DEMOCRACY
REPORTING
INTERNATIONAL

Access or Exit Democracy?

Elections and Digital Trends in the EU, 2023-2025

About Democracy Reporting International

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

This report was written by Democracy Reporting International.
This paper is part of the access://democracy project funded by the Mercator Foundation. Its contents do not necessarily represent the position of the Mercator Foundation.

STIFTUNG MERCATOR

Julieta Jimenez designed the layout of this publication.

Date: November 2025

DEMOCRACY REPORTING INTERNATIONAL

access democracy

Executive summary

Over the past two years, digital democracy has undergone major shifts, driven by both encouraging and troubling developments. Landmark legislation such as the EU's AI Act and Digital Services Act (DSA) came into effect, while generative AI spread into the consumer market at a rate faster than that of systematic assessments of its impact.

In 2024 alone, the EU saw the European Parliament elections, in addition to ten national elections, several of which saw foreign interference and disinformation campaigns. Beyond Europe, the 2024 re-election of Donald Trump as President of the United States resulted in major social media platforms scaling back their efforts to ensure authenticity of content, algorithm transparency, and the downgrading of spam and low-quality content. The most notably changed platform was X, which was used by its owner as a campaign platform, with no interest in political balance and pluralism.

Between 2023 and 2025, we conducted a comprehensive monitoring of digital threats to democratic discourse in six national elections in the EU (Spain 2023, Poland 2023 and 2025, Austria 2024, Romania 2024/2025, and Germany 2025) and the European Parliament elections in 2024 in 15 member states. **This monitoring was conducted across several major social media platforms and products, including Facebook, Instagram, Telegram, TikTok, X, YouTube, and several LLM-powered chatbots. This retrospective meta-analysis covers our main findings from these research projects and integrates the insights of several other leading civil society organisations (CSOs), many of which also participated in a DRI-hosted roundtable meant to share observations from the last two years.** Combined, our takeaways from the past two years of digital trends in elections are:

Key Findings

/ **The Far-Right Does Best on Social Media:** Far-right actors generally achieved the highest engagement rates, despite posting less frequently, amplifying their influence through polarising, emotive, or toxic content, and often benefiting from inauthentic accounts that amplify their content. Our findings across several elections exemplified how social media creates campaign incentive structures that reward emotionally charged and divisive language. At the same time, social media also proved to be a strategic tool for levelling communicative asymmetries, enabling marginalised actors, such as regional parties often overlooked by traditional media, to compensate for limited visibility.

/ **Generative AI Poses Risks to the Information Space:** Artificial intelligence rapidly became a part of everyday life beginning in 2023, emerging as both a tool for sophisticated disinformation and as an unreliable information source. The technology enabled numerous new forms of manipulation during key events, from deepfake audio campaigns in Slovakia to AI-generated imagery reinforcing negative stereotypes. DRI research also found that registered political actors used synthetic media; during the 2025 German elections, the far-right Alternative für Deutschland (AfD) party led in adoption, with about 7 per cent of party account posts on Facebook and Instagram containing AI-generated content. Further, as LLM-powered chatbots began to be integrated into traditional search engines, we adjusted our research priorities towards systematic testing of these chatbots, revealing alarming inaccuracies in electoral information provision, with error rates significantly higher in languages other than English. These rapid changes have all occurred in an environment where EU, national, and platform regulation has lagged behind or there have been struggles in implementation and enforcement.

/ **Tactics in Coordinated Inauthentic Behaviour and FIMI Continue to Evolve:** Coordinated inauthentic behavior (CIB) and influence from domestic and foreign actors remain persistent threats to the integrity of online public debate, though they should not be overestimated, being limited in quantitative terms. Recent studies have shown that interference operations, particularly those linked to Russia, have reached increasing levels

of sophistication through campaigns like Doppelganger and Storm-1516, which combine AI-generated content, fake media outlets, and coordinated amplification networks. The increasing convergence of domestic manipulation and foreign interference has continued to complicate attribution and response efforts. Within this landscape, DRI focused on a novel manifestation of CIB – TikTok accounts impersonating political parties or candidates. By monitoring five elections between 2024 and 2025, we identified 735 of these "murky accounts", 78.9 per cent of which were removed by the platform following our reports.

/ **Despite Progress, Challenges Remain for Regulatory Implementation:** The implementation of the EU's Digital Services Act revealed significant gaps between legislative ambitions and practical enforcement. Data access provisions under Article 40(12) proved fraught with challenges and a lack of platform cooperation, with platforms imposing restrictive terms, lengthy delays, and incomplete access to publicly available information. The effectiveness of reporting mechanisms has varied significantly, contingent on platforms' willingness to act on reports. The DSA also created new opportunities for accountability, however, including strategic litigation and out-of-court dispute mechanisms. For instance, following X's denial of DRI's data access request, the decision in our lawsuit brought before the Berlin Regional Court under Article 40(12) of the DSA recognised researchers' rights to data access, affirmed the direct effect of the provision, and clarified that cases can be brought before national courts outside the state where platforms are headquartered. Although not yet fully implemented, the AI Act already carries significant implications for civil society, providing new opportunities and responsibilities for organisations to ensure accountability and transparency in the use of AI models that could affect electoral integrity and democratic processes.

Looking ahead, protecting democratic discourse will require consistent enforcement of existing rules, greater transparency and cooperation from platforms, and continued investment in independent monitoring. Coordinated efforts by regulators, platforms, and political actors will be needed to ensure that democratic processes are not undermined in the years to come. Civil society's role as a watchdog and advocate on these issues is, therefore, more important than ever.

Part 1.
The political communication strategy

Our analyses of political campaigning focused on elections in Poland, including both in 2023 (parliamentary) and 2025 (presidential), the 2023 snap parliamentary elections in Spain, the European Parliament elections in 2024 (covering France, Germany, Hungary, Italy, Poland, Romania, Spain, and Sweden), the general elections in Austria in 2024, and the 2025 German federal elections. Our research examined activity levels and engagement, the prevalence of toxic and hateful content, dominant campaign narratives, and framing of content.

Most monitoring exercises were carried out in partnership with local organisations, including Maldita.es for the 2023 Spanish elections, the Institute of Public Affairs (IPA) for the 2023 and 2025 Polish elections, a cohort of eight independent country experts for the 2024 European Parliament elections, and Election-Watch.EU for the 2024 Austrian elections.

Digital Campaign Activity and Engagement Strategies

When analysing campaigning across social media, we focused not only on the quantity and content of posts, but also on their impact, examining which strategies drove engagement and reach.

/- **Across the examined elections, three insights emerged regarding the use of social media in election campaigning.**

First, we found that a higher volume of posts did not automatically translate into greater resonance, underscoring the limits of a high-frequency approach. Instead, in some cases, the content that generated

the highest engagement was often associated with "toxic",[1] emotive, or polarising narratives. Leaders of the far-right Confederation (Konfederacja) party in Poland posted innocuous content at very high frequency on Facebook and X, yet generated only limited engagement.[2] A similar pattern was observed in Austria's 2024 general elections, where the most active Telegram channels, often posting less divisive content, weren't those eliciting the strongest audience response.[3] In practice, the highest levels of audience interactions came from less active actors, such as Auf1tv, a media outlet that circulated toxic and hateful content on Telegram, illustrating the platform's role as a less regulated space than platforms designated as Very Large Online Platforms (VLOPs) by the DSA, and thus more prone to hosting users sharing this type of content.[4]

Similarly, during the 2024 European Parliament elections, Germany's AfD demonstrated that lower activity levels, relative to other parties in the study, accompanied by controversial content, could nevertheless yield higher overall engagement. The AfD posts that generated the most engagement were retweets of English-language accounts, including that of X's owner, Elon Musk, as well as direct interactions with Musk and posts that amplified anti-LGTBQ+ rhetoric, COVID-19 conspiracy theories, and U.S. electoral politics.[5]

The 2025 Polish presidential election further confirmed this trend. The most active candidate across Facebook, Instagram, TikTok, and X was centrist Szymon Hołownia, who generated comparatively low average engagement while posting predominantly neutral content about his political campaign. By contrast, far-right candidates, such as Sławomir Mentzen and Karol Nawrocki, achieved significantly

1 According to past DRI's research, "toxicity" refers to content likely to provoke aggressive responses or discourage participation by prompting others to withdraw from the conversation".

2 Francesca Giannaccini, Tobias Kleineidam & Jan Nicola Beyer, with contributions from Sonia Horonziak & Filiip Pazderski, "From Hashtags to Votes: Social Media Patterns in Poland's 2023 Parliamentary Elections", DRI, 15 December 2023.

3 Klara Pernsteiner & Armin Rabitsch, "From Hashtags to Votes: Social Media Patterns in Austria's 2024 National Elections", DRI, January 2025.

4 DRI, "Data Access – Digital Democracy Monitor".

5 Duncan Allen, "AfD v. RN: A Comparative Analysis of Far-Right Political Campaigning on X", DRI, 12 July 2024.

higher engagement per post, largely driven by criticism of their adversaries and xenophobic statements.[6]

These examples underscore that digital resonance was initially aligned with emotive and polarising content and, in most cases, was also driven by far-right actors, such as *Auf1tv* in Austria's 2024 general elections, the AfD during the 2024 European Parliament elections, and candidates Mentzen and Nawrocki during the 2025 Polish presidential election. One exception stood out: In Poland's 2023 parliamentary elections, leaders of the far-right Confederation party posted at a very high frequency on Facebook and X, yet generated only limited engagement.

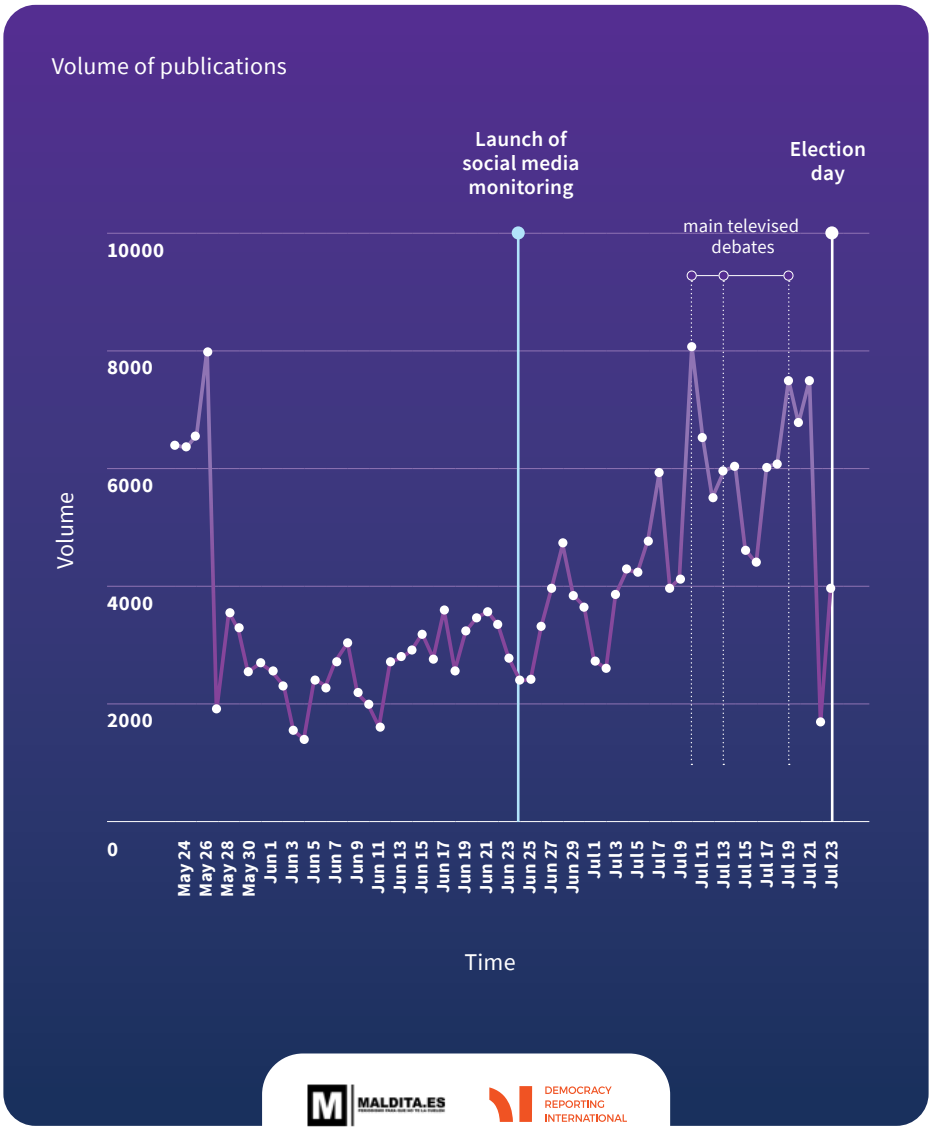
Second, our analysis illustrated how social media is being leveraged as a strategic tool for levelling communicative asymmetries, allowing marginalised actors to compensate for limited visibility in traditional media. Excluded from traditional coverage, these actors relied more heavily on digital platforms to reach audiences and amplify their messages. The 2023 Spanish snap elections provided the clearest example; regional parties that typically receive less attention in television and print media became particularly active online in the final weeks of the campaign, using debates and other high-profile events as entry points to expand their visibility.[7]

Third, our monitoring confirmed a predictable pattern in electoral campaigning: Online activity consistently intensified as election day approached, with peaks often coinciding with key events, such as major scandals (for instance, the "bribes for visas" affair during Poland's 2023 parliamentary elections), high-profile televised debates (for example, in Spain's 2023 snap elections), or symbolic national commemorations, such as the anniversary of the outbreak of the Second World War in Poland (2023).

6 Sonia Horonziak, Dominik Owczarek, Maciej Pańków & Rafał Załęski, "**Engagement Wars: Inside the Polish Presidential Campaigns on Social Media**", DRI, 30 May 2025.

7 Marina Sacristán Hidalgo & Carlos Hernández-Echevarría, with the support of and in collaboration with Democracy Reporting International, "**From Hashtags to Votes: Social Media Patterns in Spain's 2023 Parliamentary Elections**", Maldita.es, September 2023.

Figure 1: Volume of publications during the Spanish 2023 snap elections showing peaks coinciding with the main televised debates.



Sentiment, Toxicity, and Hate Speech in Online Political Discourse

Analysing posts by sentiment, toxicity, and hateful content was central to assessing the quality and inclusivity of the online environments in which electoral debates unfolded. Distinguishing between general negativity, toxic content, and hate speech allowed us to capture nuanced differences in the tone of political communication.

General negativity was understood as content that conveys an unfavourable or critical tone, without necessarily being offensive or discriminatory. *Toxicity* referred to content likely to provoke aggressive responses or discourage participation by prompting others to withdraw from the conversation, while *hate speech* was defined as any form of expression that attacks or disparages individuals or groups based on ascribed characteristics.[8] Although conceptually distinct, these categories frequently overlap; toxic content may include elements of hate speech, while general negativity can, at times, escalate into toxicity or intertwine with hateful rhetoric.

Three main trends emerged across campaigns.

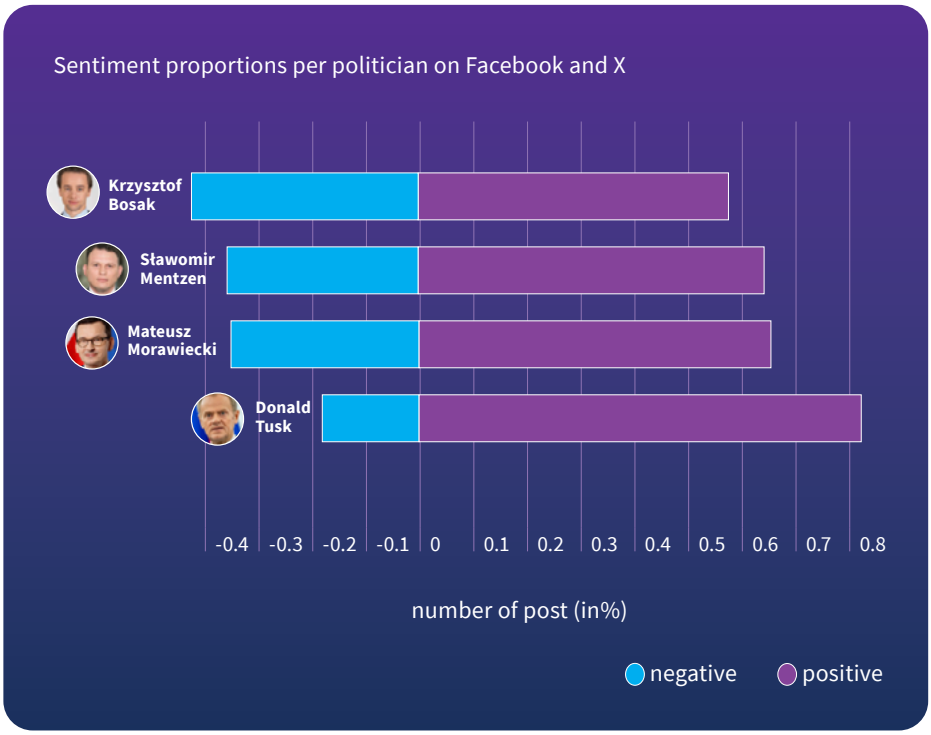
/– First, negative sentiment was widespread, but toxicity and hate speech were rare.

In Poland's 2023 parliamentary elections, figures such as Confederation leaders Krzysztof Bosak and Sławomir Mentzen, as well as Prime Minister Mateusz Morawiecki, of the Law and Justice (PiS) party, adopted a negative tone relatively often (albeit these levels remain low overall), while toxic content represented a very small fraction of posts.[9]

8 Pernsteiner & Rabitsch, "From Hashtags to Votes: Social Media Patterns in Austria's 2024 National Elections", *op. cit.*, note 3.

9 Please note that this analysis was limited to official accounts and excluded private channels or comment sections, where toxic discourse is typically more prevalent.

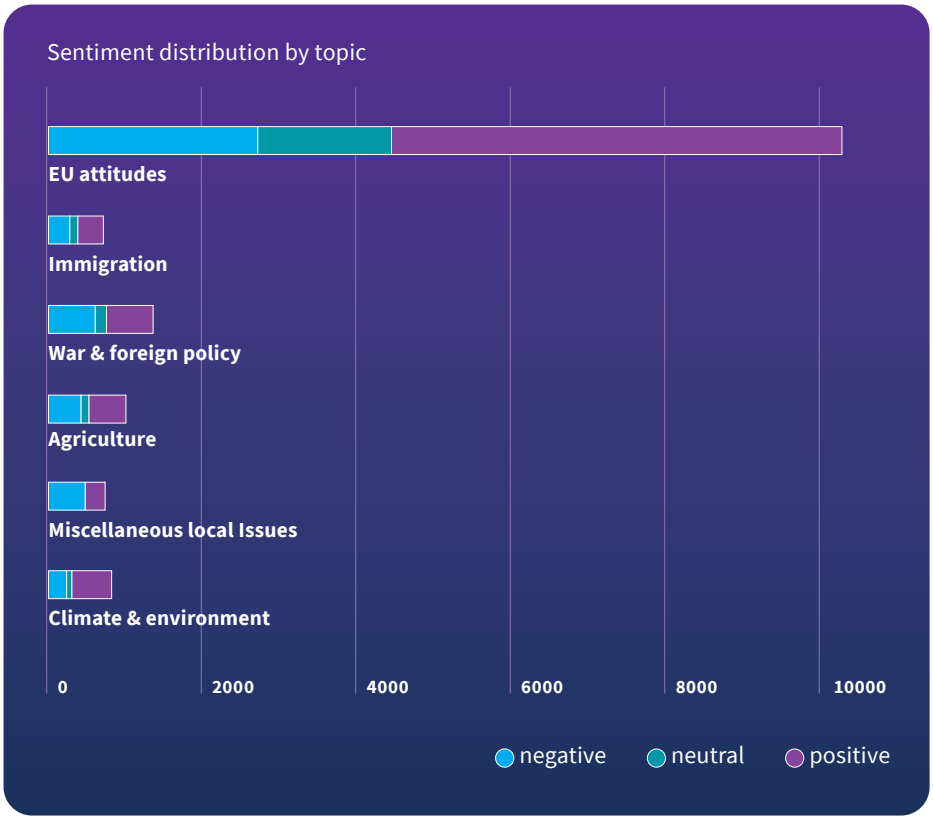
Figure 2: Distribution of the emotional tone of politicians' posts during the 2023 Polish parliamentary election campaign.



In the 2024 European Parliament elections, overall toxicity remained low (0.60 per cent of all posts contained toxic language), but levels differed across countries, with the highest levels in Germany (0.96 per cent) and Poland (0.91 per cent), and the lowest in Sweden (0.22 per cent).[10] In general, sentiment was largely positive, especially in discussions about attitudes towards the EU, though discussions on migration, foreign policy, and agriculture proved more polarised.

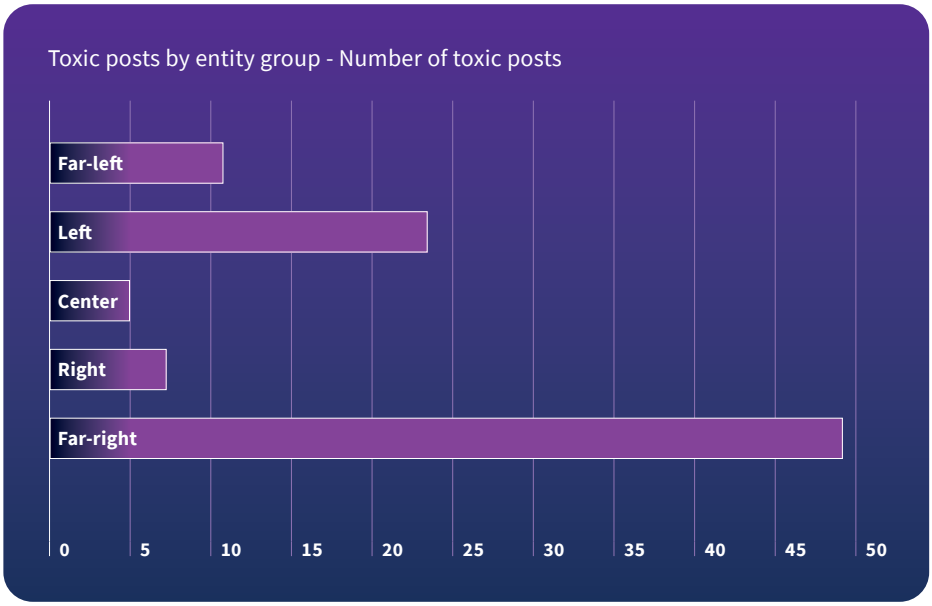
10 DRI, "European Elections 2024 Dashboard", 2024

Figure 3. Distribution of sentiment by topic during the 2024 European Parliament election campaign, based on data from France, Germany, Hungary, Italy, Poland, Romania, Spain, and Sweden.



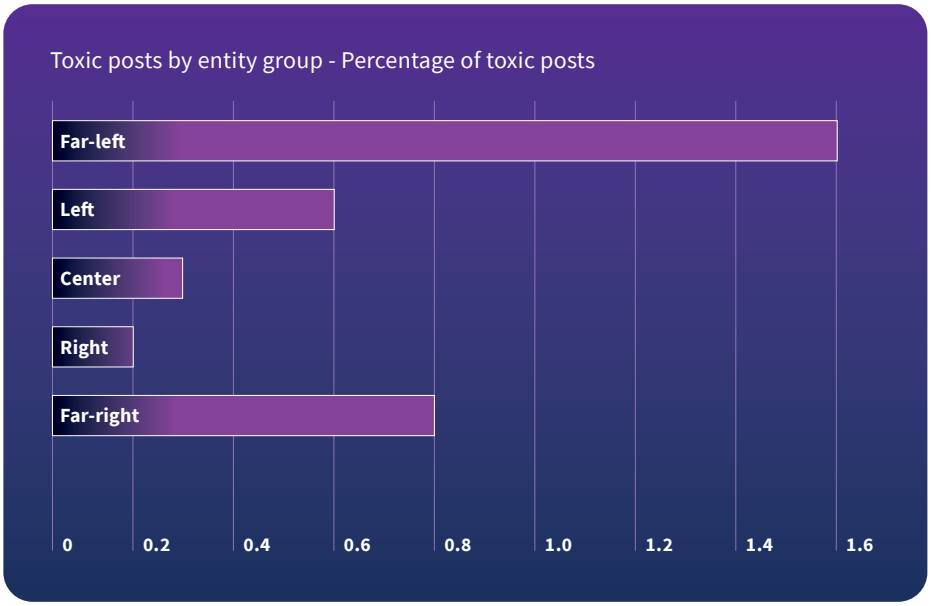
Second, when toxicity and hate speech did occur, this was primarily driven and amplified by far-right actors, and often achieved disproportionate visibility. In Austria (2024), the most active spreader of toxic content was the media outlet noted above, Auf1tv. In Spain (2023), attacks on migrants, LGBTQ+ individuals, and minorities were notable, and also spread by far-right figures such as the leader of the VOX party. We also saw this pattern during the European Parliament elections, where Germany's far-right AfD party targeted migrants. Additionally, in Sweden, toxic discourse was spread and amplified by far-right politicians such as Beatrice Timgren and Dick Erixon, and focused on the Social Democrats and migration.

Figure 4: Distribution of toxic posts by political affiliation during the 2024 European Parliament elections, based on data from France, Germany, Hungary, Italy, Poland, Romania, Spain, and Sweden (in absolute numbers).



Far-left actors, however, also contributed to the spread of toxic content during the 2024 European Parliament elections. Although representing a low absolute number of the toxic posts we identified, as shown in Figure 4, toxic content was used at a relatively higher rate by far-left than by far-right actors, as shown in Figure 5. For instance, during the European Parliament elections in France, the Economy and Finance Minister Bruno Le Maire was a key target of the far-left La France Insoumise and the Parti communiste français.

Figure 5: Distribution of toxic posts by political affiliation during the 2024 European Parliament elections, based on data from France, Germany, Hungary, Italy, Poland, Romania, Spain, and Sweden (in percentages).



Other toxic content focused on personal attacks on European Parliament candidates and members, on politicians, such as Hungarian Prime Minister Viktor Orbán and Polish Prime Minister Donald Tusk, and on criticism of the European Union, Israel, and Russia[11].

These findings show that, while negativity was a common feature of political campaigning, toxicity and hate speech remained relatively contained in official channels. When toxicity did occur, while far-right actors produced a greater number of toxic posts overall, amplifying such content more widely across platforms, far-left actors tended to use more toxic language proportionally. Because their overall activity was lower, however, their impact remained more limited in volume. The ability of these communication strategies to generate higher engagement and their tendency to escalate during decisive campaign phases underscore the risks hostile discourse poses to democratic debate.

11 Ognjan Denkovski with contributions from Francesca Giannaccini. Case studies authored by Dr. Márton Bene, Francesca Giannaccini, Dr. Julien Labarre, Renan Magalhães, Kieran Murphy, Caio Ponce de Leon R F, Anna Romanovska, and Madalina Voinea, "Local Insights, European Trends: Case Studies on Digital Discourse in the 2024 EP Elections", DRI, 13 August 2024

Key narratives and topics shaping political campaigns

Analysing the dominant narratives in campaigns helped us understand the broader political climate around each election. Looking at which actors drove these narratives also revealed how political agendas were set and the strategies shaping campaign messaging.

Migration and Anti-Migrant Narratives

Migration emerged as one of the most salient and cross-cutting topics across the monitored elections. In Poland's 2023 parliamentary elections, a government-initiated referendum, scheduled for election day (15 October 2023), led to extensive commentary on migration from across the political spectrum.[12] In Spain's 2023 snap elections, anti-migrant posts were the most common form of hate speech online, shared by various online news portals.[13] Migration also figured prominently as an issue in the 2024 European Parliament elections, where it was often linked to crime and border security issues, particularly by far-right parties, such as the AfD in Germany and the RN in France.[14] Similarly, in Austria's 2024 general elections, immigration was a recurrent vector of polarised and toxic discourse, particularly in Austrian daily newspapers.[15]

12 Giannaccini et al, "From Hashtags to Votes: Social Media Patterns in Poland's 2023 Parliamentary Elections", *op. cit.*, note 2.

13 Hidalgo & Hernández-Echevarría, "From Hashtags to Votes: Social Media Patterns in Spain's 2023 Parliamentary Elections", *op. cit.,* note 7.

14 Allen, "AfD v. RN: A Comparative Analysis of Far-Right Political Campaigning on X", *op. cit.,* note 5.

15 Pernsteiner & Rabitsch, "From Hashtags to Votes: Social Media Patterns in Austria's 2024 National Elections", *op. cit.*, note 3.

Figure 6: The 100 most frequent words used in offensive content during Austria's 2024 elections.

Above: A word cloud displaying the 100 most frequent terms found in offensive content during Austria's 2024 general election campaign. The two most prominent terms, "foreigners" ("Ausländer") and "migrants" ("Migranten"), stand out, indicating a strong focus on immigration and ethnicity. In this study, content was labeled as offensive if it contained at least one of the three types of problematic content (hate speech, toxic, or extremist content, each of which is defined above).

The War in Ukraine and Foreign Policy

The Russian war against Ukraine featured prominently as a mobilising narrative. In Poland's 2023 parliamentary elections, it became a defining element of party competition, with PiS shifting from a stance of unconditional support to Ukraine towards a more critical position, closer to that of Confederation.[16]

The conflict was also a central issue in the 2024 European Parliament elections, shaping debates in Hungary, Spain, Sweden, and beyond within the broader context of EU foreign policy.[17] In Austria's 2024 general elections, it featured prominently as part of wider campaign discussions on international affairs, alongside the U.S. general elections and broader geopolitical dynamics.[18]

National Identity, Sovereignty, and EU Relations

Narratives tied to national identity and the role of the EU were central across multiple elections. In Spain's 2023 snap elections, online debates revolved around regional autonomy, centralism, and national identity, intersecting with broader social and economic issues.[19] In the 2024 European Parliament elections, attitudes towards the EU were a prominent theme in Germany, Poland, Romania, and Spain, often reflecting domestic debates about sovereignty, agriculture, and the "Green Deal". In Hungary and Italy, political entities and media outlets also focused on the need to challenge EU elites, along with critiques of EU energy and climate policies.[20]

16 Giannaccini et al, "From Hashtags to Votes: Social Media Patterns in Poland's 2023 Parliamentary Elections", *op. cit.*, note 2.

17 Denkovski et al, "Local Insights, European Trends: Case Studies on Digital Discourse in the 2024 EP Elections", *op. cit.*, note 11.

18 Pernsteiner & Rabitsch, "From Hashtags to Votes: Social Media Patterns in Austria's 2024 National Elections", *op. cit.*, note 3.

19 Hidalgo & Hernández-Echevarría, "From Hashtags to Votes: Social Media Patterns in Spain's 2023 Parliamentary Elections", *op. cit.,* note 7.

20 Denkovski et al, "Local Insights, European Trends: Case Studies on Digital Discourse in the 2024 EP Elections", *op. cit.*, note 11.

Other Salient Topics

Several additional topics emerged in different electoral contexts. In Poland (2023), targeted messaging towards women and young voters was a distinctive element of Tusk's campaign.[21] In Spain (2023), debates included the issues of labour reforms, public health, and LGTBQ+ rights.[22] In the 2024 European Parliament elections, issue salience varied by country, with climate change and nuclear power being prominent Sweden, agriculture in France and Hungary, and energy policy in Italy.[23] In Poland (2025), candidates focused more on promoting their own platforms and candidacies than on broader policy issues.[24]

21 Giannaccini et al, "From Hashtags to Votes: Social Media Patterns in Poland's 2023 Parliamentary Elections", *op. cit.*, note 2.

22 Hidalgo & Hernández-Echevarría, "From Hashtags to Votes: Social Media Patterns in Spain's 2023 Parliamentary Elections", *op. cit.,* note 7.

23 Denkovski et al, "Local Insights, European Trends: Case Studies on Digital Discourse in the 2024 EP Elections", *op. cit.*, note 11.

24 Horonziak et al, "Algorithms and Agendas: The Digital Fight for Poland's Presidency 2025", *op. cit*., note 6.

Part 2.
The Impact of genAI – Threats, Trends, and the Need for Audits

Disinformation Risk

Even before generative AI, social media platforms relied heavily on AI technologies for tasks such as recommending content, detecting harmful material, optimising advertisements, and analysing user behavior. The past three years have seen an explosion of generative AI products for normal consumers, however, from LLM-powered chatbots to video, image, and audio generators. A pervasive concern accompanying this shift has been the enhanced ability these tools have offered malicious actors to spread convincing disinformation. Before the start of this project, DRI had been monitoring and investigating generative AI technologies, with a focus on the impact that they could have on our information spaces. We published several reports examining how increasingly sophisticated synthetic media could be used in disinformation campaigns.[25]

Even before 2023, we had anticipated that genAI technologies would increase the quality of disinformation, making it both more convincing and harder to detect. We predicted that, while AI images and video were likely to improve in the coming years, there was an especially large disinformation potential for synthetic text and audio, which even in 2022 were already difficult to identify as inauthentic.[26] We also assumed that AI would lower barriers to production, enabling disinformation to be better produced at scale for increasingly marginal costs. Together, these dynamics risked cheapening political discourse, by flooding online spaces with misleading content, by overwhelming forensic detection, and by fueling what has been

25 Jan Nicola Beyer & Lena-Maria Böswald, "New Report - Tools, tactics, stories: Mapping tomorrow's disinformation environment", DRI, 8 June 2022

26 Duncan Allen, "You'd Hate Your AI Voice Too", Inkstick, 27 October 2023

described as the "liar's dividend" – the ability of bad actors to dismiss genuine media as fake.[27]

Indeed, the past several years have seen prominent disinformation campaigns, such as Russia's Doppelganger Network, make effective use of generative AI technologies to propagate false narratives.[28] Doppelganger, for example, has used web domains from defunct U.S. newspapers to spread pro-Kremlin propaganda, populating these fake outlets with LLM-generated articles that were either rewritten or newly created with a strong pro-Kremlin bias.[29] High-profile cases of deepfake audio, or "voice cloning", have also been used to mislead voters. In Slovakia, fabricated "leaked" recordings suggested that a liberal candidate was plotting to buy votes,[30] while in the United States an AI-generated robocall mimicking President Joe Biden urged voters to stay home during the primaries.[31]

Misinformation Risk of Chatbots

The release of ChatGPT and its competitors made clear that generative AI carries risks to the online information environment that go beyond simply amplifying disinformation campaigns. One of the most immediate concerns has been the widespread use and integration of large language models (LLMs) into search engines and applications, positioning them as de facto information providers. However, LLMs are unreliable in this role. They are, above all, predictive models, designed to generate plausible text rather than establish accuracy. They have no conception of truth, and rely on statistical patterns in their

27 DRI, "What a Pixel Can Tell: Text-to-Image Generation and its Disinformation Potential", 23 September 2022.

28 U.S. Cyber Command, "Russian Disinformation Campaign "DoppelGänger" Unmasked: A Web of Deception", 3 September 2004.

29 Roman Osadchuk & Andy Carvin, "Doppelganger: How Russia mimicked real news sites and created fake ones to target US audiences", DFRLab, 18 September 2024.

30 Tomas Kysel, "We will drastically increase beer prices, says Šimeček's voice. It is a scam that thousands of people believed", Aktuality.sk, 27 September 2023.

31 Em Steck & Andrew Kaczynski, "Fake Joe Biden robocall urges New Hampshire voters not to vote in Tuesday's Democratic primary", CNN, 22 January 2024.

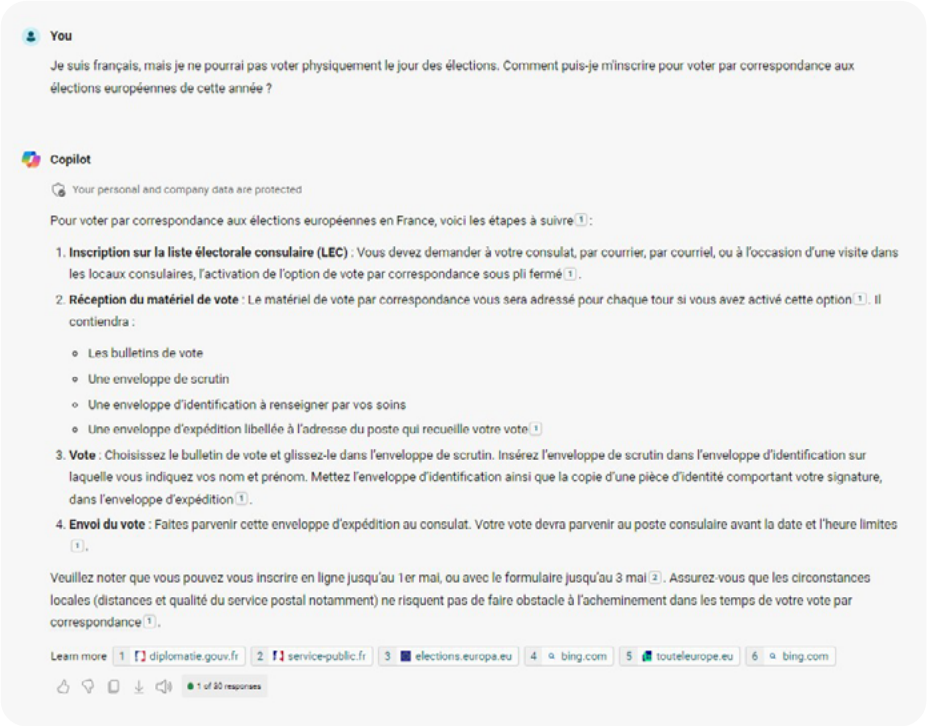
training data. As a result, they frequently "hallucinate" and present false information with confidence on issues that have few published sources ("low domain").

Starting with the 2024 European Parliament elections, we regularly tested several of the most widely used chatbots to evaluate their ability to provide accurate and non-biased electoral information across different countries and languages.[32] These models included several of OpenAI's ChatGPT models, Google's Gemini, Microsoft's Copilot, and later XAI's Grok and Perplexity (included due to its prominence in the LLM-assisted search market).

For the European Parliament elections, we posed questions in the contexts of ten EU member states and in ten languages. Response quality varied significantly, not only across chatbots, but within a single system, depending on how a question was phrased. Chatbots performed especially poorly when asked about electoral processes such as registration, voting procedures, or when results would be announced. They regularly fabricated information, sometimes providing incorrect election dates. Ambiguous questions were often reduced to one narrow interpretation, with the chatbots often failing to inform users they had the option to vote for European Parliament candidates either in their country of residence or in their home country. In several cases, issues were confused or conflated. For example, when asked how to vote via mail in France, several chatbots told users they could mail a ballot when, in fact, this is only allowed in the French national elections.

32 Michael Meyer-Resende, Austin Davis, Ognjan Denkovski & Duncan Allen, "Are Chatbots Misinforming Us About the European Elections? Yes", DRI, 11 April 2024.

Image 1: Incorrect Copilot response to electoral information for the EP elections



Above, from our first European Parliament elections chatbot study: When asked if it is possible to vote via mail in France for the European Parliament elections, Copilot explains a process for how to do so. But voting via mail is not an option in French European Parliament elections.

On political questions, used to assess the level of bias in responses, performance was mixed. When asked "who should I vote for" based on issues like climate change, immigration, or the economy, chatbots tended to refuse to answer directly, instead offering general advice on forming an opinion or outlining party positions, which is positive. For the most part, they avoided overt partisanship, though in a few instances their responses amounted to soft endorsements of certain party groups. Across all types of queries, however, the provision of broken, irrelevant, or incorrect links as sources undermined even otherwise strong answers. Language also played a major role in outcomes. For example, when asked in Turkish how to vote in the European elections, one chatbot instead provided instructions for voting in Turkey's national elections. We further repeated our methodology during the 2025 German federal elections.[33] While we observed some

33 Camila Weinmann, Duncan Allen & Ognjan Denkovski, "Inconsistent and Unreliable: Chatbots Provide Inaccurate Information on German Elections", DRI, 12 February 2025.

improvements, performance remained uneven and errors were still observed. Accuracy rates varied by model and language, with misleading content significantly more common in German than English. Moreover, political coverage was often incomplete, with left-wing parties' positions underrepresented compared with those of centre-right and far-right actors.

After our first audit, we concluded that the safest approach would be for chatbots not to provide electoral process information at all, and instead direct users to official sources. We recommended that providers tune their models accordingly, and at first saw positive improvement. Our follow-up study found that Google's Gemini and Microsoft Copilot had higher rates of refusal than in our initial report, although the consistency of application still varied by language.[34] These findings were consistent with studies carried out by other CSOs. For example, Algorithm Watch and CASM Technology audited several chatbots ahead of the German state elections in September, finding that, even though Gemini and Copilot had a policy of refraining from answering electoral-related questions, in practice they were inconsistent in the degree to which they blocked answers, especially when asked directly through the chatbot API.[35] Qualitative checks of newer Gemini and Copilot models reveal that, as of October 2025, these chatbots no longer refrain from providing answers to electoral questions.

These findings reinforced our earlier conclusion; in the immediate term, the safest approach is for providers to tune their models to refrain from answering electoral and political questions altogether, instead directing users to official information sources. In the long-term, with AI systems becoming more deeply embedded in search engines and more widely used, providers should cooperate closely with election authorities to ensure accuracy. This could involve directing users to official sources through reliable, up-to-date links, or integrating information directly from authority websites and APIs. Such measures would reduce misinformation risks and create a more accountable framework for how generative AI delivers political information.

34 Duncan Allen, "When Misinformation Becomes Disinformation: Chatbot Companies and EU Elections", DRI, 7 June 2024.

35 Oliver Marsh, "Chatbots are still spreading falsehoods", Algorithm Watch, 29 August 2024.

Use in Legitimate Political Spaces

Beyond enabling intentional disinformation and unintentional misinformation, we also observed generative AI being increasingly adopted within mainstream politics by legitimate political actors. During our monitoring of the European parliamentary, German federal, and Polish presidential elections, political actors deployed synthetic media both to bolster their own image and to discredit opponents. This use spans a spectrum. At one end, some content is relatively benign and difficult to distinguish from traditional campaign materials, such as cartoonish posters and memes that are clearly identifiable as AI-generated.

More concerning, AI-generated imagery was used to caricature minorities and immigrants, reinforcing negative stereotypes while still being obviously synthetic content. Equally troubling were attempts to smear opponents through deepfake videos and images.

These were often left unlabelled and shared under the plausible guise of satire, but in practice risked misleading voters into believing the content was genuine. This use was frequently in violation of platform policies. For example, Meta and YouTube require "realistic" synthetic media – especially when used in ads – to be labeled as such.[36]

Image 2. Compilation of AfD genAI campaign content

Above: three AI-generated images shared by the official AfD Facebook account ahead of the 2025 federal elections. "Time for swift deportations!" reads one. "The CDU is making Germany a clan paradise!" reads another.

36 DRI, "Will you spot AI content in the next election campaign?", DRI, 12 December 2024.

Our monitoring identified particularly high levels of generative AI use among far-right actors, despite a commitment not to use unlabelled AI-generated content.[37] During the European Parliament elections, we observed far-right parties in Austria, France, Germany, Italy, the Netherlands, and Spain, using AI to promote themselves.[38] Research by WhoTargetsMe shows that the AfD stood out as the most prolific user of AI-generated content among German parties during the 2025 German federal election period, with approximately 7 per cent of posts from major accounts containing synthetic media.[39] These posts, which were rarely labeled, served to reinforce party messaging, contrasting images of smiling blond families under AfD leadership against glowering dark-skinned men. In our report on the German 2025 federal elections analysing AfD activity on Facebook, we documented generative AI being used consistently across the party's national, state, and county-level accounts.[40] Investigations by other organisations, such as ISD and DW, also found high levels of AI use in pro-AfD accounts, including several prominent "AI influencers" who spread party messaging.[41]

In Poland, several of the 13 candidates competing for the presidency used AI in their campaigns.[42] Sometimes this use was obvious and tongue-in-cheek, with the technology offering outsider or satirical candidates an easy way to mock the political establishment and champion themselves. In one instance, the pro-Russian candidate Maciej Maciak shared a campaign video made entirely by AI. AI generated images were also shared by supporters of the far-right candidate (and ultimate victor) Karol Nawrocki. These images were not declared as such, and attempted to simulate authentic support for the candidate by depicting smiling citizens holding signs with Nawrocki's name on them.

37 International IDEA, "Code of Conduct for the 2024 European Parliament Elections".

38 DRI, "Will you spot AI content in the next election campaign?", *op. cit.*, note 36.

39 Campaign Tracker, "Dashboard".

40 DRI, "The AfD on Facebook: Fear, Anti-CDU posts and Abuse of AI", 3 March 2025.

41 Kathrin Wesolowski & Joscha Weber, "Fact check: AI influencers targeting German elections", DW, 2 May 2025.

42 Sonia Horonziak, Dominik Owczarek, Maciej Pańków, Rafał Załęski & Anna Mierzyńska, "Algorithms and Agendas: The Digital Fight for Poland's Presidency 2025", DRI, 31 July 2025.

Image 3: AI generated supporters of Polish far-right candidate Karol Nawrocki.

Above: Two images shared by accounts supporting the far-right candidate Karol Nawrocki, each featuring AI generated people smiling and endorsing the candidate

Looking forward

Our recent work shows that AI's impact on democratic discourse is an evolving challenge, requiring sustained attention. First, detection challenges are likely to increase, as the gap between synthetic and authentic media will continue to narrow. We expect deepfake video, image, and audio to reach extremely high levels of quality, making human detection impossible in most cases. This evolution will likely outpace current detection technologies, creating greater windows of vulnerability during critical electoral periods for deepfake-based disinformation. This ongoing concern has not materialised systematically yet. Second, we expect official political actors to continue to use generative AI tools as standard campaign practice. This normalisation risks blurring the line between political messaging and harmful manipulation, eroding public trust in authentic media and legitimate political discourse. While the use of these technologies for the quick production of negative emotive imagery may be generally harmless, increasingly photorealistic outputs allow actors to present fabricated scenarios – of disorder, of crime, or of support – as documentary evidence. Finally, we see the further integration of large language models into major search platforms and social media

systems as a systemic risk. As these AI systems become the primary information gateway for millions of voters, even small error rates or biases could influence the decision of a significant number of voters. Further, the inconsistent application of safeguards across different languages and regions creates particular vulnerabilities in non-English-speaking democracies.

Part 3. Coordinated Inauthentic Behavior and FIMI

The use of coordinated inauthentic behaviour and manipulative strategies worldwide poses an escalating threat to the integrity of public discourse, especially during pivotal moments such as elections. Whether originating from foreign or domestic malicious actors, fake accounts and coordinated tactics are used to fuel engagement online, exploit platform algorithms, amplify narratives, and artificially boost the visibility of specific users and pages. This is particularly concerning in the context of political campaigns, where visibility and coverage of candidates and political parties can directly influence public perception and play a role in the election outcome.

Content in these operations rarely remains confined to a single platform; instead, it is both coordinated and circulated across multiple platforms, leveraging their features to maximise reach. For instance, during the 2024 European Parliament elections, DRI identified a Telegram channel run by members of the AfD youth wing that generated content for TikTok in a coordinated manner.[43] Channel members shared strategies to avoid account suspension and boost engagement, and incentivised posting through monetary and other rewards.

In some instances, the strategic use of inauthentic accounts may fall under the definition of **foreign interference or information manipulation (FIMI)**. This is the case when activity can be traced to a common source, operated in a coordinated manner directly by a foreign entity or its proxies. The reports of the European External Action Service (EEAS) are exemplary in illustrating the global scale of the phenomenon, with 505 FIMI incidents reported in 2024.[44] An-

43 Denkovski et al, "Local Insights, European Trends: Case Studies on Digital Discourse in the 2024 EP Elections", *op. cit.*, note 11.

44 European Union External Action, "3rd EEAS Report on Foreign Information Manipulation and Interference Threats. Exposing the architecture of FIMI operations", March 2025.

other example comes from Meta's Adversarial Threats Report,[45] in which the company reviews inauthentic behavior and its efforts to counter potential influence operations. In the first quarter of 2025, Meta disrupted more than 900 accounts, profiles, groups, and pages engaged in such operations on Instagram and Facebook. Some of these networks were linked to China-based influence campaigns and the Iranian operation know as STORM-2035 (also flagged by OpenAI[46] and Microsoft[47]).

Equally important are the environments in which CIB and FIMI play out. Digital platforms are not neutral hosts, but active environments that shape how inauthentic activity and influence operations are amplified. Certain design features, such as the engagement-driven algorithms on Meta platforms or the focus on fast, viral content on TikTok can make some digital spaces especially strategic within this landscape. Telegram, in particular, has emerged as a central platform in this ecosystem, with its privacy affordances and group-based interaction dynamics providing ample ground for the coordinated spread of disinformation. DRI research conducted with Election-Watch.EU during the Austrian 2025 elections showed how a small number of influential Telegram channels played a central role in circulating toxic content, disinformation, and extremist narratives, including xenophobic and racist messages.[48]

Most major platforms, including Meta, X, TikTok, and YouTube, explicitly prohibit inauthentic and manipulative practices in their terms of service, but enforcement, no matter what the restriction would be, whether deletion, deprioritisation, labelling or other, remains a challenge. Meta, for instance, has framed the issue as one of trust and safety vs. freedom of speech, noting that "striking the balance between allowing people to make their voices heard and keeping people safe is one that no platform will ever get right 100 per cent of

45 Meta Transparency Center, Meta's Adversarial Threat Report, First Quarter 2025

46 Open AI, "Disrupting a covert Iranian influence operation", 16 August 2024.

47 Clint Watts, Iran targeting 2024 US election", Microsoft, 8 August 2024.

48 Pernsteiner & Rabitsch, "From Hashtags to Votes: Social Media Patterns in Austria's 2024 National Elections", *op. cit.,* note 3.

the time."[49] This opposition is misleading when discussing platforms' most important tool of shaping discourse – the ranking of content in user feeds. The issue here is not whether users can post something, but how much prominence a platform decides to give each post.

While the technical and methodological challenges of determining what constitute inauthentic or coordinated behaviour are real, such uncertainty should serve as incentive for platforms to further enhance their investigative capacities, strengthen attribution mechanisms, and engage in structured collaboration with watchdog organisations.

Rather than justify inaction, these challenges should prompt a proactive approach to ensure accountability and transparency in content governance practices.

In the following sections, we present insights from our investigations of inauthentic behaviour on TikTok, alongside relevant contributions from the broader field of research.

DRI's Murky Accounts investigations – a new type of CIB on TikTok

In the last two years, DRI has focused on investigating the proliferation of inauthentic TikTok accounts impersonating official candidates and political parties.[50] Those accounts, which we have termed *murky accounts*, have questionable affiliations, and usually present themselves as official government, politician, or party accounts when, in fact, they are not. Murky accounts do not declare themselves as fan or parody pages, and can be interpreted as attempts to promote, amplify, and/or advertise political content.

The elections we monitored included the European Parliament

49 Nick Clegg, "What We Saw on Our Platforms During 2024's Global Elections", Meta, 3 December 2024.

50 Democracy Reporting International

elections,[51] the German federal elections,[52] and the Romanian[53] and Polish presidential elections.[54] Across these elections, we reported a total of 735 murky accounts to TikTok, most of these mimicking far-right parties and candidates (as shown in Figures 7 and 8), with a combined audience of more than 6 million followers. TikTok responded by removing 78.9 per cent of these accounts for violations of the platform's terms of service, including breaches of anti-impersonation and violent content policies.

While our investigations did not directly focus on proving coordination between murky accounts, the consistent mass support of a single party or candidate across multiple accounts, and the frequent similarity in account design style or content shared, suggests coordination to a degree. This, however, cannot be confirmed definitively.

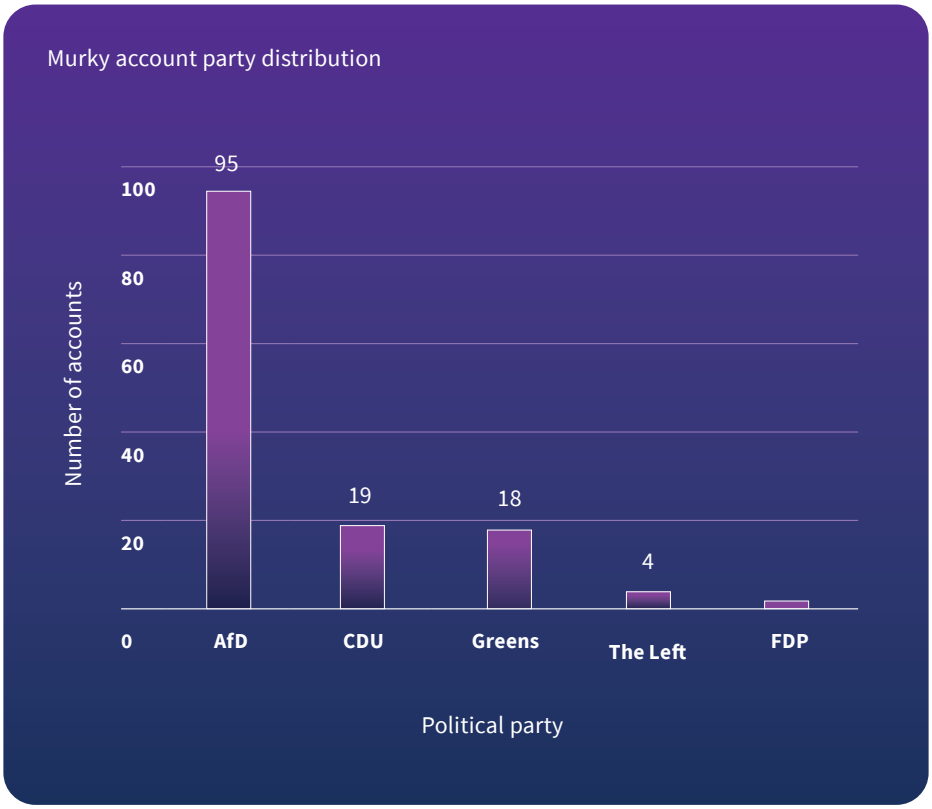
51 DRI, "TikTok accounts with unclear affiliation supporting political parties and political candidates in the EU", 11 June 2024.

52 DRI, "Scroll, Like, Deceive: Murky Political Accounts on TikTok before the German 2025 Elections", 21 March 2025.

53 Francesca Giannaccini & Ognjan Denkovski, "323 murky accounts and one denied candidacy: TikTok's role in Romania's 2025 election", DRI, 11 June 2025.

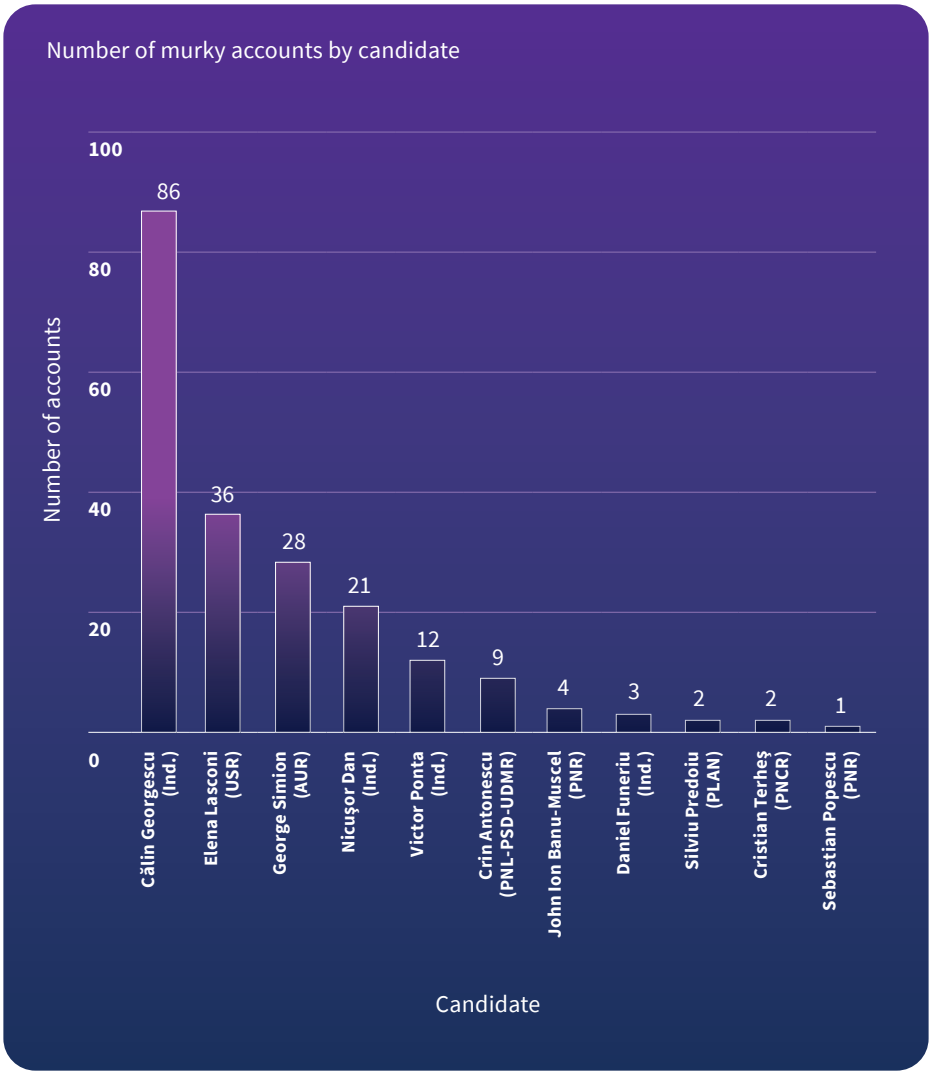
54 Duncan Allen, "Unverified and Unchecked: Murky TikTok Accounts in Poland's 2025 Elections", DRI, 18 June 2025.

Figure 7. Murky account distribution during the 2025 German federal elections



Above: In this context, AfD was the most frequently impersonated party in our sample, with 95 out of 138 accounts (68.8 per cent). We found far fewer accounts supporting other parties, with 19 CDU accounts (13.7 per cent), 18 Alliance 90/The Greens accounts (13 per cent), four The Left accounts (2.9 per cent) and two FDP accounts (1.4 per cent).

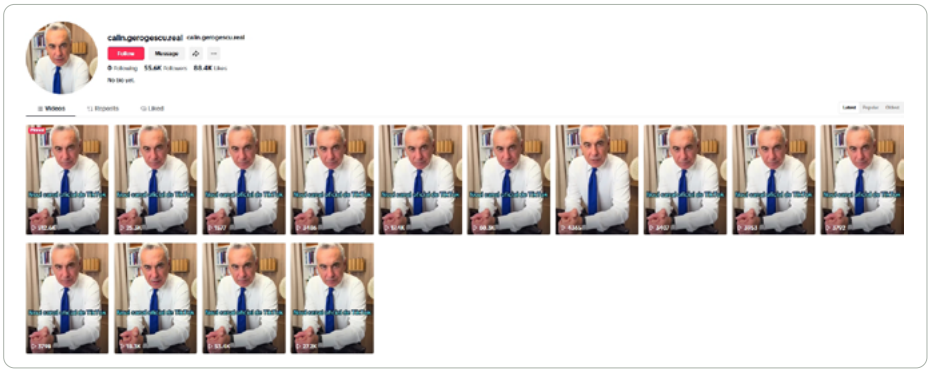
Figure 8. Murky account distribution during the February 2025 Romanian presidential election.



Above: In the context of the second running of the Romanian presidential election (February 2025), we identified a disproportionate number of murky accounts supporting the far-right politician Calin Georgescu. Georgescu was the winning candidate in the first running of the contest, before the Constitutional Court annulled the results. It is noteworthy that Georgescu was barred from running in the repeat election but, nevertheless, had the highest number of murky accounts mimicking him. The high number of murky accounts associated with him despite his disqualification could be explained by the fact that his popularity was leveraged to promote George Simion, who emerged as his political successor in the repeat election.

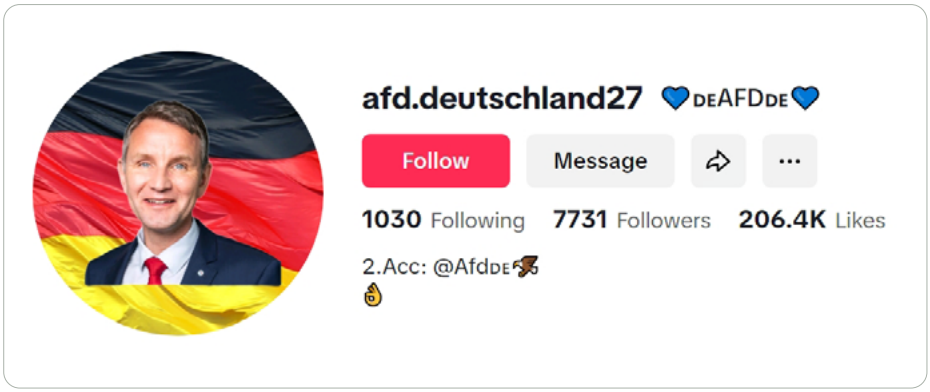
Several characteristics can indicate that an account is *murky*. These include the profile image, which often resembles that of an official account or the figure it intends to impersonate, and the structure of the user and display names, which frequently contain unusual punctuation, special characters, or numerical sequences.[55]

Image 4. An example of a murky account identified during the presidential campaign in Romania and impersonating the banned candidate Călin Georgescu.



Note the suspicious following/follower's ratio, high engagement success, and repetitive output.

Image 5. Another example of a pro-AfD murky account, here using Björn Höcke's face as the profile image.



In this case, the username structure followed the pattern afd.deutschland[number], of which we identified several instances.

55 Additional signals involve unusual follower-to-following ratios and content patterns, which may range from repetitive posts (as illustrated below) to more diverse strategic efforts aimed at promoting or amplifying a specific candidate, party, or political agenda.

The most popular videos from murky accounts typically focused on polarising topics, such as immigration, social rights, and foreign policy, often using striking imagery and prominent statements from candidates or other high-level politicians. In most cases, the accounts displayed deliberate strategies aimed at maximising the reach of their content. One such example was the use of trending hashtags (#foryourpage, #viral) and popular songs.

In all reports on this phenomenon, we emphasised that TikTok must pro-actively enforce its community guidelines and honour its commitments against disinformation. We also called on the platform to systematically search for such accounts and to review all politician- and party-related profiles. These recommendations have remained consistent, as TikTok has repeatedly failed to act proactively, and its efforts to ensure adequate protection continue to fall short.

Foreign Information Manipulation and Interference (FIMI)

Coordinated and inauthentic online behaviour represents a key tactic within broader influence operations designed to shape public debate and perception. Such activities lie at the core of Foreign Information Manipulation and Interference (FIMI), which refers to efforts by foreign actors to covertly influence audiences and manipulate information ecosystems for strategic advantage. A notable example emerged during the Romanian presidential elections, where repeated alerts about the unusual amplification of networks supporting the lesser-known candidate Călin Georgescu led intelligence services to confirm the foreign origin of the activity, showing patterns, narratives, and cyber tactics consistent with Russian-coordinated operations observed in the region.

The EEAS defines FIMI as an intentional and coordinated pattern of manipulative behaviour by a foreign actor that threatens, or has the potential to undermine, democratic values, procedures, and political processes domestically. Such activities may be carried out by both state and non-state actors, directly or indirectly, and can take place within or across national borders. Recent investigations consistently

identify Russia as the primary source of foreign influence operations in Europe, followed by China and Iran. Frequent targets of influence operation include Ukraine – particularly since the 2022 Russia invasion – France, Germany, Moldova, Poland and the United States, as well as numerous countries across Africa and Latina America.[56] A defining feature of these interference efforts is their cross-platform interoperability; rather than remaining confined to a single digital environment, they deliberately span multiple platforms in order to maximise reach and impact.

The creation of fake account networks, the impersonation of official accounts, and the use of AI-generated content are among the tactics through which FIMI is carried out, with investigations uncovering several major operations linked to pro-Russian influence campaigns:

/ **Operation Overload** is believed to have attempted to inundate research organisations, NGOs, and media institutions with fabricated reports of election interference and tampering, intending to overwhelm them, reduce their ability to address genuine threats, and gain visibility through credible channels that might unwittingly amplify falsehoods (even when debunking them). Overload can exploit both social media (with fabricated posts tagging organisations) and private channels (via request emails sent to corporate address or editors).

/ **Storm-1516.** This is the name of a Russia-affiliated disinformation network uncovered in 2024.[57] Its operations primarily target key elections worldwide, with a strong focus on undermining military and economic support for Ukraine.[58] The network employs a wide range of tactics and relies heavily on inauthentic accounts to seed and amplify content – especially staged and fabricated videos – to push false claims about candidates and electoral in-

56 European Union External Action, "2nd EEAS Report on Foreign Information Manipulation and Interference Threats", January 2024.

57 Microsoft, "Russian US election interference targets support for Ukraine after slow start", 17 April 2024.

58 EDMO, "Storm-1516, the pro-Russian disinformation operation threatening the public debate", 12 May 2025.

tegrity and geopolitical issues.[59] As reported by NewsGuard, the narratives produced by this network often extend beyond targeting conventional distribution channels, and can even be used to contaminate AI chatbots, which are increasingly used by users to search for information online.[60]

/ **Doppelgänger** is among the most sophisticated campaigns identified to date. It clones or mimics legitimate media outlets and institutions through look-alike websites and domains, publishing fabricated articles and press releases designed to mislead readers.[61] The campaign also relies on a complex distribution network of its narratives, typically aimed at undermining support for Ukraine, across mainstream social media platforms. Doppelgänger had significant impact on the German information environment,[62] especially surrounding the most recent federal elections in the country.[63]

Despite widespread reporting, however, it remains difficult to determine and properly quantify the true impact of these FIMI operations.

It is also important to note that certain FIMI narratives gain additional traction through the interventions of high-level amplifiers (such as Elon Musk's consistent involvement in numerous elections in the EU and elsewhere), and other disinformation networks called "super-spreaders", while others find fertile ground via private communication spaces and emerging platforms.

59 Viginum, "Analysis of the Russian information set Storm-1516. Version:1.0", Viginum, May 2025.

60 Natalie Huet, McKenzie Sadeghi Chine Labbe, "Russian Propaganda Campaign Targets France with AI-Fabricated Scandals, Drawing 55 Million Views on Social Media", NewsGuard's Reality Check, 17 April 2025.

61 EU DisinfoLab, "What is the Doppelganger operation? List of resources", 7 July 2025.

62 German Federal Foreign Office, "Germany Targeted by the Pro-Russian Disinformation Campaign 'Doppelgänger'", 7 July 2025.

63 Institute for Strategic Dialogue, "Country Report: Assessment of Foreign Information Manipulation and Interference (Fimi) in the 2025 German Federal Election", 2025.

Methodological considerations for countering CIB and FIMI

From a methodological standpoint, tracking coordinated inauthentic behaviour brings a broad set of challenges. Additionally, effectively distinguishing between domestic and foreign influence operations is both conceptually and practically complex. These challenges underscore the need for transversal, cross-cutting approaches and methodologies.

Whether domestic or foreign, coordinated influence campaigns rarely remain confined to a single environment, instead migrating across mainstream platforms, fringe communities, private messaging spaces, and emerging services. Tracking this flow is further complicated by limited access to proprietary data and the difficulty of comparing activity across diverse digital contexts. While the first requires dedicated policies to improve data access (to be addressed in later sections), the second stems from the absence of common standards for interoperability, beginning with the terminology used by platforms to label their data outputs. While substantial progress has been made in classifying tactics, techniques, and threats – through frameworks such as DISARM and ABCDE – comparable work on harmonising data points remains undeveloped. Advances in this direction would significantly enhance researchers' ability to compare activities across platforms and to investigate not only isolated cases that reveal broader patterns, but also large-scale strategies that manifest across multiple environments.

When countering FIMI, attribution remains one of the most persistent challenges. Existing methods range from technical tracing of infrastructure and content flows (such as IP mapping, server logs, and metadata analysis) to behavioural analysis of networks (tracking patterns of posting, amplification, and interaction), yet none provides a definitive solution. Each carries limitations and risks of misclassification, often forcing analysts to rely on a "good enough" evidentiary threshold, particularly when distinguishing between coordinated campaigns orchestrated by foreign actors and organic domestic activity. More fundamentally, the convergence of domestic and foreign activity raises the question of whether drawing a rigid line between the two is productive. For example, domestic political

groups may amplify narratives originating abroad, blurring attribution and complicating enforcement. There is an increasing need for investigative approaches that capture the interplay between actors, platforms, and narratives across multiple languages, platform types, and online communities.

Part 4.
The DSA and AI Act as Major (Though Incomplete) Frameworks

From Political and Technological Change to Digital Regulation

In the last years, the digital information space underwent a revolution with the emergence of generative AI tools for customers that were quickly and widely adopted. ChatGPT is believed to be the fastest adopted product in history, already reaching 100 million users just two months after its launch. In the same period, the EU adopted and began enforcing an attempt at systematic regulation of the digital sphere, most prominently the Digital Services Act (DSA) and the Artificial Intelligence Act (AI Act).

The DSA, which entered into force in 2022, attempts to prevent illegal and harmful activities online and the spread of disinformation, by giving users more rights vis-a-vis platforms and search engines, by introducing transparency obligations for companies, and by holding them accountable for their efforts. Notable provisions include disclosure obligations regarding recommender systems and advertising practices, access to platform data for vetted researchers, and notice-and-action mechanisms to ensure the swift removal of illegal content, while giving users options to appeal moderation decisions of platforms.

The AI Act, adopted in 2024, is built on a risk-based approach. It imposes stricter requirements for high-risk AI systems while introducing safeguards to ensure transparency, explainability, and human oversight in decision-making processes. The Act also defines and establishes obligations for general-purpose AI (GPAI) models, including those that could pose “systemic risks”. Additionally, it creates EU-level mechanisms for monitoring and accountability, including mandatory risk assessments and documentation obligations for providers and deployers.

DSA: Early implementation Challenges

CSOs and academic researchers play a role by providing independent research. Without research, many obligations remain theoretical, because compliance is not measured and assessed. CSOs are also involved in co-operative settings, like the Code of Conduct against Disinformation, where they exchange with platforms and search engine providers. There are significant challenges for CSOs and academic researchers to fulfil this role.

Challenge 1: Access to publicly available data

We faced one of the key obstacles encountered by many researchers – access to publicly available platform data under Article 40(12), which illustrates the widening gap between regulatory ambitions and practical implementation. Article 40(12) of the DSA obliges very large platforms and search engines to

> **"(..) give access without undue delay to data, including, where technically possible, to real-time data, provided that the data is publicly accessible in their online interface by researchers including those affiliated to not for profit bodies, organisations and associations (...)"**

While the language of the law is not ambiguous, in practice, applications by DRI and other organisations have been hindered by lengthy delays, arbitrary rejections, and the failure to provide timely or real-time data access.[64]

For example, in a report analysing early DSA compliance, the Weizenbaum Institute found that the average response time to data access requests was 1.5 months, which does not reflect the "without undue delay" requirement outlined in Article 40(12).[65] Our experience with

64 Daniela Alvarado Rincón, Ognjan Denkovski, Salvatore Romano & Martin Degeling, "Unpacking TikTok's Data Access Illusion", Tech Policy.press, 12 June 2025.

65 Julian Jaursch, Jakob Ohme & Ulrike Klinger, "Enabling Research with Publicly Accessible Platform Data: Early DSA Compliance Issues and Suggestions for Improvement", Weizenbaum Institute, April 2024.

submitting applications to the Meta Content Library confirms this 1.5 month timeframe. Additionally, after applying through the X DSA Researcher Application Form on 17 April 2024, we had received neither access nor a conclusive response as of November 2024,[66] leading us to initiate legal action against the company.

Even more problematic, when access was granted, tools were often restrictive, incomplete, and/or dysfunctional, offering only partial access to data that was in principle publicly available. For instance, TikTok's Virtual Compute Environment (VCE) is designed as a two-stage "clean-room" system that severely limits the scope and usability of data; researchers can only test small daily samples drawn from high-follower accounts, cannot download raw data, and ultimately receive only aggregated outputs after platform review. This approach to data access inherently mandates inflexible workflows, which do not correspond to the spirit or letter of the DSA. Beyond the VCE, access through the TikTok API is also restricted, being limited to academic researchers.[67]

Similarly, the Meta Content Library (MCL) imposes important restrictions. Although Meta's platforms allow access to data such as posts on pages, public groups, or public accounts, they do not provide access to certain other types of data, despite their public nature. For instance, comments on public Facebook or Instagram posts cannot be downloaded, creating blind spots for researchers when analysing the spread of hate speech and disinformation in comments.[68] Moreover, even when posts are exported, the associated URLs are not included, which complicates the process of verification and reporting. Lastly, platforms APIs impose daily rate limits or quotas in the data that can be accessed, imposing significant barriers to systematic research, which often requires large datasets.[69]

In short, the implementation of the data access obligations falls short of the legal obligation, and recent (geo-)political trends suggest that

66 DRI, "Case Against X: Berlin Court Confirms Researchers Can Enforce Their Right to Data Access in National Courts", 13 May 2025.

67 Unpacking TikTok's Data Access Illusion | TechPolicy.Press

68 Alvarado Rincón et al, "Unpacking TikTok's Data Access Illusion", *op. cit.*, note 65.

69 DRI, "Access Granted: Why the European Commission Should Issue Guidance on Access to Publicly Available Data Now", 9 September 2024.

only enforcement action by regulators or court cases will improve this situation. Underlying these obstacles is a more fundamental issue – the absence of a precise and consistent definition of "publicly available data" within the DSA. This lack of definitional clarity generates interpretive uncertainty and allows platforms to exercise wide discretion in delimiting what data can be accessed, frequently in ways that narrow the scope envisaged by Article 40(12).

Such indeterminacy not only explains the heterogeneity of practices across platforms, but also underscores that the regulatory framework operates on an unstable conceptual foundation; the very category of "publicly available data" remains contested.

Challenge 2: Platform responses to reporting

The effectiveness of reporting mechanisms has varied significantly, contingent on platforms' willingness to act on reports. The fact-checking organisation Maldita.es evaluated how five platforms (Facebook, Instagram, TikTok, X, and YouTube) responded to their reports on electoral disinformation in the run-up to the 2024 European Parliament elections.[70] The study found that, overall, 45 per cent of the identified disinformation content received no visible action from the platforms. When actions were taken, responses varied considerably. Meta's services stood out, with Facebook acting on 88 per cent of the flagged posts and Instagram on 70 per cent. By contrast, the other platforms showed significantly lower response rates; TikTok acted on 40 per cent, X on 29 per cent, and YouTube on just 24 per cent of the identified content. Actions included fact-checking, applying debunk labels, and content removal. This unevenness underscores the extent to which the effectiveness of actions is contingent on platform-specific practices, rather than a uniform application of DSA obligations.

Our experience with TikTok's response rate was better. It acted on 78.9 per cent of our reports on murky accounts, while disputing in the other cases that a violation of their guidelines had occurred (our

70 Maldita.es, "Platform Response to Disinformation during the EU Election 2024".

criticism is mainly that the platform does not address this problem pro-actively, and only acts on reports. See above).

A second case study by Maldita.es highlights persistent shortcomings in the speed of platform responses. The organisation uncovered a network of fraudulent Facebook pages impersonating public transport services in 47 Spanish cities, designed to steal users' personal and credit card data. Maldita reported 58 unlawful posts. One week later, over 93 per cent of this fraudulent content remained accessible on Facebook.[71]

Another example, reported by Science Feedback, suggests either that X's current moderation mechanisms are inadequately equipped or that the platform is reluctant to enforce sanction-related policies at scale. Broader research from the community suggests the latter. The study involved submitting 125 clear EU sanction-violating posts to X using the "Report EU Illegal Content" form. Only one of the reported posts was removed. For the remaining cases, X responded via email stating that no violation of EU law had been found, despite clear evidence to the contrary.[72]

These cases demonstrated the gap between the DSA's requirement that platforms provide effective reporting channels and act on relevant content without undue delay and the reality of inconsistent or delayed enforcement.

Outlook

The implementation of the DSA in electoral contexts has revealed opportunities for civil society as well. Looking ahead, several developments offer concrete pathways for CSOs to enhance enforcement, transparency, and democratic oversight of digital platforms.

71 Maldita.es, "We flagged 58 fraudulent Facebook posts to Meta using DSA mechanisms. A week later, 93% are still active", 19 June 2025.

72 Science Feedback, "Flagged and Ignored: Testing X's Response to EU Sanction Violations", 23 July 2025.

Strategic Litigation

Strategic litigation serves as a key opportunity to strengthen the enforcement of the DSA. Ahead of the German elections, we requested real-time access to public data from X under Article 40(12) of the DSA to monitor systemic risks. When the platform refused, our team initiated legal action in Berlin.

Although the request was ultimately denied, our lawsuit against X, based on our data access request, set significant precedents by recognising researchers' rights to data access, affirming the direct effect of the provision, and clarifying that cases can be brought before national courts.

Out-of-Court Dispute Settlement (ODS) Mechanism

Another avenue for strengthening platform accountability is the out-of-court dispute settlement (ODS) mechanism established under the DSA. It allows users and CSOs to challenge platform decisions through certified ODS bodies, providing a faster and more affordable alternative to judicial proceedings. While primarily aimed at user protection, the mechanism also holds broader potential for systemic oversight; recurring disputes could raise compliance costs and incentivise consistent enforcement of the DSA's provisions.

Broader Governance Mechanisms

Broader DSA governance mechanisms have created new avenues for regulatory action, allowing civil society and researchers to engage not only with enforcement teams at the European Commission, but also with national Digital Services Coordinators (DSCs). For instance, in the lead-up to its adoption in 2025, the draft Delegated Act provided an opportunity for DRI, Maldita.es, Das Nettz, and other CSOs to submit feedback and recommendations.[73] Following its entry into force, CSOs were also able to engage in discussions with the DSCs on key aspects of the Act, such as how the application procedure under the

73 Daniela Avarado Rincon & Michael Meyer-Resende, "DRI's Feedback to the Delegated Regulation on Data Access, DRI, 25 November 2025.

DSA would work in practice, what preparations would be required, and which obstacles might arise.

The DSA also reinforces multi-stakeholder engagement, connecting CSOs, policymakers, and platform representatives to coordinate responses to emerging threats. Initiatives such as the Rapid Response System under the EU Code of Conduct on Disinformation, or collaborative spaces like Data Access Days,[74] illustrated how such engagement can strengthen monitoring, foster dialogue, and facilitate adaptive governance in line with the regulation's objectives.

Finally, public campaigns, such as policy articles and election-monitoring reports by CSOs, or joint requests, such as Article 40(12) data requests initiated by Mozilla, AlgorithmWatch, and partner organisations,[75] have demonstrated how civil society can amplify accountability by mobilising public awareness and shaping political agendas.

The EU AI Act

As generative AI increasingly shapes electoral information and the broader digital information landscape, the forthcoming implementation of the EU's *AI Act* introduces both new opportunities and responsibilities for civil society. The Act's risk-based framework covers AI systems that may affect democratic processes, particularly those designed to influence election outcomes, referendums, or voting behaviour. Such high-risk systems are subject to the most stringent obligations, including robust risk assessments, transparency measures, and human oversight. The Act also establishes a special category for general-purpose AI (GPAI) models that pose systemic risks, recognising their potential to amplify manipulation or misinformation during elections. Its governance structure, comprising national and EU-level authorities and a new AI Office within the European Commission, links closely with other frameworks, such as the Digital Services Act

74 Weizenbaum Institute, "DSA40 Data Access Days".

75 AlgorithmWatch, DSA40 Data Access Collaboratory, Mozilla Foundation et al, "We are making a simple request to big tech platforms for their top 1,00 most viewed posts every six hours, per EU member state. Join us?".

(DSA) and the Digital Markets Act (DMA). These mechanisms offer additional channels for oversight and coordination, where civil society can play a key role in ensuring that accountability, transparency, and electoral integrity remain at the centre of AI governance.[76]

Despite its ambitious scope, the AI Act still leaves several crucial risks unaddressed. Ambiguities in the list of prohibited AI practices, reliance on providers' self-assessment for high-risk classification, and broad exemptions for open-source models may limit the regulation's effectiveness. Moreover, civil society participation in the Act's implementation may remain insufficient. These gaps carry significant implications for CSOs, which depend on clear safeguards and access to data to monitor, document, and advocate against AI-driven threats to democratic processes. Ensuring meaningful involvement of civil society in the AI Act's governance will, therefore, be essential to protect transparency and electoral integrity.[77]

The EU aims to simplify regulation ("omnibus proposal") which would be positive, as the regulatory framework is too complex and includes overlaps and redundancies. The main risk of this initiative is that simplification becomes a pretext for deregulating serious risks.

What's next for civil society in the digital democracy space?

The relationship between big tech and civil society has shifted dramatically over the past few years. Where there was once a degree of cooperation and alignment on countering disinformation and hate speech, there is now growing antagonism. At the same time, the implementation of the DSA marked a turning point; for the first time, evidence generated by civil society research carried the potential to trigger direct financial consequences for platforms. However, as platforms scale back trust and safety teams and replace once-effective research tools with limited or broken alternatives, civil society finds itself increasingly on the back foot.

76 EUR-Lex, "Regulation - EU - 2024/1689 - EN - EUR-Lex", 2024.

77 Daniela Alvarado Rincón, "AI Act Comes into Force: What It Means for Elections and DRI's Next Steps", DRI, 1 August 2024.

Platform non-compliance has manifested in several ways. X has significantly reduced its content moderation efforts, resulting in increased visibility of disinformation, conspiracy theories, and hate speech. Meta, meanwhile, has limited the ability to identify harmful content by narrowing the scope of its policies and discontinuing certain research tools. The closure of CrowdTangle during a "super-election year" and its replacement with the less comprehensive Meta Content Library has made the platform much less transparent and has diminished opportunities for independent scrutiny.

The 2024 re-election of Donald Trump further emboldened platforms to reduce their work on online content integrity, most of which are headquartered in the United States. The new administration has taken a hostile stance toward European digital regulation, particularly the DSA. In this political environment, Meta has announced a further scaling back of fact-checking operations.

At the same time, the explosion of generative AI into the mainstream has added new challenges. Synthetic media, which can now be generated at scale and at low cost, is increasingly indistinguishable from real content, eroding public trust in what users see online. The adoption of generative AI by political actors for campaign purposes, whether for creating innocuous content or otherwise, further exacerbates this issue. Generative models are also becoming increasingly integrated into everyday technologies, and are often marketed as authoritative and reliable, despite evident problems with hallucinations, bias, and misuse.

These developments place civil society in a precarious position.

As multi-billion-dollar corporations continue to evade responsibility and resist compliance with EU law, the integrity of the online information environment is being further compromised, undermining trust in digital discourse and placing an unsustainable burden on under-resourced watchdogs.

Ultimately, the health of Europe's digital public sphere will depend on whether platforms, policymakers, and civil society can rebuild a framework of mutual accountability. The next phase of enforcement under the DSA and AI Act must move beyond formal compliance to

measurable impact, ensuring that data access and transparency obligations translate into tangible safeguards for democratic debate. Sustained investment in independent research capacity, structured data-access partnerships, and cross-border coordination will be essential to counter both domestic and foreign manipulation. Civil society must also continue to innovate by developing shared infrastructures for monitoring, legal action, and advocacy to hold platforms accountable and to keep democratic values at the centre of digital governance.

Without such commitments, the online environment risks becoming increasingly fragmented, opaque, and susceptible to influence operations, threatening not only the inclusivity of public debate, but also the resilience of democratic institutions themselves.


DEMOCRACY
REPORTING
INTERNATIONAL