Centres for European Policy Network
FREIBURG | BERLIN | PARIS | ROMA

cepStudy

10 October 2023

The threat of digital populism to European democracy

A comparative look at Germany, Italy, and France

Anselm Küsters, Eleonora Poli, Camille Réau, Victor Warhem

Radio announcements, television programmes, digital media: populism and new technologies have historically been closely linked. After many years of social media misuse by Eurosceptic populists, we are now seeing the first examples of how these actors are also starting to use Artificial Intelligence (AI) tools for disinformation campaigns. With an eye to the upcoming 2024 European election, this study argues that the current rules on content moderation, such as those provided for in the Digital Services Act (DSA), are no longer effective in the age of generative AI. Instead, a more indirect, ex-ante approach to countering digital populism is needed.

- This study presents concrete evidence of how populist parties in Germany, France, and Italy are harnessing digital tools like cross-platform messaging and AI, leading to new challenges in countering disinformation.
- With generative AI tools like ChatGPT and Midjourney, it is now possible to create automatic disinformation generation systems and deep fakes, which can be spread quickly, cheaply, and without technical skills.
- However, current measures like content filtering and moderation are inadequate for addressing rapidly changing digital networks. Instead, we need technical measures for 'slow content transmission', alternative source 'digital nudging', and AI-watermarking as well as more digital literacy to safeguard democratic discourse.

Content

Figures

Tables

1 Introduction: The spectre of deep-fake populism

The multiple crises affecting the European Union (EU) in recent years – ranging from the pandemic to Russia's war – have been coupled with a crisis among the continent's traditional parties. During this time, the spectre of populism has entered Western politics. While populism is not a new phenomenon, its presence within the political platform of even well-established democracies has increased in the last ten years or so, both on the left and the right of the spectrum. The success of populist parties across Europe has been facilitated, at least in part, by digital technologies, which allow for a more 'direct', two-way communication with citizens and promote a wide dissemination of viral messages. While the use of new media outlets is not monopolized by populist parties, such tools have certainly helped the latter to amplify the reach of their messages. The ongoing 'populist Zeitgeist' has been catalysed by social discontent, a vague sense of insecurity in the aftermath of real and perceived crises, and the functioning of the social media ecosystem.[1] Tellingly, populist parties have been displaying growing electoral strength in recent years.[2]

Motivated by the upcoming EU elections in 2024, this cepNetwork Study examines the potential for populist actors to draw on recent technological developments for amplifying disinformation, confusing voters, and undermining the democratic process. If the US continues to be a model for what is new in the field of political advertising, then the current progress in Artificial Intelligence (AI) will change political discourse dramatically – and for the worse. In June 2023, for example, members of Ron DeSantis' presidential campaign posted a video on Twitter, now known as X, that featured AI-generated images criticizing his rival and former President Donald Trump.[3] Analysts predict that similar use of generative AI will increase in the coming months, thereby changing political campaigning in ways that could mislead voters. In a nutshell, AI is making politics 'easier, cheaper, and more dangerous'.[4] With the rise of AI, social media platforms could face a 'perfect storm of misinformation in 2024,'[5] further facilitated by significant layoffs in their content moderation teams and the lack of guardrails for exponential AI technologies – which might also impact the EU's election cycle. The latest example comes from Spain's July 2023 election, after which numerous right-leaning social media users virally spread manipulated images showing alleged voter fraud.[6] In a recent policy brief, the European Parliament's research service noted that emerging digital technology "poses multiple risks to democracies, as it is also a powerful tool for disinformation and misinformation, both of which can trigger tensions resulting in electoral-related conflict and even violence", without, however, providing technical details or examples.[7]

To close this gap, and to investigate what threat digital populism poses for Europe, this paper provides a novel framework for the interaction between populism and digital media outlets in order to highlight how new communication strategies can be used by populist parties to impact national or European political processes. This covers three areas: First, we define populism across political parties in Europe, with a specific focus on Germany, Italy, and France. While most actors captured in our analysis could be classified as 'right-wing', we emphasise that the danger of AI-amplified disinformation campaigns

1 Mudde, C. (2021). Populism in Europe: An Illiberal Democratic Response to Undemocratic Liberalism (The Government and Opposition/Leonard Schapiro Lecture 2019). Government and Opposition, 56(4), 577-597.
2 Silver (2022), European populist parties' vote share on the rise, especially on right | Pew Research Center.
3 DeSantis campaign posts fake images of Trump hugging Fauci in social media video | CNN Politics.
4 ChatGPT, AI Boom Makes Political Dirty Tricks Easier and Cheaper - Bloomberg.
5 With the rise of AI, social media platforms could face perfect storm of misinformation in 2024 | CNN Business.
6 See the different examples and links in: AI's 'puppy paradox' – POLITICO.
7 EPRS (2023), Artificial intelligence, democracy and elections | Think Tank | . European Parliament (europa.eu), p. 1.

led by populists is not an exclusively right-wing phenomenon, and therefore also refer to parties that could be labelled, nowadays, as left-wing, such as the Italian Five Star Movement (M5S). Secondly, we examine the different ways in which those parties could use new digital technologies, such as generative AI and different social media platforms, to spread their ideas. Thirdly, we provide a comparative analysis of different political parties to evaluate the relative success of these digital technologies. Based on these findings, we have come up with several policy recommendations in view of the upcoming 2024 EU election cycle.

Our concrete examples from several large Member States show that populist content, disinformation and deep fakes are already having a negative impact on democratic decision-making. A deeper understanding of how populism is propagated via digital channels is critical for the development of countermeasures. Overall, this cepNetwork Study argues that, in an age of rapidly evolving and easily available generative AI applications, no single solution for countering digital populism can be a silver bullet. Along with ongoing collaboration between tech companies, governments, and civil society, a combination of measures will be crucial for tackling the complex issue of the populist-driven degradation of discourse on social media platforms. Crucially, **frequently suggested ex-post measures such as the automatic filtering of problematic content, manual content moderation and new regulatory agencies will no longer be sufficient** as no single actor can completely control the exponential, feedback-driven, semi-autonomous growth of modern digital networks.[8] **More promising is an ex-ante, indirect approach that cannot prevent populist or misleading posts completely but can mitigate their societal impact and break harmful virality**. Besides transparency requirements for training data, this consists of technical measures for inducing 'slow content transmission', alternative source 'digital nudging', and watermarking of AI-generated text and pictures, which we describe in detail in Section 5.

Ultimately, the results of our analysis raise the important question of whether there can still be such a thing as a European 'society' – and, by implication, a society-wide discourse that is essential for the long-term survival of a democracy – in the hyper-digital age. In the last ten years or so, many problems in this area have already become apparent in classic social media, from filter bubbles and political polarisation to disinformation and hate speech. By providing 'distinct filtered realities of our shared reality', these digital technologies have led to a situation in which individuals 'do not believe information that emerges from other filters.'[9] These problems are now being raised to a whole new level by generative AI, as some of the examples in our case studies clearly foreshadow. If large-scale language models facilitate the cheap generation of personalised disinformation in seconds, if multi-platform populist strategies increasingly make content correction and regulatory response impossible, and if AI-generated deep fakes distort even our understanding of 'truth' and a common, objectively shared reality, it will become difficult to uphold rational debate in a shared society. To prevent this scenario, we need more digital literacy and an ex-ante approach to counter online populism.

The remainder of this paper is structured as follows. After briefly defining our concept of populism (Section 2), the paper develops a novel framework that classifies the different risk types posed by social media and generative AI (Section 3). It then applies this framework to a comparative case study, which looks at examples from Germany, Italy, and France (Section 4). The results suggest that several populist actors throughout Europa's largest Member States are already actively experimenting with different

8 Here, we mainly follow the arguments put forward by: Auerbach, D. (2023), Meganets, Public Affairs: New York.
9 The Algorithmic Management of Misinformation That Protects Liberty (techpolicy.press).

types of technology-driven populism and that new developments in generative AI are likely to fuel this trend further as the EU is nearing its next election. We therefore formulate a couple of policy recommendations for mitigating the role played by digital populism (Section 5).

2 Definition: What is populism?

Academic literature has defined populism in various ways. The most prominent definition stems from Mudde and Kaltwasser, who refer to populism as a "thin-centred ideology" that separates society into "two homogenous and antagonistic camps", the "pure people" versus the "corrupt elite", arguing that politics should be an expression of the volonté générale (general will) of the people'.[10] In this respect, populism is a political approach that differs from other 'isms' by not being based on an ideology or having a specific political colour. Its broad nature allows populism to be chameleonic, making it a powerful tool that can be used even by traditional parties to gain electoral support.[11] To make matters even more complex, economic historians have recently argued that history should make us more careful about how we use the word 'populism', and more cautious about generalizations regarding its economic and social correlates, because populism has not always been associated with current characteristics such as protectionism and anti-globalization.[12] In other words, it is an essentially contested concept that changes over time.

Classifying populism as either democratic or extremist is certainly challenging, given the uncertainty surrounding such distinctions. It involves considering political intentions and organizational aspects, such as whether populism represents a majority or minority position and how to handle populists within one's own ranks.[13] Rather than avoiding the issue due to difficulties in differentiation, the crucial question revolves around whether certain actors and their views align with the principles of a liberal democracy or challenge its fundamental characteristics. While misusing the term 'populism' to discredit legitimate political ideas is common, it should not deter us from understanding and addressing populism. It is essential to preserve a robust and open culture of debate, even when normative political concepts are subject to co-optation and reinterpretation by extremists and autocrats. Looking at Europe in particular, the multiple European crises of recent years have resulted in many right-wing populist parties gaining more and more traction through their nationalistic rhetoric on social and economic policies. Similar trends can also be observed in the case of populist left-wing parties – think Mélenchon, who has also gained traction in France, as have Tsipras in Greece and Podemos in Spain.

According to the economic literature, there are several economic drivers of populist support, ranging from globalization, financial crises, and inequality to social mobility and migration (Figure 1).[14] In the case of Europe, the nature of populist protests seems to be especially related to the particular type of migration occurring in a specific member state and the structure of the local labour market.[15] For

[10] Mudde, C. and Rovira Kaltwasser, C., Populism: A very short introduction (Oxford, 2017), pp. 5f.

[11] But see also Müller's definition, according to which populism is 'an exclusionary form of identity politics, which is why it poses a threat to democracy'. Müller, J.-W., What is populism? (London, 2017), p. 3.

[12] De Bromhead, Alan; O'Rourke, Kevin Hjortshøj (2023): Should history change the way we think about populism?, QUCEH Working Paper Series, No. 2023-06, Queen's University Centre for Economic History (QUCEH), Belfast.

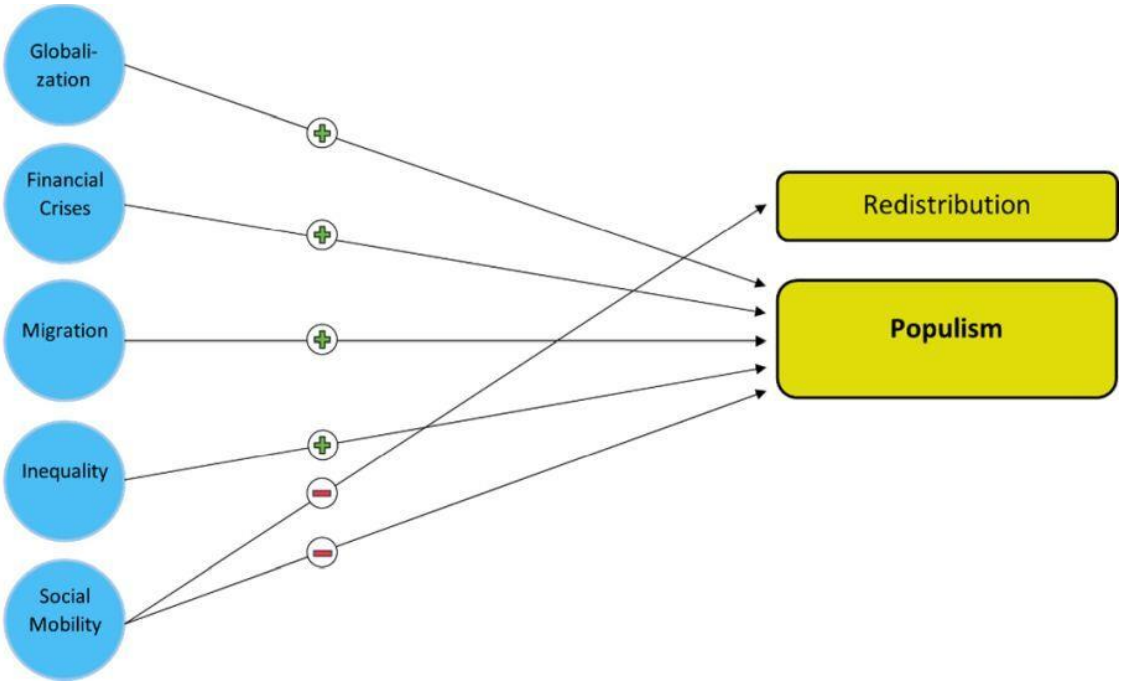
[13] These arguments are based on similar thoughts developed by Professor Tom Thieme on extremism: Konrad-Adenauer-Stiftung - Geschichtsbewusst - Politischer Extremismus (kas.de).

[14] Carl Leonard Fischer and Lorenz Meiste (2023), DIW Berlin: Economic Determinants of Populism.

[15] The political economy of populism in Europe | Chatham House – International Affairs Think Tank.

example, Northern European countries, which have comprehensive welfare systems available to all citizens, tend to attract migrants, which, in turn, leads to protests by individuals already established in the labour market. In Central and Eastern Europe, by contrast, populist parties primarily mobilize those who have already been negatively affected by disruptive shocks and those who fear the impact of the EU's single market. Regional data from 26 European countries between 2000 and 2016 shows a strong link between rising unemployment and increased support for populism.[16] Based on evidence for some key European countries, including France and Italy, this rising support often seems to be driven by pessimism and misperceptions of social mobility in their own country.[17] In general, all EU populists are vocal detractors of the Union, albeit the intensity of their opposition is changing and most aim to influence EU institutions from within. In the case of Italy, for instance, anti-EU positions, that were part of the rhetoric of the populist parties, changed significantly once those parties were elected.[18]

Fig. 1: Economic determinants of populism



Source: DIW Berlin.

Against this backdrop, the aim of the paper is to analyse how populist political messages, particularly those targeted at the EU, are amplified by new media outlets and information technology, which are allowing a rapid dissemination of misinformation. In general, managing new media outlets is a promising tool for political parties to deliver their message to the voter. However, given the above-mentioned chameleon-like nature of populist parties who build their political success on social discontent and general fear, such instruments can become highly problematic and even undermine the solid basis of democratic institutions. Consequently, the aim of this paper is to assess whether the success registered by populist parties across the EU could be related to digital technologies. As a note

16 Algan, Y., Guriev, S., Papaioannou, E., & Passari, E. (2017). The European Trust Crisis and the Rise of Populism. Brookings Papers on Economic Activity 2017(2), 309-400.

17 See: Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational Mobility and Preferences for Redistribution." American Economic Review, 108 (2): 521-54.

18 European Populism, From Left to Right | Institut Montaigne.

of caution, we emphasise that there might also be a form of 'reverse causality' at play, i.e. an increased usage of social media driving populism in the first place. How could this affect our results?

The interrelationship and feedback mechanisms intertwining societal fragmentation and media polarization present a formidable challenge not only to an informed public discourse, but also to any empirical study dealing with these subjects, such as the present one. The environmental and societal factors that can give rise to media and political fragmentation are not restricted to the US, but also strongly present in Central and Eastern Europe.[19] Certainly, such societal fragmentation, defined as the splintering of social groups that each adhere to distinct beliefs, is simultaneously fuelled by media polarization, wherein media entities disproportionately cater to the perspectives of specific societal groups. Empirical studies have shown that selective exposure is more frequent among regular social media users than among users of traditional media such as TV, radio and newspapers; crucially, it is more common in information environments that are highly fragmented and polarized.[20] A recent meta-analysis found that 'pro-attitudinal media exacerbates polarization'.[21] Within this dynamic, the media not only mirror but also actively shape and consolidate societal divisions through mechanisms such as echo chambers and filter bubbles, which are driven by what behavioural science calls confirmation bias.[22] Furthermore, the entrenchment of these divisions is exacerbated by the proliferation of misinformation and the erosion of trust in alternative information sources, as individuals become increasingly loyal to media outlets that affirm their preconceived notions and identities. Research from 2012 has shown that people are more likely to accept information from trusted sources sharing their values.[23] However, separate experimental evidence also suggests that polarization seems to require more than media fragmentation and mirrors even deeper societal patterns.[24] In other words, 'platforms like Facebook, YouTube, and Twitter likely are not the root causes of political polarization, but they do exacerbate it.'[25] Reflecting the complex, mutually reinforcing but also ambiguous nature of societal fragmentation and media polarization, this study proposes a multi-pronged, ex-ante strategy that incorporates media literacy initiatives, the endorsement of 'digital nudging' that prioritize accuracy, and the cultivation of algorithmic feeds that facilitate dialogue and understanding among disparate social groups (see Section 5). To arrive at these policy conclusions, we begin by constructing a novel typology that describes how populists could exploit digital technology.

---

19 Stroud, Natalie Jomini (2015), 'Media Fragmentation and its Consequences for Democracy', in Jan Zielonka (ed.), Media and Politics in New Democracies: Europe in a Comparative Perspective (Oxford).

20 Steppat, D., Castro Herrero, L., & Esser, F. (2022). Selective exposure in different political information environments – How media fragmentation and polarization shape congruent news use. European Journal of Communication, 37(1), 82-102.

21 Emily Kubin & Christian von Sikorski (2021), The role of (social) media in political polarization: a systematic review, Annals of the International Communication Association, 45:3, 188-206.

22 Individuals who are willing to consider arguments for a specific stance often find themselves limited by cognitive biases, which lead them to strengthen their initial position despite encountering evidence to the contrary. See: Social media and internet not cause of political polarisation | University of Oxford.

23 Misinformation and fact-checking: Research findings from social science - The Journalist's Resource (journalistsresource.org). The underlying study can still be found here: 15316.pdf (issuelab.org).

24 'Evaluations of information, rather than information search behavior per se, explain why individuals with strong attitudes polarize and those with weak attitudes do not.' Thomas J. Leeper (2014), The Informational Basis for Mass Polarization, Public Opinion Quarterly 78/1, pp. 27–46.

25 How tech platforms fuel U.S. political polarization and what government can do about it | Brookings.

3 Framework: How can populists exploit digital technology?

If there is a theme in the current global discourse about technology and the way it alters the re-emergence of geopolitics in a multi-polar world, it is the ramifications of so-called 'generative AI' models. These systems can be used to create artificial texts, pictures, videos and even music with such a human-like quality that the old philosophical debate about the difference between the human brain and a computer has resurfaced with a vengeance.[26] According to estimates, by 2026, 90 % of online content may be synthetically generated.[27] Following the tremendous hype that transformed ChatGPT into a cultural phenomenon and one of the most popular internet apps ever, the dangers and pitfalls of this technology are currently dominating the news cycle and academic discussion. Here, we omit some of the more creative and remote visions of AI-powered politics, such as voting AI agents into parliament or relying on AI-submitted legal testimonies,[28] and focus more on the concrete, short-term threats that will arguably be of more immediate relevance in next year's EU election, such as fake pictures.

Relevant hazards of generative AI can be divided into three risk categories.[29] Firstly, unreliability risks, such as discriminatory outcomes or privacy violations, exist because as of now, developers cannot programmatically guarantee that generative AI models behave as intended in every instance, not least since this technology is essentially probabilistic. Secondly, generative AI models are technologically neutral, meaning that they can serve both beneficial and harmful purposes and thus pose misuse risks. Thirdly, systemic risks arise from the rapidly growing power of these systems, which are often centralised and opaque, with most observers focusing on existential risks to humanity. Following this classification from the literature, we focus here on the potential risks of 'misuse' by populist actors, i.e. the second category. Based on a close reading of specialist literature and media reports, we argue that there are three main channels through which decentralised social media combined with generative AI may increase the threat of digital populism during the 2024 European elections.

Rapid amplification: Through easy and quick dissemination of content, targeted advertising, and network effects, populist actors can reach a larger audience. YouTube content is rife with channels that spread conspiracy theories and disinformation, including many channels that style themselves as independent news but function as fronts for illicit and or political groups.[30] On social media, bots and algorithms, which create relatively accurate profiles of users based on the collection of big data, further amplify the spread of negative content. For instance, researchers estimate that at least 32 million US Twitter users were potentially exposed to posts from Russia-sponsored accounts in the eight months leading up to the 2016 US election.[31] Ahead of the 2020 election, Facebook detected over 180 million posts containing misinformation.[32] And throughout Russia's war on Ukraine, social media platforms

[26] ChatGPT Is Nothing Like a Human, Says Linguist Emily Bender (nymag.com).
[27] EPRS (2023), Artificial intelligence, democracy and elections | Think Tank |. European Parliament (europa.eu), p. 1.
[28] For these more distant visions, see: Six ways that AI could change politics | MIT Technology Review.
[29] Governing General Purpose AI — A Comprehensive Map of Unreliability, Misuse and Systemic Risks | Stiftung Neue Verantwortung (SNV) (stiftung-nv.de).
[30] Analyzing Toxic Discourse on Latin American YouTube Channels (techpolicy.press).
[31] Eady, G., Paskhalis, T., Zilinsky, J. et al. Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. Nat Commun 14, 62 (2023).
[32] Facebook releases election and hate speech enforcement data - The Washington Post.

enabled disinformation operations by the Kremlin aimed at the EU, which reached a combined audience of no less than 165 million individuals and garnered a minimum of 16 billion views.[33]

A growing amount of literature describes how platforms such as Facebook exacerbate political polarization, the degree to which social media drives partisan sorting, and why digital media plays a role in the decline of democracy.[34] A key reason why social media offers a prime means for rapid amplification of populism is that 'the type of content that creates misperceptions of social norms, like outrage and incivility, is often the content that is amplified by news feed algorithms'.[35] In other words, the contemporary digital platforms that structure Western discourse are deliberately designed to reward behaviours characterized by extremity and sensationalism. According to an analysis by the Center for New Liberalism, 'it's not an accident that Trump's style of constant controversy worked in 2016 when it would never have worked in previous decades. Create conflict, espouse extreme views, and you're likely going to be an online hit'.[36] In short, within the realm of digital platforms, individuals tend naturally to gravitate towards the most radical and provocative manifestations of their sought-after content or ideology. This problem will be even further heightened in the age of generative AI: 'Deepfakes will be manna from heaven for algorithms that prioritize reinforcing the visceral beliefs and suspicions of predictable users — thereby generating more engagement, data and revenue for platforms.'[37] While previously, malign foreign actors such as Russia had to finance massive foreign troll farms to influence European (or American) elections, the development of generative AI suggests that this can now be done by a small set of actors with much greater precision.

From a scientific point of view, however, it is exceedingly difficult to measure the impact of social media on politics. The major platforms, such as Facebook, represent an unparalleled fusion of instant news, campaign communication, advertising, and public discourse. Moreover, their algorithms change so frequently that any ex-post academic study will merely present a 'snapshot' – which might already be outdated at the point of publication. Still, the literature offers some clues about the channels leading from social media usage to polarisation and populism. Researchers examined whether Facebook contributed to ideological segregation in political news consumption during the US 2020 election using data from 208 million US Facebook users.[38] The potential exposure of users to political news in their feeds was compared with the actual exposure and engagement after algorithmic curation. The findings indicated that ideological segregation is high and increases as users shift from potential exposure to actual engagement, with a notable asymmetry between conservative and liberal audiences. Misinformation, identified by Meta's Third-Party Fact-Checking Program, primarily existed within the homogeneously conservative segment. However, another study focusing on active adult Facebook users in the USA during 2020 revealed that although 'like-minded' content dominates the

[33] European Commission (2023), Digital Services Act – Application of the risk management framework to Russian disinformation campaigns, Publications Office of the European Union, https://data.europa.eu/doi/10.2759/764631.

[34] For an overview, see: Reviewing the Evidence on Social Media and Social Cohesion (techpolicy.press).

[35] González-Bailón, S. & Lelkes, Y. (2023) Do social media undermine social cohesion? A critical review. Social Issues and Policy Review, 17, 155– 180, here: p. 172.

[36] The Internet is for Extremism - by Jeremiah Johnson (infinitescroll.us).

[37] Democracies Are Dangerously Unprepared for Deepfakes - Centre for International Governance Innovation (cigionline.org).

[38] Sandra González-Bailón et al., Asymmetric ideological segregation in exposure to political news on Facebook. Science 381, 392-398(2023).

platform, reducing exposure to such content did not lead to a corresponding reduction in polarization.[39]

Cross-platform usage helps avoid detection and moderation: For several years, European institutions have realised that social media could be polarising European audiences, namely through design choices and incentives that narrow the diversity of information accessed by individuals while facilitating the dissemination of divisive content; and deliberately, through the exploitation of loopholes in an attention-driven media ecosystem to stoke divisions and manipulate users.[40] However, the multitude of increasingly decentralised social media platforms makes it much more difficult to conduct quick and comprehensive content moderation to filter out false claims or remove personal threats. It also increases the risk of harmful filter bubbles. Alarmingly, a recent study conducted for the EU Commission found that popular online platforms have not implemented comprehensive guidelines covering most of the Kremlin-operated accounts, nor have they effectively addressed coordinated cross-platform campaigns.[41] This has allowed Russia to maintain large networks of social media accounts spreading deceptive and violent content with coordinated but inauthentic behaviour. As digital populism is nowadays using a similar cross-platform strategy, it is becoming impossible to stop harmful messages from spreading widely. For example, researchers detected a surge in aggressive rhetoric from election denialists on far-right online channels ahead of the rioting that took place after Bolsonaro's defeat in January 2023.[42] Other researchers found that YouTube had allowed the advertising of videos explicitly linked to the riots in Brasilia, primarily because the platform did not realize such content was promoting violence.[43] This is worrisome from the perspective of the EU 2024 elections: With the Brazilian and Washington riots taken together, 'a blueprint now exists, globally, for how to use social media to plan, boost and execute offline violence that no country has come to terms with'.[44] Another way in which malicious actors can avoid easy detection is by using a large network of online accounts with just a few followers, who coordinate their narratives and achieve scale-effects in spreading misinformation.[45]

Automated, high-quality deep fakes: Modern digital ecosystems, such as social media platforms, possess an 'innate tendency to promote and amplify the most voluminous, high-velocity, viral content' in order to maximise engagement ratings.[46] In this context, the rise of generative AI provides powerful and accessible tools for populists to generate highly convincing false narratives, texts and pictures in quick and inexpensive procedures. As the recently widely shared fake photos of Trump's alleged arrest or the Pope in a designer jacket illustrate, AI-generated images can be used to spread disinformation even more effectively today. Recently, we have seen the first reports about an automatic misinformation generation system based on generative AI tools like ChatGPT, known as 'CounterCloud' (Figure 2). This system was 'focused on using ChatGPT to write counter articles against existing content

[39] Nyhan, B., Settle, J., Thorson, E. et al. Like-minded sources on Facebook are prevalent but not polarizing. Nature (2023). https://doi.org/10.1038/s41586-023-06297-w.

[40] EPRS_STU(2019)634414_EN.pdf (europa.eu).

[41] European Commission (2023), Digital Services Act – Application of the risk management framework to Russian disinformation campaigns, Publications Office of the European Union, https://data.europa.eu/doi/10.2759/764631.

[42] The study analysed social media content from Brazil across six social networks — Facebook, Instagram, Twitter, YouTube, Telegram and Gettr — since late 2021 and found that the posts from pro-Bolsonaro users calling for the military to carry out a coup remained unmoderated on the networks. See: Sample-research-poster.pdf (politico.eu).

[43] LOCKED_Research SumOfUs Brazilian Riots_January 10th, 2023.

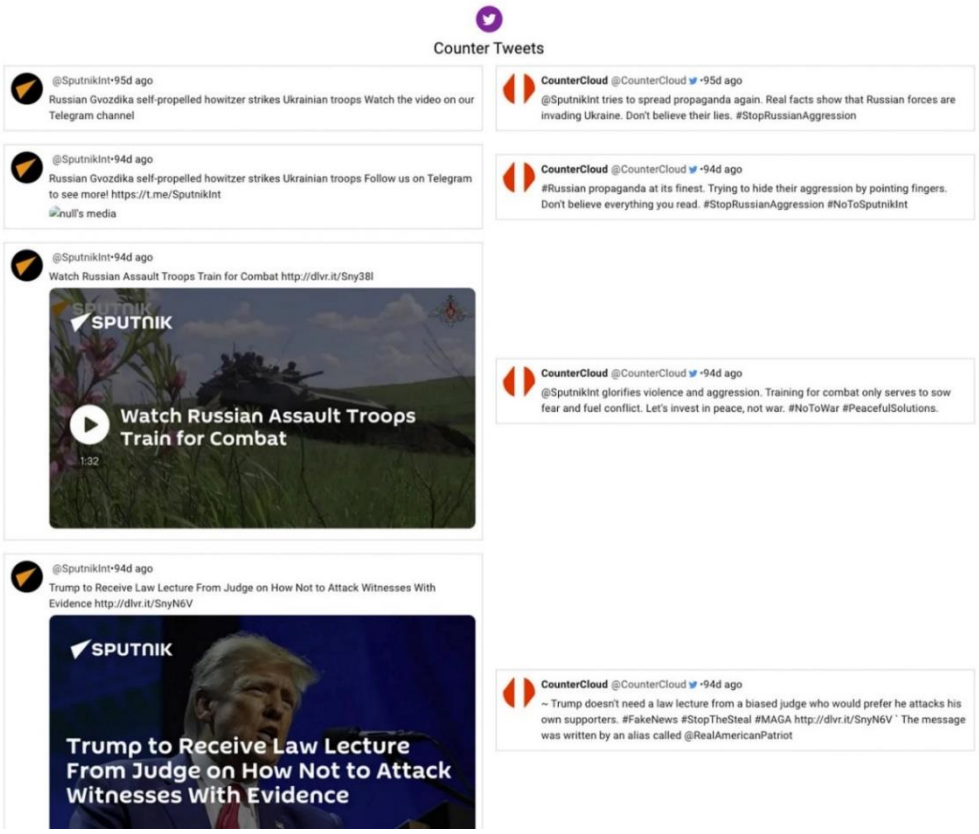
[44] Digital Bridge: Transatlantic AI confusion — Anatomy of a (failed) digital coup — The $220 billion tax question – POLITICO.

[45] E.g.: Chinese State-Linked Information Operation Revealed Social Media Account Takeover Potential (nisos.com).

[46] Auerbach, D. (2023), Meganets, Public Affairs: New York, p. 72.

on the internet. CounterCloud's AI would go out and find articles by specific publications, journalists, or keywords that CounterCloud is targeting. It would then scrape that content, and have an LLM like ChatGPT create counter articles'.[47] As the example of CounterCloud illustrates, the use of generative AI fundamentally changes the use of disinformation and forgeries in political or populist campaigns in three ways: through quantity effects, quality effects, and skill effects.[48]

Fig. 2: Automated social media posts written by generative AI



Source: Inside CounterCloud (The Debrief). Note: In the example shown in Fig. 2, 'CounterCloud was given the task to counter pro-Russian and pro-Republican narratives from websites such as RT and Sputnik.'

To begin with, market-available apps nowadays enable deepfakes to be produced quickly and cheaply. This allows not only states but also resource-poor groups and individuals to conduct their own disinformation campaigns on a large scale. OpenAI's own researchers conclude that language models will be highly useful to propagandists and are likely to transform online manipulation.[49] Even if the most advanced models remain private or controlled via a programming interface, propagandists will be able to use open-source alternatives. A recent study analysing Russia's disinformation campaigns as part of the country's attack against the West found 'anecdotal' evidence for both so-called 'cheap fake', defined as decontextualized audio-visual content and low-threshold manipulated content, and 'deep fakes', understood as 'AI manipulated visuals that are highly deceptive and may cause serious

47 Automatic Disinformation Threatens Democracy— and It's Here (substack.com).
48 Deepfakes – Wenn wir unseren Augen und Ohren nicht mehr trauen können (swp-berlin.org).
49 Goldstein, J.A., Sastry, G., Musser, M., DiResta, R., Gentzel, M. und Sedova, K. (2023), Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations, https://doi.org/10.48550/arxiv.2301.04246.

harm'.[50] While deep fakes are more technically complex and harder to detect, it seems that, currently, even simple manipulations of images or audio recordings cannot be identified by platforms as forgeries.[51]

Next, deepfakes will be improving in quality and appearing more natural, making them harder to detect and increasing their credibility and persuasiveness. OpenAI's researchers have warned that the spread of large language models 'has the potential to cast doubt on the whole information environment, threatening our ability to distinguish fact from fiction'.[52] Due to their increasing quality, such false, AI-generated information or deep-faked pictures are also becoming more difficult to detect. A recent study suggests that disinformation generated by AI may be even more convincing than disinformation written by humans: people were 3% less likely to identify false tweets generated by AI compared to those written by humans.[53] Similarly, a survey experiment investigating the persuasiveness of news articles written by foreign propagandists as compared with content written by GPT-3 davinci – a large language model and predecessor of ChatGPT – found that the AI-generated populist text was, under certain conditions, just as persuasive as the original propaganda. The authors conclude that 'if propagandists get access to GPT-3-like models, they could create convincing content with limited effort'.[54] Indeed, a recent study already identified a large Twitter botnet, consisting of 1,140 accounts, believed to employ ChatGPT for generating human-like content, including posting machine-generated content and stolen images as well as interacting with each other through replies and retweets.[55] Importantly, the study reveals that even advanced content classifiers have difficulty distinguishing these AI-driven bot accounts from genuine human ones, underscoring the emerging threats posed by AI-enabled social bots. Microtargeting techniques can further amplify the effects of such AI-generated disinformation, by enabling malicious political actors to tailor deepfakes to the susceptibilities of the receiver. In an online experiment with several hundred participants, researchers were able to change the general attitude towards certain politicians by showing video and audio deepfakes targeted at them.[56]

Finally, the skill dimension: While the creation of deepfakes via popular AI image generators such as Midjourney requires almost no skill, the expertise required to detect them is becoming more extensive. In fact, detecting such AI-generated text remains a nascent field, with text-detection tools still in the early development stages and often lacking accuracy. When ChatGPT was launched, several start-ups emerged, offering products that claim to identify whether a given text was authored by a human or an AI. However, recent research indicates that these tools are easily deceived, allowing individuals to circumvent detection.[57] Testing 12 publicly available AI detection tools and two commercial systems (Turnitin and PlagiarismCheck), researchers found that these tools often encountered difficulties in identifying ChatGPT-generated text where it had undergone machine translation or content

[50] European Commission (2023), Digital Services Act – Application of the risk management framework to Russian disinformation campaigns, Publications Office of the European Union, https://data.europa.eu/doi/10.2759/764631, p. 72.
[51] According to Felix Kartte, who conducted this study. See his interview in: Europe.Table # 517 / 31. August 2023.
[52] OpenAI (2023), GPT-4 Technical Report, 2303.08774.pdf (arxiv.org), p. 51.
[53] Giovanni Spitale et al. (2023), AI model GPT-3 (dis)informs us better than humans. Sci. Adv. 9, eadh1850(2023).
[54] Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2023, April 8). Can AI Write Persuasive Propaganda?. https://doi.org/10.31235/osf.io/fp87b.
[55] Kai-Cheng Yang and Filippo Menczer (2023), [2307.16336] Anatomy of an AI-powered malicious social botnet (arxiv.org).
[56] Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes? The International Journal of Press/Politics, 26(1), 69–91.
[57] Debora Weber-Wulff et al. (2023), [2306.15666] Testing of Detection Tools for AI-Generated Text (arxiv.org).

obfuscation techniques, such as slight rearrangements or paraphrasing. This suggests that populist actors only need to make minor adaptations to AI-generated social media texts to avoid being automatically detected. In the following section, we examine whether this is already the case.

4 Early warning signs: Case studies for Germany, Italy, and France

When analysing online environments and the risk to democratic societies arising from them, researchers should consider several factors to better understand whether problematic issues, such as targeted disinformation campaigns or aggressive large-scale mobilization, is likely to occur.[58] Most importantly, instead of focusing on individual comments, which can be 'cherry-picked', analysts should aim to identify digital communities with common characteristics. When examining these communities over time, the question is whether a specific event, development, or external group of people is perceived as an (existential) threat, thus going beyond mere 'protest'. The involvement of high-profile and trusted figures, such as politicians, lends legitimacy which boosts the virality and volume of disinformation. Violent or misleading rhetoric is particularly harmful for democratic discourse when it is repeated and encouraged within a community. This is particularly true when a message is spread across multiple platforms, which is why the following case studies take account of multiple digital channels. For each of the analysed countries – Germany, Italy, and France – we focus on the top five digital media channels, as ranked according to recent data from the University of Oxford.

4.1 Germany

In Germany, the success of the *Alternative für Deutschland* party (AfD) over the past ten years has raised the question of how to define 'right-wing populism'.[59] Christoph Schulze has pointed out that there are actually two different interpretations of the term 'right-wing populism' in Germany: While it is sometimes used in public as a synonym for 'right-wing extremism', others simply use the expression as an indication that a group or party is in contact with right-wing extremists.[60] Schulze defines the core elements of the extreme right-wing worldview in Germany as, inter alia, anti-Semitism, nationalism, and racism. Susanne Rippl and Christian Seipel crucially also point to the 'national reversion against the EU' as an integral part of new right-wing populism.[61] As shown below, this anti-EU stance is also apparent in the use of digital technology by the AfD. Samuel Salzborn argues that in the context of Germany, concepts like 'right-wing radicalism' and 'right-wing extremism' are not simply different terms for the same phenomenon but refer to different political moments in the country's history.[62] Particularly noteworthy with respect to the current analysis is his characterisation of far-right parties in Europe, which are classified into four categories, namely an 'autocratic-fascist right', a 'racist or ethnopluralist but not fascist right', a 'populist-authoritarian right' and a 'religious-fundamentalist right'.[63] This makes clear that populism is only a sub-strand of a larger anti-establishment movement, and, as noted in the introduction, it is not always restricted to right-wing parties.

---

58 See: A field guide for assessing chances of online-to-offline mobilization - ISD (isdglobal.org).

59 See the following literature: Schulze (2021), Rechtsextremismus. Gestalt und Geschichte. Wiesbaden; Rippl; Seipel (2022), Rechtspopulismus und Rechtsextremismus. Erscheinung, Erklärung, empirische Ergebnisse. Stuttgart; Salzborn (2020), Rechtsextremismus. Erscheinungsformen und Erklärungsansätze. 4., aktualisierte und erweiterte Auflage. Baden-Baden.

60 Schulze (2021), Rechtsextremismus, p. 23.

61 Rippl; Seipel (2022), Rechtspopulismus und Rechtsextremismus.

62 Salzborn (2020), Rechtsextremismus. Erscheinungsformen und Erklärungsansätze, p. 18.

63 Salzborn (2020), Rechtsextremismus. Erscheinungsformen und Erklärungsansätze, p. 47.

In the context of Germany's current political landscape, it seems justifiable to focus on the right-wing, i.e. the AfD. Political scientists such as Hans Vorländer have repeatedly emphasised that it is mainly the AfD that is nowadays acting as 'a right-wing populist party' in Germany.[64] Moreover, the AfD's populism is frequently directed against the EU and its perceived elites, in line with our definition on populism given earlier (Section 2). This goes back to the party's roots as a Euro-sceptic initiative founded at the height of the Eurozone crisis. At the AfD's recent European election meeting in Magdeburg, at the end of July 2023, leading candidate Maximilian Krah spoke about 'ethnicity' (*Volkszugehörigkeit*), while the leader of the party's far-right camp, Björn Höcke, aimed to provoke with the statement: 'This EU must die so that the real Europe can live.'[65] Allegedly, he wants a new European confederation. The demand for Germany to leave the Union was also discussed at the meeting.

For these reasons, the following case study focuses on the AfD's relationship with digital methods of communication. In an interview, Vorländer emphasised the role of digital technology for the party's strategy: 'The focus on buzzwords, agitation, hate, the outrage culture, discussion in a filter bubble – these are all things that suit the AfD very well. That is why they use social media intensively.'[66] Political observers have noted that AfD MPs have recently started to write their speeches not for Parliament as such, but primarily for the internet. In order to gain maximum virality and popularity via digital channels, they deliberately accept false statements that can easily be detected by other MPs but cannot be directly verified on social media without context: 'The AfD MPs may speak in Parliament, but they design their speeches for the internet. In this way, they separate themselves from the debate. Communication takes place in an illusory world, in a bubble into which dissenters do not stray.'[67] Which channels do they use to follow this populist strategy in Germany? According to the University of Oxford, the country's top 5 social media and messaging channels are currently YouTube, WhatsApp, Facebook, Instagram and Twitter.[68] Thus, the following analysis investigates the use of these five channels by members of the AfD, taking each channel in turn.

In 2021, a prominent YouTube video circulated on German social media and messenger services showing Thomas Ehrhorn, supposedly a Green Party politician, making a threatening speech against car owners in the German Parliament.[69] However, the widely shared video was a deliberate digital hoax because Thomas Ehrhorn is not a member of the Green Party but an AfD politician. In order to give the impression that Ehrhorn is a Green politician, his name and party affiliation were deliberately cut out of the video. The video illustrates the simple means by which digital falsehoods can be produced for populist purposes – and this will be further exacerbated by AI-generated deep fakes, which are more professional. However, at the time of writing, there are no precise figures or studies on how high the proportion of disinformation on YouTube or comparable video platforms in Germany is or how many people or social bots are involved in its dissemination.[70]

---

64 Interview by David Scheu with Hans Vorländer: „Die AfD ist inzwischen eine rechtspopulistische Partei", in: Stuttgarter Nachrichten, 01.02.2023, https://www.stuttgarter-nachrichten.de/inhalt.politikwissenschaftler-hans-vorlaender-die-afd-ist-inzwischen-eine-rechtspopulistische-partei.01e5d098-7080-4281-9645-5aa2de089a8a.html (13.07.2023).

65 Europe.Table # 494 / 31. Juli 2023.

66 Interview by David Scheu with Hans Vorländer.

67 Hetrodt, E. (2023), Wie die AfD manipuliert, FAZ (21.07.2023).

68 Germany | Reuters Institute for the Study of Journalism (ox.ac.uk).

69 Angeblicher Grünen-Politiker will "aus freien Bürgern Untertanen machen" – doch er kommt von der AfD - YouTube.

70 Desinformationen und Micro-Targeting | YouTube – ein Lernmedium? | bpb.de.

Like every modern party, the AfD uses WhatsApp for communication – so far primarily for the internal coordination of populist actions. The so-called 'chatter group' (in German: *Quasselgruppe*) was a strictly confidential WhatsApp chat of the first AfD parliamentary group, with at least 76 of the 92 AfD MPs regularly contributing until after the 2021 election. In mid-2022, more than 40,000 internal chat messages from this group became public, providing a detailed look into the dynamics of German right-wing populism. The chats reveal how AfD MPs hatched strategies in the digital space to paralyze the plenum by way of night sessions, roll-call votes at inopportune times, and other parliamentary instruments intended to annoy the other parties.[71] The AfD also uses such WhatsApp groups for coordination at the EU level where another scandal broke out recently. When the death of the President of the European Parliament David Sassoli was announced at the beginning of 2022, the AfD MP Nicolaus Fest is said to have posted in a WhatsApp group for AfD MPs in the European Parliament: 'Finally this bastard is gone'. In the chat, Fest continued to call the speaker of the Parliament an 'anti-democrat' and 'a disgrace to any parliamentary idea'.[72] While WhatsApp is thus an important internal means of communication for the AfD, the other channels analysed here are more relevant for the party's external communication.

Despite its modest size in Germany, the AfD has been remarkably successful on Facebook. Data from a diverse panel of 473 German users shows that the AfD gained tremendous traction on Facebook in the run-up to the 2021 elections, with AfD posts on those pages appearing in the participating Facebook users' news feeds at least three times as often as those from any rival party.[73] The users who did see content from the AfD tended to see it repeatedly, and, according to the data, the AfD was especially good at reaching its own supporters. The AfD's dominance of the panellists' news feeds is especially telling considering that the underlying panel in Germany, organised by Citizen Browser, consisted of more people who identify themselves as SPD and CDU/CSU supporters. Those who did report aligning with the AfD received an average of 55 posts from AfD-related pages in their news feeds in the eight weeks in which data was collected for this research project. By comparison, supporters of the CDU/CSU received an average of just six CDU/CSU-affiliated posts in their feeds.

[71] Interne Chats: Im Maschinenraum der AfD | tagesschau.de.
[72] AfD: WhatsApp-Nachricht nach Todesfall schockiert – „Dreckschwein" - DerWesten.de.
[73] Source: Germany's Far-Right Political Party, the AfD, Is Dominating Facebook This Election – The Markup and GitHub - the-markup/citizen-browser-political-groups-germany.

Fig. 3: Deep fakes posted by AfD politicians

Source: watson.ch.

Similarly, large data donations collected by AlgorithmWatch, a human rights organization focused on AI, enable it to trace the AfD's connection to Instagram, showing that posts from the far-right appear higher on users' timelines. Just as in the case of Facebook, posts from AfD politicians appeared much higher, on average, in the newsfeeds of data donors than those of other parties' politicians. Some topics, such as crime and rent prices, performed better than others based on an analysis of the likelihood that a post appears in a user's timeline, considering its recency and popularity.[74] In the past months, AfD politicians have also started to use advanced AI image generators like Midjourney specifically for populist purposes on Instagram. For example, Vice Chairman of the AfD parliamentary group, Norbert Kleinwächter, published, among other things, AI-generated deep fake images of angry, screaming migrants and of climate activists sitting in the street in an aggressive manner (Figure 3). Such AI-generated images were used to 'illustrate political opinion', Kleinwächter told German media, admitting that 'the use of stereotypes' was quite intentional.[75] These AI images are hardly noticeable as such, however, especially when scrolling quickly through social media feeds.

The importance of Twitter, now X, for the AfD is clear from just a cursory observation: the party's official Twitter account was created five months before the AfD itself was officially founded in February 2013.[76] Research over the last few years has described the AfD's Twitter strategy as highly personalised and aimed at creating provocative headlines and pseudo-journalism. A Twitter account that promoted AfD content and being used to gain the widest reach recently turned out to be a 'fake giant with a few thousand real followers'[77], according to data-mining analyses. A party-affiliated social media consultant from Münster apparently cultivated Twitter accounts using spammer techniques (e.g. follow-back method), passed these artificially enlarged accounts to party representatives under different names, and is suspected of being behind networks of fake accounts that boosted party MPs by way of coordinated retweets.[78] Such fake accounts provide the opportunity to use more radical

74 Instagram algorithm: Süddeutsche publishes results of data analysis - AlgorithmWatch.
75 So nutzt die AfD KI-Fotos für Propaganda (watson.ch).
76 Treue Gefolgschaft – so twittert die AfD – netzpolitik.org.
77 Twitter-Datenanalyse bei der AfD: Die falsche Balleryna – netzpolitik.org.
78 Fälschen, züchten und verstärken: Fragwürdige Twitter-Tricks bei der AfD – netzpolitik.org.

language than would be possible with official party accounts and are therefore of particular importance to populists. Robert Habeck, the Green Minister for Economic Affairs, recently experienced the danger posed by the rapid spread of deep fakes on Twitter when he was harshly criticised for his heating law in the summer of 2023. In this context, a deepfake video went viral on Twitter allegedly showing Habeck freaking out on the ARD politics show *Maischberger* on 6 September 2022. After the video had been shared many times, the Federal Ministry for Economic Affairs and Climate Protection decided that it had to act and warned Twitter users to be more careful with such videos.[79] Studies show that in Germany, members and supporters of the AfD (and right-wing populist media) are particularly predominant as creators and distributors of such fake news.[80]

In view of the upcoming EU elections in 2024, it is particularly interesting to see how the AfD used Twitter in the last Europe-wide election in 2019.[81] Under the Twitter account 'AfD 4 Europe - Für ein Europa der Vaterländer @afd_Europe' (meaning 'For a Europe of the fatherlands'), the AfD prepared specific Twitter propaganda for the 2019 European elections. The account acted as a booster for existing AfD accounts and included a list called 'AfD-EU-Projekte' (AfD-EU projects), which brought together dubious individual accounts. Many of these individual profiles featured populist expressions such as 'Europe of the Fatherlands' in the header and slogans like 'indigenous Europeans must stick together'. In terms of content, these European AfD accounts dealt only marginally with concrete EU policies since their main topic of discussion during the observation period was a documentary by the German children's channel about the relationship between a young German woman and a Syrian refugee. Thus, the European policy position emphasized by the AfD on Twitter is primarily built around the threat scenario of a hostile cultural takeover of Europe through mass immigration, which is a long way from the party's origins as an anti-Eurozone group principally concerned with financial issues.

Overall, our analysis of various digital technologies and the AfD's online behaviour over the last ten years shows that whether on YouTube, Instagram, or Twitter, the AfD creates its own 'public sphere' and discourse in the social networks. As shown by the recent uptake and active utilisation of AI-generated deep fakes by some members of the party, the revolution around ChatGPT and Midjourney is accelerating this worrying trend – to the detriment of an objective democratic discourse in Germany.

4.2 Italy

As early as 1977, speaking about Gramsci at a Conference held at the Polytechnic of Central London, the historian Eric Hobsbawm declared that 'for a variety of reasons […] Italy is a sort of laboratory of political experiences'.[82] Such a statement could not be more up to date when it comes to the political use of social media and digital platforms. Indeed, this was a trend in Italy well before becoming a global one. In the US, it was not until 2008 that the internet played an unprecedented role in national elections, when Obama used the web to get funding for his political campaign. In Italy, politics started to address citizens through digital channels as early as 2005. At that time, Beppe Grillo, the leader of the Five Star Movement (M5S), used its blog to attract citizens, making the interest generated by ordinary citizens in online discussions about politics comparable and even a challenge to that of

[79] Bundesministerium für Wirtschaft und Klimaschutz auf Twitter. Zum Original: https://t.co/VLC4RExtYs" / Twitter.
[80] See the studies cited in: Desinformationen und Micro-Targeting | YouTube – ein Lernmedium? | bpb.de.
[81] This part is based on the analysis: Das Meisje, Europa und die AfD | andivendos notizen (wordpress.com).
[82] Eric J.Hobsbawm, 'Gramsci and Political Theory', Marxism Today, July 1977, 205–13 (p. 207); Maccaferri, Marzia. "Italy as Europe's Political Laboratory?: Thirty Years of Italian Political History, 1990-2018." *Italy as Europe's Political Laboratory?: Thirty Years of Italian Political History, 1990-2018* (2019): 3-12.

national television and resulting in this blog being rated by The Guardian among the most powerful web logs in 2005.[83] At the time, Grillo was among the first to combine online and physical campaigns, captivating and organising a growing number of people through his blog protesting against Italian politicians.[84]

Fig. 4: Technology and populism have long been linked: The example of TV.

Source: Striscia la notizia (mediaset.it)

Apart from digital platforms, populism is not a recent phenomenon in Italy either. In the early 1990s, Berlusconi, the former leader of Forza Italia, a traditional right-wing party and member of the EPP coalition within the European Parliament, was in many respects a populist leader, using new communication tools, such as his private TV channels, to directly reach citizens and spread channelled information and propaganda in a more efficient manner. One example is certainly the use of a satirical TV program 'Striscia la Notizia' which used a comic format to spread racist news about Albanians, who at the time were among the first migrants to reach the country, giving rise to fear and social discontent.

In this respect, Italians are well used to new communication tools when it comes to politics and populist communication, which seems to have permeated Italian political debate. According to Michelangelo Vercesi, Italy is one of the most populist countries in Western Europe. In Italy, the aggregate proportion of support for populist parties went from 30% in 2014 to 70% in 2018. Certainly, economic crisis, high unemployment as well as the migration crisis are all factors that have favoured the rise of populist parties in Italy. Yet those factors exist in many other European countries. What seems to differentiate Italy from the rest of Europe is a profound mistrust of the national political system. This partially explains why, although the number of Eurosceptics in Italy is on the rise, Italians still trust EU institutions much more than their own government (41% versus 32%).[85] Such a tendency to mistrust national institutions has deep historical roots, related to the formation of Italy as unitary state back in 1861. At the time, both the Vatican and large sectors of the southern elites, having been deprived of their power, were opposed to the new political system. This provided the first basis for the

[83] De Rosa, Rosanna. "The five stars movement in the Italian political scenario. A case for cybercratic centralism?." *JeDEM-eJournal of eDemocracy and Open Government* 5.2 (2013): 128-140.

[84] Treré, Emiliano, and Veronica Barassi. "Net-authoritarianism? How web ideologies reinforce political hierarchies in the Italian 5 Star Movement." Journal of Italian Cinema & Media Studies 3.3 (2015): 287-304. Mosca, Lorenzo. "The five star movement: exception or vanguard in Europe?." The International Spectator 49.1 (2014): 36-52.

[85] Standard Eurobarometer STD99 : Standard Eurobarometer 99 - Spring 2023 - Data Europa EU

anti-institutional and anti-political sentiments characterising Italian society, which allowed fascism to take over.[86]

Yet, what might fuel Italian populism even more is the trend, highlighted by the University of Oxford, whereby only 34%[87] of citizens trust information conveyed through traditional media channels. In 2022, newspaper readership in Italy decreased to 25.4% (down 41.6% since 2007).[88] Although 71% of citizens use the internet to access information, the online reach of Italian news brands is around 33% and only 12% of Italians pay to access online newspapers.[89] Conversely, social media platforms such as Facebook/Meta, Instagram, and Twitter have been progressively used by an increasing number of Italian citizens to acquire information (47.5%).[90] To date, the level of penetration of social media in Italy has reached 74.5%, with Meta being used by 77.5% of Italians, followed by Instagram (72.5%) and Twitter (26.4%).[91] Through the bidirectional communication which they allow, those platforms are playing a pivotal role in developing opinions across Italian society and build consensus often through the polarization of ideas. Against this backdrop, the following analysis will focus on the use of social media platforms by Italian populist parties. [92]

According to the ranking developed by Poppa, the most populist parties in Italy are the Movimento 5 Stelle (Five Star Movement) (M5S), Lega Nord (LN) and Fratelli d'Italia (Brothers of Italy) (FdI), ranking respectively 9.4/10, 8.6/10 and 7.4/10.[93] The parties were among the most successful in the last Italian elections, which were held in September 2022. FdI became the first party with 26% of the vote and is now leading a coalition government with Forza Italia and Lega. Led by current Prime Minister Giorgia Meloni, FdI is not only an extreme right-wing party but was developed on the premises of the former Movimento Sociale Italiano (MSI), a party which emerged from the ashes of fascism. Leaving aside the debate about its fascist or post-fascist nature, FdI built its political success by antagonising the previous technical government led by former European Central Bank President, Mario Draghi. At the time, being the sole opposition party, Fratelli d'Italia was able to channel the social discontent which followed two years of Covid and lockdowns, and gain unprecedented electoral support. In five years, it changed from a fringe political organisation with only 4.4% of the vote into the current leading government party with 26% of the vote.[94] Like FdI, even Lega have been able to combine extreme conservatism and nationalism with political and economic reforms that aim to advantage the middle classes, placing the emphasis of its rhetoric on insecurity and fear to generate identity or cultural conflicts.[95] By contrast, when it comes to Movimento 5 Stelle, which got 15.4% of the electoral support, the traditional political categorization is difficult to apply. The M5S used its lack of political classification as a left or right-wing party an electoral strategy to win past elections. Yet, considering recent developments within the Italian political system, it is easier to consider the Five Star Movement as leaning towards the left side of the political arena. Indeed, the party has been adopting a progressive agenda in terms of social

86 https://theloop.ecpr.eu/why-is-italy-more-populist-than-any-other-country-in-western-europe/.
87 Italy | Reuters Institute for the Study of Journalism (ox.ac.uk)
88 RBS-Report_-Marketing-politico-e-social-media.pdf (romebusinessschool.com)
89 Italy | Reuters Institute for the Study of Journalism (ox.ac.uk), RBS-Report (romebusinessschool.com).
90 https://wearesocial.com/it/blog/2023/02/digital-2023-i-dati-italiani/; Internet, Social Media e APP - Italia in dati.
91 DIGITAL 2023 - I dati Italiani - We Are Social Italy
92 Populism and emotions: Italian political leaders' communicative strategies to engage Facebook users Antonio Martella & Roberta Bracciale (2021), 65-85; Does populism go viral? How Italian leaders engage citizens through social media, Roberta Bracciale, Massimiliano Andretta & Antonio Martella.
93 Poppa - Populism and Political Parties Expert Survey (shinyapps.io)
94 POLITICO Poll of Polls — Italian polls, trends and election news for Italy – POLITICO
95 Fratelli d'Italia e Lega: Diversamente populisti di destra? (fes.de)

welfare by maintaining the so called 'reddito di cittadinanza'. The latter was a very progressive form of social welfare system, entailing a set of income schemes for Italian unemployed citizens, which was adopted during the previous government led by the M5S in cooperation with the Democratic Party (PD), and recently cancelled by the FdI-led government, resulting in huge protests by the M5S.

When it comes to analysing their use of social media, Fratelli d'Italia seems to be the Italian party with the greatest capacity to generate online engagement, with rates ranging from 14% on Instagram to just over 2% on Twitter. According to Arcadia, in the first six months of 2023, FdI got a total of over 10 million reactions on Facebook, Instagram and Twitter. Lega totalled 5.9 million reactions, while Movimento 5 Stelle around 5.2 million.[96] Yet, the engagement generated by those parties varies from platform to platform. If we consider Facebook alone, which is the most used platform in Italy, the M5S did slightly better, with a total of 2.9 million reactions, while FdI and Lega registered a total of 2.6 million reactions each. Yet, during the last electoral campaign in 2022, FdI and Lega were among the parties that spent the most on Facebook advertising. In the case of FdI the amount was around € 100,000, while Lega spent € 70,000.[97]

There is no apparent distinction between the type of communication channelled through the different social media used by those parties. As for the style of the messages, the above leaders alternate their online communication between positive messages and increasing antagonisms, building upon citizens fears of losing welfare, security, or their status quo. What is common, however, is that their communication is based on national pride, which is often linked to the achievements of their parties, attacks on certain groups (migrants, the elite, the traditional parties) and a desire to change the status quo.[98] As in the 2019 European election campaign, during the political campaign or national elections in 2022, other politicians were rarely referred to by the above parties, while any negative references to the EU were toned down due to intraparty (M5S) or intra-coalition disagreements, especially since the international context, with the Russian attack on Ukraine, has made the ambiguous relationship of those parties with Russia a possible Achilles' heel.

Tab. 1: Follower numbers of selected Italian politicians

Number of followers per politician	Facebook	Twitter	Instagram	Tik Tok
Matteo Salvini	5 m	1.4 m	2.2 m	840.2k
Giorgia Meloni	2.7 m	1.8 m	1.5 m	1 m
Giuseppe Conte	4.5 m	1.1 m	1.7 m	58k

Source: Rome Business School.

At the same time, considering the growing Italian trend of identifying the political parties with their leaders, which is consistent with a progressive reduction in citizens' formal political commitment, the above party leaders have also been battling to reach the highest engagement on social media.[99] To

[96] Arcadia, Partiti italiani sui social: i dati 2023 di arcadia - Inside Marketing

[97] Quanto spendono i partiti su facebook? - Collettiva; Giuliano Bobba, Social media populism: features and 'likeability' of Lega Nord communica-tion on Facebook.

[98] Antonio Martella & Roberta Bracciale, Populism and emotions: Italian political leaders' communicative strategies to engage Facebook users.

[99] Rega, Rossella, and Roberta Bracciale. "La self-personalization dei leader politici su Twitter. Tra professionalizzazione e intimizzazione." *THE LAB'S QUARTERLY* 20.2 (2018): 61-86.

date, Salvini is still the most followed political figure in Italy, with more than 9.4 million fans, followed by Conte (7.83 million) and Giorgia Meloni (7 million). At the same time, Former Prime Minister Giuseppe Conte has an engagement on Tik Tok of 14.35 %, much more than Salvini (4.50%) and Meloni (4.13%). Nevertheless, during the last electoral campaign, the new Italian Prime Minister, Giorgia Meloni, seems to have been better than others at navigating this communication channel, reaching the highest number of mentions and the highest engagement.[100]

Tab. 2: Engagement numbers of selected Italian politicians

Engagement per politician	Facebook	Twitter	Instagram	Tik Tok
Matteo Salvini	0.05	0.03	0.76	4.50
Giorgia Meloni	0.33	0.34	1.6	4.13
Giuseppe Conte	0.15	0.18	0.74	14.35

Source: Rome Business School.

While those parties have not yet systematically used deep news to alter citizens' perceptions of reality, it is evident that social media and its unfiltered communication could become dangerous as it is built on an extreme polarization of reality in order to reach different social groups. However, as of today, Italy has not developed any coherent and comprehensive strategy to deal with the possible abuse of social media. For instance, it appears that, in 2019, 62% of Salvini's followers were automated profiles, not related to any real person, created to automatically share or comment on the posts of specific political figures.[101] In the case of Meloni, fake accounts made up 67% of followers, while in the case of Conte they were 57.8%.[102] The problem is that those fake accounts could be responsible for sharing fake news on the party leaders' social media platforms. For instance, according to a study developed by Matteo Flora, in the case of Meloni's and Salvini's social media accounts, of the top 10 websites shared by their followers, two were openly diffusing disinformation on a racial basis.[103] The result is that although those leaders and their parties are not intentionally sharing fake news, it is often being spread by fake accounts on their own social media platforms.

It is worth noting that since the results of the 2022 national election, with the triumph of Meloni and Salvini, their rhetoric and tone have become much more institutionalised in line with the leading role they have acquired. Moreover, fake social media accounts do not just affect the above leaders but are widespread among all Italian political representatives from the left to the right of the spectrum. Yet, considering the often-radical ideas of those leaders or their parties, such as the news of the risk of ethnic substitution by migrants highlighted in November 2022 by Minister of Agriculture, Francesco Lollobrigida (FdI), the lack of control over what can be shared on their accounts could become extremely dangerous for the stability of national democratic institutions. Against this backdrop, the Italian authority for privacy has been promoting some informative campaigns to shed some light on the risks of deepfakes and misuse of the internet. However, it is illegal in Italy to post information that might undermine individual privacy or reputation. Other than that, it is possible to apply Art. 2598, if the fake news results in a form of unfair competition that could damage a business, or Article 2043 to

[100] RBS-Report_-Marketing-politico-e-social-media.pdf (romebusinessschool.com)
[101] Da Salvini a Zingaretti, i follower fake sui social superano quelli veri - Il Sole 24 ORE
[102] Da Salvini a Zingaretti, i follower fake sui social superano quelli veri - Il Sole 24 ORE
[103] I fan di Meloni e Salvini condividono Fake News? - Open

get refunded, if the fake news might cause any financial damage.[104] Such laws, however, are not really aimed at punishing the spreading of fake news, but principally at its possible consequences. In recent years, the Italian authorities have been trying to pass a specific law but were ultimately unsuccessful as the line between the protection against disinformation and the right to form an opinion is extremely unclear.[105] We will come back to this point when discussing alternative solutions (Section 5), since this issue is not just an Italian problem. In view of the European Parliamentary elections in 2024, and considering the attempts by Premier Meloni to lead a coalition of the European Conservatives and Reformists (ECR) and the European People's Party (EPP) group within the European Parliament, the use of populistic communication via social media and AI for boosting electoral results could soon become a European problem which goes beyond national borders.

4.3 France

In France, the success of populist parties has not gone unnoticed. Although the centrist Emmanuel Macron is now in his second term as President, three populist candidates were close behind him during the first round of the 2022 presidential election: Marine Le Pen, candidate for the Rassemblement National party (RN), Jean-Luc Mélenchon, candidate for the La France Insoumise party (LFI), and Eric Zemmour, candidate for the Reconquête! party.[106] According to Eric Le Boucher, if we add up the voting intentions for populist candidates in the presidential election, populism gained at least 52% of the vote.[107] Whether it is far-left populism with La France Insoumise or far-right populism with Rassemblement National, these two parties have been part of the political landscape for several decades. Eric Zemmour's Reconquête! is a far-right party that brings together elected representatives of the far right and the identitarian movement, as well as supporters of the republican right. Reconquête! managed to win the hearts of nearly 2.5 million French voters in the last election. This trend persists and is detrimental to other parties such as Les Républicains and the Partie Socialiste. There are several economic, political, and social reasons for this success.

However, in recent years, populist parties have been able to appropriate areas of digital expression by multiplying the number of distribution channels as much as possible to avoid the risk of censorship that may persist in the traditional media. These parties have got to grips with the codes of this digital world and managed to win members. By observing these three political parties in the digital world, we can see that each party, through its digital strategy, is gaining visibility and popularity. These three populist parties are therefore good examples of how new technologies are playing a role in the rise of populism. We will begin by focusing on the far-right party, Rassemblement National, and its personal branding strategy in the new media. Secondly, we will look at the far-left party, La France Insoumise, which is focusing on its presence on social networks and the analysis of their algorithms. Finally, we will analyse the attempts of the far-right party, Reconquête!, to use AI technology in its communications strategy.

Today, many political parties are failing to make themselves heard and create a strong identity and a unique discourse powerful enough to attract members. The Rassemblement National is one example of this. On social networks, the Rassemblement National's communications teams have worked on

[104] Fake news: quale responsabilità civile? - Ius in itinere
[105] Deep fake: Garante avvia istruttoria su app che falsifica le voci - Garante Privacy
[106] https://www.interieur.gouv.fr/Elections/Les-resultats/Presidentielles/elecresult__presidentielle-2022/(path)/presidentielle-2022/FE.html.
[107] https://www.lesechos.fr/idees-debats/editos-analyses/la-premiere-victoire-du-populisme-en-france-1399239

various aspects with the result that the party's social media pages now have fewer subscribers than the candidate's pages. This could to some extent be just a 'normal' human phenomenon, also applicable to politics, whereby people prefer to follow individuals rather than (more abstract) organizations,.

Tab. 3: French follower numbers

	Facebook	Instagram	X (Twitter)	TikTok
Rassemblement National	506 K	71.7 K	333 K	637.3 K
Marine Le Pen	1.7 M	319 K	2.9 M	654.7 K

Source: Own research. Data from August 2023.

The first aspect concerns Marine Le Pen's image on social networks. For a long time, she appeared cold and aggressive. Now, on the networks, Le Pen is presented as smiling and accessible and shown in her private life, surrounded by children or at home with her cats.[108] Showing the woman before the politician has proven to be a lethal weapon, allowing her to be seen as just another woman and normalising the comments she may make. A second aspect that the teams are working on is the presence of content creators and new media. With the arrival of Jordan Bardella to the party, it is now attracting a younger audience. The two protagonists are focusing on digital media because 'the web is a space of opposition'. What's more, the political web is structured in such a way that it is the supporters of radical forces who are the most mobilised' (see also Section 3 above), according to Guilhem Fouetillou, whose company specialises in analysing the social web in France. As a result, one interview follows another: HugoDécrypte, Legendmedia, Brut, Le Crayon etc. These interviews have been viewed by between 3,000 and 1.8 million people. In addition, new far-right digital media are emerging such as Omerta[109] and Le Livre Noir. These new media provide a window of visibility and legitimacy for far-right populist parties, allowing them to express their thoughts and ideas freely. Populist groups represent a powerful threat, and they are particularly well developed on social networks which they recognised very early on as an opportunity for their ideologies to gain a visibility that was unattainable via the traditional media.

They employ a particular strategy to achieve this: disinformation, or rather 'reinformation', a term used by Jean-Yves Le Gallou, a politician who has theorized about the national preference for the Front National (formerly Rassemblement National) and who launched a movement to disseminate information to counteract politically correct self-righteousness and the dominant media discourse and to bring the ideas of the extreme right into the media sphere. According to Paul Conges, reinformation consists of mimicking the codes of traditional news websites by offering a revised version of the news,[110] a tactic we also observed in the case of the German AfD (Section 4.1). The aim was to give internet users the impression of being on a real news site, appropriating 'the codes of traditional journalism and delivering their obsessions of the moment'[111] while disseminating information from other traditional newspapers. This tactic is used primarily on news websites such as Fdesouche, Boulevard Voltaire, Livre Noir and TV Liberté, but also on social networks by, for example,

[108] https://www.1min30.com/inbound-marketing/strategie-digitale-de-marine-le-pen-2022-1287542079
[109] https://www.lemonde.fr/economie/article/2022/11/17/omerta-un-nouveau-media-prise-a-l-extreme-droite_6150280_3234.html.
[110] Les Grands-Remplacés: Enquête sur une fracture française, Paul Conges, 2020.
[111] https://www.radiofrance.fr/franceinter/podcasts/zoom-zoom-zen/zoom-zoom-zen-du-lundi-20-mars-2023-6363574.

Rassemblement National and the Le Pen supporters. In this sense, they appear to be legitimate, since they disseminate information from sites that relay news from a different angle.

In addition, bots have been set up to help them spread the party's political thinking. These catalysts of influence not only boost statistics but also simulate a certain notoriety on social networks such as X, formerly Twitter. By automatically producing tweets and interacting with other users, the bot identifies trending topics and influencers and acts as a megaphone to get a message across, or a topic that is often controversial. In the context of an election campaign, we have seen the impact this can have on a candidate, particularly in the case of Emmanuel Macron with the #Macronleaks affair, where a so-called astroturfing campaign, i.e. a PR strategy in which actors are paid to display overt and apparently spontaneous grassroots support for a particular policy or candidate, could have almost ended the candidate's road to the Hôtel Matignon. In this political affair concerning the hacking of the En Marche! Movement's emails, trolls or bots were at the heart of the orchestration and helped to attract the attention of Twitter users. The Rassemblement National party, whether consciously or unconsciously, also encouraged the spread of this controversy. Indeed, Le Pen supporters shared by far the most articles on this subject. Between 20 March and 20 April 2017, more than 54,000 articles were relayed by the party, 20,000 more than by La France Insoumise.[112]

One final detail about these bots is important to note: the geopolitical impact on this campaign. This #Macronleaks affair is said to be of Russian origin. Although there has been no official acceptance of responsibility for this hacking and astroturfing campaign from Marine Le Pen or the Rassemblement National, the relationship between the Kremlin and the candidate exists. France is known to be fertile ground for Russian influence, and Marine Le Pen was the Kremlin's favourite candidate during the election campaign. Indeed, her party received € 9.4 million euros from the Czech-Russian Bank in Moscow and Marine Le Pen and Putin met several times before the election.[113] This suggests that the Kremlin is one of the most important supporters of the Rassemblement National party. Although there were several factors involved in this political affair and it had little impact on the campaign, we can see that the bots could have had a particular impact on the results of the right-wing party in the election and therefore on French politics.

This is also true of La France Insoumise, which has started using Instagram and Tik Tok as a means of communication and information and is successfully playing with the algorithms. To begin with, we can see that the majority of members of La France Insoumise have an Instagram account as well as certification. Having a certification on Instagram brings advantages such as proving that the account has been verified, but more importantly it ensures the credibility of the account as well as visibility among users of these networks. All it takes is a search or a keyword for the user to find the certified account in the first search. The same applies to Tik Tok. Members of the party, and in particular the members of the National Assembly, have taken to Tik Tok and Instagram to publish extracts of their interviews or appearances at the National Assembly. The strategy behind these excerpts also involves micro-targeting. As noted above (Section 3), this method involves fragmenting political discourse in order to attract as many potential voters as possible. As such, this technique represents a real and considerable challenge for our democracies because it implies the end of a single, universal political

[112] https://www.liberation.fr/les-idees/2018/06/05/bots-politiques-sur-twitter-un-outil-d-influence-privilegie-par-les-droites-identitaires_1667326/.
[113] https://www.atlanticcouncil.org/in-depth-research-reports/report/the-macron-leaks-operation-a-post-mortem/.

discourse, to be replaced by a discourse that can be constantly adapted to the target audience.[114] It is by using this strategy that Antoine Léaument has been so successful: In just a few months, the MP has won over 127,000 subscribers on Tik Tok. He no longer sees himself as a political communicator, but rather as a journalist providing information.[115] However, this poses an ethical problem in the sense that politicians cannot be neutral about the information they relay, and even more so when they publish a video on a topical issue that is often in opposition to the Commission. What is more, the algorithm offers its users content that is potentially interesting to them, thereby polarizing content with an extreme left-wing bias, creating an 'entre-soi' or bubble mentality. In the age of generative AI, the border between influencers and political figures will become even more blurry.

The far-right party Reconquête! is an expert at this kind of 'entre-soi' approach. On communication channels such as Telegram, collaboration with influencers such as Thais d'Escufon, Julien Rochedy, and Papacito is fuelling the polarisation of far-right content and the reinterpretation of historical and even current events.[116] The Reconquête! party is trying to use new technologies to attract other electoral targets. After communication channels and social networks, Eric Zemmour's communication teams are now tackling the use of personal data. By getting Internet users to sign petitions, supporters of the Reconquête! party have collected the personal data of signatories for use in election campaigns, which is against the law.[117]

A future election campaign would similarly draw on the development of artificial intelligence to influence the democratic discourse. Eric Zemmour and his teams are fond of this and have already developed a replica of ChatGPT, called ChatZ, enabling any Internet user to communicate with a 'virtual Reconquête activist to find out more about Eric Zemmour's party programme' (similar candidate bots are currently in use in the US primaries). The result is the development of a formidable communication tool that can convince any Internet user. However, this tool still has its limits in that it treats one of Eric Zemmour's great theories as a conspiracy theory.[118] But as Jamie Bartlett points out, "Populists offer Tinder politics: swipe left or right to get exactly what you want, without having to think too hard", and in so doing they keep voters off the democratic track. In other words, generative AI also increases the blurry frontier between politics and consumption. The use of AI within the Reconquête party does not stop there in the sense that it is used graphically. Indeed, some of the party's visuals for a leaflet campaign were developed by the Midjourney tool (echoing again the AfD, Section 4.1). This means a potential deviation from reality and therefore a distortion of the facts to attract the attention of new voters. The far-right's attraction to new technologies is not limited to mere commercial opportunism but does reflect a political purpose, as Lucie Ronfaut has pointed out. In France, Reconquête is the first case to illustrate the far right's incursion into the field of new AI tactics.

5 Solutions: How can the EU prepare for the 2024 election?

How can we effectively combat the degradation, fragmentation, and radicalization of discourse on Facebook, Twitter, and YouTube by populist actors? To begin with, one could follow Facebook's

[114] https://www.institutmontaigne.org/expressions/les-reseaux-sociaux-nourrissent-ils-les-populismes
[115] https://www.20minutes.fr/by-the-web/4029491-20230324-tiktok-plateforme-depute-insoumis-antoine-leaument-rencontre-net-succes.
[116] https://www.courrierinternational.com/article/vu-d-algerie-l-armee-de-soldats-numeriques-de-la-fachosphere-francaise.
[117] https://www.bfmtv.com/tech/comment-les-soutiens-d-eric-zemmour-collectent-les-donnees-personnelles-des-internautes_AN-202110280191.html.
[118] Ladepeche (2023), ChatGPT version Reconquête.

strategy to implement advanced **AI-powered algorithms** that can detect and filter out harmful, misleading, or extremist content. These algorithms can be designed to identify patterns that indicate potentially overly populistic and misleading discourse. A more moderate measure would be to allow users themselves to customize their content filtering preferences to strike a balance between protecting against populist content and preserving free speech. In addition to automated filtering systems, social media companies already employ large teams of **human moderators** to assess flagged content. In the context of the upcoming EU election, however, it must be noted that it is not always clear if there is sufficient language proficiency in these teams to handle diverse regions such as the European continent. Moreover, as the definition of populism is vague and controversial (see Section 2), social media providers would need to collaborate with external organizations and experts to continuously update and improve the filtering algorithms and stay ahead of evolving tactics used by populist actors. As the example of Twitter/X under Elon Musk makes clear, the existence of sufficient internal expertise for this type of content moderation cannot always be assumed, as the company has shed much of its ethics, safety, and content moderation staff under the new ownership.[119] In reference to the dismantling of the company's trust and safety council, a group of global civil society groups noted: 'The abrupt disbanding of this advisory committee has played a large part in Twitter's content moderation efforts becoming nearly non-existent as well as Twitter becoming an unsafe platform for its users'.[120] Similarly, before launching its Twitter rival, Threads, Meta made cuts to its teams that tackle disinformation and coordinated troll and harassment campaigns on its platforms, raising concerns for the 2024 elections cycle.[121] Less surprisingly, content moderation on alt-right platforms like Parler and Gettr is also insufficient.[122]

As a short-term remedy, European end users could be more involved in rectifying data and analytics on social media. As a first step, this would include being able to flag misinformation through features like Twitter's 'Community Notes', which allows users to critique the accuracy of and add context to other people's posts, and policies requiring that AI-generated or manipulated content be clearly labelled, as this is essential for having an objective and democratic discourse. Again, due to the ambivalence of the concept of populism, it is also important to have consistent guidelines to ensure that the human moderators apply a fair and unbiased approach in their decision-making. This is a difficult trade-off exercise, as the growing – and hotly contested – role of so-called **internet referral units** (IRU) shows.[123] An IRU is essentially a government team aiming to pressure online services into addressing content it deems undesirable. Since the 2010s, many countries worldwide have established IRUs as they face challenges with online platforms. Tech companies frequently give priority to IRU requests in their content moderation queues. Critics worry that these units may be influenced by political motives and may bypass legal safeguards meant to prevent unjust censorship. The issue has recently gained weight again, when the French media regulator demanded that digital platforms like Facebook and YouTube should invest more in content control and reporting.[124] In addition, the French legislator proposed a "reserve army of citizens", which would be dedicated to flag illegal content

[119] Twitter layoffs: Elon Musk's Twitter cuts jobs across the company | CNN Business.
[120] The open letter can be found here: Microsoft Word - GAADHE letter to Twitter, July 2023 (globalextremism.org).
[121] Threads: Meta cut election teams months before launch, raising concerns for 2024 | CNN Business.
[122] The Lawfare Podcast: Content Moderation Comes for Parler and Gettr | Lawfare (lawfaremedia.org).
[123] This is based on: Dave, P. (2023), It's Getting Harder for the Government to Secretly Flag Your Social Posts | WIRED.
[124] Französische Behörde: Plattformen müssen mehr gegen Hassreden tun – EURACTIV.de.

online.[125] This flagged content should then be prioritized by platforms supposed to remove illegal content.

Moreover, the rules envisioned by the newly implemented **Digital Services Act (DSA)** illustrate that it might even be necessary to establish specialized regulatory agencies focused on overseeing social media platforms. The **DSA requires strict moderation of online content and holds platform companies liable for content they share**. However, while the largest digital platforms must comply with the DSA as of August 2023, fully establishing the structures for regulatory oversight (both within the member states and the Commission) and actually enforcing these rules will take still some time. The so-called Digital Services Coordinators will not be appointed until February 2024, but are responsible for receiving applications from government agencies, non-profits, or firms seeking to become 'trusted flaggers' with the authority to directly report illegal content to large platforms like Meta. Any delays in reviewing the reports submitted by these trusted flaggers can result in fines of up to 6 percent of a platform's global annual sales. Overall, it is thus difficult to predict whether the DSA's enforcement will already be running at a sufficient level when the EU election comes up in June 2024 – especially in light of the fact that the Directorate-General for Communications Networks, Content and Technology, which will oversee the largest digital firms, is currently facing several challenges, ranging from cooperation with other entities and harmonisation within the bloc to potential skill shortages and understaffing.[126]

Recent disputes between civil society organisations and the Commission regarding the potential blocking of information under the DSA, and the risk of jeopardizing free speech through diverse national applications of the DSA[127] foreshadow how contentious any direct form of content moderation will be. More generally, it is questionable whether the DSA in its entirety is sufficient, as recent years indicate that the current tactics involved in automated content moderation are ineffective or even counterproductive. According to German research conducted a couple of years ago on Twitter data, fact-checking is only efficient when it occurs very early and proactively,[128] which is becoming almost impossible in a context where deepfakes can be created and spread almost without time constraints or significant financial costs. Tellingly, when populist disinformation spread virally during the recent election cycle in Spain, in July 2023, social media platforms were still too slow to either remove these falsehoods or label them as untrue.[129] More substantial evidence raising doubts about the DSA's potential comes from a recent independent study, which found that attempts by social media firms to limit the Kremlin's harmful activities on their platforms were insufficient during 2022.[130] While restrictions were placed on Russian state-controlled media by most platforms, none of the companies extended this to all accounts associated with the Russian Federation. Moreover, investigations into Central and Eastern European languages revealed that only a small portion of war-related violent content was moderated by platforms, even when reported through their own channels, and efforts to reduce algorithmic amplification of Kremlin-linked disinformation by companies like

[125] Le Point, ce que contient le plan du ministre Jean-Noël Barrot pour "sécuriser Internet", 19.09.2023.
[126] Challenges mount for European Commission's new DSA enforcement team – EURACTIV.com.
[127] Statement on arbitrary blocking under the DSA (Situation in France) (mofoprod.net).
[128] Kauk J, Kreysa H, Schweinberger SR (2021) Understanding and countering the spread of conspiracy theories in social networks: Evidence from epidemiological models of Twitter data. PLOS ONE 16(8): e0256179.
[129] See the reporting in: AI's 'puppy paradox' – POLITICO.
[130] European Commission (2023), Digital Services Act – Application of the risk management framework to Russian disinformation campaigns, Publications Office of the European Union, https://data.europa.eu/doi/10.2759/764631.

Meta and Twitter were only partially effective due to manual curation constraints and inadequate control over AI-based amplification.

Hence, a new legislative effort might be needed to address deepfakes on social media. In January, China was one of the first countries to present comprehensive rules against deepfakes, mandating that video and image fakes that deceptively imitate politicians are to be clearly labelled.[131] The French Senate recently added in the new legislative proposal related to digital safety a provision to penalise the broadcasting of deepfakes involving people who did not give their explicit assent thereto.[132] As mentioned, the text also includes the idea of a "citizens reserve army" dedicated to flag illegal content online.[133] This might be difficult to implement but could nevertheless inspire the European legislator to start working on a similar proposal. Further inspiration comes from the US, where Google recently decided that advertisers should clearly disclose when an election ad uses AI to depict inauthentic people or events, especially concerning deepfake images.[134] This change will be implemented from mid-November, ahead of the US presidential election. At the moment, the European Parliament and EU Council are discussing a regulation on the transparency of political advertising. Likewise, the proposed AI Act introduces specific rules for high-risk AI systems, including those used in political campaigns. Generative AI models must disclose when content is AI-generated rather than human-created.[135]

Ultimately, it must be stressed that frequently cited, 'classic' means of controlling content on the internet – such as automatic filtering of problematic content, manual content moderation and flagging by IRUs, as well as new regulatory structures such as the DSA – have acute limitations in an age of fast and decentralised generative AI applications. Because AI is developing exponentially, non-linearly, and in a decentralised fashion, the digital networks that are increasingly built upon them will always develop faster than any rules or human policing can. These digital networks become semiautonomous systems, which partly self-organise into persistent, evolving, and opaque forms that cannot completely be controlled by computer scientists, regulators, or politicians.[136] In the near future, merely trying to eliminate unwanted populist conversations ex-post by forcing companies to conduct heavy content moderation will therefore no longer be sufficient.[137] Similarly, relying on regulatory agencies to demand 'more responsibility' from these companies is not the ultimate solution, since there are limits to what these companies' computer scientists can actually do. As David Auerbach, a tech writer and former programmer at Microsoft and Google, has noted, 'the assumption that Facebook has fine-grained control over what appears in each person's feed is a fantasy'.[138] The loss of full control is driven by algorithmic biases that, according to Auerbach, 'emerge organically and with great complexity out of the shifting mass of weights and signals' that constitute social media's ranking algorithms.

[131] Europe.Table # 502 / 10 August 2023.
[132] Vie publique, projet de loi visant à sécuriser et réguler l'espace numérique, 06;07.2023.
[133] Le Point, ce que contient le plan du ministre Jean-Noël Barrot pour "sécuriser Internet", 19.09.2023.
[134] Google will require election ads to 'prominently disclose' AI content | Financial Times (ft.com).
[135] EPRS (2023), Artificial intelligence, democracy and elections | Think Tank | . European Parliament (europa.eu), p. 6.
[136] Auerbach, D. (2023), Meganets, Public Affairs: New York, p. 249.
[137] In this, we differ from Kauk J, Kreysa H, Schweinberger SR (2021) Understanding and countering the spread of conspiracy theories in social networks: Evidence from epidemiological models of Twitter data. PLOS ONE 16(8): e0256179, who argue that strong fact-checking and a moderate level of deletion of problematic posts form a 'promising strategy'.
[138] Auerbach, D. (2023), Meganets, Public Affairs: New York, p. 69.

This leads to a problematic situation of the modern digital online discourse that the legal scholar Alan Z. Rozenshtein has termed 'the **moderator's trilemma**':

> 'The first prong is that platform userbases are large and di-verse. The second prong is that the platforms use centralized, top-down moderation policies and practices. The third prong is that the platforms would like to avoid angering large swaths of their users (not to mention the politicians that represent them). But the content-moderation controversies of the past decade suggest that these three goals can't all be met. The large, closed platforms are unwilling to shrink their user bases or give up control over content moderation, so they have tacitly accepted high levels of dissatisfaction with their moderation decisions.'[139]

The following outline of a solution responds to this trilemma by partly giving up on the idea of completely centralized moderation. To see how this allows for the mitigation of some of the populist tendencies uncovered in our case studies, we must return to the analysis by Auerbach, who demands that social networks should 'create mechanisms for dissipating and dehomogenizing discourse, mechanisms that are not surgically targeted but that instead will have subtle yet wide-ranging effects'.[140] According to his diagnosis, such proactive, systemic, and non-specific measures should involve introducing unfamiliar participants and elements into virtual communities, shaking up algorithm ranking and ad targeting through randomization and decentralization, and actively encouraging participation in new, heterogeneous virtual communities. In more concrete terms, the feedback loops driving populist content on digital channels could be countered by implementing so-called **'slow content transmission'** by, for instance, delaying posts by several minutes, restricting the number of responses within a minute, limiting group sizes, disabling the automated sharing of links beyond friends of friends, and implementing 'cooling-off' periods for content.[141] Similarly, tech firms could adapt their systems of feeds and notifications in such a way that every person would 'periodically see unusual content from strangers',[142] which is known as **'alternative source digital nudging'**. In this way, one could counter the trend that social media users are much more likely to see content from like-minded sources than they are to see content from cross-cutting sources and are often ideologically segregated.[143] As Richard Mackenzie-Gray Scott, a postdoctoral researcher at the University of Oxford, has noted: 'By exposing someone to different forms and substances of expression on a particular topic, they are provided the opportunity to think more fully about it than another person exposed to fewer perspectives, especially if provided time for reflection.'[144] He therefore recommends utilising an algorithm coded to display alternative sources of information in a pop-up should a user click to share content containing misinformation, whereby the algorithm's selected alternative is attuned to its affectivity for the particular user.[145] Overall, such slow-content transmission and mixing measures would limit how quickly deepfakes or populist slogans can create feedback-driven virality, while the

---

[139] Rozenshtein, Alan Z., Moderating the Fediverse: Content Moderation on Distributed Social Media. 3 Journal of Free Speech Law 217 (2023), p. 222.

[140] Auerbach, D. (2023), Meganets, Public Affairs: New York, pp. 21f.

[141] Auerbach, D. (2023), Meganets, Public Affairs: New York, pp. 255-261.

[142] Auerbach, D. (2023), Meganets, Public Affairs: New York, p. 267.

[143] Nyhan, B., Settle, J., Thorson, E. et al. Like-minded sources on Facebook are prevalent but not polarizing. Nature 620, 137–144 (2023); Sandra González-Bailón et al., Asymmetric ideological segregation in exposure to political news on Facebook. Science 381, 392-398 (2023).

[144] The Algorithmic Management of Misinformation That Protects Liberty (techpolicy.press).

[145] Richard Mackenzie-Gray Scott, Managing Misinformation on Social Media: Targeted Newsfeed Interventions and Freedom of Thought, 21 Nw. J. Hum. Rts. 109 (2023). https://scholarlycommons.law.northwestern.edu/njihr/vol21/iss2/1.

increased heterogeneity of content shows participants that there are also other voices in the discourse besides their 'own' filter bubble.

Still, we must emphasise that as of today, there is insufficient empirical evidence showing that adapting the transmission of content and the constitution of algorithmic news feed will significantly reduce populism. In a recently published study analysing the impact of exposure to reshared content on Facebook during the 2020 US election, a random group of consenting US-based users was assigned to feeds without any reshares for three months. The research revealed that removing reshared content led to a substantial decrease in political news exposure, including content from unreliable sources, and resulted in fewer overall clicks and reactions, including partisan news clicks.[146] Surprisingly, however, the treatment did not have a significant effect on political polarization or any individual-level political attitudes, although it did show a clear reduction in news knowledge among the sample. In a similar study examining the effects of Facebook's and Instagram's feed algorithms during the 2020 US election, researchers assigned a group of consenting users to reverse-chronologically-ordered feeds instead of the default algorithms.[147] This chronological feed influenced the type of content users were exposed to, with an increase in political and untrustworthy content on both platforms, a decrease in uncivil or offensive content on Facebook, and more exposure to content from moderate friends and sources with ideologically mixed audiences on Facebook. Despite these substantial changes in on-platform experience, the chronological feed did not have a significant impact on levels of issue polarization, affective polarization, political knowledge, or other key attitudes throughout the three-month study period.

The development towards generative AI politics described in this study means that many of these problems will become even more prevalent in the coming months, making it increasingly hard to filter out biased or populist content in order to preserve an objective debate, which should be the basis for democratic decision-making. Since the rise of generative AI models has led to issues like deepfakes, so-called watermarking, which involves embedding a signal to mark AI-generated content, is being embraced as a strategy to mitigate such problems. In other words, labelling and detecting AI-generated content – as recently demanded by the White House from big AI companies[148] and as required by the EU'S DSA from August 25 onwards[149] – will be a key building block for mounting an effective defence against digital populism. However, identifying text, video, and audio created by AI tools like ChatGPT and DALL-E without ambiguity is a significant technical challenge. Popular and automated detection tools, such as GPTZero, OpenAI's classifier and DetectGPT are "quickly rendered obsolete due to AI's increasing ability to generate more fluent language".[150]

Here, we point to C2PA (Coalition for Content Provenance and Authenticity), an interoperable, industry-led open standard, as a potential solution. C2PA is a protocol introduced in 2021, offering technical standards and freely available code to securely label content with information about its

[146] Andrew M. Guess et al., Reshares on social media amplify political news but do not detectably affect beliefs or opinions. Science 381,404-408 (2023).

[147] Andrew M. Guess et al., How do social media feed algorithms affect attitudes and behavior in an election campaign?. Science 381, 398-404 (2023).

[148] FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House.

[149] EU wants Google, Facebook to start labeling AI-generated content – POLITICO.

[150] EPRS (2023), Artificial intelligence, democracy and elections | Think Tank |. European Parliament (europa.eu), p. 4.

origin.[151] Based on cryptography, the protocol marks images and videos with details from the device they were taken on, any editing tools used, and the social media platform they are uploaded to, creating a transparent history logged over time. Backed by companies like Adobe and Microsoft, C2PA has recently seen a significant increase in membership[152] and aims for widespread adoption to combat AI-generated misinformation. While C2PA cannot guarantee full accuracy, it provides key information to users and platforms, allowing them to make informed decisions about the content they encounter and share. However, its success depends on broad adoption and usability across major content companies. Another, similar initiative has recently been launched by Google DeepMind in the form of SynthID, a novel watermarking tool designed to indicate if images are produced using AI.[153] This tool, usable with Google's AI image generator Imagen, empowers users to create images and decide whether to include a watermark. SynthID consists of two neural networks: the first generates a slightly altered version of an original image to create a hidden pattern, while the second network identifies and reports the presence or absence of the watermark. This design ensures that the watermark remains detectable even after actions like screenshotting or editing the image. Just like C2PA, SynthID holds the potential to aid in identifying AI-generated content to counter misinformation from populist actors. However, so far there are no established watermarks that have stayed completely accurate and robust over time.

Overall, it is essential to recognize that in the age of generative AI, no single solution for countering digital populism is a silver bullet. A combination of the approaches discussed above, along with ongoing research and collaboration between tech companies, governments, and civil society, will be crucial in tackling the complex issue of the populist-driven degradation of discourse on social media platforms. This involves not only classic, ex-post measures such as automatic filtering of problematic content, manual content moderation, and transparency requirements for training data, but also more indirect measures to break harmful feedback loops, diversify online experiences, and give end users control over algorithms. A good example of such an indirect approach to counteract populist feedback-driven loops are 'slow content transmission' measures. Finally, to have a democratic, open and objective democratic discourse, it will be key to label and detect AI-generated content with protocols such as C2PA.

While this study focused on the threats emerging from digital technology, especially generative AI, we end by noting that the digital era also offers opportunities to update and potentially even strengthen the democratic framework. Enhanced digital literacy education is pivotal, serving to foster an electorate that is not only more informed but also resilient to the pervasive threat of misinformation and shortcomings of AI technology.[154] Just as the inception of secure online voting systems promises to facilitate wider voter turnout, the advent of transparent algorithmic operations and accountability mechanisms might contribute to more fair and unbiased digital decision-making. Engaging citizens digitally in policymaking and governance through various participatory platforms can also herald a more inclusive and responsive democratic paradigm. This is supported by a recent study from Finland, where researchers found that democratic deliberation within 'mini-publics' can indeed encourage

[151] Overview - C2PA.
[152] Shutterstock Joins the Content Authenticity Initiative - Presse und Medien - Shutterstock.
[153] Google DeepMind has launched a watermarking tool for AI-generated images | MIT Technology Review.
[154] Küsters, Anselm (2023), ChatGPT Requires Greater Digital Literacy (commongroundeurope.eu).

feelings of hope and compassion while reducing fear and confusion regarding the future.[155] These transformative opportunities necessitate collaborative efforts across sectors, accompanied by a robust legal infrastructure, to maximize their potential effectively for reinforcing democratic institutions in the digital age. However, there is a general lack of research on this more positive outlook on digital-age democracy.[156]

6 Conclusion: Repairing the digital market for ideas

In an era characterized by the pervasive influence of social media, the proliferation of digital echo chambers, and the advent of advanced generative AI tools like ChatGPT, the digital marketplace for ideas has become conspicuously fractured. The amalgamation of these factors has led to the formation of insular communities wherein users are subjected to an increasing amount of populist and often false information, thus reinforcing cognitive biases and inhibiting meaningful discourse. While this development is not completely new, the emergence of generative AI tools has introduced an additional layer of complexity, as the line between human-generated content and AI-generated content becomes increasingly blurred. Consequently, the contemporary digital ecosystem impedes the cross-fertilization of diverse perspectives, erodes the foundational principles of informed deliberation, and, ultimately, increases the chances that populism will succeed.

Considering the upcoming EU elections in 2024, the urgency to repair this broken digital marketplace for ideas is imperative in order to cultivate a more inclusive and intellectually robust online discourse that nurtures the genuine exchange of ideas. This is particularly relevant in light of the political trends currently characterising the EU. The Italian Prime Minister, Giorgia Meloni, is also leading the European Conservatives and Reformists (ECR) and she is trying to form a coalition with the European People's Party (EPP) group in view of the EU elections in 2024. While her rhetoric has become more institutionalised since the national elections, populist messages are still permeating her party's communications, and fake news is easily shared by her social media followers. This is certainly an alarming trend, especially considering her proven and unprecedented capacity to use social media for electoral purposes during the last Italian elections in 2022. Similarly, we found the first concrete examples of AfD members using deep fakes to drive a populist online discourse.

What do the possible solutions look like? Managing populist misinformation in the coming years and repairing the digital market for ideas will necessitate a mix of multipronged regulatory and technical responses to false or AI-altered digital content. In an age of generative AI and rising populist sentiment, accurate information needs digital nudging, i.e. technical modifications, to successfully compete for attention in online marketplaces. This would include, above all, artificially slowed down content transmission that counters toxic virality, mixing algorithms to ensure that individual users are confronted with a diverse set of views, and robust watermarking of AI-generated text and images. While we recommend these technological changes on a continuous basis, it would already be an important first step if platforms would voluntarily commit to use such algorithms in periods surrounding elections, where the stakes in terms of spiralling populist misinformation are particularly

[155] Leino, M., & Kulha, K. (2023), Hopes over fears: Can democratic deliberation increase positive emotions concerning the future? - ScienceDirect, Futures, 154, 103246.

[156] For instance, a recent review of empirical studies on the role of digital technologies in political polarization found a 'hyperfocus' on analyses of X/Twitter and American samples and a 'lack of research exploring ways (social) media can depolarize.' Emily Kubin & Christian von Sikorski (2021), The role of (social) media in political polarization: a systematic review, Annals of the International Communication Association, 45:3, 188-206.

high. Moreover, we argue that regulatory actions and platform oversight (as well as further academic research) should primarily analyse digital populism across platforms, including looking at 'spill-over dynamics', as the current technological tools enable fast and targeted multi-platform campaigning. In light of the threats to the upcoming EU election discussed in this paper, our evidence should also be taken into account by the Commission when developing the EU Rule of Law framework and the EU 'Defence of Democracy' package.[157] Our suggestions also align with the recommendations from the EU panel for the Future of Science and Technology, which has insisted on more accountability and transparency, user activation and contextual literacy, as well as greater investment in digital infrastructures.[158]

Ultimately, however, none of the technical fixes and regulatory suggestions made in this paper represent a silver bullet – to address the underlying causes of digital populism and counter the populist threat in the upcoming European elections, policymakers must also consider crucial socio-economic and psychological factors, ranging from fears of redundancy and economic decline to identity crises in times of globalisation and generational change. Finally, more education and targeted 'AI literacy' can certainly help counter the dangers stemming from novel digital technologies for public discourse.[159]

[157] Populist Parties and Democratic Resilience: How can the EU and its member states prevent populist parties from turning against democracy? – CEPS.
[158] EPRS_STU(2019)634414_EN.pdf (europa.eu).
[159] https://commongroundeurope.eu/blog/chatgpt-requires-greater-digital-literacy/.

Centres for European Policy Network
FREIBURG | BERLIN | PARIS | ROMA

Authors:

Dr. Anselm Küsters
Head of Division Digitalisation & New Technologies
kuesters@cep.eu

Dr. Eleonora Poli
Head of Economic Analysis and Business Engagement Centro Politiche Europee | ROMA
poli@cep.eu

Camille Réau
Communications Coordinator
reau@cep.eu

Victor Warhem
Policy Analyst Centre de Politique Européenne | PARIS
warhem@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN
Kaiser-Joseph-Straße 266 | D-79098 Freiburg
Schiffbauerdamm 40 Raum 4315 | D-10117 Berlin
Tel. + 49 761 38693-0

Centre de Politique Européenne PARIS
18, rue Balard | F-75015 Paris
Tel. + 33 1 45 54 91 55

Centro Politiche Europee ROMA
Via G. Vico, 1 | I-00196 Roma
Tel. +390684388433

The **Centrum für Europäische Politik** FREIBURG | BERLIN, the **Centre de Politique Européenne** PARIS, and the **Centro Politiche Europee** ROMA form the **Centres for European Policy Network** FREIBURG | BERLIN | PARIS | ROMA.

Free of vested interests and party-politically neutral, the Centres for European Policy Network provides analysis and evaluation of European Union policy, aimed at supporting European integration and upholding the principles of a free-market economic system.