PROMPT

Narrative Report #1

PROJECT	
Project number:	LC-02629302
Project acronym:	PROMPT
Project name:	Predictive Research On Misinformation and Narratives Propagation Trajectories

Deliverable 5.2. 1st Narrative Report on the state of disinformation on the War in Ukraine, and European Elections	
Contractual date:	30/06/2025
Submitted Date:	08/07/2025
Version:	[1.0]
Contributors	Rīga Stradiņš University (RSU) (Latvia), Università degli Studi di Urbino (Italy), HUN-REN Center for Social Sciences (Hungary), Re:Baltica (Latvia), Les Surligneurs (France), Erich-Brost Institut (Germany), Wikimedia France, Asociația Digital Bridge (Romania), Orizzonti Politici APS (Italy), opsci.ai (France).

TABLE OF CONTENTS

INTRODUCTION

Even though false and misleading information has existed throughout history, disinformation in the 21st century presents a new level of previously never experienced threats and cause for concern due to the advances in technology and global connectivity. The public information and discussion ecosystem is undergoing a profound structural transformation, driven by the emergence of an automated public sphere governed by the operational logic of global platforms.

There is a recognized challenge in defining and conceptualizing disinformation and related terms like misinformation. The term "fake news" has become politicized and used to discredit news sources. There are also various sources of disinformation – from foreign states to activist networks or individuals and even companies. While research has approached the role of Russia in spreading disinformation to cause chaos and influence elections in several countries, the People's Republic of China has also been observed as a powerful source of disinformation. Groups with ideological agendas, such as climate change denial and anti-vaccination groups, may spread disinformation. Business owners, celebrities, influencers, even experts, domestic political actors or other powerful individuals may also be motivated, or follow incentives (including financial ones) to produce or spread disinformation.

Disinformation is observed to be spread through a complex interaction of social media, online news sites, traditional media, and offline spaces. It can originate online or from public figures, then be spread by sites masking as legitimate news sources, circulated via social media, and even amplified by traditional media. Disinformation can appear in various forms, from being entirely fabricated to featuring information that is reconfigured or taken out of context. The use of generative artificial intelligence (genAI), which can produce plausible text, images, video, and audio has made disinformation production become more accessible, cheaper, and easier to use.

The unprecedented scale, speed, and complexity of disinformation in a digitally connected world is contributing to creating a constant state of polycrisis. Disinformation poses a threat to freedom and democracy by manipulating public opinion and aiming to erode institutions. It can undermine the capacity of individuals to make well-informed political decisions and evaluations. The impact of disinformation on democratic elections, polarization, attitudes and behaviour, political divisions, is still being explored and is of great concern. On these grounds, PROMPT aims to examine the impact of disinformation on social media in order to contribute to broader research on its effects in the public sphere. We position ourselves as an important link in this causal chain, which is still insufficiently explored.

Long-term disinformation campaigns designed to sow mistrust, confusion, and sharpen existing socio-cultural divisions using nationalistic, ethnic, racial, and religious tensions are particularly concerning. Disinformation can also promote societal unrest and undermine

social cohesion. Disinformation can amplify negative emotions such as fear, anxiety, and fury, with negative implications for people's mental health.

Information pollution contaminates public discourse on a range of issues beyond politics, including medical information. Medical misinformation has historically posed a worldwide health threat, spreading incorrect treatment advice through various channels. The COVID-19 pandemic highlighted how fake news can have considerable consequences on public reaction, including fueling conspiracy theories, promoting risky advice, hindering clear communication, reducing trust in reliable sources, leading to vaccine unwillingness, and causing unjustified fear. Disinformation is also difficult to tackle and debunk, especially when it spreads within trusted networks.

Concerns about social media's role in spreading disinformation are felt all over the world, from the United States to Europe, North America, Asia, as well as Latin America. Media professionals consider the fight against disinformation as one of the most important challenges to their work. Citizens expect media professionals not only to be a source of trusted and verified information, but also to protect the public from the influence of disinformation. Even more, transformation in disinformation strategies and techniques have changed the way journalists process information, pushing them to adapt and adopt new working methods.

Disinformation is an ever-changing phenomenon, with new topics, players and techniques/methods being increasingly instrumentalised. Against the backdrop of these concerns, [PROMPT](#) employs AI-driven methods to create a disinformation detection and contextualisation arsenal for journalists and activists. The most important yet difficult task in disinformation studies is to decipher the impact of disinformation. **Central to PROMPT is the hypothesis that propagators of disinformation increasingly leverage culture-, community- and language- specific patterns to achieve greater impact**. One needs to understand how narratives are formed and reshaped to resonate in specific communities. PROMPT employs AI-driven methods and Large Language Models (LLM) to help monitor disinformation narratives, how they propagate and transform in the process of its movement and transformation across social platforms and local contexts.

The report outlines the current topics dealing with disinformation in a media-driven world against the background of a constant state of polycrisis. The report first gives a compact introduction into the different terms and tactics of disinformation-spreading on social media and a brief description of the various (most popular) tools currently used by media experts to identify disinformation and harmful narratives.

The main focus of the report is to unfold the **concept and mechanisms behind PROMPT** and its **contribution to the fight against disinformation and harmful narratives**.

Addressing the spread of disinformation in the digital age requires a **dynamic and multi-dimensional methodology** that accounts for how false narratives evolve, adapt to platform-specific affordances, and circulate across linguistic and national boundaries. The report gives insights into PROMPT's integrated approach when analysing the dissemination of disinformation narratives across various social media platforms. The methodological strategy combines Large Language Models (LLMs), dynamic network analysis (DNA), and supervised and unsupervised classification pipelines for detecting policy-related claims, with the main goal to capture coordinated information flows focusing on temporal, cross-platform, and cross-language propagation dynamics.

The report applies its conceptual framework and technical pipeline to **analyse the narrative state of two PROMPT topics in the public space** (social media, digital commons and digital media as observed through the project) in the project's explored linguistic zones: 1) the **war in Ukraine**, and 2) elections taking place in EU countries, specifically the re-run of the first round of the **Romanian presidential elections**, which took place on May 4, 2025.[1] Focusing on different disinformation narratives at work, the report gives insight into the communities – sources of harmful narratives, their aims and points of weakness. The main metrics include the societal impact, using social media engagement metrics (likes, shares, comments, etc.) as a proxy, the structure of narratives' dissemination, main communities as well as key opinion leaders contributing to the narratives spread and "swinging" communities, exposed to the misleading narratives.

---

[1] Given the difficulties that emerged concerning data collection on our third topic of analysis, - LGBTQIA+ issues – a study case on this subject has not yet been developed. Nevertheless, we are working on updating the data gathering process, by adapting trigger-based and narrative/claim-based queries across platforms, APIs and languages. These simplified queries have started to show promising results, gathering generic publications that demand increased filtering and detection to point out disinformation speeches regarding LGBTQIA+ topics. On these grounds, we are expecting to include a study-case on LGBTQIA+ related issues on our second narrative report.

Disinformation comes in various forms and channels; however, distinct characteristics help to identify it:

- **False or Misleading Content.** Disinformation is information that is false, incorrect, or misleading. It can range from entirely fabricated content to information that has been reconfigured or taken out of context. Disinformation is often labeled as "fake news", while this term fails to describe the complexity of information pollution. While both disinformation and "fake news" intentionally circulate false information, "fake news" is manufactured and presented *as news*, whereas disinformation is not necessarily news.

- **Deliberate Creation and Spread.** A defining feature of disinformation is that it is deliberately created and spread or intentionally circulated. While misinformation is false information shared without the intention to cause harm or mislead, disinformation is intentional. However, distinguishing between misinformation and disinformation is not always possible in practice.

- **Intent to Deceive, Mislead, or Cause Harm.** Disinformation is intended to deceive and mislead people, with the aim of misleading and influencing public opinion. The intent is to deliberately cause harm to a person, social group, organization, or country, or to gain political, personal, or financial advantage.

TOOLS AND TACTICS
Weapons of choice currently used by media experts to tackle disinformation and fact-check

Fact-checkers and other experts employ various tools, strategies and technologies to identify disinformation. Due to the scale, volume, and speed of information dissemination in the 21st century, manual identification of disinformation is often insufficient, and **automated approaches** are crucial, where possible. Experts also employ **hybrid approaches,** combining automated detection with manual analysis by subject matter, as traditional Natural Language Processing (NLP) struggles with linguistic nuances and adversarial tactics. Some studies link survey data to digital trace data to understand who is exposed and why users interact with disinformation.

There are different automated detection technologies that may help identify disinformation:

- **Machine Learning (ML) and Natural Language Processing (NLP):** These are central to detecting patterns, inconsistencies, semantic anomalies, and stylistic deviations characteristic of disinformation in large volumes of text and other data. Advanced techniques include **deep learning frameworks** like Convolutional Neural Networks

(CNN) and Recurrent Neural Networks (RNN), as well as **transformer models** (e.g., BERT, GPT, DeBERTa, xlmRoBERTa) which are effective for capturing complex relationships, understanding context, and performing semantic reasoning. ML algorithms can recognize language nuances that suggest false information.

- **Analysis of Stylistic and Contextual Features:** Algorithms can be developed to analyze texts based on their stylistic and contextual characteristics to detect disinformation.

- **Sentiment Analysis:** Identifying emotional coloring in text can indicate propagandistic or manipulative content.

- **Rhetorical Device Detection:** Identifying specific rhetorical devices, such as those intended to increase doubt, can help indicate when disinformation is present. Specific patterns in the use of certain rhetorical devices are observed more frequently in disinformation.

- **Visual and Audio Analysis:** Technologies are being developed to address fabricated, manipulated, or falsely contextualized visuals and audio, including **AI-based photo and video analysis tools** and methods to detect **deepfakes**. Reverse image search can help verify visual content.

- **Analysis of Dissemination Patterns and User Behavior:** Studying how information spreads and identifying abnormal behavior of objects or users, such as coordinated inauthentic behavior, social bots, or cyborgs, can help detect manipulative information campaigns.

- **Graph Analytics and Network Causal Inference:** These methods are used to map networks, identify influential actors in spreading narratives, and quantify their impact, accounting for network topology and removing the effects of social confounders.

There are also several **fact-checking and verification fools** that provide aid amidst the torrents of information. While some tools focus on manual verification (like Snopes), there are automated or semi-automated tools with capabilities to analyze claims and visualize dissemination chains (like Hoaxy).

PROMPT partners **Wikimedia** (France), **Les Surligneurs** (France), **Euractiv** (Romania) and **Re:Baltica** (Baltic states) are regularly involved in fact-checking. Their most frequently used tools include:

- AI4Trust – a hybrid system that leverages both human and machine intelligence to combat disinformation. In its pilot phase, AI4Trust enables users to verify the authenticity of videos, images, and text by detecting potential manipulations.

- [Metadata2GO](#) – a user-friendly tool designed to inspect the metadata of images and various file types, helping to uncover hidden details such as creation dates, device information, and editing history.
- [TinEye](#) – a reverse image search engine that continuously crawls the web to help users find, verify, and track the origin and usage of images online.
- [DeepFake Detector](#) - an AI-powered tool capable of analyzing audio and video content to determine whether it is genuine or artificially generated (deepfake).
- [Logically](#) - a comprehensive platform that integrates predictive modeling, network analysis, and narrative intelligence with expert insights to identify and anticipate potential threats.
- [Osavul](#) - advanced AI-driven platform focused on assessing information and analyzing the impact of narratives, supporting efforts to understand and counteract disinformation.
- [Osint framework](#) - a web-based resource designed to help researchers, investigators, and analysts locate and use open-source intelligence (OSINT) tools and techniques; it organizes a wide range of publicly available resources—such as search engines, databases, and investigative tools—into a structured, interactive directory, making it easier to gather information from publicly accessible sources across domains like social media, geolocation, domain names, and more.
- META fack-checking tool – identifies and flags potentially false or misleading posts, videos, and other content.
- AI tool on TikTok – uses machine learning to detect potentially misleading or false content.
- ChatGPT – useful for fact-checkers to speed up the initial stages of fact-checking and hence save time, namely by finding background information and providing translation.

PROMPT: building on previous work

Disinformation on digital platforms poses a serious threat to democratic institutions, civic trust, and minority rights. Studies have examined how false narratives spread across various themes such as armed conflict, elections, and anti-LGBTQIA+ rhetoric. Advances in computational methods, framing theory, and network science offer new ways to study these dynamics.

Textual network analysis has emerged as a crucial method for identifying latent discursive patterns within social media and news corpora (Diesner, 2015). This approach underscores the significance of relational linkages between ideas and their manifestation in public discourse, particularly during socio-political crises. These techniques have also been employed in analyzing identity-focused discourse, such as LGBTQIA+ discussions, where network structures elucidate how moralized narratives converge around exclusionary or populist ideologies.

Framing theory provides a valuable analytical perspective for understanding the propagation of disinformation narratives. Frames are defined as specific methods of structuring meaning around an issue, such as through blame attribution or emotional tone (Entman, 1993). **Narratives, on the other hand, refer to broader interpretive templates that delineate causality, intention, and moral evaluation over time.** Verbal framing analysis has been applied, for example, to examine how digital narratives concerning conflict reinforce polarisation and foster competing interpretive communities (Ptaszek et al., 2024). Furthermore, beyond conflict zones, framing has been studied in relation to anti-LGBTQIA+ rhetoric, where such narratives frequently employ child protection frames and moral panic tropes. In the context of EU elections (Bayer et al., 2019), disinformation campaigns often exploit cultural identity frames and anti-elite sentiment to undermine democratic legitimacy.

Amplification mechanisms and interaction dynamics play a critical role in shaping the structure of disinformation ecosystems. Research indicates that these networks are frequently characterized by high internal connectivity and limited cross-group interaction. This structure can impede the dissemination of misleading content across different ideological groups, yet it makes such narratives more resistant to correction. Motivated reasoning and strong identity-based commitment to group narratives often prevent fact-checks or rebuttals from penetrating ideologically homogeneous communities (Altay et al., 2023). These dynamics are further reinforced by algorithmic filtering and user-driven content curation, which enhance intra-group coherence while restricting the exposure to corrective information (Yarchi et al., 2021; Wilson & Starbird, 2020).

Recent research emphasizes **cross-lingual dissemination** in European disinformation. Major campaigns use translation, localization, and platform-specific features to adapt narratives across languages. The European Digital Media Observatory (EDMO, 2023) reports that anti-EU and pro-Kremlin disinformation spreads in synchronized bursts, starting in major Western European languages and then reaching Central and Eastern Europe. Multilingual influencers, transnational conspiracy communities, and shared hashtags help these narratives bypass national media gatekeeping. Therefore, recent findings emphasize the need for multilingual monitoring frameworks to detect semantic parallels and local adaptations of core narratives.

Integrating **natural language processing** (NLP) with **network analysis** is essential for studying digital narratives (DiMaggio et al., 2013; Starbird, 2017). Studies highlight the value of combining topic modeling with frame analysis to track narrative evolution over time. Identifying bridge users in dissemination networks is key to understanding how isolated clusters connect and spread frames between communities, expanding their reach. Dynamic network visualization tools allow real-time tracking of frame propagation, enhancing our comprehension of narrative shifts across platforms.

PROMPT addresses these challenges by implementing language-agnostic LLMs and narrative mapping tools trained on cross-lingual datasets. Recent studies indicate that combining network analysis, frame theory, and social media studies offers valuable insights into the

mechanisms of digital narrative propagation, especially in conflict scenarios. Analyzing the role of key actors, interaction types, and platform-specific dissemination practices can improve our understanding of how frames develop, compete, and persist in the online public sphere.

PROMPT builds on the **NODES** project (https://nodes.eu, 2022-2024) that also stands out for its narrative-focused, transnational approach and integration of computational and interpretive methods. Led by Re-Imagine Europa, NODES (Narratives Observatory combating Disinformation in Europe Systemically) tracked disinformation in four EU countries across three domains: climate change, migration, and COVID-19. It combined social science and computational tools to map how polarising narratives evolve and spread across linguistic and national boundaries. A core innovation of NODES was its use of relational network models, which emphasize the circulation of meaning and narrative logic over ideological alignment. These models help reveal how disinformation is transmitted, reshaped, or resisted by different communities. **PROMPT builds on this foundation by integrating Large Language Models (LLMs) and real-time network analysis**, enhancing scalability and enabling detection of semantic shifts, irony, and emotional tone. It also **broadens the scope to include disinformation in wartime, electoral disinformation and anti-LGBTQIA+ rhetorics**, aiming to establish a scalable, cross-platform narrative analysis infrastructure across Europe.

PROMPT: methodology

PROMPT leverages AI-powered techniques and Large Language Models (LLMs) to track disinformation narratives — examining how they spread, evolve, and adapt across social media platforms and within different local contexts.

First, PROMPT has developed a semantic-axiological matrix, explained in a comprehensive codebook, to analyze and categorize core disinformation narratives and the arguments that support them. Second, PROMPT implements language-agnostic LLMs and narrative mapping tools trained on cross-lingual datasets to analyze the role of key actors, interaction types, and platform-specific dissemination practices.

1. Automated content analysis using semantic-axiological matrix

Drawing on prior disinformation research, practitioner best practices, and analysis of disinformation "samples," a codebook was developed comprising 30 code groups and over 240 individual codes. Grounded in an interdisciplinary framework—spanning media studies, communication, and semantics—the coding system integrates linguistic features, rhetorical strategies, and information manipulation techniques. It also includes codes that help identify the intent behind disinformation, the motivations of its disseminators, and its potential

impact. In this codebook, the form of disinformation (how it is presented) is considered as significant as its content (what it conveys).

The core of the codebook focuses on linguistic and rhetorical categories. Based on the insight that disinformation increasingly relies on culturally and linguistically specific patterns to enhance its effectiveness, the analysis emphasizes the rhetoric of disinformation through a multimodal (text, images, video) and multilingual (eight languages) lens. These practices are decoded by categorizing persuasive mechanisms, manipulation techniques, lexical choices, and a wide array of rhetorical figures. Axiological-semantic categories are employed to identify the underlying values promoted by disinformation content, allowing to explore deeper layers of meaning behind the disinformation messages.

The codebook also facilitates the identification of content designed to incite division, spread fear, or manipulate public sentiment. This includes the use of aggressive or hostile language (e.g., "enemy," "traitor," "invader," "conflict," "threat") and discriminatory stereotypes (e.g., "catho-facho," "welfare queen," "lazy immigrant"). Alarmist vocabulary — such as "catastrophic," "urgent," "epidemic," "tragic," or "crisis" — is often employed to provoke panic or exaggerate situations. While such language elements do not necessarily entail disinformation, they are strong indicators of its potential presence.

Following the initial testing phase, the codebook was expanded to include analysis of how language in social media posts aligns with various political ideologies. This addition significantly enhances the tool's capacity to contextualize disinformation within broader political and cultural frameworks.

Powered by AI and Large Language Models (LLMs), all categories within the PROMPT framework contribute to a comprehensive disinformation analysis toolkit. This system is designed to detect subtle linguistic nuances in public discourse—nuances that may escape casual observation but can significantly alter the intended message. These shifts in meaning play a critical role in shaping how disinformation is perceived and received across digital platforms.

All PROMPT partners contributed to the development and validation of the analytical matrix. During three technical workshops, project collaborators proposed and examined disinformation examples, providing critical input that informed revisions to the matrix. A key insight from this collaborative process was t**he importance of balancing journalistic standards — such as facticity, disinformation typologies, and verifiability — with less commonly addressed dimensions in disinformation analysis, including axiological attributes and rhetorical devices.** This pluralistic approach is central to PROMPT's contribution to the field of disinformation research.

The matrix was tested using a diverse sample of country-specific disinformation cases, including debunked claims sourced from fact-checking organizations. It reflects a deliberate

integration of theoretical frameworks and empirical analysis, drawing on literature and real-world disinformation samples from six countries: Latvia, Lithuania, Estonia, Romania, Italy, and France.

Designed to account for the unique characteristics of disinformation dissemination in each of these contexts, the matrix considers linguistic nuances, the structure of national communication environments, and the influence of local socio-political and cultural narratives.

Despite the complexity of its development, the PROMPT matrix is expected to significantly advance the understanding of disinformation mechanisms—rhetorical and beyond. It offers a valuable toolset for identifying actors who seek to manipulate public opinion, polarize digital discourse, and erode democratic institutions.

Codebook categories

The codebook is divided into five sections, each containing several categories and addressing a specific question:
   1) **What** does one observe in the social media post?
   2) **Where** has the social media post originated and which geographical area does it cover?
   3) **Who** is the author and target of the social media post, and which audience is it intended for?
   4) **How** is disinformation spread through the social media post?
   5) **Why** is the social media post being distributed?

1.1. Section "WHAT"

PROMPT - and hence the codebook - focus on **three topics** of analysis: 2024 EU Elections, War in Ukraine, and LGBTQIA+ issues.

The **Russia-Ukraine conflict** is marked by an intense information war, where rhetorical strategies are used to shape public perception. Russian disinformation often relies on such rhetorical figures as metaphors (e.g., Ukraine as a "cancerous tumor") to justify aggression, hyperbole (claims of "genocide") to provoke fear, and antithesis (Russia as "liberator" vs. Ukraine as "Nazi state") to polarize opinion. Repetitive phrases in political speeches (anaphora) are used to build emotional momentum and national unity. These tactics aim to delegitimize Ukraine and frame Russia's actions as defensive. In contrast, Ukraine counters with irony and satire, using memes and humor to expose the flaws in Russian propaganda. This approach helps engage audiences and build solidarity. However, social media algorithms often amplify emotionally charged content—regardless of accuracy—highlighting the need for stronger platform-level moderation to curb the spread of disinformation.

Disinformation targeting the **LGBTQIA+ community** often uses emotionally charged language to incite fear and hostility. Metaphors like "threat to children" or "gender ideology" frame LGBTQIA+ identities as dangerous and oppressive, suggesting an imposition on traditional norms. Hyperboles are employed to exaggerate the risks of gender-affirming care or inclusive education. Derogatory terms such as "groomer" and "transtrender" are used to stigmatize and dehumanize LGBTQIA+ individuals. The bandwagon fallacy is also often employed in anti-LGBTQIA+ disinformation, suggesting widespread opposition to LGBTQIA+ rights and thereby aiming to isolate the community and its allies.

Social media algorithms can worsen the problem by creating echo chambers that amplify anti-LGBTQIA+ content. Once users engage with such material, platforms may recommend more extreme content, increasing the risk of radicalization. In response, LGBTQIA+ advocacy groups work to counter disinformation through fact-checking and personal storytelling, promoting empathy and understanding.

The third area of interest for PROMPT was initially restricted to **disinformation trends across the 2024 European Parliamentary elections in France, the Baltic region, Italy, and Romania**. However, we have since then decided to broaden our scope to include **all electoral processes taking place throughout our countries of analysis during the time frame of the project**, in order to ensure the timeliness and maximise the relevance of the project's analysis.

Disinformation during the 2024 elections in **France** used fear-based messaging, oversimplified political issues, and personal attacks. Appeals to national identity were common, aiming to sway voters emotionally. Fact-checkers played a key role in countering false claims, while debates emerged over new foreign interference laws and their impact on civil society. In the **Baltic region**, disinformation—often linked to Russian sources—used loaded language, historical revisionism, and scapegoating. Narratives framed NATO as a threat and blamed the West for local issues. In response, Baltic governments promoted media literacy, supported independent journalism, and updated legal frameworks to address evolving threats such as deepfakes.

Disinformation in **Italy** reflected populist themes, emphasizing "us vs. them" rhetorics and alarmism around immigration and EU policies. Irony was used to subtly criticize opponents. Social media allowed direct communication with voters, amplifying both genuine messages and false narratives. COVID-19-related disinformation was also present.

Presidential elections in **Romania** faced conspiracy-driven disinformation, including claims of elite manipulation and foreign control. Nationalist and anti-establishment rhetorics were widespread. Platforms like TikTok and Telegram were key channels, with influencers and bots spreading content. Low digital literacy made the public more vulnerable, prompting concern over electoral integrity.

After identifying the **narrative** of the concrete unit and providing the **argument/claim** of the message, the codebook provides several categories of disinformation to choose from, classifying the main types of disinformation based on their structure, intent, and impact. Multiple types may apply to a single case, but the focus should be on identifying the top one to three most prominent elements that shape the narrative.

- **Clickbait** refers to content that uses exaggerated or misleading headlines and visuals to attract attention, often without delivering meaningful information.
- **Conspiracy theories** involve unfounded claims that explain events as secret plots by powerful actors, typically resistant to factual correction.
- **Fabrication** includes entirely false stories designed to deceive, often mimicking legitimate news formats.
- **Misleading connection** occurs when factual elements are presented out of context or paired with unrelated visuals or headlines.
- **Hoaxes** are elaborate fabrications that blend truth and falsehoods, potentially causing real harm.
- **Biased** or one-sided content pushes a specific ideological agenda, often using emotional or polarizing language.
- **Pseudoscience** misrepresents scientific research or promotes false claims under the guise of science.
- **Rumors** are unverified or ambiguous claims that spread informally, especially on social media.
- **Fake reviews** are dishonest evaluations that do not reflect genuine consumer experiences.
- **Trolling** involves posting inflammatory or offensive content to provoke or disrupt.
- **Imposter content** impersonates credible sources, such as using fake logos or journalist names, to lend false legitimacy. If none of these categories apply, the label "Other" should be used.

Verifiability refers to the extent to which information can be confirmed using reliable sources. If the content is fully supported by trustworthy evidence, it is considered verifiable. When no credible sources are available to confirm the information, it is marked as unverifiable. In some cases, only parts of the information can be verified, while other parts remain unconfirmed due to a lack of sources or because they involve opinions, rhetorical questions, or speculation about future events.

Facticity refers to the degree to which a statement or piece of content is based on verified facts. This category is used when the verifiability of the information has already been assessed. If the statement is entirely untrue and lacks any factual basis, it is classified as false. If it contains some truth but is mostly misleading or inaccurate, it is considered mostly false. A statement that is largely accurate but may miss minor context is labeled mostly true. When the information is fully accurate and supported by verified evidence, it is categorized as true.

The next category of the codebook refers to **format**: the type of content used in a post, and multiple formats may apply. Short text includes posts with 1–50 words, while long text ranges from 50 to 200 words. Visual formats include photos, videos, drawings, photomontages (images made from parts of multiple photos), and graphics such as posters or flyers. Memes are also common—these are humorous or opinionated images or videos tied to current events or cultural references. Deepfakes are highly realistic AI-generated media that alter someone's appearance or voice, while AI-supported content shows signs of artificial intelligence use without being a deepfake. Posts may also include audio or links to external sources like websites, blogs, or podcasts. If none of these formats apply, the content is categorized as "Other."

Next, **journalistic genres** refer to the different styles and structures used in journalistic writing.
- A **news article** is a fact-based, objective report that answers the key questions—what, where, when, how, why, and who—usually following the inverted pyramid structure.
- A **feature story** explores a topic or person in greater depth and with more narrative detail, while remaining factual and informative.
- **Reportage** combines eyewitness accounts, background context, and expert insights to create vivid, immersive storytelling.
- The **interview** genre includes several types: fact-and-opinion interviews that clarify current issues; portrait interviews that reveal a person's character or worldview; explanation interviews that provide expert insight into complex topics; and vox populi interviews that gather public opinion through brief street-level responses.
- The **opinion genre** includes columns, editorials, reviews, and essays, which express the writer's personal viewpoint, supported by arguments and evidence.
- If a piece does not fit any of these categories, it is coded as **Other**.

1.2. Section "WHERE"

The category **platform** refers to the digital space where the content appears. This includes major social media platforms such as Wikipedia, Facebook, Instagram, X (formerly Twitter), TikTok, YouTube, Reddit, Telegram, and Bluesky. It also covers various types of online media outlets: privately owned outlets that follow journalistic standards; public service media funded to serve the public interest; and state-owned media, which are government-funded and often promote official narratives. Other platforms include websites run by companies, NGOs, or universities, as well as official institutional or corporate websites. Blogs are also included as a distinct category. If the content does not fit any of these, it should be categorized as "Other."

Geographical Focus of message refers to the level or scope of the message's intended audience or relevance. It can be national, targeting an entire country; local, focusing on a specific region, municipality, or community; European, addressing issues related to the European Union; or global, with relevance beyond national or regional borders. Some

messages may have a regional focus, covering multiple countries within a specific area such as the Baltic States or Central and Eastern Europe. If the geographical scope is unclear or combines several levels (excluding national + European), it should be marked as impossible to identify or other.

The **Geographical Attribution of Narrative** is an open-ended category where the specific country or region from which the narrative originates should be noted (e.g., Russia, Ukraine, Western Balkans).

Language identifies the language in which the message is communicated. Options include French, Italian, Romanian, Estonian, Latvian, Lithuanian, Russian, English, or Other if the language does not fall into the listed categories.

1.3. Section "WHO"

The next categories deal with the target of disinformation and the entity that has produced it.

Target – Victim Typology identifies who is being targeted or harmed by the disinformation. This may include politicians and their relatives, government or international institutions, business or NGO representatives, professional associations (including unions), communities, individual citizens, journalists or editors, and experts or academics. If the target does not fit these categories, it should be marked as "Other." A specific individual or group can also be named in the open-ended field Target – Victim of the Attack.

Author – Emitter Type refers to the source or publisher of the content. This could be a journalist, politician, government or international institution, expert, NGO representative, union or professional body, activist, known proxy influencer, or a random individual. If the author is unknown or doesn't fit these categories, use "Unknown" or "Other." The Author – Emitter field allows for naming a specific individual or entity responsible for the narrative.

Community describes the intended audience. If the content is meant for the general public, it is coded as Public. If it targets a specific group or is shared in a closed setting, it is marked as Restricted/Closed/Limited. Use "Other" if the audience doesn't fit these categories. A more precise group can be named in the Specific Community field.

The **Source** field captures any referenced material or publication mode, such as whether the post is an original or a shared piece, and if it includes additional commentary.

Finally, **Ideological Orientation of the Author** refers to the political or philosophical stance reflected in the content.
This may include:
- **Conservatism**, which resists radical change and emphasizes tradition and limited governance;
- **Christian Democracy**, which blends Christian ethics with democratic values and social welfare;

- and **Liberalism**, which prioritizes individual freedom, pluralism, and constitutional governance.
- **Social Democracy** supports state intervention and welfare to promote equality,
- while **Communism** seeks a classless society through centralized control.
- **Anarchism** rejects state authority in favor of decentralized, cooperative systems.
- **Economic Libertarianism** champions free markets and minimal government interference.
- **Green Ideology** focuses on environmental sustainability and democratic reform.
- **Nationalism** emphasizes national identity and self-determination, often combined with other ideologies.
- **Fascism** promotes strong leadership and national unity, often through authoritarian means.
- **Populism** pits the "pure people" against a "corrupt elite," and can align with various broader ideologies.
- **Republicanism** values civic virtue and critiques concentrated power.
- **Imperialism** supports or critiques empire-building based on ideological narratives.
- **Feminism** challenges patriarchal systems and advocates for gender equality and structural change.

1.4. Section "HOW"

The next group of categories deals with "how" disinformation is spread, forming the core of the semantic-axiological matrix and representing a key innovation of the PROMPT project. It builds on the work of the European Commission's [Joint Research Centre](#), and is expanded using scientific literature and resources of the Universities of Cambridge and Oxford, Merriam-Webster dictionary, etc.

- **Imitated Content Structure** refers to the way disinformation mimics familiar formats to appear credible or persuasive.
- A **news-like (fact)** structure presents information in a factual tone, resembling traditional news reporting.
- **News-like + comment** adds evaluative or emotional commentary to factual claims. An **argument** structure includes reasoning or claims of proof, often suggesting causation or correlation.
- **Official statements** come from institutions or organizations and may include disinformation presented as formal declarations.
- **Allegations** involve explicit accusations or claims of legal or ethical violations.
- **Opinion** pieces reflect personal views, generalizations, predictions, or rhetorical questions.
- **Individual experience** mimics personal social media posts, often sharing anecdotal stories to evoke emotion or credibility.
- If the content does not fit any of these formats, it is categorized as **Other**.

Information Manipulation and Persuasion Techniques (IMPT) are the strategies used to influence audiences—either deceptively or persuasively. Information manipulation techniques are often covert and unethical, aiming to mislead or distort reality by exploiting emotions and cognitive biases. Persuasion techniques, on the other hand, use rhetorical and linguistic tools to shape opinions and attitudes, not necessarily through deception.

The first group of IMPTs involves **attacks on reputation** with an intention **to undermine a person's credibility**. Such attacks employ various techniques.
- **Name calling** or labelling assigns derogatory terms to discredit individuals or groups.
- **Guilt by association** links someone to a negative figure or ideology.
- **Casting doubt** subtly undermines credibility without direct accusations.
- **Appeal to hypocrisy** deflects criticism by accusing the opponent of similar faults.
- **Smears or poisoning the well** preemptively discredit a source.
- **Ad hominem** attacks target the person rather than their argument.
- **Trolling**, including gaslighting and tone policing, seeks to provoke or derail discussion.
- The **genetic fallacy** dismisses information based on its origin rather than its content.
- **Pathologization** frames opponents as mentally unstable.
- The **untrustworthy communicator** technique labels someone as inherently unreliable.
- The **Galileo gambit** falsely equates being criticized with being correct.
- **Stereotyping** reduces individuals to oversimplified and often harmful generalizations.
- Finally, **hahaganda** uses humor or satire not to inform or entertain, but instead to ridicule and undermine the credibility of individuals or institutions, thereby diverting attention from their arguments by portraying their authors as inherently unacceptable through mockery. While it may appear lighthearted, its goal is to erode trust and trivialize serious matters through repeated mockery.

The second group of IMPTs refers to an **attack on the argument to justify** why something is good or bad. Such justification includes techniques to legitimize a claim, often by appealing to emotion, authority, or cultural norms rather than evidence.
- **Flag-waving** appeals to patriotism or loyalty to justify a stance.
- **Appeal to authority** uses expert opinion to support a claim, while irrelevant authority misuses expertise from unrelated fields.
- **Appeal to popularity** (or bandwagoning) suggests something is true or good because many people believe it. Appeal to values invokes shared moral or cultural beliefs, and appeal to fear or prejudice manipulates emotions to justify a position.
- **Confirmation bias** reinforces pre-existing beliefs by selectively presenting information.
- **Proof by example** uses isolated cases as evidence for broader claims.
- **Cultural biases** rely on stereotypes or assumptions rooted in cultural norms.
- **Appeal to tradition** argues something is right because it has always been done that way.
- **Argumentum ad captandum** uses emotionally charged language to win approval.

- **Appeal to ignorance** claims something is true because it hasn't been proven false.
- **Appeal to common sense** and **appeal to nature** both rely on what seems "obvious" or "natural" rather than evidence.
- **Appeal to force** justifies a claim through threats or coercion. These techniques aim to make arguments seem reasonable or justified, even when they lack factual support.

The third group of IMPTs employs an **attack on the argument to distract** the attention from the topic. Such distraction includes techniques that divert attention away from the main issue, often by misrepresenting or shifting the focus of the discussion.
- One common method is the **strawman fallacy**, where an argument is invented, distorted or exaggerated to make it easier to counter it. Instead of addressing the original point, a weaker version is refuted.
- **Red herring** introduces unrelated information to distract from the topic at hand, steering the conversation in a different direction.
- **Whataboutism** shifts focus by pointing to another issue—often unrelated or worse—to deflect criticism or avoid accountability.

The fourth group of IMPTs refers to an **attack on the argument to simplify** the logic behind causes and consequences. Such **simplification** involves reducing complex issues to overly superficial explanations, often distorting the logic behind causes and consequences.
- **Causal oversimplification** attributes an outcome to a single cause, ignoring other contributing factors.
- **False dilemma** (or black-and-white fallacy) presents only two extreme options, excluding nuanced alternatives.
- **Consequential oversimplification**, such as the slippery slope, suggests that one action will inevitably lead to extreme consequences.
- **Relativization** downplays the seriousness of an issue by comparing it to something else.
- **Generalization** makes broad claims based on limited evidence.
- **Cherry picking** or selective omission highlights only favorable facts while ignoring contradictory ones.
- **False cause** assumes a causal link where none exists,
- and **false attribution** misrepresents the source of a claim.
- The **politician's syllogism** uses flawed logic to justify action: "Something must be done; this is something; therefore, this must be done."
- **Vagueness** avoids clarity to obscure meaning, while circular reasoning uses the conclusion as its own proof.
- **Causal reductionism** simplifies complex causal relationships, and far-fetched hypotheses introduce implausible explanations to distract or mislead.

The fifth group of IMPTs involves an **attack on the argument to encourage thinking in a particular way.** Such calls include techniques designed to steer thinking in a particular direction by using emotionally charged or manipulative cues.

- **Slogans** are short, memorable phrases that distill complex ideas into persuasive soundbites, designed to evoke emotions and guide people's thinking in a specific direction. Repeated frequently, they reinforce a message and simplify nuanced issues.
- **Conversation killers**, also known as thought-terminating clichés, are phrases used to shut down debate or critical thinking (e.g., "It is what it is" or "That's just how things are").
- **Appeal to time** (Kairos) emphasizes urgency or the "right moment" to act, often pressuring audiences to respond without reflection.
- **Astroturfing** refers to the creation of fake grassroots movements or public support, giving the illusion of widespread legitimacy and consensus.

The sixth group of IMPTs refers to an attack on the argument by manipulative wording. Such techniques involve using emotionally charged or strategically vague language to influence perception and trigger emotional responses.
- **Loaded language** uses words with strong positive or negative connotations to sway opinion.
- **Obfuscation**, **intentional vagueness**, and **confusion** present unclear messages, making it harder for audiences to critically assess their content.
- **Exaggeration** or **minimisation** distorts the scale or importance of an issue, often to provoke or downplay concern.
- **Repetition** reinforces messages through constant exposure, making them more memorable and seemingly credible.
- **Fearmongering** uses alarming language to incite fear and urgency.
- **Dog whistles** are coded messages that appear neutral but carry specific meanings for targeted groups.
- **Puns** and **phrasemes** involve the creative manipulation of language—either by playing with words' sounds, spellings, or meanings (puns), or by using common expressions and idioms (phrasemes)—to subtly introduce double meanings or insinuations. In disinformation and hate speech, such techniques are often used to convey coded or offensive messages without making them explicit. For example, altered words like "iSSrael" or repurposed idioms can disguise hateful content while preserving deniability, making these tools particularly effective for indirect persuasion or veiled attacks.

IMPTs can be further understood through **rhetorical figures**, as these are strategically employed in the implementation of manipulation techniques. Disinformation narratives contain numerous rhetorical figures, as well as language and speech techniques used to influence or persuade an audience. In the context of the PROMPT project, these techniques are analyzed for how they are used to make disinformation or misinformation more convincing and emotionally impactful.

Multiple rhetorical figures can be present in a single message.

- **Amplification** and **exaggeration** are rhetorical techniques used to intensify communication, making it more persuasive and emotionally engaging.
- **Hyperbole** involves deliberate overstatement to provoke strong reactions (e.g.: "This election could change Europe forever!" or "This is the worst policy in our country's history,").
- **Climax** builds momentum by arranging ideas in increasing importance, as in "We will fight for justice, for equality, for freedom."
- **Epanorthosis** reinforces an idea by immediately correcting or rephrasing it for emphasis, such as "The war is terrible—no, catastrophic." These figures are commonly used in political and persuasive discourse to produce a lasting impact.

Comparison, **contrast**, and **false equivalencies** are rhetorical strategies that link or oppose ideas—often misleadingly—to persuade or evoke emotion.
- **Allusion** are indirect references to historical or cultural events or symbols to evoke associations or draw comparisons—sometimes to persuade or dramatize a situation. For instance, calling a challenge "our Dunkirk moment" or referring to a threat as "a Trojan horse" connects present events to powerful past narratives, often shaping perception subtly.
- **Antithesis** presents opposing ideas in a balanced way, such as "Freedom vs. oppression" or "We fight for peace, they seek war."
- **Oxymoron** combines contradictory terms—e.g., "peacekeeping forces" or "deafening silence"—to provoke thought or mask contradictions.
- **Paradox** uses seemingly self-contradictory statements that reveal deeper truths, like "We are at war for peace" or "The only constant in politics is change."
- **Metaphor** implies a comparison, as in "The war is a ticking time bomb," while simile makes comparisons using "like" or "as," such as "Misinformation spreads like wildfire." These devices often blur distinctions or create emotional resonance by equating things that aren't truly alike.

Repetition is a powerful rhetorical tool used to emphasize key messages and make them more memorable, reinforcing ideas through repeated phrases, such as "They lie; they lie; they lie!".
- **Epistrophe** repeats words at the end of successive clauses (e.g: "We must fight for truth; we must fight for our future.").
- **Anaphora** starts consecutive clauses with the same phrase (e.g., "We need truth. We need transparency.").
- **Epanalepsis** repeats the beginning of a sentence at its end—"The war must end, the war."
- **Epizeuxis** uses immediate repetition for emphasis: "Misinformation, misinformation, misinformation!"
- **Symploce** combines anaphora and epistrophe: "When there is misinformation, we must fight it. When there is truth, we must spread it."

- **Adnomination** repeats root words or derivatives, like "Fight for freedom, fight for fairness."
- **Anadiplosis** links clauses by repeating the last word of one at the start of the next: "Fear leads to anger. Anger leads to hate."
- **Polyptoton** repeats variations of the same root, as in "We must fight the fighters and defeat the defeated." These techniques reinforce messages, stir emotion, and increase rhetorical impact.

Questioning is a rhetorical strategy used to provoke thought, create doubt, or guide the audience toward a specific conclusion.
- **Hypophora** involves asking a question and immediately answering it, such as "What can we do about misinformation? We can educate ourselves and demand accountability," or "Is this policy effective? Absolutely not." This technique steers the audience by framing both the question and the answer.
- **Rhetorical questions**, on the other hand, are asked for effect and don't require an answer—like "How can we trust the government when they keep changing their story?" or "Is it really safe for our kids?" These questions imply doubt or criticism and encourage the audience to reflect or agree without needing a direct response. Both forms are commonly used to challenge assumptions, stir emotions, and reinforce a message.

Sound and rhythm are rhetorical tools that enhance the memorability and emotional impact of language.
- **Alliteration** repeats initial consonant sounds to create a catchy, rhythmic effect — "Fake news fuels fear and frustration."
- **Assonance: the similarity in sound between two syllables that are close together, created by the same vowels but different consonants (e.g. "back" and "hat"), or by the same consonants and different vowels (e.g. "hit" and "hat")**
- **Onomatopoeia** uses words that imitate sounds - e.g.: "the buzz of misinformation" - to make messages more vivid and sensory.
- **Homoioteleuton** involves repeating similar word endings for a rhythmic or poetic feel, as in "The war is a bore, a chore, and a snore." These techniques help grab attention and make messages penetrate in the audience's mind.

Deceptive substitution and **indirection** are rhetorical strategies used to mislead or obscure meaning.
- **Euphemism** softens harsh realities with milder terms, like calling civilian casualties "collateral damage" or disinformation "alternative facts."
- **Irony** subtly exposes contradictions between appearances and reality, often by stating the opposite of what is meant, to highlight inconsistencies or hidden truths. For example, saying "We value honesty in politics" when dishonesty is evident points to underlying hypocrisy.

- **Metonymy** replaces a concept with something closely related—e.g., using "The Kremlin" to refer to the Russian government.
- **Synecdoche** substitutes a part for the whole, such as "All hands on deck" to mean all crew members.
- **Paralipsis** draws attention by pretending to ignore something: "I won't even mention the misinformation spread by certain media outlets." Similarly, apophasis mentions something by stating it won't be mentioned, like "I won't talk about the lies told by our opponents." These techniques subtly shape perception and can manipulate understanding.

Storytelling and **Anecdotes** use personal or relatable stories to illustrate a point in an engaging way. An anecdote is a brief, often amusing or insightful account of a real event that helps connect emotionally with the audience. For example: "I remember when misinformation led to panic buying during the pandemic." This technique adds authenticity and makes abstract issues feel more tangible and human.

Clever derisive comments use humor, irony, or mockery to criticize or highlight the absurdity of a situation.

- **Irony** can be employed in this context as a way of saying the opposite of what is meant to emphasize contradictions or absurdities, often using humor or sarcasm to highlight hypocrisy or provoke reflection—for instance, remarking, "It's ironic that those who claim to protect democracy spread the most misinformation."
- **Sarcasm** is a sharper, often mocking form of irony, where the speaker deliberately says the opposite of what they mean to ridicule or provoke, making their true intention obvious—for example: "Yeah, because a little 'special military operation' is nothing to worry about." This technique is often used to undermine opponents or stir strong emotional reactions.

Subtle downplaying expressions are rhetorical techniques used to soften the impact of a statement, often to convey modesty, irony, or politeness.
- **Understatement** minimizes the importance or severity of something, as in "It's just a small issue," even when the issue may be significant.
- **Litotes** is a form of understatement that affirms something by negating its opposite—for example, saying "not bad" to mean "good." These expressions can make messages seem more measured or indirect, while still conveying a clear point.

Wordplay and clever language use creative phrasing to engage, entertain, or persuade—often drawing on cultural or historical references.
- **Puns (paronomasia)** rely on playful wordplay, using words with similar sounds or multiple meanings to create humorous, witty, or rhetorically engaging effects—for instance, "Vote for change, not small change," or "trans-forming society." This technique is often used to make messages more memorable, entertaining, or clever.

- **Neologisms** are newly coined terms that capture emerging ideas or movements—examples include "fake news," "Brexit," or "wokeism."
- **Idioms** are familiar expressions with figurative meanings that are often culturally specific and must be learnt as a whole - for instance, "elephant in the room" or "barking up the wrong tree," often used to simplify complex issues or make messages more relatable. These techniques add flair and memorability to communication.

Logical reasoning uses structured argumentation to support a conclusion.
- A **syllogism** is a classic form of reasoning where a conclusion is drawn from two premises. For example: "Only aggressive nations invade other countries. Russia is not an aggressive nation (false premise). Therefore, Russia did not invade Ukraine." This structure can appear logical on the surface, but if the premises are flawed or misleading, the conclusion will be too. Syllogisms are often used to give arguments an appearance of rationality, even when based on false assumptions.

Parallelism and **structure** are rhetorical techniques that use repetition and symmetry to create rhythm, clarity, and emphasis.
- **Chiasmus** reverses the order of words or concepts for impact, as in "Ask not what your country can do for you—ask what you can do for your country."
- **Antimetabole** is a specific type of chiasmus that repeats exact words in reverse order, like "Russia defends its borders, its borders defend Russia."
- **Isocolon** uses parallel structure in successive phrases or clauses of equal length, such as "Many will enter, few will win."
- **Tricolon** builds momentum with three parallel elements, as in "Government of the people, by the people, for the people." These patterns make messages more memorable and persuasive through balance and rhythm.

Conjunctions and **omissions** are rhetorical techniques that shape the rhythm and emphasis of a sentence.
- **Polysyndeton** uses multiple conjunctions in close succession to slow the pace and build intensity—for example: "We have lies and deceit and corruption and greed."
- **Asyndeton**, by contrast, omits conjunctions to create a faster, more dramatic effect, as in "We came, we saw, we conquered." Both techniques influence how a message is received, either by overwhelming the listener or delivering a punchy, memorable impact.

Breaking off and pausing is a rhetorical technique used to create suspense, convey strong emotion, or suggest unspoken consequences.
- **Aposiopesis** occurs when a sentence is deliberately left unfinished, allowing the audience to fill in the rest. For example: "If misinformation continues, then…" This sudden break can imply urgency, frustration, or a threat without stating it outright, making the message more dramatic and emotionally charged.

Reversal and **inversion** is a rhetorical technique that rearranges the usual word order to create emphasis, add drama, or draw attention to specific ideas.

- **Anastrophe** inverts the typical structure of a sentence, often giving it a poetic or striking tone. For example: "Powerful you have become." This unusual phrasing slows the reader down and highlights the key word or concept, making the message more memorable and impactful.

Emphasis and **redundancy** are rhetorical techniques used to simplify and reinforce messages, making them more memorable and persuasive—often by repeating ideas or using catchy phrasing.

- **Slogans** can be crafted to simplify and amplify political or cultural messages, making them easy to remember and repeat—for instance, "Make America Great Again" or "Take Back Control." They are often used to unify supporters and reinforce campaign themes.
- **Pleonasm** uses more words than necessary, like "new innovation," to emphasize a point even if it may be redundant.
- **Tautology** repeats the same idea in different words, as in "free gift." These techniques can make statements sound more convincing or authoritative, even when they lack substance.

Sensory descriptions use vivid language to engage the senses and evoke strong emotional responses.

- **Synesthesia** is a technique where one sense is described in terms of another, creating rich and memorable imagery. For example: "The silence was as thick as a blanket." This blending of sensory experiences helps paint a more powerful picture in the audience's mind, making the message more immersive and emotionally resonant.

Diminishment and **minimization** are rhetorical strategies used to downplay the seriousness or importance of an issue.

- **Tapinosis** achieves this by using belittling or trivializing language—for example: calling a significant battle "just a skirmish." This technique can make major events seem minor, reducing their perceived impact and shaping public perception.

Abuse and criticism are rhetorical strategies used to provoke strong negative emotions and discredit a target through harsh or contemptuous language.

- **Bdelygmia** involves expressing disgust or hatred by using abusive descriptions—for example: "You are a liar, a cheat, and a fraud," or "Those corrupt politicians are ruining our country." It can also include labeling groups or individuals with derogatory terms, such as calling disinformation "toxic lies" or using dehumanizing language to attack marginalized communities. This technique aims to demonize opponents and rally emotional support through outrage.

Imagery and symbolism use vivid language and figurative comparisons to convey deeper meanings and stir emotions.

- **Metaphor** compares two unrelated things without using comparative terms such as "like" or "as", thereby creating strong visual or symbolic associations—for example: "Draining the swamp" to suggest eliminating corruption, or describing Ukraine as a "puppet state." Other examples of metaphors include "The EU is a ship navigating stormy seas" or "The fight for LGBTQ+ rights is a cancer on society."
- **Personification** gives human traits to non-human entities, making abstract ideas feel more relatable or emotional—such as "Fear gripped the nation" or "Europe stands at a crossroads." These techniques help shape perception through powerful, emotionally charged imagery.

Humor and ridicule use wit, irony, and exaggeration to criticize or mock individuals, ideas, or institutions.

- **Parody** imitates a style or speech in a humorous way to highlight its flaws or absurdity—for example, a fake headline like "Donald Trump Says the Earth is Flat," or a mock political speech that exaggerates clichés.
- **Sarcasm** uses biting or exaggerated irony to express criticism or scorn, typically through statements that sound positive but clearly imply the opposite—such as: "Oh great job on the economy, really fantastic," or "Sure, let's just let anyone identify however they feel!" This form of humor aims to ridicule or provoke reactions through sharp, cutting remarks.
- **Satire** blends humor, irony, and exaggeration to expose and criticize, often seen in political cartoons or satirical articles like "The EU will make us all eat bugs!" These techniques entertain while delivering pointed critiques.

Emotional manipulation involves using language and tone to influence feelings and sway opinions, often by appealing to fear, contempt, or false modesty.

- **Accismus** refers to the rhetorical device in which someone pretends to reject something they actually desire—for example: "Oh, I don't care about winning the election, but it would be nice if I did," or dismissing an issue as unimportant while continuing to promote it through jokes or memes.
- **Bdelygmia** uses harsh, abusive language to express disgust or hatred, such as calling opponents "corrupt" or describing disinformation as "toxic lies."
- **Dysphemism** replaces neutral or positive terms with harsh, offensive ones—for instance, saying "illegal aliens" instead of "undocumented immigrants," or calling a peace treaty a "surrender." These techniques are designed to provoke strong emotional reactions and shape public perception through negativity and manipulation.

Lastly, **cultural and historical references** use familiar events, phrases, or symbols to persuade by tapping into shared knowledge and emotions.

- In this context, **allusions** are employed to refer to well-known historical, cultural, or literary events, figures, or symbols to create connections and resonance with the

audience. Expressions like "This is our Dunkirk moment" or "A Trojan horse in our midst" draw on shared cultural memory to frame current situations in a familiar and impactful way.

- **Idioms** are common expressions with figurative meanings, like "elephant in the room" or "barking up the wrong tree," often used to simplify complex issues in a relatable way.
- When employed as cultural and/or historical references, **slogans** encapsulate political, social, or ideological messages, drawing on cultural familiarity to make them instantly recognizable and memorable—for example: "Make America Great Again" or "Take Back Control."

Categories of Lexicon refer to the specific words or phrases used in a message. The focus is on analyzing the actual language—individual terms or expressions—rather than the overall tone or emotional feel of the content. Multiple lexical categories may apply to a single message.

The Axiological-Semantic Categories include three types of word connotations.
- **Positive connotation** refers to words that express favorable or uplifting meanings, such as political ideals ("freedom," "democracy," "unity"), social justice terms ("equality," "inclusivity," "rights"), humanitarian values ("compassion," "support," "help"), national pride ("patriotism," "heritage," "sovereignty"), and progressive change ("innovation," "progress," "future").
- **Negative connotation** includes words that evoke criticism, fear, or hostility, such as those used for demonization ("traitor," "enemy of the state," "terrorist"), marginalization ("deviant," "pervert," "radical"), victimization ("victims of political correctness," "silenced majority"), fearmongering ("invasion," "collapse," "chaos"), and stigmatization ("groomer," "indoctrinator").
- **Neutral connotation** encompasses emotionally neutral or objective terms, including descriptive words ("election," "conflict," "community"), technical vocabulary ("legislation," "diplomacy," "policy"), statistical terms ("percentage," "rate," "average"), and geographical names ("Ukraine," "Brussels," "Berlin").

The **Axiological Framing Dichotomy** category captures oppositional framings where individuals or groups are positioned against perceived "others," whether political, institutional, or identity-based. It includes contrasts such as: "we/me vs the national political elite," "we/me vs the EU or global political elite," "we/me vs a particular country," "we/me vs a specific institution or company," "we/me vs a particular ethnic group," "we/me vs a particular race," "we/me vs a particular religious group," "we/me vs any part of the LGBTQ+ community," "we/me vs a particular culture," and "we/me vs unspecified others." If no such dichotomy is present, it should be noted accordingly.

Lastly, **Categories of Language** identify different types of language use based on intent and rhetorical effect.

- **Indicative language** presents statements as factual or verifiable, often used to frame misleading claims as truth, such as asserting that "most European citizens support extreme right parties" during the 2024 EU Elections, which oversimplifies complex realities.
- **Imperative language** issues commands or calls to action, like "Support our troops now!" or "Join the fight against Russian aggression!" which aim to mobilize public sentiment.
- **Conditional language** suggests outcomes based on hypothetical scenarios, often to instill fear or urgency, as in "If the EU does not act against Russia, we risk losing our sovereignty."
- **Debitive language** expresses obligation or necessity, framing issues as urgent or morally imperative, such as "We must protect LGBTQIA+ rights at all costs!" which may marginalize dissenting views.
- **Conjunctive language** links ideas to imply associations that may not exist, for example, "The rise in LGBTQIA+ rights is directly linked to the decline of traditional family values," which attempts to provoke emotional reactions.
- **Presumptive language** assumes shared beliefs or facts without evidence, such as "Everyone knows that Ukraine is winning the war," which shapes perception by implying consensus where it may not exist.

1.5. Section "WHY"

The last segment of categories deal with the reason **why disinformation is spread**. Our categorisation here is aligned with the 5D Framework, initially developed by Nimmo (2015) and later updated by Collab (2019), to classify the strategic objectives of disinformation campaigns. This classification is widely accepted and adopted by other disinformation frameworks and renown institutions, namely DISARM or the EEAS. First, the **intention behind the information** may or may not be identifiable. If an intention is present, the **aim or purpose of the message** can fall into several categories, including:
- to **dismiss** criticism or **discredit** sources,
- to **distort** facts or narratives,
- to **distract** by shifting attention or blame,
- to **dismay** by instilling fear or threats,
- or to **divide** by deepening social or group conflicts.
- Any other purpose not captured by these categories is marked as "other".

Additionally, the **level of intended effect** (also multiple selections allowed) identifies the target of the message's impact: it may be aimed at an individual, an institution, society at large, a specific community, or fall into an unspecified or other category.

Altogether, the previously described categories of the codebook (semantic-axiological matrix) form the basis for automated content analysis in the PROMPT project, aimed at uncovering not only disinformation narratives themselves but also the complex mechanisms through which they are constructed.

2. Network analysis

In PROMPT project's network models, **nodes** represent entities such as individual users, social media accounts, content (e.g., tweets, URLs, memes), or abstract narrative units. **Edges** capture relationships based on shared behaviors — like reposting the same content, using identical hashtags, or referencing similar narratives — and are weighted by factors such as co-share frequency, synchronicity, or mutual engagement. Temporal dynamics are integrated through:

1. **Time-stamped edges** – each interaction includes a timestamp.
2. **Time-sliced networks** – the network is divided into snapshots for comparison across events or periods.

This enables PROMPT to track shifts in coordination, narrative spread, and community structure over time.

In PROMPT, network construction uses **multilayer modelling** to capture coordinated behavior across content types, modalities, and platforms. Each layer represents a distinct coordination mode, such as URL sharing, hashtag reuse, image duplication (via perceptual hash), narrative alignment (from LLM annotations), textual near-duplicates (via cosine similarity) and cross-platform actions (e.g., synchronized posts on Facebook and X/Twitter). As of now, the implementation works in silo for each of these layers and the next update merges them for the complete network.

CooRTweet supports this architecture by detecting coordination separately for each modality (e.g., URLs, images) within a shared time window, then stacking the results into a composite coordination graph. This graph reflects multimodal coordination, where actors may engage across different content types without sharing identical objects. The combined results feed into the generate_coordinated_network routine, producing a graph with edges weighted by number of shared objects, edge symmetry and mean time delta between shares. This results in a **multimodal, multilayer network**, where actors can appear in multiple layers, each representing a different coordination mechanism. By supporting diverse object types — from URLs to images to semantic narratives — CooRTweet enables unified, cross-platform network analysis. This is especially valuable for disinformation research, where mixed media formats often reinforce shared narratives.

CooRTweet is a lightweight, content- and platform-agnostic tool for identifying coordinated activity. It defines coordination as repeated, near-synchronous sharing of uniquely identifiable objects (Pacheco, Flammini, & Menczer, 2021; Righetti, N., & Balluff, P., 2023). The tool requires four fields:

1. **object_id** – Unique identifier for the content (e.g., URL, hashtag, narrative ID)
2. **account_id** – ID of the account sharing the content
3. **content_id** – ID of the post containing the object

4. **timestamp_share** – Time of sharing (UNIX format)

This minimal setup allows CooRTweet to work across platforms and content types. In PROMPT analysis, **object_id** represents narrative units, enabling detection of coordination around shared narratives—not just identical content. This is made possible by upstream use of **Large Language Models (LLMs)**, which classify each post into a narrative cluster (defined by topic and stance). These narrative IDs are then used as **object_id** in CooRTweet to detect coordination.

PROMPT methodology offers two entry points:
1. **Narrative-first approach**: Starting from a dataset of posts (e.g., tweets about the war in Ukraine), we detect narratives via LLMs, assign narrative IDs to each post, and run CooRTweet to identify users who share the same narratives synchronously.
2. **Actor-first approach**: Beginning with a list of known or suspected coordinated accounts (e.g., from prior disinformation campaigns), we monitor their output in real time, detect high-performing posts, and iteratively discover new actors sharing similar content.

This dynamic and recursive process allows us to construct and update coordination networks over time, even as actors and narratives evolve.

The PROMPT project's approach improves on earlier methods (e.g., NODES) by: using **LLMs for scalable, objective narrative annotation**, detecting **semantic coordination**, not just identical content, adding a **temporal layer** via CooRTweet's time-windowed detection and focusing on **repetition and synchronicity** to reveal genuine coordination, ideal for spotting influence operations and disinformation.

2.1. Temporal Network Analysis Following Saqr

To complement content-based coordination detection methods such as CooRTweet, the PROMPT project also employs temporal interval network analysis, drawing on the framework introduced by Saqr (2024). This approach explores the temporal proximity of user activity rather than relying on semantic or content-based alignment. It is particularly effective for capturing the rhythms of engagement, identifying spontaneous attention bursts, and tracing the temporal evolution of online interactions.

Unlike methods that detect coordination through repeated content like hashtags or URLs, **Saqr's approach links users based on activity within the same time window** (e.g., 30 seconds, 5 minutes, 1 hour). These time-based connections suggest shared attention to external events—such as breaking news or disinformation—regardless of the specific content used. This makes the method especially effective in high-traffic settings (e.g., Twitter during crises), where users often respond to the same topic without using identical language or media.

Constructing temporal networks using the Saqr method involves several adjustable parameters that shape how user interactions are detected and analyzed. One key parameter is **time resolution**, which determines the granularity of time intervals—such as 1 minute, 5 minutes, or 1 hour. Finer resolutions are more sensitive to rapid bursts of activity, while coarser intervals help filter out noise and emphasize sustained engagement. The ideal resolution depends on the platform (e.g., Twitter vs. YouTube) and the nature of the event, whether it's a sudden news break or a slow-developing narrative.

Another important aspect is the **windowing logic**, which defines how time intervals are structured. Fixed windows divide time into non-overlapping segments, making computation straightforward and comparisons consistent. Sliding windows, which overlap, are better suited for capturing rolling bursts of activity. Event-based windows, triggered by specific incidents like viral tweets or public announcements, allow analysts to focus on activity immediately before and after key moments. The choice of windowing logic influences whether the resulting network emphasizes responsiveness or coordination.

Edge formation rules determine how users are connected within these time windows. In symmetrical pairing, all users active in the same window are linked to each other, forming complete subgraphs. Asymmetrical logic connects the first user in a window to all subsequent ones, highlighting potential influence or seeding effects. Alternatively, thread-based logic links users only if they engage with the same message or tag the same account, focusing more on direct engagement than on timing alone.

Edges in the network can be either binary or weighted. **Edge weighting** schemes vary: some reflect how often two users appear in the same time window, others apply a decay function based on how close their activity times are, and some incorporate content similarity, giving more weight to users who share topics or media types. These variations help distinguish between loose synchrony and strong behavioral alignment.

To reduce noise, **node filtering and thresholding** can be applied. Users may be included only if they exceed a minimum level of activity or appear in a certain number of time windows. This helps isolate core participants in dynamic narratives.

Finally, once the network is built, its **temporal dynamics** can be explored in different ways. The network can be aggregated into a single structure, divided into snapshots (such as daily segments) for comparative analysis, or animated to visualize how coordination and activity evolve over time.

Together, these features make the Saqr method a flexible and powerful framework for studying temporal coordination—whether the goal is to detect emerging narratives, identify key influencers during crises, or compare coordination patterns across disinformation campaigns.

In addition to its technical flexibility, the Saqr method offers several analytical advantages There are also several analytical advantages to the Saqr method for analyzing the temporal

dynamics of online interaction. One of its key strengths is identifying **temporal centrality**—users who act as initiators, early responders, or connectors during critical time windows. These individuals often shape the direction of discourse and amplify narratives early on, even without large followings, making them vital for detecting emerging communities and fresh audience engagement.

The method also excels at revealing **cluster emergence**, where groups of users temporarily align around specific events or issues. These clusters often reflect situational coordination or opportunistic behavior, especially in fast-moving contexts like policy shifts, crises, or viral controversies. By analyzing **temporal patterns**, the method uncovers recurring motifs such as echo bursts, cascades, or delayed reactions. These patterns help explain how attention flows and whether narrative uptake follows predictable rhythms.

Another strength is detecting **narrative reactivation**—when dormant topics resurface, either organically or through strategic prompting. For example, reviving anti-LGBT rhetoric during election cycles. This is especially relevant for fact-checkers, who often encounter recycled narratives with slight variations over time. The method also supports **crisis reactivity tracking**, capturing how users respond in real time to external events like protests, military actions, or court rulings. By mapping these surges, analysts can identify which actors or narratives are primed for rapid mobilization.

When combined with semantic annotations—such as topic, stance, or sentiment labels from language models—temporal networks become even more powerful. This hybrid approach links timing with content, enabling researchers to trace the life cycle of narratives even when users avoid explicit keywords or use varied language.

The Saqr method is especially **valuable for disinformation research**, where coordination is often subtle, adaptive, and dispersed rather than overt or uniform. By focusing on t**emporal co-occurrence rather than content repetition**, it reveals behavioral alignments that traditional models may overlook.

Temporal network analysis helps surface low-visibility synchrony in several key ways. It captures the rapid spread of emotionally charged content that varies in wording but is shared within the same moment—common in disinformation targeting sensitive topics like migration, crime, or political elites. It also detects reactivity to signals from influencers, bots, or foreign media, where users respond to shared cues without using identical language, exposing distributed amplification networks that rely on timing rather than content.

The method is particularly effective during flashpoint events—such as riots, legislation, or scandals—where spontaneous framing alignments emerge without shared hashtags or URLs. Temporal clustering can reveal this early-stage coordination before semantic patterns become visible.

Additionally, it supports the detection of **narrative seeding and testing**, where disinformation actors introduce themes in one region or language and monitor short-term uptake elsewhere.

This helps identify cross-platform or cross-language "narrative probes" designed to test resonance.

As disinformation campaigns increasingly mimic organic behavior, detecting temporal convergence without content duplication becomes a critical early-warning tool. It enables researchers to flag emerging threats before they solidify into coherent narratives, enhancing the PROMPT project's ability to track and counter disinformation before it gains traction.

2.2. Comparison of CooRTweet and Saqr Methods

While both CooRTweet and the Saqr-based temporal method aim to uncover how narratives spread, they focus on different aspects of coordination. CooRTweet is designed to detect explicit coordination by tracking repeated sharing of identifiable content—such as links, hashtags, or claim IDs—within short time frames. This makes it especially effective for spotting orchestrated campaigns, automated behavior, and overt information operations. In contrast, the Saqr method captures **implicit co-attention**, linking users based on shared activity within the same time window, regardless of content overlap. This makes it better suited for identifying spontaneous or emergent engagement, including the activation of swing users or peripheral audiences who may not use the same language or media (Saqr, 2024; Pacheco et al., 2021).

Together, these methods offer a complementary view: CooRTweet highlights structured, content-driven synchronization, while Saqr reveals behavioral convergence over time. Their integration within the PROMPT pipeline enables a more nuanced understanding of how disinformation narratives spread—whether through deliberate coordination or organic diffusion.

2.3. Intra-Platform vs. Cross-Platform Analysis of Disinformation Narratives

Disinformation research typically follows two main analytical paths: **intra-platform analysis** (Nahon, K., & Hemsley, J., 2023; Ahmed, H. et al., 2022), focusing on how narratives evolve within a single platform, and **cross-platform analysis** (Barbero et al., 2023; Starbird, K., & Wilson, T. 2020), which examines how narratives spread and transform across multiple digital ecosystems.

Intra-platform analysis allows for detailed insights into platform-specific behaviors. Because platforms like Twitter or Facebook offer structured metadata—such as retweets, replies, and hashtags—researchers can build fine-grained network models that reflect how content circulates within that environment. This approach also helps assess how platform features, like recommendation algorithms or moderation tools, shape the visibility and amplification of disinformation. However, focusing on a single platform can create blind spots. It may overstate the coherence or ideological unity of a community, while missing how narratives often originate on fringe platforms and later migrate into mainstream spaces. Many coordinated campaigns unfold across multiple platforms, making a single-platform view too narrow.

Cross-platform analysis addresses this by tracing how narratives move, mutate, and gain traction across different digital spaces. It captures the broader information ecosystem, revealing how the same narrative may be framed differently depending on the platform's format (e.g., video on TikTok vs. text on Twitter) and audience. This approach is especially useful for understanding how fringe content becomes mainstream through cycles of exposure and adaptation (Phillips & Milner, 2021).. That said, cross-platform analysis comes with challenges. Data access and formats vary widely between platforms, making interoperability difficult. Metrics like follower count aren't directly comparable across platforms due to differences in visibility algorithms and user behavior. Aligning timestamps and resolving user identities across platforms adds further complexity. Despite these hurdles, **cross-platform research is essential for understanding the full lifecycle of disinformation—especially in multilingual or transnational contexts** (Hunt & Papakyriakopoulos, 2023).

In the PROMPT project, both approaches are combined. Intra-platform analysis reveals detailed patterns of interaction and narrative development within individual platforms. Cross-platform analysis, still in development, aims to track how content spreads across linguistic and cultural boundaries. When supported by rich, interoperable metadata, this dual approach offers a powerful lens for understanding how disinformation circulates in today's complex digital landscape.

Narrative Identification Methodology

As detailed earlier, the PROMPT project uses dynamic network analysis to study how disinformation narratives spread online. **A critical first step in this process is reliably identifying and classifying those narratives**. To achieve this, the project employs a multi-stage, multilingual methodology designed to detect, categorize, and trace policy-related claims and disinformation frames across social media platforms.

Given the complexity and linguistic diversity of online discourse the project adopts a **hybrid approach** that balances conceptual rigor with computational scalability. This strategy combines a theory-driven supervised classification pipeline with a semantic similarity–based filtering method. Together, these techniques offer both interpretability and flexibility, addressing the challenges of large-scale, multilingual data analysis.

The methodology integrates human expertise with advanced AI tools and consists of four interconnected components (see table below), each contributing to the accurate detection and tracking of evolving narratives.

Custom Codebook Development

To analyze discursive patterns in social media posts, the PROMPT project applied a quantitative approach powered by advanced AI techniques. The first step involved developing a custom coding scheme tailored to the project's focus. This scheme, structured as a codebook, draws on established frameworks such as the Comparative Agendas Project

(Baumgartner et al., 2019; Bevan, 2019) and the MARPOR/Comparative Manifestos Project (Budge et al., 2001; Klingemann et al., 2006).

Each unit of analysis was assigned a single code, hierarchically embedded within broader thematic domains. The codebook categories were developed through a combination of theoretical grounding and empirical iteration based on the characteristics of the dataset. Applied to the war in Ukraine, the final version consists of eight distinct narrative categories, which can be further grouped into six overarching thematic domains.

Justification	Narrative	Description	Claim
Cultural-Religious	Identity and Cultural Threat	The conflict serves as a defence against perceived cultural or ideological threats to traditional values or Russian identity.	The West's interference in Ukraine spreads liberal propaganda and threatens Russian heritage.
Economic and social	Economic Fallout/Domestic Welfare Neglected	Criticises the conflict's economic impact, the diversion of resources to Ukraine, including sanctions and the exploitation of Ukraine's resources, at the expense of domestic welfare.	We spend billions on Ukraine while our citizens suffer.
Legal/Political	Violation of Russian Sovereignty/Western geopolitical meddling	Russia seeks peace and is legitimate/has a moral obligation to defend itself and stabilise its neighbourhood against the threat of the West, whose involvement can lead to World War III.	NATO expansion violates Ukraine and Russia's sovereignty—it's illegal interventionism.
	Illegitimate and corrupt Ukrainian leadership	Claims Ukraine's government lacks legitimacy and is a puppet of Western powers. Zelensky and his clique are installing a corrupt and incompetent dictatorship in a failed state.	The Ukrainian government is just a puppet regime of the West, not democratically elected.
Security	Ukrainians and Ukraine are a military threat and aggressive war-mongers that threaten EU stability and security.	Warns of the conflict escalating into a global war (e.g Ukraine plans on invading neighbouring countries) and highlights risks from Ukrainian refugees or terrorism.	The war in Ukraine costs lives for other Western countries.
Humanitarian	Western Propaganda and Civilian	Frames allegations of war crimes and civilian suffering as outcomes of Western influence	Western arms prolong civilian suffering and spread fake narratives about Russia.

	Suffering	or propaganda.	
Historical-Ideological	Historical Betrayal of Russia	Frames the conflict in terms of broken promises and historical grievances.	Ukraine belonged to Russia.
	Ukraine/Nazi Allegation	Frames Ukraine or its government as being ideologically aligned with or directly supporting Nazi principles, often used to justify Russian actions as a continuation of an anti-fascist struggle.	The Ukrainian government is a neo-Nazi regime glorifying fascists, and Russia must stop them to protect the region.
None	None	None	

Table 1. PROMPT Codebook for topic 1 – War in Ukraine

LLM-Supported Supervised Classification Pipeline

At the core of the PROMPT project's classification system is a theory-driven, expert-defined codebook designed to capture recurring discursive patterns related to the Russia–Ukraine war. Grounded in research on propaganda, strategic narratives, and information warfare, the codebook defines eight primary narrative categories grouped into six broader thematic domains. **Each category includes clear criteria and real-world examples, serving as both an analytical guide and the foundation for model training.**

To scale this process across large, multilingual datasets, a supervised learning pipeline is implemented using fine-tuned transformer-based language models. Trained coders manually annotated a stratified sample of posts according to the codebook. These annotations were then used to fine-tune a multilingual model under few-shot learning conditions, enabling accurate classification even in low-resource settings.

Model performance is evaluated using standard metrics such as precision, recall, and F1-score. Annotation consistency is also assessed through inter-annotator agreement—both among human coders and between human and machine outputs—ensuring reliability and linguistic generalizability.

To maintain interpretability and prevent model drift, a human-in-the-loop validation process is used. Experts regularly review and correct model outputs, especially for ideologically sensitive or ambiguous content, preserving transparency and trust in automated decisions.

All classification results are stored in a searchable, metadata-rich database that includes actor, platform, document type, date, source, and narrative code. This infrastructure supports comparative and longitudinal analysis of narrative diffusion and enables modeling of spatial and temporal dynamics, including narrative spread, reactivation, and persistence.

The current model is based on **XLM-RoBERTa**, a state-of-the-art multilingual transformer optimized for low-resource classification (Liu et al., 2019). It is integrated into the **Babel Machine** platform (https://babel.poltextlab.com), which processes user-uploaded corpora and outputs structured, codebook-aligned classifications enriched with metadata (Sebők et al., 2024)..

Recent evaluations show the model achieves a macro-averaged F1 score of **0.75** and an overall accuracy of **75.6%** across the eight narrative categories. Performance is strongest in high-frequency categories like security and historical-ideological frames, with lower precision in underrepresented classes—confirming the model's robustness for downstream analysis.

Automated Filtering Using Multilingual Embeddings

To manage the scale and diversity of social media content, the PROMPT project incorporates in parallel a filtering layer based on sentence embeddings. A curated set of around 200 fine-grained policy claims—each representing specific instances of broader narrative frames—serves as the reference dataset. Using multilingual embedding models, the system calculates cosine similarity between each post and these reference claims. Only posts exceeding a defined similarity threshold are retained for further analysis.

After benchmarking several models, including Jina v3, bge-m3, Cohere Multilingual v3, Snowflake Arctic, and Voyage 3, **Jina v3** was selected for its strong performance in cross-lingual similarity tasks.

When no close match is found, two fallback strategies are used: prompting a large language model (LLM) to generate new labels in a few-shot setup, or applying **BERTopic**, a semi-supervised clustering method seeded with existing claims (Bommasani et al., 2021; Grootendorst, 2022). These approaches help surface emerging narratives and maintain adaptability.

This embedding-based layer complements the supervised classification pipeline by detecting semantically similar or newly emerging expressions that may fall outside the predefined codebook. While the supervised model ensures conceptual clarity, the embedding model adds flexibility and discovery potential.

Both methods are anchored in the same unified codebook. The 200 reference claims used in the embedding model are aligned with the broader narrative categories, ensuring consistency across approaches. This alignment supports integrated analysis, narrative tracking, and iterative refinement of both the codebook and the models.

By combining top-down theoretical structure with bottom-up empirical sensitivity, the framework enhances coverage, adaptability, and explanatory power—crucial for tracking disinformation in dynamic, multilingual environments.

Network Analysis of Narrative Propagation

After posts are classified using the narrative codebook, **dynamic network analysis is applied to model how specific frames and claims spread across social media.** Using metadata such as retweets, replies, and quote tweets, co-dissemination graphs and actor-interaction networks are constructed to reveal the structure and pathways of narrative diffusion.

This approach helps identify influential actors, clusters of coordinated amplification, and key bridging users who connect otherwise separate communities. These insights directly support PROMPT's detection strategies and enhance real-time monitoring of high-risk narratives.

The following section presents a case study on the war in Ukraine, a high-stake geopolitical context that provides a rich environment for examining the spread and transformation of disinformation. By tracing how specific claims circulate within and across online communities, the case study demonstrates how the integrated methodology—combining LLM-based classification, semantic filtering, and network analysis—can be operationalized to uncover patterns of narrative diffusion and coordination.

CASE STUDY I: WAR IN UKRAINE
Mapping the network dynamics of policy frames in the Ukraine conflict on X

Russia's full-scale invasion of Ukraine in February 2022 not only reshaped Eastern Europe's geopolitical landscape but also triggered one of the most intense information battles of the digital age. Platforms like X (former Twitter) quickly became arenas where state actors, partisan groups, and everyday users competed to shape public perception of the conflict. In these digital spaces, policy frames—the rhetorical lenses through which events are interpreted—are used to justify actions, discredit opponents, and rally support both domestically and internationally.

While previous research has highlighted the spread of disinformation and competing narratives during the war, it often treats framing as static or confined to individual posts or outlets. In reality, framing is dynamic and networked: frames emerge, evolve, and fade across user clusters and interaction types—retweets, quotes, replies. Temporal network analysis helps uncover how these frames spread, which ones gain traction, and how they move—from central "hub" accounts to peripheral users, or the other way around—and how long they persist.

This case study addresses two key gaps in the literature. First, while many studies examine frame prevalence or static network snapshots, few explore the **temporal dynamics** of frame diffusion—how frames emerge, evolve, fragment, or consolidate over time, and the roles different user types play in these processes. Second, existing research rarely combines **multilingual LLM-based classification** with **dynamic network modeling** to analyze how narratives spread across languages and platforms. By integrating a theory-informed codebook with an LLM-enhanced classification pipeline and both contact- and interval-based temporal network analysis, this study offers a comprehensive view of how policy frames related to the Ukraine war are constructed, disseminated, and sustained on Twitter.

Using a dataset of over 70 million tweets (Feb 2022–June 2023), PROMPT:
1. Identifies dominant narrative frames across multiple languages.
2. Maps how these frames spread via retweets, quotes, and replies.
3. Analyzes the roles of key actors—hubs, bridges, and peripheral users—in narrative diffusion.
4. Tracks how narrative clusters shift in response to major events and media coverage.

The analysis reveals how digital narratives are amplified and polarized, offering practical insights for policymakers, platform designers, and civil society actors working to counter disinformation. The ongoing Russian aggression against Ukraine has generated a flood of competing narratives across political, media, and digital spheres.

The literature review synthesizes key research on major thematic dimensions: Cultural-Religious Identity, Security, International Relations, Economic and Social Impact,

Legal-Political Issues, and Political Leadership—while incorporating recent computational and discourse-analytical methods for detecting and classifying wartime narratives on social and mainstream media:

- A dominant strand in the literature concerns depicting **Western and liberal democratic values as cultural threats**. Russian state-aligned narratives often portray the conflict as a defensive effort to protect Russian identity and heritage. Identity-based threats are strategically employed to mobilise cultural resistance.
- **Security**-oriented narratives frequently frame Ukraine as an aggressor or military threat, a theme central to Russian justifications for intervention.
- **Geopolitical narratives** often present Russia's actions as legitimate responses to Western expansionism. Russian strategic interests are often framed as rational and defensive, grounded in a *realpolitik* worldview. Narratives are frequently tailored across national contexts to justify Russia's position in **international relations**.
- Narratives concerning **economic decline**, sanctions, and domestic hardship play a central role in public communication about the war. Economic decline is often linked to broader narratives of betrayal and victimhood.
- **Legal and political narratives** often portray NATO and the EU as violating Russia's sovereignty.
- Narratives challenging the **legitimacy of the Ukrainian government** are widespread in both state media and social platforms. Topical research has discussed the longstanding use of corruption as a delegitimising tool. Ukraine's leadership has also been portrayed as Western-controlled and morally bankrupt.

Recent research highlights the dominance of cultural-ideological and security threat narratives in social media discourse about the Russia–Ukraine war (Hunter 2023; Lawrence, 2022). Cultural-ideological narratives frame the conflict as a defense against Western liberalism, portraying Russia as the guardian of traditional values. These narratives often depict the West as morally and culturally corrupt. Security narratives portray Ukraine—and NATO—as military threats, justifying Russian aggression as defensive (Korostelina, 2025; Ventsel et al., 2021). These narratives claim that Western arms prolong the war and civilian suffering. Public monitoring efforts, such as the EUvsDisinfo database maintained by the European External Action Service, have tracked these narratives since 2015. The database is a key resource for validating and supplementing academic and policy research on pro-Kremlin disinformation.

These findings indicate that narratives centered on cultural decline and existential security threats dominate online discourse and are especially effective on algorithm-driven platforms. Their emotional and ideological intensity makes them more likely to be shared, engaged with, and believed—reflecting broader trends in strategic communication and disinformation. The PROMPT case study of the Ukraine conflict addresses these central research questions:

RQ1: How do the structural dynamics of retweet and quote networks evolve during high-attention periods, and what patterns of centralisation, fragmentation, or bridging emerge over time?

RQ2: How do varying forms of user participation—such as initiating, amplifying, or bridging—contribute to the observed patterns of narrative diffusion and network structure during peak periods of activity?

Data

This case study uses a multilingual Twitter dataset on the Russia–Ukraine conflict, collected via the Tweepy API and shared on Kaggle. It includes both high-frequency and high-impact users. The key features of the dataset include, first, an asymmetric structure. The follower distribution is heavily skewed—median: 168 followers; mean: 4,500—due to a few influential accounts (e.g., media, influencers). This power-law pattern is crucial for network analysis, as "hub" users can significantly shape narrative spread. Second, the dataset has linguistic diversity: English dominates (~11M tweets, ~66%), followed by Spanish, French, German, Italian, Ukrainian, and Russian. About 1.7M tweets are labeled "undetermined," reflecting the global and regional nature of the discourse.

In terms of user activity, the patterns are similarly uneven. Most users posted only a few tweets, while a small minority were extremely active, posting hundreds or thousands. This skew is shown in logarithmic plots and reflected in engagement metrics. The follower count had weak correlation with likes and retweets, suggesting visibility doesn't guarantee engagement. To identify influential users, a composite engagement score (likes + retweets ÷ tweet count) was used—highlighting users who gained high impact with relatively few posts.

Hashtag analysis confirms the dataset's topical focus, with dominant tags like **#ukraine, #russia, #putin, #bakhmut**, and polarized expressions such as **#standwithukraine** and **#russiaisaterroriststate**. These patterns make the dataset ideal for studying narrative spread, influence, and information warfare in a multilingual conflict context.

The data spans a period from February 2022 to June 2023, covering the first 16 months of the full-scale Russian invasion of Ukraine. It includes over 1.2 million tweets, enabling large-scale analysis of public discourse, sentiment dynamics, and narrative dissemination during the conflict. The dataset includes a rich set of metadata, which allows for network, temporal, and content-based analyses. The primary variables are:

- **User metadata**: userid, username, acctdesc, location, following, followers, totaltweets, usercreatedts
- **Tweet metadata**: tweetid, tweetcreatedts, text, retweetcount, favorite_count, hashtags, language, coordinates
- **Interaction structure**: is_retweet, original_tweet_id, original_tweet_userid, original_tweet_username, is_quote_status, quoted_status_id, quoted_status_userid,

quoted_status_username,   in_reply_to_status_id,   in_reply_to_user_id,
in_reply_to_screen_name

- **Technical/logging variables**: extractedts (timestamp of data collection)

Analytical Sample: First Anniversary of the Ukraine War

From over 70 million tweets, we selected a high-activity 7-day window marking the first anniversary of the Russia–Ukraine war—an ideal period for analyzing narrative spread and engagement. This subset includes **697,473 English tweets** from **219,753 users**, most of whom posted only once (median = 1), while a few hyperactive accounts posted over 1,000 times. Activity follows a **power-law distribution**, typical of social media. **Follower stats** show similar skew: median = 321, mean = 13,496, with the top user exceeding 39 million followers. This heavy-tailed structure calls for **log-transformed or rank-based metrics** to avoid distortion by outliers, when modelling influence and engagement. **Data quality challenges** persist—especially in user location, which is self-reported and highly inconsistent: more than 53,000 unique string values appear, ranging from precise addresses to fictional or vague entries. Engagement patterns reveal weak correlations between follower count and likes/retweets. Notably, **hashtag overuse** negatively impacts retweets (ρ = –0.213)**,** while **Tweet length** shows a weak positive link to engagement, suggesting informativeness matters slightly more than brevity.

Overall, this anniversary-week subset offers a high-resolution snapshot of content production, user behaviour, and engagement dynamics during a politically and emotionally salient period. Its structure—marked by extreme variance in user activity and influence—necessitates robust, non-parametric modelling strategies, particularly in research focused on the detection and diffusion of disinformation or narrative framing.

Network analysis overview

The Twitter dataset tracks user activity related to the Ukraine–Russia conflict across multiple time periods. Each tweet—identified by a unique tweetid—is categorized as an original tweet, retweet, or quote (handled separately due to distinct interaction patterns). While group affiliations aren't labeled, follower and following counts act as proxies for user influence. These metrics support filtering and weighting in the network analysis.

Network Building Principles

Following Saqr's (2024) framework for temporal network analysis, we build directed dynamic networks where: nodes represent users, and. edges represent time-stamped interactions (retweets, quotes, replies), indicating information flow. Each edge assumes minimal cognitive engagement and includes user metadata (follower/following counts). If the source user isn't in the dataset, their metadata is marked as missing (set to 0).

Temporal Network Typology: Retweet/Quote Network

To analyze how information spreads over time, we use Saqr's (2024) interval temporal network model, where each edge has a start and end time, representing the lifespan of shared content. This allows us to measure temporal metrics like density, centralization, connectedness, and transitivity over a 7-day period.

We define nodes as users, edges as retweets or quotes, indicating information flow from the original poster to the interacting user, timestamps mark when the interaction occurred; if both retweet and quote are present, two edges are created, and follower/following counts are included; missing data is set to 0. Edges are grouped into **tweet chains** using a depth-first search (DFS) algorithm. Each chain ends with the timestamp of its last tweet.

From a two-day subset, we generated **473,221 edges, 232,469 nodes** and **61,149 tweet chains** (mean size: 7.4; 75% have ≤4 users; largest: 7,542 tweets). We focus on **high-impact clusters**—chains with ≥10 unique users—to study meaningful content diffusion. These clusters highlight sustained or cascading resharing, and we analyze their structure using metrics like **density** and **connectedness**.

Temporal Evolution of Retweet Network Structure During the Ukraine War Anniversary

Figure 1 presents a series of interval temporal retweet networks for tweet groups comprising more than ten unique retweeters, recorded daily between February 23 and 28, 2023. This time frame corresponds to the week surrounding the first anniversary of the full-scale Russian invasion of Ukraine, a period marked by intensified public attention and narrative contestation in online spaces. The temporal snapshots reveal distinct phases in the structure and dynamics of content dissemination, suggesting an evolving pattern of narrative propagation on Twitter.
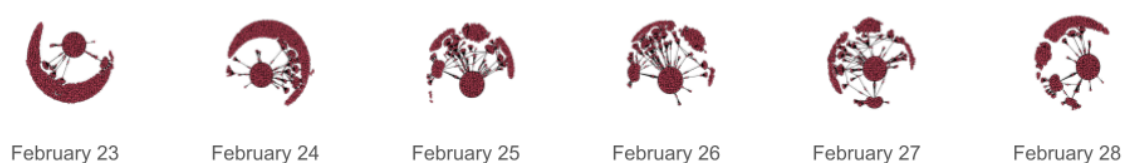

February 23   February 24   February 25   February 26   February 27   February 28

Figure 1: Interval temporal user networks of tweet-groups with over 10 unique users retweeting

Narrative Diffusion Phases:
   Phase 1: Centralized Launch (Feb 23–24)
      The network is compact and highly centralized, with most retweets clustered around a single dominant account—typically institutional or high-profile. This reflects a top-down dissemination pattern, with minimal organic spread or secondary hubs.
   Phase 2: Viral Spread & Fragmentation (Feb 25–27)

Tweet volume peaks, and the network becomes more complex, featuring multiple hubs and loosely connected clusters. Independent actors begin circulating similar narratives within their communities. Bridge users emerge, linking clusters and enabling cross-community diffusion. Influence becomes decentralized, with no single dominant voice.

Phase 3: Fragmented Decline (Feb 28)

Activity drops, and the network fragments further. Small, disconnected clusters persist, but no new central figures emerge. Bridge users become inactive, limiting cross-cluster flow. The narrative loses momentum, shifting from broad public discourse to isolated, localized engagement.

Content-Centric Network Analysis: Co-Retweet Structures of Narrative Clusters

Alongside user-based networks, we applied a **co-retweet model** that treats **tweets as nodes**, linking them when **retweeted by the same user**. This approach reveals **discursive proximity**—how tweets cluster based on shared engagement, not just content similarity. Tweets retweeted by the same users often reflect related themes or frames, forming **narrative clusters**. This model helps identify how users group and amplify specific messages, offering insight into the structure of collective discourse.

Structural Features of the Co-Retweet Network

We constructed the co-retweet network using the same 7-day analytical window (February 23 – March 1, 2023), limited to English-language tweets with at least 10 unique retweeters. The resulting graph contained [X] nodes and [Y] edges (fill in with your actual values), and displayed a modular structure with a large central cluster and numerous smaller, disconnected components.
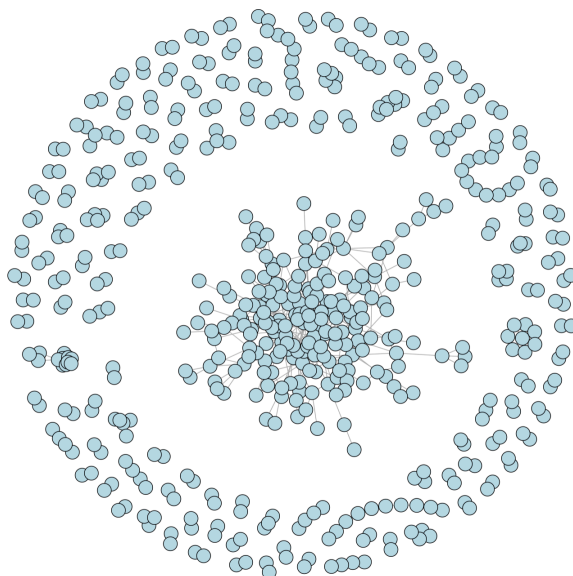

Figure 2: Network of High-Engagement Tweets During the Ukraine War Anniversary Week

The central cluster comprised tweets that were widely retweeted by a diverse user base, including high-profile accounts, institutional actors, and verified journalists. These tweets

often expressed general pro-Ukraine solidarity or commemorated the war's anniversary—indicative of high narrative convergence. Peripheral clusters tended to include more niche or polarising tweets—often either critical of Western policy or promoting alternative narratives—that were retweeted primarily within specific ideological communities. Bridging tweets, i.e. those that connected otherwise separate clusters, often took a critical but balanced tone, sometimes linking humanitarian concerns with geopolitical claims. These tweets functioned as points of discursive overlap between otherwise disconnected audiences.

Interpretive Value

The co-retweet network reveals structural alignment of message content based on collective user behavior, offering a lens to detect de facto narrative coalitions. Tweets grouped together not by textual similarity but by shared audience engagement provide evidence of interpretive compatibility and strategic bundling. Whereas the user-based retweet network captures who engages with whom, the co-retweet network uncovers what kinds of content are cognitively or ideologically co-selected. The combination of both network views allows us to analyse not only how frames travel across users, but also how tweets coalesce into frame constellations through shared attention.

The distribution of the most frequent claims

While the structure of retweet and quote networks reveals how narratives spread, it's equally important to examine which specific claims gain traction. Using supervised classification and embedding-based similarity methods, we identify and categorize distinct claims in tweets. This content-level analysis connects interaction patterns with the meaning of the messages. The next section maps the most frequent claims and highlights those driving the largest dissemination clusters. By combining network dynamics with narrative types, we gain a deeper understanding of why certain frames dominate public discourse during key moments of the conflict.

The most widely shared claims center on cultural-ideological and security threat narratives. Nearly half of the top 30 claims fall into **Historical-Ideological** and **Security** categories, accounting for about 75% of all claim mentions. This suggests these narratives are both varied and disproportionately amplified. Visual data supports this trend, with high-frequency claims like "Ukraine belongs to Russia" and "Ukraine is preparing to invade Russia", "Russian people are sacrificing themselves again to defeat Nazism", and reflecting historical or security framings. Even less frequent claims often echo themes of escalation, terrorism, or global instability. Together, the data shows that cultural and existential threat narratives dominate the discourse, making them key channels for war-related disinformation.

During the week around the war's first anniversary, we identified four major dissemination clusters in retweet and quote networks, each centered on a specific disinformation claim. Table 2 outlines these clusters, showing the dominant claim and tweet volume as a measure

of engagement. The largest, with over 3,000 tweets, promoted the idea that "Russia is trying to save Ukraine," while others framed Ukraine as illegitimate, corrupt, or anti-peace.

All four clusters reflect historical-ideological or security narratives, emphasizing existential threats or grievances like anti-Western sentiment. Although based on a short time frame, the strong link between narrative type and network scale suggests that certain claims are more likely to gain traction and spread widely online.

Group	Claim	Count
12	Russia is trying to save Ukraine.	3 275
156	Zelensky and his wife are enjoying a luxurious lifestyle while the Ukrainian people are dying.	1 093
594	Ukraine and EU countries are disrupting peace talks.	179
329	Ukrainians hate Zelensky.	168

Table 2: Representative claim clusters with their frequencies

This pattern shows that disinformation spreads not just because it's false, but because of how it's framed and the emotions it triggers. Narratives tied to identity, security, or historical injustice are especially effective at driving engagement and maintaining cohesion across networks. These dynamics highlight the role of narrative structure in shaping what spreads, how, and why—revealing the mechanics of virality in geopolitical conflicts.

Network-Based Claim Analysis Using Retweet and Quote Structures

This analysis uses a retweet/quote network to trace how claims spread around the first anniversary of the Ukraine war. Each tweet forms a directed edge from the original author to users who retweet or quote it, with timestamps enabling structural and temporal insights.
To reduce noise, users with only one incoming edge and no outgoing edges (nodes with in-degree one and out-degree zero) — those not actively spreading content — are removed. The final network, built using NetworkX, includes 58,884 users and 216,815 connections, capturing key pathways of information flow.

Measuring User Exposure to Claims

User exposure is defined probabilistically: if a user posts a tweet linked to a claim with at least 65% confidence, they are marked as exposed. A user-claim matrix is built by aggregating daily tweet data over seven days, assigning +1 for each claim per user per day. These daily matrices are merged into a cumulative table to track exposure trends over time. Some claims, like "Ukraine is mobilising," show steady growth, while others, such as "Ukraine belongs to Russia," plateau quickly. Notably, major dissemination begins around February 26—two days after the war's anniversary—indicating delayed viral spread.

Claim Correlation and Cluster Dynamics

Using total exposure scores over seven days, Pearson correlations between claims reveal thematic clusters — such as strong links between EU-critical narratives. For example, users promoting claims about EU leaders supporting war crimes also tend to frame Ukrainian refugees as a cultural threat, suggesting ideological alignment. Correlating claim exposure with users' follower and following counts shows two things: (1) follower and following counts are highly correlated, as expected; (2) neither metric correlates meaningfully ($r < 0.1$) with claim exposure. This indicates that high spreaders of the claims identified in this analysis cannot be described as having a certain level of follower or following counts.

Visualisation and Network Dynamics

Claim diffusion is also visualised over the static network (i.e., a time-agnostic graph). Using cumulative daily exposure scores, the spread of specific claims, can be visualised step by step, highlighting how claims move through the user network.
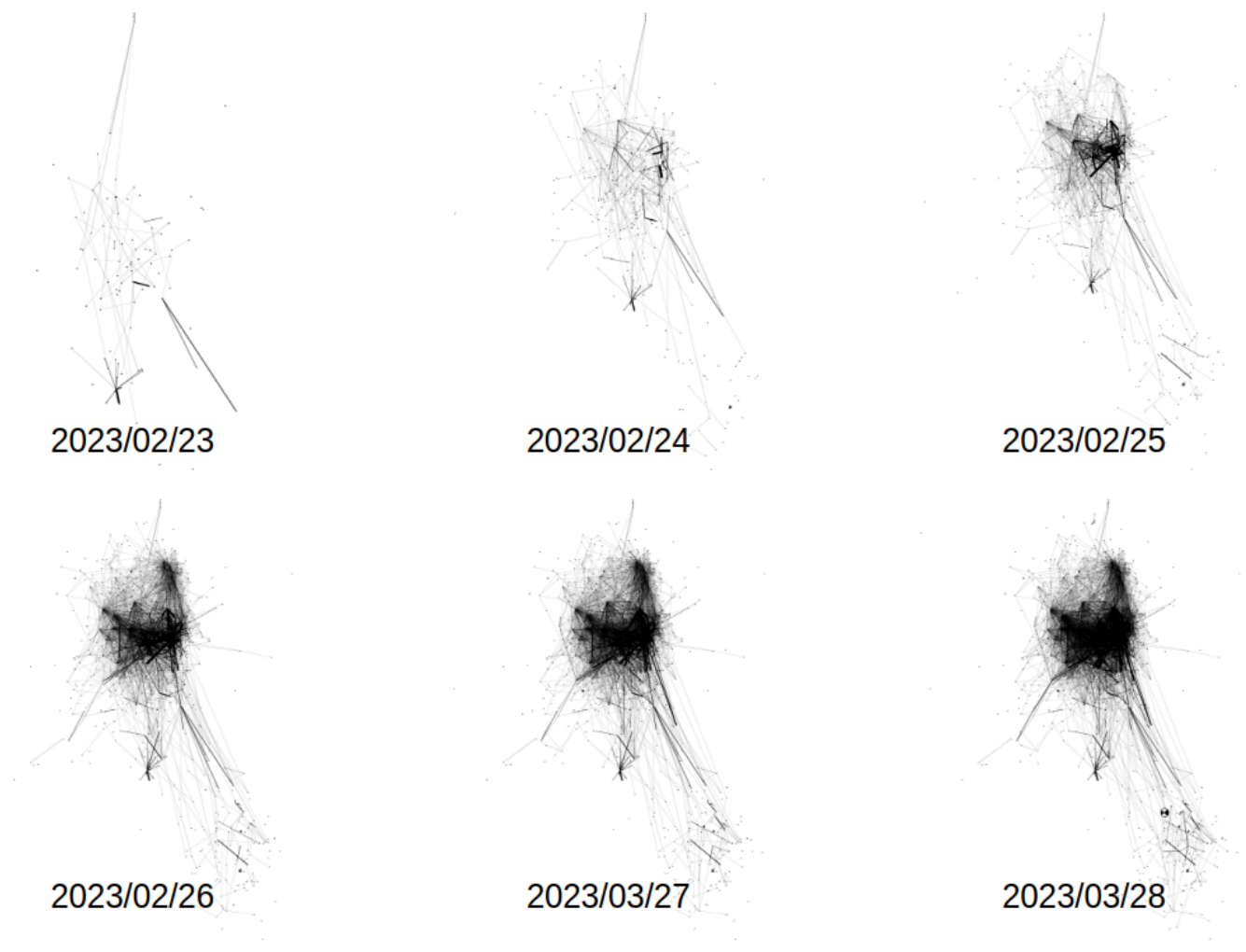

Figure 3: User exposure spread of a specific claim

Identifying Key Actors: Claim Owners and High-Exposure Users

To understand how claims spread, we identify two key user types based on exposure counts:

1. Claim Owners - users who post original tweets (not retweets or quotes) linked to a claim with at least 65% confidence. As originators of narratives, their role is crucial — they introduce new content and may hold central positions in the network.
2. High-Exposure Users – users who frequently share claim-related content—whether through original tweets, retweets, or quotes. While not always the source, they act as amplifiers, helping claims reach broader audiences over time. This method focuses on consistent amplifiers but may miss occasional or one-time participants, a limitation tied to the short observation window. Capturing such "swing" users would require longer-term or more adaptive tracking methods.

Comparing Network Metrics

Once these two groups are identified, various network metrics (such as degree centrality, betweenness centrality, clustering coefficient, etc.) can be used to assess and compare their positions in the network. The aim is to explore, (1) whether they hold central positions in the network, (2) whether they act as bridges between clusters, (3) whether they belong to tight-knit groups or span across communities, and (4) how their connectivity and influence differ. This comparison allows researchers to uncover organisational patterns behind claim diffusion, identify key nodes for potential intervention, and better understand narrative propagation mechanics, particularly in disinformation or coordinated messaging campaigns.

Conclusion and limitations

The Ukraine war case study demonstrates the framework's capacity to detect ideologically polarising narratives that shape discourse during high-intensity geopolitical moments.

The analysis reveals how narrative spread evolved during a peak period of the Russia–Ukraine war. Initially, retweet and quote networks were centralized around high-reach institutional accounts, then shifted to decentralized viral spread, and eventually fragmented. These phases reflect changing patterns of user engagement that shape narrative visibility. Claim initiators often sit at the network's edges but spark broader cascades. In contrast, high-exposure users are more central and sustain narrative momentum. A smaller group of bridging users connects otherwise separate communities, helping narratives cross ideological boundaries. These distinct roles highlight the need to examine not just what spreads, but who drives the spread and how.

Several limitations should be noted. First, the classification system is optimized for English, limiting accuracy in other languages unless further adapted. While the embedding model is multilingual, the fine-tuned LLM is primarily trained on English data. Second, the one-week time frame restricts insights into long-term trends or sporadic "swing" users. Third, applying

this method across platforms is difficult due to inconsistent APIs, non-standard metrics, and challenges in matching users across ecosystems.

Despite these challenges, the methodological infrastructure developed here provides a strong foundation for future expansions. Crucially, PROMPT project aims to extend narrative detection beyond English, fine-tuning the classification pipeline in additional languages. This will involve integrating new annotated corpora, multilingual LLM fine-tuning, and region-specific codebook extensions. Expanding the system's cross-linguistic capacity is essential for building a truly European infrastructure for narrative monitoring and disinformation research.

This case study was designed as **a real-time investigation into the Romanian electoral context during 2024–2025,** with the objective of rapidly identifying coordination dynamics and dominant narratives across public discourse. The approach aimed to gather the broadest https://docs.google.com/document/u/4/?authuser=4&usp=docs_webpossible set of social media data related to the Romanian elections, in order to assess the presence of semantically aligned behaviours within the evolving media ecosystem.

To achieve this, two datasets were compiled and merged. The first was generated using a set of targeted keywords, including names of political figures (Antonescu, Simion, Ponta, Lasconi, Georgescu, Terheș, Sandru), electoral terms (elections, vote, party, candidate, campaign), ideological references (deep state, traditional values, family, social cleansing), and geopolitical issues (Ukraine, Russia, war, refugees). This dataset spans from 1 January 2024 to 6 May 2025 and consists of approximately 88,000 public posts.

The second dataset consisted of approximately 96,000 posts published between August 2024 and 23 April 2025 by problematic Romanian Facebook accounts, identified and provided by a local NGO. These posts were retrieved through Meta's Content Library.

The two corpora were deduplicated and combined to form a unified dataset. Coordination patterns were then analysed separately using CooRTweet, as detailed in the following section.

The merged dataset of 164,762 public Facebook posts related to the Romanian elections (Jan 1, 2024 – May 6, 2025) was analyzed using CooRTweet to detect coordinated behavior.

Using CooRTweet, we identified 97 groups of accounts exhibiting signs of coordinated posting behavior. The detection was configured with the following parameters:
- **min_participation**: 2
- **time_window**: 10 seconds
- **edge_weight**: 0.9 (quantile filter)
- **subgraph**: 1 (focus on strong coordination)

We identified **97 groups** of accounts showing signs of coordination.

To better understand the nature of these groups, we selected the largest coordinated communities — prioritizing those with the highest number of unique objects (i.e., shared posts) and unique accounts — and extracted the textual content they disseminated.

Narrative extraction was conducted using ChatGPT, powered by the GPT-4o model. For each coordination community, a small, randomly selected sample of post texts was compiled to avoid exceeding context window limitations. ChatGPT was then prompted to infer and summarize the dominant themes and discursive patterns in each community.

community	avg_time_delta	avg_edge_symmetry_score	unique_objects	unique_vertices
2	5.7438782285345	0.5	212	2
42	0.949649721465653	0.5	154	3
8	4.11465525424801	0.5	117	12
1	5.47910732499677	0.5	90	11
4	0.502923976608187	0.5	77	2

Table 3. Communities and key indicators

Next, we dive into the key narratives and tendencies emerging from each major coordinated community.

Community 2 – Sarcastic Nationalism and Anti-Western Ridicule

This corpus is marked by a sarcastic and culturally conservative discourse targeting liberal democratic institutions, Western alliances (EU/NATO/USA), and progressive values. Humor, irony, and ridicule are central rhetorical devices, deployed to delegitimize centrist, pro-Western actors such as Ciolacu, Geoană, Iohannis, and Lasconi. In contrast, figures aligned with anti-globalism or sympathetic to Russia and BRICS receive neutral or favorable portrayals. The narrative structure reflects a deep right-wing populist orientation, with recurrent attacks on gender politics, mainstream media, and liberal norms, promoting a nationalist, sovereigntist worldview.

Entities & Themes

Pro-Western Institutions & NATO/EU/USA

The commentary targets pro-Western institutions such as NATO, the EU, and the USA, with a tone that is consistently critical and sarcastic. These entities are frequently portrayed as hypocritical, ineffective, or morally compromised. For example, one post mocks U.S. military power with the line, "US News informs us which is the most powerful military on the planet. Babies, dance!" — a sarcastic jab suggesting the unseriousness of American strength. Another dismisses Western celebrity activism with the remark, "Actors are not always smart." A third example critiques local media's deference to the U.S., referencing a post by the American ambassador and sarcastically quoting a Romanian journalist: "Thank you, beautiful allies!"

The overarching narrative is shaped by Euroskeptic, anti-globalist, and anti-NATO sentiments, often infused with right-wing populist rhetoric.

Ukraine & Zelensky

The tone toward Ukraine is predominantly hostile and mocking. Posts often portray the country as incompetent, deceitful, or merely a puppet of Western powers. One sarcastic remark reads, "Zelensky's victory plan succeeded — we're just waiting for confirmation," expressing disbelief in Ukraine's military achievements. Another post mocks early optimism about Ukraine's war efforts: "Dictator Zelensky embarrassed himself by not taking Moscow in 3 days." Additionally, claims such as "Zelensky banned the Orthodox Church in Ukraine" echo common Russian disinformation narratives.

The broader narrative is pro-Russian and anti-Ukrainian, often sympathetic to Vladimir Putin. It draws on familiar disinformation tropes, including the denial of Ukrainian sovereignty, ridicule of Western alliances like the G7, and efforts to undermine international support for Ukraine.

Romanian Politicians (Ciolacu, Iohannis, Ciucă, Lasconi, Geoană)

The tone toward domestic political figures ranges from satirical to openly hostile, with frequent accusations of incompetence, corruption, or dishonesty. One post mocks Nicolae Ciucă's perceived lack of ambition or capability by quoting him saying, "I don't see myself as president." Another targets USR politicians, claiming, "Lasconi lies about her education like Bot lies about his," aiming to discredit them personally. Meanwhile, Mircea Geoană is ridiculed for allegedly copying NATO's branding, with the sarcastic remark, "Geoană plagiarizes NATO's colors to the comma."

The overarching narrative is strongly anti-establishment, particularly critical of centrist and pro-European parties such as PSD, PNL, and USR. The tone reflects populist and anti-globalist attitudes.

Western Media & Culture

The tone is heavily sarcastic and often hostile, particularly toward liberal or progressive ideals. Posts frequently mock concepts related to gender identity and Western cultural norms. One example reads, "Your husband today can be someone's wife tomorrow," a dismissive take on gender fluidity. Another post sarcastically asks, "Am I the only one noticing that Geoană plagiarizes NATO colors exactly?"—again ridiculing pro-Western symbolism. The broader narrative is staunchly anti-progressive and anti-"woke," rejecting identity politics and liberal Western values. It reflects a strong alignment with right-wing cultural conservatism.

Russia & Allies (Putin, BRICS, etc.)

The tone in this category is mixed but generally leans positive or neutral, often using ironic glorification. Some posts express subtle admiration for figures like Vladimir Putin, as in the remark, "Hmm, that smile gives the West a headache," which suggests a kind of backhanded praise. Others highlight the economic performance of non-Western alliances, such as BRICS, with posts sharing data like "GDP growth data of BRICS vs. G7 in 2024," implicitly promoting the idea of a shifting global power balance. The underlying narrative is pro-Russian and aligned with BRICS, subtly endorsing a multipolar world order over Western dominance.

Recurring Themes & Rhetorical Patterns

A common motif across posts is the repeated use of phrases like "Vezi primul comentariu" ("See the first comment") or "Neapărat vezi primul comentariu" ("Definitely see the first comment"). These serve as cues directing users to linked or hidden content, which is often satirical, conspiratorial, or propagandistic in nature.

Another notable pattern is the use of satirical "soldier journals" under titles like *"Un ostaș în slujba țării "*("A soldier in service of the country"). These ironic pseudo-diaries humorously depict military incompetence, poking fun at both Romanian army culture and NATO-style training. They typically end in absurd scenarios, undermining the seriousness of military institutions and channeling anti-militarist populism.

There is also a recurring focus on sex and gender, particularly through satire aimed at transgender athletes or gender nonconformity. These posts often reflect a reactionary or conservative stance. For instance, one post reads, "Halteforila noastră a câștigat argintul la bărbați" ("Our weightlifter won silver in the men's category"), while another notes, "Imane Khelif și-a șters profilul [...]" ("Imane Khelif deleted her profile [...]"), both implying criticism or mockery of gender diversity in sports.

Political and Ideological Framing

The discourse reflects a broadly right-wing populist orientation, marked by strong nationalist, anti-globalist, and pro-sovereignty themes. There is consistent opposition to NATO and the European Union, often expressed through ridicule of progressive politics and Western institutions. Irony, satire, and meme-like rhetoric are frequently used to challenge mainstream narratives, particularly those associated with the liberal international order. At times, the content aligns with or echoes Putin-friendly narratives. This is evident in the derision of Ukraine's efforts, the amplification of BRICS economic achievements, and attempts to delegitimize NATO's actions and authority.

Summary

The overarching discourse is characterized by a sarcastic, populist critique of liberal democratic institutions, Western alliances (NATO/EU/USA), and progressive social norms. The language and framing strategies draw heavily on humor, irony, and ridicule, often with an intent to delegitimize perceived elite narratives.

While Romanian politicians are attacked across the spectrum, centrist and pro-Western figures (Ciolacu, Geoană, Iohannis, Lasconi) are especially targeted. In contrast, figures or narratives sympathetic to Russia, anti-globalism, or alternative geopolitical poles (e.g., BRICS) receive relatively favorable or neutral treatment. Gender politics, Western liberal values, and mainstream media are consistent targets of derision, indicating a deeply right-wing, nationalist, and culturally conservative ideological current.

Community 42 – Electoral Fragmentation and Media-Driven Crisis Discourse

This corpus, largely sourced from Antena 3 CNN, captures a fragmented and ideologically tense electoral climate. Key narratives oscillate between anti-elite populism (Simion, Georgescu) and reformist centrism (Lasconi, Dan), against a backdrop of internal disintegration within PSD, PNL, and USR. The tone is shaped by scandal, delegitimization, and heightened anxiety over foreign interference, especially from Russia. Media framing is ambivalent: while AUR and Simion are scandalized, they are also positioned to appeal to anti-system sentiment; Lasconi and USR are alternately endorsed and undermined, often via gendered conflict frames; and mainstream parties are depicted as ineffective or imploding. The overarching rhetoric reflects a rise in sovereigntist sentiment, media dramatization, and political distrust.

Entities and Narrative Directions

George Simion / AUR

George Simion and the AUR party are portrayed through nationalist, populist, and euroskeptic lens and a strong anti-establishment tone. Posts suggest that AUR engages in controversial tactics, such as allegedly attempting to bribe priests before elections by sending them religious notes with money. Simion is also shown accusing authorities of trying to rig the vote, reinforcing a narrative of institutional distrust. In another instance, he is depicted calling Gigi Becali live on air, urging him to mobilize voters—an appeal to populist, grassroots support.
Overall, Simion and AUR are framed as anti-system challengers to the political elite. Their rhetoric emphasizes election integrity, religious symbolism, and direct appeals to "the people." The suggestion that AUR restricted press access further underscores a tense relationship with media transparency and democratic norms.

Nicușor Dan

Nicușor Dan is presented as a technocratic, center-right figure with a generally pro-European and anti-corruption stance. He is positioned as a key contender in the presidential race, notably facing off against George Simion. However, his image is not without controversy. Posts highlight that the Electoral Authority has requested an investigation into his campaign financing, and that a poll he commissioned ranked another candidate, Georgescu, in first place—raising questions about credibility and campaign strategy. While Dan is largely framed as a rational, reform-oriented candidate, these narratives suggest attempts to complicate or undermine his image, portraying him as part of a fragmented center-right landscape that is not immune to scrutiny or political maneuvering.

Elena Lasconi / USR

Elena Lasconi is portrayed as a reformist, pro-European figure navigating internal party conflict and public scrutiny. In one instance, she responds to accusations from Georgescu, who claimed she wanted to send young people to war—an example of the gendered and confrontational rhetoric she faces. Another post notes that she remains "president without a party," highlighting a leadership vacuum and internal fragmentation within USR. Her own words—"It's ridiculous, a manipulation technique"—reflect a defensive stance against perceived political attacks. Overall, Lasconi is framed as a candidate under siege, emblematic of a reformist platform struggling with cohesion and credibility, and subject to subtle gender-based discreditation.

Călin Georgescu

Călin Georgescu is depicted as a deeply anti-establishment figure, often surrounded by controversy and suspicion. Posts warn of the dangers his leadership might pose, with comparisons to totalitarian regimes and concerns about foreign influence, particularly from Russia. His party's symbolic post-election fasting and his declaration of zero campaign expenses reinforce an image of radical simplicity and outsider status. Georgescu is framed as both a disruptor and a potential destabilizing force, relying on low-budget, anti-system messaging to appeal to disillusioned voters.

Victor Ponta

Victor Ponta is cast as an opportunist and a destabilizing presence within the PSD. His expulsion from the party by its National Political Council signals internal conflict, while statements such as "A vote for Ponta is a vote for Nicușor Dan" suggest that his political maneuvers are seen as indirectly aiding rivals. Ponta's ideological ambiguity and perceived disloyalty position him as a divisive figure undermining party unity from within.

Crin Antonescu / Coalition (PSD–PNL–UDMR)

Crin Antonescu and the governing coalition are portrayed as centrist but increasingly fragmented and under pressure. Antonescu's call to treat all parties equally reflects an attempt to maintain democratic decorum, yet he appears isolated amid growing tensions. Reports of PSD members demanding Ciolacu's resignation and PNL and UDMR shifting support to Nicușor Dan suggest a coalition on the verge of collapse. The narrative paints a picture of political instability and eroding leadership within the ruling alliance.

External Influences (Russia, Trump, Ukraine)

A broader geopolitical narrative runs through the discourse, emphasizing Romania's vulnerability to foreign interference. Posts claim that Russia funded a €69 million disinformation campaign, while others invoke globalist conspiracies to explain electoral disruptions—such as the assertion that elections were canceled to prevent peace in Ukraine. These narratives often link figures like Trump, NATO, and domestic candidates in ways that heighten strategic anxiety and reinforce nationalist sentiment. The overall effect is to cast doubt on the legitimacy of democratic processes and amplify fears of external manipulation.

Summary

The media content—primarily sourced from Antena 3 CNN—reflects a fragmented and ideologically charged electoral climate in Romania. Several dominant rhetorical patterns emerge across coverage.

First, there is a clear tension between anti-elite populism, represented by figures like George Simion and Călin Georgescu, and institutional reformism, embodied by candidates such as Elena Lasconi and Nicușor Dan. **This dichotomy frames the election as a clash between system challengers and technocratic modernizers.**

Second, traditional parties like PSD, PNL, and USR are portrayed as internally fractured, with frequent dramatizations of betrayal, leadership crises, and coalition instability. **These narratives contribute to a broader sense of political disintegration.**

Third, **anxiety over foreign interference—particularly from Russia—is a recurring theme.** This includes both cyber and ideological threats, often tied to anti-globalist rhetoric and fears of manipulated democratic processes.

Fourth, all major candidates are subject to delegitimization efforts. Scandals, accusations, and discrediting narratives are used to erode public trust, regardless of political affiliation.

In terms of framing preferences, AUR and Simion are often targeted with scandalous or inflammatory allegations, yet the tone may paradoxically resonate with anti-system voters.

USR and Lasconi receive mixed treatment—highlighted as reformist but undermined through narratives of internal betrayal and gendered conflict. Georgescu is depicted with a blend of ridicule and alarm, casting doubt on his legitimacy. Meanwhile, the mainstream coalition parties are consistently portrayed as ineffective or on the verge of collapse.

Overall, the ideological trends point to a rise in sovereigntist and anti-establishment rhetoric, a decline in coalition unity, and an increase in media dramatization and adversarial political tone.

Community 8 – Polarized Electoral Framing and Strategic Delegitimization

This merged corpus reveals a polarized and adversarial political discourse, where ideologically loaded binaries dominate: reform vs. corruption, Europe vs. Russia, civic modernity vs. extremist regression. Elena Lasconi and USR emerge as the main vehicle for pro-European, anti-corruption, civic reform discourse. PSD and PNL, though formally pro-European, are heavily attacked as corrupt, stagnant, and complicit in maintaining a captured state. AUR and SOS România are clearly framed as extremist threats, tied to Kremlin interests, and socially regressive.

The broader public narrative leans strongly anti-system, yet diverges sharply in proposed solutions: populist-nationalist vs. liberal-reformist. There is overarching distrust in institutions, particularly domestic ones, with selective trust placed in the EU and NATO. The texts exhibit a clear center-right, reformist, pro-Western orientation, heavily favoring Elena Lasconi and USR, while delegitimizing both the traditional elite (PSD, PNL) and radical alternatives (AUR, SOS) through a mixture of anti-corruption, anti-Kremlin, and civic moralization.

Entities & Themes

USR (Uniunea Salvați România) / Elena Lasconi

USR, under the leadership of Elena Lasconi, is positioned as a centrist to center-right reformist force with a strong pro-European and pro-NATO orientation. The party's messaging emphasizes technocratic governance, anti-corruption populism, and a firm rejection of Russian influence. Lasconi is consistently framed as the only credible and principled alternative to Romania's entrenched political class.

Her campaign rhetoric highlights civic engagement and national urgency, with slogans like "România pentru toți, nu doar pentru unii" ("Romania for all, not just for some"), "lupta generației noastre" ("our generation's fight"), and "votul poate salva democrația" ("your vote can

save democracy"). These messages frame the election as a defining moment for the country's democratic future.

A sharp East–West divide is central to the narrative: Russia is portrayed as a symbol of regression and authoritarianism, while the EU represents modernity, freedom, and prosperity. Lasconi's campaign also strongly rejects the "old system" of PSD and PNL, labeling them as corrupt and oligarchic. The Romanian diaspora is frequently invoked as a vital democratic force, and strategic anti-extremism messaging targets figures like George Simion, Călin Georgescu, and Diana Șoșoacă.

Examples of this framing include statements such as: "Elena Lasconi is the only right-wing candidate with a chance to enter the second round and defeat PSD," and "Vote for the only good-faith candidate who is not part of the corrupt PSD-PNL system." Another post warns, "On December 1 and 8, we choose whether to remain with our NATO and EU partners… or take the path of Russia and extremism."

In summary, USR is portrayed as a modernizing, anti-system—but not anti-democratic—movement. Lasconi is cast as the moral and political counterweight to both domestic corruption and foreign authoritarianism, with the electoral choice framed as a matter of national survival.

PSD (Partidul Social Democrat) / Marcel Ciolacu

PSD is broadly positioned as a center-left party with populist leanings and establishment characteristics. Its messaging traditionally emphasizes social welfare and economic stability, often accompanied by implicit nationalist undertones. However, within the analyzed media corpus, PSD is overwhelmingly framed in a negative light.

The party is frequently delegitimized through allegations of corruption, such as its association with the Nordis scandal. It is also accused of engaging in "blat politic" (political rigging), particularly in connection with AUR, suggesting covert alliances that undermine democratic competition. Additionally, PSD is portrayed as either enabling or tolerating Russian influence, further distancing it from pro-Western values.

Narratives often depict PSD as part of the entrenched political elite, maintaining privilege at the expense of the public. For instance, one post claims, "Ciucă and Ciolacu have known for two years about Iohannis's villa," implying complicity and concealment. Another criticizes the government for prioritizing elite benefits over public welfare: "Instead of cutting privileges for party cronies, the Ciolacu 2 Government stops the increase in child allowances." A third post bluntly states, "We know about the Ciolacu–Simion arrangement," reinforcing suspicions of backroom deals.

In summary, while PSD's official messaging may focus on economic justice, the dominant framing in this corpus casts the party as corrupt, anti-democratic, and increasingly out of step with Western democratic norms.

AUR (Alianța pentru Unirea Românilor) / George Simion

AUR, led by George Simion, is positioned as a nationalist-populist party with far-right leanings. It is strongly euroskeptic and often associated with russophile rhetoric. The party presents itself as anti-establishment, defending national sovereignty, traditional values, and anti-corruption—though often with xenophobic undertones.

Within the analyzed media corpus, AUR is consistently framed as extremist and subversive. Posts link the party and its leadership to Russian intelligence and portray it as a threat to democratic institutions. For example, one post criticizes the electorate, stating, "George Simion, pushed into Parliament along with Șoșoacă through a reckless vote by 500,000 irresponsible people." Another highlights accusations of foreign ties: "Chris Terheș... now reports the AUR leader for meetings with Russian spies." A third post adds, "Putin's trolls in Romania are turning on each other," reinforcing the narrative of Kremlin influence.

Overall, AUR is not merely depicted as a populist or nationalist force, but as a destabilizing actor aligned with foreign interests. The framing suggests that the party poses a direct challenge to Romania's democratic integrity and Western alliances.

Diana Șoșoacă / SOS România

Diana Șoșoacă, leader of SOS România, is portrayed as an ultranationalist figure with strong pro-Russian views and a staunchly anti-EU, anti-NATO stance. Within the analyzed media corpus, she is consistently framed as unstable and politically harmful.

Her rhetoric and affiliations are frequently linked to Russian propaganda, notably through figures like Dmitri Rogozin, who has amplified her statements on Russian platforms such as VK. One post describes her as "a public danger," while another highlights Rogozin's publication of her message, reinforcing the perception of her alignment with Kremlin narratives.

Șoșoacă is positioned as a fringe extremist figure, with explicit ties to Kremlin-linked discourse. She is not seen as a serious institutional actor but as a symbol of democratic dysfunction and foreign subversion.

PNL (Partidul Național Liberal) / Klaus Iohannis / Nicolae Ciucă

PNL is ideologically positioned as a center-right, pro-European party with a technocratic and economically liberal orientation. However, in the analyzed media corpus, it is portrayed as politically stagnant and complicit in preserving the status quo.

President Klaus Iohannis is frequently criticized for passive leadership, with one post noting that he "reappeared... only to congratulate the National Team," implying disengagement from substantive governance. The party is also accused of forming a self-serving alliance with PSD, described as "The Hydra plan of PSD-PNL for the coming years," suggesting a coordinated effort to maintain power through institutional entrenchment.

Despite its theoretical alignment with reformist values, PNL is framed as compromised and ineffective. Its legacy of technocratic governance is overshadowed by allegations of hypocrisy, media manipulation, and soft authoritarian tendencies. The narrative suggests that PNL, rather than offering a genuine alternative, has become a pillar of the very system it once claimed to reform.

General Political Institutions (Parliament, Government, CCR, EU)

A number of recurring themes emerge across the analyzed content, reflecting widespread public disillusionment with national institutions and growing concerns about democratic integrity. Chief among these is a deep sense of distrust—particularly toward Parliament and the Constitutional Court (CCR)—which are frequently portrayed as corrupt, unrepresentative, or manipulated for political ends.
One post bluntly states, "Parliament is full of thieves who don't represent the people," capturing the prevailing anti-elite sentiment. Another accuses the CCR of engineering electoral outcomes: "CCR... removed an extremist candidate from the race... to engineer the second round," suggesting judicial interference in the democratic process.

Amid this institutional skepticism, the European Union is selectively embraced—especially by reformist voices—as a symbol of modernity, rule of law, and democratic hope. This contrast highlights a broader narrative tension: while domestic institutions are seen as compromised, external anchors like the EU are viewed as potential safeguards for Romania's democratic future.

There is a deep anti-institutional sentiment across the board, but particularly focused on national institutions perceived as compromised (Parliament, CCR), while EU institutions are selectively embraced — especially by reformist voices — as external anchors for Romanian democracy.

Summary

The merged corpus reveals a deeply polarized and adversarial political discourse in Romania, structured around stark ideological binaries: reform versus corruption, Europe versus Russia, and civic modernity versus extremist regression. These oppositions shape the framing of nearly all major political actors and narratives.

Elena Lasconi and USR emerge as the primary carriers of a pro-European, anti-corruption, and civic reform agenda. Their messaging is consistently framed as modern, democratic, and aligned with Western institutions. In contrast, PSD and PNL—despite their formal pro-European positions—are heavily criticized as corrupt, stagnant, and complicit in maintaining a captured state apparatus.

AUR and SOS România are depicted as extremist threats, closely tied to Kremlin interests and promoting socially regressive agendas. Their presence in the political landscape is framed as a danger to democratic norms and Western alignment.

The broader public narrative leans strongly anti-system, but diverges sharply in its proposed solutions. On one side is a populist-nationalist current, skeptical of Western influence and rooted in traditionalist rhetoric. On the other is a liberal-reformist vision, advocating for transparency, civic engagement, and alignment with the EU and NATO.

Institutional distrust is a unifying theme, particularly toward domestic bodies such as Parliament and the Constitutional Court. However, this skepticism is tempered by selective trust in external institutions—especially the EU—which are seen by reformist voices as safeguards for democratic integrity.

Overall, the corpus leans toward a center-right, reformist, and pro-Western orientation. It strongly favors Elena Lasconi and USR, while delegitimizing both the traditional political elite (PSD, PNL) and radical alternatives (AUR, SOS) through a blend of anti-corruption, anti-Kremlin, and civic moralization narratives.

Community 1 – Populist Victimhood and Sovereigntist Framing

This corpus revolves around a populist-nationalist narrative portraying George Simion and AUR as persecuted defenders of national sovereignty. The discourse presents institutions like the judiciary and police as corrupt instruments of a foreign-aligned elite, fueling themes of victimization, electoral fraud, and media manipulation. The rhetoric is emotionally charged, invoking patriotism and historical grievance to legitimize anti-establishment sentiment. Ideologically, the corpus aligns with right-wing populism, nationalism, and euroskepticism, positioning AUR as the authentic voice of the people against a captured and unjust system.

Entities & Themes

George Simion / AUR

George Simion and the AUR party are framed within a strongly right-wing nationalist-populist and euroskeptic narrative. Posts consistently portray Simion as a persecuted political figure,

targeted by state institutions and the media for his opposition to the political establishment. This framing constructs a narrative of political martyrdom, positioning him as a victim of systemic repression.

The rhetoric is emotionally charged, emphasizing themes of injustice, betrayal, and censorship. Simion is depicted as standing alone against a corrupt elite that has ruled Romania for decades. One post highlights the perceived absurdity of his prosecution: "George Simion has a criminal case because he wrote with chalk on the asphalt. Those who have sold the country in the last 34 years receive special pensions and positions in the state." Another declares, "George Simion and AUR stand up to the treacherous system that has ruled Romania for 34 years!" A third post lists a litany of alleged suppression tactics: "They tried everything: criminal cases, secret police, banning from polling stations, TV, fake polls, gendarmes, sold-out press..."

This narrative reinforces AUR's self-image as the only truly sovereignist and anti-system force in Romanian politics. It draws a sharp line between "patriots" and "traitors," echoing classic populist binaries. The tone is conspiratorial, suggesting coordinated efforts by state institutions, security services, and media outlets to silence dissent and maintain elite control. These posts construct a narrative of political martyrdom around George Simion, characterizing him as a target of a corrupt elite precisely because of his opposition to the status quo. The rhetoric is strongly populist and conspiratorial, often suggesting coordinated efforts by state institutions and media to silence dissent.

PSD and Marcel Ciolacu

In the analyzed posts, PSD and its leader Marcel Ciolacu are framed through a strongly critical, populist lens. The party is portrayed as a defender of entrenched power and a threat to national sovereignty, often accused of acting in the interest of foreign powers or suppressing democratic freedoms.

Ciolacu is specifically targeted with accusations of censorship and authoritarian behavior. Posts draw historical parallels to Romania's communist past, invoking the 1946 elections to suggest that current governance mirrors totalitarian tactics. One post warns, "Ciolacu, banning a party is the beginning of dictatorship! That's how the communists started in 1946!" Another adds, "A prime minister who wants to ban a political party just because he doesn't like it is a national disgrace."

These narratives reflect a broader hostility toward PSD, framing the party as ideologically repressive and disconnected from democratic values. The use of emotionally charged historical comparisons serves to heighten public anxiety and reinforce the image of PSD as a corrupt and authoritarian force within Romanian politics. References to communist-era tactics serve to amplify fears of authoritarianism and frame current governance as ideologically oppressive.

General Anti-Establishment / Nationalist Narrative

This narrative reflects a strongly populist, nationalist, and sovereigntist worldview, often accompanied by euroskeptic undertones. Posts emphasize the defense of national sovereignty against both foreign influence and domestic elites, portraying Romania as a nation under occupation—politically, economically, or ideologically.

Historical references are frequently invoked to legitimize opposition and protest, drawing on figures like Tudor Vladimirescu to frame current political resistance as part of a patriotic legacy. The language used reinforces a binary worldview: patriots versus traitors, the people versus the corrupt elite.

One post declares, "Romania is today a colony that must be liberated!"—a dramatic call to reclaim national autonomy. Another accuses authorities of electoral manipulation: "They kicked AUR members out of the polling stations. I haven't seen anything like this even in movies. A huge fraud is being prepared." A third post proclaims, "Courage is reborn. Romania is reborn!"—a rallying cry for national revival.

Overall, this rhetoric promotes a nationalist-populist worldview and a conspiratorial and emotionally charged vision of politics. There is also a conspiratorial tone suggesting systemic electoral fraud and media manipulation.

Romanian Institutions (Justice, Police, Electoral Process)

The posts analyzed reflect a deep distrust of Romanian institutions, particularly the justice system, law enforcement, and the electoral process. These institutions are frequently portrayed as tools of political repression, used to silence dissent and target opposition figures—especially those affiliated with AUR.

Allegations of abuse of power are common. For example, multiple posts highlight the criminal investigation of George Simion for writing with chalk on the pavement, framing it as an absurd and politically motivated act of persecution. One post sarcastically refers to Romania as "the state of the righteous," mocking the justice system's priorities. Another post claims, "The polling stations have become military units. We were not allowed access," suggesting that the electoral process is being militarized and manipulated to exclude opposition voices.

This rhetoric aligns with a broader anti-system narrative, portraying legal and procedural actions not as neutral enforcement of the law, but as targeted suppression. By emphasizing selective justice and institutional overreach, these posts work to delegitimize state authority and reinforce populist claims of elite control and democratic decay.

Summary

The collected discourse reveals a dominant populist-nationalist narrative, centered around the persecution of George Simion and the AUR party by what is framed as a corrupt, foreign-controlled or elite-dominated system. The tone is consistently anti-establishment, with strong emotional appeals to patriotism, victimization, and historical injustice. Institutions such as the judiciary, police, and electoral authorities are portrayed as tools of suppression rather than neutral bodies, and there is frequent invocation of conspiracy-like themes (e.g., electoral fraud, media manipulation).

The ideological leanings reflected are right-wing populist, nationalist, and sovereigntist, with hints of euroskepticism. The discourse strongly opposes both mainstream political parties (especially PSD) and institutional authority, while elevating Simion and AUR as the legitimate representatives of "the people."

Community 4 – Conspiratorial National Romanticism and Anti-Globalism

This corpus blends nationalist-romantic imagery with anti-Western and, at times, antisemitic conspiracy theories. It glorifies ancient Dacian and Romanian legacies, fostering cultural pride through historical revisionism and geopolitical victimhood narratives. Content includes recurrent critiques of the U.S., Israel, global finance, and modern economic systems, suggesting covert manipulation of Romanian sovereignty. The discourse promotes skepticism toward global institutions and Western influence, situating itself within a broader nationalist-populist and sovereigntist ideological frame, laced with cultural essentialism and anti-modern economic views.

All posts originate from the website **dni.org.ro** and its social media presence. The site positions itself as an educational or heritage-promoting platform, but a closer examination of themes and rhetorical patterns reveals a distinct ideological orientation.

Entities & Themes

Romanian Heritage, Achievements, and Natural Resources

This narrative emphasizes national pride, cultural heritage, and historical self-reliance. Posts highlight Romania's early technological achievements and rich natural resources, often framed in a romanticized and nationalist tone. The country is portrayed as historically innovative and resource-rich, yet underappreciated or exploited by external forces.

Examples include references to Romania's pioneering role in the global oil industry: "Romania, the country that had the world's first oil well, the first refinery, and the first country to export gasoline." Other posts celebrate ancient Dacian heritage, such as "The Kosons of the Dacians, gold coins from Sarmizegetusa Regia," and "The Dacian gold treasure from Perșinari,

Dâmbovița County." Additionally, the Apuseni Mountains are described as home to "the largest precious metal deposit in Europe."

These messages support a nationalist-conservative worldview, using historical and cultural symbols to inspire pride and assert Romania's legacy of independence and strength. The underlying tone suggests that Romania has been unjustly underestimated or constrained by foreign interests, reinforcing calls for sovereignty and national revival.

Anti-Western and Anti-American Geopolitical Narratives

This narrative promotes a deep skepticism toward U.S. foreign policy and Western geopolitical influence, often through historical revisionism and conspiracy-leaning interpretations of past events. Posts highlight controversial or covert U.S. actions to question the legitimacy of American global leadership and to frame the West as manipulative or imperialistic.

Examples include references to "Operation Ajax," which claims the U.S. orchestrated the 1953 coup in Iran, and "Operation Northwoods," alleging that the U.S. once planned terrorist attacks against its own citizens in 1962. Other posts list decades of U.S. military interventions (1950–2011) and assert that the 2003 Iraq War was based on lies about weapons of mass destruction, describing it as a war for oil.

These narratives support a strongly anti-American and anti-globalist worldview. By emphasizing historical grievances and covert operations, they aim to undermine trust in Western institutions—particularly the United States—and challenge the moral authority of its foreign policy. The overall tone is accusatory and distrustful, reinforcing broader themes of sovereignty, resistance, and opposition to global hegemony.

Anti-Israel and Antisemitic Rhetoric

This narrative combines geopolitical hostility toward Israel with overt antisemitic tropes and historical revisionism. Posts frame Israel as an instigator of violence and manipulation, while also invoking conspiratorial interpretations of Jewish involvement in historical events.

One post references the Lavon Affair, claiming it "demonstrated Israel's ability to use terrorist operations," suggesting a pattern of covert aggression. Another cites the 1967 attack on the USS Liberty during the Six-Day War to portray Israel as a reckless or hostile actor. More troublingly, a post revisits the 1920 bombing of the Romanian Senate, attributing it to "a group of Jewish communists," which reflects classic antisemitic framing that blames Jewish communities for acts of terrorism and political subversion.

These narratives go beyond criticism of Israeli policy and veer into antisemitic ideology. By repeatedly invoking such examples, the discourse promotes a worldview in which Jewish

individuals or the state of Israel are cast as inherently dangerous or deceitful. This framing is not presented as neutral historical analysis but as part of a deliberate ideological agenda.

Economic and Monetary Criticism

This narrative reflects a deep skepticism toward modern banking and global financial systems, often framed in populist or conspiratorial terms. Posts criticize the foundations of contemporary monetary policy, portraying it as inherently exploitative and disconnected from real economic value.

One post claims that "modern bankers borrowed this system from medieval goldsmiths," suggesting that today's financial practices are rooted in historical deception. Another blames fiat currency and fractional reserve lending for economic instability: "The main cause of economic crises [...] is fiat money [...] the system called 'fractional reserve lending'."

These critiques align with broader anti-global finance rhetoric, commonly found in nationalist or conspiracist circles. While not always explicitly antisemitic, the language and framing often echo historical tropes that have been used to target Jewish communities, particularly in critiques of banking and finance. The overall tone is distrustful, portraying the financial system as a rigged structure that benefits elites at the expense of ordinary people.

Historical Revisionism and Conspiratorial Framing

This narrative centers on challenging official historical accounts and promoting alternative interpretations rooted in conspiracy. Posts often invoke terms like "false flag operations" or allege covert foreign influence, such as "The CIA spy network in Romania," to suggest that major historical events have been manipulated or misrepresented by powerful actors.

A repeated example is the claim that "Israel attacked the American ship USS Liberty," which is used to question the integrity of Western alliances and official military narratives. These posts do not merely critique policy—they aim to undermine the legitimacy of mainstream historical understanding.

This framing is characteristic of revisionist and conspiratorial discourse, often aligned with nationalist or authoritarian worldviews. It casts doubt on Western democratic norms and institutions, encouraging skepticism toward established facts and promoting a worldview in which hidden agendas and elite conspiracies shape global events.

Summary

The overall discourse promoted by the site and its shared content blends nationalist-romantic themes with strong anti-Western, anti-American, and anti-Israel rhetoric. In several instances, the messaging also includes antisemitic undertones. The

ideological orientation is distinctly right-leaning and nationalist-populist, with frequent alignment to sovereigntist and conspiratorial worldviews.

Recurring rhetorical patterns include the glorification of ancient Dacian and Romanian achievements to foster national pride, often framed as evidence of historical greatness and self-sufficiency. Historical revisionism is used to portray Romania as a victim of foreign manipulation, particularly by Western powers. Geopolitical conspiracies involving the United States, Israel, and global finance are common, reinforcing a narrative of external control and betrayal.

There is also a strong emphasis on skepticism toward globalist institutions and modern economic systems, especially banking and fiat currency. These narratives are typical of Eastern European alternative media ecosystems that promote historical revisionism, cultural nationalism, and distrust of liberal democratic norms.

CONCLUDING REMARKS

This first PROMPT narrative report highlights the urgency and complexity of addressing disinformation across various sociopolitical domains. Through in-depth case studies focused on the Russia–Ukraine conflict and the 2024–2025 European electoral landscape, the report provides a nuanced understanding of how disinformation narratives emerge, spread, and evolve across platforms, languages, and cultural contexts.

The methodological framework developed by PROMPT—combining AI-driven large language models, semantic-axiological coding, and dynamic network analysis—has demonstrated its strength in capturing both explicit coordination and subtle, time-based behavioral synchrony in online spaces. These tools allow for a deeper comprehension of narrative dynamics, amplification mechanisms, and key actors involved in shaping public discourse.

While significant progress has been made in identifying and tracing disinformation related to armed conflict and electoral manipulation, challenges remain—particularly in capturing narratives targeting vulnerable groups such as the LGBTQIA+ community. Data collection and methodological adaptation are ongoing in this area, with promising preliminary results that will be further explored in the second narrative report.

As disinformation becomes increasingly multimodal, multilingual, and emotionally charged, PROMPT's findings offer valuable insights for journalists, fact-checkers, policymakers, and civil society actors. By equipping these stakeholders with data-driven tools and narrative intelligence, the project seeks not only to detect and interpret harmful narratives but also to strengthen democratic resilience and information integrity across Europe.

The next phase of the project will continue to expand on these foundations, integrating additional case studies and refining detection techniques to ensure timely, context-sensitive responses to the evolving disinformation landscape.

REFERENCES

Ahmed, H., Traore, I., & Saad, S. (2022). Disinformation detection on social media: An integrated approach. Multimedia Tools and Applications, 81, 36965–36999. https://doi.org/10.1007/s11042-022-13129-y

Altay, S., Berriche, M., Heuer, H., Farkas, J., & Rathje, S. (2023). A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. Harvard Kennedy School Misinformation Review, 4(4), 1-34.

Barbero, F., op den Camp, S., van Kuijk, K., García-Delgado, C. S., Spanakis, G., & Iamnitchi, A. (2023). Multi-modal embeddings for isolating cross-platform coordinated information campaigns on social media. arXiv. https://arxiv.org/abs/2309.12764

Baumgartner, Frank R., ed., Christian BREUNIG, ed., Emiliano GROSSMAN, ed. (2019). Comparative Policy Agendas: Theory, Tools, Data. Oxford: Oxford University Press, pp. 3-16. ISBN 978-0-19-883533-2. Available under: doi: 10.1093/oso/9780198835332.003.0001

Bayer, J., Bitiukova, N., Bard, P., Szakács, J., Alemanno, A., & Uszkiewicz, E. (2019). Disinformation and propaganda–impact on the functioning of the rule of law in the EU and its Member States. European Parliament, LIBE Committee, Policy Department for Citizens' Rights and Constitutional Affairs.

Bevan, S., & Jennings, W. (2019). The UK policy agendas project. In Comparative policy agendas: Theory, tools, data (pp. 176-183). Oxford: Oxford University Press.

Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. Stanford Institute for Human-Centered AI. https://arxiv.org/abs/2108.07258

Budge, Ian, et al. (2001). Mapping policy preferences: Estimates for parties, electors, and governments 1945-1998. Vol. 1. Oxford University Press, USA.

Diesner, J. (2015). Small decisions with big impact on data analytics: A case study in textual network analysis. 2015 IEEE International Conference on Big Data (Big Data), 2879–2888. https://doi.org/10.1109/BigData.2015.7364089

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. Poetics, 41(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

EDMO. (2023). Disinformation in Russian language spreading through the EU. https://edmo.eu/publications/disinformation-in-russian-language-spreading-through-the-eu/

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of communication, 43(4), 51-58.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with class-based TF-IDF and embeddings. arXiv preprint arXiv:2203.05794. https://arxiv.org/abs/2203.05794
Hunt, A., & Papakyriakopoulos, O. (2023). An agent-based model of cross-platform information diffusion and the simulated cross-platform impacts of interventions. Social Network Analysis and Mining, 13(1), 45. https://doi.org/10.1007/s13278-024-01305-x

Hunter, R. (2023). The Ukraine crisis: Why and what now?. In Survival February-March 2022 (pp. 7-28). Routledge.

Klingemann, H.-D., Volkens, A., Bara, J., Budge, I. and McDonald, M. D. (2006) Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Central and Eastern Europe, European Union and OECD, 1990–2003. New York: Oxford University Press.

Korostelina, K. V., Zakharchenko, A., Rohalov, O., Molotsiian, V., Fursenko, M., Zakharchenko, O., & Bakhtiarova, K. (2025). Dynamics of national identity narratives in Ukraine. National Identities, 1-25.

Lawrence, P. (2022). Russia's war in Ukraine. Journal of Global Faultlines, 9(2), 198-211.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

Nahon, K., & Hemsley, J. (2023). Disinformation on digital media platforms: A market-shaping approach. New Media & Society. https://doi.org/10.1177/14614448231207644

NODES. (2023). Narratives Observatory combatting Disinformation in Europe Systemically. Retrieved May 7, 2025, from https://nodes.eu/

Pacheco, D., Flammini, A., & Menczer, F. (2021). Uncovering coordinated networks on social media: Methods and case studies. Political Communication, 38(1-2), 130–150. https://doi.org/10.1080/10584609.2020.1834120

Phillips, W., & Milner, R. M. (2021). You are here: A field guide for navigating polarized speech, conspiracy theories, and our polluted media landscape. MIT Press.

Ptaszek, G., Yuskiv, B., & Khomych, S. (2024). War on frames: Text mining of conflict in Russian and Ukrainian news agency coverage on Telegram during the Russian invasion of Ukraine in 2022. Media, War & Conflict, 17(1), 41-61.

Righetti, N., & Balluff, P. (2023). CooRTweet: Coordinated Networks Detection on Social Media. R package version 2.1. 0.

Saqr, M. (2024). Temporal Network Analysis: Introduction, Methods and Analysis with R. In Learning analytics methods and tutorials: a practical guide using R (pp. 541-567). Cham: Springer Nature Switzerland.
Sebők, M., Máté, Á., Ring, O., Kovács, V., & Lehoczki, R. (2024). Leveraging Open Large Language Models for Multilingual Policy Topic Classification: The Babel Machine Approach. Social Science Computer Review, 0(0). https://doi.org/10.1177/08944393241259434

Starbird, K. (2017, May). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1, pp. 230-239).

Starbird, K., & Wilson, T. (2020). Cross-platform disinformation campaigns: Lessons learned and next steps. Harvard Kennedy School Misinformation Review, 1(6). https://doi.org/10.37016/mr-2020-49

Ventsel, A., Hansson, S., Madisson, M. L., & Sazonov, V. (2021). Discourse of fear in strategic narratives: The case of Russia's Zapad war games. Media, War & Conflict, 14(1), 21-39.

Wilson, T., & Starbird, K. (2020). Cross-platform disinformation campaigns: Lessons learned and next steps. Harvard Kennedy School (HKS) Misinformation Review, 1(1). https://doi.org/10.37016/mr-2020-002

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. Political Communication, 38(1–2), 98–139. https://doi.org/10.1080/10584609.2020.1785067