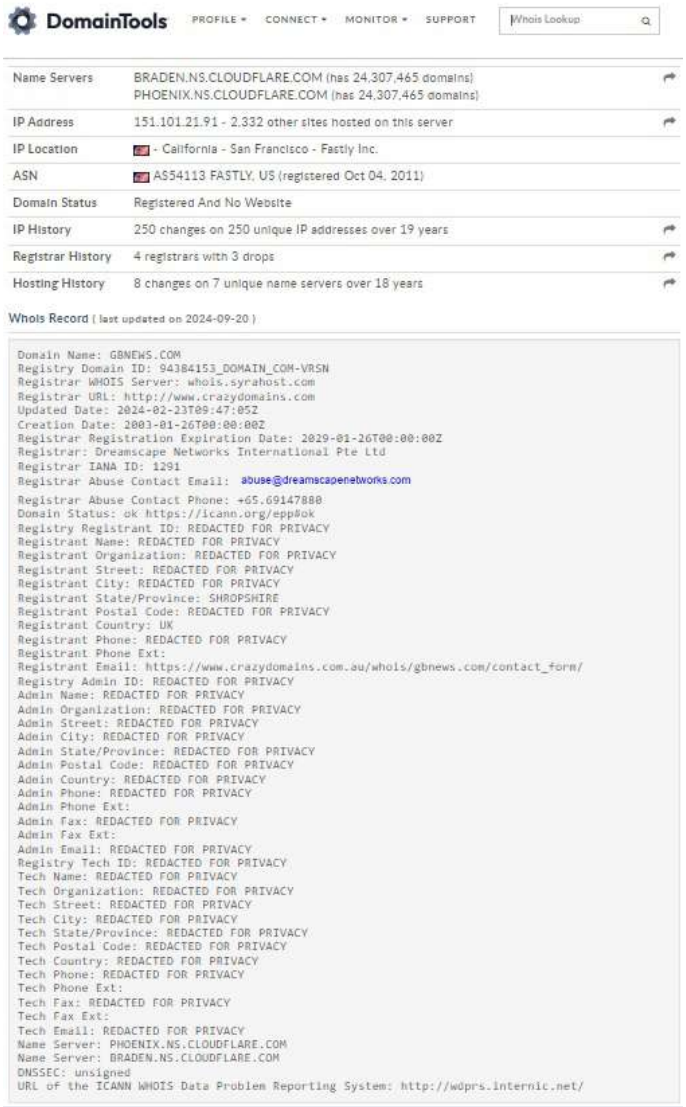
Investigating websites and domains

Several resources can aid in further investigation when the web domain is available. These tools help verify domain registration details, check global domain name service (DNS) records, and identify servers and devices linked to the disinformation network. They also enable researchers to trace the ownership and history of domains involved in spreading disinformation, helping to uncover the authors behind an incident or campaign.

Tools to investigate domains, websites, and IPs:

- Censys.
- DNSlytics.
- DomainTools.
- ICANN.
- WHOIS.

These tools typically operate by requesting the domain to investigate or, in the case of a reverse domain lookup, by using elements like an IP address or an analytics tracker (as mentioned earlier with tools like DNSlytics). It is always recommended to consult multiple sources to gather the most comprehensive information possible.



Example: To investigate the domain of a website that published a transphobic hoax about the opening of the Paris Olympics, one would simply enter the outlet's name into the search box (NewsGuard's Reality Check, 2024)[42]. DomainTools shows the domain record for GB News' website, responsible for spreading the false information.

Additional clues can often be found in metadata (as explained in section 3.3 on authenticity assessment). Tools that analyse metadata associated with images or documents can reveal important insights regarding the information source, such as location details, including significant activity from locations or timezones that do not align with the user base, a strong hint of foreign involvement.

IMPACT ASSESSMENT TOOLS

The motivations behind IBD-fuelled FIMI attacks are complex, potentially leading to real-world harm and long-term negative consequences. Identity-based disinformation primarily impacts vulnerable groups such as women, LGBTIQ+ individuals, and racial or ethnic minorities. In some instances, the goal is to push these groups out of the public sphere; in others, it is to disrupt peaceful coexistence and weaken democratic societies based on equality, tolerance, and solidarity. Nonetheless, measuring the full extent of the damage and establishing a direct cause-effect relationship is highly challenging and rarely feasible.

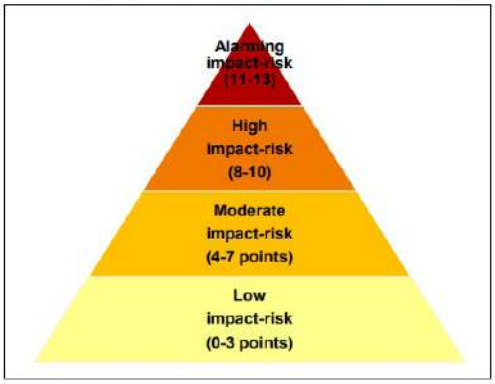
Despite these difficulties, certain indicators can offer insight into the potential impact of hoaxes, incidents, or campaigns. These indicators are mainly based on the outreach and distribution of harmful content but also consider factors such as direct calls to action, which can draw a clearer link between an online message and an offline event.

- The **Breakout Scale** is well-known among researchers. There, Ben Nimmo (2020) proposed measuring the impact of influence operations based on their distribution across communities and platforms. The severity increases as the operation moves beyond closed communities or a single platform and reaches a broader audience, including amplification by public figures. At its peak, the campaign may include a call for violence and demand a policy response.

Example: During the 2021 German federal election, multiple platforms, communities, mainstream media outlets and public figures amplified a foreign-led disinformation campaign targeting Green Party candidate Annalena Baebock (Smirnova et al., 2021)[43]. The IBD case would be classified as "category five" on the Breakout Scale.

- On a smaller scale, to assess the impact of individual hoaxes, EU DisinfoLab developed the **impact-risk index** (Miguel Serrano, 2022). This methodological tool measures a single hoax's potentially harmful impact and offline risk based on various criteria ranging from media outreach to a call to action. By evaluating these indicators and assigning different scores, the tool produces a calculated impact-risk assessment of the hoax, categorised as low, moderate, high, or alarming.

INDICATORS	MEASURES	POINTS
1. Engagement	0 – 1.000 shares and reactions = 0 points 1.001 – 10.000 shares and reactions = 1 point 10.001 – 100.000 shares and reactions = 2 points More than 100.001 shares and reactions = 3 points	0-3
2. Exposure	0 – 1.000 views = 0 points 1.001 – 10.000 views = 1 point 10.001 – 100.000 views = 2 points	0-2
3. Number of platforms	Content shared on one or two platforms = 0 points Content shared on more than two platforms = 1 point	0-1
4. Number of languages	Content circulated in one language = 0 points Content circulated in more than one language = 1 point	0-1
5. Media outreach	The content did not reach mainstream media = 0 points The content reached at least one mainstream media = 1 point	0-1
6. Type of actor	The transmitter/amplifier is not a public figure of any sorts = 0 points The transmitter/amplifier is a public figure and/or a recurrent disinformer who has been fact-checked before = 1 point	0-1
7. Formats	Content spread in one format exclusively = 0 points Content spread in more than one format = 1 point	0-1
8. Call to action	The content does not contain any exhortation = 0 points Engagement is 0. Then, engagement x exhortation is 0 x 0 = 0 call to action: 0 points Engagement is 1. Then, 1 x 0 = 0 -> call to action: 0 points Engagement is 2. Then, 2 x 0 = 0 -> call to action: 0 points Engagement is 3. Then, 3 x 0 = 0 -> call to action: 0 points The content contains an exhortation = 1 point with multiplier effect: Engagement is 0. Then, engagement x exhortation is 0 x 1 = 0 call to action: 0 points Engagement is 1. Then, 1 x 1 = 1 -> call to action: 1 point Engagement is 2. Then, 2 x 1 = 1 -> call to action: 2 points Engagement is 3. Then, 3 x 1 = 3 -> call to action: 3 points	0-3 with multiplier effect



Example: Correctiv debunked the claim that Mike Tyson wanted to fight Imane Khelif as a way to falsely imply that she is not a woman (Marinov, 2024)[44]. The hoax obtained 5.4k likes on Threats alone (1 point for engagement) and over 30k views between X and Telegram (2 points for exposure). Spread on multiple platforms (1 point) and formats (1 point), the hoax would be considered to have a moderate impact-risk of the index.

Outreach and monitoring tools

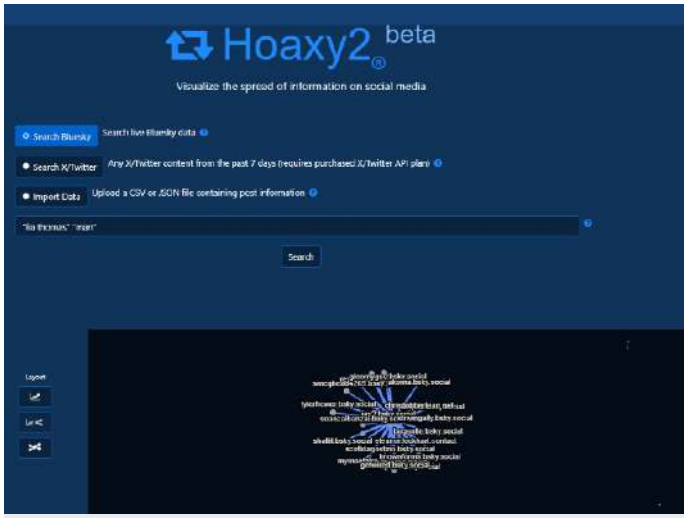
One of the key aspects of these methodologies is evaluating the reach of a given piece of content. The right tolls make it easier to search for key elements like interactions, responses, and engagement with the original content, as outlined in these guidelines. These tools are also valuable for conducting deeper content investigations, helping to uncover the source, the intentions of bad actors, potential calls to action, and even whether public figures played a role in amplifying the message.

Outreach tools:

- BuzzSumo tracks content virality and analyses social media content performance.
- Hoaxy visualises the spread of information across several platforms and tracks claims and fact-checks dating back to 2016. Specifically, it extracts data from Bluesky and X, while also allowing data import from other platforms.
- Open Measures is useful for exploring the outreach of a piece of content on fringe platforms such as 4chan, Gab, or Gettr.
- Until mid-August 2024, Meta supported CrowdTangle, an incredibly valuable tool for tracking virality on Facebook, Instagram, and X, but it is unfortunately no longer available.

Many social media monitoring tools – typically used for marketing tools and often paid – can also be leveraged to visualise the outreach of specific content and track social media activity, such as:

- Hootsuite
- Meltwater
- NewsWhip
- Social Blade



Example: Hoaxy could be used to track the narratives against the trans swimmer Lia Thomas that circulated during the 2024 Paris Olympics. The image shows the visualisation of Bluesky accounts using the words “Lia Thomas” and “man” in the same post.

CONCLUSIONS AND RECOMMENDATIONS

The rise of identity-based disinformation within the broader context of information manipulation and interference by foreign actors who exploit deep societal fractures should be recognised as a serious threat. These IBD-oriented FIMI campaigns not only aim to undermine democratic values, processes, and institutions but also inflict tangible harm on individuals and communities based on gender, orientation, race, and ethnicity. As FIMI perpetrators adapt their strategies in response to events and counter-disinformation efforts, defenders must adopt dynamic, multi-faceted investigative approaches. Therefore, it is imperative for OSINT practitioners to maintain an identity-focused lens throughout their activities and avoid overlooking the specificities of marginalised groups.

Moreover, those targeted by IBD do not always have immediate access to external help, making it urgent for them to equip them with basic knowledge and tools to be gather evidence and defend themselves. In this context, OSINT is vital for detecting and analysing disinformation. These guidelines reinforce the resilience of potential targets by providing accessible tools to archive gathered evidence, uncover coordination, verify authenticity, trace sources, and assess impact of IBD-focused FIMI operations.

To strengthen these efforts, a set of strategic recommendations guide the defender community in enhancing their investigative capabilities and ensuring a more effective response to the menace.

- IBD-related FIMI researchers must continuously map and adapt to evolving malicious tactics. As the nature of disinformation shifts, investigative strategies should evolve accordingly to remain effective.
- Collaboration among multiple stakeholders is critical to counter IBD and FIMI successfully. A coordinated approach can ensure that efforts are aligned and not duplicated.
- While the availability of free tools is a positive development, more advanced solutions are needed to address increasing data access restrictions due to platform policy changes. Accessing data is essential for mapping threats beyond anecdotal evidence.
- OSINT investigations must operate within legal and ethical bounds. Investigators should reflect the diversity of the communities targeted by IBD, incorporating gendered, racial, and queer perspectives to ensure a victim-centric approach. Those targeted by IBD have a lot to contribute and should be actively engaged in the fight against these attacks.
- Finally, these guidelines emphasise the importance of establishing global standards in OSINT investigations, particularly for IBD-related FIMI. It crucial to balance the need for standardisation with a deeper understanding of identity-specific experiences. These standards should prioritise interoperability, transparency, and accountability to strenghten whole-of-society resilience against interference and manipulation.

APPENDICES

LIST OF TOOLS REPOSITORIES

- Bellingcat's Online Open Source Investigation Toolkit
- Exposing the Invisible – The Kit
- IntelligenceX
- IntelTechniques (paid)
- OSINT Essentials
- OSINT Framework
- RAND's tools finder
- Thomson Reuters tools recap

LIST OF ARCHIVING TOOLS (SECTION 3.1)

TOOL	KEY USE	PLATFORM	ACCESS
archive.today	Digital archiving	Web-based	Free
ArchiveBox	Digital archiving	Web-based	Free
Arquivo.pt	Digital archiving	Web-based	Free
Arweave	Decentralised archiving	Blockchain	Paid
FDOWN.net	Downloading Facebook videos	Web-based	Free
Getfvid	Downloading Facebook videos	Web-based	Free
Ghost Archive	Digital archiving	Web-based	Free
HTTrack	Downloading entire websites for offline use	Windows, Mac, Linux	Free
Hunch.ly	Local drive archiving	Local drive	Paid
Perma.cc	Digital archiving	Web-based	Limited free plan
Savefrom.net	Downloading YouTube videos	Web-based	Free
SnapSave	Downloading Facebook videos	Web-based and app	Free
SSSTWITTER	Downloading X videos	Web-based	Free
Wayback Machine	Digital archiving	Web-based	Free
Webrecorder	Interactive recordings of web sessions	Windows, Mac, Linux	Free

Wget	Downloading entire websites for offline use and resuming interrupted downloads	Windows, Mac, Linux	Free
Stone	Screen captures and webcam commentaries for research transparency	App	Free

LIST OF COORDINATION ASSESSMENT TOOLS (SECTION 3.2)

TOOL	KEY USE	PLATFORM	ACCESS
CooRnet	Detecting coordinated link-sharing behaviour (CLSB)	Self-hosted	Free
CooRTweet	Detecting coordinated networks on social media	Self-hosted	Paid
Cytoscape	Network analysis, visualisation, and data integration	Self-hosted	Free
DNSlytics	Searching domains sharing the same Analytics ID	Web-based	Limited free plan
Followerwon	X accounts network analysis	Web-based	Free
Gephi	Social network analysis, exploring relationships between accounts	Self-hosted	Free
Maltego	Network analysis, visualisation, and data integration	Self-hosted	Paid
Meltwater	Social media monitoring	Web-based	Paid
NodeXL	Network analysis and visualisation	Plug-in	Paid
Twiangulate	X accounts network analysis	Web-based	Free
InVID-WeVerify verification plugin	Social network analysis	Extension	Free (upon registration)

LIST OF AUTHENTICITY ASSESSMENT TOOLS (SECTION 3.3)

TOOL	KEY USE	PLATFORM	ACCESS
Botometer X	Automated X account detection	Web-based	Free
Bot Sentinel	Automated X account detection	Web-based	Free
Database of Known Fakes (DBKL)	Fact-checking database	Web-based	Free (restricted access)
Deepware Scanner	AI-generated content detection (video)	Web-based	Free

EUvsDisinfo Database	Fact-checking database	Web-based	Free
EXIF Purge	Metadata remover for photos	Self-hosted	Free
ExifTool	Metadata analysis	Self-hosted	Free
FotoForensics	Image forensic analysis	Web-based	Free
Forensically	Image forensic analysis	Web-based	Free
Google Fact-Check Tools	Fact-checking database	Web-based	Free
Google Reverse Image Search Tool	Reverse image search	Web-based	Free
Irfanview	Metadata analysis	Self-hosted	Free
Jimpl	Metadata analysis	Web-based	Free
Metadata2Go	Metadata analysis	Web-based	Free
Microsoft Bing Visual Search	Reverse image search	Web-based	Free
Pangram Labs	AI-generated content detection (text)	Web-based	Limited free plan
Reveal Image Verification Assistant	Image forensic and metadata analysis	Web-based	Free
RevEye	Reverse image search	Extension	Free
TinEye	Reverse image search	Web-based	Free
InVID-WeVerify verification plugin	Synthetic image detection Video deepfake detection Face swapping in images Image forensic and metadata analysis Keyframe video segmentation Reverse image search	Extension	Free (upon registration)[45]
Yandex	Reverse image search	Web-based	Free

LIST OF SOURCE ASSESSMENT TOOLS (SECTION 3.4)

TOOL	KEY USE	PLATFORM	ACCESS
Censys	Domain name search	Web-based	Limited free plan
DomainTools	Domain name search	Web-based	Limited free plan
DNSlytics	Domains sharing the same Analytics ID search	Web-based	Limited free plan
Epieos	Email and phone reverse lookup	Web-based	Limited free plan
FreeCarrierLookup.com	Telephone number search	Web-based	Free
Fonefinder	Telephone number search	Web-based	Free
Gephi	Social network analysis	Web-based	Free
Hunter	Email search	Web-based	Limited free plan
ICANN	Domain name search	Web-based	Free
Instant Username Search	Username search	Web-based	Free
Namecheckr	Username search	Web-based	Free
Namech_k	Username search	Web-based	Free
OSINT Industries	Personally identifiable information search	Web-based	Paid
Truecaller	Telephone number search	Web-based	Limited free plan
theHarvester	Email search	Self-hosted	Limited free plan
UserSearch	Username search	Web-based	Paid (free trial)
InVID-WeVerify verification plugin	Social network analysis	Extension	Free (upon registration)
WhatsMyName	Username search	Web-based	Free
WhoCalld	Telephone number search	Web-based	Free
WHOIS	Domain name search	Web-based	Free

LIST OF IMPACT ASSESSMENT TOOLS (SECTION 3.5)

TOOL	KEY USE	PLATFORM	ACCESS
BuzzSumo	Tracking content performance	Web-based	Limited free plan
Hoaxy	Social media content visualisation	Web-based	Limited free plan
Hootsuite	Social media statistics and analytics	Web-based	Paid
Meltwater	Tracking content performance	Web-based	Paid
NewsWhip	Tracking content performance	Web-based	Paid
Open Measures	Fringe platform content outreach	Web-based	Free
Social Blade	Social media statistics and analytics	Web-based	Limited free plan

REFERENCES

1 Rid, T. (2021). Active measures. The secret history of disinformation and political warfare. Picador.

2 EEAS (October 2023). FIMI targeting LGBTIQ+ people: Well-informed analysis to protect human rights and diversity. Available at: https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-LGBTQ-Report-03-Digital%201.pdf

3 Sessa, M.G. (26 January 2022). What is gendered disinformation?. Heinrich-Böll-Stiftung. Available at: https://il.boell.org/en/2022/01/26/what-gendered-disinformation

4 Sobieraj, S. (22 October 2019). Disinformation, democracy, and the social costs of identity-based attacks online. MediaWell. Available at: https://mediawell.ssrc.org/articles/disinformation-democracy-and-the-social-costs-of-identity-based-attacks-online/

5 Alaphilippe, A. (2022). Sources ouvertes et lutte contre la désinformation: un chantier démocratique. Hérodote, 186(3), 69-83. https://doi.org/10.3917/her.186.0069

6 Pamment, J. (September 2020) The EU's Role in Fighting Disinformation: Crafting a Disinformation Framework. Working Paper of the Carnegie Endowment for International Peace. https://carnegieendowment.org/files/Pamment_-_Crafting_Disinformation_1.pdf

7 Nimmo, B. and Hutchins, E. (10 November 2022) Presentation on Overarching Online Operations Kill Chain. Cyberwarcon Conference. https://www.csoonline.com/article/3680149/meta-s-newkill-chainmodel-tackles-online-threats.html

8 Mezzofiore, G. (3 July 2024). Deepfake video targeting Zelensky's wife linked to Russian disinformation campaign, CNN analysis shows. CNN. Available at: https://edition.cnn.com/2024/07/02/europe/deepfake-video-zelensky-wife-intl-latam/index.html

9 Nimmo, B. (September 2020) The Breakout Scale: Measuring the impact of influence operations. Working Paper of the Brookings Institution. https://www.brookings.edu/articles/the-breakout-scale-measuring-the-impact-of-influence-operations/

10 Miguel Serrano, R. (10 June 2022). Towards an impact-risk index of disinformation: Measuring the virality and engagement of single hoaxes. EU DisinfoLab. Available at: https://www.disinfo.eu/wp-content/uploads/2022/09/20220610_IndexImpactAssessment_Final-1.pdf.

11 Brady, K. (October 2021). Online trolls direct sexist hatred at Annalena Baerbock. DW. Available at: https://www.dw.com/en/germany-annalena-baerbock-becomes-prime-target-of-sexist-hate-speech/a-57484498

12 Stover, E., Koenig, A., & Freeman, L. (2022). Berkeley Protocol on Digital Open Source Investigations: A practical guide on the effective use of digital open source and information in investigating violations of international criminal, human rights and humanitarian law. International Human Rights Center. Available at: https://www.ohchr.org/sites/default/files/2024-01/OHCHR_BerkeleyProtocol.pdf.

13 ObSINT (March 2023). Guidelines for public interest OSINT investigations. Available at: https://obsint.eu/wp-content/uploads/2023/04/Guidelines-for-Open-Source-Intelligence-Organisations.pdf.

14 Millett, E. (5 December 2023). Deploying OSINT in armed conflict settings: law, ethics, and the need for a new theory of harm. ICRC blog. Available at: https://blogs.icrc.org/law-and-policy/2023/12/05/deploying-osint-in-armed-conflict-settings-law-ethics-theory-of-harm/

15 Koenig, A., & Egan, U. (March 2021). Power and privilege: Investigating sexual violence with digital open source information. Journal of International Criminal Justice 19(1), 55–84. https://doi.org/10.1093/jicj/mqab014

16 Lead Stories (20 February 2020). Fake news: A viral image does NOT show Rep. Ilhan Omar without a head scarf. Available at: https://leadstories.com/hoax-alert/2020/02/fake-news-a-viral-image-is-NOT-Rep-Ilhan-Omar-without-a-head-scarf.html

17 Tufekci, Z. (25 July 2016). WikiLeaks put women in Turkey in danger, for no reason (update). HuffPost. Available at: https://www.huffpost.com/entry/wikileaks-erdogan-emails_b_11158792

18 Miguel Serrano, R., Sessa, M.G., & Alaphilippe, A. (24 May 2024). Assessing cost-effectiveness: Responses to the Doppelganger operation. EU DisinfoLab. Available at: https://www.disinfo.eu/publications/assessing-cost-effectiveness-responses-to-the-doppelganger-operation

19 ISO/IEC 27037:2012, Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence. Edition 1. Available at: https://www.iso.org/standard/44381.html

20 European Commission (30 April 2024). Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373

21 Dang, S. (6 November 2023). Exclusive: Elon Musk's X restructuring curtails disinformation research, spurs legal fears. Reuters. Available at: https://www.reuters.com/technology/elon-musks-x-restructuring-curtails-disinformation-research-spurs-legal-fears-2023-11-06/

22 Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S., & Kaufmann, Z. (2021). Malign creativity: How gender, sex, and lies are weaponized against women online. Wilson Center: Science and Technology Innovation Program. Available at: https://www.wilsoncenter.org/sites/default/files/media/uploads/documents/Report%20Malign%20Creativity%20How%20Gender%2C%20Sex%2C%20and%20Lies%20are%20Weaponized%20Against%20Women%20Online_0.pdf

23 Zaveri, M., & Fortin, J. (6 May 2019). Russian efforts exploited racial divisions, State of Black America report says. The New York Times. Available at: https://www.nytimes.com/2019/05/06/us/russia-disinformation-black-activists.html

24 Janulewicz, L., & Balint, K. (15 December 2021). The global network working to curb the rights of women and sexual minorities. ISD. Available at: https://www.isdglobal.org/digital_dispatches/the-global-network-working-to-curb-the-rights-of-women-and-sexual-minorities/

25 Simmons, C., & Martiny, C. (24 January 2024). Networks of dissuasion: Mapping online attacks on reproductive rights in France. ISD. Available at: https://www.isdglobal.org/wp-content/uploads/2024/01/Networks-of-Dissuasion-Mapping-Online-Attacks-on-Reproductive-Rights-in-FR.pdf

26 Marchlewska, M., & Cichocka, A. (23 March 2020). How a gender conspiracy theory is spreading around the world. The Conversation. Available at: https://theconversation.com/how-a-gender-conspiracy-theory-is-spreading-around-the-world-133854

27 Brown, A. (19 August 2019). The myth of Eurabia: How a far-right conspiracy theory went mainstream. The Guardian. Available at: https://www.theguardian.com/world/2019/aug/16/the-myth-of-eurabia-how-a-far-right-conspiracy-theory-went-mainstream

28 Hurst, L. (20 October 2023). Generative AI fueling spread of deepfake pornography across the internet. Euronews. Available at: https://www.euronews.com/next/2023/10/20/generative-ai-fueling-spread-of-deepfake-pornography-across-the-internet

29 #ShePersisted (May 2024). Big tech and the weaponisation of misogyny in Moldova's online ecosystem. An assessment of digital threats to women in public life. Available at: https://she-persisted.org/wp-content/uploads/2024/06/ShePersisted-Moldova-Report-ENG.pdf

30 Warren, T. (14 October 2024). The Internet Archive is back as a read-only service after cyberattacks. The Verge. Available at: https://www.theverge.com/2024/10/14/24269741/internet-archive-online-read-only-data-breach-outage

31 Witness (2013). Activists' guide to archiving video. Available at: https://library.witness.org/product/activists-guide-to-archiving-video/

32 Bellingcat (8 March 2022). How to archive Telegram content to document Russia's invasion of Ukraine. Available at: https://www.bellingcat.com/resources/how-tos/2022/03/08/how-to-archive-telegram-content-to-document-russias-invasion-of-ukraine/

33 A complete user guide is available at: https://archiveweb.page/guide.

34 Romero Vicente, A. (8 August 2024). Coordinated Inauthentic Behaviour detection tree. EU DisinfoLab. Available at: https://www.disinfo.eu/publications/coordinated-inauthentic-behaviour-detection-tree/

35 A complete list of Google Dorks is available at: https://www.boxpiper.com/posts/google-dork-list.

36 Atanasova, A., Lesplingart, A., Poldi F., & Kuster, G. (June 2024). Operation Overload. CheckFirst & Reset.Tech. Available at: https://checkfirst.network/wp-content/uploads/2024/06/Operation_Overload_WEB.pdf

37 Silverman, C. (11 July 2023). What the rollout of Google Analytics 4 means for website investigations. Digital Investigations. Available at: https://digitalinvestigations.substack.com/p/what-the-rollout-of-google-analytics

38 Miguel Serrano, R. (10 February 2021). "Tierra Pura, product of the pandemic: A new Spanish-language disinformation outlet with connections to the Epoch Times ecosystem. EU DisinfoLab. Available at: https://www.disinfo.eu/publications/tierrapura-product-of-the-pandemic/

39 Kravets-Meinka, D. (15 March 2024). The sad fate of Yandex: From independent tech startup to Kremlin propaganda tool. Zois–Centre for East European and International Studies. https://www.zois-berlin.de/en/publications/zois-spotlight/the-sad-fate-of-yandex-from-independent-tech-startup-to-kremlin-propaganda-tool

40 More information on EXIF data is available at: https://www.photographymad.com/pages/view/exif-data-explained.

41 Thompson-Fuller, T. (7 October 2021). Deepfake video appears to show New Zealand PM 'smoking crack'. AFP Fact Check. Available at: https://factcheck.afp.com/doc.afp.com.9P97TM

42 NewsGuard's Reality Check (30 July 2024). False Olympics claims: Terrorism, crime, and wardrobe malfunction. Available at: https://www.newsguardrealitycheck.com/p/false-olympics-claims-terrorism-crime?open=false#%C2%A7one-more-thing-no-an-olympics-opening-ceremony-dancer-did-not-expose-his-privates

43 Smirnova, J., Winter, H., Mathelemuse, N., Dorn, M., & Schwertheim, H. (September 2021). Digitale Gewalt und Desinformation gegen Spitzenkandidat:innen vor der Bundestagswahl 2021. ISD. Available at: https://www.isdglobal.org/wp-content/uploads/2021/09/Digitale-Gewalt-und-Desinformation_v5.pdf

44 Marinov, V. (14 August 2024). Nein, Mike Tyson hat keinen Kampf mit algerischer Boxerin Imane Khelif angekündigt. Correctiv. Available at: https://correctiv.org/faktencheck/2024/08/14/nein-mike-tyson-hat-keinen-kampf-mit-algerischer-boxerin-imane-khelif-angekuendigt/

45 While most functions require access approval, image forensics, metadata analysis, keyframe fragmentation, and reverse image search are freely available to all within the verification plugin.


European Union
EXTERNAL ACTION