PROMPT

Narrative Report #2

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

Deliverable 5.2 - 2nd Narrative Report on the state of disinformation narratives on the Moldovan Parliamentary elections, the war of aggression against Ukraine and LGBTQ+ rights and communities	
Contributors	Rīga Stradiņš University (RSU) (Latvia), opsci.ai (France), Università degli Studi di Urbino (Italy), HUN-REN Center for Social Sciences (Hungary), Erich-Brost Institut (Germany), Asociația Digital Bridge (Romania), Andra-Lucia Martinescu and Marius Dima (Diaspora Initiative)

PROMPT is a pilot project co-financed by the European Union under Grant Agreement CNECT/LC-026293

TABLE OF CONTENTS

INTRODUCTION

Disinformation is an ever-changing phenomenon, with new topics, players and techniques being developed and instrumentalised. Against this backdrop, the second European Narrative Observatory - PROMPT - employs AI-driven methods to help monitor disinformation narratives, how they propagate and transform across social platforms and local contexts.

This second narrative report delves into the evolving landscape of disinformation, with a distinct focus on the recent Moldovan parliamentary election. It also presents PROMPT's findings on the disinformation landscape regarding the war of aggression against Ukraine and the LGBTQ+ community. A special emphasis is placed on the impact of the war in Ukraine and its reverberations within the electoral discourse in Moldova, highlighting how external conflicts can be leveraged to manipulate local sentiments and polarize communities. Across these chapters, we point out the coordinated dissemination of disinformation narratives, as well as the persuasion techniques, rhetorical devices and emotional triggers mobilised to propagate them. Building on engagement metrics, these analyses provide insights on the most active accounts being followed, engaged with, shared and reacted to on each topic.

The report also explores the experience and perspectives of journalists working on disinformation in Romania, Italy, Estonia, Latvia, Lithuania and France. Besides shedding light on country-specific disinformation dynamics, they also focus on their day-to-day realities of identifying, countering, and reporting on disinformation, and the ways in which evolving strategies and platform dynamics are reshaping their professional practices.

By combining analyses of major disinformation narratives, their dissemination patterns and firsthand journalistic insights, this report aims to provide an in-depth understanding of the mechanisms, sources, and societal impact of contemporary online information manipulation.

The technical annex to the report details the process of data collection and analysis, as well as technical limitations to data interpretation at this stage of the project.

1. MOLDOVA'S 2025 PARLIAMENTARY ELECTIONS: DISINFORMATION AS A GEOPOLITICAL BATTLEGROUND

Elections have become a battleground. They are in the global spotlight in countries that did not previously make international headlines. Elections are scrutinised by global audiences in a borderless informational space that expands well beyond traditional constituencies. This visibility, however, is Janus-faced. On the one hand, it empowers democratic oversight, lending civil societies - and monitoring/disinformation detection efforts - a necessary leverage. On the other hand, it offers fertile ground for hostile operations that exploit the very same attention pathways and transnational purview. In this environment, **global audiences themselves have become the object of dissection and micro-targeting,** and individuals entrenched in fragmented streams of information that mirror, and often magnify, the fractures within the societies they observe.

Platforms that claim to democratise access to information facilitate these dynamics through design logics that privilege amplification over verification, and algorithmic engagement over integrity. A consistent policy response to this unchecked asymmetry has yet to materialise,[1] leaving electoral ecosystems structurally vulnerable to manipulation by malign actors who acutely understand the mechanics of virality and outrage.

1.1. Concepts, structure and main findings

Our approach[2] to the Moldovan elections helps reconceptualise contemporary geopolitics as a competition for the governance of information environments. Power depends on the capacity to structure visibility, circulation, and credibility. Interference is no longer pre-eminently aimed at persuasion or outright ideological conversion, but at **participatory deterrence**, or the depletion of civic engagement through fatigue, cynicism and the pre-emptive delegitimisation of electoral choice.[3]

By controlling the propagation of narratives, identities and publics, influence operations reshape the preconditions of public participation, not by force, but by re-organising the space in which democracy is deliberated; and with the aim not to convert but to erode trust.

The **analysis offers a radiography of electoral interference**, both online and offline. It examines the Moldovan context before analysing the disinformation mechanisms at work, looking specifically at:

- the **wider geopolitical and socio-economic landscape**, emphasising how hybrid interference exploits existing domestic vulnerabilities, historical rifts and/or regional alliances.
- **a cross-platform examination of the online environment**, using targeted data collection to chart the scale and transnational reach of manipulative interventions across social media and the web via networked geographies.
- **the Tactics, Techniques and Procedure**s (TTPs) through which digital geopolitics are enacted, blurring the lines between origin, affiliation and attribution. Such patterns include templated amplification, synchronous vs long-term temporal coordination, narrative laundering, and obfuscated attribution, showing how circular flows of validation are engineered to simulate consensus or grievance within domestic debates. Furthermore, mapping the TTPs reveals how influence operations function less as isolated campaigns than as self-reinforcing loops:

[1] The Digital Services Act (DSA), while an important step forward, ultimately depends on enforcement by domestic regulators whose resources and expertise vary widely across jurisdictions, often producing uneven implementation.
[2] Principal Investigator: Andra-Lucia Martinescu (The Diaspora Initiative / The Foreign Policy Centre), Cognitive AI & Data Architecture: Marius Dima (Qriton), with support from the PROMPT consortium, Attila Biro and the investigative team from Context.ro, Vladimir Buruiana (Moldovan civic diaspora), and more largely, the civil societies and watchdogs on the digital frontlines. An interactive version of this analysis is available at: https://elections.igov.ro/moldova.html.
[3] To situate this shift, we propose a holistic three-layer spatial framework that links infrastructures, operational behaviour (TTPs – Tactics, Techniques, Procedures), and strategic effects on democracy (see full picture in Technical appendix 2)

narratives are tested in one online environment, refined through transnational propagation, and reintroduced domestically as evidence of societal fracture.
- **the knowledge infrastructures:** by applying composite risk metrics to Wikipedia, we reveal how coordinated editing or citation gaps can transform encyclopaedic content into a vector for information manipulation.

The analysis:
- confirms the presence of hostile influence operations but exposes an architecture far more complex and adaptive than current academic or policy frameworks account for.
- argues that electoral interference now functions less as communicative persuasion than as **a geopolitical spatial strategy routed through digital infrastructures**. Rather than advancing along conventional territorial boundaries, influence is exerted through networked geographies composed of platform architectures, language corridors and algorithmically mediated publics, of which the diaspora is one conduit of many. In such environments, disinformation actors are deliberately masked by obfuscated origin and mutable affiliation, where visibility itself becomes weaponised.
- that the objective pursued by such operations is not persuasion but **participatory deterrence**, achieved by exhausting civic agency rather than converting opinion.

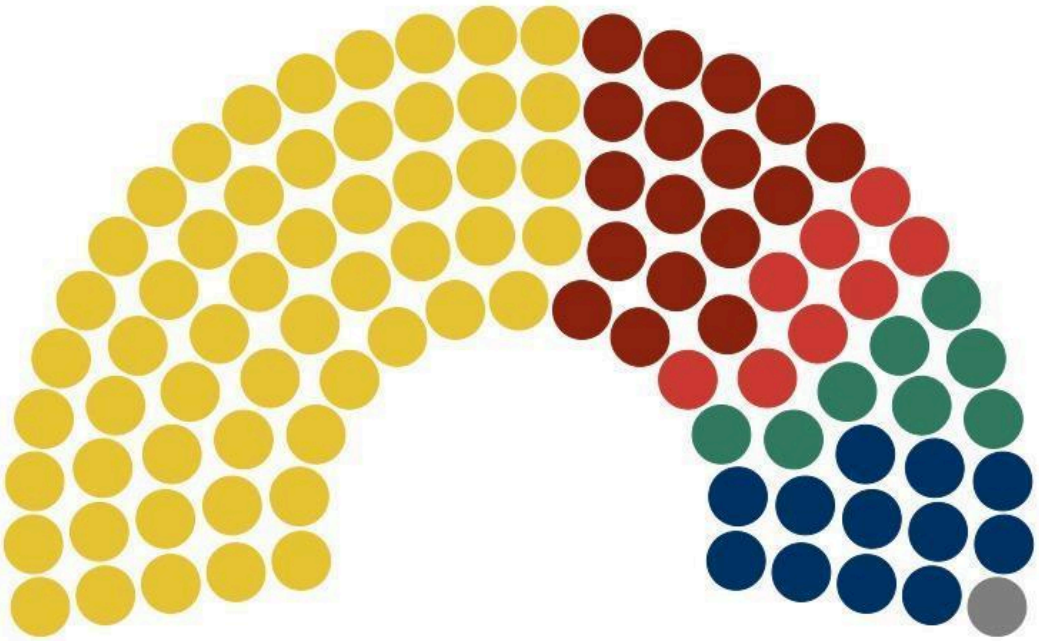
1.2. Contextual analysis

The patterns of interference preceding and accompanying Moldova's parliamentary election cannot be understood solely through the lens of online manipulation and propaganda. Digital operations thrive in conditions already embedded in domestic and regional political arenas. Dynamics such as the geopolitics of 'neutrality' (long instrumentalised by Russia) to proxy candidates, (covert) financial streams, do not simply co-exist with disinformation; they enable and magnify it, conferring influence operations both entry points and receptive audiences.

To situate the 2025 vote in Moldova within a **genealogy of subversion** is to understand Moldova's political contestation as more than a domestic struggle over governance. Positioned at Europe's Eastern edge, **this small, landlocked post-Soviet state functions simultaneously as a nodal point in Europe's security architecture and as a (historical) laboratory for hybrid destabilisation.**

1.2.1. A vote at the fault line of Europe

Moldova entered its September 2025 parliamentary elections at a moment of rare consequence, with rising geopolitical pressures magnifying what was already perceived as a pivotal vote. In a parliamentary regime, where control of the legislature determines both government formation and strategic direction, the vote assumed historic weight. Would Moldova consolidate its European path or succumb to Russian influence? The outcome delivered a measure of clarity. The pro-European Party of Action and Solidarity (PAS) secured an outright majority with 55 of 101 seats, a mandate sufficient to govern without coalition partners and to stabilise the political landscape. Alongside PAS, the new parliamentary configuration will include the *Patriotic Bloc* (*Blocul Patriotic*) with 26 seats, the *Alternative Bloc* (Blocul Alternativa) with eight seats, *Our Party* (Partidul Nostru) and the *Democracy at Home Party* (Democratia Acasa - PPDA) with six, reflecting a fragmented but contained opposition totalling 46 seats.



Party	Seats
Partidul Actiune si Solidaritate (PAS)	55
Partidul Socialistilor din Republica Moldova (PSRM)	17
Partidul Comunistilor din Republica Moldova (PCRM)	8
Blocul Alternativa (BA)	8
Partidul Nostru (PN)	6
Partidul Democratia Acasa (PPDA)	6
Independenti	1

While the ballot was tallied, the pro-Russian Patriotic Bloc leadership, accompanied by a small group of supporters, staged a demonstration outside the Central Electoral Commission, threatening to reject the outcome after claiming their own victory earlier that evening.[4] Such contradictions were a mainstay on the campaign trail, instrumentalised to mobilise grievance, while preserving contestation as a political resource irrespective of the outcome. No incidents were reported, and the protest soon defused. Yet, no matter how anti-climactic or short-lived, **this rally was the latest expression of a months-long hybrid destabilisation campaign with an unprecedented degree of coordination and intensity.** It may not be the last in a country already strained by mounting economic pressures – inflation hovers at 9%, energy prices are surging; and economic precarity endures, coupled with persistently low salaries and pensions.[5] **Moldova did exhibit greater preparedness than some of its regional allies**, but it was not completely immune to the cross-border spillovers that shaped Romania's recent suite of presidential elections.

One striking example was **the unexpected ascendance of Vasile Costiuc's *Democratia Acasa* Party (PPDA)** to pass the 5% threshold and win six parliamentary mandates, running on a 'sovereignist'

[4] 'Protest nocturn la CEC al Blocului Patriotic' (28 September 2025) *in Ziarul de Garda* (ZdG). Available online at: https://www.zdg.md/stiri/protest-nocturn-la-cec-al-blocului-patriotic-daca-in-noaptea-asta-vor-fi-falsificari-noi-maine-nu-vom-recunoaste-alegerile/.

[5] The International Monetary Fund Report notes that 'inflationary and energy-related pressures continue to strain institutional capacity. International Monetary Fund (July 2024). Republic of Moldova – Fifth Review under the ECF/EFF. Available online at: https://www.imf.org/en/Publications/CR/Issues/2024/07/11/Republic-of-Moldova-Fifth-Reviews-Under-the-Extended-Credit-Facility-and-Extended-Fund-551687. Average gross earnings in Moldova reached 15,470.6 MDL per month in Q2 2025 (≈ €794 at 19.5 MDL/EUR, or ~€1,320 in PPP-adjusted terms), compared to an EU full-time adjusted average of ~€3,158/month in 2023 (Eurostat), underscoring persistent wage gaps despite nominal growth.

platform reinforced by the support networks of the far-right Alliance for the Union of Romanians (AUR). Costiuc himself has been tied to a number of Russian ventures and to dubious alliances with Vlad Plahotniuc, who was recently extradited from Greece to face corruption charges.[6] Pre-electoral polls had failed to credit PPDA with any realistic chance of clearing the threshold, even as the party's messaging and visibility became increasingly evident across online ecosystems[7] - echoing the sudden prominence of (fringe) radical currents during Romania's presidential race, where algorithmic amplification and digitally fuelled mobilisation shifted the political centre of gravity.

The broader aftermath, however, is just as crucial. In both Romania and Moldova, political contests not only exposed but also accelerated societal polarisation, rooted in perceptions of disenfranchisement and otherwise legitimate socio-economic grievances that had long simmered just beneath the surface. These fractures were magnified by hybrid pressures, **weaponising discontent across transnational information spaces and driving online-offline mobilisation cycles**. The PPDA breaching the parliamentary threshold (after three failed attempts) could produce ripples in the longer term. Procedural legitimacy may afford the populist platform enough room to mainstream its narratives, beyond its platform-engineered grassroots activism – an operation that nevertheless succeeded in reaching a critical mass, both domestically and abroad.

The Moldovan diaspora, estimated at 1.2 million individuals, emerged as a decisive political force. In recent electoral cycles, the state has expanded overseas voting, opening a record 301 polling stations across 41 to 45 countries for the parliamentary election, and offering postal voting in designated states.[8] Turnout abroad was relatively high, with over 275,000 casting a ballot by closing time. Preceding that day, however, the diaspora was targeted by a vast information warfare campaign that involved the Matryoshka networks (translated as 'nested doll') — layers upon layers of cloned media, proxy outlets, and visible or anonymous online personas mutually reinforcing each other.[9] The aim was to demobilise diaspora participation and to undermine trust in democratic processes by staging or urging protests both abroad and at home.

1.2.2. Proxy mobilisation in the electoral arena

In (brief) retrospect, September's parliamentary race unfolded against the backdrop of a fragmented political landscape with fifteen political parties, four electoral blocs and four independent candidates vying for control of the legislature.[10] There were indeed some surprising twists of events. On the cusp of voting, the Central Electoral Commission (CEC) barred the *Moldova Mare* (Greater Moldova) party, led by former prosecutor Victoria Furtuna,[11] from running, amid sweeping investigations into illegal financing and vote-buying schemes tied to Russia, the fugitive oligarch Ilan Shor, and proxy

[6] Gabriel Mateescu (28 September 2025). 'Democratia Acasa, partidul sustinut de AUR la alegerile din Moldova, urca pe locul al patrulea' *in Context.ro* (an investigative journalism outlet). Available online at: https://context.ro/democratia-acasa-partidul-sustinut-de-aur-la-alegerile-din-moldova-urca-pe-locul-al-patrulea-dupa-procesarea-a-aproape-jumatate-din-sectiile-de-vot/.

[7] The PROMPT consortium in collaboration with investigative journalists from Context.ro (also part of FACT hub) analysed the TikTok surge associated with PPDA and leader Vasile Costiuc, prior to the ballot. Analysis available online at: https://context.ro/1000-de-tehnici-de-manipulare-pentru-alegerile-din-republica-moldova-cazul-costiuc/.

[8] Adept Association (August 2025). *Moldova Parliamentary Elections 2025: Polling Stations Abroad* (UNDP Report). Available online at: https://www.undp.org/sites/g/files/zskgke326/files/2025-09/adept_note_on_polling_stations_abroad_2025.pdf

[9] Gabriel Gavin (August 2025). 'Russia targeted voters across EU, Moldova warns' *in Politico*, available online at: https://www.politico.eu/article/russia-moldova-voting-elections-candidates-west-kremlin/.

[10] According to the Parliamentary Elections Portal (September 2025), available online at: https://alegeri.md/w/Alegerile_parlamentare_din_2025_%C3%AEn_Republica_Moldova#Concuren.C8.9Bi_electorali.

[11] In July 2025, Victoria Furtuna was placed under EU sanctions (2025/1434 OJ L202501434). Available online: https://data.europa.eu/apps/eusanctionstracker/subjects/177594. Also see, Thomas Rowley (15 September 2025). 'Fugitive Moldovan tycoon recruits top Russian bankers to run sanctions-busting crypto firm: leak' in Reporter London, available online at: https://reporter.london/?p=1484.

infrastructures such as the Evrazia foundation.[12] While the platform assumed a 'sovereignist', ostensibly nationalist rhetoric, the probes revealed direct coordination with Russian curators, amongst them Anton Tregub and Alexandr Petrov, who funnelled hundreds of thousands of euros into the party's campaign operations, including vast promotion activities on social media (Facebook, Instagram, TikTok), and Google.

In parallel, Irina Vlah's party *Inima Moldovei* (Heart of Moldova), of the Patriotic Electoral Bloc (Blocul Patriotic – a coalition of pro-Russian, post-communist factions led by former president Igor Dodon), was also struck from the ballot, with the Court citing bribery and illegal financing.[13] Vlah was the former governor of autonomous Gagauzia, a predominantly Turkic enclave of Orthodox belief that, in the aftermath of Russia's 2022 invasion of Ukraine, emerged as the preferred staging ground for Russian influence operations, and a stronghold for Shor's funded destabilisation efforts. The importance of this region, an economically deprived sliver of land in southern Moldova, cannot be underestimated in Russia's and, by default, its proxies' strategic calculus: Gagauzia's territorial status has long been gamed to undercut Moldova's fragile sovereignty, echoing the 1990s when its separatist mobilisation unfolded in tandem with Transnistria's, albeit with different outcomes.

In August 2025, Evghenia Gutsul, Vlah's successor and close affiliate of Ilan Shor, was sentenced to seven years in prison for funnelling Russian funds into the Shor Party between 2019 and 2022,[14] including illicit subsidies (*i.e.*, cash-based payments) used to orchestrate protests and propaganda activities, at times opportunistically reviving the separatist rhetoric.[15] In 2023, Gutsul was elected *bashkan* (governor) of Gagauzia on the *Victory* (Pobeda) *Bloc*'s ticket, while serving as its executive secretary.[16] This umbrella alliance was founded by Shor (also its chairman) as a surrogate vehicle after the Constitutional Court officially banned his party in 2023; by then, he had already fled to Russia. The Bloc's inaugural congress was ominously held in Moscow, and assembled Gagauz officials, remnants of the banned Shor Party and disparate pro-Russian factions, 'supporting Moldova's accession to the Eurasian Economic Union [a Russian-led structure]', closer ties with the Community of Independent States (CIS) and '(...) traditional partners and neighbours', explicitly referencing Russia.[17] In other words, the Bloc positioned itself to obstruct Moldova's European trajectory by leveraging captive pro-Russian constituencies in Gagauzia and beyond, embedding itself in the country's most vulnerable political and socio-economic fault lines to sustain Russian influence despite institutional bans.[18]

---

[12] Journalistic investigation conducted by Deschide.md (Moldovan News Outlet). Cristian Reznic (26 September 2025). 'Victoria Furtuna, coordonata de Moscova si Ilab Sor: Cum Rusia a finantat activitatile partidului Moldova Mare' in Nord News (MD), available online at: https://nordnews.md/stiri-nationale/social/finantare-ilegala-moldova-mare-rusia/.

[13] Stephen McGrath (Associated Press). Moldova bars two pro-Russian parties from high-stakes parliamentary election' in PBS News (26 September 2025). Available online at: https://www.pbs.org/newshour/world/moldova-bars-two-pro-russian-parties-from-high-stakes-parliamentary-election.

[14] In 2023, Ilan Shor – a Moldovan oligarch and former mayor of Orhei - was convicted in absentia by the Chisinau Court of Appeal for his role in the '$1 billion bank fraud' (Source: The Guardian), receiving a 15-year prison sentence and the confiscation of assets (Source: Reuters). Shor had already fled Moldova in 2019 while under investigation and was believed to be residing in Israel at the time of the ruling. In the months following his conviction, he was sanctioned by both the EU and the US for acts of corruption and destabilisation of Moldova's democratic institutions. Council Implementing Regulation (EU) 2022/2408 of December 2022. Official Journal: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2408. U.S. Treasury Department (OFAC), 'Sanctions of Corrupt Oligarchs and Kremlin-Linked Actors' (26 October 2022):https://home.treasury.gov/news/press-releases/jy1054.

[15] Saman Nazari (December 2024). *Shor's Echo: Influence Operations Targeting Moldovan Gagauzia* (Alliance4Europe Report). Available online at: https://alliance4europe.eu/shors-echo-influence-operations-targeting-moldovan-gagauzia.

[16] Madalin Necsutu (May 2023). 'Pro-Russian's Governorship Win in Moldova's Gagauzia Upheld' in Balkan Insight, available online at: https://balkaninsight.com/2023/05/23/pro-russians-governorship-win-in-moldovas-gagauzia-upheld/.

[17] Shor's opening remarks quoted by Infotag, a Moldovan news portal, available online at: https://www.infotag.md/politics-en/315485/.

[18] The infusion of illicit capital driving the 2023 gubernatorial elections fringed on the surrealist, which, setting aside the entrenched corruption, vote-buying schemes and pyramid-like financial operations traced to sanctioned Russian banks (i.e.: PSB),# also witnessed the unveiling of a flashy amusement park, GagauziyaLand, spun as a philanthropic initiative of Shor's, now commanding a dilapidated rural landscape - the soft facade of yet another Potemkin village camouflaging systemic capture, much like its other iteration in Orhei (OrheiLand) - Sarah Rainsford (19 October 2024). 'Russian cash-for-votes flows into Moldova as nation heads to polls' in BBC News, available online at: https://www.bbc.co.uk/news/articles/c23kdjxxx1jo.

Foreshadowing the elections that were to follow, **Gagauzia served as a laboratory for tactics later scaled across Moldova**, fusing illicit financial pipelines, on-the-ground patronage networks, as well as digital mobilisation and coordinated disinformation campaigns.

Although Moldova's CEC reiterated its 2024 decision to prohibit the *Victory Bloc* from participating in the 2025 parliamentary ballot, its presence still loomed large, resurfacing in online spaces and recalibrating its weight behind amenable politicians and parties, exposing how proxy infrastructures adapt and re-embed under new guises. Even from his safe haven abroad and despite the weight of Western sanctions, Ilan Shor, the convicted oligarch ensconced in Russia, continues to cast a long shadow over Moldovan politics - not as a mere appendage of Russian influence but as **one of the Kremlin's privileged conduits for interference**.

1.2.3. The road not taken – and why it matters

What would have happened if the September parliamentary vote had yielded a different result? Had no party secured an outright majority, coalition building would have become critical, but also the most volatile factor shaping the next government. Previous attempts at power sharing had repeatedly collapsed under the weight of mistrust, corruption and competing geopolitical loyalties.

Uneasy alliances oscillated between (at times) reformist, European-leaning platforms and opportunistic political arrangements of the old guard, newly emergent elites, or oligarchic clans, all variously bound by crippling corruption schemes and the long reach of Russia's patronage networks.[19] At their most malign, these coalitions periodically resuscitated bids for reintegration into Russian-led structures, triggering parliamentary dissolutions, sudden realignments and prolonged episodes of institutional paralysis.

Such moments of fracture reliably yielded geopolitical dividends for Moscow, which treated political/domestic instability not necessarily as a by-product but as a strategic asset — a calibrated opportunity to stall, dilute, and reverse even the most incremental alignment with Europe or the West. Every crisis became a reset point, one step forward towards integration, two steps back into the grey zone of neutrality, and externally managed stagnation. The ongoing war in Ukraine, however, has added an extra layer of urgency for Russia. What we may consider isolated incursions can be better understood as **components of a broader hybrid coercion approach**, whereby political interference, electoral disruption and kinetic probing (including the recent drone incursions)[20] are deployed in tandem, not only to test reaction but also to exhaust/overwhelm state and alliance-level responses.

This is precisely what has been contested. Under the incumbent administration, which has retained a governing majority since the July 2021 snap parliamentary elections, when the Party of Action and Solidarity (PAS) secured 63 out of 101 seats, the country experienced, for the first time since independence, **an uninterrupted pro-European parliamentary-executive alignment**. This was the first legislature to explicitly align its agenda with the European integration process.[21] Politically, it

---

[19] For instance, in 2019, the short-lived PSRM-ACUM coalition formed to dismantle the oligarchic control of Vlad Plahotniuc, collapsed five months later when PSRM joined the Democratic Party in a no-confidence vote that toppled Maia Sandu's government.# The successor Ion Chicu cabinet (aligned with pro-Russian President Igor Dodon) promptly soft-pedalled justice reforms, revived Moscow-centric initiatives (including a Russian state loan later struck down by the Constitutional Court), and reoriented energy and diplomatic channels eastward - see Eugen Urusciuc (27 September 2025). 'Parlamentul R Moldova (...)' *in Radio Free Europe Moldova*, available online at: https://moldova.europalibera.org/a/parlamentul-r-moldova-de-la-agrarieni-si-comunisti-coalitii-monstruoase-aliante-beton-si-binoame-pana-la-majoritate-proeuropeana/33539642.html.

[20] 'Russian drone incursions' *in the Guardian* (15 October 2025) available online at: https://www.theguardian.com/world/2025/oct/15/russian-drone-incursion-tactically-stupid-and-counterproductive-says-polish-minister.

[21] Roughly one in eleven laws passed over the four-year mandate carried the EU imprint, amounting to around 140 acts harmonising national legislation with European standards across multiple sectors. RFE/RL (Sep 2025):

went even further by adopting the Parliamentary Declaration on Moldova's accession to the EU, reaffirming an irreversible commitment to European integration (since 2024 it is also enshrined in the Constitution).[22]

In effect, Moldova moved tangibly westward, securing EU candidate status in 2022 and formally opening accession talks in June 2024. Brussels has since kept the enlargement track active. This geostrategic tilt was reinforced on the security and energy fronts; Moldova's grid was synchronised with continental Europe in March 2022, and gas interconnectivity with Romania (Iași-Ungheni-Chișinău) expanded alternatives to Russian supply. In parallel, Moldova served as a transit corridor in the EU's Solidarity Lanes, routing Ukrainian exports via Moldovan rail and the Giurgiulesti/Danube axis to Romanian ports.[23] Domestically, Chisinau demonstrated resolve in dismantling Russian-linked networks, including the 2023 ban of the Shor Party over its role in orchestrating destabilisation.[24]

1.2.4. The geopolitics of neither east nor west – Europe as a battleground

Yet, the trajectory of domestic reform has been neither linear nor universally embraced, particularly in disaffected constituencies where economic hardship weighs more heavily than geopolitical aspirations. However, it is precisely this tension between strategic reorientation and structural vulnerability that renders Moldova's European turn susceptible to external sabotage and interference. Such efforts extend beyond the manipulation of electoral outcomes, seeking instead to erode public confidence in the very assumption that integration can generate tangible socio-economic or democratic dividends, thereby transforming latent discontent into a lever of geopolitical obstruction. The analysis of cross-platform (mostly Russian-affiliated) disinformation ecosystems substantiates this trend, showing how **the EU has been actively targeted, reframed as either predatory, ineffectual or outright destabilising.**

In line with Russia's information warfare doctrine, rooted in the concept of *reflexive control* and the fusion of psychological, informational and political instruments,[25] **Moldova's policy achievements were deliberately recast as vulnerabilities in a bid to legitimise their reversal**: EU alignment depicted as a loss of sovereignty, security cooperation and the support afforded to Ukraine as provocation, energy diversification as economic sabotage, NATO as a deliberate war proxy, and so forth. The objective is not persuasion in any conventional sense, but rather the systematic erosion of societal resilience, aimed at fragmenting public support and fostering confusion.

Nor were these distortions confined to policy. **Narratives surrounding governance, social cohesion, and even the integrity of elections were relentlessly targeted**, ensuring that democratic participation itself became a site of contestation, mistrust and manipulation. In this sense, Moldova was not merely a receptacle of propaganda but an operational theatre of Russia's hybrid strategy, where the information domain could be weaponised to shape choices before they were even made. The methods deployed drew on a repertoire that had been tested in Ukraine, Georgia, the Western Balkans, and, increasingly, Romania, amongst others.

Since the country's independence, geopolitics has played a disproportionate, albeit valid, role in Moldova's politics, but not necessarily in a coherent manner or as an expression of geopolitical

https://moldova.europalibera.org/a/parlamentul-r-moldova-de-la-agrarieni-si-comunisti-coalitii-monstruoase-aliante-beton-si-binoame-pana-la-majoritate-proeuropeana/33539642.html

[22] During the campaign, Russian-aligned parties and leaders constantly threatened to back-track on this constitutional provision and organise a referendum that would herald a return to the status quo - disinformation outlets amplified this narrative across platforms and the web.

[23] Solidarity Lanes: Moldova and Ukraine (European Commission), available online at: https://international-partnerships.ec.europa.eu/policies/global-gateway/solidarity-lanes-moldova-and-ukraine_en.

[24] Alexander Tanas (June 2023). 'Moldova bans pro-Russian Shor party after months of protests' in Reuters, available online at: https://www.reuters.com/world/europe/moldova-bans-pro-russian-shor-party-after-months-protests-2023-06-19/.

[25] Keir Giles (2016). *The Next Phase of Russian Information Warfare* (Riga: NATO StratCom COE). *Passim*.

conviction. The language of East and West had long functioned as a revolving instrument of leverage, deployed for electoral gain, coalition bargaining, or legitimacy-seeking, often obscuring deeply entrenched transactional governance or clientelism beneath ideological posturing. From Russia's vantage point, fostering controlled fragmentation into political blocs, as seen in this parliamentary race, was less about elevating a single ally than about sustaining volatility. This opportunistic strategy targeted cohesion across the broader European integrationist camp, deliberately diluting and even confusing the pro-European message. By cultivating multiple political actors simultaneously, the approach ensured that parallel channels of influence remained active even if parties or leaders were discredited or excluded from the race. Our analysis of online disinformation and propaganda ecosystems across multiple platforms confirms this pattern.

1.2.5. Nothing Is what it seems – incursions into rhetorical camouflage

An illustrative case is the *Alternative Electoral Bloc* (Blocul Electoral Alternativa - BA), ostensibly (self-declared) as pro-European, but in fact operating as a pro-Russian conduit, aligned with Moscow's strategic interests and official posturing. The Bloc's leadership includes a number of controversial figures, amongst them, Alexandr Stoianoglu (presidential contender in 2024), Ion Ceban (formerly a member of the pro-Russian Socialist Party, PSRM, who was denied entry in Romania and the Schengen area on grounds of national security risks), and Mark Tkaciuk, a communist ideologue who persistently advocated for Moldova's integration into the Eurasian Union and adherence to the Kozak Plan.[26] Throughout its campaign trail, BA avoided a clear positioning on core geopolitical issues, including Russia's aggression against Ukraine, Moldova's relationship with NATO, and its EU accession path. Thus, the adoption of a pro-European rhetoric may be deemed as an electoral tactic designed to appeal to moderate voters without alienating a core pro-Russian base. Essentially, 'nothing is what it seems': ideological lines become deliberately blurred, with 'sovereignist' movements reframing Kremlin positions as nationalism or anti-establishment resistance, and self-declared pro-European blocs in fact treading a carefully curated ambiguity that obscures external alignment. In practice, **such political formations employ camouflage strategies, rebranding hostile agendas in a pro-European vernacular** to preserve influence under shifting electoral and geopolitical constraints.

Furthermore, as evidenced by our data, a substantial share of the information manipulation arsenal was channelled into sustaining 'sovereignist' (far-right) parties and blocs, amplifying their messaging and political foothold. The *Moldova Mare* (Greater Moldova) Party was ultimately banned from running on the ballot, but *Democrația Acasă* - PPDA (under the leadership of Vasile Costiuc) won six mandates/seats, using an aggressive TikTok campaign that propelled its transnational outreach.[27] The PROMPT consortium, in collaboration with the FACT EU Hub, conducted an analysis of PPDA's online ecosystem and rhetoric prior to the elections, forestalling the party's resurgence. At the same time, connections between far-right populist/irredentist movements across the region have increasingly displayed converging agendas and thematic overlaps. Such cross-border spillovers were particularly forceful in the case of Moldova and Romania.

These political hybrids blend nationalist rhetoric with populist tropes, allowing them to exploit domestic grievances while opportunistically tapping into transnational ideological currents, including newly imported slogans and nominal affiliations to the MAGA and its European offshoot, MEGA (Make Europe Great Again) movements.[28] **The circulation of such narratives has relied on an ecosystem of foreign influencers and political technologists, some visible, others concealed behind online**

[26] Reuters (July 2025). *'Moldovan Mayor Barred From Romania Over Security Concerns'*, available online at: https://www.reuters.com/world/romania-bans-moldovan-mayor-border-free-schengen-area-ministry-says-2025-07-09/.
[27] '1000 manipulation techniques in Moldova's elections. The Costiuc Case' (22 Sep 2025), available online at: https://context.ro/1000-de-tehnici-de-manipulare-pentru-alegerile-din-republica-moldova-cazul-costiuc/.
[28] MEGA Scandal la Chisinau. Mai multi participant la o conferinta internationala – interzisi in R. Moldova' (28 July 2025) in Ziarul de Garda (ZdG), available online at: https://www.zdg.md/importante/mega-scandal-la-chisinau-mai-multi-participanti-la-o-conferinta-internationala-interzisi-in-r-moldova-sis-evenimentul-ar-avea-legaturi-dubioase-cu-gruparea-criminala-sor/.

avatars and proxy accounts, with many also operating from the United States (where the origin could be traced).[29] In both cases, unfounded accusations of election fraud and vote theft were intended to pre-emptively discredit the result and incite civil unrest. This is just one of numerous other examples that amassed substantial transnational engagement – a form of ideological franchising, repackaged for local consumption. Such strategies form part of a broader repertoire of political uncertainty, whereby parties or leaders disavow firm geopolitical allegiance while signalling de facto alignment through narrative cues and coalition patterns.

Partidul Nostru (Our Party), which secured six parliamentary seats under the leadership of Renato Usatîi (the pro-Russian former mayor of Bălți in northern Moldova), exemplifies another strand of camouflaged political ambiguity. Though Usatîi claimed to be 'neither with the Russians nor the Europeans', his positioning mirrors a familiar pattern: **adopting 'neutrality' while normalising pro-Russian preferences** beneath a veneer of pragmatism. This posture dovetails with a wider effort by Russian-aligned actors to re-legitimise neutrality as a structural constraint on Moldova's foreign and security policy.

From a geopolitical perspective, **Moldova's neutrality clause** (stipulated in the Constitution) has been a significant point of contention ever since the country's independence.[30] More recently, in April 2024, the Socialist and Communist parties tabled a draft law that would have redefined neutrality to explicitly prohibit all forms of military or security cooperation with Euro-Atlantic structures. Such a move would not merely reaffirm Moldova's non-alignment but effectively institutionalise it as a buffer state – one in which Russia, already maintaining troops in separatist Transnistria, could exploit by freezing the country's strategic options and blocking deeper integration with the West, a limbo with profound regional reverberations.[31] From an operational standpoint, such a posture would also constrain Romania's role as NATO's principal staging and transit hub on the Black Sea-Danube axis, amplifying risk across a supply network that underpins both Ukraine's resilience and regional security. A constellation of narratives was subsequently deployed to stoke public fears about Moldova's involvement in the war in Ukraine, with an imminent attack upon Transnistria, discursively framed as a NATO/Western proxy.

1.2.6. Strategic recalibration or a genealogy of subversion

In 2024, the Kremlin's coordinated attempts to influence the presidential election and Constitutional referendum (enshrining EU accession as a strategic objective) were narrowly thwarted - but only temporarily so. A fragmented electoral arena and the prospect of fractional negotiations tangibly expanded Russia's opportunities for interference, otherwise commensurate with its strategic bid to reestablish control over Chisinau.

However, Russia's tactical approach has visibly adapted, and some lessons have been learnt. This recalibration reflects a broader shift within the Kremlin's power vertical. Dmitry Kozak's removal as Deputy Chief of Staff around mid-September and the rise of Sergei Kiriyenko within the ranks of the Presidential Administration have signalled a decisive turn in Russia's management of the near abroad,

---

[29] In 2024, during Romania's presidential elections, Jackson Hinkle, an American commentator, openly aligned with Russian state media, played an active role in amplifying polarising frames and has since directed similar messaging toward Moldova. Jackson Hinkle made unfounded allegations of electoral fraud and the repression of opposition leaders, otherwise a common narrative thread amplified by Russian-affiliated disinformation networks: https://x.com/jacksonhinklle/highlights. Also see coverage of Hinkle's participation at a forum held in Moscow, 'Romanian extremists Calin Georgescu and George Simion praised at pro-Russian Moscow forum' in G4Media (Romanian news outlet), available online at: https://www.g4media.ro/romanian-extremists-calin-georgescu-and-george-simion-praised-at-pro-russian-moscow-forum.html.

[30] Vladimir Socor (August 2022). 'Moldova's Bizarre Neutrality: No Obstacle to Western Security Assistance (Part One)' *in Eurasia Daily Monitor* 19(123), Jamestown Foundation. Available online at: https://jamestown.org/program/moldovas-bizarre-neutrality-no-impediment-to-western-security-assistance-part-one/.

[31] 'Russia Continues Efforts to Regain Influence Over Moldova' (September 2025) *ISW Brief*, available online at: https://understandingwar.org/research/russia-ukraine/russia-continues-efforts-to-regain-influence-over-moldova/

with profound implications for Moldova in the immediate lead-up to the vote. While Kozak seemed to court a more transactional approach, akin to elite brokerage, energy/commercial inducements, or formal architectures such as the federalisation plan for Moldova,[32] designed to play out over the long term, **Kiriyenko brings a hardened political technology edge practised both domestically and across the occupied territories of Ukraine.** Undeniably, each aimed to exert control over this small Eastern European polity. What set these strategies apart was the tempo. Grassroots-organisations' testimonies and a robust body of evidence attest to an accelerated destabilisation campaign that cuts across neighbourhood dynamics, fuelled by information warfare and proxy networks – all substantially funded.

In effect, the hybrid pressure was visibly intensified, favouring deniable levers to sway votes, polarise public opinion, and destabilise Moldova's pro-Western orientation and incumbent majority. The modus operandi combined, amongst others, coordinated influence operations sustained through illicit funding infrastructures, proxy mobilisation (*i.e.* in Russian-speaking localities) and recruitment to stir public unrest, in Moldova and abroad. Mirroring Kiriyenko's interventions in Russia and Ukraine, constellations of covertly funded NGOs and Orthodox religious networks were deployed as levers, weaving narratives of religious persecution with material incentives to cultivate support and loyalty among transnational communities of faith. In parallel, clerical hierarchies and parish priests (most subservient to the Moscow Patriarchate) were mobilised to sermonise against European integration, framing it as a spiritual threat, and to seed narratives across coordinated Telegram channels and local media.[33]

To complicate the fragile balance of prospective coalition negotiations or even contest an unfavourable outcome, Russian-affiliated networks and local proxies have incited public unrest and mobilisation in the vote's immediate aftermath. Prior orchestrations targeted the Moldovan diaspora. The manufacturing of public outrage in online spaces was correspondingly reinforced by the tactical training of saboteurs in various locations across the western Balkans - **coercive auxiliaries prepared to embed protest movements and foment violent escalation.**[34] Although at least one such network was dismantled prior to the vote, the scope of sabotage campaigns and the level of penetration cannot be fully gauged.

Setting aside the almost instant proliferation capacity of online ecosystems, the weaponisation of ecclesiastical networks echoes historically in the KGB's (well-documented) playbook, whereby Soviet front organisations such as the Christian Peace Conference (CPC) and the World Peace Council provided religious facades for influence operations abroad. Through these platforms and numerous others, Moscow cultivated its relations with the clergy, legitimised Soviet foreign policy in ecumenical forums, and penetrated international institutions under the guise of interfaith dialogue and peace activism[35] - similar to how Russian-funded activist NGOs instrumentalised claims of religious persecution within UN fora (for instance, in the Committee for Human Rights).[36]

This genealogy of subversion reminds us that what appears to be new is often deeply rooted. By failing to connect present-day influence operations with their historical precedents, much of the

[32] Anton Troianovski (August 2025). 'The Quiet Technocrat Who Enacts Putin's Ruthless Agenda' *in The New York Times*, available online at: https://www.nytimes.com/2025/08/10/world/europe/putin-russia-ukraine-war-sergei-kiriyenko.html.
[33] Mihaela Tanase, Marionela Toma (2 September 2025). 'Persecutia ortodoccsilor: Operatiune ruseasca dedicate alegerilor parlamentare din Republica Moldova' *in Context*, available online at: https://context.ro/persecutia-ortodocsilor-operatiune-ruseasca-dedicata-alegerilor-parlamentare-din-republica-moldova/
[34] 'Moldova arrests 74 over "Russian plan to incite mass riots"' (23 September 2025) in The Times, available online at: https://www.thetimes.com/world/europe/article/moldova-elections-2025-news-p97tw7wvl.
[35] A recommended read into the history of the KGB's modus operandi, Christopher Andrew, Vasili Mitrokhin (2000). *The Mitrokhin Archive. The KGB in Europe and the West* (Penguin: London). PP.: 634-5.
[36] Mihaela Tanase, Marionela Toma (2 September 2025). 'Persecutia ortodoccsilor: Operatiune ruseasca dedicate alegerilor parlamentare din Republica Moldova' *in Context*, available online at: https://context.ro/persecutia-ortodocsilor-operatiune-ruseasca-dedicata-alegerilor-parlamentare-din-republica-moldova/

scholarship on information threats risks treating symptoms in isolation while missing the structural persistence of a modus operandi – an omission that distorts both assessment and response.

1.3. The online ecosystem

1.3.1. Methodology

Data collection was conducted over a four-month period, from June 1st to September 23rd, 2025, capturing the digital information environment in the critical run-up to Moldova's parliamentary elections. Extraction and monitoring processes occurred at regular intervals across multiple online platforms, including **Telegram, TikTok, X, Facebook, VK, YouTube, and select web sources**. The collection strategy combined targeted keyword tracking with data collection based on pre-identified problematic accounts linked to known influence operations, or accounts identified with electoral blocs/parties and leaders.

The methodological approach was selective rather than exhaustive, privileging strategic relevance over total volume saturation – consistent with hybrid threat analysis and cross-platform forensics. Particular attention was paid to content that exhibited signals of strategic coordination or repetition, foreign amplification or attribution to known influence networks, as well as the narrative engineering around divisive and security-sensitive themes. The keyword framework combined three sources:

- General thematic terms such as "Moldova", "elections", "vote", and "parliament" to capture the mainstream discursive terrain.
- Specific names and hashtags associated with candidates, parties and political coalitions – including those flagged in investigative journalism or civil society reporting.
- Inductively refined terms that emerged during preliminary rounds of data collection, allowing the search corpus to evolve in tandem with the information ecosystem itself.

The overarching aim was to capture not only the content of influence operations, but also their Tactics, Techniques, and Procedures (TTPs) – **the behavioural layer**.[37] Albeit not exhaustive, these include the use of coordinated cross-posting, platform-specific manipulation strategies, disguised amplification tactics, and camouflaged affiliations (*i.e.:* pro-Russian actors posing as neutral, pro-European or *sovereignist* entities). The merged dataset was therefore curated to enable forensic inquiry into:

- The provenance and dissemination chains of strategic messaging (narratives).
- The temporal evolution of narrative clusters (i.e. anti-EU, neutralist, irredentist, anti-establishment).
- The cross-platform architecture of influence operations and coordinated disinformation campaigns.
- The deployment of known or novel TTPs in digital manipulation, including misattribution, recycled narratives, and proxy amplification.

With 5,200 entries, the cross-platform dataset integrates content metadata (timestamps, platforms, engagement metrics), actor attribution (including compromised or disinformation-linked entities), and origin tracing (where identifiable), enabling both granular and structural mapping of Moldova's pre-electoral information space. A separate TikTok dataset was parsed for analysis, in partnership with investigative journalists from the FACT EU Hub, and extracted using FactorY, an in-house AI-based software. The narratives were then processed in the **PROMPT Corpus Analyser** to identify persuasion techniques, rhetorical devices and emotional triggers associated with political discourse and patterns of memetic amplification. The insights drawn from just over 2,100 TikTok posts revealed the mechanisms behind the populist party's (PPDA's) ascendance.

[37] Inspired by and building upon the DISARM Framework: https://disarmfoundation.github.io/disarm-navigator/.

Another separate dataset focused on the Wikipedia corpus of Moldovan-related Wikipedia pages prior to the elections. Using the **PROMPT Wikipedia Sensitivity Barometer** and statistical analysis across 16 variables, we probed the extent to which Wikipedia's public knowledge ecosystem is exposed to manipulation and how. This pioneering approach advances **electoral integrity research**, by looking at where information credibility is shaped long before it reaches social platforms. By quantifying composite metrics such as manipulation and sourcing risks, as well as behavioural volatility, our analysis reveals **how coordinated editing can subtly recalibrate what counts as factual consensus within the digital public record**.

Furthermore, **to reinforce analytical reliability in detecting coordination patterns**, the datasets underwent cross-validation **using the Qriton Data Pipeline**, a purpose-built analytical environment designed to process large-scale, time-sensitive data streams.[38] For anomaly detection, we deployed **Hopfield networks.**[39] These flagged approximately **35 coordinated campaigns**[40], indicative of narrative recycling, templated amplification and message discipline among known and proxy accounts.

The proposed framework bridges the gap between traditional content analysis and AI-powered forensics, mediating a granular understanding of how electoral manipulation is architected, amplified, and legitimised across online platforms and the web.

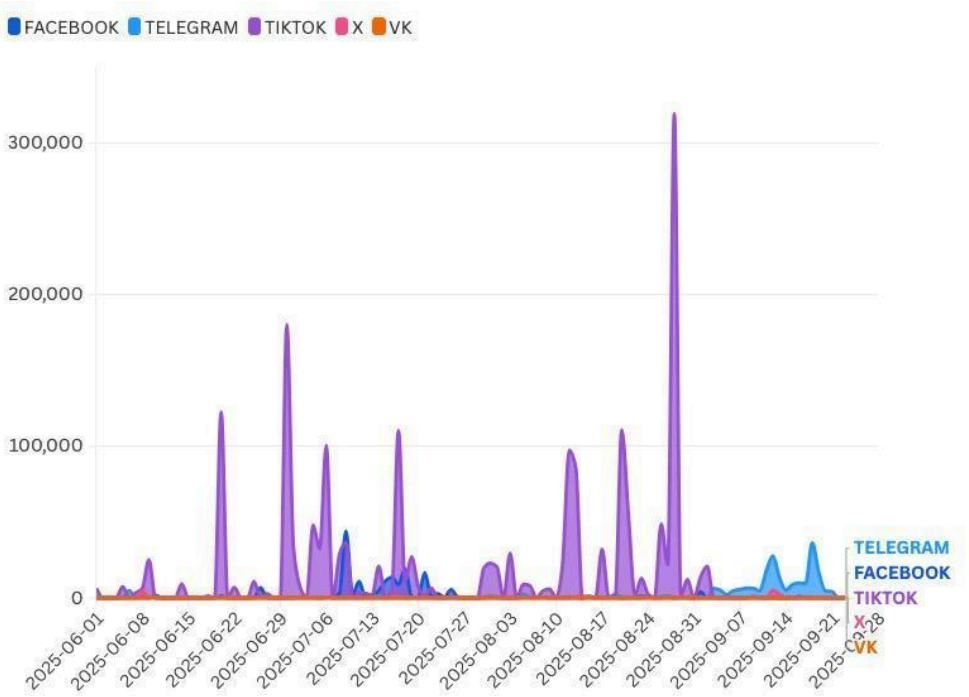
1.3.2. The scale and centre of gravity (cross-platform)

We strategically curated a slice of the broader information environment, focused on relevance, engagement and disinformation-linked activity. Within this scope, approximately 3600 posts were tagged as disinformation or coordinated influence operations, based on a range of indicators, including state-affiliation (i.e., government-controlled/sponsored media outlets), recycled narratives, origin tracing, and previously documented networks of malign actors. Telegram emerged as the central operational layer, with over 3,000 posts, more than 77% of which are attributed to disinformation-linked actors or compromised accounts. It serves as both a primary channel for initial dissemination and a redistribution vector across other platforms.

[38] Qriton's (qriton.com) infrastructure supported multiple forensic functionalities in the detection of coordinated behaviour. The Annex section includes validation statements from our cross-platforms datasets. Through the Smart Prompt Builder (an AI-driven agent workflow generator), we successfully selected data sources and automated aspects of exploratory data analysis, ensuring consistency and replicability.

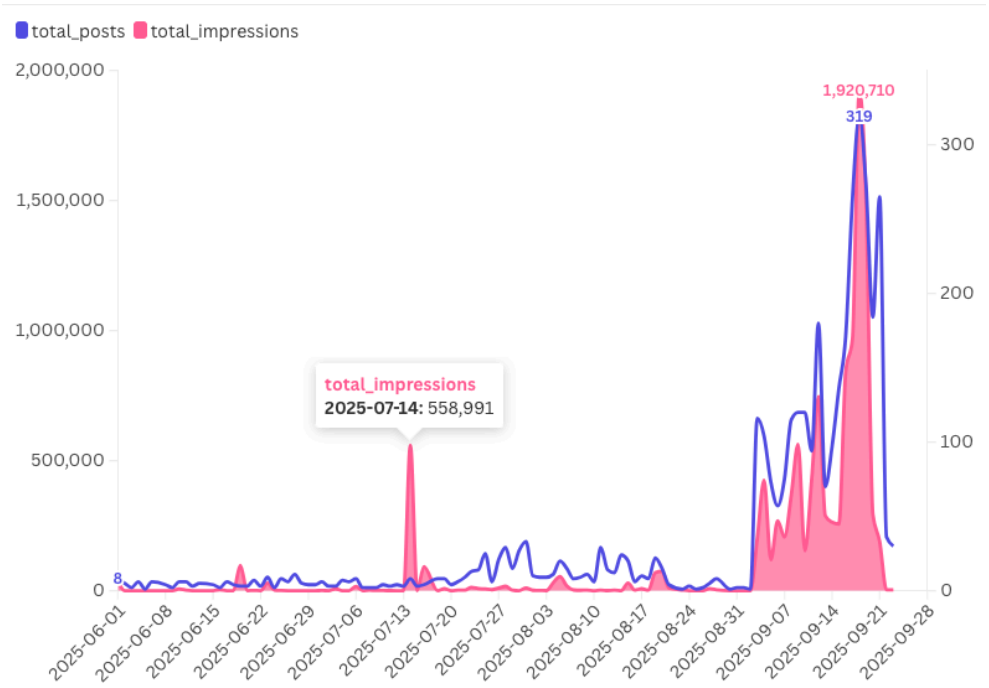
[39] These are a form of evolved neural network optimised for pattern recognition in time-evolving graphs, particularly subtle forms of coordination or repetition across fragmented digital spaces (i.e.: clustered posting behaviour, account synchronisation, and temporal validation)- Ramsauer, H., Schafl, B., Lehner, J., Seidl, P., Widrich, M., Gruber, L., Holzleitner, M., Pavlovi'c, M., Sandve, G.K., Greiff, V., Kreil, D.P., Kopp, M., Klambauer, G., Brandstetter, J., & Hochreiter, S. (2020). *Hopfield Networks is All You Need. ArXiv, abs/2008.02217* (Cornell University).

[40] In practical terms, the pipeline validated the results of earlier analytical layers, including semantic clustering (0.75 similarity threshold) and actor attribution, confirming coherence within the broader temporal structure of the dataset. The confidence threshold of 60% or higher was considered substantially significant and cross-referenced with known threat actor profiles.



While the dataset captures only a partial facet of TikTok activity, preliminary figures indicate high engagement intensity, **with over 17.8 million views and 1.85 million reactions from just 217 entries.** These trends highlight TikTok's disproportionate visibility and capacity for mobilisation. Comparatively, Telegram content amassed 13.7 million views and 229,000 reactions, with narratives often originating from high-audience Russian or proxy channels.

In the lead-up to the vote, we observe a heightened online activity, with clear signals of increased platform engagement, narrative seeding, and audience micro-targeting. Cumulatively, the **disinformation segment of the ecosystem** exceeds 11 million measured impressions (views and reactions), with activity surges clustered around late July and mid-September, coinciding with offline mobilisation attempts, decisions from electoral authorities, and/or proxy campaign escalations (see Graph below).



1.3.3. Beyond hard borders - online geographies & diasporic spaces

Where we could trace the country of origin for compromised actors, the cross-platform data sample revealed a concentration of activity in Moldova and Russia. Moldovan accounts were responsible for the highest volume of disinformation-tagged posts, while Russian-origin accounts generated disproportionately high engagement relative to their output – the latter, 559 posts and over 5,750,000 measured impressions. This suggests the use of amplification infrastructures or pre-established/captive audiences. A smaller but non-negligible share of disinformation-linked content also originated from accounts geolocated across Europe - in Romania, Germany, the United Kingdom, France, as well as the United States, etc.. These were likely tied to diaspora clusters, proxy amplification, or regionally coordinated influence assets. In a nutshell, the vast geographical footprint emphasises the transnational character of the ecosystem, which blends localised seeding with cross-border mobilisation tactics.
A distinctive feature of Russian-affiliated activity was the production and dissemination of narratives in multiple languages through channel/account spin-offs, and often with near-identical semantic structures. These spin-offs were tailored for consumption across Romanian, French, German and English-speaking audiences, to match local discursive frames.[41]

However, online geographies are often strategically misleading, and intentionally so. Attribution in digital environments is a notoriously cumbersome process, especially when disinformation/influence operations actors actively employ obfuscation tactics to mask provenance.[42] The result is a layered disinformation strategy that couples linguistic localisation with centralised message control. In order to surface operational linkages, our approach **fused geolocation analysis with behavioural, semantic and temporal forensics**.

The demographic presence of diaspora communities needs to be factored in when assessing the geographical origin and transnational propagation of manipulative content. Diaspora populations, particularly across Europe (in Romania, Italy, Germany, France, the United Kingdom, and so forth), **represent not only a consequential electoral bloc – eligible to participate via external voting mechanisms – but also a strategic vector for influence operations**. These communities were frequently targeted through platform-specific content (disseminated in multiple languages) that was framed to undermine trust in democratic participation, question its legitimacy, or exploit perceived disconnections between diaspora preferences and domestic sentiment. On the voting day, offline destabilisation tactics, including bomb threats reported at polling stations abroad, were deployed to induce fear, suppress turnout, and provoke administrative disruption.

However, the targeting of diasporic constituencies forms only one axis of a broader operational strategy. Influence campaigns concurrently seek to reshape perceptions within host European publics, subtly embedding narratives that portray Moldova's democratic processes as unstable, externally manipulated or geopolitically compromised. This convergence of digital and physical pressure points reflects a hybridised tactic: seeding doubt and polarising the Moldovan polity, while simultaneously eroding international confidence in Moldova's democratic resilience and progress towards integration. Furthermore, offline events and/or procedural decisions from the Central Electoral Commission (CEC), such as the opening of polling stations abroad, including in separatist Transnistria and Russia (for those extraterritorial constituencies), were consistently instrumentalised to sow distrust in electoral/democratic processes and incite civil unrest prior to the vote.

[41] We also identified clusters in Japanese and Arabic for regionally targeted dissemination, emanating from Russian-origin channels.
[42] Channels or accounts that appear to operate from the UK or Germany are likely controlled by actors/units elsewhere, using VPNs, spoofed metadata, or leased digital infrastructures to shield through plausible deniability. Geographical ambiguity is common among state-aligned or proxy operations, to enable even broader narrative reach while concealing operational origins. In effect, the digital topology of influence operations is not fixed by borders but instead is defined by strategic dispersion and language-based targeting to appear culturally or politically native.

1.3.4. Under the microscope – Transnistria

Between June 4th and September 23rd, 2025, approximately 263 posts referencing voting and separatist Transnistria were disseminated across five major platforms, contributed by 78 unique accounts – a disinformation campaign sustained over a four-month period. The narrative thread spanned Telegram, TikTok, X, and affiliated web domains. While 35 accounts lack identifiable geolocation metadata, the majority originated from Moldova (54), followed by the United States (37), Romania (34), Russia (27) and various European states, including Italy (25), the United Kingdom (21), and Spain (10). However, when it comes to online geographies and origin attribution, 'nothing is what it seems' either.

The data reveals a deliberate pattern of propagation, with cross-platform, multi-language seeding designed to manufacture reach and legitimacy.[43] A significant subset of posts was seeded and/or amplified by entities (accounts, channels, etc.) affiliated with Russian influence operations, such as *Rybar, Slavyangrad,* and *The Islander,* often through language-specific offshoots targeting Moldovan, Romanian, and even more prominently, global audiences.

The **recurrent narrative template framed the reduction of polling stations in Transnistria as ethnic disenfranchisement**, strategically contrasted with the expansion of diaspora voting across Western Europe. This was paired with allegations of Western, Romanian, and Ukrainian interference in Moldova's electoral processes, as well as militaristic disinformation – for instance, claims of an impending aggression towards Transnistria and Moldova, orchestrated as a NATO proxy to the war in Ukraine. Activity peaked on September 21st, cumulating 55 posts in just one day, and an hourly burst of 15 posts on September 19th (at 18:00 UTC). During the week of September 15th, 138 posts were pushed in a concentrated burst. The use of identical messaging across accounts/channels, coupled with concentrated timing, supports a pattern of synchronised dissemination.

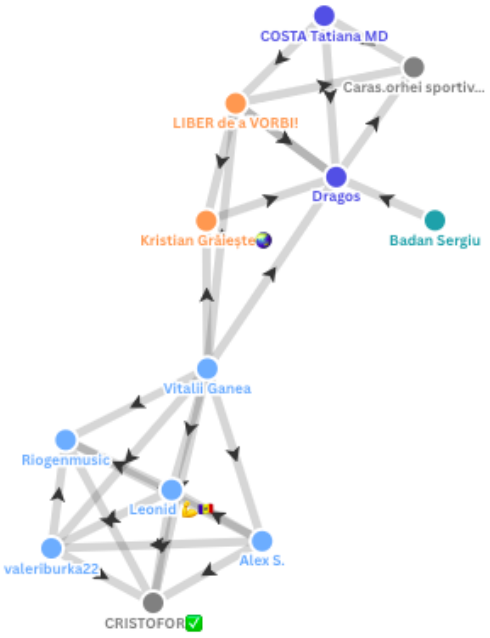
Actors such as *Rybar, the Islander* etc., seeded manipulative content which was then repeated and magnified by accounts/channels ostensibly based in the United States, United Kingdom, Australia or Spain. Russia was the origin for the majority of the content. Though listed as under the United States, '*RT and Sputnik News*' operates as an English-language extension of Russian state media. Additional clusters (i.e.: *DD Geopolitics*, *Two Majors – English Channel*, *Eurasia & Multipolarity*) recurrently amplified multiple seeding actors, indicating cross-cluster redundancy. This pattern suggests that the ecosystem is not only cross-platform and multilingual but densely interconnected, enabling reinforcement and repetition at scale. **The effect is a strategic saturation of the information environment, particularly around manipulative narratives such as voter suppression, NATO aggression, and Western/proxy (electoral or political) interference.**

For example, the original message invoking the suppression of voting rights, particularly the 'redrawing of the electoral map by PAS' (the Party of Action and Solidarity, founded by Maia Sandu), was posted by Rybar (1.3 million followers) – a well-documented disinformation actor/channel associated with military blogger placed under sanctions, Mikhail Zvinchuk, who is also tied to Russia's Ministry of Defence.[44] Public EU documents also attest to his participation in a high-level working group convened in 2022 by Vladimir Putin to coordinate Russia's mobilisation against Ukraine.[45] The channel has expanded its reach significantly, with spinoffs in multiple languages across a vast transnational geography. Zvinchuk's offline presence has accrued in recent years. Media investigations placed him in the Balkans, in Republika Srpska (Bosnia & Herzegovina), where he

[43] This is to say that even if an account appears to be based in Spain and consistently posts in Spanish, in fact, it acts as a localised spin-off of a much larger foreign influence network.
[44] https://tgstat.ru/channel/@rybar_in_english/23795
[45] Council Implementing Regulation (EU) 2023/1216 of 23 June 2023, restrictive measures in respect of actions undermining or threatening the territorial integrity, sovereignty and independence of Ukraine. Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32023R1216

discussed, in cooperation with government agencies, the opening of a media school for journalists and held training courses on the use of Telegram.[46]

A parallel amplification strand around voter suppression in Transnistria was picked up by *The Islander* Telegram channel, a well-documented disinformation outlet linked to Gerry Nolan and Chay Bowes, both Irish nationals with extensive histories in geopolitical influence operations, notwithstanding direct affiliations with Russian state media (i.e. RT and Sputnik).[47] Active across Telegram and X, *The Islander* functions as a narrative laundering node, repackaging Russian-origin messaging into Anglophone spaces. Posts referencing Moldova's electoral process amassed over 450,000 views across platforms, with targeted messaging in English. Mimicking spontaneous dissent, the same tactic of narrative laundering and scripted mobilisation extended to Moldova's domestic information space via TikTok. Some of the most reactive content amplifying the voter suppression narrative emerged from a coordinated TikTok cluster of just four accounts.[48] Each circulated an identical video post, denouncing the reduction of polling stations in Transnistria, with the messaging linguistically tailored for a Moldovan audience.[49] The amplification pattern suggests a deliberate attempt to manufacture virality while mimicking spontaneous, grassroots dissent (*astroturfing*). The accounts are linked to Tatiana Costachi, a Moldovan propagandist who reiterates distinctly pro-Russian narratives, often camouflaged in an ethno-nationalist, profoundly anti-Western, and anti-establishment veneer.[50]



Furthermore, we identified two adjacent clusters (see left graph) with similar architecture but distinct narrative frames: one centred on **anti-EU mobilisation**, and another on **diaspora scapegoating.** Both are linked to the amplification network associated with Tatiana Costachi.

The anti-EU cluster comprises 18 posts from 12 accounts, generating over 1.37 million impressions, while the diaspora scapegoating cluster includes 23 posts from 11 accounts, reaching approximately 1.88 million impressions. Despite their thematic divergence, both clusters replicate the dissemination pattern observed in earlier content: high semantic uniformity, identical/copy-pasted or minimally altered scripts, released in tight temporal windows (within 24-48 hours).

Messaging frames the EU as an exploitative colonial threat undermining Moldova's economy and sovereignty, while

[46] Irvan Pekmez (November 2024). 'Putin's Messenger: Russia's Rybar to Open Media "School" in Bosnia's Serb Entity' in Detektor Media, available online at: https://detektor.ba/2024/11/05/ruski-ribar-i-vlasti-republike-srpske-pokrecu-medijsku-skolu-propagande/?lang=en

[47] 'Prominent Irish Blogger Amplifies Kremlin-Aligned Claims About NATO Expansion' in Disinfowatch, available online at: https://disinfowatch.org/disinfo/prominent-irish-blogger-amplifies-kremlin-aligned-claims-about-nato-expansion/

[48] Dragos, Caras.orhei sportiv 7777, LIBER de a VORBI!!, and COSTA Tatiana MD. Together, the posts amassed 161,000 impressions, an unusually high volume for content that exhibited no narrative or visual variation.

[49] At the centre of amplification is this message: "Asociatia Juristilor din Republica Moldova si-au pronuntat ingrijorarea si condamna ferm decizia de a pune mai putine sectii de vot in regiunea transnistreana. (...) Este antidemocratic. Incalca drepturile omului. Guvernarea PaS, incompetenta, nu permite votul celor din Rusia si impune obstacole pentru diaspora din Europa. Prin observatorii controlati de ei, vor face ce vor cu voturile. Oare aceste alegeri vor fi corecte sau din nou fraudate? (...)". Translation: The Association of Jurists of the Republic of Moldova has expressed concern and strongly condemns the decision to reduce the polling stations in Transnistria. (...) This is anti-democratic. It violates human rights. The current PaS government, incompetent, does not allow those in Russia to vote and imposes obstacles for the diaspora in Europe. Through their controlled observers, they will do what they want with the votes. Will these elections be fair, or once again fraudulent?

[50] https://stopfals.md/ro/article/profil-de-propagandist-tatiana-costachi-costa-moldovanca-teze-sovietice-de-statalism-antiromanesc-falsuri-despre-nato-romanizarea-militarizarea-si-atragerea-r-moldova-in-razboi-181000

simultaneously mobilising viewers to engage in economic nationalism (i.e. 'Buy only local products to escape European colonists'). Meanwhile, the diaspora cluster redirects frustration inward, portraying Moldovans abroad as a parasitic force, responsible for inflating real estate prices, manipulating elections through absentee ballots, or abandoning Moldova while abusing its services. Such inflammatory rhetoric exploits economic anxiety, generational divides, and post-Soviet identity fractures to deepen resentment and sow division between citizens at home and abroad. Despite thematic variation, both clusters converge on a unifying logic of betrayal, reinforcing a wider disinformation ecosystem rooted in institutional distrust, anti-Western sentiment, and identity-driven polarisation. Despite being operated by relatively low-follower, low-engagement accounts, the two clusters' coordinated structure enabled disproportionate visibility, pushing total reach above 3.2 million impressions. A similar tactic could be observed in a parallel TikTok subset associated with Vasile Costiuc, leader of the populist PPDA party, suggesting the strategic use of TikTok as an amplification engine for (emotionally triggered) mobilisation.

1.3.5. Under the microscope – PPDA and Vasile Costiuc

Vasile Costiuc, president of the *Democratia Acasa* (Democracy at Home) political platform (PPDA), boosted his profile visibility through a network of affiliated TikTok accounts, spreading manipulative content and outright disinformation. The TikTok dataset, consisting of 2,171 entries, was extracted and parsed through the AI-based Factory software deployed by the investigative outlet Context.ro (part of the EU FACT Hub). The rhetorical analysis was derived using the PROMPT Corpus Analyser.

The PPDA's strategy simultaneously targeted Moldova, Romania, and Ukraine through vast, transnationally spanning disinformation networks and coordinated influence operations. The cross-border spillovers were evident. George Simion, the far-right leader of AUR (a Romanian political party with spinoffs in Moldova), is a vocal supporter of Costiuc and his political platform. The rhetorical arsenal follows similar if not identical patterns across an identifiable repertoire of topics/themes: from the grievances of local farmers and producers, to asserting sovereignty and 'taking back control' of the government, economy, whilst dismantling the 'illegitimate, corrupt establishment'.

The targets of this messaging (emanating from Costiuc and affiliates) extend beyond the political sphere, into civil society, media, and other domains (prompting libels against his network from reputed civic activists, journalists, etc.). In fact, this is a crucial conceptual shift, as it demonstrates how **propagandists (similar to Costiuc) are not merely competing in electoral politics but are actively working to reshape an entire ecosystem of public discourse** – essentially redefining who is trustworthy, which narratives are legitimate, and which forms of civic participation are acceptable. The discourse, therefore, moves beyond winning parliamentary elections, into a socio-political purge and change of paradigm.

Coordination & amplification patterns

There are approximately 337 duplicate entries or identical content repeats across multiple TikTok accounts (2171 total entries in the dataset). These are not merely occasional overlaps; they indicate a systematic practice of cross-posting similar or identical scripts. In practice, the accounts may be either centrally managed or follow a coordinated distribution pipeline. At least 11 TikTok accounts recycled their own content heavily, with nearly 200 unique transcripts reposted, adding up to 274 redundant pushes. Even without cross-account coordination, single accounts try to game TikTok's algorithm by reposting the same script multiple times – a tactic of content flooding.

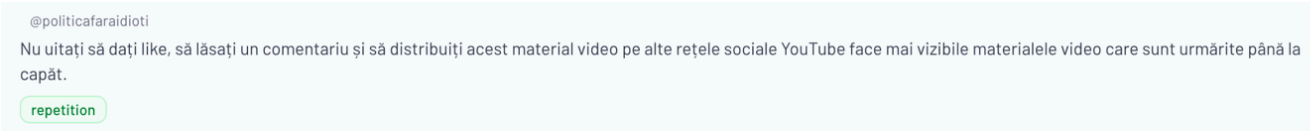
Furthermore, some account pairs share dozens of transcripts, forming the 'spine' of the network, the hubs that recycle narratives most aggressively. In many cases, the same accounts reappear across multiple strong edges (connections), suggesting a core cluster of operators tied together by repeated scripts. Two distinct coordination patterns emerge:

- **Twin accounts**: pairs of accounts that consistently post almost identical content, suggesting they are controlled by the same operator or team. These are essentially 'mirrors' of each other, used to multiply the visibility of identical messages.
- **Cluster accounts**: larger groups where each account is strongly linked to multiple others through shared scripts. This dynamic creates a dense web of cross posting that gives the impression of a grassroots swarm and organic support, while in reality, it is a tightly managed cluster.

Thus, the network expands by replicating the same narratives across a web of interconnected accounts. The most influential ties reveal which accounts are moving in unison, demonstrating that what appears to be a chorus of voices is, in reality, a centrally controlled echo-chamber. Intriguingly, the **narratives and rhetorical devices mirror deep emotional appeals**: the struggles of local farmers and producers, localised economic grievances, victimhood and persecution, and so forth. Many of the repetitive scripts also centre on personal tragedies – a husband who lost his toes and survives on a modest disability pension, parents unable to afford surgeries, a family evicted from their home – aiming to attract a vast (outraged) audience and trigger emotional responses. Also, the Moldovan diaspora and its plight 'far from home' is heavily referenced across repetitive scripts.

Other posts deploy symbols of everyday life and rural identity, asserting classic nativist tropes: references to local grapes, honey, pears and schoolchildren eating local fruit, contrasted with foreign-imported bananas. Together, these narratives produce a carefully calibrated sense of victimhood and betrayal, which demands regime change and the mass mobilisation against the corrupt government.

A significant feature of this network is the direct mobilisation of users as amplifiers. The example below, processed through the PROMPT social data analysis tool, shows how audiences are instructed to act: to repost content, to flood TikTok with new accounts, to share clips widely, to overwhelm perceived opponents online. In this particular content, amplification is framed not just as engagement, but as a form of political struggle.



Picture 1 – TikTok post (by one of the affiliated accounts) mobilising audiences to engage, share, etc (PROMPT Corpus Analyser)

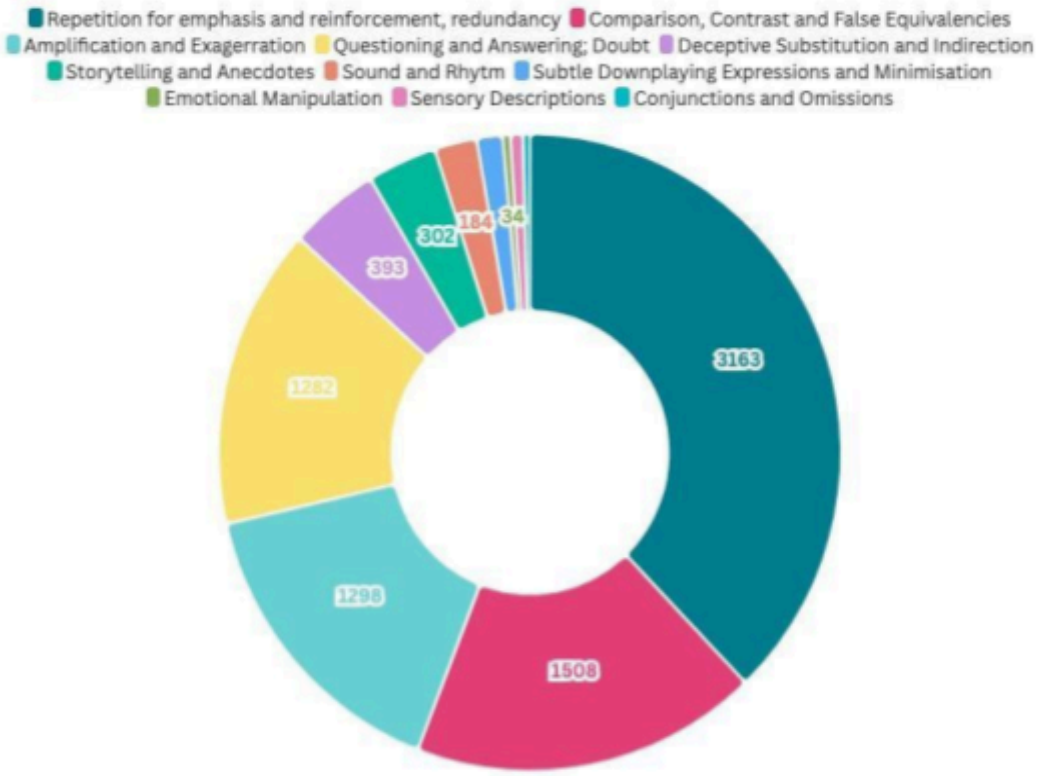
These calls to action are reinforced by a steady stream of posts claiming persecution and censorship of the *Democratia Acasa* political party, of Vasile Costiuc and his affiliated networks. TikTok suspensions, alleged state surveillance or blocked content are portrayed as proof of the ruling government's desperation – otherwise a ubiquitous trope used by wider disinformation networks – and election rigging. Alongside mobilisation appeals, the network repeatedly targets civil society organisations and independent media. Watchdog groups are cast as corrupt agents of foreign interests (i.e. Soros), allegedly funded to protect the regime. Journalists and NGOs are lumped together with the government as part of a 'system' that systematically ignores the people's suffering. This rhetorical strategy delegitimises independent oversight and positions civil society as a collective enemy – collapsing all institutional counterweights into a single hostile bloc. Similar patterns and rhetorical devices are noticeable in Romania, featuring prominently in the far-right discourse.

We also identified 8587 rhetorical devices, their distribution consistent with wider coordination and amplification patterns, particularly the use of repetitions and redundant constructs for emphasis and reinforcement. The PROMPT Corpus Analyser also detected over 1390 persuasion techniques (see examples in the Figure below) and rhetorical devices present in the TikTok corpus, amongst the most frequent:

Exaggeration & hyperbole	Inflating threats or hardships (eg: portraying Moldova as ruined) to magnify urgency
Scapegoating	Assigning blame to government, civil society, or external partners as the singular cause of complex problems.
Emotional exploitation	Centering stories on deprivation, sick children, evicted families, or impoverished pensioners to provoke outrage and sympathy, then linking emotion to political action
Delegitimisation	Framing instituions, NGOs, watchdogs, media as corrupt or foreign-controlled, thereby stripping them of credibility.
Conspiracy framing	Claims of censorship, persecution, or secret financial schemes, which portray the political bloc as both victim and heroic resistance.

Figure 1 – Examples of persuasion techniques taken from the PROMPT Corpus Analyser Codebook

The use of rhetorical devices shows how style reinforces substance, not incidentally, but converging with persuasion techniques to maximise memorability, emotional impact, and the viewers' engagement or even political mobilisation.



*Graph SEQ Graph * ARABIC 3 – chart depicts the distribution of rhetorical figures by number of occurrences. Processed by authors from PROMPT Corpus Analyser*

In a nutshell, the constellations of rhetorical devices function together as amplifiers of persuasion. Emotional storytelling and sensory descriptions make hardship vivid and relatable, while contrasts and false equivalencies reduce complex (lived) realities to stark binaries. Repetition and redundancy, featuring most prominently, reinforce the messaging until it becomes self-evident, while rhetorical questioning simulates dialogue, guiding audiences (usually) towards pre-set conclusions.

Building on the cross-platform dataset, Vasile Costiuc, the leader of Democratia Acasa political bloc, also featured in the Russian-spun Pravda network and Romanian language affiliates, at least 7 times

between July and August 2025. These web outlets, part of a much wider disinformation and propaganda ecosystem, amplified the same narratives of victimhood and persecution circulating on TikTok (through the network of coordinated accounts), portraying Costiuc as being silenced, harassed, or marginalised by politically complicit state institutions and hostile media. The Sputnik Telegram channel follows the same pattern. In fact, **the Pravda network repeatedly cites as sources Telegram channels that are themselves notorious vectors of disinformation and influence operations**: *Triunghiul Basarbean*, *Sputnik necenzurat*, *Gagauz News* etc. By layering these citations, this multiplatform propaganda network produces an illusion of corroboration, transitioning from Telegram, TikTok, and onto the web, while masking central coordination.

As we examined in Costiuc's and PPDA's TikTok case, the significance of templated amplification patterns extends beyond mere repetition. From an operational standpoint, it serves as **a force multiplier for disinformation and information manipulation by manufacturing an illusory consensus**. When identical or near-identical framings appear synchronously across channels, languages and domains, they simulate organic public outrage, thereby coercing undecided audiences into perceiving certain narratives as dominant or even inevitable. In fragile electoral environments such as Moldova or Romania, trust in institutions is fragmented at best, and political/ideological affiliations are fluid and often camouflaged. The appearance of 'multiple independent sources' repeating similar headlines (see collage above, of Pravda-linked domains), enables unverified claims to cross the threshold from speculation to perceived fact – even when they trace back to Telegram clusters spreading disinformation.

In this sense, **templated amplification** functions not merely as propaganda, but **as cognitive infrastructures**, shaping how events are interpreted, which actors are trusted, and which features are deemed plausible. It also provides a measure of operational efficiency for hostile actors: **once a narrative template proves effective, it is redeployed with new variables (different political candidates, electoral settings, countries, crises and so forth), thereby reducing the cost of influence campaigns while expanding their lifespan.**

1.3.6. Regional targeting, semantic geographies and templated amplification

The online ecosystem revealed not an isolated country-based interference effort, but a regionally integrated information strategy simultaneously targeting Moldova, Romania, Ukraine, as well as the EU/NATO, and other European countries, fused into a single geopolitical battlespace (albeit online). Elections have been systematically framed from procedural democratic events into externally orchestrated power contests, frequently referenced alongside Romania's 2024-25 presidential ballot, creating a manufactured sense of electoral interdependence or even shared illegitimacy. Conversely, allegations of Ukrainian and Western interference also abounded. The claims of external meddling were paired with sustained attacks on (diaspora) voting processes, depicted as manipulated, externally controlled, or irrelevant.

A prominent tactic involved framing **Romania's impending austerity crisis as a direct cost of supporting Moldova and Ukraine** (i.e. through energy exports, military/humanitarian aid, in-country refugee assistance), depicting solidarity as self-inflicted harm. This was particularly potent across Romanian-language channels and web platforms associated with the far-right, where supposed crisis narratives are wrapped in anti-European/anti-establishment and conspiratorial rhetoric.

Clustering approach and narrative lifecycles

The cross-platform dataset covers 693 semantic clusters and approximately 2300 disinformation posts.[51] The clusters collectively generated 112 million impressions and involved an average of 8-9

[51] Each cluster aggregates identical or similar posts across languages (Romanian, Russian, English, French, Italian, even Japanese, etc.) and records the number of posts, impressions, duration and up to three narrative categories (the latter, processed manually).

unique actors (median=3) per semantic cluster. Most activity unfolded within tight/synchronous amplification cells, with a small subset showing high actor (account) dispersion (50-120 contributors each) – a pattern indicative of coordinated mass-push moments.

Taken together, the observed amplification modes, ranging from synchronous bursts to strategically reactivated baselines, demonstrate that influence operations around Moldova's elections were not merely reactive or opportunistic but structured to maintain a long-term narrative scaffolding across borders and languages.

One of the main narratives identified by PROMPT - **election interference and voter suppression** - is the most frequent narrative (main) observed in the context of these elections.[52] Nearly half of all clusters sought to neutralise/invalidate the election before it occurred. This narrative took the form of four dominant claims:

- **Administrative abuse and censorship** – *Opposition repression/censorship* (289 instances/posts) was the single most frequent subcategory. The closure of some media outlets and CEC decisions banning participation in the ballot were routinely labelled as authoritarian, framing opposition blocs as 'political dissidents', dissenters or anti-establishment resistance. As expected, pro-Russian parties, including populist platforms, and their leaders were the most amplified across all platforms and the web. The arrest of Evgenya Gutsul (Shor's associate and former governor of autonomous Gagauzia) prompted accusations of undue process and human rights infringement across a vast disinformation ecosystem - Russian-affiliated with spin-offs in multiple countries and languages.
- **Voter disenfranchisement and diaspora instrumentalisation** – roughly 260 posts accused the government of voter disenfranchisement, particularly referencing Transnistria, as well as Gagauzia, in conjunction with appeals to defend the identity of these 'long-ignored' provinces, intentionally discriminated against for their anti-European/pro-Russian orientation, language or ethnicity. There were also numerous conjoined references to the Moldovan diaspora residing in Russia and the low number of polling stations, in contrast to the diaspora located in Western Europe. Often, the defence of Orthodox/traditional identity (Identity and Sovereignty main_category) would be included in the narrative mix, building on allegations of Orthodox persecution, which, otherwise, was a thematic mainstay.
- **Fraud & manipulation** – 141 posts alleged pre-planned ballot stuffing, postal vote rigging, or 'white vans full of bribed voters', without providing evidence, but instead relying on repetition. These claims would be paired with pre-emptive mobilisation, incitement to protest and civil unrest, and allegations of government orchestrated repression ('Violence & chaos' narrative category).
- **Institutional capture & foreign meddling** – 114 posts portrayed electoral authorities/governmental institutions, law enforcement, media, and civil society as captive, subservient to the incumbent party [PAS], while denouncing a plethora of injustices against the opposition. Over 40 posts reversed allegations of foreign interference, from Romania, Ukraine, the EU and other European countries (*i.e.:* France, the United Kingdom).

Temporal analysis shows that over 80% of these narratives appeared as high-intensity bursts (<12h windows, with a significant subset reinforced over time.[53]

[52] 311 unique clusters and 334 appearances (in the case of hybrid messages where multiple categories and narrative subcategories could be present)

[53] One of the most active clusters (cluster_id 3) textually repeating the allegations of institutional capture (within the electoral interference & voter suppression category) involved 12 unique accounts and 13 disinformation posts within a span of ~1.5 days (between 2025-07-31 07:17:52 and 2025-08-01 18:51:14, UTC-standardised, with a duration of roughly 35 hours). One of the sample_texts within the cluster: ['📝Sandu will decide for everyone📝\n\nThe President of Moldova, Maia Sandu held a meeting of the Supreme Security Council (SSC) to discuss the "unprecedented attempt of interference in the parliamentary elections" by Russia.\n\nSandu claims that Russia allegedly plans to use political parties, financial

Narratives on the loss of **identity and sovereignty, particularly at the hands of the EU,** spans 183 unique semantic clusters and a total of 189 appearances (for clusters combining two or more categories). The dominant claims crystallised around:

- Swaying **geopolitical orientation and public opinion towards Russia** (with approximately 122 posts), invoking traditional (at times, fraternal) linkages, while urging the defence of religious (Orthodox)/traditional identity (with 37 contributions). Numerous such instances prefigure a return to the status quo ante and Moldova's participation in Russian-led structures (*i.e.:* the Eurasian Economic Union, Community of Independent States (CIS) or BRICS). Pro-Russian parties and leaders are often quoted as intent on reestablishing Moldova's neutrality (*i.e.*: excluding the European clause from the Constitution).
- Inducing the **perception of loss of sovereignty and ethno-cultural identity**, through territorial annexation by Romania, portrayed as an existential threat with 'imperialist' ambitions, or by Western supranational structures seeking to nullify national identities through absorption (around 44 posts).
- Messages claiming the **disintegration of Moldova under pro-EU leadership** often plugged into fears of sovereignty/identity loss, and the marginalisation of language & ethnic communities, particularly those espousing traditional/religious values (25 posts). Anti-LGBTQI posturing was also present, with concurrent streams amplified from Romania and Moldova (example, below). Certain posts, albeit not yet categorised, invoked historical revisionism in various forms, particularly with references to the Great Patriotic War and how it is being erased from the national/collective memory (the 9th of May appropriated as a European celebration).

Temporal behaviour in this narrative category exhibited a hybrid amplification structure. Approximately 78% of semantic clusters appeared as synchronous short-bursts (<12h), suggesting rapid response messaging tied to specific (media) hooks.[54]A smaller share (roughly 5-6%) persisted across several weeks, as a baseline substrate reactivated throughout the campaign.

Functionally, narratives often mingled and reinforced one another: identity and sovereignty' frames were most often appended to the 'Election interference & voter suppression category', reinforcing the notion that a rigged election was not merely illegitimate, but part of a broader and more intrusive civilisational plot.

Similar to the narratives circulating on Ukraine's government, stories about **external influence and occupation** are present across 158 semantic clusters (with 159 appearances), and includes:

- **Proxy war framing** (324 posts) to stoke public anxiety regarding an impending attack orchestrated by NATO (most commonly), but also the EU and other Western powers. Posts insinuated Moldova was being prepared as 'the next Ukraine' with Romanian/Western involvement, casting the country as a sacrificial buffer. NATO exercises were referenced in conjunction with Transnistria as a theatre for staged provocations. A notable subset also alleged Moldova's active participation in the war in Ukraine, through covert special forces units.
- **Foreign puppet-master framing** spanned approximately 85 posts and sought to nullify the incumbent government's agency, but most forcefully the President's, portrayed as a foreign pawn, 'Romanian agent' sponsored by the 'hypocritical West/Europe' to plunge the country into a regime of occupation, also equated with the EU integration trajectory.

instruments, propaganda, and other methods to influence the electoral process….', originally seeded by the Russian-associated Rybar Telegram channel, then distributed across a vast eco-system.

[54] For instance, declarations from Russian officials, interviews with Moldovan leaders, published in Russian state media, etc.

Temporal behaviour within this category was more durable than average. While roughly 70% of clusters followed rapid burst patterns[55], a small set of high-dispersion clusters persisted for weeks. The proxy war scenario was concentrated in the period closer to the vote, around mid-September.

Narratives related to **the economic and social crisis** category appear across 62 semantic clusters, with a distinct subset originating in Romania (5-6 clusters), from segments of the populist far-right. The messaging crystallises around:

- **Economic collapse & hardships, as well as a lack of public support** for the government and domestic (pro-EU) reforms, with 141 posts. A narrative subcategory includes alleged punitive measures that a pro-EU government might take in response (64 posts) – for instance, mandatory conscription for women, heightened taxes, etc. Content mostly focuses on inflation, energy prices, food shortages, corruption, and alleged economic mismanagement, with few blaming Ukrainian refugees for the economic rifts impacting Moldova.[56]
- The **austerity crisis affecting Romania, with blame shifted onto Moldova and Ukraine** (for energy exports, military/humanitarian aid, etc.), highlights how Romania was selectively deployed as a cautionary tale against Moldova's integration path. Although the number of posts (around 45 in total) may seem negligible within the overall narrative architecture, their persistence is among the longest in the dataset, cyclically resurfacing beyond two-week periods. In this case, Romanian audiences were primarily targeted in a bid to convert economic anxiety into geopolitical resentment. However, the same discourse was mirrored in Moldova in relation to energy prices and the cost of diversification projects.

Over 250 semantic clusters were assigned two or more narrative categories, commensurate with the hybrid nature of the messaging. The pairing patterns are quite evocative. So far, the highest frequency (based on co-occurrences in the 3 main narrative categories) emerged between 'Election interference & voter suppression' and 'Identity & sovereignty', where the procedural delegitimisation of the vote was reinforced as an existential/civilisational betrayal.
To a lesser extent, another high-frequency combination included 'Election interference' and 'External influence & occupation' categories, reframing domestic actors as foreign agents staging a controlled takeover, or blending territorial paranoia with the imminence of war. Interestingly, a smaller but consistent category 'Violence & chaos' appends to other narratives to escalate emotional cues: to incite protests or, conversely, depict protests as widespread disaffection (mostly staged from abroad, such as the one in Russia in response to the low number of polling stations), and in general, to foment pre-emptive destabilisation, with allegations of violent repression of civil unrest.

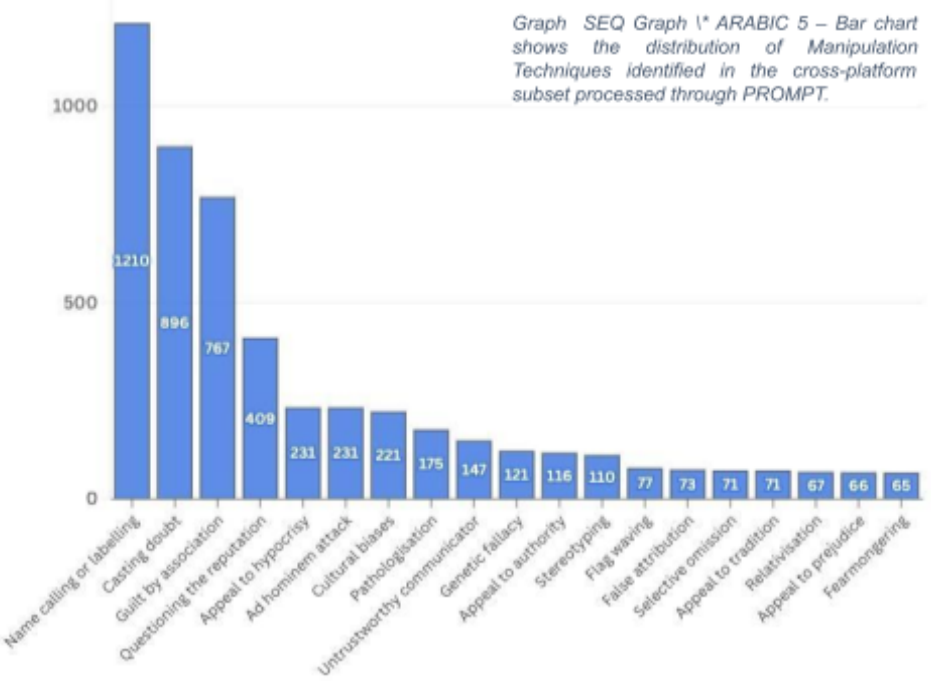
1.3.7. Manipulation through rhetorical devices

To understand how these narratives and claims are made persuasive, we further processed a segment of the cross-platform corpus (approximately 2349 entries) in the PROMPT Corpus Analyser, which detected a total of 6823 persuasion techniques and 3671 rhetorical figures embedded in the discourse. An important aspect, **posts usually layer multiple manipulation techniques and rhetorical devices, at times in a single sentence.**

Across the coded sample, the most recurrent persuasion techniques were *name-calling/labelling* (in roughly 1210 instances), casting doubt (896), and *guilt by association* with 767 instances, respectively – all of which function to erode trust in institutional actors/leaders while pre-emptively discrediting alternative viewpoints.

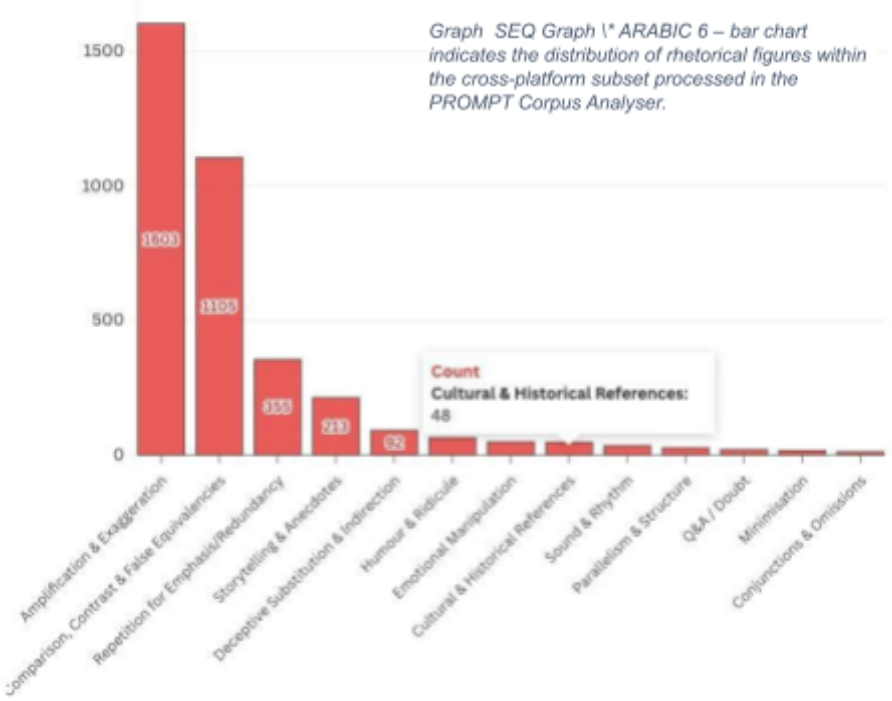
[55] A cluster alleging *'Moldova's Suicide Pact: The Coming War for Transnistria (...)'* was repeated in tightly coordinated bursts (through Telegram and the Pravda web affiliates) in at least 30 posts (all disinformation-linked) by 14 unique accounts, spanning ~1.2 days (from *2025-09-20 11:52:18* to *2025-09-21 16:45:28* – UTC standardised).
[56] A smaller subset presented the results of unsubstantiated election polls, showing the opposition in the lead, or alleging the lack of public support for government reforms.

These techniques operate less by argument than by positioning: targeted political figures are reduced to hostile archetypes ('globalists', 'foreign puppets', 'traitors', etc.), and communities tarnished through associative blame (for instance, civil societies portrayed as captive to incumbent power, or entirely deprived of agency). Casting doubt functions more subtly, often through grammatical ambiguity and certain rhetorical cues.



*Graph SEQ Graph * ARABIC 5 – Bar chart shows the distribution of Manipulation Techniques identified in the cross-platform subset processed through PROMPT.*

In parallel, the dominant groups of rhetorical devices relate to tactics of *amplification and exaggeration* (1603 instances), followed by *false equivalencies* (1105), *repetition/redundancy* (355), and *anecdotal storytelling* (213). The repertoire of stylistic devices served as a force multiplier, intensifying emotional resonance while lending speculative claims a sense of inevitability.
Exaggeration inflates procedural disputes into civilisational collapse, with anecdotal inserts (such as those framing a proxy war) presented as systemic proof. Together, such elements produce messages that feel persuasive irrespective of empirical substantiation.



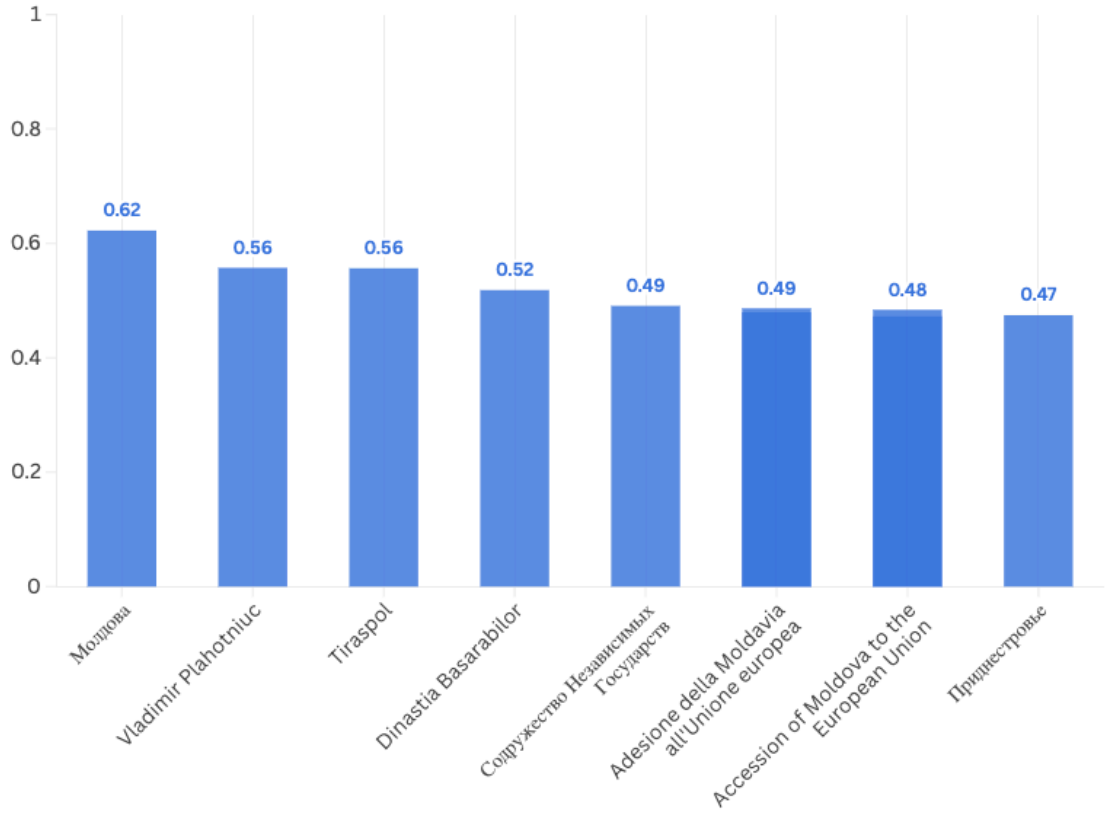
*Graph SEQ Graph * ARABIC 6 – bar chart indicates the distribution of rhetorical figures within the cross-platform subset processed in the PROMPT Corpus Analyser.*

1.3.8. The Wikipedia Sensitivity Moldova Barometer and composite risks

The Moldova Wikipedia Sensitivity Barometer offers a unique opportunity in the empirical study of Wikipedia as a vector for information warfare and epistemic instability, with a focus on contextualised, election-related narratives. It has been commonplace to assume that Wikipedia acts as a neutral, crowd-sourced knowledge environment; however, various reports have shown that it is increasingly targeted by coordinated editing efforts, both overt and covert.[57] The Moldovan Barometer dataset offers a quantitative, multi-factor lens in assessing the health and integrity of articles related to Moldova across multiple language editions. Particularly valuable, the convergence of editorial behaviour, sourcing quality, and attention dynamics (typically studied in isolation) across 16 distinct features enables the modelling of multi-dimensional risk vectors for disinformation.

To this end, three core composite metrics were developed using normalised indicators to summarise various classes of risk:

The Manipulation Risk Score (MRS): captures potential coordinated editing behaviour – *sockpuppets*, *edit spikes*, *view spikes*, *edits revert probability*, *anonymity*, *contributor add/delete ratio*.[58] The MRS could be deployed to identify pages with unusual or aggressive editing patterns that may point to manipulation attempts.



Disinformation actors do not necessarily engage in 'open vandalism', instead, they operate through patterned editorial behaviours in an attempt to simulate organic/grassroots participation (similar to other environments).

[57] Marco Silva (November 2021). 'Climate change: Conspiracy theories found on foreign-language Wikipedia' in BBC News, available online at: https://www.bbc.co.uk/news/technology-59325128. See also, García-Méndez, S., Leal, F., Malheiro, B., & Burguillo, J. C. (2025). *Identification and explanation of disinformation in Wiki data streams. Integrated Computer-Aided Engineering*. https://doi.org/10.1177/10692509241306580

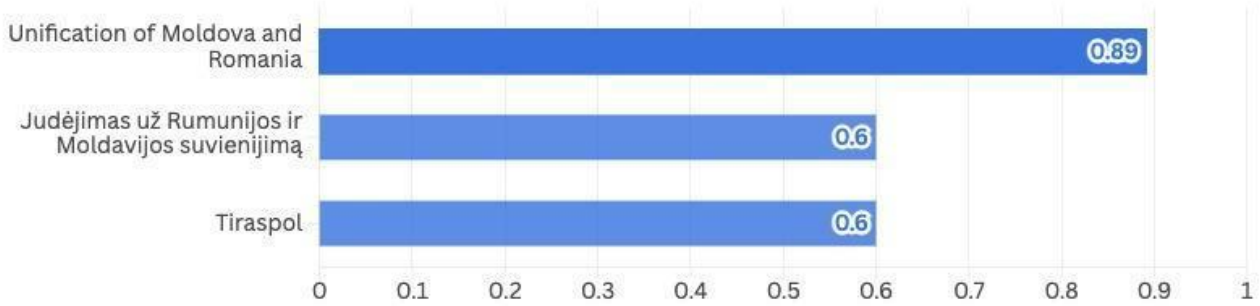
[58] Indicators were scaled using MinMax normalisation mapping all values between 0 and 1 for uniform weighing. Normalised value = (X – min(X)) / (max(X) – min(X)), where X represents the raw feature value per article/entry.

Preliminary findings: top-ranked pages on the MRS include entries on 'Молдова' (Moldova, Russian entry), Vladimir Plahotniuc (Romanian), Tiraspol (English page), Dinastia Basarbenilor (Romanian page). These pages, with politically sensitive content or historically/geopolitically contested content witnessed intense editing by both registered and anonymous users. Coupled with sockpuppet flags, this pattern *may*[59] suggest strategic manipulation.

Upon closer inspection of the Romanian-language page associated with the Basarab Dynasty, editorial patterns reveal **the insertion of claims tied to unsubstantiated or entirely absent historical sources.** Multiple additions reference disputed dynastic origins with 'citation needed' tags left unaddressed. **Such behaviour is consistent with revisionist strategies**, whereby history is selectively reinterpreted to legitimise current political aims, especially claims surrounding the ethno-historical roots of Moldova.

Similarly, a review of the English-language Wikipedia entry on Moldova's Accession to the EU exhibits narrative shifts linking EU integration with the prospect of unification with Romania, while mentioning fears and lack of public support in autonomous Gagauzia toward EU accession. Although unification is not part of the accession negotiation framework, the article includes statements that advance this possibility as a direct outcome or concern, reframing an otherwise technical process into a national identity issue. These edits also reflect a core pattern in the interference campaign levelled against Moldova, premised on stoking regional division, while amplifying ethno-national anxieties. Moreover, the page references an opinion article from a news/media outlet (*Globe Banner*), that masquerades as an independent journalistic source, although it lacks editorial transparency, verifiable authorship, or a track record of legitimate reporting – hallmarks of inauthentic or fabricated media designed to lend false legitimacy to narrative framing.[60]

The issue of reliable sources is elaborated further in the **Sourcing Risk Score** (SRS), which evaluates epistemic integrity using *citations gaps*, *suspicious sources*, and *source concentration*.[61] Higher values may indicate precarious sourcing and potential narrative fragility.Wikipedia's sourcing model makes it vulnerable to plausible sounding claims backed by unreliable or cherry-picked sources, especially for regional topics with limited journalistic or academic coverage.



Preliminary findings: articles with the highest SRS scores showed not only a lack of reliable sourcing but also complete citation voids (predominantly, for historical and ethnic identity topics).

[59] "May" is an important nuance. For example, the intense editorial activity associated with Vladimir Plahotniuc could be attributed to developments in his extradition from Greece, which garnered significant public scrutiny. In normalised terms, a score of 0.62, as seen for the Russian entry on Moldova, reflects an elevated behavioural risk profile, pointing to intense or irregular editorial activity, which may be consistent with coordinated narrative shaping.

[60] Thomas Sparrow (March 2025). 'Fact check: How to spot fabricated news reports' in Deutsche Welle, available online at: https://www.dw.com/en/fact-check-how-to-spot-fabricated-news-reports/a-71992819.

[61] Citations gaps - weighted 40%, suspicious sources - weighted 40%, and source concentration - inverse, weighed 20%. The weighing attribution reflects the relative importance of lacking citations and use of unverified sources in propagating disinformation.

The English-language entry on the 'Unification of Moldova and Romania' (0.89), indicates persistent citation gaps, lack of source diversity, and/or reliance on potentially problematic references. Although the bar chart aggregates SRS at entry/article level, in fact, the high score was achieved over several months, spanning January-August 2025. The temporal repetition is analytically significant because it highlights not just an isolated instance of poor sourcing, but **a chronic pattern of under-referenced, manipulable content.** Upon a cursory review, the page displays a constellation of sources (for the 2025 edits) that (again) mimic the appearance of legitimate journalism, but are in fact cloned or fabricated outlets tracing back to Russia (for instance, orenada-news, Caliber.Az).

The Behavioural Volatility Index (BVI) assesses editorial irregularities, using *sporadicity*, *contributors concentration* and *add/delete ratio* (equally normalised and averaged). It helps to capture erratic engagement patterns that may suggest chaotic or conflict-driven content changes, and all potential markers of editorial disruption or reactive content disputes.

Preliminary findings: pages with high BVI scores did not always align with entries with high manipulation or sourcing risk. This finding supports the idea that **behavioural chaos (mass edits clustered around political events or crises) is not inherently manipulative but may still provide an entry point for hostile actors.**

The article 'Moldovlased' (Estonian language entry for *Moldovans*) ranks highest with a BVI score of 1.0, exhibiting the maximum observed volatility across the dataset (registered in August 2025).[62] It showed a pattern of sustained editorial instability, most likely suggesting ongoing contestation and unresolved editorial disputes. The French-language article on Moldova's EU accession procedure and the Russian-language entry on Moldova also rank highly, with BVI values above 0.70.

1.4. Conclusion

It is worth reminding that the corollary of hybrid influence operations is simple: **if you can convince people that democracy does not work, you do not even need to convince them who to vote for.** Perhaps this is the most suitable adage for what we observed during Moldova's parliamentary elections.

The monitoring and analysis of Moldova's 2025 parliamentary elections reveals a strategically fragmented and hybridised disinformation ecosystem, where influence operations exploit both platform-specific dynamics and cross-domain synergies. Information manipulation is not confined to overt propaganda, but materialises in more subtle epistemic disruptions, ranging from coordinated editing on Wikipedia (a collaborative knowledge production space) to narrative laundering via short-form video and diaspora-targeted messaging, among others.

We found evidence of coordinated campaigns, some of which involved transnational dissemination strategies with adaptations across a constellation of languages and geographies. This pattern underscores the transboundary nature of digital threat architectures, where geographic indicators can be deceptive and digital attributions strategically camouflaged. We also observe the rising role of platform convergence: manipulation campaigns were often mirrored across multiple ecosystems, with TikTok serving as a launchpad for memetic messaging, while Telegram acted as an operational coordination hub. Content was repackaged, cross-posted, and seeded in ways that blur the lines between organic expression and tactical influence. Therefore, our findings advocate for a broader, integrated understanding of electoral interference, one that includes not just what is said, but how, where, and through which architectures it circulates.

[62] Although the bar chart is aggregated at article level, the entry appears multiple times in the previous months from January through July 2025.

This is the result of a granular, months-long investigation, involving the coordination of transnational teams of researchers, journalists and grassroots civic groups, affords us a candid diagnosis. Our capabilities for detection and monitoring increased significantly. Multifaceted data collection and analysis software tools enabled real-time threat detection. At least four such instruments were deployed for an in-depth interpretation of coordination patterns.

As civil societies, we are better positioned to understand the hybrid battlefield leveraged against democratic integrity. And yet, there is a prevailing belief amongst many of us that, despite all these analytical capabilities and innovations, we trail behind, more insecure than ever. States, supranational structures and alliances still operate in policy and decision-making silos, marked by the hard borders that continue to govern geo-strategic thinking. Uncomfortable truths are cushioned and curated to suit political sensibilities, while adversaries move fluidly across systems, exploiting precisely those gaps in vision and resolve. Platforms, too, play an increasingly destructive role. Those ostensibly designed to democratise speech, now systematically amplify what enrages, divides and confuses – algorithmic incentives and democratic interests are no longer aligned.

We are engaged in an asymmetric warfare, where grassroots counter-response remains dwarfed and powerless in comparison. And yet, it is within these grassroots citizen-led communities of practice that the first alarms are often sounded and where democratic values are lived rather than merely legislated. If we are to withstand the pressures of this evolving assault on democracy, we must elevate local defences, give them a resolute voice, and the capabilities to pre-emptively respond – not in echo chambers but on the digital frontlines.

2. WAR IN UKRAINE: THE ENDURANCE OF DISINFORMATION TRENDS

2.1. Main findings

The war in Ukraine continues to dominate the European disinformation landscape, not only as a geopolitical crisis but as a persistent source of polarizing narratives across social media. This overview of disinformation on war in Ukraine draws on the analysis of coordinated inauthentic behaviour and main narratives to identify how social media communities frame the war. By mapping engagement patterns and network structures, the analysis also reveals how narratives are amplified.

The analysis reveals that:

- Coordinated activity across platforms centres on **contesting President Zelensky's legitimacy.** Zelensky's leadership is repeatedly questioned through claims that portray him as corrupt, extremist, or responsible for prolonging the war, while alternative figures - notably Donald Trump - are elevated as potential peace brokers to reinforce a contrasting image of pragmatic, solution-oriented leadership.
- The war in Ukraine is widely depicted as a **proxy struggle between the West and Russia**, reinforced by claims portraying Ukraine as a Western puppet, economic instability as an inevitable consequence of the war, and global power dynamics as shifting. Economic narratives further amplify this by linking the war to energy crises, sanctions, and fears of broader societal or economic collapse.
- **Twitter/X** functions as a rapid-fire arena where polarized narratives and information-warfare themes converge, producing a highly charged and often adversarial portrayal of Zelensky.
- **Facebook** amplifies emotionally-driven, ideologically-rigid narratives that cast Zelensky within moralized and populist critiques.
- **Instagram** presents a more curated, diplomatic, and institutionally framed image of Zelensky, emphasizing polished international engagements and restrained, policy-oriented commentary.
- **Total engagement** around the Ukraine war is similar across platforms, but arises from different dynamics—high-efficiency, lower-volume posting on Facebook versus sustained, high-volume activity on Twitter/X—with attention peaking at different times and influence concentrated in key hub accounts or established pages rather than evenly distributed.
- The **high-engagement ecosystem on Twitter/X** is shaped by a mix of political figures, specialized media, activists, and highly polarized commentators, with engagement concentrated among mid-level-follower, frequently posting accounts rather than solely high-follower politicians or news outlets, highlighting a disconnect between follower count and audience impact.

2.2. Methodology

The analysis of disinformation narratives and their coordination on social media rests on several operations. We collected 1,656,205 social media posts across 3 platforms - Twitter/X, Facebook and Instagram. Our analysis of coordinated behaviour[63] yielded 60 communities on

[63] Coordinated behaviour refers to situations in which two or more social media accounts repeatedly perform actions involving the same uniquely identifiable content within a predefined time interval (Righetti & Balluff, 2025). To detect CIB, we used the CoorTweet package.

Facebook, 8 on Instagram and 6 on Twitter/X, active from April to August, 2025.[64] We also filtered the main disinformation narratives, using the PROMPT Corpus Analyser, resulting in a smaller dataset (3137 posts) that we reviewed manually. We ran an additional network analysis to compare and contrast the evolution of online conversations around this topic.[65]

2.3. Disinformation narratives and online coordination on the war in Ukraine

Narratives on the war in Ukraine are different across Twitter/X, Facebook and Instagram, shaped by the platforms' distinct communicative styles.

On **Twitter/X**, the discourse is fast-paced, fragmented, and often confrontational. Posts tend to be short, punchy, and heavily reliant on rhetorical devices that maximize emotional impact. Zelensky is portrayed in polarized terms—either as a symbol of resistance or as a corrupt figurehead—depending on the community. Allegations of corruption are common. The platform also amplifies information warfare narratives, with posts warning of "underground PR agencies" and "information sabotage," suggesting a battlefield of perception as much as policy.

Facebook fosters emotionally intense and ideologically entrenched narratives. The platform's longer format allows for more elaborate storytelling, often infused with grassroots activism and populist sentiment. Zelensky is frequently criticized, with posts accusing him of prolonging the war and mismanaging resources. Posts include moral condemnation, delegitimization, and emotional appeals, such as portraying leadership as disconnected from public suffering. The discourse is marked by binary framing—East vs. West, good vs. evil—and appeals to collective identity, often mobilizing outrage against elites.

Instagram, by comparison, supports more curated and strategic messaging. Criticism against Zelensky is more restrained and policy-focused. Military developments in the war and sanctions are often correlated with broader negative economic consequences, notably on energy markets. The tone is informative rather than inflammatory, reflecting media-driven content and institutional messaging.

The analysis of coordinated inauthentic behaviour detects coordination around the following narratives and claims:

- There is a pervasive focus on the **Zelensky government's legitimacy,** though the framing varies by platform. On Facebook, this manifests through a Russia-driven disinformation narrative portraying **the Ukrainian government as losing its legitimacy**, with claims that Zelensky has dismantled anti-corruption laws and betrayed democratic principles. This connects to a harsher framing on Facebook where **Zelensky is depicted as an aggressive war-monger** who sacrifices his population, reflecting a broader disinformation narrative on Ukraine and Ukrainians (including refugees) being

[64] An edge_weight threshold of 0.99 was used so as to analyse the themes that actually emerge from coordination, without forcing the query-to-theme assignment. We limited our qualitative analysis to a random sample of 30 posts per community.

[65] For network analysis, three groups of indicators were used - engagement indicators (number of posts, total engagement, engagement per post, peaks); network indicators (parent relationships through retweets/reposts); actor-level metrics (see theTechnical Appendix 1 for detailed explanation). Several methodological and data-related constraints must be acknowledged: 1. network reconstruction was only fully possible on Twitter/X, as Facebook datasets lacked explicit resharing or reply metadata. This means that network-based comparisons across platforms must be interpreted cautiously and should not be generalised beyond the available interaction types 2. engagement data cannot be interpreted as public opinion or sentiment, only as interactional behaviour (likes, comments, reshares).

aggressive war-mongerers who pose a threat to European and global security. It is exemplified by a post stating: "Zelensky envoie les jeunes mourir" (Zelensky sends the youth to die). In contrast to Zelensky, other political figures such as Donald Trump are sometimes portrayed as potential peace brokers. For example, one post speculates about a meeting between Trump, Putin, and Zelensky in Serbia, suggesting that Trump's involvement could lead to a resolution of the conflict. This narrative positions Trump as a pragmatic leader capable of negotiating peace, while Zelensky is depicted as obstructive or unwilling to compromise. On Instagram Zelensky is often portrayed in a contested light. One narrative presents Zelensky as a proactive leader seeking international support. For example, a post describes his meeting with German Chancellor Friedrich Merz in Berlin, emphasizing Ukraine's efforts to secure further military aid. This framing positions Zelensky positively as a diplomatic actor navigating complex alliances, and it aligns with a broader narrative of Ukraine as a nation defending its sovereignty against aggression. However, this must be balanced with strong criticisms against the Ukrainian President. In one instance, former US President Donald Trump is quoted as accusing Zelensky of prolonging the war, referring to Ukraine as a "killing field." This rhetorical move shifts the focus from external aggression to internal leadership, suggesting that Zelensky's decisions may be contributing to continued violence. This framing connects with a broader narrative on accountability and legitimacy, in which Zelensky is not only a victim of geopolitical conflict but also a participant whose actions are subject to scrutiny. On Twitter/X, the delegitimization is more ideological and polarized, often reflecting the Kremlin's narrative of the **supposed allegiance of the Ukrainian government to Nazism**[66]. Other posts, including several citing Tucker Carlson, claim that Zelensky "has all the characteristics of a dictator". Across all platforms, criticisms directed towards President Zelensky are often **emotionally charged.** 1 out of 3 posts reviewed in the filtered dataset (3437 posts) included one or several rhetorical devices of emotional manipulation[67]. For example, a lament from a resident of Kharkiv, for instance, describes the aftermath of a bombing that killed 17 people, including four children. Such vivid imagery can be appropriated to evoke outrage, reinforcing the narrative that Zelensky's leadership has led to unnecessary suffering.

- Across the platforms is often found the narrative that **Ukraine is a platform for the West in its geopolitical fight against Russia**[68]. On Facebook, Zelensky is explicitly framed as a **puppet of Western powers**, with texts implying his actions align more with NATO or US interests than with the Ukrainian people. This narrative is reinforced by posts that describe mass protests against his government, suggesting a growing domestic discontent. One such post reads: "Crise en Ukraine: les Ukrainiens se retournent contre Zelensky, manifestations partout" (Crisis in Ukraine: Ukrainians turn against Zelensky, protests everywhere). Using vivid metaphors - such as Russia striking "the brain of NATO," the EU trembling after a Trump–Putin handshake, or references to a "coup fatal" and a "séisme géopolitique" - posts dramatize events to heighten the conflict's perceived stakes and imply that global power dynamics are changing. This sentiment is echoed on Instagram, where discussions of US-Russia relations and sanctions support the narrative that Ukraine is a puppet of the West in its geopolitical fight against Russia, **framing the war as a larger struggle between global powers**. Zelensky's role within this

[66] This narrative represents two out of three posts in the filtered dataset (3137 posts), and essentially posts on Twitter/X.

[67] These are the following "emotional" rhetorical devices: accismus, bdelygmia, dysphemism, loaded language and phrasemes

[68] This narrative is closely associated with the communities showing the strongest coordinated activity on Facebook specifically. By contrast, fewer communities coalesce around other war-related narratives (e.g. "The conflict comes at the expense of domestic welfare", Community 33; "Illegitimate Kyiv government", Community 12).

context is both symbolic and strategic . He is central to debates about military aid, peace negotiations, and leadership ethics. While Facebook users utilize metaphorical language about a "geopolitical earthquake" to describe these dynamics, Twitter/X users engage in a battleground of perception and identity, where the geopolitical struggle is often simplified into a moral dichotomy of "good" Ukraine versus "evil" Russia. or conversely, dismissed as Western propaganda.

- The war is also discussed through economic lenses. Facebook communities engage with economic narratives that **correlate the war to energy crises and trade disruptions.** On Twitter/X, this theme is visceral and fear-based, emphasizing a narrative of **economic collapse or social division stemming from the war in Ukraine.** Here, the narrative is dressed in exaggerated language about "bloody business plans" or suggests that Ukraine's involvement threatens the internal stability of other nations. Alternatively, the brutalities of war are often dismissed as Ukrainian and Western propaganda[69]. On Instagram, while economic consequences are mentioned, they are often tied to broader discussions of international diplomacy and sanctions rather than the fear-driven "collapse" narrative seen on Twitter/X.

2.4. Engagement and network dynamics

2.4.1. Engagement patterns

The war in Ukraine is an **enduring and geopolitically structured topic.** Conversations on Ukraine on X and Facebook unfold around a (now) long-standing international conflict, in which both news media and political figures maintain sustained communication over time.

Across the observed six months, more than 1.1 million posts addressed the war across the two platforms. Twitter/X produced the majority of content with 649,532 posts, while Facebook accounted for 454,121 posts. Despite Twitter/X's higher posting volume, the total engagement was nearly identical on both platforms: 125,611,952 interactions on Twitter/X and 125,602,144 on Facebook. This symmetry in total engagement masks two entirely different platform dynamics. Whereas Twitter/X achieved its engagement through sustained high-volume activity, Facebook generated an almost identical level of public interaction with 30% fewer posts, indicating significantly higher efficiency per post.[70]

The platform-specific temporal patterns also diverged: Facebook reached its highest weekly engagement per post in week 2025-W14, with 364.18 interactions per post, while Twitter/X's peak appeared later, in week 2025-W22, with 277.38 interactions per post (Figure 1). This temporal mismatch suggests asynchronous mobilization across platforms rather than a coordinated or simultaneous spike in attention.

[69] This narrative is the third largest in the filtered dataset.

[70] Indeed, when engagement is normalized by volume, the contrast becomes clear. Facebook recorded an average engagement per post (EPP) of 271.94, compared to 185.89 on Twitter/X.

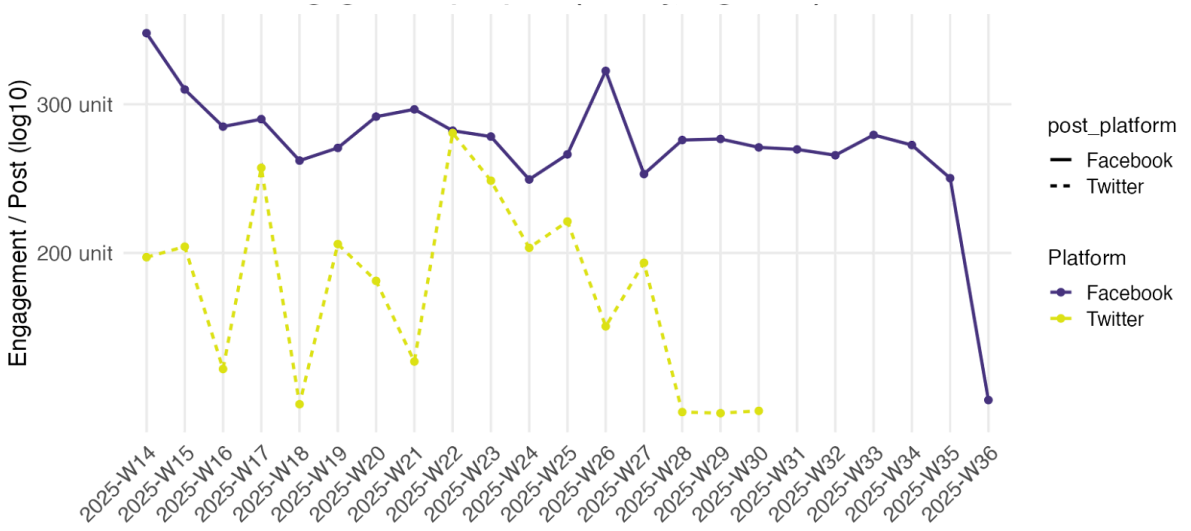


Figure 1: Weekly engagement per post – Topic War in Ukraine

These findings suggest that, within Ukraine-related discussions, Facebook serves as a high-impact, low-volume environment, whereas Twitter/X operates as a high-volume, continuous information network. This contrast aligns with platform affordances: Facebook's algorithmic feed amplifies highly engaged content even if fewer posts are produced, while Twitter/X favors rapid, frequent messaging, particularly among journalists, politicians, and activists.

2.4.2. Top influencers

On Twitter/X, the discussion around the war in Ukraine is highly concentrated,[71] polarized and consistently emotional in nature.

It is dominated by a mix of political actors,[72] polarized commentators,[73] and specialized media focused on the conflict. These may not necessarily have converging views on how to solve the conflict. Several mainstream or specialized media promoting factual coverage of the war, or pro-Ukrainian accounts,[74] also generate significant engagement. This suggests a strong polarization in the conversation around this topic, or at minimum that parallel conversations are taking place around this topic.

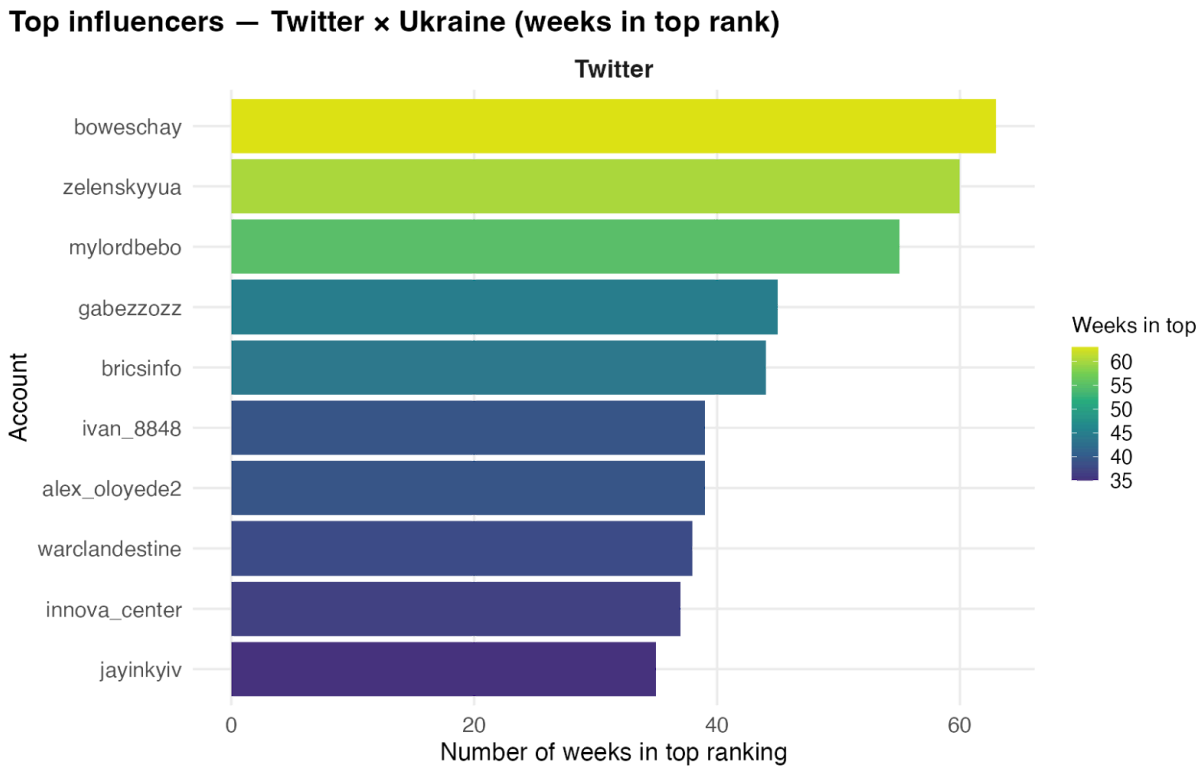
There is also a significant disconnect between massive follower counts and top engagement metrics:

- **Low engagement rate for high followers:** Accounts with the largest follower bases (Rep. Marjorie Taylor Greene: 4.9M followers) do not top the engagement list. For example, Chay Bowes, with "only" 231K followers, generated 342,035 engagements from 184 posts.
- **High engagement rate for mid-level followers:** Highly polarized, frequently posting commentators generate disproportionately high engagement. Figures like Chay Bowes

[71] The top hubs, measured by weighted in-degree, were accounts receiving the highest number of incoming reposts or mentions. The leading hub accumulated a weighted in-degree of 3750.9 (raw in-degree 7,519), followed by others with 2354.0, 771.9, 744.4, and 631.1, respectively. These accounts appear in the export as numeric identifiers (e.g., 1.720665e+18) because no screen names were attached in the file. This suggests that network influence is highly concentrated, but the absence of account names prevents content-level interpretation.
[72] Such as Rep. Marjorie Taylor Greene or Jordan Bardella.
[73] Such as Chay Bowes, Spetsnaℤ 007 or Alex Jones
[74] Unsurprisingly, for example, Volodymyr Zelensky's account.

(231K followers) lead in total engagements, suggesting their content is exceptionally successful at driving user activity within their targeted audiences.

- **Volume vs. impact:** An account like Alexander Ivanov posted 2,854 times, achieving 111,076 engagements, demonstrating that high post volume can yield high aggregate engagement, but does not necessarily translate to top-tier engagement per post compared to accounts like Chay Bowes.[75]



Among suspected disinformation proxies, a substantial portion of accounts express anti-establishment orientations.[76] Emotional intensity also appears in more expressive or combative formulations - e.g. mentions that "Truth matters — even when they want us silent."

The metadata indicates that platform mechanics play a central role in shaping how discourse is produced, shared, and interpreted. Several structural features appear to facilitate rapid circulation, high visibility, and the use of compressed identity markers.

- **Rapid dissemination and high engagement**: The very high metrics associated with certain accounts—reaching up to 16.8 million views, 342,035 engagements, and substantial numbers of shares—demonstrate that the platform enables fast and widespread dissemination of content. These indicators suggest that posts can be amplified quickly through user interactions and algorithmic distribution.
- **Hashtag-based thematic organization:** Hashtags function as organizational tools that consolidate discourse around specific issues or affiliations. Examples such as #Frexit show how users categorize their posts within recognizable thematic clusters, enabling rapid aggregation of related content.

[75] A similar trend is observed on Facebook.

[76] These uses phrases like "The System Has Been Compromised" and "Exposing Elites," indicating persistent scepticism toward institutional actors. Some accounts adopt a critical stance toward media and authority through statements including "Media is a Virus," "Non Compliance is the Cure," or ideological self-descriptions like "Anti woke – Anti fake news – Anti hypocrisy."

- **Identity signalling through flags and symbols:** The frequent use of national flags (🇷🇺) in usernames and biographies serves as an immediate visual marker of political or national identification. These symbols allow users to position themselves within particular interpretive communities, influencing how their posts are received and circulated.
- **Anonymity and pseudonymity:** Many highly active accounts operate under pseudonyms. The absence of personal identification allows users to adopt stylized or exaggerated personas, which may facilitate more direct, provocative, or unfiltered forms of expression.

2.5. Conclusion

The analysis of disinformation surrounding the war in Ukraine demonstrates the persistence and adaptability of hostile narratives across platforms. These narratives are not limited to military developments but extend into identity and economic frames, often coordinated across online communities. Prominent examples include claims that Ukraine is a Western puppet in a proxy war against Russia, allegations that President Zelensky dismantled anti-corruption laws and sacrifices his people for foreign interests, and fear-driven narratives linking the war to Europe's imminent economic collapse and energy shortages. Such narratives are amplified through synchronized bursts on Twitter/X and Facebook, frequently paired with emotionally charged rhetoric portraying Zelensky as a dictator or associating Ukraine with Nazism. Their durability and cross-platform migration underscore the challenge of countering disinformation in protracted crises, where the informational battlefield becomes as enduring as the physical one.

Platform-specific dynamics amplify these narratives in distinct ways. Twitter/X operates as a high-volume, rapid-fire environment where polarized discourse and ideological framing dominate, while Facebook fosters emotionally charged, identity-driven narratives that embed disinformation within moralized critiques of leadership and governance. Instagram, by contrast, offers a more curated and diplomatic portrayal, yet still reflects contested interpretations of Zelensky's role and Ukraine's trajectory. These differences reveal that disinformation does not spread uniformly but adapts to the affordances and audience expectations of each platform, complicating detection and mitigation strategies.

Engagement analysis point to few structural patterns: the concentration of influence within a small cluster of highly active accounts, the absence of synchronized engagement peaks across platforms, and the reliance on emotionally loaded rhetorical devices to sustain attention. These patterns confirm that disinformation about Ukraine is not an incidental by-product of conflict but a deliberate, networked strategy aimed at eroding trust, polarizing publics, and reframing geopolitical realities. Addressing this challenge requires integrated responses that combine technical monitoring with narrative-level interventions, capable of disrupting both the content and the coordination mechanisms that enable its persistence.

3. US VS. THEM, GOD'S DESIGN AND ELITE-RESENTMENT: DISINFORMATION AGAINST LGBTQ+ INDIVIDUALS AND COMMUNITIES

3.1. Main findings

In recent years, transgender and gender-diverse individuals — particularly trans women — have become central targets of coordinated disinformation campaigns[77]. These efforts, often spearheaded by right-wing and fundamentalist groups, have subjected LGBTQ+ communities to sustained attacks on their identities and human rights. The EU is starting to frame this dynamics as a structural problem. Under the Digital Services Act, the European Board for Digital Services now identifies gender-based violence as one of the main systemic risks that very large platforms must assess and mitigate.

Given the documented intertwinement between anti-LGBTQ+ disinformation and gender-based violence,[78] this chapter analyses the main narratives of disinformation targeting LGBTQ+ individuals, the rhetorical devices and persuasion techniques employed, and the emotional triggers these narratives seek to activate. It also examines the most active accounts on X involved in anti-LGBT disinformation and evidence of coordinated behaviour across platforms.

It finds that LGBTQ+ narratives are constructed and disseminated in different ways across platforms:

- **On Twitter/X, protection frames are often employed as covers for exclusion**: child-safety, school "indoctrination" claims, paired with women's sports fairness and the idea that "gender ideology" has taken over institutions and is undermining cultural purity are often invoked through protective lenses to promote exclusion of LGBTQ+ individuals.
- **These X-forged frames also surface on Facebook inside broader threads** (e.g., Ukraine or elections), where they are recycled to moralize geopolitics and mobilize audiences.
- **The same persuasion techniques are used across both platforms** — labeling ("biological males," "groomers"), false equivalence (equating gender-affirming care with FGM), and appeals to authority (court rulings, agency probes) — to trigger emotional reactions, such as outrage at elites, resentment about fairness in women's sports, fears for physical safety, and concerns about health risks, especially around trans athletes and healthcare.
- **A prominent framing employed in these campaigns is "negative othering":** LGBTQ+ individuals are portrayed as outsiders whose existence threatens the values and cohesion of an imagined "us." This us-versus-them logic is pervasive, often framing supporters of transgender rights as adversaries to traditional or national values. These narratives are reinforced by associating LGBTQ+ advocacy with Western or elite influence. Official statements and institutional actions — such as court rulings or government investigations — are frequently cited to legitimize these exclusionary

[77] Gender disinformation in the context of LGBTI communities, Submission to the Special Rapporteur on freedom of opinion and expression, 7 July 2023. Available at https://www.ohchr.org/sites/default/files/documents/issues/expression/cfis/gender-justice/subm-a78288-gendered-disinformation-cso-ilga-world.pdf

[78] See *The inextricable link between Gender Disinformation and Gender-Based Violence* in Gender disinformation in the context of LGBTI communities, Submission to the Special Rapporteur on freedom of opinion and expression, 7 July 2023. Available at https://www.ohchr.org/sites/default/files/documents/issues/expression/cfis/gender-justice/subm-a78288-gendered-disinformation-cso-ilga-world.pdf

narratives and intensify social divisions, particularly around contentious issues like sports participation, education, and healthcare.

- **The need to protect the "natural family structure" or "natural order" is also a central narrative.** This frame appears in both religious and secular forms.[79] Despite their different 'wraps', both variants rely on similar rhetorical devices— antithesis (us vs. them), hyperbole, and anecdote/metonymy — and evoke overlapping emotional triggers, such as outrage at elites, resentment over perceived injustice, fear of cultural loss, and a sense of lost control.

Together, these intertwined narratives and rhetorical strategies form a complex ecosystem of persuasion, shaping public attitudes and policy debates around LGBTQ+ rights.

3.2. Methodology

The original dataset contains 1,711,649 posts, of which 845,469 from Twitter/X, 751,658 from Facebook, and 114,522 from Instagram, spanning April-July/August 2025. It results from keyword-based queries in the eight PROMPT languages[80] and lists of (problematic) accounts pre-identified by civil society activists and journalists. A coordinated behaviour analysis (CIB)[81] and network dynamics analysis[82] was applied to the dataset, which yielded 44 semantic communities. The dataset was also filtered with the PROMPT LGBTQ+ narrative taxonomy, resulting in a dataset comprising 128,594 posts. 9485 posts were randomly selected and analysed through the PROMPT Corpus Analyser and reviewed qualitatively.

3.3. Disinformation narratives and online coordination targeting LGBTQ+

The analysis of the broad corpus and filtered dataset confirms a well-established finding in LGTBQ+ and disinformation scholarship:[83] isolating narratives conceptually is helpful to build taxonomies, but in practice **narratives co-occur, interlock, and mutually reinforce one another across online communities**. For example, narratives portraying the LGBTQ+ community as a threat to child safety often overlap with claims that an imagined "gender ideology" is dominating Western liberal democracies, particularly in educational settings[84].

---

[79] Religious variants invoke concepts like "God's design," "sin," and "moral decay," often amplified by faith leaders and scriptural references; secular variants emphasize "biology," "common sense," demographic concerns, and parental rights

[80] For Facebook and Instagram, datasets were not filtered by language, nor was any language constraint applied during the download via the Meta Content Library. The broader corpus remains however largely dominated by social media posts in English.

[81] 27 communities on X; 9 communities on Facebook; 8 communities on Instagram. Samples of up to 30 posts per community were identified. A coordination threshold of edge_weight ≥ 0.99 was applied to ensure themes arise from coordination (not isolated virality).

[82] For network analysis, several methodological and data-related constraints must be acknowledged: 1. network reconstruction was only fully possible on Twitter/X, as Facebook datasets lacked explicit resharing or reply metadata. This means that network-based comparisons across platforms must be interpreted cautiously and should not be generalised beyond the available interaction types 2. engagement data cannot be interpreted as public opinion or sentiment, only as interactional behaviour (likes, comments, reshares).

[83] Strand, C., & Svensson, J. (2021). Disinformation campaigns about LGBTI+ people in the EU and foreign influence(Briefing PE 653.644). Directorate-General for External Policies of the Union, European Parliament. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/653644/EXPO_BRI(2021)653644_EN.pdf

[84] This is illustrated by the CIB analysis of Facebook: the community showing the strongest coordinated activity on LGBTQ+ topics (Community 3) simultaneously mobilised three of our narratives: (1) an imagined "gender ideology" dominating institutions, (3) the LGBTQ+ community as a threat to child safety, and (4) LGBTQ+ people as a corrupting influence on morally "pure" societies. This configuration empirically confirms that anti-LGBTQ+ disinformation narratives tend to be tightly intertwined in practice.

The **most frequently present narratives identified by the PROMPT Corpus Analyser are also those which are the object of online coordinated inauthentic behaviour.** They focus on the putative corrupting influence of LGBTQ+ on 'pure' societies (890 items), the domination of gender ideology across liberal democracies (503), the need to protect the natural family/order (299). They also present LGBTQ+ rights as a tool of Western cultural colonialism" (255); as a public-health hazard (195); and as a threat to child safety (175).

Each of these narratives rests on different arguments:

- **LGBTQ+ has a corrupting influence on morally "pure" societies**: this frame casts LGBTQ+ visibility as contamination of a "healthy" social body and is the backbone into which other claims are nested. On both Twitter/X and Facebook, it stands in the background of other topics. The PROMPT Corpus Analyser confirms its centrality: it is the most frequent narrative (890 items) and it regularly travels with anti-West/EU and "gender ideology" narratives. This narrative is easy to package with others. Whatever the issue, elections, foreign policy, or education, it acts as a moral compass. It also has an insidious role: it supplies a ready moral rationale that normalizes exclusion as cultural self-defense.

- **An imagined "gender ideology" is dominating Western liberal democracies:** the narrative claims that key institutions (schools, courts, federations, media, regulators) have been "captured," turning "gender ideology" into imposed orthodoxy. Coordinated behaviour shows, for example, that a story acts as backbone for diffusion. The story argues that schools are "out of control," women's sports "unfair," and recent rulings or guidelines are "going too far." It travels easily from X, where it is reposted heavily, into Facebook, where it permeates inside other conversations (e.g., elections administration, Ukraine/war threads). **LGTBQ+ issues therefore remain in the background even when not the main conversation topic.** The analysis of coordination also shows that this storyline often pairs with elite-resentment narratives. According to it, a corrupt and/or distant elite is imposing "gender ideology" on ordinary people, which then serves to legitimate concrete crackdowns (bans, investigations, funding cuts) at school-board, ministry, or federation level. In the filtered dataset, this narrative is the second most frequent (503 items), and its qualitative review shows that it mixes and mingles with other polarised topics and issues.

- **The "natural family structure" / "natural order" must be protected**: this narrative argues that heteronormative families exclusively safeguard social stability and demographic continuity. It appears in both religious (divine design, moral decay) and secular (biology, common sense, parental authority, demographic anxiety) tropes. It is coordinated via different clusters which relate it to school controversies, parental rights, and women's sports. On Facebook, it often piggybacks on electoral or Ukraine threads. The fact that it often appears alongside other anti-LGBTQ+ narratives - such as "gender ideology as imposed orthodoxy", or anti-Western/anti-EU cues, suggests that it acts as a backbone identity frame which supports other narratives, rather than a standalone narrative.

- **LGBTQ+ rights as Western cultural colonialism:** This narrative casts LGBTQIA+ inclusion as a foreign imposition, an evidence of a morally decaying West (often "the EU" writ large) exporting corrosive values. It pairs quickly with EU-skeptic and anti-Ukraine frames ("Gayropa," "West in decline"), meaning that even inside an LGBTQ+ query you find adjacent geopolitical storylines bundled together. This bundling shows up at scale in the PROMPT corpus (255 items). This narrative is especially visible on Facebook, where LGBT cues are grafted onto broader debates about sovereignty, corruption, and war.

The effect is to recode anti-LGBT messaging as national self-defence: resisting "Western ideology" is to protect tradition, identity, but also sovereignty.

- **LGBTQ+ identities are a public-health hazard**: A quieter but strategic thread treats identity and care as clinical dangers. Communities cross-post to pathologize LGBTQ+ rights. Under claims of clinical harm (e.g. analogies of mutilation), this storyline moves the debate from "morals" to an alleged health emergency (even though moral panic is never fully displaced). This smaller narrative is qualitatively important: it upgrades the narrative on "protecting the children" into a seemingly evidence-based and scientifically-backed rationale for bans, funding cuts, and audits. It often shows up alongside the "institutional capture" storyline, implying that regulators or hospitals are no longer acting independently.

- **The LGBTQ+ community is a threat to child safety**: this storyline claims that the promotion of LGBTQ+ rights endanger children, via "indoctrination," "grooming," or exposure to "inappropriate" content. Communities are particularly well-structured on Twitter/X, but the narrative also appears on Facebook inside broader threads (elections, Ukraine). There, child-protection is used as a moral geopolitical compass: the focus shifts from strategic questions ("who is winning the war?") to civilizational ones ("is the West in moral decline, exporting 'corrupt' values to our families?"). Foreign-policy or electoral debates are reframed as tests of virtue to protect "values under siege" and children. Coordination and rhetorical analysis also reveal frequent pairing with the "gender ideology" narrative: school cases and curricula are cited as proof that institutions are "captured," which intensifies the sense that children need shielding.

Alongside these main narratives, and their supporting claims and arguments, several storylines also feature in the analysed datasets:

- **LGBTQ+ rights, and transgender rights in particular, are criticized behind the veil of the integrity and safety of women's sports.** Coordination analysis shows that sports debates open up the conversation on other targets - schools, bathrooms, books, healthcare; etc. These conversations draw a lot of engagement. They have a strong emotional appeal, tapping into resentment over perceived injustice. **They are used as a platform for calls of action not only to better regulate women's sports, but other sensitive spheres** (bathrooms, library shelves, hospitals, etc.). In short, sports provides a socially acceptable front door: once the fairness premise is accepted, adjacent anti-LGBTQ+ positions are easier to advance and defend across platforms.

- While not an LGBTQ+ narrative, **"elite-resentment" is a mobilizing force for anti-LGBTQ+ communities.** Many posts express frustration at powerful institutions — such as courts, ministries, school boards, public broadcasters, universities, "Big Tech," or hospital administrators — and channel this frustration into support for anti-LGBTQ+ goals. This framing suggests that a "corrupt elite" is imposing "gender ideology" on ordinary people, which is then used to justify crackdowns, including bans, investigations, and funding cuts[85].

- While PROMPT focusses on anti-LGBTQ+ narratives, it is worth noting that the analysed datasets also surfaces pro-rights discourse.[86] While polarisation around LGBTQ+ rights

[85]The opposition to elites is observable across multiple, heterogeneous narratives. In the Community 11 identified on Facebook, the narratives "Ukraine is a platform for the West in its geopolitical fight against Russia" and "An imagined 'gender ideology' is dominating Western liberal democracies" are both present, and both designate a common enemy: Western liberal elites and institutions.

[86] For example, one community rallied around court reviews or "forced outing" bills, framing them as threats to equality and recognition and urging collective action.

is unsurprising, PROMPT's linguistic analysis shows that **pro- and anti-LGBTQ+ actors draw on a very similar toolkit — persuasion techniques[87] and rhetorical devices[88]. They activate similar emotional triggers[89], but with opposite goals.** For example, pro-rights posts channel outrage to defend measures safeguarding LGTBQ+ individuals' rights; while their opponents propose to restrict them. This underscores the polarisation of online discourse and the adaptability of persuasive strategies supporting opposing narratives.

3.4. Persuasion techniques, rhetorical devices and emotional triggers mobilised in anti-LGTBQ+ discourse

The analysis of both the Corpus Analyser and the CooRTweet package uncovered **a recurring set of persuasion techniques and rhetorical devices in anti-LGBT narratives, each closely linked to specific emotional triggers.**

Casting doubt (6154 posts) is the most prevalent persuasion technique observed. Together with **appeal to authority** (1463 posts), these two persuasion techniques are used to redirect anger and indignation toward institutions, as in posts - derived from the CIB reports - framing schools, courts, or media as "captured/woke," or by citing "investigations," "AG probes," and "federal action" to signal that something harmful is happening, especially to children - as shown by the CIB reports. These strategies also evoke a sense of loss of control or powerlessness, with phrases like "they're forcing this on us," "parents sidelined," or "no say". These all suggest that ordinary people are unable to influence outcomes.

Name-calling and **labeling** (5981 posts) are the second largest group of persuasion techniques identified. Within online communities, we found that terms like "woke," "groomers," and "radical leftist" are used to provoke anger, disgust, and a sense of identity threat, reinforcing an "us vs. them" mentality. For instance, posts have described LGBTQ+ activists as "dangerous LGBTQ extremists" or accused them of flying a "child mutilation flag," directly invoking fear and disgust.

Narratives concerning schools & education, public health & hospitals and sports employ persuasion techniques such as:

- **slippery slopes** (16 posts) - posts arguing that if trans girls are allowed to compete, women's sports will "end"
- **false dilemmas** (31 posts) - posts arguing that we should either protect women's sports by banning 'biological males' or sacrifice fairness
- **overgeneralization** (1002 posts) - a single viral incident at one school is used to declare that "schools are captured by gender ideology," followed by calls to investigate/pull funding for the entire district/system.

These findings are contextualised in Chapter 4 of this report, in which Italian and Romanian fact-checkers reflect on 'generalization' as a commonly employed technique of LGBTQ+ disinformation in both countries.

Fearmongering (1931 posts) and **appeals to prejudice** (1610 posts) are also often mobilized. Claims conveying these persuasion techniques include the idea that prestigious institutions are pressuring young children to declare pronouns every year, suggesting that all schools are next.

[87] Appeal to authority, labeling/loaded lexicon, casting doubt.
[88] Anecdote/metonymy, hyperbole, antithesis/us-them.
[89] Outrage at elites, resentment over perceived injustice, fear of systemic chaos / loss of control, and fear for physical safety/health

Others portray trans girls as a safety risk and argue that allowing them into teams/locker rooms "endangers girls," pushing bans as the only way to keep women safe.

To support these emotionally-loaded persuasion techniques, several rhetorical devices are often mobilised: hyperbole, antithesis, rhetorical questions, and anecdote/metonymy are used to intensify emotional responses. **Loaded language** and **hyperbole** often appear together to invoke fear and disgust, as seen in statements such as "This is GROOMING." **Antithesis** is used to reinforce binary oppositions — such as "parents vs elites," "biological women vs men in women's sports," and "tradition vs woke" — which fuel anger and strengthen in-group cohesion. **Anecdotes and metonymy** are often paired with **overgeneralization** to frame isolated incidents into alleged evidence of a broader civilizational danger. For example, a post highlights a single headline about an FBI probe into a children's hospital and pairs it with Feminine Genital Mutilation (FGM) language, then generalizes to claim that *"children's hospitals/medicine"* are "butchering kids" and must be shut down or investigated system-wide**.**

Overall, these rhetorical and persuasive techniques are carefully orchestrated to elicit **strong emotional reactions,** — namely fear, anger, disgust, and anxiety — mobilize audiences, and reinforce exclusionary attitudes toward LGBTQ+ individuals.

3.5. Engagement and network dynamics

3.5.1. Engagement patterns

Unlike the topic of the war in Ukraine (see Chapter 2), which is characterized by sustained geopolitical discussions, LGBTQ-related discussions reflect moral, identity, and rights-based controversies that often **trigger sharp spikes in engagement.**

Engagement levels show a disproportionate concentration on Twitter/X, where 173,261,481 total interactions were generated, compared to 26,944,882 on Facebook. At first glance, this suggests that Twitter/X dominates the conversation about LGBTQ+ issues. Yet when the number of posts is taken into account, **both platforms demonstrate a remarkably similar level of efficiency per post.**[90]

Twitter/X dominates in total engagement largely due to its significantly higher posting volume, which reflects its role as a fast-paced arena for live commentary, activism, and media dissemination. Facebook, although generating far fewer posts, achieves a slightly higher average engagement per post (392.95) and the highest recorded weekly EPP of the entire dataset (619.42 in 2025-W29). This suggests that LGBTQ+-related posts on Facebook—particularly those that go viral—achieve intense audience reactions, potentially driven by sharing within personal networks and community pages.

This places the LGBTQ+ discourse at the top in terms of attention density, surpassing Ukraine (EPP: Facebook ~272, Twitter/X ~186). **The intensity of engagement is not only high but also sharply peaked**. On Facebook, the highest EPP was recorded in week 2025-W29, with 619.42 interactions per post, while Twitter/X's peak occurred earlier, in week 2025-W14, at 558.55 (Figure 2)

[90] Specifically, the average engagement per post (EPP) reached 386.51 on Twitter/X and 392.95 on Facebook, the highest average values across all three topics studied.

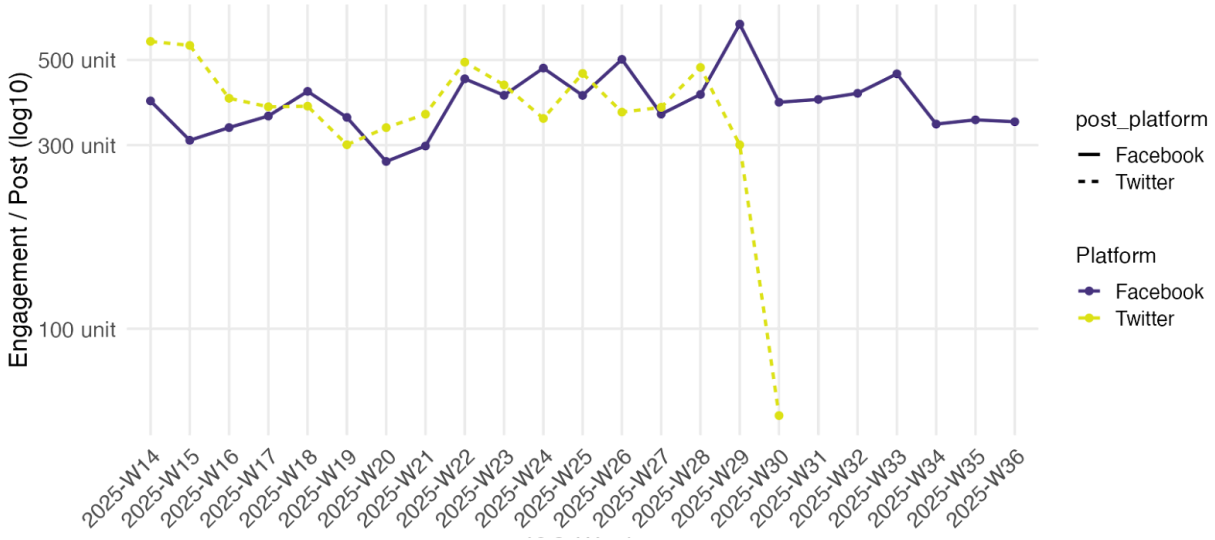


Figure 2: Weekly engagement per post – Topic LGBTQ+

There is no temporal synchronization between platforms: Facebook's peak appears in late July (W29), while Twitter/X's occurs in early April (W14). Though this would require external validation, asynchronous peaks suggest that attention to LGBTQ issues is not triggered by a shared media event across platforms but is instead driven by platform-specific viral moments—possibly linked to local political debates, pride events, or controversies.
Overall, online discussions on LGBTQ+ issues suggest a shift toward high emotional engagement, as shown in the narratives, persuasion techniques, and rhetorical devices mobilised by disinformation actors (but not only) when interacting on social media.

3.5.2. Top influencers

Twitter/X[91]

The LGBTQ+ discourse on Twitter/X is strongly polarized and shaped by conservative or reactionary voices. Accounts like *libsoftiktok*, *elonmusk*, and *jk_rowling* appear most frequently among the top influencers, indicating that much of the attention is driven by critical or confrontational positions rather than advocacy efforts. Supportive voices exist, but are less coordinated and less centrally influential.

Account Typology	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Most followers	J.K. Rowling (*@jk_rowling*)	Libs of TikTok (*@libsoftiktok*)	Matt Walsh (*@mattwalshblog*)	Collin Rugg (*@collinrugg*)	Billboard Chris (*@billboardchris*)	Gays Against Groomers (*@againstgrmrs*)	Sall Grover (*@salltweets*)	Jeff Younger (*@jeffyoungershow*)	Dr. Kevin M. Young (*@kevinmyoung*)	ralph 🏳️‍🌈 (*@therocketralph*)
Most engagement	Libs of TikTok (*@libsoftiktok*)	Sall Grover (*@salltweets*)	Gays Against Groomers (*@againstgrmrs*)	Dr. Kevin M. Young (*@kevinmyoung*)	Billboard Chris (*@billboardchris*)	ralph 🏳️‍🌈 (*@therocketralph*)	Jeff Younger (*@jeffyoungershow*)	Matt Walsh (*@mattwalshblog*)	J.K. Rowling (*@jk_rowling*)	Collin Rugg (*@collinrugg*)

[91] Network data were rich for Twitter/X: interaction graphs on Twitter/X allowed for detailed weekly network construction, modularity measurement, and the identification of hub actors who repeatedly appeared in the top influencer lists across several weeks.

Most shares	Libs of TikTok (*@libsoftiktok*)	Gays Against Groomers (*@againstgrmrs*)	Sall Grover (*@salltweets*)	Jeff Younger (*@jeffyoungershow*)	Billboard Chris (*@billboardchris*)	Dr. Kevin M. Young (*@kevinmyoung*)	J.K. Rowling (*@jk_rowling*)	Collin Rugg (*@collinrugg*)	ralph 🏳️‍🌈 (*@theroc ketralph*)	Matt Walsh (*@mattwalshblog*)
Most reactions	Libs of TikTok (*@libsoftiktok*)	Sall Grover (*@salltweets*)	Dr. Kevin M. Young (*@kevinmyoung*)	Gays Against Groomers (*@againstgrmrs*)	ralph 🏳️‍🌈 (*@theroc ketralph*)	Billboard Chris (*@billboardchris*)	Matt Walsh (*@mattwalshblog*)	Jeff Younger (*@jeffyoungershow*)	J.K. Rowling (*@jk_rowling*)	Collin Rugg (*@collinrugg*)

Among distributors and high-reach pages, **Libs of TikTok (@libsoftiktok)** stands out for sheer impact: it tops engagements, reactions, and shares within the top-10 and operates as a cross-platform brand with rapid video clipping and reposting that favors immediate amplification. **J.K. Rowling (@jk_rowling)** and **Matt Walsh (@mattwalshblog)** provide very large follower bases, giving the ecosystem significant audience reach, while **Collin Rugg (@collinrugg)** plays the role of influencer-aggregator, frequently surfacing fast-moving items to a broad audience. Mainstream TV/radio desks are largely absent here. Instead, activist brands and public figures anchor both visibility and spread of anti-LGBTQ+ discourse.

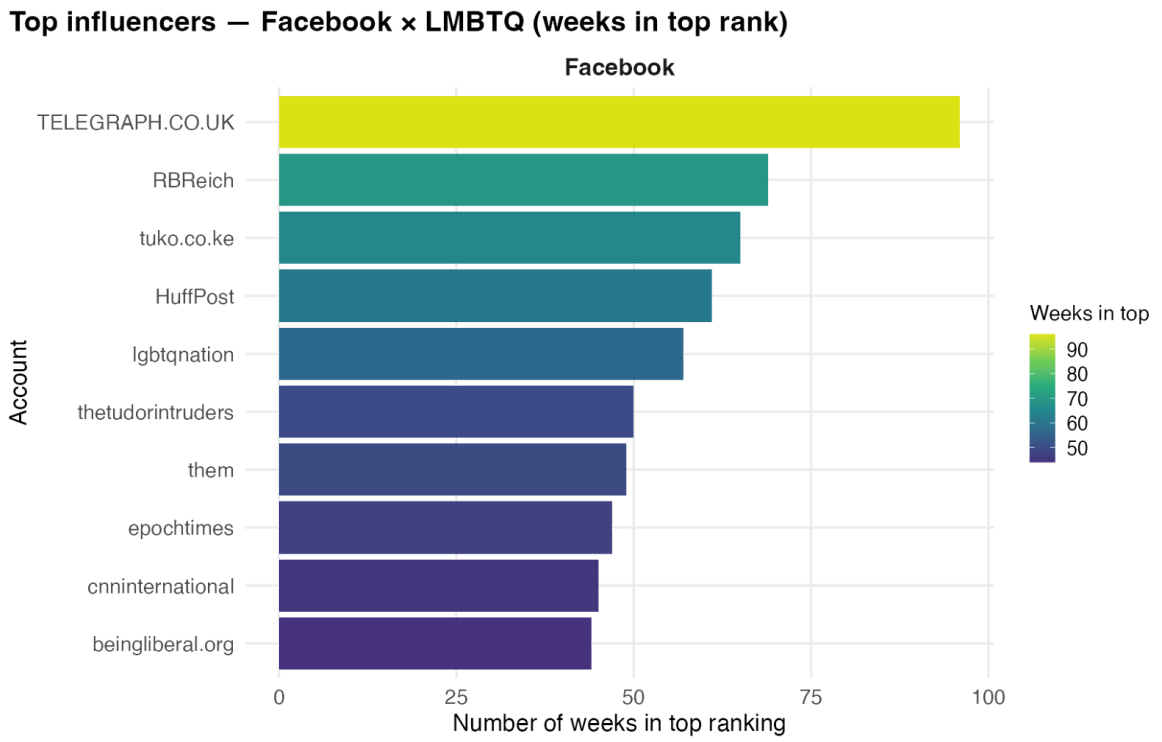
Within activist organizations and aligned personalities, a compact cluster drives most of the engagement. **Gays Against Groomers (@againstgrmrs)** and **Sall Grover (@salltweets)** consistently rank near the top for engagements and shares, often acting as second-tier amplifiers behind Libs of TikTok. **Billboard Chris (@billboardchris)** and **Jeff Younger (@jeffyoungershow)** also feature prominently, with steady posting that translates into reliable reactions and frequent resharing. The distribution of shares is notably steep—far higher for Libs of TikTok than for peers—indicating a hub-and-spoke pattern where one account seeds or curates content that others pick up.

Among individual figures, **Rowling** and **Walsh** underscore the split between **reach** and **activation**: both have many followers, but their engagement is typically outpaced by activist accounts who opt for high-frequency posting of short formats. Two outliers illustrate the dynamics of episodic virality: **Dr. Kevin M. Young (@kevinmyoung)** posts very little yet registers high reactions, suggesting one or two posts spiking to the top of the feed.

Overall, the LGBTQ+ top-10 is **activist-led rather than media-led**. High-reach public figures supply audience ceilings, but activist brands and influencer-aggregators—especially Libs of TikTok, followed by Gays Against Groomers and Sall Grover—are the primary drivers of engagement, sharing, and real-time diffusion in our corpus. Aggregator-style accounts such as Collin Rugg help bridge broader audiences, while low-volume spikes from figures like Dr. Kevin M. Young can briefly reorder attention. In short, agenda-setting and amplification are concentrated in a small cluster of activist/influencer accounts that excel at rapid, cross-platform distribution and engagement optimization.

Facebook

In contrast, the LGBTQ+ narrative on Facebook is dominated by media organizations and professional news sources. Pages such as *TELEGRAPH.CO.UK*, *HuffPost*, *lgbtqnation*, and *RBReich* rank among the top actors, suggesting that the conversation is structured more around journalistic content and less around individual influencers. The narrative here is less polarized and more embedded in mainstream media flows.



3.6. Conclusion

The analysis shows that anti-LGBTQ+ disinformation is less a collection of separate narratives than a tightly interwoven field. The same core storylines recur across platforms and communities: LGBTQ+ people have a corrupting influence on "pure" societies, an imagined "gender ideology" has captured institutions, LGBT ideology is threatening the "natural" family order, etc. These strands constantly overlap and borrow from each other rather than appearing in isolation. Sports, schools, and healthcare act as entry points: once "fairness," "protecting children," or "common sense" are invoked, it becomes easier to plug in adjacent claims about institutional capture, demographic decline, or moral decay.

Across these narratives, the same persuasion techniques and rhetorical devices appear over and over again, be it when discussing schools, women's sports, or hospitals. These strategies are explicitly emotional: they seek to trigger fear (for children, for bodily safety, for social order), anger and resentment (towards "elites," institutions, and activists), disgust (through pathologisation and mutilation imagery), and a diffuse sense of loss of control. Pro- and anti-rights actors draw on very similar tools and emotional triggers, but in opposite directions, strongly indicative of polarisation.

Finally, platform and network dynamics matter for how these narratives travel. On X, the space is activist-led rather than media-led: a compact cluster of accounts sets the pace, including visible public personalities such as J.K. Rowling. On Facebook, by contrast, the LGBTQ+ conversation is anchored more in media outlets and professional pages, even if anti-LGBTQ+ frames are grafted into broader threads about elections, corruption or Ukraine. Both platforms however show similarly high engagement per post on LGBTQ+ content, signalling that this topic is discussed intensely. In summary, a small number of overlapping narratives, carried by a shared repertoire of emotionally loaded techniques and amplified by different influencer ecologies across platforms, shapes not only how LGBTQ+ rights are discussed, but also how wider political and geopolitical questions are 'moralised' online.

4. FACT-CHECKERS ON THE FORCE OF DISINFORMATION IN EUROPE: ELECTION LEGITIMACY WHISPERS, THE DOMESTICATED WAR IN UKRAINE AND THE CULTURE-CLASH FRAMES OF LGBTQI+

Merle van Berkum *(Erich-Brost-Institut) and* ***Sara Mercereau*** *(Opsci) brought together six fact-checking organizations – Les Surligneurs (France), Re:Baltica (Latvia), Delfi Lithuania (Lithuania), Facta News (Italy), Delfi Estonia (Estonia) and Euractiv Romania (Romania) – to delve into the subject of disinformation and journalism.*[92]

The aim of interviews of said fact-checking organizations is to better capture country-specific disinformation dynamics — together with the narrative structures and rhetorical strategies that sustain them — and to understand the journalistic and fact-checking practices developed in response. The findings of Merle and Sara are shared in a conversation with Rīga Stradiņš University (RSU), highlighting both the commonalities and the country-specific particularities of disinformation mechanisms, trends, and impacts.

4.1. Country dynamics & shared patterns

RSU: *You mentioned that PROMPT focuses on disinformation narratives across three main themes. Looking at the broader picture, which* ***narratives*** *stand out as most dominant across the six countries within these themes?*

MVB & SM: We observe three recurring issues: election distrust, LGBTQIA+ culture-war frames, and the war of aggression against Ukraine reframed as a domestic cost/risk analysis. Election distrust refers to narratives that undermine the integrity or legitimacy of voting systems and democratic processes. LGBTQ+ culture-war frames use gender and sexuality as polarizing wedges to fuel broader social conflict. Ukraine as domestic cost/risk highlights how the war is reframed not in terms of geopolitics, but as a burden on national security, economy, or everyday life. These are shared, but each country activates them through local levers. That's why the same topic – election distrust – can have very different materializations, such as e-voting opacity in Estonia, climate change skepticism and Green Deal resentment in Latvia, short-lived rumours questioning the legality of candidates or results in Lithuania, elite-conspiracy/EU-failure frames in France, and recurrent hoaxes about military mobilisation for the war in Ukraine in Romania. Beyond our topics of analysis, climate and vaccines were also recurring disinformation themes, with journalists from France and Italy highlighting vaccine disinformation as a persistent and emotion-heavy disinformation topic.

RSU: *Let's start there. We know that during election campaigns, disinformation can target many different issues, from electoral processes themselves to adjacent topics linked to the economy, the environment, health... What main trends have you observed across the countries you analyzed?*

MVB & SM: Disinformation during electoral procedures is a recurring feature in all analyzed countries. However its intensity and impact vary widely depending on national contexts. In Estonia, e-voting technology is repeatedly questioned, namely through the argument that Estonia is the only country in Europe using e-voting at such a scale. In Latvia, claims of stolen

[92] Throughout the discussion, the names of the countries are used interchangeably with those of the fact-checking organizations representing them, on the rationale that each organization's insights reflect the disinformation dynamics of its national context.

or incorrectly counted votes emerged during the 2024 European Parliament elections, and reemerged in the 2025 municipal elections, with several parties now using these claims as routine campaign fuel.

In Lithuania, election distrust tends to be short-lived: after the presidential elections, there were a few online posts questioning whether President Gitanas Nausėda had met the legal requirements to run in the elections, but these rumours did not gain sustained traction.

Claims of rigged elections and/or election fraud also occur in France. Disinformation regarding electoral processes is however mostly framed less as procedural fraud than as an elite conspiracy/EU failure: the main narratives blame national political and economic elites or the European Union for manipulating outcomes, betraying citizens' interests, or eroding national sovereignty.

In Romania, electoral moments trigger fear spikes (war, mobilisation rumours), anti-elite talk and nostalgia of the communist era. In Italy, disinformation about vaccines, climate, and immigration is often amplified by mainstream media. This makes fact-checkers' work harder during election cycles.

RSU: *PROMPT also focuses on* ***LGBTQIA+ disinformation****. What did you learn there?*

MVB & SM: What really stood out is that LGBTQ+ disinformation doesn't carry the same weight everywhere. You can think of it as a spectrum: in some countries it plays a central role, in others it's almost absent. At the high-end lies Romania, where it remains persistent, bound up with faith and family frames, and regularly deployed against political opponents. In Lithuania, it tends to come in spikes — such as during debates on the partnership law, when "protect the children" or "protect the family" claims are especially prominent — but it subsides once the debate fades.

In Latvia, it's more situational and meme-driven: the Istanbul Convention was reframed as an "LGBT law," and after the elections "rainbow coalition"[93] memes circulated. In France, it tends to revolve around personalities and elites — for instance, the Brigitte Macron rumor, which continues to resurface. In Estonia it's much more muted since marriage equality laws passed in 2024, despite some occasional flare-ups around debates on trans athletes and 'fairness' in women's sports.

Finally, Italy often imports trans/LGBTQ frames from the US debate. Because LGBTQ+ issues are less visible in mainstream debates, dog-whistle memes[94] and generalization travel without encountering many counter-voices in mass media. Across these contexts, though, the underlying frames remain strikingly similar, embedded in claims such as "protect the children" or "this is being imposed by elites." What varies is the timing and the format. At times, the trigger is a legislative calendar; at others it takes the shape of memes or "gossip" about individuals. These factors determine whether the issue dominates the conversation or just stays at the margins.

RSU: *And regarding* ***Ukraine****?*

[93] After a coalition involving two ideologically opposite parties in Latvia was formed, memes emerged depicting politicians kissing with rainbow imagery, falsely suggesting a "gay coalition" or "rainbow coalition".
[94] Dog-whistle memes are humorous or ironic images that use coded symbols, phrases, or in-jokes to convey hostile or exclusionary messages (for example against LGBTQIA+ people) in a way that is clearly understood by in-groups, while remaining deniable as "just a joke" to wider audiences.

MVB & SM: Everywhere, the foreign war is domesticated in the digital space, reframed from a distant geopolitical conflict into immediate kitchen-table concerns, primarily related to the cost of living, safety, and identity. A foreign topic thus becomes more personal and local. In Lithuania, where the war in Ukraine is said to be the main disinformation topic, narratives often claim that Lithuania's support for Ukraine comes to the detriment of Lithuanian citizens; and that aid organizations are corrupt. Other prominent narratives include Soviet nostalgia, victim-blaming of Ukrainian soldiers, mockery of Western politicians, and claims that Lithuanians are fleeing the country due to fear of war. In Estonia, where the topic is highly salient, propaganda aims to turn the public against NATO, aid to Ukraine, and Ukraine's potential EU membership. Online voices often invoke the impact of the war on the cost of living; and raise the fear of an escalation of the conflict into a continent-wide war. In Latvia, there are fewer concrete false stories about Ukraine that can be clearly debunked, but fact-checkers observe a persistent anti-Ukrainian tone woven into broader anti-government grievances.

During the 2024 European Parliament elections, disinformation spread about the EU's actions in Ukraine. In Romania, anti-West/EU narratives are paired with mobilisation hoaxes, - false claims that ordinary citizens would soon be forcibly conscripted and sent to fight in Ukraine. Narratives about the war in Ukraine primarily focus on fear-mongering and questioning intervention, rather than the war itself, due to initial public empathy for refugees. In France, Ukraine resurfaces only episodically, with narratives often criticising EU institutions and falsely linking Ukraine to Nazism — though, since October 2024, attention has been eclipsed by the war in Gaza. As for Italy, networks that formed around vaccines frequently shift back and forth to the topic of Ukraine. Much of the content is imported from US/Russian ecosystems and staged on Telegram before being more widely circulated.

4.2. Platforms, flows, and rhetoric

RSU: *Now that we have an idea of the substance of disinformation campaigns in these countries, let's explore where this content actually travels. In this line, which* ***platforms or types of media*** *play the biggest role in spreading disinformation in each country?*

MVB & SM: In Estonia, the infosphere is influenced by language: Estonian-language audiences are on Facebook and Instagram; Russian-speakers are on TikTok and Telegram. In Latvia, TikTok is at the frontline because, unlike Facebook or Instagram, there is no structured third-party fact-checking partnership on the platform, so no local team systematically reviews or labels viral false content. Telegram is central to Russian-language channels. Lithuania still anchors reach on Facebook/YouTube, with TikTok rising; Kremlin-adjacent "alt-news" sites seed disinformation on social media, and a handful of super-spreaders drive virality.

In France the model is "public-towards-private" messaging. Seeds are planted on X, then rumors consolidate on WhatsApp and Telegram channels. Romania shows multi-channel synchronicity between TV, radio, TikTok and Telegram. There churches and opinion leaders act as amplifiers, and parts of the diaspora re-circulate the disinformation content. In Italy, Telegram has become the primary origin hub where disinformation networks are built and coordinated as many disinformation influencers moved there after being banned from Facebook. Nevertheless, the latter platform is still used to amplify the content to a wider audience.

RSU: *You mentioned the* ***Romanian diaspora as an amplifier of disinformation.*** *In what ways does the community abroad help reinforce these narratives?*

MVB & SM: Diaspora communities were mainly mentioned by the Romanian fact-checking organization. The Romanian team flagged the diaspora as especially vulnerable to

disinformation because of distance, precarious work abroad, and the search for a community. These provide openings for disinformation emotive frames targeting the West, mobilization fear - i.e., spreading fear of being forcibly mobilised for the war in Ukraine - and the (lack of) morality of LGBTQ+ activists and laws. The same messages surface, often simultaneously, across TikTok, TV and radio commentary, Telegram, and church networks. Because much of it is wrapped in humor or satire, it feels "safe to share", while the diaspora serves both as target and relay, feeding material back into domestic feeds. It's a good example of how distribution channels -not just content - shape the impact of disinformation.

RSU: *Recently some experts have been discussing the emerging weight of* ***"disinfotainment"*** *- disinformation presented in entertaining formats (memes, humor, satire, influencers) - in sharing and spreading disinformation. How present do you think this phenomenon is in your countries of analysis, and what is its role in the dissemination of disinformation?*

MVB & SM: Disinfotainment mostly works as a force multiplier, in that it does not usually create new narratives of its own, but strengthens existing narratives instead and circulates them faster. Memes, short clips and irony lower the social and reputational cost of sharing and make claims harder to debunk. For example, in Latvia, meme-mockery is considered a problem: disinformation often takes the form of sarcastic images or short ironic posts on TikTok and Facebook, especially in the Russian-language spaces, where Ukraine support is framed as absurd or Latvia as a "failed state." In Romania, disinfotainment often takes the form of talk show clips, which evoke similar narratives as those spread on social media, generating near-simultaneous cycles across broadcast and social media.

France has a distinct humor layer moving from X to private messaging groups. Lithuania describes meme-mockery as "definitely a problem" and often impossible to fact-check. Estonia sees spikes within specific communities when a meme wave crests. That format advantage helps both LGBTQ+ frames and election/Ukraine claims travel fast. Italy, in turn, witnesses the rise of coordinated meme communities (e.g.mattonisti) who deliberately organize online to push disinformation through memes which often look harmless on the surface but carry hidden or coded messages ('dog whistles') understood only by in-groups, which gives them plausible deniability. Furthermore, comics and stand-up comedy are also being used to spread dangerous narratives under the guise of humor, making it hard for fact-checkers to intervene.

4.3. Who creates it vs. who spreads it

RSU: *Much attention has been paid to the authors of disinformation narratives, but some argue that the* ***propagators*** *– rather than creators - of those narratives have the most prominent role in its* ***dissemination****. What did you find on the creators and propagators of disinformation in those countries?*

MVB & SM: Most of the fact-checkers we spoke to describe it as a kind of chain reaction — starting "upstream" and then flowing "downstream". Upstream are Kremlin-linked or Kremlin-friendly outlets, like those "alternative news" sites in Lithuania or the foreign portals flagged by the Estonian team. Italy, which is said to operate as a *"second-level market"* for disinformation, imports US and Russian frames on topics such as woke/cancel culture or NATO/Ukraine, which are then translated (sometimes poorly, even with grammar mistakes) and adapted into the Italian context.

Downstream, the dynamics vary by country. In Latvia, you see political party pages or individual politicians picking up these claims and making them part of routine campaigning. In France, they often begin as rumors on X before migrating into private spaces like WhatsApp or

Telegram, where they really gain traction. In Romania, the same narrative can show up almost at the same time on TV, radio, TikTok, and Telegram — blending mainstream credibility with viral speed. And quite often, the same actor isn't merely repeating a line but coining it and then amplifying it across its entire network.

RSU: *Russia has been identified as one of the major countries authoring and propagating disinformation in European countries. Drawing on your contact with fact-checkers, how would you characterize* ***Russia-linked disinformation*** *across the countries?*

MVB & SM: From what the fact-checker teams told us Russia plays two roles at once: as a source of frames and as a pipeline others tap into. You see state- and para-state storylines along with Kremlin-adjacent "alt-news" sites setting the tone — Ukraine fatigue, NATO as a risk, the EU as weak or overbearing—but those lines don't land as "Moscow says…". Instead, they're quickly localized: reframed as bills, safety concerns, or identity issues, and pushed through whichever channels prove most effective in each place.
The mechanics differ by country. In Estonia, fact-checkers describe a consistent pattern where local actors draw from Russian portals (alongside some US and Hungarian sources) and repackage the material — often around anti-NATO themes or cost-of-living comparisons that cast support for Ukraine as a domestic burden. In Latvia, even after Russian TV and major sites were cut off, the flow shifted to TikTok and Russian-language Telegram; there, party pages and public figures pick up the narratives and weave them into campaign messaging. In Lithuania, Kremlin-adjacent "alt-news" outlets seed the stories, and a small group of super-spreaders can propel them onto Facebook, YouTube, and increasingly TikTok, with Russian-speaking audiences proving the hardest to reach with corrections.

France is different: the Russia label is less explicit; with narratives often seeded on X before hardening in WhatsApp and Telegram groups as EU-burden or elite-conspiracy talk. In Romania, pushes are nearly simultaneous across TV, radio, TikTok, and Telegram — including mobilization hoaxesv—vwith the diaspora serving as both a target and a relay back into the domestic sphere. Italy picks up a lot of narratives originated in Russia, namely Russian frames about NATO and Ukraine, which are then translated and adapted into the Italian context.

So, yes, Russia plants a lot of the seeds—but traction comes from local hands. Politicians, party pages, influencers and admins translate, time, and format those narratives for their own audiences. This is why the same core ideas can look like anti-NATO "peace" talk in Tallinn, TikTok-first election slurs in Riga, super-spreader lifts in Vilnius, elite-conspiracy riffs in Paris, and broadcast-to-social choruses in Bucharest.

RSU: *Given their geographical and historical contexts, the* ***Baltics*** *are particularly in the crosshairs of Russian disinformation. How would you describe the way Russian disinformation operates in the Baltics specifically?*

MVB & SM: All three countries share Russian-language pipelines and the rise of short-video and memes, but the hooks differ. In Estonia, recurring attacks target e-voting and amplify anti-NATO escalation frames in Russian-language TikTok and Telegram channels. Latvia reports institutionalized campaign disinformation — mixing process fraud with Brussels resentment — circulating on TikTok and in AI-edited memes, with Telegram central for Russian-speaking audiences. In Lithuania, entrenched super-spreaders and Kremlin-adjacent "alt-news" sites seed material that is later repackaged for Facebook, YouTube, and increasingly TikTok. Across all three, broadcast restrictions didn't stop the flow; they displaced it onto social and messaging platforms.

4.4. Practices, tools, impact—what helps, what's missing

RSU: *Finally, what kinds of* ***tools*** *are fact-checking teams actually using in their work, and what do they say about* ***impact*** *and* ***unmet needs****?*

MVB & SMa: Across countries, the basics are similar: teams rely heavily on OSINT techniques like reverse image search, archiving, and geolocation. Some also use specialized tools to identify AI-generated content, but all interviewees stressed that AI is not trusted for verification tasks such as confirming authenticity.
Estonia relies on basic information analysis tools like Excel, but also increasingly experiments with AI-based tools such as Google's Notebook LM for summarization and analysis. Larger datasets are handled with the support of data journalists using Python or R. Still, they emphasize financial and data-access barriers and expressed a need for more sophisticated partners or tools to analyze the origin of narratives. Latvia explained that Meta's dashboard for Facebook has lost much of its value, leaving most social media monitoring as a manual process. They are particularly concerned about disinformation on TikTok — especially among Russian speakers — and are hoping for an AI-powered monitoring tool like the one their Romanian colleagues are developing. Lithuania uses InVID and WeVerify – tools developed by fellow EU projects focusing on tackling disinformation -, archiving tools and both Google and Yandex reverse image search, alongside an in-house monitoring tool developed by MATA. But they note that general monitoring tools work poorly with smaller and more complex languages such as Lithuanian. They also struggle to reach audiences who most need fact-checking, as Russian-speaking communities often resist their content.

In France, Les Surligneurs rely mainly on manual processes and collaborative spreadsheets for issue tracking. They are currently developing a pattern-checking tool to automatically flag when new statements repeat previously fact-checked legal or political claims, linking them to prior analyses. They also expressed interest in a monitoring tool to track specific personalities or groups by topic, though the sheer volume of subjects might limit its practical use compared to their pattern-based approach.

Romania underlined the need for AI tools that can both track viral disinformation in real time and generate counter-content. They see partnerships and access to monitoring platforms like Osavul as crucial, given the current lack of instruments to effectively counter large-scale false narratives.

Italy depends on open-source tools for reverse image searches, satellite mapping, transcription, and translation. They also experiment with participatory formats such as a WhatsApp chatbot where readers can send questions or links. What they miss most is reliable access to platform data: since the decline of tools like CrowdTangle and advanced searches on X and Facebook, in-depth investigations have become increasingly difficult.

The bigger picture is that everyone agrees fact-checking alone isn't enough. Impact is measured through traffic and engagement, whether debunked posts get removed, and how far broadcast segments travel. Some teams are experimenting with short-form video to reach younger audiences. But interviewees consistently emphasized that fact-checking needs to be combined with media literacy, stronger newsroom ties, and platforms that actually enforce their own rules. The hardest part remains the same everywhere: disinformation narratives are increasingly circulating in private groups, where measurement and countering are far more difficult.

RSU: *Summarizing everything we've discussed, how do these findings advance PROMPT's mission and strengthen the project overall?*

MVB & SM: Looking at the bigger picture, these interviews add two important layers for PROMPT. First, they *deepen* the signals we already track. Instead of treating narratives as just keywords, they help us refine our country-specific maps of the frames (issues of fear, identity, belonging) that make stories stick — why e-voting doubt gains traction in Estonia, why "protect the children" resonates in Lithuania at specific legislative moments, why disinfotainment spreads so easily, and how diaspora or private-group dynamics shape the felt impact.

That context strengthens our ability to train the PROMPT AI-tool to detect not just strings of text but also narratives, rhetorical moves (as interviewees themselves described them), and local triggers (election calendars, legal debates, energy prices).
Second, they translate practice into product. We heard very concrete needs — TikTok/RU-language lanes in the Baltics, super-spreader mapping in Lithuania, X to WhatsApp/Telegram hand-offs in France, broadcast ↔ social synchrony in Romania, small-language constraints—so we can priorities features like cross-language tracking, "synchronized drops" alerts, and problematic actor monitoring.

CONCLUDING REMARKS

Disinformation operates through coordinated, adaptive mechanisms. It exploits the architecture of digital platforms, the vulnerabilities of democratic institutions, and the emotional predispositions of audiences. Our analyses of Moldova's parliamentary election, the war in Ukraine, and LGBTQ+ debates (Chapters 1-3) demonstrate that these campaigns are not isolated incidents but interconnected strategies designed to undermine trust, polarise societies, and manipulate political outcomes. Reflections from practitioners (Chapter 4) help contextualise these observations across different national contexts and digital habits.

In the Moldovan election, hybrid interference combines geopolitical objectives with digital manipulation. Disinformation narratives are not merely about electoral preferences; they are embedded in broader frames of sovereignty, security, and cultural identity. By leveraging local grievances and amplifying them through transnational networks, adversarial actors transformed a domestic electoral process into a proxy battlefield for regional influence. This case underscores the fragility of small democracies and the need for tailored resilience strategies that address both the technical and societal dimensions of vulnerability.

The war in Ukraine remains the most persistent source of disinformation across the analysed platforms. Unlike episodic electoral campaigns, conflict-driven narratives exhibit remarkable durability. Narratives often combine issues—linking military developments to energy security, economic hardship, and migration fears—creating a multidimensional frame that sustains engagement. Their endurance highlights the challenge of combating disinformation in protracted crises, where the informational battlefield becomes as enduring as the physical one.

LGBTQ+ narratives further illustrate how identity politics are weaponized to deepen societal divides. These campaigns often frame LGBTQ+ rights as existential threats to traditional values or national sovereignty, using emotional language to provoke outrage. They are amplified through coordinated networks across platforms, creating the illusion of widespread dissent. By exploiting cultural sensitivities, disinformation actors redirect public attention from governance and policy issues toward manufactured moral conflicts, making identity-based polarization a strategic tool for destabilization.

The report also shows that the topics of Ukraine and LGBTQ+ circulate differently across social media. The topic of the war in Ukraine is the most sustained and extensive conversation across platforms,[95] with overall stable engagement. The discourse is highly networked, centralized around key accounts, and unfolds as a continuous, evolving conversation rather than episodic bursts. The LGBTQ+ topic differs in scale but not in intensity - it displays the highest, volatile, engagement. This suggests that LGBTQ+ debates trigger highly emotional, polarizing, and shareable content, leading to strong user reactions, especially around specific events (e.g., Pride Month, legislative controversies). These differences also reflect platform-specific logics. Twitter/X is consistently the primary site for networked political diffusion; and Facebook produces high engagement without visible reshare networks. Furthermore, engagement peaks are not synchronized across platforms, indicating that each platform follows its own temporal attention patterns.

Last but not least, the report shows that narratives are hybrid, blend different themes - sovereignty, security, and identity - to ensure adaptability across contexts and maximum reach. These "meta-stories" are similar, yet come with variants, which facilitate their propagation.

[95] Over the analysed period (April–September 2025), it generated more than 4.5 million posts on Twitter/X, 454,121 on Facebook, and 80,518 on Bluesky, with consistently high engagement levels.

TECHNICAL APPENDIX

This appendix details the PROMPT methodology. It also presents the additional technical parameters to obtain and process different sources of data for Chapter 1 - MOLDOVA'S 2025 PARLIAMENTARY ELECTIONS: DISINFORMATION AS A GEOPOLITICAL BATTLEGROUND.

T1 - PROMPT DATA COLLECTION AND PROCESSINGS

To support open science and research transparency, and allow replicability and uptake in the disinformation ecosystem, this appendix describes the main mechanisms behind the PROMPT online dashboard.

- Data collection, from query formulation to data storage
- Processings: from data cleaning to producing outputs

Both of these features support the core tools of the PROMPT dashboard:

- The Corpus Analyser
- The Disinfo Scanner
- The Wikipedia Sensitivity Meter
- The Wikipedia Sensitivity Barometer

Data collection

PROMPT works across 3 topics, 6 country case-studies, 8 platforms and 8 languages. It analyses textual input from different social media platforms, and interactions around these items (likes, reposts, etc.) between 1st January 2024 to October 2025 (and ongoing). The below figure summarizes the parameters for data collection within the project:

	European elections	War in Ukraine	LGBTQI+ rights and freedoms	
CORPUS				**PLATFORMS**
France				X
Italy				TikTok
Romania				YouTube
Estonia				Facebook, Instagram
Latvia				Wikipedia
Lithuania				Bluesky
Global				Telegram

Data are needed both for (1) model training and (2) data analysis to support the work of analysts.

Data collection for training

To train AI-models, access to quality datasets - multilingual, disinformation-focussed - for training is limited. Several benchmarked and standard datasets are however available and commonly used for benchmarking.

Name	Description	Link	Rationale
MultiClaimNet	Academic dataset containing disinformation claims	https://zenodo.org/records/15100352	Used to test the agentic and embedding-based model for disinformation detection
Twitter/X "Sunset" Ukraine-Russian Crisis Dataset	Twitter/X "sunset' dataset on the topic of the war in Ukraine	https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows	Used to develop the dynamic network analysis modelling and test the model embeddings
EUvsDisinfo Dataset	Dataset of Kremlin-back disinformation social posts and news items	https://euvsdisinfo.eu/disinformation-cases/	Used to test the semantico-axiological matrix, to benchmark embeddings in narrative detection; and benchmark LLMs in rhetorical detection

Data collection for analysis

With the exception of Wikipedia, collecting data that is representative of the prevalence of disinformation on social media platforms is a challenge. As captured in Chavalarias' analogy, while it is next-to-impossible to evaluate the depth of the digital sea, it is still possible to measure whether some given monitored (disinformation) content increases or decreases, becomes viral or not, across time and topics, by setting a fixed threshold. To do so, several challenges must be addressed:

- Access to social media platforms is uneven and greatly constrained by lack of VLOP cooperation despite the DSA (with the exception of Wikipedia). This has implications for comparative studies, as mentioned in the 1st report of the SIMODS project: "to compare how permeable each platform is to misleading content (...), indicators must be defined in a way that is comparable across platforms and stable over time so that progress, or deterioration, can be quantified".
- Access to quality metadata - including user demographics - is very limited, though recent EU rules should help address
- Systematic data collection across languages is complicated by:
    - the uneven distribution of user-generated content because of (1) digital habits (platforms are more or less popular across countries) and (2) population size and engagement on social media platforms
    - the cultural, geographic, national and linguistic markers enshrined in user-generated content - i.e posts on the topic EU elections are likely to debate different issues or people across 8 languages.

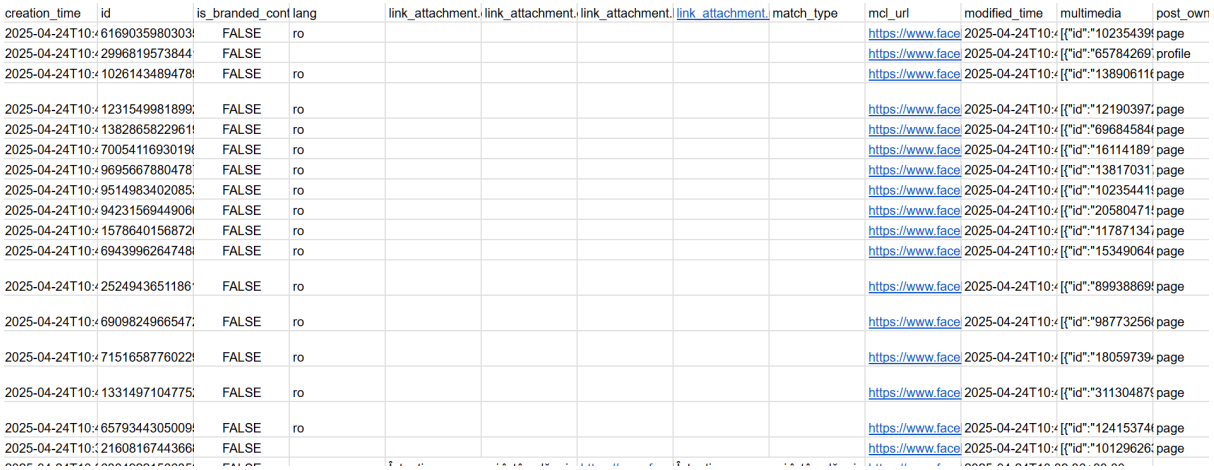
The full list of challenges and mitigation measures are detailed in the PROMPT White Paper, authored by the University of Urbino Carlo Bo - The State of Social Media Research APIs & Tools in the Digital Services Act Era.

To address these issues, PROMPT combines different approaches to collect social data. It has collected more than **6 million posts across 6-months window frames**, by using:

- lists of (problematic) accounts pre-identified by civil society activists and journalists

Fact-checkers, civil society activists and opsci.ai contributed lists of influential 'problematic accounts' to be included in the data collection. These influential accounts were determined on the basis of their social media visibility. Using a snowballing strategy, we added to this initial list throughout the project, enabling us to retrieve an increasing number of social media posts based on pre-identified known disinformation actors.

AI de Noi dataset	An independent, volunteer-driven civic project—engaged in bottom-up monitoring of inauthentic behaviour on Facebook (Romania), **manually flagging compromised groups and accounts** for takedown.	**Not public - terminated**	Used for analysis of propagation and coordination in the wake of the Romanian presidential elections



creation_time	id	is_branded_cont	lang	link_attachment.	link_attachment.	link_attachment.	link_attachment.	match_type	mcl_url	modified_time	multimedia	post_own
2025-04-24T10:4	61690359803035	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"10235439	page
2025-04-24T10:4	2996819573844	FALSE							https://www.face	2025-04-24T10:4	[{"id":"65784269	profile
2025-04-24T10:4	1026143489478	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"13890611	page
2025-04-24T10:4	12315499818992	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"12190397	page
2025-04-24T10:4	1382865822961	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"69684584	page
2025-04-24T10:4	70054116930196	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"16114189	page
2025-04-24T10:4	9695667880478	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"13817031	page
2025-04-24T10:4	9514983402085	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"10235441	page
2025-04-24T10:4	94231569449066	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"20580471	page
2025-04-24T10:4	15786401568726	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"11787134	page
2025-04-24T10:4	6943996264748	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"15349064	page
2025-04-24T10:4	2524943651186	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"89938869	page
2025-04-24T10:4	6909824966547	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"98773256	page
2025-04-24T10:4	71516587760225	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"18059739	page
2025-04-24T10:4	1331497104775	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"31130487	page
2025-04-24T10:4	6579344305009	FALSE	ro						https://www.face	2025-04-24T10:4	[{"id":"12415374	page
2025-04-24T10:3	21608167443666	FALSE							https://www.face	2025-04-24T10:4	[{"id":"10129626	page

Example of a pre-formed dataset relevant to the Romanian elections based on lists of social media accounts aggregated by the AI de Noi initiative

- Keyword-based queries

We selected approximately 50 keywords per language and topic - the war of aggression against Ukraine; LGBTQIA+ rights and issues, and EU and European national elections. These words were chosen for their

relevance to the public conversation on each of the topic and their connection with misleading claims or local issues in our target countries. Given various spellings, we developed short directories of each main keyword (e.g. LGBT/LGBTQ/LGBTQI etc) to ensure broader coverage.

The keyword lists were developed by opsci.ai and reviewed by country-experts and fact-checkers within our consortium. They were divided into two different queries whose results were merged:

- 'open' keywords and hashtags (by topic)

To ensure that our data collection covered each topic comprehensively, we developed an open keyword/hashtag search. This mixed usual disinformation keywords ('deep state') and neutral terminology ('LGBT').

(LGBT OR LGBTQ OR LGBTQI OR LGBTQIA OR gender OR transgender OR "non-binary" OR queer)(agenda OR narrative OR indoctrination OR hoax OR fraud OR propaganda OR ideology OR scam OR "deep state" OR extremists OR dictatorship OR lobby OR "moral decay" OR family OR school OR threat OR children OR tyranny OR decay OR wokism OR woke)

#LGBTPropaganda OR #AgendaLGBT OR #LGBTDictatorship OR #StopLGBT OR #GenderIdeology OR #RainbowLobby OR #TransScam OR #LGBTGrooming OR #ProtectOurKids

Example of an open query formulation for the topic 'LGBTQI+' in English

- 'closed' keywords and hashtags (by topic)

'Closed keywords lists' help cover compound words, popular tropes, names of public figures - we used an exact match search whenever possible. Some of these expressions are concepts used almost exclusively by users defending a world vision specific to them (e.g. "green madness"). Such catchphrases encapsulate a story or a theory in themselves and do not need to be combined with any other keywords. The use of such keywords harvests almost exclusively relevant publications, but requires frequent updates as new stories emerge.

("Valdis Dombrovskis" OR "Roberts Zīle" OR "Elīna Pinto" OR "Harijs Rokpelnis" OR "Ivars Ijabs" OR "Nils Ušakovs")

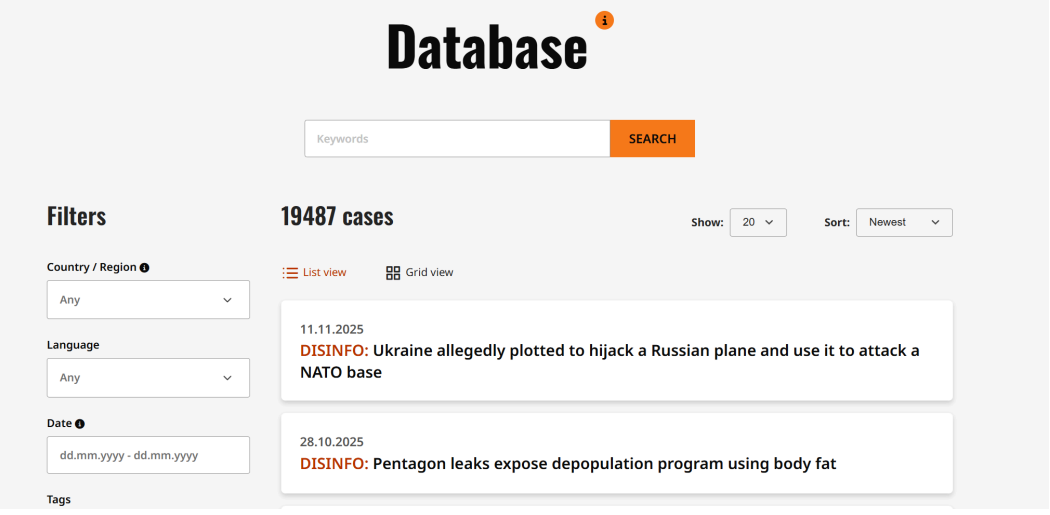
Example of a closed query on political leaders for the topic 'European Elections' in Latvian

("Climate colonialism" OR "European economic interests" OR "Neocolonial commercial practices" OR "Pseudo-Green ideology")

Example of catchphrases and concepts leading to relevant content.

- Full thematic pre-formed datasets relevant to the 3 topics

Whenever possible, PROMPT combined its own data collections with datasets collected by fact-checkers and analysts. This helped complement lists of social media accounts and web domains considered problematic in each country, thereby allowing to improve the data collection efforts.



EUvsDisinfo Database curated by the EEAS East Stratcom Task Force

Additional measures mitigate certain inherent challenges to social media data collection:

- To compensate for the lack of access to certain social media platforms (X in particular), PROMPT combined data collection through VLOP API programs for researchers with alternative scraping tools.

- To mitigate linguistic bias, "open keywords" keywords were translated identically across the eight languages of the study. "Closed keywords", on the other hand, were adapted to the national context in each country to ensure relevance - in the case of public political figures for the topic of national elections for example.
- To address the trade-off between recall (the extent to which all relevant content on the topic is retrieved) and precision (the extent to which the retrieved content is actually relevant to the topic), PROMPT worked with iterative query formulation - combining or separating queries into 'packages' or 'chunks'.
- In line with PROMPT's overview of social data collection using VLOP APIs, query formulation required distinct standards and iterative fine-tuning. While certain platforms permitted the use of Boolean operators and phrase-based queries, others restricted searches to simple flat keyword strings. As queries became broader, the risk of inconsistency and false positives increased, we balanced recall with precision and refined queries, while keeping the overall number limited in order not to exceed quota budgets and collection times, yet still ensuring sufficiently large datasets for analysis.

The case of Youtube

In the case of YouTube, data collection followed a two-fold process:

- results are first retrieved through search.list calls (10 quota units each)
- results are complemented by videos.list calls (1 unit per video) to obtain full engagement metrics.

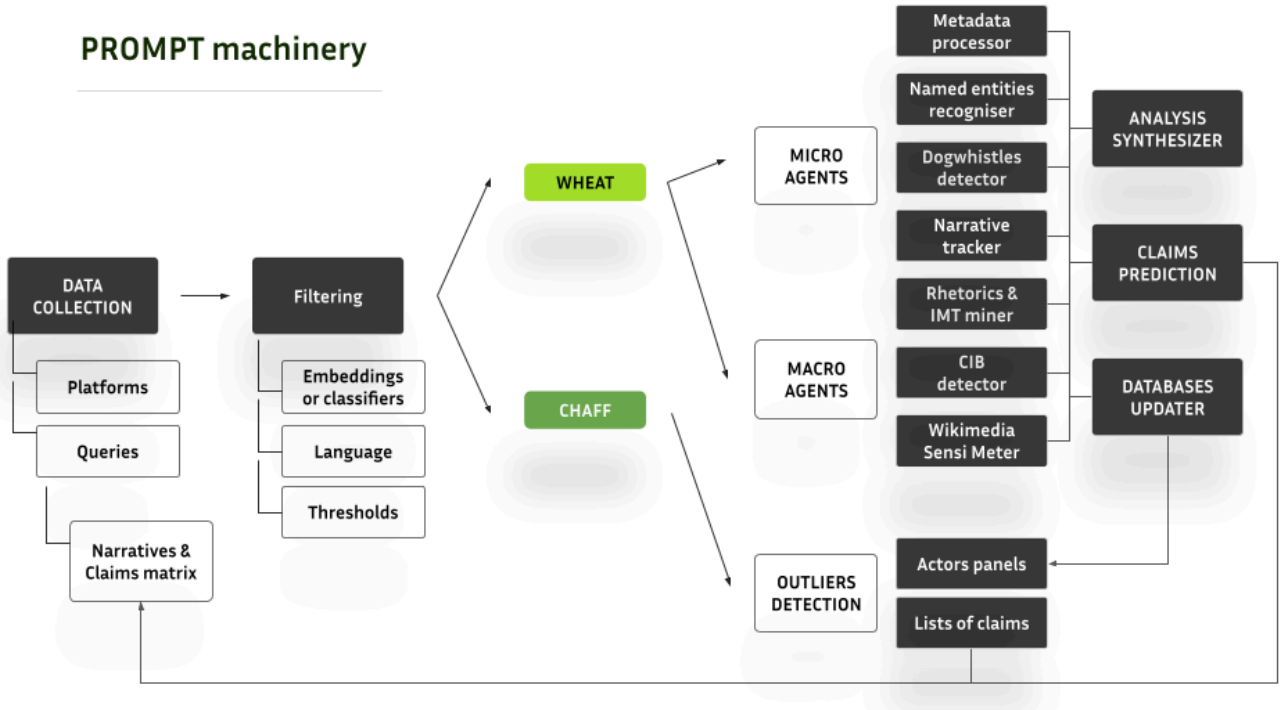
This structure quickly exhausts the available query budget and, as a consequence, makes the overall process lengthy. As a result, collecting consistent datasets from YouTube requires careful optimisation under strict quota limits.

Data processings

Collecting data on social media with the previously presented methodology provides us with a large volume of content to check. One of the main pillars of the ENO-PROMPT project is that, in order to detect disinformative content in these datasets, analysts can be helped by technical tools specially tailored for certain types of signal. These signals rely on coordination, consolidation of panels of known disinformative actors, narratology and rhetorical analysis. While taken separately, they cannot provide attribution, they can provide enough information for experts to conduct their investigations **faster** and **at-scale**.

The technical pipeline used through the ENO-PROMPT project for processing collected data comprises **four main categories that tackle four types of signals**. Each of these tools can be enacted at a micro-level — that is, at a document level (single social media post) — or at a macro-level — a whole dataset. Micro-level processings are always available to analysts, even for an ad-hoc analysis of a singular piece of content. Conversely, macro-level processings always require a dataset or collection of documents as they rely on distributional cues to 'raise flags'. These macro-level tools are tailored for large-scale investigations.

The technical pipeline represents an integrated approach to disinformation detection and analysis, combining multiple methodological perspectives — rhetorical, actor-based, narrative-focused, and coordination — into a coherent analytical framework. Each component provides distinct signals which, when synthesized, offer analysts comprehensive visibility into the diverse landscape of information manipulation. The system's design prioritizes both immediate operational utility for ongoing investigations and the progressive refinement of detection capabilities through continuous learning from analyst feedback and evolving threat landscapes. **Future improvements and research venues include prebunks, knowledge graphs and additional mathematical modelisations for behavioural features.**



The architecture of the pipeline incorporates computational efficiency through selective processing and use of frugal models, both enabled through the structural choice of agents:

- benchmark evaluations and model fine-tuning ensure frugality in resource allocation and matching between specific tasks and smaller specialized models rather than a monolithic one.
- Not all analyses are necessarily applied to all content types.

This selective approach allows the system to scale effectively while maintaining analytical depth where it matters most. The next sections provide an overview of the main 'agents' deployed in the PROMPT architecture.

Narratology

Disinformation frequently operates through narratives — coherent storylines that structure how information is interpreted and remembered. PROMPT distinguishes ***narratives*** (stable, high-level interpretive frameworks) from ***claims*** (specific factual assertions that may or may not align with evidence). This distinction enables analysts to track how particular false claims

serve broader narrative purposes, and how narratives persist even as specific claims are debunked.

Topic modeling to build a narrative taxonomy*:* PROMPT provides a comprehensive taxonomy of disinformation narratives across the project's three core topics. The taxonomy was developed manually and is now being compared with results obtained by computational topic modeling approaches, particularly BERTopic, which enable inductive discovery of thematic structures within large corpora of potentially disinformative content.
BERTopic's approach — generating contextual embeddings via transformer models, reducing dimensionality through UMAP, clustering via HDBSCAN, and extracting topic representations through class-based TF-IDF — proved particularly well-suited to the heterogeneous and evolving nature of disinformation discourse. The resulting topic models can provide analysts with data-driven starting points for narrative identification, which are then refined through expert interpretation and theoretical grounding.

The taxonomy organizes narratives hierarchically, distinguishing between frames (or meta-narratives,broad interpretive frameworks such as "institutional distrust" or "national sovereignty under threat"), narratives (more specific storylines such as "international organizations undermine national interests"), and claims (or sub-narratives, particular instantiations within specific contexts). This hierarchical structure enables both broad pattern recognition and granular tracking of narrative variants.
Note that this topic modeling is of interest when joined with the filtering — see below — in order to provide to analysts already debunked content similar to theirs, but also for later use in order to obtain a similar output for any narrative on any topic, beyond the scope of PROMPT.

Narrative filtering: With the narrative taxonomy established, the technical challenge becomes automatically classifying new content into this framework. The classification system has evolved significantly over the project lifecycle, reflecting advances in both the field and our specific requirements. Initial approaches leveraged BERT-based classifiers developed in collaboration with CSS partners, training on analyst-labeled examples to identify frame alignment. These models performed adequately but exhibited limitations as they struggled with novel framings of established narratives and could not — due to lack of annotated data — work at the narrative and claim level.

The current system employs embedding models combined with similarity metrics in order to find the closest item — similarly to a search engine. We additionally fine-tuned models using contrastive learning techniques in order to obtain better results and make use of annotated data. The contrastive approach works by learning representations where content expressing the same narrative is embedded nearby in semantic space, while content expressing different narratives is pushed apart. This is achieved through a training regime that presents the model with positive pairs (different texts expressing the same narrative) and negative pairs (texts expressing different narratives). For this, we specifically created a perturbation pipeline that, from a single text, creates variations in multiple languages — as we need to ensure that the model is multilingual — with code-switching and "user-edits" to make it similar to content posted on social media, for robustness purposes.

Thanks to the contrastive learning framework, our model offers several advantages compared to the original, BERT-based approach: better generalization to unseen phrasings of known narratives, improved robustness to stylistic variations, and more interpretable decision boundaries. The model currently used is based on the Jinav3 embedding model.

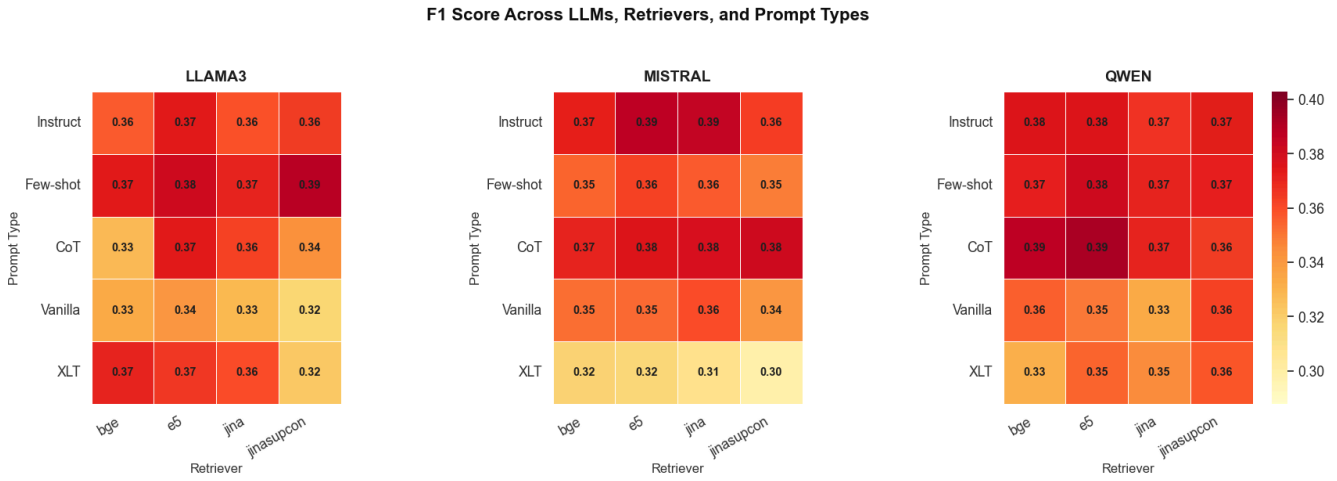
Classification operates for now at multiple confidence thresholds, with high-confidence matches automatically labeled and low-confidence cases flagged for analyst review. This human-in-the-loop approach ensures continuous improvement while the project is still in progress.

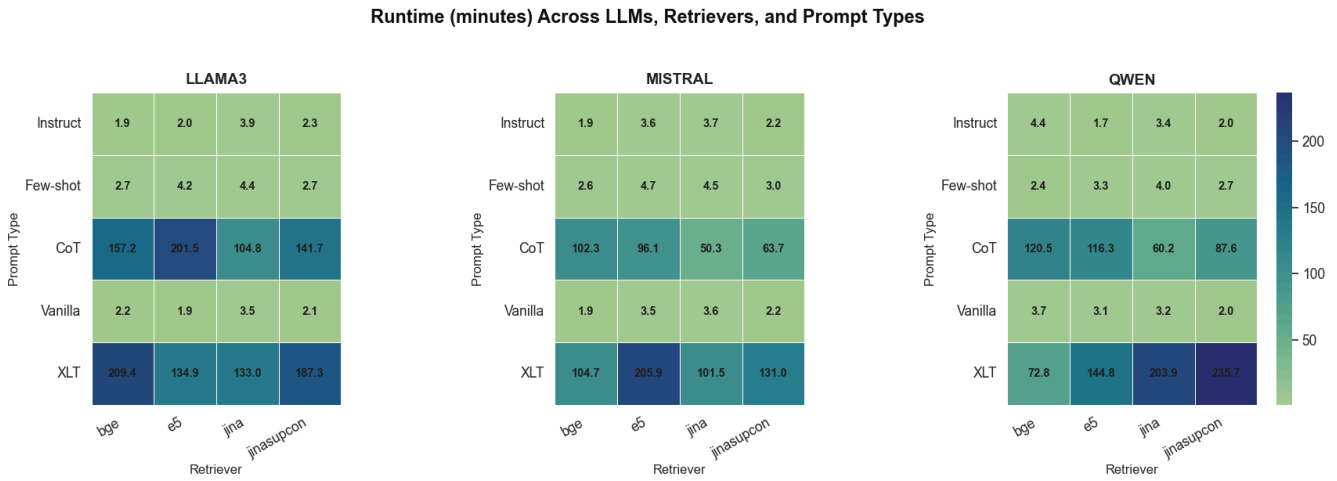
We evaluated our similarity-retrieval component on a manually aligned part of the AMC16k dataset (2,714 claim–post pairs). The data cover several languages, include both matching and non-matching pairs, and reflect the kind of noise you expect on social media. We test different embedding models and prompting styles, compared their retrieval quality, looked at where they made mistakes, and checked how long each setup takes to run so we know what's realistic for actual use.

Each claim–post pair was evaluated under four retrievers (BGE-M3, e5-multilingual, Jina v3, and our supervised-contrastive JinaSupCon variant) and five prompting formats (Vanilla, Instruct, Few-shot, Chain-of-Thought, and Cross-lingual Transfer).
LLMs tested: LLaMA-3-8B-Instruct, Mistral-7B-Instruct, and Qwen-2-7B-Instruct.

The evaluation pipeline reproduced – and expand to LLM usage – the workflow of the PROMPT narrative-filtering:

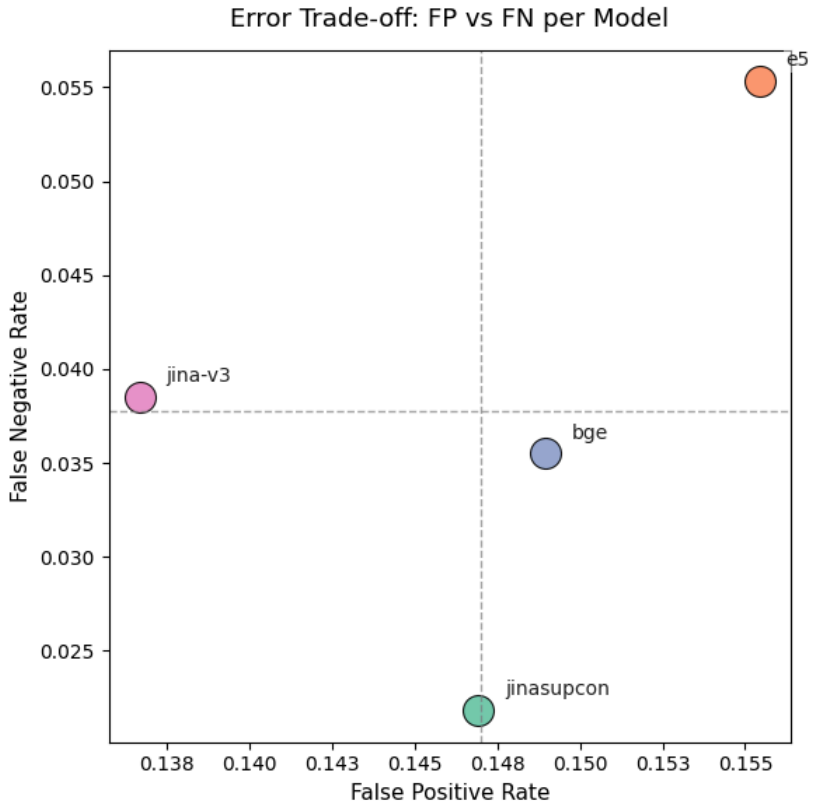
1. retrieve top-k candidates using the embedding model,
2. filter using cosine similarity threshold,
3. apply LLM scoring under different prompting strategies,
4. compute classical classification metrics (F1, precision, recall).

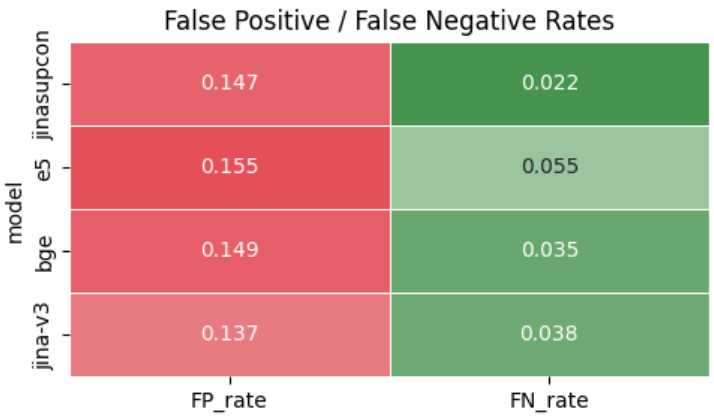




Across all configurations, the best overall setup came from combining strong embeddings (Jina v3 and JinaSupCon, the fine-tuned version of Jina v3) with Qwen or LLaMA3, using Few-shot or Instruct prompting. These combinations gave the most stable F1 scores while keeping runtime low enough to be practical. Jina v3 was often one of the strongest retrievers across models, and JinaSupCon performed similarly, especially in cases where capturing more relevant matches mattered.

Few-shot and Instruct prompting stood out as the most reliable styles: they regularly matched or outperformed the other prompt types without extremely long runtimes of Chain-of-Thought. When put together, *Jina v3 / JinaSupCon + Qwen (or LLaMA3) + Few-shot/Instruct* delivered the best balance of accuracy, speed, and consistency.





To evaluate the embeddings on their own, without any influence from prompting or LLM scoring, we also measured false-positive and false-negative rates using direct cosine similarity on the full multilingual dataset.
This analysis gives a clearer view of each model's strengths and weaknesses. Jina v3 shows a good balance between the two error types, while the supervised-contrastive version (JinaSupCon) achieves the lowest false-negative rate overall, meaning it is better at catching relevant matches. BGE and e5 show higher error rates on both sides. This result aligns with our earlier findings: the Jina models provide the most stable and reliable embedding space, with JinaSupCon having an advantage when high recall is essential.
All in all, this means that our fine-tune is of interest in specific regimen and if the embedding model is used on its own, while the generic model provides good results without additional fine-tuning if coupled with a LLM that works with what the embedding retrieves.

New narrative detection: Detecting genuinely new narratives — those not yet captured in the matrix taxonomy - represents one of the project's most challenging methodological problems. The core difficulty lies in distinguishing between novel narratives and novel framings of existing narratives. The project's operating assumption is that narratives should be relatively stable constructs, serving as stable anchors to which multiple evolving claims can attach.

The detection system employs a multi-stage approach:

- Outlier detection: Content that scores poorly across all known narrative categories is flagged as potentially representing new 'narrative territory'. This uses the contrastive learning model's embedding space, identifying content that sits far from any established narrative cluster.
- Semantic coherence analysis: Flagged content is analyzed to determine whether multiple pieces cluster together semantically, suggesting a coherent new narrative rather than idiosyncratic content. This leverages HDBSCAN clustering in the embedding space with careful parameterization to avoid fragmenting known narratives.
- Analyst validation: Automatically detected candidate narratives are presented to analysts for validation, refinement, and formal incorporation into the narrative taxonomy. Analysts assess whether the detected pattern represents a genuinely new interpretive framework or a variant of existing narratives.

This methodology is under continuous refinement, with particular attention to the theoretical question of narrative boundaries and the operational question of detection sensitivity thresholds.

Coordination

Coordinated behavior represents a key indicator of inauthentic activity and organized influence operations. Unlike organic discourse, where users independently decide what to post and when, coordinated operations exhibit temporal, behavioral, and content synchronicity that can be detected through statistical analysis.

Coordination detection: Coordinated behaviour refers to situations in which two or more social media accounts repeatedly perform actions involving the same uniquely identifiable content within a predefined time interval (Righetti & Balluff, 2025). Coordination is detected when accounts share identical or equivalent objects, such as URLs, posts, hashtags, or images, in a near-synchronous manner. While single simultaneous shares may occur randomly, repeated synchronous actions from a stable group of accounts indicate a non-random pattern that suggests strategic, centralized or automated activity. The concept includes both explicit coordination and forms of organized amplification that emerge from repeated, quasi-synchronous sharing behaviours.

Coordinated behaviour, therefore, is grounded in two key criteria:

- synchronicity, meaning that accounts share the same content elements within a specific time window;
- repetition, meaning that the same account pairs synchronously share the same objects multiple times.

When these conditions co-occur, the accounts become increasingly likely to be part of a coordinated network rather than engaging in organic activity.

The fundamental insight is that coordinated actors tend to post similar content within compressed time-windows, a pattern unlikely to occur by chance in organic discussion. CooRTweet operationalizes the definition of coordinated behaviour by identifying all account pairs that share the same objects within a chosen time threshold. It then measures how often these synchronous shares occur and constructs a weighted network where edges represent recurrent co-sharing of objects. High edge weights indicate strong coordinated activity. The package also provides tools to select only the most coordinated pairs, for example by filtering edges above the ninety fifth or ninety ninth percentile of the edge weight distribution, and to isolate the fastest coordinated clusters using narrow time windows such as ten seconds.

The conceptual framework implemented in the CooRTweet R package generalizes coordination detection across platforms, content types and modalities, and allows researchers to analyse both mono-modal and multi-modal networks.

PROMPT's core coordination detection methodology employs temporal clustering based on the CooRTweet package, augmented by semantic similarity analysis.

The technical pipeline operates as follows:

- Embedding generation: All content items are processed through transformer-based embedding models (such as multilingual sentence-BERT variants) to generate semantic representations that capture meaning rather than surface-level text similarity. Note that, from a computational point of view, this step is already done for other features so there is little to none additional cost.
- Similarity computation: Pairwise semantic similarities are calculated across content items within temporally bounded windows. This reduces computational complexity from $O(n^2)$, which is reprehensive, across entire datasets to manageable chunks while capturing coordination that manifests over hours or days rather than months.
- Temporal clustering: Content items that are both semantically similar (exceeding a threshold derived from empirical calibration) and temporally proximate are identified as potential coordination clusters.
- Actor network analysis: For validated coordination clusters, the system identifies which actors participated, revealing coordination networks that may span multiple content items and temporal windows.

This content-based approach avoids reliance on engagement metrics or follower graphs, which may be less available or less reliable across different platforms and data collection contexts. However, they can be additional signals of interest.

Influencers shift of interest (in progress): The detection of behavioral anomalies in "magic-middle" social media influencers represents one of the project's most prospective research directions. These influencers — operating below the threshold of systematic monitoring while maintaining trusted relationships within specialized niches — serve as critical intermediaries in contemporary information warfare.

The mathematical framework developed for this analysis focuses on behavioral pattern analysis rather than content classification, enabling politically neutral detection. The approach models influencer digital footprints using semantic embeddings of their posting histories, processed through hierarchical stochastic processes:

Hidden Markov Models (HMMs) capture inter-topic transitions, modeling how influencers shift between different topical areas over time. Each influencer's posting history is represented as a sequence of topic states, with transition probabilities learned from their historical behavior. Sudden shifts to atypical topics—especially those associated with known disinformation narratives—trigger anomaly flags.
Ornstein-Uhlenbeck (OU) processes model intra-topic semantic drift, capturing how an influencer's treatment of a particular topic evolves over time. The OU process, borrowed from physics and quantitative finance, describes mean-reverting stochastic dynamics. Each influencer has a characteristic "semantic position" within each topic space that serves as a mean-reverting attractor. Deviations from this characteristic position suggest external influence or deliberate repositioning.

The framework provides anomaly detection through "cost of postage" metrics — quantifying how surprising a particular post is given the influencer's historical patterns. Posts with high

cost of postage are flagged for analyst review. Additionally, the system employs Wasserstein distances to compare the entire probability distributions of influencer behavior over time, enabling detection of gradual behavioral shifts that might not trigger threshold-based alerts.
This content-agnostic methodology enables detection of influence operations even when the specific narratives involved are novel or when influencers maintain plausible deniability by framing content as personal opinion or genuine inquiry.

Community shift of interest: Parallel to individual influencer analysis, the system tracks collective behavioral shifts within identified communities or network clusters. Communities are defined through network analysis of interaction patterns (mentions, replies, retweets/shares) or through content-based clustering of frequently co-engaged audiences.

The analysis applies similar stochastic modeling approaches at the community level, asking whether the aggregate behavior of a community has shifted in ways inconsistent with historical patterns. Community-level detection offers complementary value to individual influencer tracking: while individual anomalies may represent organic changes in interests, coordinated shifts across community members provide stronger evidence of organized influence operations.

Dynamic Network analysis: The core objective is the systematic identification, classification, and diffusion analysis of claims and disinformation narratives in large-scale, multilingual textual corpora related to a given topic and its global discursive environment. Our methodological framework consists of four interrelated components:

1. Custom Codebook Development: We developed a novel coding scheme tailored to our research focus. This scheme takes the form of a codebook, which draws inspiration from established frameworks such as the master codebook of 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 (in our case, an individual post) 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.

2. Large Language Model (LLM)-Supported Multilingual Classification: To apply the codebook across multiple languages and large datasets, we fine-tune transformer-based multilingual language models using supervised learning techniques. Human-annotated training sets are used to train and validate the models in a cross-lingual setup. This enables accurate narrative classification at scale, while maintaining transparency and reproducibility through active learning loops and inter-annotator agreement testing.

To generate annotated data, trained annotators manually code a subset of the corpus. We annotated a sample of claims to fine-tune large language models (LLMs) under few-shot learning conditions, building on recent research in low-resource classification (Mate et al. 2023). These human-coded examples serve as the basis for fine-tuning multilingual LLMs for claim detection at scale.

We evaluate model performance through standard metrics such as precision and recall, while also assessing inter-annotator agreement with human coders and the models' robustness across languages. To mitigate model drift and ensure interpretability—particularly for ideologically sensitive claims—we implement a human-in-the-loop validation strategy. This iterative process includes expert review and correction of LLM outputs, especially in cases involving ambiguity or borderline classifications. The approach improves both the reliability and the transparency of the automated classification system.

3. Automated Filtering Based on Multilingual Embeddings: Given the scale of the data and the diversity of textual sources, we employ multilingual sentence embeddings to filter and cluster thematically relevant content. This embedding-based semantic filtering step enables the detection of latent topic clusters and narrative variants across linguistic contexts, helping to prioritize content for manual review and model refinement.

- We compiled a hand-picked list of relevant claims, based on real-life examples of discourse found in the media or on social networks. The list contains around 200 claims that provide finer-grained detail within each Narrative Frame. We use embeddings to measure the similarity between posts and the predefined claims and frames. For each social media post, the system assigns the five most similar claims, provided they surpass a predefined threshold using cosine similarity.
- Recent literature and advances in multilingual modeling support the use of language models across different linguistic contexts. In practice, we benchmarked several models, including Jina v3, bge-m3, Cohere Multilingual v3, Snowflake Arctic, and Voyage 3. We calibrated the acceptance/rejection thresholds by testing claims and their translations to ensure consistency (cf. heatmaps). This process aims to achieve sufficient contrast between claims within the same narrative and those across different clusters. Based on these evaluations, we selected Jina v3 as the final embedding model. On a consumer-grade GPU (NVIDIA GeForce RTX 4070), processing one million posts takes an average of 16.3 minutes. By including both claims and frames, we can track areas where refinements are missing. In such cases, we apply one of two automated approaches: either prompting a large language model (LLM) to generate a new label in a few-shot setting for posts without relevant claims, or using a semi-supervised BERTopic approach. The latter uses the existing claim list as a seed and augments it with new clusters detected by the algorithm.

4. Network Analysis of Narrative Propagation: Building on the classified corpus, the project conducts network analyses to study the patterns of claim dissemination and narrative spread. Using metadata from social media platforms (e.g., retweet, quote, and reply structures), we construct user interaction networks and narrative co-dissemination graphs. We examine how disinformation frames propagate across user clusters, identify influential actors, and map the structural positions of bridging nodes responsible for cross-cluster diffusion.

4.1 Network Building Principles

Following Saqr's (2023) framework for temporal network analysis, we construct directed dynamic networks by defining nodes and edges, with time-stamped interactions forming the edge attributes. Two main conceptualisations guide our approach:

- Conversation-Based Network: Here, conversations—defined as an original tweet and all replies (including nested replies)—are treated as nodes. The underlying assumption is that users contributing to the same conversation are likely exposed to other posts in that thread, thus forming a virtual interaction space. We construct edges between these nodes based on retweets across conversations, drawing on methods proposed by South et al. (2022), who define directed links from an original post to quoting or retweeting accounts.
- User-Based Information Flow Network: In a second model, we treat users as nodes and define directed edges as instances of information transfer, occurring when one user retweets, quotes, or replies to another. This model assumes that each of these actions entails at least minimal cognitive uptake of the source content. Edges are time-stamped and annotated with user metadata (follower/following counts). In cases where quoted or retweeted users are not otherwise present in the dataset, their metadata is recorded as missing (set to 0).

4.2 Temporal Network Typology and the re-tweet/quote network

- To evaluate information spread over time, we construct an interval temporal network based on the approach developed by Saqr (2023). In Saqr's (2023) interval temporal network approach, edges have both a start (onset) and end time (terminus), representing the duration of potential information availability. This setup allows for the analysis of longer information lifespans across user chains and enables the calculation of temporal density, degree centralization, connectedness, and transitivity across the seven time points spanning the seven days.
- Based on the variables describing the identifier of the tweet itself, the time of the tweet's creation, the identifier of the user who posted the tweet, and equivalent information regarding the original tweet in case of retweeting or quoting other contributions, a strictly defined interval temporal network can be constructed to represent the flow of information across users over time. This retweet and quoting network can be conceptualised as a user network where connections are made when information passes from one use to another. As evident from its naming, information exchange is presumed in two cases: first, when user A retweet's a tweet from user B, in which case we presume information flow from user B to user A, and second, when user A quote's the tweet of user B, in which case, similarly, we presume information flow from user B to user A. It is worth noting that less strict conceptualisations are possible, such as including conversations – a tweet and subsequent replies posted to it directly, or to its derived responses – where a level of information flow could be presumed between the user who tweeted and those who posted replies. Although such a network could include more users and, with appropriate weighting, reveal alternative user relations, in the case of information flow, the "reply" relationship would assume that such interactions work towards furthering the original message, which in many cases would be incorrect.
- Network edges are created by iterating through the dataset of tweets and noting a connection between users when there is a tweet-retweet relationship or a quoting relationship. Edges are annotated with the timestamp of the current tweet as a variable indicating the time of information flow. In case of tweets which are retweets and

include a quote as well, two edges are created for both relationships. Nodes are annotated with follower and following counts for later analysis; however, for original users whose tweets are quoted or retweeted, and their users are not in the dataset as "tweeters", follower and following counts are set as 0. Edges are grouped into tweet chains and annotated with the group number they belong to. This grouping is done with a second iteration of the dataset using a DFS (depth-first search) algorithm that traces chains of retweets and quotes to create the given groups. In our conceptualisation the end time (terminus) of each edge is defined as the timestamp of the last tweet within a retweet/quote group.

Linguistic analysis

PROMPT contends that the disinformation discourse contains persuasion techniques, stylistic and rhetorical devices that can be screened to tag suspicious content. The PROMPT semantico-axiological matrix classifies content through these various properties. The processings of this category, at a micro-level, leverage Generative AI (especially LLMs and other NLP tools) to automatically fill out this matrix for each provided piece of content.

Semantic-axiological matrix: a comprehensive codebook operationalizes rhetorical and stylistic markers characteristic of disinformative discourse. It serves both as a resource for analysts rooted in linguistic research and as the foundation for the semantico-axiological matrix, capturing dimensions such as the emotional triggers of disinformation, argumentative structure, source attribution patterns, and audience mobilization tactics. Key elements include markers for:

- Emotional triggers and axiological framing: Fear appeals, moral outrage, in-group solidarity, out-group derogation
- Persuasion techniques: False equivalences, whataboutism, ad hominem attacks, strawman constructions
- Rhetorical devices: used in combination with manipulatory techniques to convey messages favourably to targeted audiences.

The automated completion of this matrix leverages Large Language Models in a structured prompting framework. The system uses Pydantic, a type-verification framework, to guide the LLM through each dimension of the codebook while minimizing hallucination rates and type mismatch due to the stochastic nature of LLMs. For high-volume processing environments, we are currently developing BERT-based classifiers fine-tuned on analyst-labeled examples. These lighter models offer significant computational advantages — e.g. CPU compatibility — while maintaining acceptable performance on core dimensions. Our hypothesis is that ensemble approaches combining LLM assessments for complex rhetorical features with BERT classifiers for more straightforward stylistic markers optimize the accuracy-efficiency tradeoff.

Dogwhistle detection: Dogwhistles — coded language that conveys specific meanings to target audiences while maintaining plausible deniability — represent a significant challenge in content moderation. The project has developed a multilingual dictionary of dogwhistle terms and phrases, building on existing resources from EU monitoring efforts and extending coverage to non-English languages included in the project. The current implementation uses regex patterns

and Levenstein-based distances for known dogwhistles, providing rapid scanning capabilities across large datasets. However, the evolution of coded language necessitates more adaptive approaches. We are exploring LLM-based detection in zero-shot configurations, though the computational costs and reliability of such approaches remain under evaluation.

Wikipedia NER: The system incorporates Wikipedia-based Named Entity Recognition to enhance analytical workflows. While this functionality primarily serves user interface and quality-of-life improvements for analysts, it plays a crucial role in contextualizing content and enabling the Wikipedia sensitivity analysis described below. The NER system identifies entities mentioned in social media content and links them to their corresponding Wikipedia articles, providing analysts with immediate access to contextual information and enabling downstream analytical processes.

Wikipedia sensitivity

Wikipedia serves as a contested battleground for information influence operations, with nefarious actors often collecting information for narrative and claim creation, and establishing source legitimacy through Wikipedia editing before broader dissemination. This hypothesis suggests that Wikipedia activity can provide weak signals and early indicators of emerging disinformation campaigns. We have developed a suite of metrics that leverage Wikipedia's unique characteristics — edit histories, talk page discussions, source citations, etc— and made it available for cross-language article comparison to assess the "sensitivity" of entities mentioned in social media content. These metrics are divided into three categories.:

- Heat Risk Indicators: This category measures the intensity and volatility of activity around an article. It detects abnormal spikes in pageviews and edits, evaluates the probability that modifications will be reverted, examines the level of editorial protection applied, and quantifies the intensity of debates on talk pages. These signals reveal topics under tension or receiving unusual attention, potentially indicating ongoing influence operations.
- Quality Risk Indicators: This category assesses the reliability and maturity of encyclopedic content. It analyzes the article's official classification (from “stub” to “Featured Article”), identifies the presence of unreliable or blacklisted sources, detects citation gaps and unsourced claims, measures content staleness, and examines the diversity versus concentration of cited sources as well as the balance between content additions and deletions. These metrics help identify articles vulnerable to manipulation due to poor editorial quality.
- Contributor Behavior Risk Indicators: This category detects suspicious or coordinated contribution patterns. It identifies the presence of sockpuppets, measures the rate of anonymous edits, evaluates editorial concentration (monopolization by a small number of contributors), analyzes the regularity of editors' activity, and examines imbalances in their editing behaviors. These indicators reveal abnormal community dynamics that may signal orchestrated manipulation attempts.

These metrics are calculated through a combination of Wikipedia API queries, statistical analysis of edit histories, and automated parsing of article metadata and talk page (‘discussion’)

structures. Several metrics build upon approaches already used by parts of the Wikimedia community for vandalism detection and edit quality assessment.

Panels

Nefarious actors frequently persist across multiple disinformation campaigns. The relative stability of these threat actors, as well as the disproportionate influence they have on the digital space, allows for significant computational optimization and knowledge dissemination. By maintaining panels of known problematic accounts, we can save the most resource-intensive processing for content produced by these actors. This enables proactive monitoring, alerting analysts to newly created content from known sources without scanning entire social media platforms.

Metadata filtering: The initial stage of panel processing involves metadata-based filtering to reduce the volumetry of content requiring deeper analysis. The system ingests comprehensive metadata collected during social media scraping operations, including account identifiers, temporal information, engagement metrics, posting patterns, and network relationships. Expert-knowledge lists of problematic actors, continuously refined through analyst feedback and previous investigations, serve as the baseline for filtering operations ; and previous and continuous results serve as regular updates. By rapidly identifying content originating from or associated with known panels, the system prioritizes analytical resources toward the highest-value targets. This however does not mean that we rely only on this as we recall that *it is only one flag among many.*

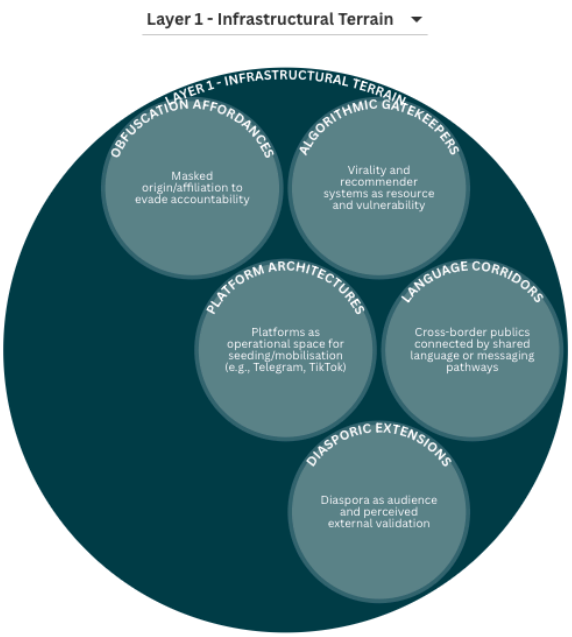
T2: MOLDOVA'S 2025 PARLIAMENTARY ELECTIONS: DISINFORMATION AS A GEOPOLITICAL BATTLEGROUND

Conceptual framework

We propose **a holistic three-layer spatial framework** that links infrastructures, operational behaviour (TTPs – Tactics, Techniques, Procedures), and strategic effects on democracy

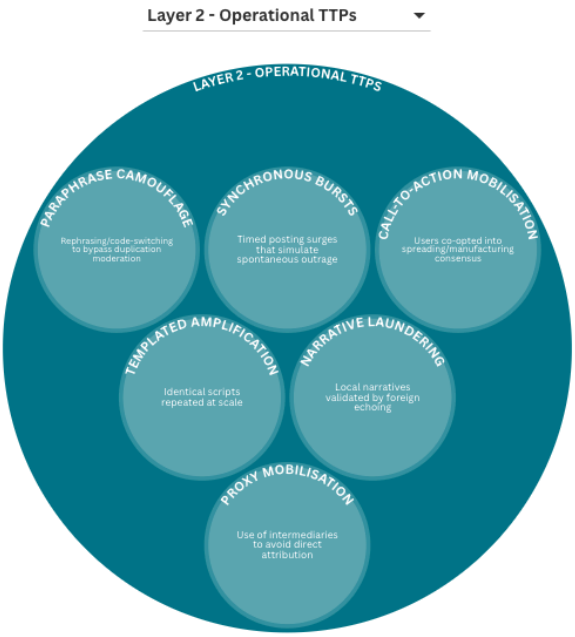
Layer 1 – Infrastructural Terrain (where interference operates)

Platform architectures	Each with specialised functions (Telegram for seeding, TikTok for mobilisation, web ecosystems for laundering and longevity
Language corridors	Publics interconnected semantically rather than territorially
Diasporic extensions	Transnational communities serving simultaneously as audience and manufactured evidence of societal sentiment
Algorithmic gatekeepers	Virality and recommender systems as resources and vulnerabilities
Anonymity & obfuscation affordances	Enabling mutable origin and plausible deniability

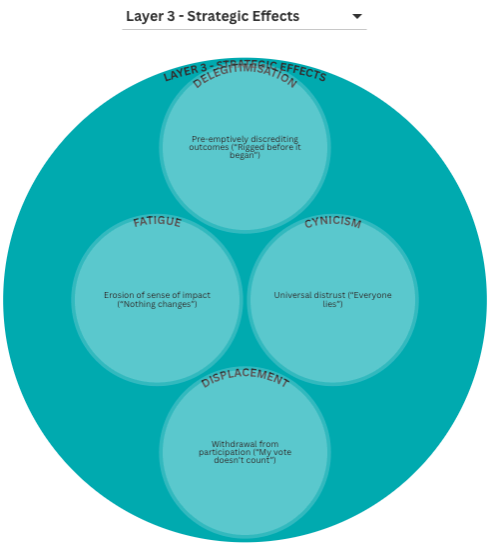


Layer 2 – Operational Behaviours (how interference is executed)

Templated amplification	Identical scripts reposted at scale via coordinated timing to simulate spontaneous or long-term outrage
Narrative laundering	Domestic grievances validated via 'foreign echoes'
Proxy mobilisation	Deniable amplification via covert intermediaries / origin masking
Paraphrase camouflaging	Code-switching to avoid moderation and detection
Participatory calls to action	Enrolment of audiences as force multipliers. Transition from coordinated or engineered to organic engagement



Layer 3 – Strategic effects on democracy (what interference seeks to achieve)



Data Validation

In a nutshell, the validation process involved:

- **Anomaly detection via Hopfield networks**, capable of storing and retrieving associative patterns. The model could detect subtle temporal repetitions and latent associative memory structures within disinformation/information manipulation campaigns. This approach was particularly well-suited to datasets where coordination is obscured by decentralised or proxy entities obscure coordination.
- **Semantic clustering**: posts and messages were grouped using semantic similarity at a 0.75 cosine threshold, flagging clusters of near-identical phrasing and structure, indicative of templated or recycled scripts.
- **Temporal actor validation**: we mapped actor-content-time relationships across platforms to flag accounts exhibiting abnormal behaviour patterns, such as coordinated bursts of messaging, mirroring of external sources or behavioural anomalies across multiple identities.
- **Confidence scoring**: a coordination confidence level above 60% was deemed strategically relevant and used to filter high-risk content for deeper inspection. Each flagged cluster was cross-referenced with known disinformation proxies or previously identified influence operators.

Narrative Clustering

The cross-platform dataset (5,286 entries) covers 693 semantic clusters and approximately 2300 disinformation posts which circulated between June and mid-September 2025. Each cluster aggregates identical or similar posts across languages (Romanian, Russian, English, French, Italian, even Japanese, etc.) and records the number of posts, impressions, duration and up to three narrative categories (the latter, processed manually). Why opt for this clustering approach?

While surface-level textual analysis highlights bursts of synchronised posting, a closer examination of thematic recurrence reveals a more deliberate and sustained strategy of narrative reinforcement through distributed paraphrasing. To differentiate between isolated viral bursts and coordinated, long-term amplification efforts, each post in the dataset was grouped into semantic similarity clusters using an embedding-based similarity threshold (0.75 cosine distance). By correlating start and end date timestamps within each cluster, we could measure narrative persistence over time (duration). Clusters with short life spans (<12 hours) and high posting density suggested synchronous coordination, where multiple entities push the same message simultaneously for rapid visibility – as evidenced in the Transnistria use case. Clusters with long duration (days or weeks) and periodic reactivation indicate strategic narrative anchoring, with messages being reintroduced in the information space deliberately

over time. So far, by examining timeline behaviours, three temporal life-cycle patterns and distinct modes of amplification could be identified:

Amplification mode	Identifiers	Example pattern
Synchronnous burst	Multiple identical posts appearing within hours across social media accounts/channels and web domains (seemingly unrelated)	'PAS redrawing the electoral map'
Strategic reactivation	Clusters suddenly reactivated around campaign events, official decisions, in the proximity of the ballot date	Opposition repression/ Evgenya Gutsul 'unlawful' sentending
Baseline dripfeed (narrative substrate)	Same narrative frame repeated intermittently over months, sustained through paraphrasing	Moldova proxy war escalation, EU/NATO occupation

Viral bursts alone do not necessarily prove coordination. However, when identical or near-identical framings recur episodically across discrete time windows, languages and platforms (and the web) may signal templated orchestration, both automated and human-in-the-loop. It also reveals a critical insight for election monitoring: a coordinated influence campaign is no longer defined by volume, but by controlled repetition with strategic latency.

The clusters collectively generated 112 million impressions and involved an average of 8-9 unique actors (median=3) per semantic cluster. Most activity unfolded within tight/synchronous amplification cells, with a small subset showing high actor (account) dispersion (50-120 contributors each) – a pattern indicative of coordinated mass-push moments. Temporal behaviour was measured using the observable lifespan of each cluster, defined as the interval between the earliest and latest appearance of semantically similar posts within the dataset. However, durations represent visible persistence rather than full lifecycle certainty; some narratives likely circulated earlier or resurfaced beyond the observed window. Based on these intervals, 82.5% of clusters qualified as short-lived synchronous bursts (<12h), 14,4% sustained activity for 12-72 hrs, while 3% persisted for multiple days or weeks, typically via intermittent reactivation rather than continuous posting.

The narrative categories in the parsed dataset were assigned manually. While semantic clustering grouped posts by lexical/structural similarity (cosine 0.75), this approach captures surface-level phrasing rather than intent or functional role. In practice, two messages may express near-identical wording but serve different strategic narratives depending on context, platform or speaker (i.e., 'EU integration destroys sovereignty' may function as economic grievance, identity threat, or foreign occupation depending on framing cues). Conversely, logically equivalent narratives are often paraphrased beyond the threshold of machine-detected similarity, meaning purely embedding-based methods underestimate narrative continuity when hostile actors purposely rephrase or code-switch. Within the cross-platform dataset, we observed several evasion patterns or anti-repetition camouflage.

In the example below, **the obfuscated variant** does not rewrite or paraphrase the message, it retains over 90% lexical overlap with the seed/initial post. Instead, it introduces emoji substitutions, bullet reformatting and punctuation encoding changes to produce a technically

distinct post that will bypass naïve duplication or similarity filters, while carrying the same function.[96]

Templated Repetition	Obfuscated Version	Narrative Function
"Not All Moldovans Are Equal – The Central Election Commission of Moldova has published a preliminary draft on the opening of polling stations abroad [...]."	"Voices for the Diaspora only – The Central Election Commission of Moldova announced they plan to reduce the number of polling stations..."	Pre-emptive deligitimisation of voting outcome

Feature	Seed Cluster ('Not all Moldovans are Equal...)	Obfuscated Cluster ("Voices for the Diaspora only...)	Technique
Emoji style	Uses headline framing emojis (📝, 🚩, ❗, ⭐, 💸)	Uses slightly softer / more mixed palette (📝, 🖍️, 🏳️, ❗, 💸)	Suggests format-targeting (the first in a more alarmist tone)
Bullet formatting	Mix of hyphen bullets and Unicode dots (•)	More structured Telegram bullet icons (▪ , 🖍️, 🚩)	Indicates cross-app formatting – WhatsApp/Telegram vs bot-published post
Quotation marks & apostrophes	Mix of ASCII and curly Unicode quotes (" " / ')	Mostly straight ASCII quotes (" ")	Suggests different keyboard origin/forwarding pipeline
Hashtag / Tag structure	Ends with stable signature (#elections #Moldova #Russia ⭐ @rybar)	Same signature but sometimes shuffled or spaced differently	Standard for bot amplification / reposting plug-ins
Spacing / Line breaks	Blocks of text separated with double line spacing	More compact paragraph format before emoji-inserted breaks	Indicates intentional format compression for readability in crowded feeds

Wikipedia Metrics Correlation Matrix

To better understand how the three Wikipedia indicators -MRS, SRS and BVI - relate to one another, we computed a correlation matrix across all composite risks and underlying (source) features, previously normalised.[97] While the three primary scores were constructed to represent separate dimensions of editorial manipulation, sourcing fragility and editorial instability, the matrix was used to assess whether risk types overlapped. The statistical analysis confirmed that the risk indicators have near-zero correlation, particularly between MRS and SRS (r=+0.03), and MRS and BVI (r=-0.005), **which largely indicates independent operational modes.** There was a slight alignment between SRS and BVI (r= +0.09), which implies that sourcing gaps may attract unstable editing. **In other words, disinformation is not a singular,**

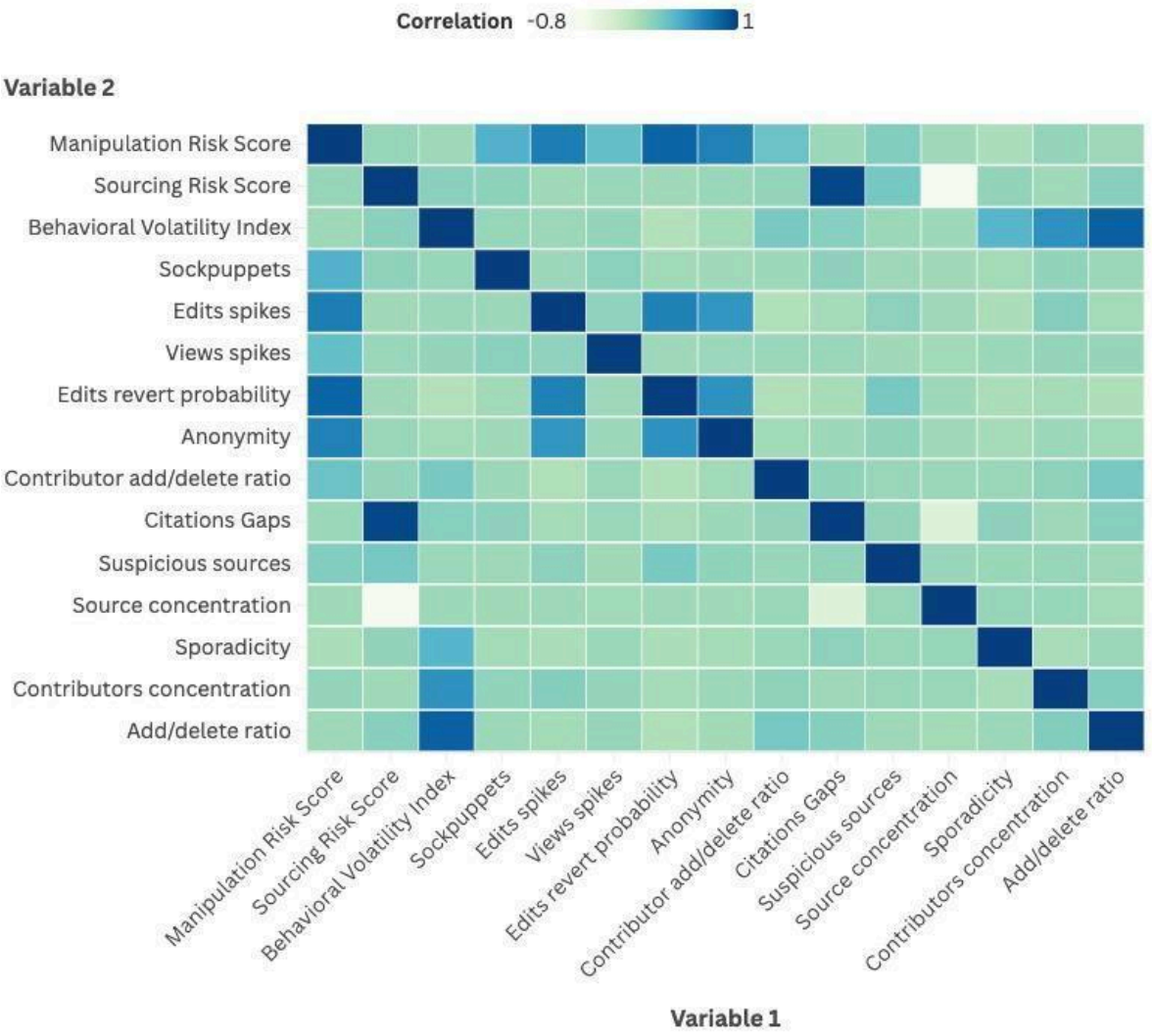
[96] Both variants originate from the @rybar (Rybar) disinformation ecosystem, initially seeded on Telegram and repackaged across other platforms and formats (August 2025). The second instance is a reformatted/emoji enhanced version, with minimal text mutation (except the headline).

[97] A Pearson correlation was computed between all variables including composite risk scores and source metrics, to identify linear relationships between features, overlaps and distinct behaviour among indicators.

but a multi-modal phenomenon. Some campaigns may seek to edit persistently, others to destabilise content, or undermine citation legitimacy, often without overlap.

However, within the composite metrics, strong internal associations emerged whereby each score showed robust coherence with its underlying components:

- The Behavioural Volatility Index (BVI), capturing editorial instability, was most closely aligned with *Add/Delete Ratio* (r=0.82) and *Contributors Concentration* (r=0.55), reflecting environments where a small number of editors rapidly rewrites content.
- For the Manipulation Risk Score (MRS) – editorial manipulation – the highest contributing features were *Edits Revert Probability* (r=0.80), *Edit Spikes* (r= 0.66), and *Anonymity* (r=0.64). This shows that manipulation risk tends to emerge in environments of frequent content reversals and low contributor transparency.
- The Sourcing Risk Score (SRS) – epistemic fragility – was overwhelmingly driven by *Citations Gaps* (r=0.95) and negatively associated with *Source Concentration* (r=-0.66), which means that high-risk articles often lack citations and rely heavily on very few sources.

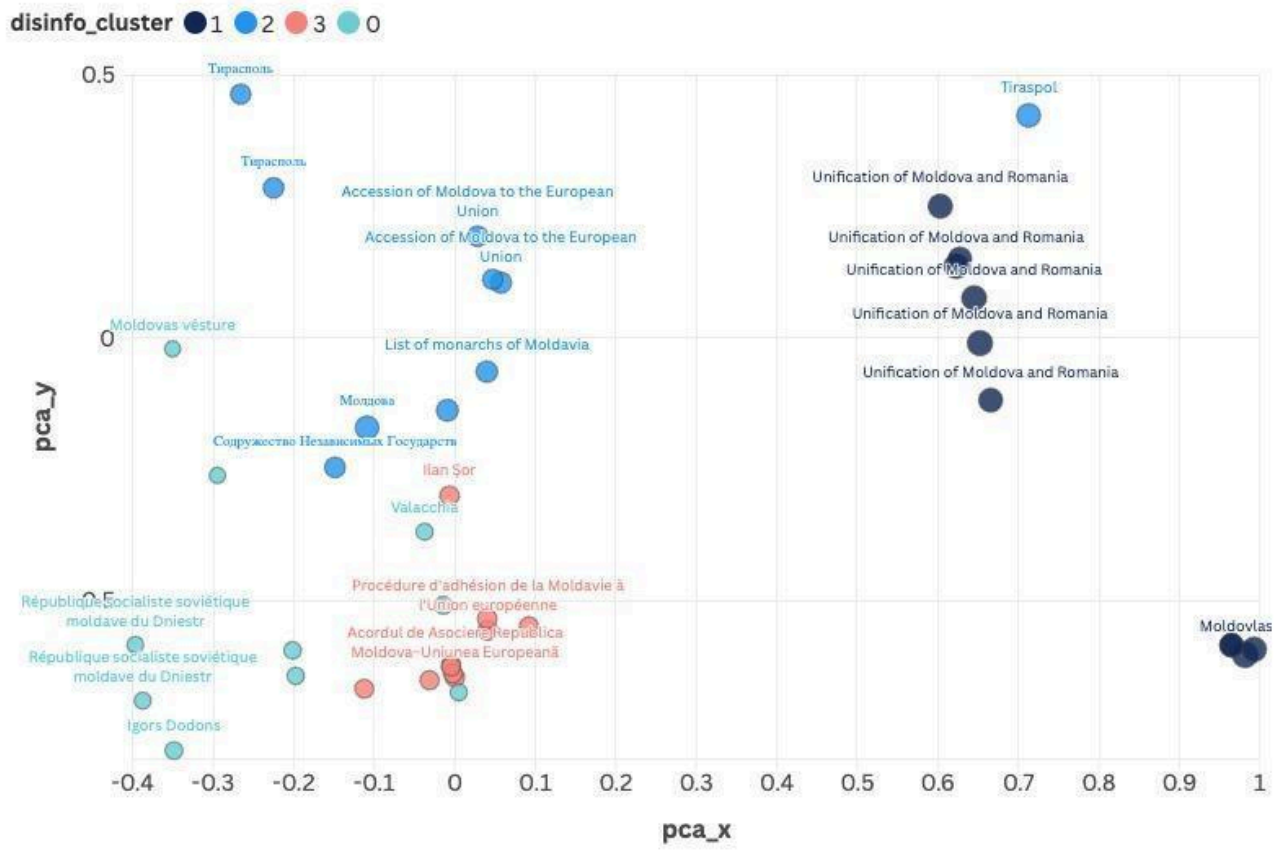


Moving beyond individual risk indicators to uncover broader patterns of disinformation exposure, an unsupervised clustering analysis was conducted. Using normalised values from all behaviour and source-based variables, a K-Means algorithm grouped articles into four distinct

clusters based on their combined profiles. The resulting clusters reveal noteworthy dynamics across the risk spectrum:

- **Cluster 0** exhibited moderate values of Behavioural Volatility (BVI = 0.31) and Manipulation (MRS = 0.20), pointing to sporadic but likely organic editorial activity, with low sourcing risk.
- **Cluster 1** emerged as the most epistemically vulnerable, with a Sourcing Risk Score (SRS) average of 0.60, indicative of widespread citation gaps and poor source diversity.
- **Cluster 2** displayed the highest Manipulation Risk Score (MRS) with 0.24, suggesting a prevalence of coordinated/anomalous editorial behaviour, including *sockpuppetry* and *revert-heavy editing*.
- **Cluster 3** recorded the lowest scores across all three indicators, suggesting articles in this cohort are comparatively stable, better sourced, and less exposed to editorial interference.

To better interpret the operational patterns behind the most exposed and vulnerable clusters, we examined the ten highest-risk articles within each group based on their combined risk scores. For example, Cluster 1 contains articles such as 'Unification of Moldova and Romania' (English entry) and 'Moldovlased (Estonian), analysed previously. Both display elevated risks across all three indicators, signalling potential manipulation. In a nutshell, politically charged, historical, and/or identity-linked topics are most vulnerable to disinformation tactics that blend source-based fragility with behavioural disruptions. Cluster 2 includes pages such as 'Tiraspol' and 'Transnistria War' where the dominant signals stem from manipulation metrics (i.e.: coordinated editing, anonymous contributions, and revert-heavy histories), which points to active efforts of shaping content through editorial control.



PROMPT

www.disinfo-prompt.eu