European Commission

Joint Research Centre

FRACTURED REALITY

HOW DEMOCRACY CAN WIN THE GLOBAL STRUGGLE OVER THE INFORMATION SPACE

JOINT LEAD AUTHORS:
MARIO SCHARFBILLIG
STEPHAN LEWANDOWSKY

AUTHORS:
SACHA ALTAY
MARSHALL VAN ALSTYNE
ANASTASIA KOZYREVA
RALPH HERTWIG
PHILIPP LORENZ-SPREEN
RENEE DIRESTA
SEBASTIAN VALENZUELA
STEFANIE EGIDY
WALTER QUATTROCIOCCHI
AMY ORBEN

2026

Contact information
Name: Mario SCHARFBILLIG
Email: Mario.Scharfbillig@ec.europa.eu

Joint Research Centre
https://joint-research-centre.ec.europa.eu

JRC144603
EUR 40652

PDF ISBN 978-92-68-38011-6 ISSN 1831-9424 doi:10.2760/9358883 KJ-01-26-115-EN-N

Luxembourg: Publications Office of the European Union, 2026

Generative AI (ChatGPT, Mistral) has been used for improving the readability of texts in various chapters as well as in the formatting of references. Everything has been critically checked by the authors.

Cover and inside page illustrations, the visuals of this report were generated using artificial intelligence.

How to cite this report: Scharfbillig, M., Lewandowsky, S., Altay, S., van Alstyne, M., Kozyreva, A. et al., *Fractured reality – How democracy can win the global struggle over the information space*, Publications Office of the European Union, Luxembourg, 2026, https://data.europa.eu/doi/10.2760/9358883, JRC144603.

CONTENTS

ABSTRACT

The report examines the evolving impact of digital technologies on European democracy. In an era of a global struggle to control the information space, this report offers critical insights into the fracturing of perceived realities, the rise of the ”fantasy-industrial complex,” and the systemic risks posed by the attention economy. It exposes how platforms algorithmically prioritise engagement over accuracy, reinforcing ideological echo chambers and amplifying mis- and disinformation that erode democratic resilience. The report highlights how structural incentives and foreign control of the most important players in the information space undermine information integrity, collective knowledge and civic discourse. In response, the report presents a number of recommendations from fostering alternative public spaces and crowd-sourced knowledge systems to reforming business models, restoring user agency, and advancing EU digital sovereignty through decentralised infrastructure.

ACKNOWLEDGEMENTS

Mario Scharfbillig and Stephan Lewandowsky jointly led the report on equal terms.

We gratefully acknowledge the contributions of Jacopo Nudo to this report. We also acknowledge contributions from colleagues at the Joint Research Centre: Sven Schade, Julien Theron, Paulo Rosa, Laurent Bontoux, Mathew Lowry, Eimear Farrell and Jens Linge.

We greatly appreciate comments provided on a previous version of this report by anonymous internal reviewers from the JRC Editorial Review Board (JERB), Mirela Marcut, Nestor Brown and by our external reviewers: Prof. Sander van der Linden (University of Cambridge); Prof. David Garcia (University of Konstanz); Prof. Alan Jagolinzer (University of Cambridge); Prof. Adam Berinsky (Massachusetts Institute of Technology); Prof. Lisa Fazio (Vanderbilt University); Dr. Nir Grinberg (Ben Gurion University).

The creation of this report was facilitated and partly shaped by funding from European research agencies. The following projects deserve particular mention: European Research Council (ERC) Advanced Grant No 101020961 (PRODEMINFO); European Commission Horizon 2020 grant 101094752 (SoMe4Dem); European Commission Horizon 2020 grant 101004509 (CHANSE); Horizon 2020 grant 883121 (PRODIGI).

EXECUTIVE SUMMARY

THE NEED TO OVERCOME EVER MORE FRACTURED PERCEIVED REALITIES

Today's information space is managed through an "attention economy" based on algorithmic satisfaction of individuals' inferred preferences. The business model of platforms is based on engagement maximisation. Platforms present users with content that is likely to extend their dwell time without regard to the quality of the information, with the purpose of selling more personalised advertisements. **The attention economy thus structurally — albeit inadvertently — privileges content that threatens democracy because human attention favours information that is negative, emotional, and conflictual.**

Platforms are also instrumental in the formation of ideologically like-minded user groups, where individuals mutually reinforce their opinions without questioning their own views. While there has never been a time where everyone has shared the same perceived reality, **technology has enabled a previously unseen proliferation and fragmentation of perceived realities**. Democratic institutions and democracy itself cannot survive without some degree of shared reality.

Thus, the attention economy structurally enables low quality information to spread. Recent years have seen a massive shift towards social media becoming an important news source for many, especially the younger generations. A major concern in this context has been tackling mis- and disinformation. But disinformation has evolved from contesting specific claims to "flooding the zone". Whereas information manipulation conventionally involved "systemic lies" that were meant to be believed, the current situation is a mix of dis-and misinformation, deception, misleading information, and kernels of truth. The term "fantasy-industrial complex" is used to describe this self-organizing and loosely coordinated system. **The goal of information manipulation today is often not to make individuals believe certain false claims but to distract and generate distrust, and to activate anti-democratic norms, authoritarian instincts and behaviours**. Blaming individuals for being misinformed is a misdiagnosis of the underlying cause.

In this context, mis- and disinformation are simultaneously causes, mediators and symptoms of harms in society, and they operate within feedback loops where underlying attitudes and online content reinforce one another. **Other underlying factors—e.g., income and wealth inequality, persistent socioeconomic insecurity, entrenched partisan identities, growing institutional distrust—predispose individuals** and communities to seek out alternative information sources, which can then lead to behavioural changes, including potentially harmful ones like vaccine refusal. These underlying causes need to be addressed by policy measures that go beyond the information space.

Estimates of the prevalence of misinformation vary considerably based on the level of analysis, how misinformation is identified, and the sampling space. **Estimates of misinformation exposure as a percentage of civic/news content consumed online tend to hover around 1%-10%, with prevalence increasing considerably, to 10%-30%, for content involving contentious topics (i.e., climate, health, Russia/Ukraine)**. The prevalence of false beliefs—as opposed to mere exposure to false content—is higher still, especially in certain target populations (e.g., 47% of older adults hold misinformed beliefs about important health issues).

Many of these risks are further amplified by generative AI, which enables the large-scale manipulation of information for political purposes through persuasive text, videos and voice content. **Although there is only limited evidence to date that generative AI has been a game changer in recent elections, AI poses a distinct challenge** because it can create the illusion of knowledge, especially when blended with social media, potentially degrading further a sense of shared reality. Generative AI also offers potential to enhance democracy including for deliberation, and large-scale knowledge translation, requiring careful exploration based on democratic principles.

Democracy needs a healthy information space that includes sufficient accurate and verifiable information, a political plurality of opinions and voices, diverse media ownership, and low levels of harmful content and misinformation if some sense of shared reality is to be preserved.

Recommendations:

- *Create alternative public spaces that do not depend on the attention economy*: Personalisation will persist and will continue to create distinct realities under current business models. New spaces of shared reality—also in the physical world—need to be created.
- *Reinforcement of crowd-sourced knowledge*: Collective knowledge production and sharing, as in Wikipedia and other decentralised wiki models, has proved to be one of the most reliable and unbiased information sources. Similar initiatives should be supported by creating independent environments for deliberating competing opinions, validating key statistics, in a cooperative and inclusive manner.
- *Regulations for more user agency in the online world*: Improving platform design from behavioural science insights, including for example cool-offs, accuracy prompts, encouraging lateral reading, coordination help, empowerment of citizen choice architects, awareness campaigns to help citizens helping themselves. This includes media literacy.
- *Better and more far-reaching fact-checking implementation*: Current fact-checking does not necessarily reach the audiences who were exposed to the misinformation in the first place. Mechanisms could be put in place to target exposure to those who saw misinformation and increase demand.
- *Demonetise disinformation actors:* Mechanisms exist, such as reducing or withholding of ad-revenues for those pushing clearly-identified misinformation. As long as misinformation is profitable, it will persist.

None of these measures is sufficient if alternatives to the current engagement-based business models do not emerge.

THE NEED FOR ALTERNATIVE BUSINESS MODELS THAT PROMOTE CITIZEN AGENCY

The negative consequences of current platform design can be changed. Digital platforms offer online community and the potential for new forms of political direct influence, which could better support free expression, democratic participation and deliberation. But this requires value-based models of social media, in contrast to the predominant current attention economy model that rewards actors—from grassroots movements, influencers, media, to political elite—for pursuing extreme, divisive and emotive positions, crowding out the middle ground and reducing the possibility of shared reality.

Furthermore, individuals' agency online is compromised by current business models. Online behaviour is largely constrained by platforms without knowledge or input of the user, and algorithms can dynamically steer them towards

decisions in the platform's interest. The design is opaque, lacking public scrutiny or democratic control. **Platforms enable free speech by permitting the production and dissemination of most content, but they shape the most prominent content amplifying what is in line with their own interests.**

Recommendations:

- *Encourage business model changes*: Platforms rely on maximising attention and time spent on the platform to sell ads. Alternative business models like subscription-based or non-profit could be incentivised through increased advertising taxation, fees/taxes based on prevalence of harmful/misinformation content in samples similar to financial audits, or the implementation of "speech markets" incentivising content creators to use monetary stakes to signal information quality and contestability.
- *Algorithms that work for democracy rather than against it*: Algorithms can be designed to push or to withhold anti-democratic content, as users become less polarised and their emotions less negative if anti-democratic content is down-ranked.
- *Restoring user autonomy*: Users could be given more choice over recommender algorithms, which could be achieved via in-situ data rights (right of users to be served by other operators using their data), interoperability for recommender algorithms enabling the choice from a menu of algorithms, or user control over entire feeds or specific topics/features. This would increase competition, decrease power of platform owners over content, and enable new content providers and curators.

Even if the attention economy business model can be addressed, an important impediment to the implementation of the proposed solutions arises from the fact that most relevant platforms are controlled by entities outside EU jurisdiction.

THE NEED FOR EU DIGITAL SOVEREIGNTY

The challenges for democracy posed by the current infor-mation space are an integral part of the current geopolitical **struggle over information integrity**.

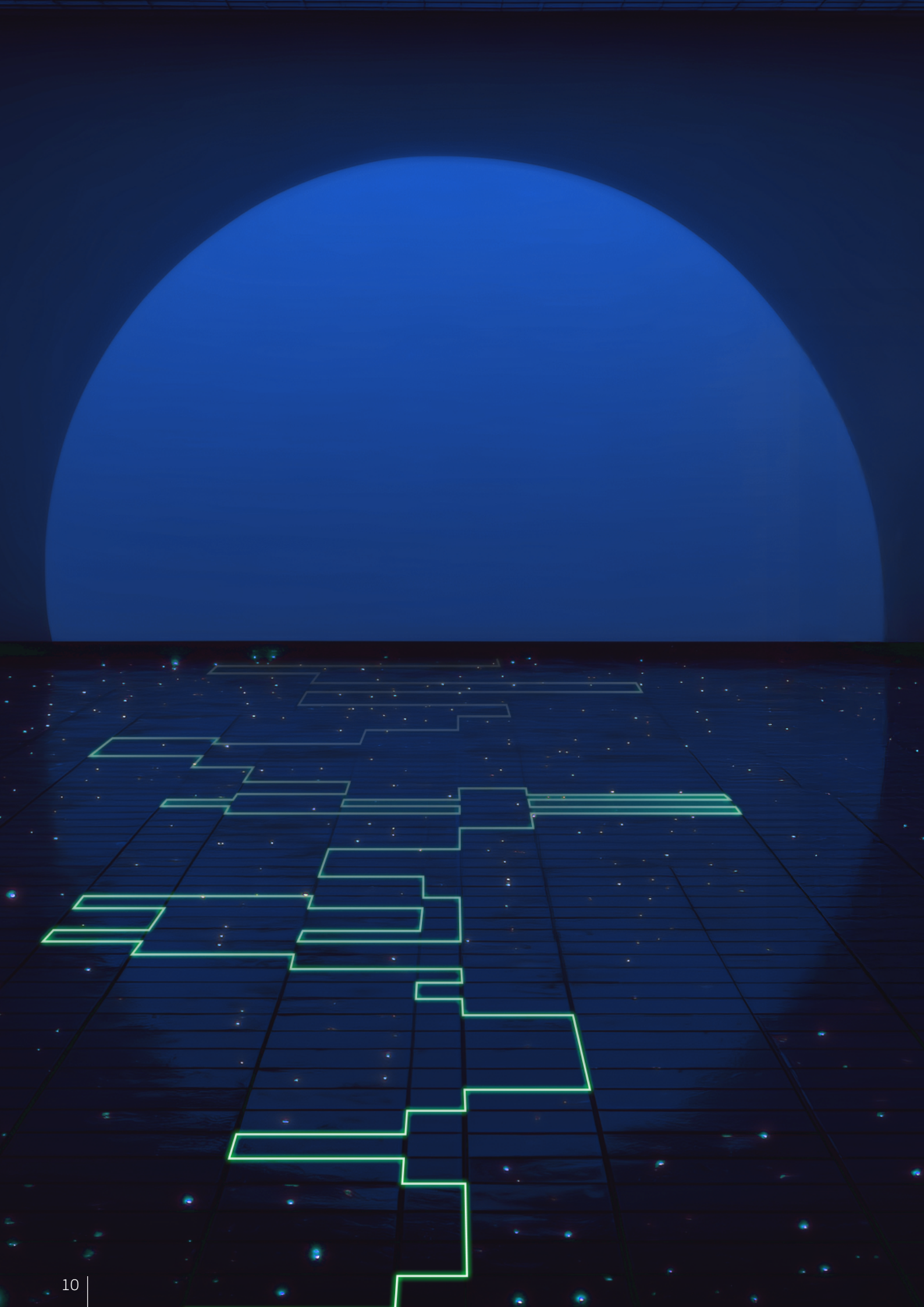
Research suggests that algorithms have been instrumentalised by platform owners in favour of their interests. They can also be gamed by inauthentic (i.e., non-human) behaviour, including orchestrated campaigns by state actors undermining democratic discourse and sovereignty.

Europeans' information diets are thus in the hands of foreign actors who may not share European democratic values. Despite the earlier recommendations, **those values cannot be fully protected without EU digital sovereignty and autonomy.**

Recommendations to achieve digital sovereignty:

- *Achieve digital sovereignty through European digital, physical, and online infrastructure*: Decentralised systems like Atmosphere, FediVerse, Mastodon, Eurosky, which also need EU cloud infrastructure, computing power, etc. to give autonomy back to citizens, take the decisions out of the hands of foreign corporate owners, and can allow exploration of alternative democratic models in real time. Financial support, R&D funding, encouraging more EU-based private funding, and public early adoption (e.g., requiring public administrations to also use alternative platforms, training civil servants) are needed to make them viable competitors of established centralised incumbents, whose R&D budget dwarfs funding currently available in the EU.
- *Decentralised alternatives require support*: A whole new decentralised ecosystem of applications, ranging from public services to commercial ones, including social media and entertainment, could be supported, based on alternative economic models. Information could be treated as the "public commons" that it is, and remunerated according to non-commercial mechanisms (e.g., public subsidies, reputation schemes, local currencies, etc.).

The pace of change challenges the ability of research to keep up and investigate impacts, with platform data kept mostly private. Following the need for faster research and evaluations of digital policies, a European "CERN for Data and Democracy" to bring together fragmented capacities in Europe for continued research on platforms at industrial scale is needed. **A global innovation race without examining its societal implications may outpace the ability of societies to absorb those changes peacefully.**



1

INTRODUCTION

1.1 DEMOCRACY AND THE INFORMATION SPACE

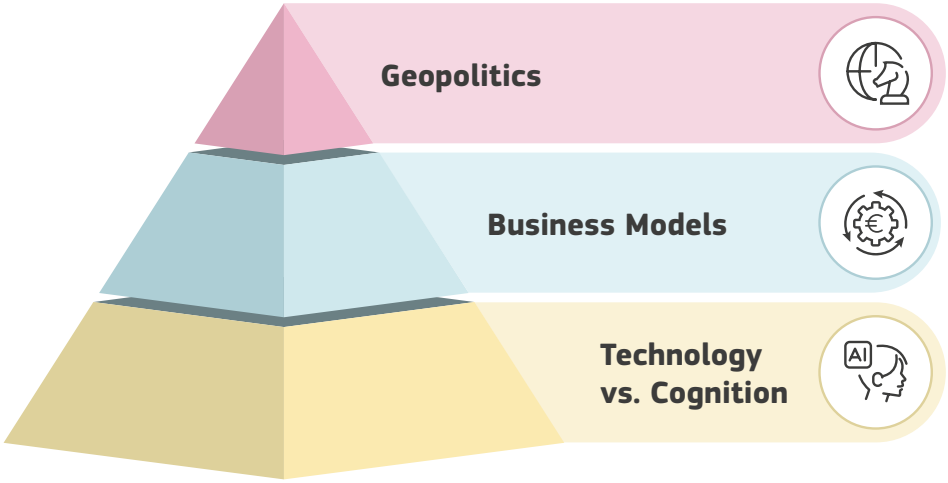
Democracy is essential for the European Union (EU) not only for peace and security, but also for social cohesion as a foundation for long-term economic development. These principles have been affirmed in the recent European Democracy Shield (EDS) (JOIN (2025) 791). Democracy is multi-dimensional, touching on questions from local civic engagement all the way up to the international geopolitical order, but all of these are currently fundamentally being reshaped by changes to the information space.

This report re-examines four interrelated drivers of the effect of technology on the integrity of the information space (Figure 1.1). The report begins by analysing the impact of technology on individual human cognition when interacting with information. It then goes on to show that these impacts are the result of an optimised interaction between human behaviour and technology driven by the platforms' business models. Finally, the report shows how online operators have become power players in shaping democracy. Therefore, each of the drivers at the three levels (technology interacting with human cognition, business models, geopolitics) creates specific risks and harms that challenge the EU's stated goals but that also present opportunities to make democracy in the EU more resilient.

> **"Without high-quality information citizens will have difficulty holding their governments to account."**

Figure 1.1: The challenges to the EU's democratic information spaces online operate on three levels: "Technology vs. Cognition" (Chapter 2), "Business Models" (Chapter 3), and "Geopolitics" (Chapter 4) at the top of the pyramid.

CHALLENGES AND SOLUTIONS TO CREATE MORE DEMOCRATIC INFORMATION SPACES ONLINE



Key challenges	Solutions
• Struggle over information integrity	• Achieve digital sovereignty through European digital, physical and online infrastructure
• Algorithms are instrumentalised by platform owners	• Decentralised alternatives require support
• Lack of EU technological autonomy	

Key challenges	Solutions
• Facilitation of low quality, sensationalist, extremist and toxic content	• Encourage business model change
• Lock-in through digital addiction and network effects	• Algorithms that work for democracy rather than against it
• Severely limited user autonomy on platforms	• Restoring user autonomy

Key challenges	Solutions
• Information overload	• Creating alternative public spaces that do not depend on the attention economy
• Ideological segregation into echo chambers and platforms	• Reinforcement of crowd-sourced knowledge
• Reality distortion into conflictual camps	• Regulations for more user agency in the online world
• Misinformation as cause, mediator and symptom of societal harms	
• Emergence of the fantasy-industrial complex	

Source: Own elaboration

The power of the attention economy

The attention economy is a system in which organisations and creators **purposefully structure content or experiences to acquire, retain, and monetise human attention**—a scarce cognitive resource—by directing it toward externally supplied stimuli.

In this system, attention seekers (such as media producers, platforms, advertisers, entertainers, and publishers) take strategic actions to attract attention, while individuals allocate their finite attentional capacity among both externally initiated demands and self-directed uses.

Activities chosen primarily for internal goals—such as rest, reflection, social connection, or walking in nature—are not part of the attention economy, even though they consume attention, because they do not involve a third party pursuing attention as a commercial objective.

For democracy to function, citizens need to be accurately informed. Without high-quality information provided and channeled through media outlets and social media, citizens will have difficulty holding their governments to account [1]. Unfortunately, many societies have seen a divergence in citizens' beliefs [2, 3] and accompanying high levels of affective polarisation [4, 5]. This fracturing of perceived reality[1] and dislike of other groups will make it harder and harder for civic deliberation, agreement and collective intelligence to materialise. This may eventually lead to a breakdown of democratic discourse altogether [6].

There has been a massive global shift of news consumption patterns away from directly receiving information from legacy media (newspapers, TV) to new forms on information mediation through social media, including video-based services such as TikTok [7]. Although this trend is global, it is slower in Europe than in the US, with less than 20% of European respondents indicating that social media is their main source of news, compared to 34% in the US. Because younger people indicate much higher percentages, this trend can be expected to continue [7].

Social media have frequently been implicated as a main vector of polarisation and information threats such as dis- and misinformation [8, 9, 10, 11], and there is evidence that polarisation online has increased considerably between 2020 and 2024 [12]. A recent systematic review of the effects of social media on democracy concluded that, especially in established democracies, social media use

was associated with increasing polarisation and exposure to misinformation [13]. At the heart of social media, and its impact on society, is the engagement-driven business model which relies on people remaining on the platform to view advertisement. This business model is part of the "attention economy"; see textbox.

Concerns about the adverse effects of social media and information technologies have been further amplified by the advent of generative artificial intelligence (GenAI), which can cheaply automate content generation. Large language models (LLMs) can generate personalised, persuasive text in almost any language at high speed for anyone interested

[1] We use the fracturing of perceived reality and realities synonymous throughout the report. There was certainly never just one perceived reality for people, but the emphasis in this report lies on the manifold increase of conflictual realities through technology.

"This report shows that there is already evidence of identified risks of social media and other technologies that should be acted upon by rapidly designing and testing suitable interventions."

in misusing it [14]. A recent large-scale evaluation of 466,000 AI-generated claims shows an alarming trade-off, where increased persuasion was based on less factually accurate content [15].

1.2 A RAPIDLY CHANGING WORLD REQUIRES RAPID EVIDENCE

The accelerating pace of change especially in the digital space has profound implications for science and its role in evidence-informed policymaking. Science, by design, is careful and cautious. But careful and cautious are insufficient in a world in which "move fast and break things"[2], has become the mantra of one of the most powerful industries globally.

Much of the analysis in this report is based on hard-won insights involving researchers in an adversarial stance vis-a-vis big tech platforms. Data access has been hard to come by and collaborations with platforms are fraught with difficulty and risks [16].[3] There has also been recent concern that collaborations with platforms and funding arrangements have led to undue industry influence downplaying risks in social media research [19].

In times when democracies and the international order are increasingly fragile, and citizens' views and even health are changing sometimes rapidly, awaiting conclusive evidence before changing policies may not be an option. Experts have therefore argued that one should move to a regime of "Minimum Viable Evidence". Under this regime, tests of ways to mitigate a potential harm commence before the harm has been established by a rigorous evidence threshold. In fields of continuously evolving evidence, the legal principle of proportionality requires a continuous monitoring and updating of interventions in line with new evidence (e.g., [20]), highlighting again the importance of effective real-time data access to speed up the cycle, reduce harms, and at the same time avoiding disproportionate interventions against businesses.[4]

Throughout this report, the minimum viable evidence regime is applied. This report shows that there is evidence to already recognise and act upon, by designing and testing suitable interventions, the identified risks of social media and other technologies. In this case evidence includes

[2] https://www.businessinsider.com/mark-zuckerberg-2010-10

[3] The Delegated Act on data access under the DSA (C(2025)4340) should change this in the future, but it remains to be seen if platforms comply in spirit with the regulation and if the requirements are feasible for academic researchers. The "common research support framework" announced in the European Democracy Shield (JOIN(2025) 791) can improve this further, which is essential as research has revealed significant gaps in the translation of the DSA into data access and reality [17]; see also guidance for data access for researchers [18].

[4] The European Court of Justice also acknowledge the societal need to respond when a risk is recognized and allow for the 'precautionary principle' to guide policy-making ([21, 22, 23])..

academic preprints by reputable teams with a proven track record, even if they are not yet peer-reviewed.

This report uses an earlier report on the influence of technology on democracy as a departure point [24]. The previous report identified four pressure points between online platforms and human cognition with adverse impacts on democracy: the attention economy, choice architectures, algorithmic content curation, and misinformation. All four are still at the core of the debate around democratic resilience, and new evidence presented in this report supports the original findings while adding important nuances and new directions. Many potential threats highlighted before have become reality despite various legislative changes in Europe, which are documented in this report.

This report is therefore intended as a re-affirmation of the need to implement and enforce already adopted legislation as well as the need to take additional actions, especially in the area of building EU digital sovereignty, by encouraging alternative business models, and overcoming conflict-driven echo chambers and platforms. Actions are primarily recommended at the EU level, due to the legal priority given to EU laws in areas in which the EU has already legislated. However, several of the recommended actions are also possible to implement and reinforce at Member State Level, and the European Centre for Democratic Resilience announced in the European Democracy Shield can serve as a way to coordinate action in this space.[5]

"This report reaffirms the need to implement and enforce already adopted legislation as well as the need to take additional actions."

Structure of this report and methodology

The report is structured around the three layers of the influence of technology on democracy, each explored in a separate chapter. Each chapter targets a different layer of drivers (Figure 1.1) starting from the bottom of the hierarchy to explain the impacts on citizens first, before moving up the ladder of relevant interventions. Each chapter and comprises three parts:

- *What?* Describes the challenge to democracy and a strategic goal.
- *Why?* Presents the scientific evidence in support of the goal.
- *How?* Proposes regulatory or political means to pursue the goal.

The report is the result of a narrative review by a collection of domain exports from different disciplines, including Legal Studies, Psychology, Cognitive Science, Decision Science, Economics, Data/Computer Science, Media and Communication Science.

The authors were selected by the JRC as a follow-up from the previous report from 2020 on Technology and Democracy with respect to continuation of expertise and the evolution of the online sphere since then. The report was extensively reviewed by experts in the JRC and by external experts as well as by policymakers working in the European Commission to reflect the state-of-the-science and societal challenges as accurately as possible.

[5] https://digital-strategy.ec.europa.eu/en/news/european-democracy-shield-and-eu-strategy-civil-society-pave-way-stronger-and-more-resilient

2

THE NEED TO OVERCOME EVER MORE FRACTURED PERCEIVED REALITIES

KEY TAKEAWAYS

What?

- The health of European democracy is threatened by a fracturing of public opinion into multiple subjective "realities" and high levels of polarisation that defy easy reconciliation.
- Information manipulation has evolved from identifiable and circumscribed falsehoods to a combination of outright falsehoods, misleading statements, and kernels of truth, amplified by algorithms to create fractured realities.
- A principal danger of dis- and misinformation now lies not in creating specific false beliefs but their corrosive effect on the common ground on which democracies depend.

Why?

- The abundance of information available online overloads our cognitive system, which responds that are detrimental to the civil discourse democracy depends on.
- The fracturing into online echo chambers—and, increasingly, "echo platforms"—is characterised by an increase in toxic, hateful and out-group denigrating use of language. Evidence increasingly suggests a causal role of online platforms in this trend.
- The harms from misinformation are multidimensional: false and misleading content can shift individual beliefs, intentions, and sometimes even behaviours, as well as reinforce broader vulnerabilities by increasing polarisation and undermining institutional trust. Causality is complex: misinformation can act as a cause, consequence, and amplifier within feedback loops.
- The reshaping of the information space and misinformation has led to a "fantasy-industrial complex" involving interactions between politicians, corporate actors, platforms, legacy media, influencers, and citizens to create fabricated versions of reality.

How?

- The digital information space is an engineered space, and like all engineered spaces policies can help redesign it to support, rather than undermine, democracy.
- Fractured realities can be overcome by having spaces for citizen contact and deliberation online and offline that are designed to allow civil contact between polarised communities.
- Countermeasures to misinformation are complementary: system-level interventions reshape incentives and reduce supply; individual-level tools build resilience and competences; and fact-checking provides a corrective backstop albeit with limited reach. The goal is not to eliminate falsehoods altogether but to build resilient information environments where truth can emerge and citizens can deliberate on shared facts.

2.1 WHAT?

A functioning democratic society depends on a minimal common ground: that facts exist, that truth is attainable, and that consensus can be built around shared observations. These assumptions make disagreement possible and productive, bringing together a shared reality with people's diverse values, opinions, worldviews, needs and fears [25]. Competing opinions only have meaning when participants also recognize some basic premises as true—otherwise debate collapses into mutual incomprehension [26]. A plurality of concerns still needs to be grounded in a shared reality [27].

Social systems—democratic institutions, public health, justice—depend on shared frameworks of reality. They do not require perfect truth, but they do need a stable, common frame of reference. When that fractures, collective action falters and trust deteriorates. In democratic systems, the fragmentation of shared reality—what some scholars call "epistemic polarisation"—is among the most dangerous consequences of misinformation [28]. One manifestation of this polarisation is the widespread decline in citizens' exposure to differing political views. The 2026 Edelman Trust Barometer found that in several EU Member States (France, Ireland, Sweden, Netherlands, Germany) only between 25% and 30% of the public are exposed to sources with a different political leaning, and decreased by nearly 9% compared to the year before.[6]

> **"If you want to rip the heart out of a democracy, you go after the facts."**
>
> Nobel Peace Prize laureate Maria Ressa

Today's digital information environment governed by the attention economy creates a fundamental shift in how information is consumed because of how technology and human cognition work together. First, the human tendency of confirmation bias and the digitally enhanced selective consumption of information creates opposing certainties. Confirmation bias is the cognitive tendency to search, favour, recall, and interpret information in ways that confirm one's pre-existing beliefs, while discounting or ignoring contradictory evidence [29, 30]. An essential part of this is the searching and paying attention to information aligned with existing views. This tendency has always existed. However, what has changed is that attention is guided via algorithms reacting to our cognitive biases, the tendency is multiplied in extreme ways. Citizens increasingly have the perception of not even having to search for the news to be well informed – something that is called the "News Finds Me" perception [31]. This creates a fundamental challenge for democracy when opposing groups subscribe to incompatible perceived realities.

Second, on top of biased information-processing and curation in the digital information space, misinformation and misleading information contribute further to the gulf between online "tribes". Blurring the line between facts and opinions turns public debate into a contest between polarising narratives. Deliberate lying, especially at scale, obstructs cooperation and corrodes trust in society. While human cognition tolerates uncertainty and occasional error, systematic and repetitive deception fuels cynicism and exploits our basic cooperative instincts that assume communication is generally honest [32, 33].

[6] https://www.edelman.com/sites/g/files/aatuss191/files/2026-01/2026%20Edelman%20Trust%20Barometer%20Global%20Report_Final.pdf

Definition of Mis- and Disinformation

Mis- and Disinformation The European Commission's Communication on the European Democracy Action Plan recognizes the following distinction:

- *Misinformation* is false or misleading content shared without malign intent though the effects can be still harmful.
- *Disinformation* is false or misleading content that is spread with an intention to deceive or secure economic or political gain or to cause public harm.

This report uses the term "misinformation" as an umbrella term for content that is factually inaccurate (either fully or partially) and capable of misleading, misinforming, or influencing public views based on flawed representations of reality. A survey of 150 academic experts broadly agree with that definition (noting that satirical or parodical content is excluded) [34]. The term "disinformation" is used when the deceptive intent is apparent and the term "Foreign Information Manipulation and Interference (FIMI)" applies when disinformation can be identified to come from foreign actors.

Taken together, these changes are best described as a move from one of shared reality with occasional challenges by disinformation to an individually constructed reality, where it is not facts that are being debated, but entire diverging realities based on what the audience wants to hear and what powerful actors want them to hear. These actors and their products have been described as the "fantasy-industrial complex" [35], which is best understood as an emergent, incentive-driven attention market in which political entrepreneurs and creators package identity-affirming narratives; platforms reward engagement; and audiences participate through remixing and amplification. The "industrial" aspect arises from the repeatable, monetizable supply chain made possible by the ecosystem. Crucially, the fantasy-industrial complex is not a centrally-coordinated initiative, it is a product of the current online system and its properties.

Many debates in politics focus solely on the role of dis- and misinformation and the need for "information integrity" [36], defined for example by the OECD as "...the result of an information environment that promotes access to accurate, reliable, evidence-based, and plural information sources and that enable individuals to be exposed to plural and diverse ideas, make informed choices, and better exercise their rights" [37, p. 7]. This report shows that simply fighting misinformation to improve information integrity is largely to treat symptoms, rather than the underlying causes. Instead, this report argues that an exclusive focus on misinformation is missing the over-arching point that citizens are also living in different perceived realitiesWhat is needed is to realign the information environment such that the information citizens receive is high quality, politically diverse, and civil rather than toxic so that democratic debates can be conducted in a fruitful manner. Thus, the strategic goal of policies should be to establish technology that enables democracies, and at the same time increase autonomy of well-informed citizen.

2.2 WHY?

2.2.1 The fallout from information overabundance

The amount of information available on the internet has risen by 20% annually for the past 5 years[7] and "AI slop" is beginning to further increase the amount of information available. Whatever harms may affect democracy, scarcity of information is not one of them. Although abundant information was heralded as one of the benefits of the Internet by early enthusiasts [38, p. 33], there are several cognitive and digital design reasons why information abundance is in fact a problem for democracy [39].

First, when information is abundant, people's attentional processes focus on information with the following attributes [40]:

[7] https://explodingtopics.com/blog/data-generated-per-day

- belief consistency (seeking information that confirms one's held beliefs);
- negativity (overestimating risks and danger);
- sociality (being guided by others' choices), and
- predictiveness (being led astray by spurious correlations).

This focus leads to increased polarisation, myopic focus on a subset of political or social options, emotional outrage, out-group hate, and reliance on statistical noise and low-quality content ("bullshit"), respectively [40]. In addition, under overload and time-pressure, people's ability to differentiate between true and false information decreases [41].

Second, these individual-level effects scale up into large-scale trends, such as an observed decline of the public's collective "attention span" over time. For example, the lifetime of popular Twitter/X topics decreased by 32% between 2013 and 2016 [42]. An extended collective attention span, however, is crucial to tackle big societal issues such as "wicked" problems where quick solutions are elusive [43, 44]. Reduced public attention likely also entails diminished political accountability: if a politician's transgression is forgotten or superseded by the next scandal after a few hours, public appetite for accountability is unlikely to persist [45, 46].

Third, the platforms' widespread recent withdrawal from fact-checking offloads the vetting of information onto users [47] and creates new complications. It forces users either to neglect accuracy or to take on the role of fact-checker themselves, prompting further engagement but multiplying the social costs of verification across the entire population. Moreover, platforms enable a new kind of information warfare. Social media "signal jamming" occurs when well-funded or influential parties use their networks to drown out competing voices [48]. This phenomenon has been observed in the US where many moderate voices in public political discourse have been drowned out by "troll armies" [49]. This strategy is also argued to be used by the Chinese government [50, 51].

Arguably, therefore, the abundance of information, combined with false and misleading information, and digitally curated distraction, is imperilling democratic accountability online because meaningful conversation is lost in the noise.

2.2.2 Creation of conflictual echo chambers and echo platforms

While citizens are overloaded with information, they do not see the same information, or more precisely, they do not see the same information through the same lens. Content curation algorithms identify patterns in user-content interactions to predict preferences and provide personalised suggestions. Algorithms do so by recommending to each user the items preferred by similar users. Meta's platforms, including Facebook and Instagram, heavily rely on social network structures to curate and rank content in users' feeds. These algorithms leverage ties among users and amplify moral-emotional content through social reinforcement mechanisms, optimising for engagement such as reactions, comments, and shares ([52]; see Chapter 3). In contrast, TikTok's "For You" feed (the default and most widely used TikTok algorithm) prioritizes individual user interactions with content, such as watch time and likes, largely ignoring the social graph. X (formerly Twitter) uses a hybrid algorithm focusing on content virality and confrontation dynamics within network clusters [53, 54]. Several Very Large Online Platforms (VLOPs) or Very Large Online Search Engines (VLOSEs) mention echo chambers as risks or enablers of systemic risks to civic discourse, electoral processes, and public security in their Digital Services Act (DSA) risk reporting[8], but as the evidence presented below shows, it remains questionable if the solutions implemented by the platforms are sufficient.

The personalisation of information is not new. Offering more personalised or specialised channels for entertainment

[8] https://digital-strategy.ec.europa.eu/en/news/press-statement-european-board-digital-services-following-its-16th-meeting

and political news started with "new" information sources decades ago (cable news, talk radio and blogging) [55], and directly influence people's voting behaviour [56]. Personalisation also makes social media more enjoyable leading to longer app use [57].

Platforms today, however, go one step further and actively select content for users, who have limited choice over what they see on the platforms. The continuous accommodation of users' beliefs and preferences reduces cognitive effort but fosters ideological segregation.[9] That is why, although information is overabundant, attention tends to concentrate on a limited set of sources [58], leading to a reduction in informational diversity. This process reinforces tendencies for similar individuals to flock together and, over time, contributes to the emergence of ideologically uniform spaces—commonly known as echo chambers [59, 60, 61]. Cross-partisan political content is penalised by algorithms; posts that bridge partisan divides receive less visibility and engagement. Optimisation for user attention can inadvertently discourage content that seeks to permeate echo chambers [62]. Knowledge of echo chambers is not new and recent research shows at least partial awareness of them among younger people, but user agency to counter ideological homogenisation is limited by platform's recommender system design and insufficient digital literacy [63].

"Platforms today actively select content for users, who have limited choice over what they see."

The distorted echo provides a toxic lens at other perspectives

These echo chambers are not only about information and beliefs, they also help with the creation of perceived enemies. For example, Facebook's algorithms have been implicated in increasing polarisation [64]. A stronger example is a recent field experiment investigating the Twitter/X algorithm by re-ranking content that expressed antidemocratic attitudes and partisan animosity [65]. When antidemocratic content was down-ranked, participants' outgroup animosity and negative emotions declined compared to a control group that was exposed to the standard Twitter/X algorithm. The experimentally observed changes were comparable in magnitude to three years of increasing affective polarisation in the US [65].

These dynamics do not necessarily arise from deliberate algorithm design but emerge organically from the cumulative effects of personalisation strategies.[10]

Echo chambers are therefore not isolation chambers, but conflicting camps creating and reinforcing opposing identities [68]. They help align individuals' opinions on politics and channel political conflict increasingly into one or two principal issues, which may help explain the increasing perception of alignment between opposing views on migration, climate, vaccination and many other issues that are, in substance, quite different. Perhaps ironically, users of digital platforms may be exposed or connected to *more* diverse content [52, 69, 70, 71, 72, 73], but in a conflictual manner [74], which helps explain why even experiments on online platforms forcing people to "burst their bubble" can increase, rather than decrease, polarisation [75].

Cognition and the public sphere – the move to echo platforms and artificial realities

Over time, the interplay between selective exposure, algorithmic amplification, and content-moderation-induced migration has led to a deeper transformation, from fragmented communities within platforms to the emergence of ideologically distinct platforms themselves. These so-called "echo platforms" [76] are socio-political environments where user demographics, content, and narratives are aligned by design

[9] Ideological segregation refers to the tendency of individuals to consume information that aligns closely with their existing beliefs and viewpoints, resulting in limited exposure to diverse perspectives.

[10] An earlier study [66] found that altering the Facebook algorithm to a chronological feed did not depolarize people in comparison to a control group with a conventional newsfeed. The findings of that study must be interpreted with caution, however, because the newsfeed in the control group was changed during the study period to reduce prevalence of misinformation in the lead-up to the 2020 US election. Moreover, a further recent study showed while algorithms can demonstrably shift attitudes, this effect is not reversed by subsequently withholding the algorithm [67]. Thus, the absence of depolarization when an algorithm is withheld does not negate the possibility—now repeatedly confirmed—that the algorithm polarized people to begin with.

and by self-selection (e.g., X vs. Bluesky). Platforms should thus no longer be treated as isolated systems[11]; rather, they constitute an interconnected ecosystem. Evidence shows increasing toxicity in longer conversations and a progressive simplification of user language, phenomena observed across multiple platforms and temporal spans [77, 78]. The object of analysis for democracy should no longer be the individual platform, but the online ecosystem as a whole.

A further profoundly powerful technology contributing to fracturing of reality is generative AI. To illustrate, a recent experiment found that deceptive explanations generated by the most recent generation of LLMs are more persuasive than accurate, honest explanations and can amplify the belief in false news [79]. Another study found that AI-generated misinformation is more trusted than human-generated misinformation [80]. This power can also be used for good, e.g., for dissuading people in believing unsupported conspiracy theories [81], but platforms have no natural incentive to do so when toxicity contributes to greater user engagement, calling for policy intervention. Furthermore, LLMs *are designed* to create the appearance of knowledge even where there is none, which creates a cognitive trap for users: the illusion of knowledge [82]. Unlike human experts, LLMs do not say "I don't know" in response to a query that they cannot answer, see textbox for details.

Echo chambers are not purely the result of digital influence. A recent megastudy disentangled the roles of algorithms and human self-selection in creating segregated online communities and found that self-selection is an important factor. Analysis of news consumption data from around 208 million Facebook users [85] revealed that ideological segregation is already high in the content presented to

[11] In EU regulation such as the DSA, platforms are required to perform risk assessments of their own influence. However, the evidence over echo platforms suggests that regulators should take a broader perspective and think about the information system as a whole.

When Generative AI goes social: Generative AI and epistemia

Generative AI today is seamlessly integrated into several social media platforms (Grok on X, Meta AI on Facebook/Instagram/Whatsapp, Sora 2 on Sora, Truth Search on Truth Social).

An audit by NewsGuard found that one of the most recent additions of genAI — *Sora 2* — produced videos that advanced provably false claims in 16 out of 20 trials (80%).[a]

The implications of this integration go beyond the generation of specific disinformation videos: the blending and-integration of generative AI with social media extends the incentives of the attention economy to software that, by design, is not grounded in reality.

LLMs do not retrieve facts. They generate text, by predicting the most statistically probable continuation of a sequence. This is a statistical process, with plausible linguistic patterns distilled from massive corpora of human language [83]. Likewise, video generators such as Sora 2 operate without regard to a ground truth—their entire purpose is to create a plausible illusion to increase engagement.

The direct or indirect integration of genAI into powerful social media sites thus creates an increasingly relevant cognitive trap: the *illusion of knowledge* created by an LLM's fluent language, omnipresence, and realistic videos does not actually resolve informational gaps, but nonetheless gives the user the sense that those gaps have been filled.

When combined with the business models of platforms, this creates the conditions for a distinct new informational regime which is termed *epistemia* [82]. Epistemia lowers the threshold of responsibility in content creation ("the AI said it" becomes a deflection, not a citation) and may foster a false sense of competence in users, who weave LLM outputs into their reasoning without noticing that the substance is hollow. Evidence already shows how human language shifts towards words preferentially used by chatbots [84].

[a] https://www.newsguardrealitycheck.com/p/openais-sora-2-is-a-willing-hoax

users (by platforms) through algorithmic curation and rises even further when moving to active engagement (initiated by users), measured by likes and comments, in line with findings on motivated reasoning [86, 87]. As a rough guide, about a third of segregation can be attributed to the algorithm, with the remaining two thirds resulting from people preferentially engaging with attitude-aligned content.

Recent evidence has also found that echo chambers play a crucial role in facilitating the spread of harmful content: purveyors of hate dominate such environments, acquiring central positions in online networks and leveraging echo-chamber interactions to amplify hate speech diffusion [88].

Thus, echo chambers and platforms cater to user preferences and reinforce them further, while preparing the ground for less careful investigation of facts and truths, thereby raising the levels of toxicity and offline harms.

2.2.3 The impact of "conventional" dis- and misinformation

Platform algorithms facilitate the merging of true, misleading and false information making it ever harder for audiences to know what is true. However, understanding how this threatens societies requires disentangling causal mechanisms, the prevalence of falsehoods, and possible remedies. Misinformation is simultaneously an outcome of prior conditions (e.g., polarisation; [89]) and a cause that affects not only people's attitudes and behaviours directly, but also those very conditions, creating feedback loops that complicate any simple causal inference of harms [90, 91]. To navigate this complexity, it helps to think in three nested causal levels: individual, information-environment, and societal.

At the level of the individual, experimental evidence shows that direct exposure to false claims can shift beliefs (in facts), attitudes (i.e., the evaluation of something as good or bad), intentions, and, in some contexts, behaviour [e.g., 92], with the effect size being greatest for beliefs and smallest for behaviours. The information-environment level embeds those individual effects within a wider media ecosystem shaped by different actors and various types of media and their "affordances" or structural elements (e.g., algorithmic curation of content in social media newsfeeds). Finally, the societal level comprises structural drivers—trust in

"Platform algorithms facilitate the merging of true, misleading and false information making it ever harder for audiences to know what is true."

institutions, economic inequality, legal frameworks, platform business models—that condition both the supply of and demand for misinformation and are often the more important parts to address.

Crucially, misinformation can be a cause, effect, mediator (i.e., the process through which it affects other variables), or moderator (i.e., an amplifier of the effect) within the same causal chain [93]. For example, economic inequality (societal) can erode institutional trust (information-environment), heightening individual susceptibility to conspiratorial narratives that, in turn, intensify polarisation, thereby impacting the political culture that originally enabled misinformation. Conversely, a surge of coordinated misinformation about alleged election fraud (information-environment) can shape individual perceptions especially when coming from "trusted" sources, eroding confidence in electoral procedures. Acknowledging the multi-layered causal architecture outlined above is essential to avoid the false dichotomy that frames misinformation as either a mere symptom of deeper social challenges or an autonomous driver of current societal harm (see also [91]). In

what follows, the report focuses on the evidence of harms from false and misleading information on two causal levels: the level of the broader information environment and the level of individual or collective exposure to misinformation.

Harms on the level of information environment

Multiple studies illustrate an association between misinformation and adverse outcomes. For example, a panel study found that a loss of trust in traditional media and a move towards populist parties in Germany were predicted by disinformation beliefs [94]. Likewise, belief in misinformation predicts lower trust in traditional media over time and that, in turn, predicts more misinformation exposure over time [95]. Low trust in conventional media and belief in misinformation seemingly reinforce each other in a feedback loop, likely mediated by social media.

The mediating, but nonetheless causal, role of social media can be illustrated in the following example: Even though anti-refugee sentiment is a necessary background factor for hate crimes, it is not their sole cause. One causal analysis [96] shows that these crimes are influenced by social media consumption. When access to social media is temporarily cut off due to an internet outage and the inciting posts from far-right parties are unavailable, fewer hate crimes were recorded [96].

The role of social media platforms in this interaction is not politically neutral. Although misinformation is usually not directly promoted by platforms, as mentioned above, content that includes negativity, outrage and negative outgroup sentiment is particularly visible and algorithmically successful [97, 98, 99]. Misinformation in turn has been shown to evoke more outrage than do trustworthy sources [100]. Thus, recent studies have found that social media algorithms disproportionately favour parties spreading more misinformation whereas politicians who shy away from inflammatory language are experiencing much lower engagement [101, 102, 103].

Harms at the level of individuals or groups

Measurable harms from misinformation are readily observable in laboratory experiments [104, 92]. One limitation of those studies is that they give participants no choice about the material they consume, thus preventing them from using strategies to resist misinformation, such as avoiding poor quality sources which many citizens do [105]. There are, however, numerous field studies that show that these effects also hold outside the lab [13]. The textbox below summarizes the evidence of those harms by policy domain.

Misinformation also exploits existing vulnerabilities—such as polarisation and institutional distrust—and channels them into concrete outcomes, from changes in health behaviour or reduced support for climate action to erosion of electoral trust or activating extremists to act on their beliefs. From this perspective, a particularly consequential danger of misinformation is not the belief in specific falsehoods but the follow-up actions from those beliefs which may erode the very conditions under which societies can deliberate, cooperate, and correct course.

In other words, the fight against misinformation cannot be a replacement for fighting the root causes that make misinformation attractive.

“Low trust in conventional media and belief in misinformation seemingly reinforce each other in a feedback loop, likely mediated by social media.”

Overview of harms arising from misinformation by domain

It can be difficult to define "harms" in an unequivocal way, since consequences such as growing support for a favoured political direction, or the increased belief in a contentious claim, may be welcomed by some groups in society. The authors acknowledge this issue and list a broad range of harms as identified in the scientific literature:

Democracy and elections

- Reduced trust in elections by false voter fraud accusations, in particular among politically aligned voters [106, 107].
- Incitement of political violence; e.g., cascades of leaders' tweets predicted bursts of violence during the January 6 Capitol riot [108].
- Belief in misleading political claims, especially when politically aligned, but limited and inconsistent effects on broader political attitudes or behaviours [109, 110, 111, 112, 113].

Public health

- Reliable shifts of beliefs in false claims, variable impacts on vaccination intentions [92, 114, 115, 116, 117, 118, 119, 120].
- Misinformation exposure is leading to reduced vaccination rates in Italy [121] and increased hesitancy in the US [121].
- Exposure to information that downplayed the severity of the pandemic has also been causally linked to a greater number of COVID-19 cases and deaths [122].
- Anti-vaccination content on Facebook flagged as false by fact-checkers has been estimated to reduce COVID-19 vaccination intentions by 1.36 percentage points [116], although the prevalence of such false content is low (0.3% of vaccine-related views) [116].
- By contrast, factually accurate but deceptive or misleading anti-vaccination content reached over 20% of US Facebook users and hence was estimated to be 46 times more consequential for vaccine hesitancy than false information that was flagged by fact-checkers [116].

Climate action

- Reduced acceptance of evidence for climate change [123, 124].
- Varying impacts on attitudes [125, 126].
- Partially lower pro-environmental intentions [127, 117], self-reported behaviour [128], and actual behaviour [129].
- Shifts in public narratives away from the reality of climate change towards costs, feasibility and fairness [130].

Hate crime, extremism

- Causal evidence that misinformation increased participation in killings during the Rwandan genocide [131].
- Causal evidence that Facebook exposure facilitated hate crimes in Germany [96].

Erosion of trust and social norms

- Erosion of trust in news media [132, 95].
- Overconfidence in own ability to discern information [132].
- Increased willingness to exhibit extreme behaviours such as racism [133, 134, 135].

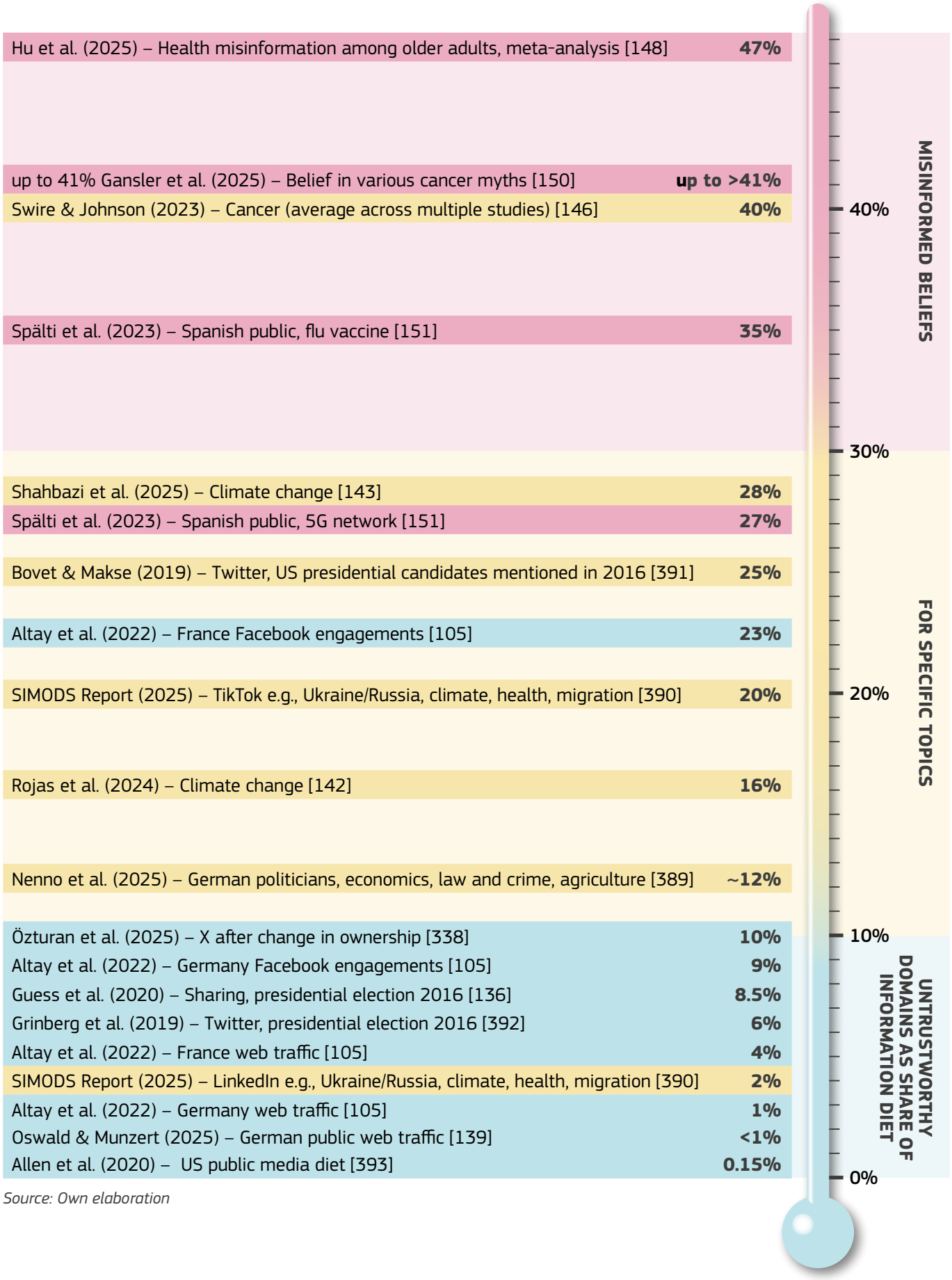
Prevalence of misinformation

The question of harms is inseparable from the question of prevalence. Even if exposure to a single false claim produces only modest effects, the consequences at scale depend on how widespread such content is, how often people encounter it, and how many come to believe it.

Researchers have used three different approaches to estimate the prevalence of misinformation:

- Untrustworthy domains as share of information diet.
- Misinformation encountered for specific topics (e.g., during a targeted search, or by using a hashtag).
- Prevalence of misinformed beliefs in the population.

Figure 2.1: Prevalence of misinformation in untrustworthy domains in general, specific topics, and misinformation beliefs.



Source: Own elaboration

Figure 2.1 summarizes the estimates obtained using those three approaches.

Prevalence of untrustworthy sites in overall information diet. These estimates rely on ensembles of sites that are considered untrustworthy. Those ensembles are carefully curated and cross-checked against fact-checkers [136]. One limitation of using such lists is that they are often relatively small, ranging from 21 [137] through 490 [136] for studies that examined the 2016 US presidential election [138]—although some more recent studies rely on thousands of domains [105].

Between 2017 and 2021, in the UK and Germany, untrustworthy websites represented less than 1% of all visits to news websites [105]. In the US and France, the figure rose to 4% [105]. These findings have been confirmed recently in Germany, where untrustworthy sources made up less than 1% of the news diets of the German population between 2017 and 2024 [139]. The numbers increase considerably when social media platforms such as Facebook are analysed. During the period 2017-2021, in France and the US, untrustworthy websites captured 23% and 19%, respectively, of all engagement on Facebook [105]. In Germany and the UK these numbers were lower (between 2% and 9%) [105]. Similarly, outright false information was relatively infrequent on Facebook in Germany during the COVID-19 pandemic (about 1%; [140]).

When assessing those estimates, it should be remembered that they exclude numerous known vectors of misinformation, from misleading political ads to false statements by politicians or articles reported in more mainstream media. Some research has considered those other vectors. For example, a study of nearly 14 million Facebook posts during the 2020 US election that focused on images, rather than text, estimated the prevalence of misinformation in political content to be around 20% (albeit including some humorous and satirical content) [141].

Prevalence of misinformation for specific topics. Several studies have quantified false information in content relating to climate change. Estimates range from 15.5% among relevant tweets [142] to 28% on a large dataset of posts drawn from Twitter, Reddit, and YouTube [143] An examination of YouTube's "up-next feature" revealed that 16% of the top 100 related videos in response to the search term "global warming" contained misinformation [144].

"Between 2017-2021, in France and the US, untrustworthy websites captured 23% and 19%, respectively, of all engagement on Facebook."

For health misinformation, a meta-analysis of 65 studies reported estimates ranging from 30% (for medical treatments) to 97% (for recreational drugs), with vaccines (43%), eating disorders (36%) and noncommunicable diseases and pandemics (40%) occupying the middle ground [145]. Focusing specifically on cancer, estimates were summarized by [146]. Across nine studies, prevalence estimates ranged from 12% for podcasts about prostate cancer to 100% for the top 50 videos on prostate cancer screening on TikTok. The average estimated prevalence across studies was 40%.

Overall, it is clear that misinformation is a notable concern when considered as a share of information available online on important topics. In addition, harms from misinformation cannot be captured by studying exposure to blatantly false "fake news" alone. As Allen and colleagues [147] argue, misperceptions (i.e. persistent, incorrect belief contradictory to best available evidence) are often fuelled by accurate-but-misleading content and elite-driven rhetoric, which are far more prevalent than fabricated headlines. The next section looks into the prevalence of misperceptions.

“One large multi-country study shows that nearly half of older adults hold misinformed beliefs about important health issues.”

Prevalence of misinformed beliefs. Ultimately, for misinformation to matter, it has to be believed. Relevant research reveals a problematic picture. For example, one large study from multiple countries including in the EU shows that nearly half of older adults (aged above 50) hold misinformed beliefs about important health issues [148].

A meta-analysis on health literacy among older adults in the EU identified a prevalence rate of 47% of the population with limited (“insufficient or problematic”) health literacy [149]. In a representative sample of Americans, only 25% of respondents were found to reject all of 5 common misconceptions about cancer [150], with acceptance rates ranging from 7% to 41% across items. In a recent and large sample of over 5000 Spaniards, misperception about genetically modified foods and general health risks of 5G technology were particularly prevalent [151], with the latter subject often being implicated in online debates. Thus, at least some misperceptions are widely spread, while others are less so, with a current lack of clear evidence on the reasons for the variation. It is hard to pin down misperceptions as being caused by social media, but a recent pre-print shows that the attention economy incentivises writers to reduce the information they communicate, which in turn increases biases in beliefs especially among people with lower interest [152].

2.2.4 The move from isolated misinformation to the “Fantasy-Industrial Complex”

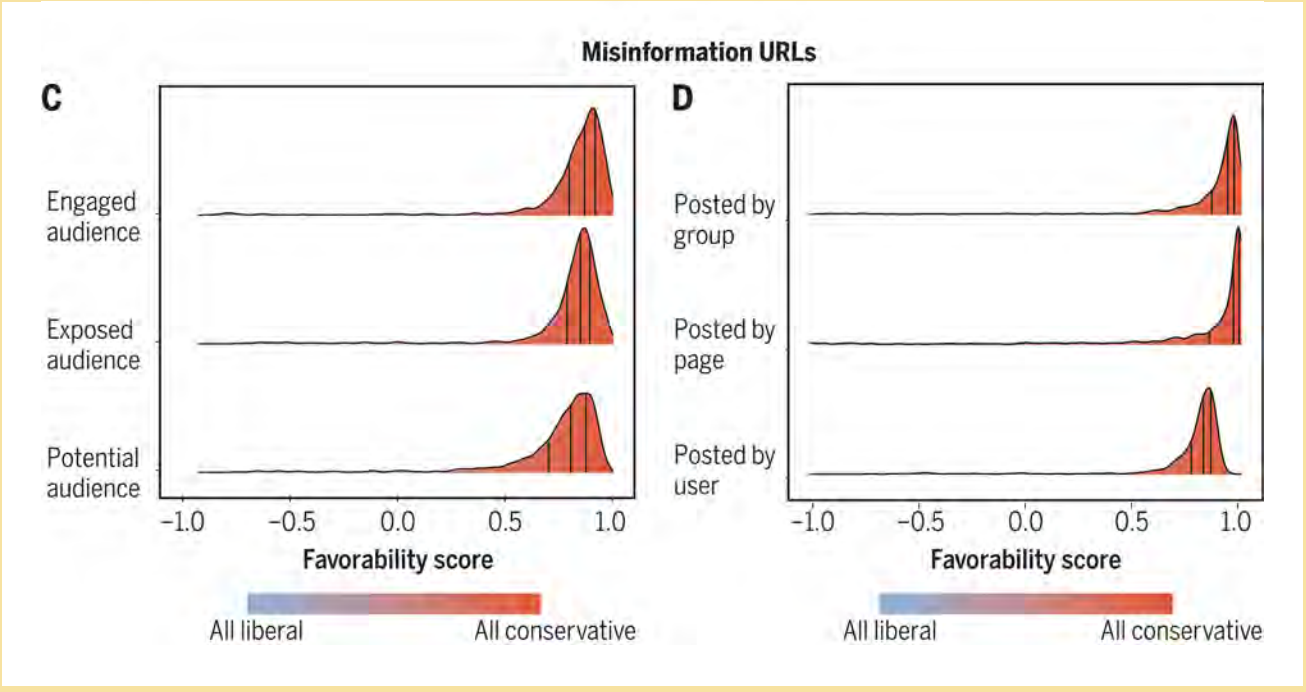
Beyond individual harms of misinformation, information overabundance and echo chambers and echo platforms have enabled another profound shift in the information space. This shift is best illustrated by considering an historical example. Today, it is known with considerable confidence that there were no Weapons of Mass Destruction (WMD) in Iraq immediately before the US-led invasion in 2003,

The skewness and political asymmetry of misinformation spreading and consumption

On Facebook, 20% of users with the most conservative news diet accounted for 62% of all exposure to fake news sources [109]. On Reddit, a small minority of active users (≃ 3%) accounts for most of the toxic comments posted on the platform [153]. On YouTube, 6.3% of users account for 79.8% of exposure to extremist channels [154]. Likewise, visits to vaccine-skeptical websites are heavily skewed [155], such that people with favourable views towards vaccines are extremely unlikely to visit them (2% of vaccines websites visited), while people with very unfavourable views of vaccines are much more likely to visit them (18% of vaccines websites visited) and account for most of the visits.

Exposure to misinformation is concentrated on the political extremes but not uniformly [156]. An analysis of content from political parties across 26 countries found that radical right populists (not left-wing populists) spread more misinformation than other parties [157]. This ideological slant of online misinformation has been frequently reported in terms of production [158, 159], as well as for exposure [160, 161]. Consistent with this pattern, Oswald and colleagues [139] find that exposure to untrustworthy websites in Germany in 2017 and 2024 is concentrated among far-right users and associated with lower democratic satisfaction. Figure 2.2 provides a striking illustration of this slant.

Figure 2.2: Most news stories labelled as misinformation on Facebook (97%) have audiences that are conservative on average (left panel). A similar pattern is observed for sharing (right). Figure based on data from more than 200 million Facebook users.



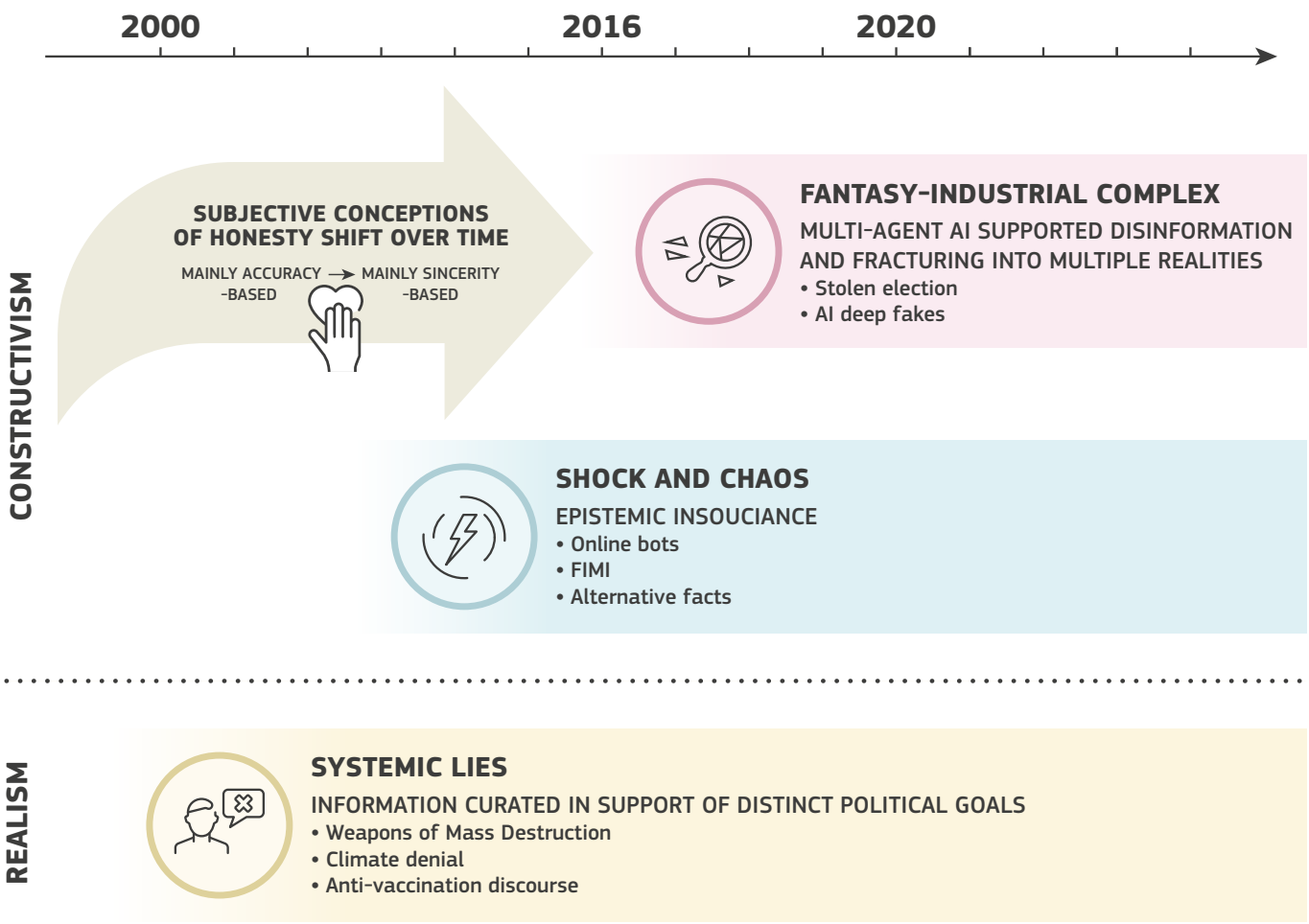
Source: [85]

even though political leaders had expressed certainty at the time that those weapons existed and constituted a threat [162, 163].

The original claims about WMDs constitute a textbook case of misinformation, which was later corrected: Many of the same sources that promulgated the original misinformation acknowledged the error later, confirming that when actors contest the truth with a shared commitment to reality, misinformation can often be identified and corrected [164] and, in principle at least, actors can be held accountable for their rhetoric.

Arguably, there has been a shift in the information space since the early 2000s that makes such accountability increasingly challenging and that has reshaped the entire notion of misinformation. Figure 2.3 provides a conceptual overview of this evolution. The figure illustrates an apparent shift in the type of disinformation deployed by some prominent political actors in western democracies. This new type of disinformation can be called a "shock and chaos" regime [165]. This regime is characterised by indifference of the actors to the truth and relies on an extreme form of constructivism in which "truth" is entirely in the eye of the beholder, something that is particularly easy to achieve with the aid of social media, assisted by echo chambers and platforms. A characteristic attribute of shock and chaos is that specific falsehoods are not meant to be believed; instead, their function appears to

Figure 2.3: Conceptual overview of the transformation of misinformation from "systemic lies" to "shock and chaos" which has given rise to the fantasy-industrial complex.



Source: Own elaboration

undermine the possibility of finding common ground. When shock and chaos is practiced by a multitude of actors, from politicians to influencers to members of the public, the resultant "fantasy-industrial complex" can create entirely fictional alternative realities. These are built from selected pieces of reality, misrepresentations thereof, and outright falsehoods.

In a fight for attention, corrections and fact-checks rarely travel far. Shock and chaos disinformation has therefore also been described as the "firehose of falsehood" propaganda model [166].

2.2.5 The new world of authenticity in the fantasy-industrial complex

Fighting individual falsehoods and susceptibility to misinformation is important, but demand-side interventions do not absolve platforms or political elites. By shaping the broader information environment, they can normalise problematic beliefs, shift perceived norms, and amplify fringe narratives. The public in liberal democracies overwhelmingly values honesty in their politicians [167] and they want impartial, transparent, and accurate news reporting [7]. It is therefore seemingly paradoxical that voters in several countries have elected candidates who are not known for truthfulness and integrity.

This paradox can be explained by considering what being "honest" actually means. In addition to accuracy, a perceived component of honesty by citizens is the authenticity of expression of one's belief [168, 169]. It is not so much about what is said, but how it is said. People may thus consider a politician or celebrity to be honest despite being wildly inaccurate when expressing a sincerely held—if mistaken—political belief. In contrast, factually correct statements by public figures may be seen as dishonest when answered in a way that may be technically accurate but seemingly lacking authenticity or evading the question.

This reframing of honesty from accuracy to authenticity or sincerity is strongly connected to the shock and chaos disinformation regime, with famous terms such as "alternative facts" and "truth isn't truth" [170] constituting part of the vocabulary of the fantasy-industrial complex. The following actors can contribute to the fantasy-industrial complex:

Politicians

- Social media gives politicians direct public access, shifting agenda-setting power from traditional media to themselves [171, 172].
- One example is research showing a politician's ability to use Twitter to divert media attention from topics that would supposedly harm them politically to those favouring their political strengths [173].
- Platforms algorithmically prioritise content over personal connections, amplifying sensationalist or emotional posts, giving power to those politicians who embrace this new logic [174].

Corporate Actors

- Platform owners can reinforce one-sided political agendas: After X's owner endorsed the Republican candidate for president in July 2024, Republican-leaning accounts received a boost in visibility relative to Democrats, and the owner's posts gained 17.1 billion views (July–November 2024), surpassing all political campaign ads on X [175, 176].
- A recent experiment confirmed that the X algorithm is promoting conservative content and demotes posts by traditional media, shifting users' political opinions to the right [67].
- In 2025, several platforms reduced fact-checking and moderation, impacting democratic discourse and human rights.[12]

[12] https://www.europarl.europa.eu/doceo/document/E-10-2025-000423_EN.html

"The fantasy-industrial complex allows multiple self-organising actors to blend kernels of true—but hyperpartisan—information with false and misleading information to create a fictional view of reality."

Influencers

- Influencers are online content creators who position themselves not as media but as ordinary relatable people [177].
- Political campaigns have been collaborating with influencers to tailor messages for specific audiences for several years [35, 178].
- "Mega-influencers" (e.g., HugoDécrypte in France, Joe Rogan in the US) reach millions, sometimes surpassing traditional media. In countries with limited press freedom, influencers can also provide independent news coverage [7].
- According to a recent Eurobarometer, 74% of young Europeans (age 15-24) follow influencers, for product reviews, tutorial but also commentary on social and political affairs.[13]
- Virality depends more on emotional resonance [179], and alignment with audience beliefs than credibility of the institution [180], risking a wider and easier spread of misinformation.
- Studies show that many influencers are open to promoting high-quality content when provided with resources (e.g., mental health communication toolkits) [181]. For example, the recent arrival of scientists and medical experts on TikTok, aided by a World Health Organization initiative, points to how influencers can be a force for good [182].

The fantasy-industrial complex thus emerges from the interrelated problems of an information overabundance that triggers individuals to focus on negative and belief-consistent content, echo chambers/platforms that preferentially supply people with information confirming their position, and an increase of toxicity. This lead to the illusion of understanding [183] and false consensus effects where people perceive their views to be the norm, no matter how extreme. This way, the fantasy-industrial complex allows multiple self-organising actors to blend kernels of true—but hyperpartisan—information with false and misleading information to create a fictional view of reality, typically with a specific political tendency.

Taken together, everyone can paint political opposition as deranged and misinformed, by focusing on outliers, latest gaffes, and isolated events that go viral as proof of their position. At the same time, truly destructive political forces can hide behind the outrage machine where today's scandal is inevitably followed by the next outrage tomorrow. Studies have repeatedly shown that individuals' perceptions of opposing party supporters are more extreme and negative than their positions actually are [184]. This often justifies undemocratic behaviours—even among those valuing democracy—to win over the opposition [185].

2.3 HOW?

2.3.1 Systemic reform rather than individual responsibilisation

The harms identified above emerge from the interaction between people and the architecture that determines how content is surfaced, ranked, and circulated [154, 186]. Their decisions are not necessarily guided by ideology or specific intent (although see Chapter 4), but by optimisation tuned to measurable outcomes: increase engagement, retention, click-through rates, and so on.

It follows that blaming individuals for being misinformed or caught in an echo chamber is not only potentially unproductive—it is misleading. The responsibility lies with the platform providers—including new AI chatbot providers—whose design choices are enabled or at least tolerated by legal frameworks and incentive structures that prioritise the monetisation of attention. Hence, the problems are *systemic* and hence require a systemic solution.

[13] https://europa.eu/eurobarometer/surveys/detail/3592

2.3.2 Transforming an engineered space

The digital public and private information space is an engineered space. Like any infrastructure, it must be governed through enforceable standards, public oversight, and shared responsibility—including citizens—if democracy is to remain viable. When it is collectively acknowledged that responsibility is shared between platforms, institutions, and publics, it is possible to establish the conditions for a digital environment that supports deliberation, diversity, and informed democratic participation.

The resilience of Wikipedia

Transformation of the digital information sphere requires new spaces for deliberation. Collective human intelligence can be harnessed to achieve great things through digital commons—a resource that is in under-supply if left to market forces alone. Wikipedia is a particularly illustrative example because it has proven resilient to disinformation campaigns, even concerning content surrounding highly polarised issues [187] and contested events such as the Russia-Ukraine war [188].

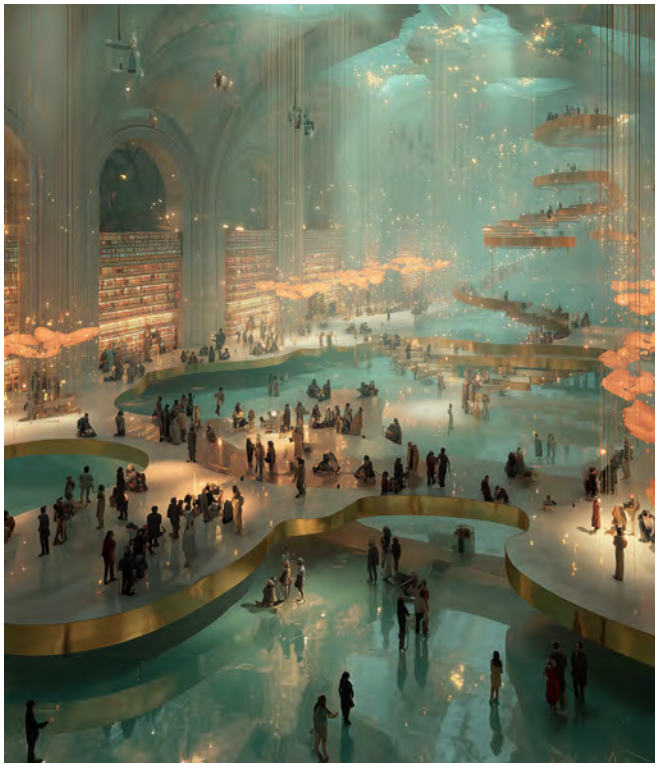
The success and resilience of Wikipedia rests on its layered and extensive set of rules and processes [189]. These range from a finely graded system of editorial privileges based on a contributor's prior record to the use of bots to identify vandals. These allow the site to adhere to a strict information value hierarchy [190]. A simulation study identified two crucial factors for Wikipedia entries to be resistant to disinformtion attacks: the presence of administrators and the ability to revert an edit [191]. The existence of an open platform in which everyone can participate that is nonetheless resilient to attacks is an important confirmation of the potential of the Internet to inform and support, rather than undermine, the public space.[14]

Consensus reality requires a consensus infrastructure. However, the current business models and the dominance of foreign players largely make EU bottom-up innovation and alternatives unlikely to emerge on a larger scale, which the next two chapters will explain in more detail. Nevertheless, more digital commons can be made available as a step in the right direction.

Deliberative architectures for restoring societal trust

One of the core challenges of an online business model based on engagement is the inability to foster constructive debate. There is evidence that implementation of the right structures and interventions for interaction, especially offline, can reduce affective polarisation in general [192, 193]. Applying these insights to democratic debate, there is growing evidence that deliberative bodies comprising randomly chosen citizens, when properly moderated and facilitated, can ameliorate polarisation and "post-truth" discourse [194]. The involvement of citizens in deliberative assemblies or "mini publics" has been identified as an important countermeasure to the corrosive effects of disinformation campaigns by malicious actors [195]. A recent field experiment with six districts in Germany with 435 participants shows positive effects on political trust, efficacy and participation among participants, and suggestive evidence for a reduction in receptiveness to conspiracy theories if representatives from multiple parties participated in the events [196].

In a survey of 105 organisers of citizen engagement exercises in 20 European countries, 97% or respondents would



[14] Nonetheless, Wikipedia's resilience can at least temporarily be undermined by paid bad-faith actors who act on behalf of wealthy clients, https://www.thebureauinvestigates.com/stories/2026-01-14/london-pr-firm-rewrites-wikipedia-for-governments-and-billionaires.

conduct an engagement again, showing the high value to public administrations, but also significant capacity gaps, especially in terms of skills to conduct or procure such work [394]. More public investment and training is therefore needed to scale these democratic innovations beyond their current, often rather localised, effects.

To scale, deliberation likely has to be done at least partially digitally or in a hybrid fashion, and successful precedents exist. An advanced and well-established platform for consultation and deliberation is used in Taiwan called vTaiwan.[15] The platform combines online and offline interactions through various channels for engagement, including an online consultation platform, community page, slack channel, weekly online meetings, and occasional hackathons, enabling citizens, government officials, representatives, and other stakeholders to discuss legislative proposals and generate non-binding policy solutions. There are two noteworthy design features that contribute to part of vTaiwan's success. First, users are unable to reply to comments, thus preventing vitriolic exchanges. Second, the platform clusters users based on their opinions by statistical techniques [197], which makes argumentative dividing lines, as well as space for consensus, visible [198].

In Europe, *Decidim*[16] offers a digital platform for participatory democracy, available as open source software, which has been deployed in multiple municipalities (e.g., Barcelona, Helsinki) and at regional (e.g., Catalonia in Spain) and national (e.g., Japan) levels. As of 2024, there were more than 450 deployments of *Decidim* in 30 countries [199]. In Iceland, a novel online platform was specifically designed to facilitate public participation in the constitutional reform process that took place in 2011 [200]. The wiki-based platform allowed citizens to contribute and edit the constitution in a collaborative manner.

The European Commission's own Citizens' Engagement Platform[17], launched in February 2024, is another example [201]. The platform allows citizens to share their ideas and proposals on relevant EU policy, in principle giving all EU citizens an equal opportunity to participate, promoting an inclusive and accessible democratic process. This potential can only become reality, however, if the platform and others like it are used and if the outcomes of the online deliberations are integrated into policymaking. Lessons learned include the need for inclusiveness, multilingualism, accessibility, and transparency, while the content needs to be relevant, well-moderated, and relevant to political decisions [201]. The use of AI in deliberative digital platforms may further support this goal. AI can analyse and synthesise vast amounts of citizen contributions, identify emerging trends and priorities, and provide policymakers with actionable insights. AI can also facilitate analysis, identify areas of agreement and disagreement and help discard spam, ultimately fostering an inclusive and participatory governance—if biases are kept at bay. However, the use of AI in deliberative democracy also poses significant challenges, including concerns about ethics, trust, algorithmic bias, data quality, and transparency.

The EDS proposes the EU-wide promotion of citizen and youth participation. If done wrong, citizen participation can lead to further disillusionment [202]. Thus, for the network to be effective, it needs to follow evidence-informed strategies of engagement along which political buy-in. Outcomes of those engagements need to be implemented

"In a survey of 105 organisers of citizen engagement exercises in 20 European countries, 97% of respondents would conduct an engagement again."

[15] https://info.vtaiwan.tw/
[16] https://decidim.org
[17] https://citizens.ec.europa.eu/index_en

Successful deliberative democracy: The case of Geraldton, Western Australia

The City of Geraldton is a regional community in Western Australia. The city faced significant "wicked problems"—economic uncertainty, climate and sustainability challenges, and low public trust in local government. In response, the city implemented a four-year "Deliberative Collaborative Governance" (DCG) program [203]. The initiative aimed to involve citizens directly in shaping policies to increase transparency and rebuilding mutual trust between the community and government.

The DCG program. The program employed a range of "deliberative democracy" techniques designed for inclusive, informed, and ongoing participation:

- *Alliance Governance Group.* Brought together government, business, Indigenous, and community representatives (though later disbanded due to limited decision power).
- *Community Champions.* Local volunteers trained to host small-group "World Cafés" and "Community Conversations" to gather grassroots perspectives.
- *Digital Engagement.* Online platform *CivicEvolution* allowed residents to propose and deliberate community projects.
- *Minipublics.* Assemblies of citizens such as Citizens' Juries and Deliberative Participatory Budgeting Panels where randomly selected residents deliberated on complex local issues.

Outcomes and impacts.

- *Improved trust and engagement.* Surveys and interviews showed increased citizen confidence that the city listened and acted on public input. Government officials gained greater respect for residents' capacity to deliberate on complex topics.
- *Better policy and decision-making.* DCG led to tangible policy outcomes such as the "Digital Futures Plan", new urban design frameworks, and long-term participatory budgeting plans.
- *Democratic reform.* DCG evolved from one-off consultations to an ongoing, institutionalised decision-making model. The approach built civic capacity, reduced adversarial politics, and promoted shared responsibility.

and followed up. The textbox below presents an illustrative case study from Western Australia of successful deployment of online and offline deliberation techniques [203].

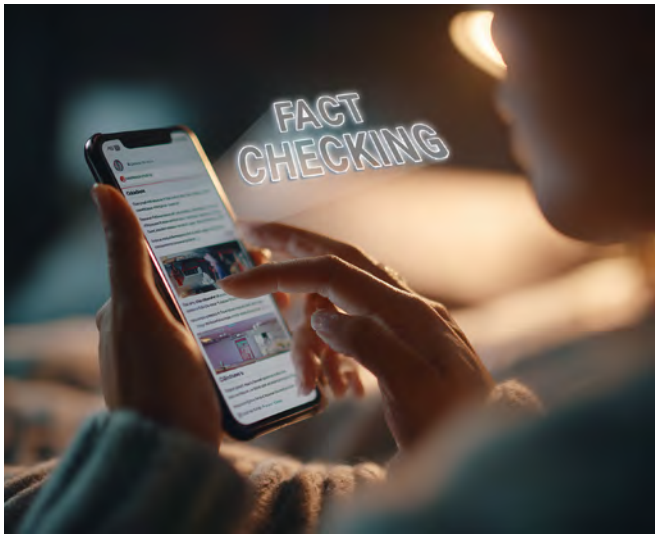
The EDS also suggests the setup of an EU Civic Tech Hub. Following the evidence presented above, it should be founded on the principle that the digital public sphere is a shared infrastructure requiring collaborative governance among platforms, institutions, and citizens. The hub could prioritise the creation of open and secure platforms that are modelled after successful examples such as Wikipedia.

The hub can help guide investing in digital commons, funding for R&D, and setting standards that ensure public interest prevails over the damaging engagement (at all cost) logic. The hub can help blend online and offline spaces to encourage constructive debate and reduce polarisation in a scalable fashion across the EU, by encouraging platform design that clusters and synthesizes opinions, and prevents toxic exchanges. To ensure lasting impact, the hub should go beyond technology and focus on building capacity, securing political commitment, and fostering a culture of shared responsibility. This involves training public administrators and civil society in participatory methods, ensuring that deliberative outcomes are acted upon, and continuously evaluating and adapting approaches based on evidence. The use of AI can help these efforts, but only if deployed ethically and transparently to avoid reinforcing biases or eroding trust.

2.3.3 System-level countermeasures to misinformation

At the platform level, such interventions can target producers of misinformation, for example, by demonetising accounts that disseminate falsehoods, downranking content flagged

as false by independent fact-checkers, or suspending repeat offenders and superspreaders of misinformation. Although certain actors often frame these measures as threats to freedom of expression, the public and experts view system-level measures favourably, opening a space for more regulatory intervention. Multiple surveys show that the public in different countries, including Europe and the US, largely approves of content moderation and other on-platform interventions against harmful content (such as misinformation or hate speech) [204, 205, 206]. A survey conducted in 2024 among 412 experts on the information environment [207] also found strong support to prioritise system-level interventions to improve the information environment and to support free and independent media. A majority of experts also supported expanding content moderation on social media platforms, deplatforming problematic actors, and treating platforms as publishers. Concerns about chilling effects of such measures on free speech are further mitigated by evidence of the skewed distribution of misinformation. As shown earlier (Figure 2.2), false and misleading content is highly concentrated: a small minority of highly active users and outlets account for the vast majority of shares. System-level moderation therefore would primarily affect this small minority of "supersharers", rather than the general public, making proportional interventions not only feasible but also likely to be effective with limited impact on ordinary users' speech [208].



One problem with the supply of misinformation is that it is often profitable. For example, a report by NewsGuard in 2021 suggested that top brands alone are spending $2.6 billion annually on advertising on misinformation sites.[18] A report from Reuters claims that 10% of Meta's 2024 revenue is projected to come from fraudulent ads.[19] Research shows that the misinformation-advertising link is ubiquitous and often occurs without platform awareness [209].[20] As long as this is the case, there is an incentive by malign actors to generate and spread misinformation. Stronger regulatory measures and enforcement, for example under the upcoming Digital Fairness Act[21], could be used to address this problem. For example, advertising distribution companies could be required to scrutinise fraudulent ads in collaboration with consumer protection agencies and to offer advertisers tools so they can choose to bypass websites that are found to be violating standards of information integrity established in the EU.

Empirical evidence on the effectiveness of system-level interventions remains limited, largely due to restricted access to platform data. Still, several causal studies suggest that regulation and platform design can curb harmful content. Germany's NetzDG law has been found to reduce the toxicity of refugee-related tweets by nearly 30% and slightly lowered anti-refugee hate crimes in high-exposure municipalities [212] without, however, chilling freedom of expression [213]. Complementary evidence was obtained in a large-scale search engine experiment across ten countries, where algorithmic downranking of low-credibility websites reduced engagement with misinformation by more than half [214]. The EU could develop a system of assigning credibility to websites by democratic oversight—such as through national citizen assemblies—that platforms could then use to adapt their algorithms.

Additionally, crises put an extra burden on public institutions with respect to social media platforms and AI, where misinformation travels rapidly and trustworthy sources may lag behind. Empirical research on crisis communication suggests that social media use by organisations during crises is associated with lower perceived responsibility for

[18] https://www.newsguardtech.com/special-reports/brands-send-billions-to-misinformation-websites-newsguard-comscore-report/
[19] https://www.reuters.com/investigations/meta-is-earning-fortune-deluge-fraudulent-ads-documents-show-2025-11-06/
[20] For example, Google supports the placement of ads based on users' search terms, which opens the door to purveyors of "alternative" pseudoscientific treatments to place their ads prominently on the results page when a user searches for terms such as "stage 4 cancer treatment" [210, 211].
[21] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14622-Digital-Fairness-Act

"Germany's NetzDG law has been found to reduce the toxicity of refugee-related tweets by nearly 30% and slightly lowered anti-refugee hate crimes in high-exposure municipalities without, however, chilling freedom of expression."

the crisis compared to communication via traditional media [215], and that coordinated and timely responses on social media can improve public perceptions of the organisation and outcomes in crisis situations [216]. Evidence from the misinformation and moderation literature further indicates that timing is critical: delayed interventions are often ineffective once content has reached large audiences [217], and trust in organizations tends to be higher when responses are issued more quickly [218]. Taken together, the evidence supports the case for a dedicated crisis protocol, such as those foreseen in the DSA and EDS, that would require expedited platform processes during crises (e.g., shorter timelines for moderation, fact-check flagging, takedowns of illegal content, and responses to complaints), while also strengthening the capacity of public authorities through enhanced data access and transparency (e.g., real-time monitoring dashboards, API access for authorised researchers and authorities, and structured reporting on actions taken, timing, and rationale to enable ex post evaluation). The same may apply during election times, where FIMI actions are ramped up by undemocratic actors.

Finally, professionalising the influencer industry, especially with methods to engage with facts and scientific evidence, could improve trustworthy content on sensitive topics. For example the EU Influencer Legal Hub and the EDS network of influencers could be used for this purpose.[22]

2.3.4 Individual-level measures

Individual-level interventions are directed at citizens within the fantasy-industrial complex. These measures aim to enhance users' ability to detect, scrutinise, and resist false or manipulative content but could be extended to escape fractured reality. Individual-level interventions on false and manipulative content have been subject to extensive empirical investigation. For instance, an international team of researchers recently developed a comprehensive toolbox of such interventions, drawing on findings from over 80 studies [219]. The toolbox distinguishes three main categories: nudges that shape behaviour at the point of engagement (e.g., accuracy prompts [220] or friction that slows impulsive sharing [221], or "nutrition labels" that can alert people to problematic content [222]); boosts and educational strategies that build competences (e.g., inoculation against manipulation tactics [223], or lateral reading to verify sources [224]); and refutation strategies that recalibrate beliefs once misinformation has been encountered (e.g., corrective debunking and fact-checking [225, 226], or source credibility indicators [227]). These approaches are complementary: some limit impulsive spread, others build durable skills, and yet others correct existing misconceptions. Importantly, their effectiveness is adaptive to the surrounding environment: lateral reading, for instance, is only useful if reliable external information is accessible on search engines, whereas accuracy nudges are most effective when users already possess the skills to recognize trustworthy content but benefit from reminders of the importance of accuracy.

A recent mega-study provides the clearest comparative evidence to date. It tested nine prominent interventions side by side under identical conditions, ranging from accuracy nudges and inoculation to debunking, fact-checking, and media literacy tips [228]. Almost all interventions improved both accuracy and sharing discernment. Media literacy tips were particularly effective, outperforming several others on both measures, while debunking and pre-emptive fact-checking also produced robust improvements. Another recent meta-analysis of 33 inoculation interventions showed their effectiveness [229].

[22] https://commission.europa.eu/live-work-travel-eu/consumer-rights-and-complaints/influencer-legal-hub_en

Various work strands in the EDS are promoting citizens' skills and media literacy, which are effective as shown above and also do not require governments to determine the "truth". Similarly, the EU democracy guide for citizens could contain the toolbox elements shown to be particularly effective as self-guided learning offers.

Beyond the toolbox, complementary approaches on the individual level seek to assist people in structuring their overall information environment. Strategies such as self-nudging and critical ignoring encourage individuals to consciously curate their information diets, for example by setting rules for when and where to check news, muting low-quality sources, or ignoring manipulative clickbait altogether [230, 39]. Enabling better self-nudging, for example by mandating platforms to provide tools like "one sec"[23] by default would enable citizens to be their own choice architect [231]. These tools do not directly target belief accuracy or sharing behaviour but instead empower users to manage their attention and reduce exposure to harmful content at the source. They could also be part of the EU democracy guide for citizens. Alternatively, the EU could make self-management tools mandatory for platforms to provide to its users for free and well visible, thus enabling citizens to take more control themselves.

Fact-checking

Fact-checking occupies an intermediate position between system- and individual-level responses. Among all interventions, fact-checking has consistently been shown to be effective in countering falsehoods and improving the accuracy of people's beliefs, even after a single exposure [232, 225], without, however, eliminating the effect of falsehoods altogether. Its effectiveness depends heavily on context, audience, and delivery format [233]. For instance, corrections tend to work better on health-related topics than on politically charged issues[234]. Format matters as well: affirmations of truth are often more persuasive than direct negations of falsehoods [235]. Fact-checking can also generate indirect effects: the prospect of being scrutinised has been shown to discourage individuals (e.g., politicians) from making false claims in the first place [236]. Fact-checking labels can also reduce the visibility and the sharing of false claims [237].

"Strategies such as self-nudging and critical ignoring encourage individuals to consciously curate their information diets."

Despite these advantages, fact-checking faces important limitations. Epistemologically, determining "truth" can be challenging, especially for misleading statements that are biased, selective, or taken out of context [238, 239]. Empirically, misinformation often spreads faster and further than corrections, meaning that many individuals never encounter the fact-check [240, 241]. Moreover, while experiments demonstrate robust belief-correction effects, these may not translate into changes in attitudes or behaviours [242, 243, 244, 245]. In real-world settings, exposure is also limited: in 2025, a global survey of 97,000 online users found that 25% of them would use a fact-checking website to verify content they suspect is false and 35% prefer an official source (e.g., government website) [7]. Likewise, during the 2016 US presidential election, 25% visited a fact-checking website, a number that decreased to 13% in the 2020 election [246]. Furthermore, due to inattention, low motivation and selective exposure, accurate corrections often do not reach voters [247]. Targeted exposure, making sure that everyone who engaged with flagged content also receives the correction, may be an option and is technologically possible.

In response to these challenges, new models of verification have emerged. Crowd-sourced approaches aim to harness collective judgment to assess and contextualise claims in real time, potentially increasing reach, speed, and perceived neutrality [248]. The most prominent of these is X's "Community Notes", which attaches context, written and rated by users, to potentially misleading posts. Early research suggests mixed outcomes: while notes can improve belief accuracy among those who read them [116], and reduce sharing of flagged content [249], they have limited effects on engagement with misinformation and face challenges

[23] https://one-sec.app/

of coverage, timeliness, and vulnerability to manipulation [250]. The strength of community notes lies less in comprehensive correction than in fostering shared standards of accountability and transparency, complementing expert-led fact-checking rather than replacing it.

Another development is the application of LLMs to automated fact-checking and debunking. Recent studies explore whether LLMs can generate timely, context-sensitive corrections at scale. While some findings suggest they can produce factually accurate and persuasive content, concerns remain about reliability, bias, and the ethical risks of delegating epistemic authority to automated systems [251]. However, the persuasive effectiveness of LLMs in laboratory experiments has been shown repeatedly [81, 252]. These systems may function best as supports for human fact-checkers, expanding capacity while retaining professional oversight. They also need to be safeguarded against political influence for their ideological slant.

The EDS proposes setting up a European Network of Fact Checkers, which would be in line with research presented in this section. Members of that network should design and execute fact-checks in line with the latest research [226, 104]. Additionally, the network alone will not be effective if its fact-checks are not widely disseminated, making it necessary to work with platforms to create a feasible and low-effort way for users to see and engage with fact-checks as well as giving enough resources to the network for dissemination (e.g., via journalists and influencers, but also feeding into AI training, or into community notes).

2.3.5 Offline solutions to online problems

Beyond digital system solutions, there should also be an increasing recognition that not all digital problems can be solved digitally. If all online engagement-based social media platforms facilitate the same pattern of toxic debate, then encouraging more “real life” interaction may be part of the solution. The decrease of social interaction offline, declining participation in social clubs and an increasing lack of activities that bridge groups of different political preferences, beliefs and backgrounds is a fundamental problem for democracy [253]. People who vote for radical parties have been shown to be more socially isolated, have low trust, and are less likely to be members of clubs and associations [254, 255, 256]. A recent Eurobarometer shows that this relationship holds especially on the political right. [257].

Therefore, a strategy is needed to raise awareness, increase public investment, encourage more participation in real-life social clubs and associations, align urban and rural infrastructure planning for more encounter and encourage companies to invest in activities that bring people together [258]. For example, a recent meta-analysis on public spaces shows that details matter when it comes to constructing encounters that promote social cohesion, such as accessibility, mixed land use, and presence of street furniture [259]. Importantly, these activities should be inclusive in terms of cultural background and political orientation to bridge divides. People accept and appreciate personalisation of entertainment, but only partially of politics [260], yet their actions guided by platform affordances drives them apart.

2.3.6 Countermeasures and building better platforms and spaces is not enough

Overall, the evidence suggests that no single countermeasure is sufficient on its own. System-level interventions such as regulation of algorithms can reduce exposure to harmful content, especially when targeting the small minority of superspreaders. Better facilitating debate and encouraging deliberation, both online and offline, is another necessary step to safeguard democratic discourse.

Individual-level strategies, from accuracy nudges and inoculation to media literacy, reliably improve information discernment, though their effects are modestFact-checking, provides one of the strongest correctives for false beliefs but has limited reach and requires more efforts by platforms to implement and disseminate them.

Experts are generally optimistic about the effectiveness of both individual-level measures (media literacy, fact-checking, content or source labelling) and system-level actions (platform design changes, algorithmic adjustments, and content moderation) [34, 261].

Nonetheless, all those interventions are likely dwarfed by the current engagement-based business model and the fantasy-industrial complex which it supports. The next chapter will show that this business model is at the root of problems arising from social media. A democratic information environment is unlikely to be achievable without reform of the underlying business model.

3

THE NEED FOR ALTERNATIVE BUSINESS MODELS AND USER AGENCY

KEY TAKEAWAYS

What?

- Harmful externalities—violence, polarisation, harassment, lost trust in institutions—occur off-platform. These social costs are not internalised into transactions between advertisers and platforms.

Why?

- Platform businesses based on engagement seek to capture user attention and optimise for time on platform rather than quality or value of information.
- Currently dominant platforms also curtail user agency by combining their services and providing users with only one gateway—namely a proprietary algorithm—to their content (e.g., newsfeeds).

How?

- Interoperability and in situ data rights, which provide users with choices of algorithms, present options to foster competition and innovation.
- Taxing digital advertising may encourage platforms toward a business model optimised for content quality rather than attention capture.
- A tax on certain types of content (e.g. disinformation) could hold platforms accountable for a flow rate of harms analogous to taxing polluting effluent of industrial firms.
- An innovative but currently still experimental solution could involve trading of information quality rights. These could decentralise the solution for harms such that no platform, powerful individual or government has the power to bias content in their favour.

3.1 WHAT?

> **"If something is free, you're not the customer — you're the product."**
>
> Bruce Schneier, in *Data and Goliath*

The previous chapter has shown that the negative consequences of online platforms on democracy, such as increased affective polarisation, the emergence of echo chambers and echo platforms, and the prevalence of mis- and disinformation, is mainly tied to their common business model.

The revenue logic of many online platforms is centred on advertising facilitated by users' behavioural data. Platforms, especially social media platforms and search engines, collect and analyse massive volumes of user interactions to sell targeted advertising, a model that has been described as "surveillance capitalism" [262] or the "attention economy" [263]. Because advertising revenues scale with engagement, platforms design their algorithms to maximise attention, virality, and time spent online. Engagement-based ranking tends to amplify sensational or polarising content, given its higher propensity to elicit reactions [52]. This creates incentives for political social-media entrepreneurs to act more extreme [264], even when citizens expressively do not want this [265].

Despite the original marketing of online platforms, they are not designed around connecting people and social sharing but are explicitly oriented toward entertainment [266]. This process may foster an overabundance of information—with the negative consequences outlined in Chapter 2—and a corresponding decrease in informational quality [267, 268]. Online platforms face a clear trade-off in content presentation between low-toxicity content with lower engagement and high-toxicity content with increased engagement that maximises profit [269]. When online platforms are not based on the engagement model (e.g., Reddit), however, exposure to diverse content is much more likely while differences between platforms with the same engagement model have minor impact [59].

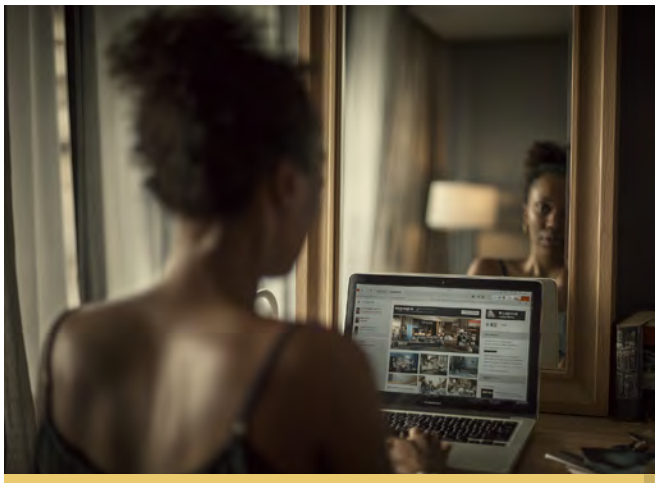
The dominant position of online platforms generates a number of market failures with consequences for democracy:

- *Market dominance and competition failures.* Social media firms are not just participants in the "marketplace of ideas"; they often are the marketplace [270]. Platforms are shaped fundamentally by network effects and scale economies. The value of the service increases with the number of users and their "friends", creating powerful feedback loops [271, 272]. A large user base attracts content creators and advertisers, which in turn attracts more users. Once entrenched, network effects generate high barriers to entry; even firms with considerable resources, such as Google, have failed to dislodge incumbents in social networking [273, 274]. In consequence, large platforms shape what many million Europeans see, yet face limited accountability for these editorial choices. At the same time, platforms soak up profits from traditional media companies. For example, the market dominance over advertising revenues can be seen in the expansion of Craigslist, which has been shown to lead to a reduction in staff of local newspapers and reduction in coverage of politics. This reduction of local media, in turn, resulted in a decline in newspaper readership and an associated increase in partisan voting and success of ideologically extreme candidates in US elections [275].

- *Externalities.* As shown in previous chapters, information disseminated online can cause significant off-platform damage—from election interference and incitement to violence from terrorist recruitment to undermining of institutions—without bearing these costs. While the DSA requires large platforms (with more than 10% of EU population as users) to conduct risk assessments and implement mitigation strategies (Art. 34 and Art. 35) [276], it does not create any liability for resulting societal harms affecting the resilience of democracy even if the causation between harm and action could be clearly shown. If a social media platform enabled violence, only the individuals involved can be held liable to the people injured. This is true even though the algorithms are magnifying the potential for such harms in society, as shown in the case of violence against refugees in the previous chapter [96]. Unlike traditional media, which face both editorial freedom and corresponding liability, social media platforms enjoy editorial control while externalizing consequences to society.
- *Curtailing user autonomy.* Users cannot meaningfully choose alternative content curation systems or escape algorithmic manipulation and choice architectures designed in the platform's interests [270]. Choices that voters, consumers, and users make almost always occur in the context of a choice architecture that defines key parameters of the decision context, including the number of available options, the representation of the options' attributes, the arrangement of the options, or the availability of a pre-selected (default) option.

Although choice architectures are indispensable in an information-rich environment, they can be abused for manipulation. "Dark patterns" are designs that coerce, steer, or deceive users into making unintended and potentially harmful decisions [277]. Dark patterns are banned by the DSA (Art. 25) but continue to be widespread [278, 279], pointing to the difficulty of detection and enforcement.

Social media platforms often employ these patterns and affordances (e.g. preselected defaults) to steer users onto their dominant feeds. For example, TikTok has been shown to steer users toward passive consumption, limit transparency about how curation is done, and discourage users from disabling data collection [280]. Vulnerable populations, such as young people with mental health difficulties, are exploited by personalisation to "trigger" their vulnerability. Users often blame themselves for failing to overcome those triggers [281].



"Vulnerable populations, such as young people with mental health difficulties, are exploited by personalisation to "trigger" their vulnerability."

These market failures present multiple interconnected problems for EU policymakers who wish to uphold democracy. The EU Charter of Fundamental Rights guarantees both freedom of expression (Art. 11) and protection from exploitation, crime, and discrimination (Art. 2 et seq and Art. 20 et seq). Yet, while the DSA prohibits targeting advertisements using sensitive personal data (Art. 26, para 3), targeting of minors (Art 28, para 2), and the political advertising regulation (EU 2024/900) requires transparency over political advertising,[24] it does not address the broader issue of algorithmic control over information flows that are having negative consequences for democracy.[25] Societal

[24] However, microtargeted political advertising—often assumed to be highly effective—is in practice loosely applied and exhibits limited but replicable impact [14, 282, 111].

[25] There are currently ongoing efforts to audit these algorithms by the European Centre for Algorithmic Transparency (ECAT), with several cases under investigation against X, TikTok and Facebook, see https://algorithmic-transparency.ec.europa.eu/index_en.

problems arising from this trade-off, then, are not "rough edges" to be ironed out but the natural consequences of a business model based on digital ads.

The three main issues that flow from the business model are (1) a disregard for information quality, (2) a high degree of (user generated) toxicity of content, and (3) a natural tendency to personalise content which facilitates the emergence of echo chambers and platforms. The DSA and the Digital Markets Act (DMA), are intended to recalibrate platform incentives. The DSA in this context requires platforms to address systemic risks, while the DMA focuses on contestability and fairness at the level of business-to-business interaction. Full enforcement has yet to play out, but addressing the three negative consequences of these business models is likely to prove challenging. EU Member States have expressed the desire for more effective implementation and enforcement especially for the protection of youth, for example expressed in the Jutland Declaration signed by 27 European governments.[26]

One particularly challenging issue may be that, even though the DSA-related e-Commerce Directive gives conditional liability for illegal content,[27] *editorial* liability for traditional print and broadcast remain higher than for platforms. While platforms are conduits for third party generated content — thus putting them on a different level of liability — their algorithms encourage certain types of content to emerge. The difference in liability creates an economic imbalance in their respective business models of presenting and moderating content, helping to tilt the playing field against high quality content generation in society. Adding to those concerns, Chapter 2 has shown that, increasingly, people are moving to echo platforms, creating the problem that individual platforms alone may not be able to counter the negative effects of personalisation on fractured realities, requiring policies beyond individual platform enforcement. Finally, the toxicity found on platforms is creating real-world harms, but some occur off platform, making it difficult to make causal inference on who is responsible, especially without sufficient access to platform data by researchers. Thus, the strategic goal should be to encourage business models for information spaces that ensure information integrity.

3.2 WHY?

3.2.1 Impact on Societal Resilience and EU competitiveness

Further addressing the market position of large platforms is essential for democratic resilience, but interestingly also for economic competitiveness of the EU. As these platforms are locking in users through network effects, addictiveness and promotion of engaging content irrespective of quality, they are providing essential avenues for mis- and disinformation not only about politically charged topics, but also about health, product scams, or environmental issues, thereby promoting a general decline of health and human and social capital.

These negative externalities will ultimately present costs for the EU: less healthy citizens, higher health care costs, lower labour productivity, more environmental destruction. For example, emerging evidence suggests that social media exposure to drinking does not simply reflect people's behaviour, but causes it, thereby potentially increasing risky drinking behaviour [283, 284, 285]. Another concrete estimate for the costs comes from a recent modelling study of the COVID-19 pandemic in Canada provided an estimate for the societal cost of COVID-19 misinformation

[26] https://danish-presidency.consilium.europa.eu/en/news/eu-ministers-united-minors-must-be-prote cted-better-online/
[27] https://digital-strategy.ec.europa.eu/en/policies/e-commerce-directive

[286]. The study estimated that misinformation contributed to vaccine hesitancy in over 2.3 million people in Canada during the vaccine rollout in 2021, estimating that immediate vaccination would have led to 28% fewer hospitalisations, and 2,800 fewer deaths, and that the costs of misinformation-induced hospitalisation were estimated at CAD$300 million. Finally, on the macro-economic level, the proliferation of misinformation has been estimated (in a pre-print) to lead to macroeconomic uncertainty, increased unemployment, and lower industrial production [287].

In a global competition, in which some countries tolerate immersion of their own population in low-quality information, those countries that decide to promote information integrity may arguably become more competitive in the long-run.

3.2.2 Are engagement based social media platforms worth having?

Applying the economic lens to the product of social media, it is important to ask if social media is a "good" (i.e., a product people want to have), or a "bad" (i.e., something people would pay to avoid). Recent studies have examined this question by asking people to name their price for discontinuing social media usage [288, 135, 289]. The studies show that people require compensation to turn off social media—in line with the argument that social media is providing a valuable good to them. However, when asked about the price tag for turning off social media for people including their friends, users were actually willing to pay money for this [289]. These results suggest that social media locks people in, e.g., for fear of missing out, but that many people might be happier without it—which would make social media a bad, not a good [288].

"These results suggest that social media locks people in, e.g., for fear of missing out, but that many people might be happier without it—which would make social media a bad, not a good."

"It has been estimated that about 31% of the use of social media can be pinned down to habit formation that exhibits some (but not all) attributes of addiction."

Multiple lines of research support the idea of social media on average being more in line with a "bad" than a "good", even though it likely provides positive effects for some people:

- It has been estimated that about 31% of the use of social media can be pinned down to habit formation [290] that exhibits some (but not all; [291]) attributes of addiction.
- People's average time on a platform is relatively unresponsive to advertising load, showing how the environments seamlessly blend purchased and organic content [292, 293].
- Restricting access to apps on people's phones (by randomly blocking e.g. Instagram or YouTube) increases non-phone time use by a wide margin, suggesting "that these applications have considerable (short-run) power over participants' time" [294, p.2].
- Features such as autoplay, "likes", read receipts, news feed algorithms, and "freemium" game mechanics encourage continued use and spending [295].
- Other reviews have defined addictive designs in higher-order and cognitive terms, including reinforcement schedules, behavioural cues, overvaluation processes, cognitive interference, partial goal fulfilment, and biased decision architectures. These features exploit psychological mechanisms that drive habit formation, distorted beliefs, and reduced self-control, thereby encouraging excessive engagement [296].
- These habit-forming design elements have been linked to negative outcomes, including loneliness, body image concerns, and excessive smartphone use, though such designs can also be used to promote positive behaviours such as physical activity [297, 298]. The business models underlying the

Emerging threats: Manosphere and gaming platforms

There are two new arenas that constitute emerging threats. Although they are under-researched and not fully understood, the potential for harm suggests they should not be ignored.

The first arena is the fracturing of realities by gender related to the "manosphere" and a female counter-movement tied to an ideological gender divide among young people, at least in some countries [301, 302]. Recent evidence from 30 countries shows a new emerging gulf between young men and women on the question of gender equality, a stark difference that has never before been found in surveys to such a degree. Misogyny and gender-based political violence is already a problem for political participation in Europe and worsens in time of crisis [303]. Evidence of the broader causal impacts of the "manosphere" is still insufficient. Research is beginning to map the online presence and impact on platforms that dehumanise women and present men as victims [304, 305, 306]. Other evidence suggests that the manosphere may be seen more as a grift [307], where online influencers generate the perception of threat, and position themselves and their products as the solution to it exploiting insecurities in men.

The second emerging problem area involves virtual worlds and adjacent gamer platforms that in principle follow somewhat similar business models but whose core product is usually not information but games and entertainment. Newer research is investigating the potential of these platforms as room for growing extremism and even as recruitment arenas for extremist organisations [308]. However, correlational evidence currently does not associate the intensity of gaming with higher rates of participation in, or positive views towards, violent extremism or terrorism [309, 310, 311]. Nevertheless, most communication happening within games is not available for public scrutiny due to its oral nature, thus potentially creating a blind spot for researchers and public authorities.

design of addictive apps are the same as those that drive the increased negative impacts for democracy through real-world isolation and loneliness, coupled with excessive exposure to extremist and toxic content online.

Welfare costs also arise indirectly, through selling of social media accounts to enable malign influence operations [299]. The platforms' ability to detect and counter such fraudulent activities has shown only minimal improvements since the adoption of the DSA [300].

3.3 HOW?

There is a manifest need for new business models that avoid the market failures just outlined. Many of the phenomena surveyed thus far apply to any reasonably-sized online platform, not only to large platforms identified in the DSA and DMA. The remainder of this chapter outlines three alternative potential interventions together with supporting behavioural evidence.

3.3.1 Interoperability and *in situ* data rights

Interoperability refers to the ability of different information systems or platforms to seamlessly exchange or integrate data.[28] For example, a form of interoperability might be achieved if one's followers on one platform could automatically become "friends" on another one. Requiring interoperability between social networking platforms will increase competition by breaking down the network effects that currently entrench incumbents.Because using multiple networks is costly in time and effort, users tend to "single-home," remaining on the dominant platform even if dissatisfied. This dynamic deters entry of competitors, because rivals cannot persuade users to replicate their networks elsewhere [289]. Mandating greater interoperability—whether vertical (e.g., third-party curators of content on top of incumbent infrastructure) or horizontal (rival platform infrastructure connecting via standardized APIs)—would lower these barriers, enabling new services to compete on dimensions such as content curation, algorithmic design, and business models [312, 313].

[28] Chapter VIII of the Data Act already demands interoperability to some extent.

The Commission is responsible for evaluating whether interoperability requirements spelled out in Art. 7 DMA should also apply to online social networking services according to article 53(2) DMA. Updating the DMA in this way, interoperability could replicate the pro-competitive effects seen historically in telephony and email, where open standards allowed multiple providers to coexist. New entrants could offer differentiated services and not only content visibility across platforms—such as enhanced privacy, prosocial algorithms, or subscription-based models—while still allowing users to connect with friends on dominant platforms.

The introduction of *in situ* data rights could extend the benefits of interoperability. An *in situ* data right entitles users to import an algorithm of their choice into infrastructure where that user's data resides [314]. This entitlement shifts the point of control to the user and away from the platform, thus going beyond mere interoperability. Under interoperability, the negotiation effectively remains between the providers of two interoperable standards—rival to rival—bypassing the user who is most concerned and whose data is providing the value. Under *in situ* data rights, every potential content curator, including the native provider, would need to reward the user in order to secure that user's permission for access.

Both ideas are, in principle, aligned with the idea of the Digital Omnibus[29], which seeks to increase EU competitiveness by giving a competitive advantage to responsible businesses. To truly reduce the costs for businesses to comply, the EU could develop standardized APIs and protocols that all social networking platforms would have to adhere to, reducing costs for compliance and increasing the ease and profitability for European companies to use them creating new innovation and growth potential. Scaling in situ data rights would require clear guidelines on how users can exercise these rights across different platforms.

3.3.2 Shifting business models through a progressive digital advertising tax

Proponents of a digital advertising tax have the explicit goal of pushing platforms away from attention addiction and toward subscription-based models, "where revenues are instead dependent on the sustained quality of content and user experience" [315, 316]. For example, two Nobel laureates, Daron Acemoglu and Simon Johnson, have proposed a 50% digital ad tax on revenues above $500 million [315]. A tax on digital advertising—particularly when designed progressively to scale with firm size—has several advantages over traditional content moderation or competition remedies:

1. *Realignment of incentives.* An advertising tax would shift platform incentives away from surveillance-driven, engagement-maximizing models toward subscription-based or non-advertising alternatives.
2. *Structural competition benefits.* A progressive tax—where larger firms pay a higher marginal rate—discourages excessive concentration by increasing the cost of scale.
3. *Content-neutrality and legal feasibility.* An ad tax is not premised on the nature of the content itself. It is business-model-based rather than speech-based, which helps insulate it from legal and political challenges grounded in freedom of expression.
4. *Regulatory efficiency.* An ad tax is less costly than regulating algorithms or content moderation.

[29] https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal

However, there are some limitations to consider with the subscription model, including a reduced ability of platforms to create value from matching content with user preferences [274] and questions about equality, as not all users may be able to pay for all app subscriptions.

To better internalise social harms, legislation can refine the advertising tax to account for the *concentration* of harmful or problematic content in a platform's output—akin to taxing the toxicity of industrial effluent rather than its sheer volume. This modification aligns with well-established environmental principles and can be operationalised without engaging in speech regulation, although great care must be taken to circumscribe what is considered "harmful".[30] The underlying idea is to follow financial auditing principles of not assessing all individual posts, but rather to sample a representative flow of content from the platforms and test them for the prevalence of disinformation, harmful content, or legal violations. This approach could be in line with DSA principles and ease the efforts to categorize platforms as either being in violation or compliance of the DSA. The Centre on Democratic Resilience announced in the EDS could help align Member States' interests along the involvement of the European Data Protection Board for oversight and compliance.

© Image: iStock

"To better internalise social harms, legislation can refine the advertising tax to account for the *concentration* of harmful or problematic content in a platform's output."

3.3.3 Markets for speech

The final intervention involves classical free-market approaches for speech in a decentralised setting [317, 318]. Applied to digital speech, speakers could voluntarily signal content quality on attention markets (e.g., being free of disinformation) by staking resources (reputation, or monetary assets) to attest to their claims. Such a mechanism can be built from existing infrastructure, using only advertising and crowdsourcing. As with an ad, the speaker would place the stake with a platform prior to message distribution, and it would increase in the size of the audience reached (e.g. adding an amount to the advertising cost per impression). By voluntarily risking more, the source distinguishes its content from standard ad content as well as any unstaked content. The stake, however, returns to the speaker if no crowd-judging system finds the facts are false. Such systems already exist as "Community Notes" on X, Facebook and Instagram, and as "Footnotes" on TikTok.

Listeners or regulators could then filter or challenge false content, with verification of the claims handled by decentralised peer juries or politically balanced citizen panels [319, 320], which have been shown to perform as well as fact-checkers when adjudicating content [321, 322, 248]. This approach balances speaker rights (the ability to reach audiences with free speech) and listener rights [323], while preserving fundamental rights. The surety pledged by speakers can be used to amplify claims (e.g., algorithmic promotion) whereas unstaked content could be de-emphasized. When staked content is successfully challenged, the challengers win the stake and the amplification ceases.

[30] One recent initiative in this dimension is the idea of a "polarization footprint", a metric for the degree of polarization stemming from platform use that could be used as a measure for taxation. See https://howtobuildup.medium.com/the-polarization-footprint-part-1-8a85fdaf35b1.

This should create an equilibrium, where truthful speakers signal credibility (by pledging a surety) while false claims are penalised (speakers lose their surety), thus ensuring that higher quality information has an advantage—in stark contrast to the current business models. The system would need to be calibrated to ensure that more financially well-off actors cannot easily absorb costs from putting out many wrong claims.

The approach has been tested in several experiments with encouraging results [324, 321, 320]. In one study in which true and false headlines were presented to participants and they had to decide whether or not to share them [324]. In a condition where "interestingness" was rewarded monetarily, people shared more false than true content. In another condition, participants could optionally choose to warrant the accuracy of a headline before sharing it, in line with the above-mentioned principles. In this condition, considerably more true headlines than false headlines were shared. These results align well with the observation that around 80% of users are concerned about their reputation when they share false information, suggesting that the majority of users would be sensitive to market signals [325].

More recently, the market-based approach has been extended to AI agents with similarly promising results [326]. The roll-out of this approach still needs to be fully tested, for example, through citizen engagement exercises to collect in-depth views on such a change to the information system. As a first step, the EU could invest in an exploratory voluntary regime to explore feasibility and scalability.

3.3.4 Solutions exist but require full EU digital autonomy

The preceding chapter identified the problems facing the European information space and proposed a range of solutions. The present chapter has aimed at the core issue, which is the engagement-based business models of the platforms and how their reform may radically improve the European information landscape.

There is, however, one practical impediment to the implementation of those solutions, which arises from the fact that almost all large platforms are owned and controlled by entities outside EU jurisdiction. Although this does not preclude EU regulatory actions, it makes control of the information space more difficult. The next chapter will show that digital sovereignty and autonomy for the EU represents an essential lever to ensure democratic information spaces.

"Around 80% of users are concerned about their reputation when they share false information, suggesting that the majority of users would be sensitive to market signals."

4

THE NEED FOR EU DIGITAL SOVEREIGNTY AND AUTONOMY

KEY TAKEAWAYS

What?

- The EU is a global leader in digital regulation and legislation but most digital infrastructure is controlled by entities outside the EU.
- This dependency enables several risks for the EU, such as direct and indirect interference, the threat of going dark, a highway of influence for foreign actors and transparent citizens.

Why?

- Although foreign entities must comply with EU law, in practice, compliance is imperfect or slow and at least some digital platforms have, arguably, interfered in European elections and imperilled the integrity of public discourse.
- Foreign threat actors have engaged in hybrid attacks on the EU using this digital infrastructure, including during the German federal election in 2025.

How?

- The path to digital sovereignty is complex but several steps are possible and necessary for democracy, such as the support of decentralised or a mix of decentralised and centralised architectures, aligned public procurement and communication and development of European AI, as well as alternative software and hardware.

4.1 WHAT?

The European Union is a global leader in digital legislation. Democracies require high-quality information to harvest collective intelligence, negotiate public values and hold politicians to account. At present, however, the most important digital information intermediaries in Europe are in a few foreign hands, based mainly in the US and China. Europeans' information diet is therefore shaped by actors who may not share Europe's democratic values and by algorithms that are controlled offshore, creating the additional enforcement burden of holding corporate entities to account abroad.

Democracy is under threat worldwide. Democracy is a set of core values, rights and freedoms, mechanisms upholding those rights, the rule of law, and a vibrant culture of citizen participation and functioning public discourse [327, 328, 329, 330]. Many of those elements are under attack and powerful countries play a central role in this process [329, 331]. The role of the EU in defending its democracy is therefore more important now than ever.

There are five levels of threat to EU digital sovereignty and autonomy:

- *Direct interference:* Large platforms have the ability—which they have sometimes exercised—to directly interfere in the integrity of political discourse and democratic decision-making by algorithmically shaping what content is produced, seen and shared.
- *Indirect interference:* Large platforms create the conditions for information overload, toxic and sensationalist debates, and low-quality information to out-compete high quality information and moderate views, thus polluting public democratic discourse.
- *Threat of going dark:* Platforms, perhaps at the behest of foreign governments, have the de facto power to turn off a large part of the existing public discourse in the EU by curtailing access to their platforms.
- *Highway of influence for foreign actors:* Digitalisation and seamless international integration enables users and (state) actors from outside the EU to reach EU citizens unhindered.
- *Transparent citizens:* Access to personal data held on platforms by foreign agents for influence operations even off-line.

All of these threats are only possible through the digitalisation of the information space and endanger the resilience of EU democracy. The increasing global disharmony and the global wave of autocratisation create a pressing need for policy action, not only in the regulation of digital platforms, but in a broader need to rethink the integrity of the information space in Europe. This challenge has been recognised by the European Commission in the State

Digital sovereignty or autonomy?

European Commission President Ursula von der Leyen expressed digital sovereignty as the "capability to make its [Europe's] own choices, based on its own values, respecting its own rules".[a] By this criterion, the EU is digitally sovereign already when it comes to passing laws that regulate the digital space, and indeed the EU is arguably the most powerful global actor in the field of digital regulation [333]. A new JRC brief defines "Digital Sovereignty" on four layers: (1) People, (2) Digital Product Markets, (3) Digital Infrastructure, Software & Data, and (4) Digital Governance [334].

An alternative term is "digital autonomy" which subsumes more than legal sovereignty and also includes European control of software, hardware, and other infrastructure such that it is beyond the reach of non-European governments or corporate bodies. This broader definition is more in line with the conceptualisation of digital sovereignty as "a provisional assemblage of technical, legal, and cultural components that stabilize sovereign claims in the digital realm" [335].

Autonomy also means that non-European actors cannot isolate Europe from the rest of the world through denial of service or other anti-competitive measures.

[a] https://ec.europa.eu/commission/presscorner/detail/en/ac_20_260

of the Digital Decade report (COM(2025)290), and calls for digital sovereignty have been issued by the European Parliament[31] and go hand in hand with the need to protect European competitiveness.[32] Scholars agree with the need for a digital public space that follows European values [332], and this agreement is echoed in an expert survey conducted for this report that is reported in the Conclusion (Chapter 5).

4.2 WHY?

4.2.1 Foreign intrusions into EU sovereignty

Although non-European corporations must legally comply with European laws if they operate within the EU, the increasing global discord has created additional challenges as exemplified by the linking of digital regulation with defence by non-European political actors [336]. Additional problems arise if the laws applicable in different countries or different EU Member States impose conflicting obligations on the responsible actors [337].

EU digital sovereignty cannot be discussed without examining the role of platforms and their owners, not only as core members of the fantasy-industrial complex introduced in Chapter 2, but also as agents that directly impact European democracy.

To illustrate, the platform X/Twitter has been shown to take a direct role in shaping political discourse in Europe and the US after it changed ownership in October 2022. A longitudinal analysis of information quality on the platform based on more than 18 billion tweets revealed a distinct breakpoint in information quality from the change in ownership onward. Whereas information quality, measured by NewsGuard scores, had gradually increased on the platform in the eight months preceding the acquisition, a sharp decline ensued in the following five months [338]. This pattern has been independently confirmed by another team of researchers [339].

In a similar fashion, a new study (not yet peer-reviewed) showns that the Facebook algorithm decreased news

“X/Twitter has been shown to take a direct role in shaping political discourse in Europe and the US after it changed ownership in October 2022.”

© Image: iStock

visibility between 2021 - 2024 [340]. Reactions to news declined by 78% during that time, with a parallel increase in reactions to non-news pages, indicating targeted suppression of news visibility. When Facebook ended the down-ranking of political content in January 2025, engagement with news again increased but only for low- and medium-quality outlets [340]. Taken with the evidence that platform algorithms increase toxicity and create fractured realities (Chapter 2), this decrease in information quality will likely have a direct effect on the ability of the public and individual citizens to make informed decisions.

Beyond the effect of the quality and diversity of information as a challenge for democratic discourse, an even more important aspect for sovereignty may be the tilting of the playing field in one specific political direction. In Europe, X has been implicated in boosting far-right content, and French prosecutors launched a criminal investigation into X in

[31] https://www.europarl.europa.eu/news/en/agenda/briefing/2025-10-06/3/promoting-and-protecting-d igital-sovereignty-in-the-eu
[32] https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en

July 2025 over allegations that X manipulated its algorithm for the purposes of "foreign interference" in elections.[33] X has so far refused to comply with prosecutors' requests for data, claiming the prosecution is politically motivated [341]. During the 2025 German election, the owner of X used his platform to interview and actively promote the leader of a far-right party [342]. An analysis by the Potsdamer Social Media Monitor for the German Bertelsman Foundation (not peer-reviewed) shows that during the German Election 2025, the AfD received the most visibility on X and TikTok (e.g. 50% of all suggested content was AfD related, relative to 15% for the second most suggested party CDU/CSU). The authors claim that the outsized exposure cannot be explained by social media strategies (such as using emotions) which they found to be common across parties [343]. These findings were replicated in another study which showed that the X algorithm disproportionately amplified content by extreme parties [103].

These data and examples reveal the power of corporations over the information diet of the European (and global) population, and by implication the political system as a whole.

4.2.2 Online platforms as part of the hybrid threat agenda

The actions of big platforms are not only of concern in their own right but may have additional downstream consequences. Digital technologies are now used by governments

[33] https://www.politico.eu/article/france-opens-criminal-probe-into-x-for-algorithm-manipulation/

Digital legislation and freedom of expression

Existing EU legislation has sought to balance competing demands for freedom of expression or speech and protection against harms such as child exploitation. For example, the DSA empowers users to object to content moderation decisions made by platforms, thus protecting the public against arbitrary platform decisions defending freedom of expression in the EU. Since the DSA came into force, 30% of 165 million content moderation decisions that users appealed through the platforms' internal mechanisms have been reversed and the platforms restored the content.[a]

At the same time, the DSA mandates that platforms take down illegal content, such as scams and markets for illegal goods. This necessarily creates a heterogeneous enforcement landscape and it is possible that an expansive definition of "illegal content" may lead to potential encroachment on freedom of expression [344].

The DSA requires platforms to file annual systemic risk assessments of their operations and how they might impact democracy, especially with regard to the exercise of fundamental rights and the public health and security of European citizens. The risk assessments reveal that the platforms are aware of the implications of their algorithms, with Bing, X, Snapchat, and TikTok all highlighting the risks of echo chambers, illegal content and hate speech in their reports.[b]

Overall, survey evidence from Europe and the United States consistently shows majority approval for platform interventions against misinformation and harmful content [204, 345, 206]. Approval is high even for restrictive measures such as content removal and account suspensions crosses ideological lines, though support is stronger among left-leaning respondents. The consistent finding is that citizens are not categorically opposed to speech restrictions online; rather, they support interventions that are seen to be effective, proportionate, and fairly implemented. This aligns with expert views that emphasize transparency, contestability, and due process rather than categorical bans, helping to ensure that content restrictions are proportionate to the potential harms [346, 347, 348].

[a] https://digital-strategy.ec.europa.eu/news/two-years-digital-services-act-allows-50-million-content-moderation-decisions-platforms-be-reversed

[b] https://cdn.table.media/assets/europe/first_article_352_dsa_report_on_systemic_risks_and_mitigations_final_ddxkzxhwga8vftvj3unr0mgkwqvk_121707.pdf

and non-state actors within and from outside the EU to repress rights and freedoms of citizens [349]. Some scholars suggest that society has entered the third age of digital authoritarianism [350], where the online information space is actively used by authoritarian forces to shape the narrative and public discussion in their favour [349]. In so far as few social media platforms and companies dominate the information environment, as well as the advertising market that thrives on surveillance capitalism, this dependence is not only an issue for EU competitiveness but also for its security because sensitive data about EU users is often available to those actors [351].

Hostile state and non-state actors try to undermine democracies in the European Union and elsewhere by exploiting weaknesses in a wide range of domains, with the digital dimension being essential [352, 353]. Social media platforms and, increasingly, AI Chatbots [354] allow new means to sap democratic societies' stability by providing digital highways for foreign and domestic information manipulation and interference. The strategies include the strategic use of fake social media accounts and social media algorithms to evoke nostalgia and "anti-elite" sentiment, thereby preparing the ground for further public opinion shifts [355].

Hybrid threats and the German federal election in 2025

A recent analysis of FIMI in the 2025 German election identified most hybrid strategies [356]:

- focus on divisive issues (e.g., migration, Ukraine)
- involve widespread use of AI-generated material (e.g., deepfakes) and impersonation of institutions to capture credibility
- include misconduct by local actors (e.g., by far-right party)
- involve foreign state influence (e.g., RT maintained influence despite sanctions)
- weak or lacking enforcement by platforms and regulators

Platforms are carriers of other information threats for democratic resilience. External actors, such as coordinated propaganda networks, can manipulate these same systems from the outside. By engaging in what is often called "coordinated inauthentic behaviour," they create and manage multiple fake or semi-automated accounts that act in a synchronized way to simulate genuine mass engagement [357]. This artificial activity can deceive algorithms into amplifying the targeted content, inflating its visibility and creating the illusion of widespread support. (See Chapter 2 for more detail.)

Empirical evidence suggests that, thus far, the direct influence of coordinated inauthentic accounts remains limited in scale [358]. However, evidence collected by several organisations on the Romanian elections in 2024/25 suggests foreign actors can blend active FIMI operations via popular non-EU platforms (Facebook, TikTok) by exploiting the algorithms' tendency for emotional and toxic content, the platforms' negligence in applying the laws on political advertising and transparency, bot networks, and hired influencers to harness real and imagined grievances of citizens.[34] Foreign and coordinated interference thus warrants careful monitoring and capabilities [359], not only from the perspective of fighting disinformation, but even more so because of the lack of platforms' compliance and their algorithms' enabling of campaigns against established democratic institutions.

4.2.3 The need for action now

The goal of digital autonomy is becoming more pressing.

AI is becoming increasingly integral to online media and the way citizens get information about politics and public administration (see also Chapter 2). While AI has been shown to potentially have positive effects for political knowledge [360, 361], and satisfaction with public administration [362], reliance on non-European-owned companies may reduce democratic resilience because of their established political biases and their susceptibility to information manipulation and exploitation due to foreign control [363]. To illustrate, in November 2023, the platform X integrated an AI "chatbot" known as Grok. Grok has been reported to communicate antisemitic tropes, praise Adolf Hitler, and

[34] https://edmo.eu/blog/algorithmic-influence-on-elections-insights-from-romanias-case-study/
https://www.sgdsn.gouv.fr/files/2025-02/20250204_VIGINUM_Rapport%20public_Elections_Roumanie_r isques_France_ROU.pdf
https://tinyurl.com/bp6kwwv7

question the extent of the Holocaust [364]. Furthermore, in October 2025, the same company developing Grok published the "Grokipedia", described as an AI-based alternative to Wikipedia without "wokeness". Grokipedia is reported to have a strong political slant to the far right [365, 366].

This does not mean that AI use for information is per se negative. For example, a recent preprint showed that fact-checks elicited on X by querying two alternative LLMs, Grok and Perplexity, are of sufficient accuracy to be valuable, even though neither model matches performance of professional fact checkers [367]. A further strength of AI can be to curate large amounts of information to the public, as is done in Grokipedia, which goes beyond what humans may be able to provide in a reasonable time frames. However, AI replaces deliberation with automation[35], which becomes a problem if it is in the hand of only one person or corporation and not responsive to counterclaims. As mentioned in Chapter 2, the digital world is an engineered space, and the system will only be truly democratic and efficient if the large-scale AI content suggestions are embedded in democratic deliberation for verification.

4.3 HOW?

Achieving digital autonomy is a massive undertaking that goes far beyond a single silver bullet, requiring investment in a multitude of physical and digital infrastructures. This infrastructure may be largely invisible to the public but is currently not under European control even if installations are nominally based in the EU. A lot of digital traffic runs through the US [368, 369], often giving the US technical and legal authority over data regardless of data storage location with the US CLOUD act (Clarifying Lawful Overseas Use of Data Act; H.R. 4943) and the Foreign Intelligence Surveillance Act [370, 371]. To illustrate, in 2025, Microsoft disabled the email accounts of the chief prosecutor of the International Criminal Court (ICC), Karim Khan, and his staff in connection with US government sanctions against the ICC.[36] Thus, products sold by US-based entities, even if contracts stipulate sovereignty, still fall under jurisdiction outside of the EU that, de facto, overrides EU legislation (e.g., GDPR). In response, the ICC has switched to the Open source OpenDesk, a move improving their digital sovereignty.[37]

A careful unpacking of the high-level concept of digital sovereignty and autonomy, including discussions about the expected changes in power-relationships, is required in the near future [334]. There is currently a risk that an increased push for digital sovereignty at the technology layer may result in a missed opportunity for a better democratic framework and might replicate a European version of the same problems, especially if business models are not aligned with democracy.

4.3.1 European digital infrastructure

There are several partially competing and partially complementary ideas developed for achieving digital autonomy for the EU while also safeguarding democratic debate. One example is the EuroStack initiative: an initiative for Europe on digital sovereignty and a larger role in the global digital economy in line with democratic principles.[38] Other approaches in the EU include the Digital Commons European Digital Infrastructure Consortium (DC-EDIC), a new

[35] https://www.reneediresta.com/source-wars-and-bespoke-realities-wikipedia-grokipedia-and-the-b attle-for-truth/
[36] https://www.computerweekly.com/opinion/Microsofts-ICC-email-block-reignites-European-data-sov ereignty-concerns
[37] https://interoperable-europe.ec.europa.eu/collection/open-source-observatory-osor/news/interna tional-criminal-court-invests-open-inn frastructure
[38] https://www.euro-stack.info/docs/EuroStack_2025.pdf

"The push for digital sovereignty should exploit the opportunity for new business models that are more aligned with democracy."

instrument to jointly develop, deploy and operate cross-border digital infrastructures with dedicated governance, like the CitiVERSE[39] This initiative is underpinned by a layered model that acknowledges the entire value chain of digital technologies, from the extraction of critical raw materials, via chip production and infrastructure development, all the way to sophisticated applications of AI for the common good. By putting in place and using a "digital stack" based on European rights, principles and values, and using competences within the continent, the EuroStack initiative seeks the development of technologies that "do no harm and fundamentally strengthen democracy".

Additionally, although the inclusion of ethical principles, and values in model and system design is a commendable goal, those values can differ considerably between different actors—as exemplified by the recently published American AI Action Plan,[40] as compared to European AI policy,[41] requiring more work to better align these principles (see for example the discussion on "openness" in the context of AI in [372]).

It also matters how new systems are implemented and scrutinised; for example, calling for Open Source software does not define the governance structure including the involvement of stakeholders in software development and deployment in a democratic way. Research on inclusive data governance does provide parts of the required insights [373], but there is a lack of systemic approaches (building on different models of democracy) to ensure true human-centricity, as well as a lack of economic incentives for widespread adoption. For example., the EU Data Governance Act[42] and Data Union Strategy[43] are designed to promote individual data empowerment, build trust and stimulate innovation in this field.

But above all, the business model will still influence technology development since profit maximisation will push technology in interaction with human cognition to deliver what is optimised. In the attention economy, this is attention, with all the negative side effects mentioned in the previous chapters. To avoid those negative effects, sovereign solutions also require a business model change.

Public administration and governance play an important role in the progress towards EU digital resilience. The public sector should demonstrate leadership by example, aligning institutional practices with the EU's vision for technological autonomy by investing in development and use of sovereign technologies.

Simultaneously, public procurement needs to facilitate this new way of running public administration and governments and favour EU-owned systems and infrastructures. Public procurement represents €2 trillion (around 13.6% of GDP)[44]—an important lever in advancing digital autonomy. This might include centralised purchases and solution provisions for re-use (at higher levels of public administration), open-source, interoperability, but also the clustering of purchases on the same level of administration (within or across borders). One such idea could be the re-use of the Citizens' Engagement Platform used for the Conference on the Future of Europe (CoFE)[45].

[39] https://digital-strategy.ec.europa.eu/en/news/commission-launch-digital-commons-edic-support-s overeign-european-digital-infrastruce ture-and
[40] https://www.ai.gov/action-plan
[41] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[42] https://digital-strategy.ec.europa.eu/en/policies/data-governance-act
[43] https://digital-strategy.ec.europa.eu/en/policies/data-union
[44] https://single-market-economy.ec.europa.eu/single-market/public-procurement/digital-procurement/public-procurement-data-space-ppds_en
[45] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democ racy/conference-future-europe_en

Communicating (with and without digital tools) in a timely manner and offering possibilities for engagement are central to such novel approaches to re-enforce democratic governance processes [374].

4.3.2 Supporting decentralised information architectures

At present, many public institutions, including at EU level, are primarily active on the same social media platforms that do not seem to be aligned with EU values and are vulnerable to coordinated misinformation campaigns. This creates an undesirable lock-in effect for the public who want to legitimately know what their governments are doing. The development of healthier public spaces should be supported by ensuring that public bodies are present on other platforms as well. For example, the EU could invest in tools that allow public communicators at all governance levels to easily cross-post information on all social media platforms (e.g., *Fedica.com*). Whereas currently, communicators often focus on the biggest platforms based on concerns of reach[46] and efficiency, having low cost and easy-to-use cross posting would reduce lock-in effects for users, and at the same time give a boost for start-ups with different business and governance models by posting valuable content. This measure would also counteract the increasing reality of echo-platforms, as discussed in Chapter 2.

"At present, many public institutions, including at EU level, are primarily active on the same social media platforms that do not seem to be aligned with EU values and are vulnerable to coordinated misinformation campaigns."

Furthermore, supporting decentralised, protocol-based social media such as the ATmosphere (used inter alia by BlueSky) and the FediVerse (used inter alia by Mastodon) may help reduce the impact of foreign control and biases in their algorithms. These decentralised approaches are not just new platforms but rather apps running on open-source protocols that, by design, cannot be fully shut down or centrally controlled. Moreover, apps on these protocols cannot achieve (full) social lock-in, thereby giving users greater sovereignty and incentivising app developers to align with their users' well-being. They do not just replace incumbent businesses with new ones, but they also change the way social media environments could be governed. The most important part of these protocols is that they are permission-less; there is not one owner of the content created and thus no dominant owner that may have to be regulated. Because of this feature, decentralised platforms circumvent the dominant market power of big platforms which combine their various services with providing content, thus preventing competition over "middleware". For example, middleware can consist of independent fact-checking services for content views, the choice over tailored personal feeds, or filters adjusting shopping website search results to prefer domestically produced, sustainable or human-rights certified products [375].

These alternative protocol-based social media apps may offer new innovation spaces for EU entrepreneurs and innovators for middleware and interoperability (see also Chapter 3), but it will also require careful management of potential threats to data security and privacy [376]. However, existing platforms hold natural monopolies due to multiple network effects (see Chapter 3). Monetisation and incentive structures for participants in decentralised systems are complex. They require support and guidance to grow alternatives and create positive momentum for switching for citizens.

One such effort involves Eurosky[47] which "is a public-interest infrastructure project that puts control in the hands of users, businesses, and European society,"[48] based on the realization that "a sovereign continent cannot outsource its public communications layer."[49] Eurosky seeks to combine

[46] This is even more important given that impression numbers on some of the big platforms are artificially boosted by those same companies, see https://www.cnbc.com/2021/02/18/facebook-knew-ad-metrics-were-inflate d-but-ignored-the-problem-lawsuit-claims.html.
[47] https://www.eurosky.social/
[48] ibid
[49] ibid

"European cloud infrastructure with open standards and democratic governance",[50] to create "a new ecosystem where innovation thrives, moderation is transparent, and no single company or country can dictate the rules."[51]This effort is the type of initiative that can help establish European digital autonomy and sovereignty and is worthy of strong support by the EU.

4.3.3 Supporting European AI based on EU Values

Political parties and candidates increasingly use AI for the creation of campaign collateral—this includes positive messages as well as AI-generated negative collateral disparaging their opponents [377]. A recent study examining 215 incidents in 50 democratic countries that held national elections in 2024 found that 80 percent had some sort of "incident"—defined as "instances of political GenAI usage"—involving generative AI [378].

Exploration of the full impact of AI on society is beyond the scope of the present report [379]. Other research deals more in-depth with the outlook and regulation of genAI in the EU [380, 381]. This report limits input on providing a rationale of a sovereign European approach to AI that would help safeguard democratic principles.

- A number of scientists have mobilized against the uncritical roll-out, adoption and use of AI in academia, highlighting the hubris, hype and dangers of pursuing fundamentally changing technology without fully understanding it [382].
- Heed lessons from the regulation of social media. Initial academic warnings about social media were largely ignored and regulation postponed until societal impact became widespread [383, 384].
- Insist on transparent and ethical use of data that respects copyright and avoids exploitation.[52] The recent LLM released by the Swiss Federal Institute of Technology
- [385] provides an example of ethical and transparent AI.[53]
- In December 2025, the #EuroLLM consortium released EuroLLM-22B, a large, fully open-source language model trained in Europe and covering all 24 official languages of the European Union. The model was trained from scratch and on a European corpus.

Assuming it lives up to its considerable promise, this model deserves to be supported by the EU.[54]

- Thoroughly test and enforce guardrails against generation of problematic content (e.g., hate speech). Current LLMs offer too little protection against circumventing of guardrails [386].
- Proper benchmarking of AI models that focuses on broader societal issues and not just technological ability [387].

On balance, this report does see great potential for AI in many dimensions of democracy, be it for information generation, synthesis, transparency, enabling personalisation of civic engagement or amplification of marginalised voices. But, at the same time, AI also enables or amplifies existing risks through increased potential for spreading disinformation at scale, deepening societal polarisation from excessive personalisation, enabling mass surveillance, reinforcing algorithmic bias, and eroding trust in institutions via opaque decision-making.

[50] ibid
[51] ibid
[52] https://www.theguardian.com/technology/article/2024/jul/06/mercy-anita-african-workers-ai-artificial-intelligence-exploitation-feeding-machine
[53] https://www.swissinfo.ch/eng/swiss-ai/fact-and-fiction-about-the-swiss-ai-model-apertus/90110034
[54] https://www.centralesupelec.fr/en/news/Sortie-d%E2%80%99EuroLLM-22B

5

CONCLUSIONS

The digital information environment—notwithstanding its numerous promises—is increasingly becoming a battleground for information integrity, marked by geo-political tensions, and the manipulation of human cognition through engagement-based business models. The development of a fantasy-industrial complex, the fracturing of reality, and the erosion of trust in democratic institutions are pressing challenges that threaten democracy and society itself. In this constantly evolving global environment, the EU stands as one of most important entities capable of taking decisive action to foster a healthy information ecosystem while preserving citizens' digital autonomy.

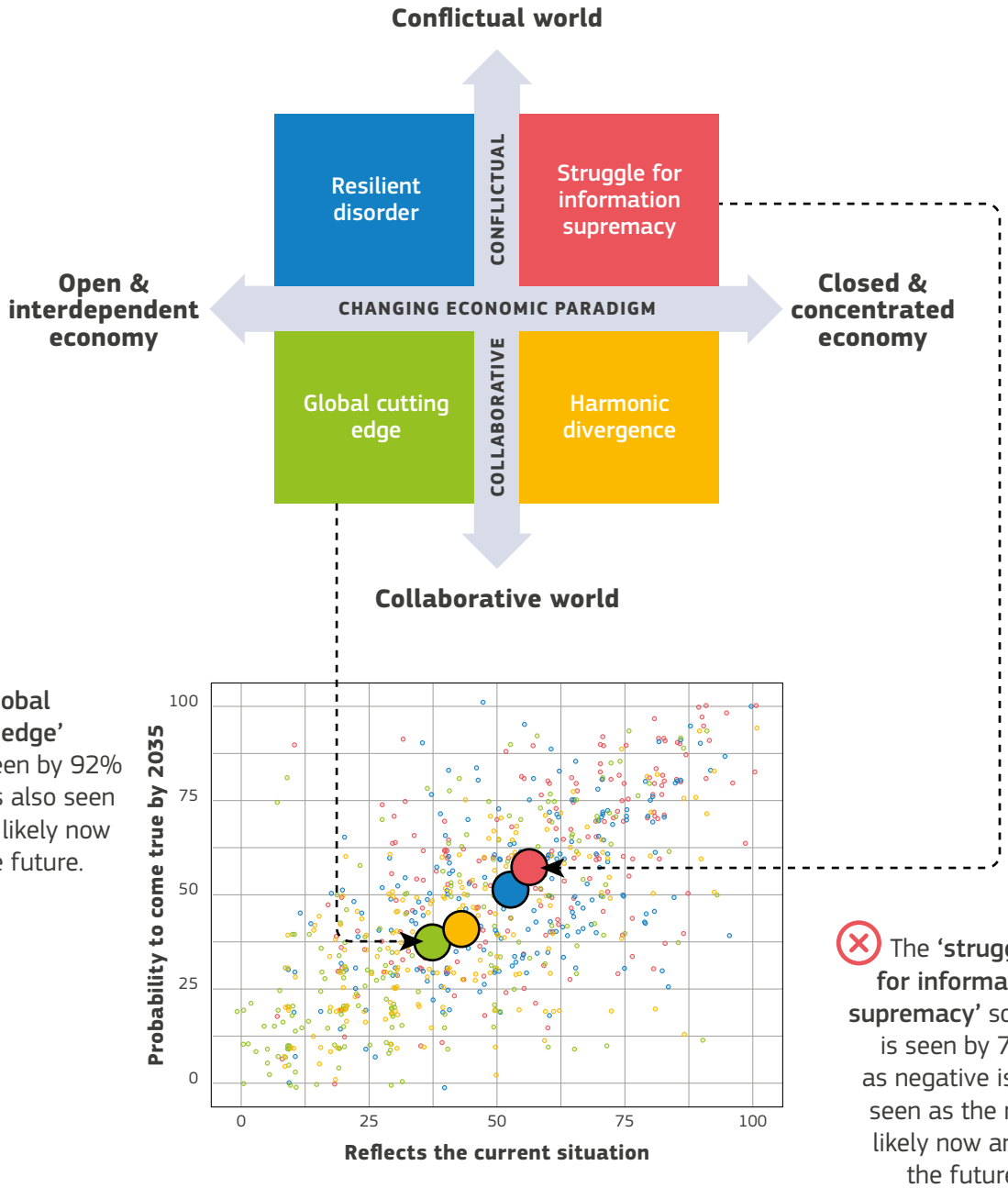
The EU's unique position as a champion of democratic values and human rights, encapsulated by the "Brussels effect" [388], places it at the forefront of addressing these challenges. The current information environment is dominated by foreign-controlled digital infrastructures and platforms that prioritise engagement and profit over the dissemination of accurate information or the well-being of its users.

The need for action is widely recognized within the expert community as well, echoing the main points of this report. In collaboration with the authors of this report, the International Panel for the Information Environment (IPIE) conducted an expert survey (in August /September 2025) that queried their views on the desirability and likelihood of current and future information scenarios. The scenarios were developed during an expert foresight analysis in 2020 [reported in 24]; see Figure 5.1.

The consensus among experts on the desirability of the scenarios is striking (textbox in Figure 5.1): the "global cutting edge" scenario is identified as most desirable, whereas the "struggle for information supremacy" scenario is almost universally rejected.

A similar consensus emerged when experts considered the present applicability of the scenarios and their future likelihood. The scatterplot in Figure 5.1 shows the responses for all scenarios, where scenarios further to the right were judged to be more prevalent now, and those further to the top were judged to become

Figure 5.1: Possible future scenarios developed during an expert elicitation in 2020 (on the left) and expert views on their present prevalence and future likelihood (right). The scatterplot shows responses to questions about the extent to which the scenario "reflects elements of the current situation in Europe" (X-axis) and "the probability that the scenario will come true by 2035" (Y-axis) on a scale from 0-100 (n = 201).



The **'global cutting edge'** scenario is seen by 92% as positive is also seen as the least likely now and in the future.

The **'struggle for information supremacy'** scenario is seen by 77% as negative is also seen as the most likely now and in the future.

Scenarios

- **Struggle for information supremacy:** assumes that the future European information space will be marked by high degrees of conflict and economic concentration. The EU is struggling to keep at least some degree of coherence in the European information space. Less than 3% of experts considered this to be beneficial for society.
- **Resilient disorder:** the EU has fostered a competitive, dynamic and decentralised information space with strong international interdependence, but is facing continuous threats from sophisticated disinformation campaigns and cyberattacks. Seen by 25% as beneficial for society.
- **Harmonic divergence:** assumes a world in which strong regulatory differences and economic protectionism between national and regional actors have resulted in a fractured, but not entirely conflictual, global information space. Seen by 60% as beneficial for society.
- **Global cutting edge:** foresees a world in which societal and geopolitical conflict have been reduced significantly, while high degrees of competition and innovation have led to the emergence of a dynamic, global information space. More than 90% of experts consider this as beneficial for society.

Source: Own elaboration

more likely in the future. The contrast between experts' evaluation of desirability for society and prognoses of likelihood is striking. The most desirable scenario, "global cutting edge", is considered to be the least likely to materialize. By contrast, the least desirable scenario, "struggle for information supremacy", is identified as the prevailing present state as well as the most likely future.

The experts were also asked to indicate in their own words "What are the most urgent actions that the EU should take?" The free-text responses were analysed with the aid of an LLM. Figure 5.2 summarizes the results. Antitrust enforcement and digital sovereignty were cited specifically by 18% of the experts, and most of the other top choices tacitly entail the need for sovereignty or stronger enforcement of existing rules: platform and AI regulation (cited by 25% of experts), and enforcement of EU digital laws (18%) go hand in glove with sovereignty.

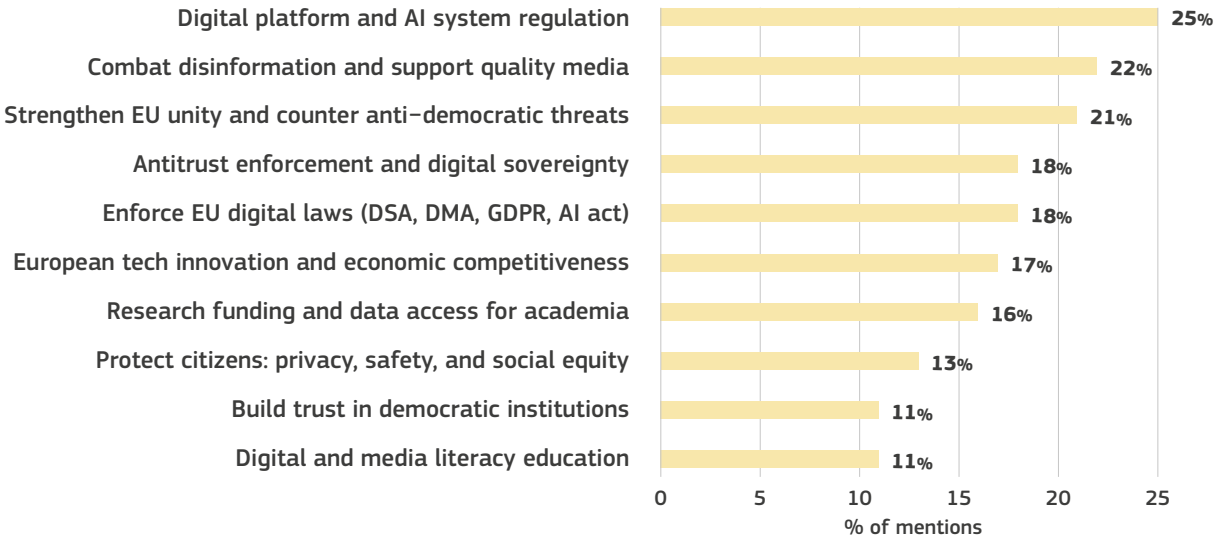
The need for digital sovereignty is not an additional layer of policy intervention, but the enabling condition for all existing and the proposed new interventions on platforms, business models, and information integrity developed in the previous chapters. Digital autonomy is a necessary condition for democratic survival rather than a desirable but non-essential goal.

By taking a leading role, the EU can set a global standard for democratic resilience in digital information spaces, ensuring that technological advancements serve the interests of citizens rather than a few powerful individuals in a winner-take-all competition. Solely focusing on winning the technological competition is not an answer to challenges to democracy.

The use of the "minimal viable evidence" for the digital information space also requires faster research and policy evaluation cycles. These could be addressed in a CERN for digital information environments that investigates alternatives to the currently dominating social media, AI and virtual worlds models. One important dimension of this is the investigation of alternative algorithms and markets for algorithms. Creating thriving markets for those algorithms by mandating interoperability or data rights would create a whole new strand of innovation where Europe could be regaining ground.

In conclusion, the challenges outlined in this report underscore the urgent need for the EU to demonstrate leadership in building digital democratic resilience and information integrity. Through strategic interventions and a commitment to democratic principles, the EU can navigate the complexities of the digital age and create an information landscape that supports a consensual yet plural understanding of reality. The time for action is now, and the EU must rise to the occasion.

Figure 5.2: Expert views on the most urgent actions for the EU to undertake. DSA refers to Digital Services Act (EU 2022/2065). DMA refers to Digital Market Act (EU 2022/1925). GDPR refers to General Data Protection Regulation (EU 2016/679), and AI act EU 2024/1689.



Source: Own elaboration

REFERENCES

[1] C. Achen and L. Bartels. "Democracy for realists: Holding up a mirror to the electorate". In: *Juncture* 22.4 (2016). ISBN: 2050-5868, pp. 269–275.

[2] R. Rekker. "Factual belief polarization between Democrats and Republicans: source or epiphenomenon of ideological and affective polarization?" In: *Frontiers in Political Science* 6 (2024). ISBN: 2673-3145, p. 1254826.

[3] R. Rekker. "How populist parties fuel science skepticism: Evidence from a 15-year panel study". In: *Acta Politica* (2025). ISSN: 1741-1416. DOI: 10.1057/s41269-025-00391-9.

[4] D. Garzia, F. F. d. Silva, and S. Maye. "Affective Polarization in Comparative and Longitudinal Perspective". In: *Public Opinion Quarterly* 87 (2023), pp. 219–231. DOI: 10.1093/poq/nfad004.

[5] Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. "Cross-country trends in affective polarization." *Review of Economics and Statistics* 106.2 (2024): pp. 557-565. DOI: 10.1162/rest_a_01160.

[6] E. J. Finkel et al. "Political sectarianism in America". In: *Science* 370 (2020), pp. 533–536. DOI: 10.1126/science.abe1715.

[7] N. Newman et al. *Reuters Institute digital news report 2025*. Reuters Institute for the Study of Journalism, 2025. DOI: 10.60625/RISJ-8QQF-JT36. URL: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025.

[8] Y. Lelkes, G. Sood, and S. Iyengar. "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect". In: *American Journal of Political Science* 61 (2017), pp. 5–20. DOI: 10.1111/ajps.12237.

[9] G. Eady et al. "How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data". In: *SAGE Open* 9 (2019). DOI:10.1177/2158244019832705.

[10] S. Iyengar et al. "The Origins and Consequences of Affective Polarization in the United States". In: *Annual Review of Political Science* 22 (2019), pp. 129–146. DOI:10.1146/annurev-polisci-051117-073034.

[11] J. J. Van Bavel et al. "How social media shapes polarization". In: *Trends in Cognitive Sciences* (2021). DOI: 10.1016/j.tics.2021.07.013.

[12] P. Törnberg. "Shifts in US Social Media Use, 2020-2024: Decline, Fragmentation, and Enduring Polarization". In: *arXiv* (2025). DOI: 10.48550/ARXIV.2510.25417.

[13] P. Lorenz-Spreen, L. Oswald, et al. "A systematic review of worldwide causal and correlational evidence on digital media and democracy". In: *Nature Human Behaviour* 7.1 (2023), pp. 74–101. ISSN: 2397-3374. DOI: 10.1038/s41562-022-01460-1.

[14] A. Simchon, M. Edwards, and S. Lewandowsky. "The persuasive effects of political microtargeting in the age of generative AI". In: *PNAS Nexus* 3 (2024), pgae035. DOI: 10.1093/pnasnexus/pgae035.

[15] K. Hackenberg et al. *The Levers of Political Persuasion with Conversational AI*. 18, 2025. URL: https://doi.org/10.48550/arXiv.2507.13919(visited on 08/20/2025).

[16] K. Mimizuka et al. "Post-Post-API Age: Studying Digital Platforms in Scant Data Access Times". In: *Journal of the Association for Computing Machinery* 37 (2025).

[17] M. Z. van Drunen and A. Noroozian. "How to design data access for researchers: A legal and software development perspective". In: *Computer Law & Security Review* 52 (2024), p. 105946. DOI: 10.1016/j.clsr.2024.105946.

[18] J. Vinagre et al. "Data Access for Recommender Systems Research: leveraging the EU's Digital Services Act". In: *Proceedings of the Nineteenth ACM Conference on Recommender Systems*. RecSys '25. New York, NY, USA: Association for Computing Machinery, 2025, pp. 1425–1426. ISBN: 9798400713644. DOI: 10.1145/3705328.3748004.

[19] J. Bak-Coleman, J. West, et al. "Industry Influence in High-Profile Social Media Research". In: *arXiv* (2026). DOI: 10.48550/ARXIV.2601.11507.

[20] K. Meßerschmidt. "COVID-19 legislation in the light of the precautionary principle". In: *The Theory and Practice of Legislation* 8.3 (2020). Publisher: Routledge_eprint: https://doi.org/10.1080/20508840.2020.1783627, pp. 267–292. ISSN: 2050-8840. DOI: 10.1080/20508840.2020.1783627.

[21] A. Alemanno. "The Precautionary Principle". In: *The Handbook of EEA Law*. Ed. by C. Baudenbacher. Cham: Springer International Publishing, 2016, pp. 839–851. ISBN: 978-3-319-24343-6. DOI: 10.1007/978-3-319-24343-6_38.

[22] G. C. Leonelli. "Acknowledging the centrality of the precautionary principle in judicial review of EU risk regulation: Why it matters". In: *Common Market Law Review* 57.6 (2020), pp. 1773–1818.

[23] S. Rose-Ackerman, S. Egidy, and J. Fowkes. *Due Process of Lawmaking: The United States, South Africa, Germany, and the European Union*. Cambridge: Cambridge University Press, 2015. ISBN: 978-1-107-33884-5. URL: https://doi.org/10.1017/CBO9781107338845.

[24] S. Lewandowsky, L. Smillie, et al. *Technology and democracy: understanding the influence of online technologies on political behaviour and decision making.* European Commission, 2020. DOI: 10.2760/709177.
[25] M. Scharfbillig et al. *Values and Identities – a policymaker's guide*. Luxembourg: Joint Research Centre, European Commission, 2021. DOI: 10.2760/022780.
[26] S. Lewandowsky, U. K. H. Ecker, J. Cook, S. van der Linden, J. Roozenbeek, and N. Oreskes. "Misinformation and the epistemic integrity of democracy". In: *Current Opinion in Psychology* (2023), p. 101711. DOI: 10.1016/j.copsyc.2023.101711.
[27] A.-G. Keizer et al. *Evidence-informed Policymaking: A pathway to increasing trust in democratic institutions and boosting competitiveness*. Luxembourg: Publications Office of the European Union, 2025. URL: https://data.europa.eu/doi/10.2760/3905455.
[28] M. Hameleers and T. van der Meer. "The Implications of Epistemic Polarization and Factual Relativism for Misinformation Research and Democracy". In: *Political Communication* 0.0 (2025). _eprint: https://doi.org/10.1080/10584609.2025.2514595, pp. 1–7. ISSN: 1058-4609. DOI: 10.1080/10584609.2025.2514595.
[29] R. S. Nickerson. "Confirmation bias: A ubiquitous phenomenon in many guises". In: *Review of general psychology* 2.2 (1998). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 175–220.
[30] K. E. Stanovich, R. F. West, and M. E. Toplak. "Myside Bias, Rational Thinking, and Intelligence". In: *Current Directions in Psychological Science* 22 (2013), pp. 259–264. DOI: 10.1177/0963721413480174.
[31] H. Gil de Zúñiga and Z. Cheng. "Origin and Evolution of the News Finds Me Perception: Review of Theory and Effects". In: *Media Influence on Opinion Change and Democracy: How Private, Public and Social Media Organizations Shape Public Opinion*. Ed. by M. Goyanes and A. Cañedo. Cham: Springer Nature Switzerland, 2024, pp. 151–179. ISBN: 978-3-031-70231-0. DOI: 10.1007/978-3-031-70231-0_10.
[32] M. Tomasello. *Origins of Human Communication*. Red. by F. Recanati. Jean Nicod Lectures. Cambridge, MA, USA: MIT Press, 13, 2010. 408 pp. ISBN: 978-0-262-51520-7.
[33] T. R. Levine. "Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection". In: *Journal of Language and Social Psychology* 33.4 (2014), pp. 378–392. ISSN: 0261-927X. DOI: 10.1177/0261927X14535916.
[34] S. Altay, M. Berriche, et al. "A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field". In: *Harvard Kennedy School Misinformation Review* (2023). DOI: 10.37016/mr-2020-119.
[35] R. DiResta. *Invisible Rulers: The People Who Turn Lies into Reality*. Hachette UK, 2024. ISBN: 1-5417-0339-1.
[36] N. Bentzen. *Online information manipulation and information integrity: An overview of key challenges, actors and the EU's evolving response*. 2024. URL: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)762416(visited on 07/07/2025).
[37] OECD. *Recommendation of the Council on Information Integrity*. 2024. URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0505(visited on 01/12/2026).
[38] NRENAISSANCE Committee. *Realizing the Information Future: The Internet and Beyond*. Washington, D.C.: National Academies Press, 1994. DOI: 10.17226/4755.
[39] S. Lewandowsky and R. Hertwig. "Critical Ignoring When Information Abundance Is detrimental to Democracy". In: *Current Opinion in Psychology* (2025). DOI: 10.1016/j.copsyc.2025.102128.
[40] T. T. Hills. "The Dark Side of Information Proliferation". In: *Perspectives on Psychological Science* 14 (2019), pp. 323–330. DOI: 10.1177/1745691618803647.
[41] M. Sultan et al. "Time pressure reduces misinformation discrimination ability but does not alter response bias". In: *Scientific Reports* 12 (2022), p. 22416. DOI: 10.1038/s41598-022-26209-8.
[42] P. Lorenz-Spreen, B. M. Mønsted, et al. "Accelerating dynamics of collective attention". In: *Nature Communications* 10.1 (2019), p. 1759. ISSN: 2041-1723. DOI: 10.1038/s41467-019-09311-w. URL: https://www.nature.com/articles/s41467-019-09311-w(visited on 07/14/2025).
[43] M. Mormina. "Knowledge, Expertise and Science Advice During COVID-19: In Search of Epistemic Justice for the 'Wicked' Problems of Post-Normal Times". In: *Social Epistemology* 36 (2022), pp. 671–685. DOI: 10.1080/02691728.2022.2103750.
[44] B. G. Peters. "What is so wicked about wicked problems? A conceptual analysis and a research program". In: *Policy and Society* 36 (2017), pp. 385–396. DOI: 10.1080/14494035.2017.1361633.
[45] D. Landa and R. Pevnick. "Political Meritocracy and Democracy". In: *Annual Review of Political Science 28,* (2025), pp. 253–271. DOI: 10.1146/annurev-polisci-050123-113309.
[46] S. Lewandowsky and P. Pomerantsev. "Technology and democracy: a paradox wrapped in a contradiction inside an irony". In: *Memory, Mind & Media* 1 (022), e5. ISSN: 2635-0238. DOI: 10.1017/mem.2021.7.
[47] C. Kang and M. Isaac. *Defiant Zuckerberg says Facebook won't police political speech*. Oct. 17, 2019. URL: https://www.nytimes.com/2019/10/17/business/zuckerberg-facebook-free-speech.ht%20ml.
[48] M. Van Alstyne. "A Consequentialist Approach to Free Speech". Boston University, Mar. 13, 2025.
[49] W. Tim. "Is the First Amendment Obsolete". In: *Michigan Law Review* 117.3 (2018), p. 548.

[50] G. King, J. Pan, and M. E. Roberts. “How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument”. In: *American Political Science Review* 111 (2017), pp. 484–501. DOI: 10.1017/s0003055417000144.
[51] E. Ulusoy et al. “Flooding the Zone: How Exposure to Implausible Statements Shapes Subsequent Belief Judgments”. In: *International Journal of Public Opinion Research* 33 (2021), pp. 856–872. DOI: 10.1093/ijpor/edab022.
[52] E. Bakshy, S. Messing, and L. Adamic. “Exposure to ideologically diverse news and opinion on Facebook”. In: *Science* 348 (2015), pp. 1130–1132. DOI: 10.1126/science.aaa1160.
[53] A. Sharma and D. Cosley. “Twitter’s recommendation system”. In: *Proceedings of the 25th International Conference on World Wide Web*. 2016, pp. 133–142.
[54] X. E. Team. *Inside X’s content recommendation pipeline*. X Corp, 2023.
[55] M. A. Baum and T. Groeling. “New Media and the Polarization of American Political Discourse”. In: *Political Communication* 25.4 (2008). pp. 345–365. DOI: 10.1080/10584600802426965.
[56] G. J. Martin and A. Yurukoglu. “Bias in cable news: Real effects and polarization”. In: *American Economic Review* 107.9 (2017), pp. 2565–99.
[57] C. A. Dekker, S. E. Baumgartner, and S. R. Sumter. “For you vs. for everyone: The effectiveness of algorithmic personalization in driving social media engagement”. In: *Telematics and Informatics* 101 (2025), p. 102300. ISSN: 0736-5853. DOI: 10.1016/j.tele.2025.102300.
[58] M. Cinelli, E. Brugnoli, et al. “Selective exposure shapes the Facebook news diet”. In: *PloS one* 15.3 (2020), e0229129.
[59] M. Cinelli, G. D. F. Morales, et al. “The echo chamber effect on social media”. In: *Proceedings of the National Academy of Sciences* 118 (2021). DOI: 10.1073/pnas.2023301118.
[60] D. Hartmann et al. “A systematic review of echo chamber research: Comparative analysis of conceptualizations, operationalizations, and varying outcomes”. In: *Journal of Computational Social Science* 8.2 (2025), p. 52.
[61] A. Mahmoudi, D. Jemielniak, and L. Ciechanowski. “Echo chambers in online social networks: A systematic literature review”. In: *IEEE Access* 12 (2024), pp. 9594–9620.
[62] K. Garimella et al. “Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship”. In: *Proceedings of the 2018 world wide web conference*. 2018, pp. 913–922.
[63] M. Ahmmad et al. “Trap of Social Media Algorithms: A Systematic Review of Research on Filter Bubbles, Echo Chambers, and Their Impact on Youth”. In: *Societies* 15.11 (2025). Multidisciplinary Digital Publishing Institute. ISSN: 2075-4698. DOI: 10.3390/soc15110301.
[64] R. Levy. “Social Media, News Consumption, and Polarization: Evidence from a Field Experiment”. In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3653388.
[65] T. Piccardi et al. “Reranking partisan animosity in algorithmic social media feeds alters affective polarization”. In: *Science* 390.6776 (2025). Publisher: American Association for the Advancement of Science, eadu5584. DOI: 10.1126/science.adu5584.
[66] A. M. Guess, N. Malhotra, et al. “How do social media feed algorithms affect attitudes and behavior in an election campaign?” In: *Science* 381 (2023), pp. 398–404. DOI: 10.1126/science.abp9364.
[67] G. Gauthier et al. “The political effects of X’s feed algorithm”. In: *Nature* (2026). ISSN: 1476-4687. DOI: 10.1038/s41586-026-10098-2.
[68] P. Törnberg. “How digital media drive affective polarization through partisan sorting”. In: *Proceedings of the National Academy of Sciences* 119 (2022), e2207159119. DOI: 10.1073/pnas.2207159119.
[69] E. Colleoni, A. Rozza, and A. Arvidsson. “Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data”. In: *Journal of Communication* 64 (2014), pp. 317–332.
[70] A. Abadie, A. Diamond, and J. Hainmueller. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program”. In: *Journal of the American Statistical Association* 105 (2010), pp. 493–505. DOI: 10.1198/jasa.2009.ap08746.
[71] A. Guess, B. Nyhan, and J. Reifler. “Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign”. In: *European Research Council* 9.3 (2018), p. 4.
[72] C. Vaccari et al. “Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of Twitter”. In: *Social media+ society* 2.3 (2016). ISBN: 2056-3051 Publisher: Sage Publications Sage UK: London, England, p. 2056305116664221.
[73] P. Barberá et al. “Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data”. In: *American Political Science Review* 113 (2019), pp. 883–901. DOI: 10.1017/s0003055419000352.
[74] A. Keuchenius, P. Törnberg, and J. Uitermark. “Echo chambers are defined by conflict, not isolation”. Mimeo.
[75] C. A. Bail et al. “Exposure to opposing views on social media can increase political polarization”. In: *Proceedings of the National Academy of Sciences* (2018). DOI: 10.1073/pnas.1804840115.
[76] E. Di Martino et al. “Ideological Fragmentation of the Social Media Ecosystem: From echo chambers to echo platforms”. In: *arXiv preprint arXiv:2411.16826* (2024).

[77] M. Avalle et al. “Persistent interaction patterns across social media platforms and over time”. In: *Nature* (2024). DOI: 10.1038/s41586-024-07229-y.

[78] N. Di Marco, E. Loru, et al. “Patterns of linguistic simplification on social media platforms over time”. In: *Proceedings of the National Academy of Sciences* 121.50 (2024), e2412105121.

[79] V. Danry et al. “Deceptive Explanations by Large Language Models Lead People to Change their Beliefs About Misinformation More Often than Honest Explanations”. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM, 2025, pp. 1–31. ISBN: 979-8-4007-1394-1. DOI: 10.1145/3706598.3713408.

[80] E. R. Spearing et al. “Countering AI-generated misinformation with pre-emptive source discreditation and debunking”. In: *Royal Society Open Science* 12.6 (June 2025), p. 242148. ISSN: 2054-5703. DOI: 10.1098/rsos.242148.

[81] T. H. Costello, G. Pennycook, and D. G. Rand. “Durably reducing conspiracy beliefs through dialogues with AI”. In: *Science* 385.6714 (2024), eadq1814. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.adq1814.

[82] E. Loru et al. “The simulation of judgment in LLMs”. In: *Proceedings of the National Academy of Sciences* 122 (2025). DOI: 10.1073/pnas.2518443122.

[83] E. M. Bender et al. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT ’21. New York, NY, USA: Association for Computing Machinery, Mar. 1, 2021, pp. 610–623. ISBN: 978-1-4503-8309-7. DOI: 10.1145/3442188.3445922.

[84] H. Yakura et al. *Empirical evidence of Large Language Model’s influence on human spoken communication*. 2025. DOI: 10.48550/arXiv.2409.01754. arXiv: 2409.01754[cs].

[85] S. González-Bailón et al. “Asymmetric ideological segregation in exposure to political news on Facebook”. In: *Science* 381.6656 (2023), pp. 392–398. DOI: 10.1126/science.ade7138.

[86] D. M. Kahan et al. “Science curiosity and political information processing”. In: *Advances in Political Psychology* 38 (2017). DOI: 10.1111/pops.12396.

[87] C. S. Taber and M. Lodge. “Motivated Skepticism in the Evaluation of Political Beliefs”. In: *American Journal of Political Science* 50.3 (2006). pp. 755–769. ISSN: 1540-5907. DOI: 10.1111/j.1540-5907.2006.00214.x.

[88] V. Goel et al. “Hatemongers ride on echo chambers to escalate hate speech diffusion”. In: *PNAS Nexus* 2 (2023). DOI: 10.1093/pnasnexus/pgad041.

[89] L. Jenke. “Affective Polarization and Misinformation Belief”. In: *Political Behavior* 46.2 (June 2024), pp. 825–884. ISSN: 0190-9320, 1573-6687. DOI: 10.1007/s11109-022-09851-w.

[90] J. B. Bak-Coleman, S. Lewandowsky, et al. “Moving towards informative and actionable social media research”. In: *arXiv* (2025). DOI: 10.48550/ARXIV.2505.09254.

[91] L. Q. Tay et al. “Thinking clearly about misinformation”. In: *Communications Psychology* 2 (2024). DOI: 10.1038/s44271-023-00054-5.

[92] P. Schmid, S. Altay, and L. D. Scherer. “The Psychological Impacts and Message Features of Health Misinformation”. In: *European Psychologist* 28 (2023), pp. 162–172. DOI: 10.1027/1016-9040/a000494.

[93] P. Nickl, M. Sultan, et al. *Global Crisis or Overblown Problem? Three Tools to Clarify Contentious Issues in Misinformation Research*. 2025. DOI: 10.31235/osf.io/4vhwq_v1.

[94] F. Zimmermann and M. Kohring. “Mistrust, Disinforming News, and Vote Choice: A Panel Survey on the Origins and Consequences of Believing Disinformation in the 2017 German Parliamentary Election”. In: *Political Communication* 37.2 (Mar. 3, 2020), pp. 215–237. ISSN: 1058-4609, 1091-7675. DOI: 10.1080/10584609.2019.1686095.

[95] S. Valenzuela, D. Halpern, and F. Araneda. “A Downward Spiral? A Panel Study of Misinformation and Media Trust in Chile”. In: *The International Journal of Press/Politics* 27.2 (2022), pp. 353–373. ISSN: 1940-1612, 1940-1620. DOI: 10.1177/19401612211025238.

[96] K. Müller and C. Schwarz. “Fanning the Flames of Hate: Social Media and Hate Crime”. In: *Journal of the European Economic Association* 19 (2021), pp. 2131–2167. DOI: 10.1093/jeea/jvaa045.

[97] C. E. Robertson et al. “Negativity drives online news consumption”. In: *Nature Human Behaviour* (2023). DOI: 10.1038/s41562-023-01538-4.

[98] S. Rathje, J. J. V. Bavel, and S. v. d. Linden. “Out-group animosity drives engagement on social media”. In: *Proceedings of the National Academy of Sciences* 118 (2021). DOI: 10.1073/pnas.2024292118.

[99] P. Nickl, M. Moussaïd, and P. Lorenz-Spreen. “The evolution of online news headlines”. In: *Humanities and Social Sciences Communications* 12.1 (Mar. 13, 2025), pp. 1–13. ISSN: 2662-9992. DOI: 10.1057/s41599-025-04514-7.

[100] K. L. McLoughlin et al. “Misinformation exploits outrage to spread online”. In: *Science* (2024). DOI: 10.1126/science.adl2829.

[101] H. Ibrahim et al. *TikTok’s recommendations skewed towards Republican content during the 2024 US presidential race*. 2025. DOI: 10.48550/arXiv.2501.17831.

[102] S. Milli et al. “Engagement, user satisfaction, and the amplification of divisive content on social media”. In: *PNAS Nexus* 4 (2025). DOI: 10.1093/pnasnexus/pgaf062.

[103] T. T. Prama et al. *Political Biases on X before the 2025 German Federal Election*. 2025. DOI: 10.48550/arXiv.2503.02888. arXiv: 2503.02888.
[104] U. K. H. Ecker, S. Lewandowsky, et al. "The psychological drivers of misinformation belief and its resistance to correction". In: *Nature Reviews Psychology* 1 (2022), pp. 13–29. DOI: 10.1038/s44159-021-00006-y.
[105] S. Altay, R. K. Nielsen, and R. Fletcher. "Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic". In: *Journal of Quantitative Description: Digital Media* 2 (2022). DOI: 10.51685/jqd.2022.020.
[106] K. Clayton et al. "Elite rhetoric can undermine democratic norms". In: *Proceedings of the National Academy of Sciences* 118 (2021). DOI: 10.1073/pnas.2024125118.
[107] N. Berlinski et al. "The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections". In: *Journal of Experimental Political Science* 10.1 (2021), pp. 34–49. ISSN: 2052-2649. DOI: 10.1017/xps.2021.18.
[108] Q. Li, B. G. King, and B. Uzzi. "Quantifying social media predictors of violence during the 6 January US Capitol insurrection using Granger causality". In: *Journal of The Royal Society Interface* 21 (2024). DOI: 10.1098/rsif.2024.0314.
[109] A. M. Guess, D. Lockett, et al. ""Fake news" may have limited effects on political participation beyond increasing beliefs in false claims". In: *Harvard Kennedy School Misinformation Review* 1.1 (2020). DOI: 10.37016/mr-2020-004.
[110] E. Porter and T. J. Wood. "Political Misinformation and Factual Corrections on the Facebook News Feed: Experimental Evidence". In: *The Journal of Politics* 84.3 (2022), pp. 1812–1817. ISSN: 0022-3816. DOI: 10.1086/719271.
[111] M. Aggarwal et al. "A 2 million-person, campaign-wide field experiment shows how digital advertising affects voter turnout". In: *Nature Human Behaviour* 7.3 (2023), pp. 332–341.
[112] A. Coppock, S. J. Hill, and L. Vavreck. "The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments". In: *Science Advances* 6 (2020), eabc4046. DOI: 10.1126/sciadv.abc4046.
[113] H. Allcott, L. Kiefer, and P. Tangkitvanich. *The Effects of Facebook and Instagram on Political Outcomes for the Average User. Available at SSRN 5341362* (2025). DOI: 10.2139/ssrn.5341362.
[114] P. Schmid and H. Bauer. "Impact of Exposure to Health Misinformation on Belief in Health Misinformation: A Meta-Analysis of RCTs". In: *Health Communication* (2025). pp. 1–11. ISSN: 1041-0236. DOI: 10.1080/10410236.2025.2536772.
[115] S. Loomba et al. "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA". In: *Nature Human Behaviour* (2021). DOI: 10.1038/s41562-021-01056-1.
[116] J. Allen, D. J. Watts, and D. G. Rand. "Quantifying the impact of misinformation and vaccine-skeptical content on Facebook". In: *Science* 384 (2024), eadk3451. DOI: 10.1126/science.adk3451.
[117] D. Jolley and K. M. Douglas. "The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions". In: *PLOS ONE* 9 (2014), e89177. DOI: 10.1371/journal.pone.0089177.
[118] L. Chen, Y. Zhang, et al. "Effects of Vaccine-Related Conspiracy Theories on Chinese Young Adults' Perceptions of the HPV Vaccine: An Experimental Study". In: *Health Communication* 36.11 (2021). pp. 1343–1353. ISSN: 1041-0236. DOI: 10.1080/10410236.2020.1751384.
[119] C. M. Greene and G. Murphy. "Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation". In: *Journal of Experimental Psychology: Applied* 27.4 (2021). pp. 773–784. ISSN: 1939-2192. DOI: 10.1037/xap0000371.
[120] C. de Saint Laurent et al. "Measuring the effects of misinformation exposure and beliefs on behavioural intentions: a COVID-19 vaccination study". In: *Cognitive Research: Principles and Implications* 7.1 (2022), p. 87. ISSN: 2365-7464. DOI: 10.1186/s41235-022-00437-y.
[121] F. Pierri et al. "Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal". In: *Scientific Reports* 12.1 (2022), p. 5966. ISSN: 2045-2322. DOI: 10.1038/s41598-022-10070-w.
[122] L. Bursztyn, A. Rao, et al. "Opinions as Facts". In: *The Review of Economic Studies* 90 (2022), pp. 1832–1864. DOI: 10.1093/restud/rdac065.
[123] S. van der Linden et al. "Inoculating the Public against Misinformation about Climate Change". In: *Global challenges (Hoboken, NJ)* 1.2 (2017), p. 1600008. ISSN: 2056-6646. DOI: 10.1002/gch2.201600008.
[124] J. Cook, S. Lewandowsky, and U. K. H. Ecker. "Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence". In: *PLOS ONE* 12.5 (2017), ISSN: 1932-6203. DOI: 10.1371/journal.pone.0175799.
[125] H. Schmid-Petri and M. Bürger. "The effect of misinformation and inoculation: Replication of an experiment on the effect of false experts in the context of climate change communication". In: *Public Understanding of Science (Bristol, England)* 31.2 (2022), pp. 152–167. ISSN: 1361-6609. DOI: 10.1177/09636625211024550.
[126] C. Drummond, M. Siegrist, and J. Árvai. "Limited effects of exposure to fake news about climate change". In: *Environmental Research Communications* 2.8 (2020). ISSN: 2515-7620. DOI: 10.1088/2515-7620/abae77.
[127] C. M. Greene, M. Brassil, et al. "Evaluating real-world effects of one-off fake news exposure". In: *Scientific Reports* 15 (2025). DOI: 10.1038/s41598-025-13291-x.

[128] T. Spampatti et al. “Psychological inoculation strategies to fight climate disinformation across 12 countries”. In: *Nature Human Behaviour* 8 (2023), pp. 380–398. DOI: 10.1038/s41562-023-01736-0.
[129] S. van der Linden. “The conspiracy-effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance”. In: *Personality and Individual Differences* 87 (2015), pp. 171–173. DOI: 10.1016/j.paid.2015.07.045.
[130] E. Elbeyi et al. *Information Integrity about Climate Science: A Systematic Review*. 2025. DOI: 10.61452/BTZP3426.
[131] D. Yanagizawa-Drott. “Propaganda and Conflict: Evidence from the Rwandan Genocide”. In: *The Quarterly Journal of Economics* 129 (2014), pp. 1947–1994. DOI: 10.1093/qje/qju020.
[132] S. Altay, B. A. Lyons, and A. Modirrousta-Galian. “Exposure to Higher Rates of False News Erodes Media Trust and Fuels Overconfidence”. In: *Mass Communication and Society* (2024), pp. 1–25. DOI: 10.1080/15205436.2024.2382776.
[133] V. Valentim. *The Normalization of the Radical Right: A Norms Theory of Political Supply and Demand*. Oxford University Press, (2024). ISBN: 978-0-19-892671-9. DOI: 10.1093/9780198926740.001.0001.
[134] L. Bursztyn, G. Egorov, and S. Fiorin. “From Extreme to Mainstream: The Erosion of Social Norms”. In: *American Economic Review* 110 (2020), pp. 3522–3548. DOI: 10.1257/aer.20171175.
[135] L. Bursztyn, G. Egorov, R. Enikolopov, et al. *Social media and xenophobia: evidence from Russia*. National Bureau of Economic Research, 2019. DOI: 10.3386/w26567.
[136] A. M. Guess, B. Nyhan, and J. Reifler. “Exposure to untrustworthy websites in the 2016 US election”. In: *Nature Human Behavior* 4 (2020), pp. 472–480. DOI: 10.1038/s41562-020-0833-x.
[137] A. M. Guess, J. Nagler, and J. Tucker. “Less than you think: Prevalence and predictors of fake news dissemination on Facebook”. In: *Science Advances* 5 (2019), eaau4586. DOI: 10.1126/sciadv.aau4586.
[138] A. Kozyreva, S. Lewandowsky, and R. Hertwig. “Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools”. In: *Psychological Science in the Public Interest* 21 (2020), pp. 103–156. DOI: 10.1177/1529100620946707.
[139] L. Oswald and S. Munzert. Exposure to untrustworthy news media then and now: Declining news media quality over 7 years. 2025. DOI: 10.31235/osf.io/5ndjx_v1.
[140] S. Boberg et al. *Pandemic Populism: Facebook Pages of Alternative News Media and the Corona Crisis – A Computational Content Analysis*. (2020). DOI: 10.48550/
[141] Y. Yang, T. Davis, and M. Hindman. “Visual misinformation on Facebook”. In: *Journal of Communication* 73 (2023), pp. 316–328. DOI: 10.1093/joc/jqac051.
[142] C. Rojas et al. “Hierarchical machine learning models can identify stimuli of climate change misinformation on social media”. In: *Communications Earth & Environment* 5 (2024), pp. 1–8. DOI: 10.1038/s43247-024-01573-7.
[143] Z. Shahbazi, R. Jalali, and Z. Shahbazi. “AI-Driven Framework for Evaluating Climate Misinformation and Data Quality on Social Media”. In: *Future Internet* 17 (2025), p. 231. DOI: 10.3390/fi17060231.
[144] AVAAZ. *Why is YouTube Broadcasting Climate Misinformation to Millions?* 2020. URL: https://avaazimages.avaaz.org/youtube_climate_misinformation_report.pdf.
[145] V. Suarez-Lledo and J. Alvarez-Galvez. “Prevalence of Health Misinformation on Social Media: Systematic Review”. In: *Journal of Medical Internet Research* 23 (2021), e17187. DOI: 10.2196/17187.
[146] B. Swire-Thompson and S. Johnson. “Cancer: A model topic for misinformation researchers”. In: *Current Opinion in Psychology* (2023). DOI: 10.1016/j.copsyc.2023.101775.
[147] J. Allen, G. Pennycook, and D. G. Rand. “Addressing misperceptions takes more than combating fake news”. In: *Trends in Cognitive Sciences* (2025). DOI: 10.1016/j.tics.2025.07.002.
[148] B. Hu et al. “Prevalence and intervention strategies of health misinformation among older adults: A meta-analysis”. In: *Journal of Health Psychology* 30 (2024), pp. 1427–1443. DOI: 10.1177/13591053241298362.
[149] K. Sørensen et al. “Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU)”. In: *European Journal of Public Health* 25.6 (2015), pp. 1053–1058. ISSN: 1101-1262. DOI: 10.1093/eurpub/ckv043.
[150] T. Gansler et al. “Sociodemographic determinants of cancer treatment health literacy”. In: *Cancer* 104 (2005), pp. 653–660. DOI: 10.1002/cncr.21194.
[151] A. K. Spälti et al. “Partisanship and anti-elite worldviews as correlates of science and health beliefs in the multi-party system of Spain”. In: *Public Understanding of Science* 32.6 (2023), pp. 761–780. ISSN: 0963-6625. DOI: 10.1177/09636625231154131.
[152] M. Serra-Garcia. *The Attention–Information Tradeoff*. CESifo Working Paper, 2025.
[153] D. Kumar et al. “Understanding the Behaviors of Toxic Accounts on Reddit”. In: *Proceedings of the ACM Web Conference 2023*. WWW ’23: The ACM Web Conference 2023. Austin TX USA: ACM, 2023, pp. 2797–2807. ISBN: 978-1-4503-9416-1. DOI: 10.1145/3543507.3583522.
[154] A. Y. Chen, B. Nyhan, et al. “Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels”. In: *Science Advances* 9.35 (2023).
[155] A. M. Guess, B. Nyhan, Z. O’Keeffe, et al. “The sources and correlates of exposure to vaccine-related (mis)information online”. In: *Vaccine* 38.49 (2020), pp. 7799–7805. ISSN: 0264-410X. DOI: 10.1016/j.vaccine.2020.10.018.

[156] C. Budak et al. “Misunderstanding the harms of online misinformation”. In: *Nature* 630 (2024), pp. 45–53. DOI: 10.1038/s41586-024-07417-w.
[157] P. Törnberg and J. Chueri. “When Do Parties Lie? Misinformation and Radical-Right Populism Across 26 Countries”. In: *The International Journal of Press/Politics* (2025), p. 19401612241311886. ISSN: 1940-1612, 1940-1620. DOI: 10.1177/19401612241311886.
[158] J. Lasser et al. “Social media sharing of low quality news sources by political elites”. In: *PNAS Nexus* (2022), pgac186. DOI: 10.1093/pnasnexus/pgac186.
[159] S. Baribi-Bartov, B. Swire-Thompson, and N. Grinberg. “Supersharers of fake news on Twitter”. In: *Science* 384 (2024), pp. 979–982. DOI: 10.1126/science.adl4435.
[160] F. Hjorth and R. Adler-Nissen. “Ideological asymmetry in the reach of pro-Russian digital disinformation to United States audiences”. In: *Journal of Communication* 69.2 (2019). ISBN: 0021-9916 Publisher: Oxford University Press, pp. 168–192.
[161] M. Mosleh and D. G. Rand. “Measuring exposure to misinformation from political elites on Twitter”. In: *Nature Communications* 13 (2022), p. 7144. DOI: 10.1038/s41467-022-34769-6.
[162] D. L. Altheide and J. N. Grimes. “War programming: The propaganda project and the Iraq war:” in: *The Sociological Quarterly* 46 (2005), pp. 617–643.
[163] L. Artz and Y. R. Kamalipour. *Bring ’em on: Media and politics in the Iraq war*. Lanham, MD: Rowman & Littlefield Publishers, 2004.
[164] S. Lewandowsky, U. K. H. Ecker, J. Cook, S. van der Linden, J. Roozenbeek, N. Oreskes, and L. C. McIntyre. “Liars know they are lying: differentiating disinformation from disagreement”. In: *Humanities and Social Sciences Communications* 11 (2024), p. 986. DOI: 10.1057/s41599-024-03503-6.
[165] A. M. McCright and R. E. Dunlap. “Combatting Misinformation Requires Recognizing Its Types and the Factors That Facilitate Its Spread and Resonance”. In: *Journal of Applied Research in Memory and Cognition* (2017). DOI: 10.1016/j.jarmac.2017.09.005.
[166] C. Paul and M. Matthews. *The Russian “Firehose of Falsehood” Propaganda Model*. RAND Corporation, 2016.
[167] N. Allen, S. Birch, and K. Sarmiento-Mirwaldt. “Honesty above all else? Expectations and perceptions of political conduct in three established democracies”. In: *Comparative European Politics* 16 (2018), pp. 511–534. DOI: 10.1057/s41295-016-0084-4.
[168] B. Cooper et al. “Honest Behavior: Truth-Seeking, Belief-Speaking, and Fostering Understanding of the Truth in Others”. In: *Academy of Management Annals* 17 (2023), pp. 655–683. DOI: 10.5465/annals.2021.0209.
[169] S. Lewandowsky, D. Garcia, et al. “When liars are considered honest”. In: *Trends in Cognitive Sciences* (2024). DOI: 10.1016/j.tics.2024.03.005.
[170] C. McVittie and A. McKinlay. “‘Alternative facts are not facts’: Gaffe-announcements, the Trump administration and the media”. In: *Discourse & Society* 30 (2018), pp. 172–187. DOI: 10.1177/0957926518816196.
[171] M. E. McCombs and D. L. Shaw. “The agenda-setting function of mass media”. In: *Public opinion quarterly* 36.2 (1972), pp. 176–187.
[172] M. McCombs. “A Look at Agenda-setting: past, present and future”. In: *Journalism Studies* 6 (2005), pp. 543–557. DOI: 10.1080/14616700500250438.
[173] S. Lewandowsky, M. Jetter, and U. K. H. Ecker. “Using the president’s tweets to understand political diversion in the age of social media”. In: *Nature Communications* 11 (2020), p. 5764. DOI: 10.1038/s41467-020-19644-6.
[174] E. L. Glaeser, G. A. M. Ponzetto, and J. M. Shapiro. “Strategic extremism: Why Republicans and Democrats divide on religious values”. In: *The Quarterly Journal of Economics* 120 (2005), pp. 1283–1330. DOI: 10.1162/003355305775097533.
[175] Center for Countering Digital Hate. *Musk’s political posts*. Center for Countering Digital Hate, 2024.
[176] T. Graham and M. Andrejevic. “A computational analysis of potential algorithmic bias on platform X during the 2024 US election”. In: *Preprint* (2024). URL: https://eprints.qut.edu.au/253211/1/A_computational_analysis_of_potential_algorithmic_bias_on_platform_X_during_the_2024_US_election-4.pdf.
[177] A. Hearn and S. Schoenhoff. “From celebrity to influencer: Tracing the diffusion of celebrity value across the data stream”. In: *A companion to celebrity* (2015). Publisher: Wiley Online Library, pp. 194–212.
[178] S. C. McGregor. ““Taking the Temperature of the Room””. In: *Public Opinion Quarterly* (2020). DOI: 10.1093/poq/nfaa012.
[179] G. Etta et al. “Characterizing engagement dynamics across topics on Facebook”. In: *Plos one* 18.6 (2023), e0286150.
[180] E. Sangiorgio et al. “Followers do not dictate the virality of news outlets on social media”. In: *PNAS nexus* 3.7 (2024), pgae257.
[181] M. Motta, Y. Liu, and A. Yarnell. ““Influencing the influencers:” a field experimental approach to promoting effective mental health communication on TikTok”. In: *Scientific Reports* 14 (2024). DOI: 10.1038/s41598-024-56578-1.
[182] C. Clarke. “The science influencers going viral on TikTok to fight misinformation”. In: *Nature* 650 (2026), pp. 542–544. DOI: 10.1038/d41586-026-00472-5.

[183] P. M. Fernbach et al. “Political Extremism Is Supported by an Illusion of Understanding”. In: *Psychological Science* (2013). DOI: 10.1177/0956797612464058.
[184] K. Ruggeri et al. “The general fault in our fault lines”. In: *Nature Human Behaviour* (2021). DOI: 10.1038/s41562-021-01092-x.
[185] M. H. Pasek et al. “Misperceptions about out-partisans’ democratic values may erode democracy”. In: *Scientific Reports* 12.1 (2022), p. 16284. DOI: 10.1038/s41598-022-19616-4.
[186] H. Hosseinmardi et al. “Examining the consumption of radical content on YouTube”. In: *Proceedings of the National Academy of Sciences* 118 (2021). DOI: 10.1073/pnas.2101967118.
[187] F. Shi et al. “The wisdom of polarized crowds”. In: *Nature Human Behaviour* 3 (2019), pp. 329–336. DOI: 10.1038/s41562-019-0541-6.
[188] L. Kurek, C. Budak, and E. Gilbert. “Wikipedia in Wartime: Experiences of Wikipedians Maintaining Articles About the Russia-Ukraine War”. In: *Proceedings of the ACM on Human-Computer Interaction* 9 (2025), pp. 1–27. DOI: 10.1145/3711107.
[189] I. Johnson et al. “Recommended Practices for NPOV Research on Wikipedia”. In: *arXiv* (2025). DOI: 10.48550/ARXIV.2510.21526.
[190] Z. J. McDowell and M. A. Vetter. “It Takes a Village to Combat a Fake News Army: Wikipedia’s Community and Policies for Information Literacy”. In: *Social Media + Society* 6 (2020). DOI: 10.1177/2056305120937309.
[191] V. Lageard and C. Paternotte. “Trolls, bans and reverts: simulating Wikipedia”. In: *Synthese* 198 (2018), pp. 451–470. DOI: 10.1007/s11229-018-02029-0.
[192] X. Fang, S. Heuser, and L. S. Stötzer. “How in-person conversations shape political polarization: Quasi-experimental evidence from a nationwide initiative”. In: *Journal of Public Economics* 242 (2025), p. 105309. ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2025.105309.
[193] J. G. Voelkel et al. “Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity”. In: *Science* 386 (2024), eadh4764. DOI: 10.1126/science.adh4764.
[194] N. Curato et al. “Twelve Key Findings in Deliberative Democracy Research”. In: *Daedalus* 146 (2017), pp. 28–38. DOI: 10.1162/daed_a_00444.
[195] S. McKay and C. Tenove. “Disinformation as a Threat to Deliberative Democracy”. In: *Political Research Quarterly* (2020). DOI: 10.1177/1065912920938143.
[196] T. Wappenhans et al. “The impact of citizens’ assemblies on democratic resilience: Evidence from a field experiment”. In: *Preprint, OSF. https://doi. org/10.31219/osf. io/hnp8k* (2024).
[197] Y.-T. Hsiao et al. “vTaiwan: An empirical study of open consultation process in Taiwan”. In: *SocArXiv* (2018).
[198] C. Horton. “The Simple But Ingenious System Taiwan Uses to Crowdsource Its Laws”. In: *MIT Technology Review* (2018).
[199] X. E. Barandiaran et al. “Decidim: A Brief Overview”. In: *Decidim, a Technopolitical Network for Participatory Democracy*. Springer Nature Switzerland, 2024, pp. 1–33. DOI: 10.1007/978-3-031-50784-7_1.
[200] A. Hudson. “When Does Public Participation Make a Difference? Evidence From Iceland’s Crowdsourced Constitution”. In: *Policy & Internet* 10.2 (2018), pp. 185–217.
[201] P. Rosa and A. Guimaraes-Pereira. *Harnessing innovation on online deliberation: lessons learnt from the on line citizen engagement process of the conference on the future of Europe*. Publications Office of the European Union, 2025. URL: https://data.europa.eu/doi/10.2760/1524483(visited on 12/08/2025).
[202] L. Van Dijk and J. Lefevere. “Can the use of minipublics backfire? Examining how policy adoption shapes the effect of minipublics on political support among the general public”. In: *European Journal of Political Research* 62.1 (2023), pp. 135–155. ISSN: 0304-4130, 1475-6765. DOI: 10.1111/1475-6765.12523.
[203] R. Weymouth and J. Hartz-Karp. “Deliberative Collaborative Governance as a Democratic Reform to Resolve Wicked Problems and Improve Trust”. In: *Journal of Economic and Social Policy* 17 (2015).
[204] A. Kozyreva, S. M. Herzog, et al. “Resolving content moderation dilemmas between free speech and harmful misinformation”. In: *Proceedings of the National Academy of Sciences* 120 (2023), e2210666120. DOI: 10.1073/pnas.2210666120.
[205] S. Munzert et al. “Citizen preferences for online hate speech regulation”. In: *PNAS Nexus* 4.2 (Feb. 2025). Ed. by K. Ognyanova. ISSN: 2752-6542. DOI: 10.1093/pnasnexus/pgaf032.
[206] Y. Theocharis et al. *Content Warning: Public Attitudes on Content Moderation and Freedom of Expression.* Content Moderation Lab at TUM Think Tank, 2025. DOI: 10.17605/OSF.IO/F56BH.
[207] International Panel on the Information Environment. *Expert Survey on the Global Information Environment 2024: Searching for Solutions*. Zurich, Switzerland, 2024. DOI: 10.61452/IQVC8768.
[208] K. Müller and C. Schwarz. *The Effects of Online Content Moderation: Evidence from President Trump’s Account Deletion*. Rochester, NY, 2023. DOI: 10.2139/ssrn.4296306.
[209] W. Ahmad et al. *The Role of Advertisers and Platforms in Monetizing Misinformation: Descriptive and Experimental Evidence*. National Bureau of Economic Research, 2024. DOI: 10.3386/w32187.

[210] N. Maani et al. “The Politics and Profit of Disinformation in Public Health”. In: *Annual Review of Public Health* (2025). Publisher: Annual Reviews. ISSN: 1545-2093. DOI: 10.1146/annurev-publhealth-071723-124408.
[211] M. Zenone et al. “Google allows advertisers to target the sensitive informational queries of cancer patients”. In: *Harvard Kennedy School Misinformation Review* (2024). DOI: 10.37016/mr-2020-170.
[212] K. Müller, C. Schwarz, and R. Jiménez-Durán. *The effect of content moderation on online and offline hate*. 2022. URL: https://cepr.org/voxeu/columns/effect-content-moderation-online-and-offline-hate.
[213] S. Maaß, J. Wortelker, and A. Rott. “Evaluating the regulation of social media: An empirical study of the German NetzDG and Facebook”. In: *Telecommunications Policy* 48 (2024), p. 102719. DOI: 10.1016/j.telpol.2024.102719.
[214] B. Kaiser and J. Mayer. *It’s the Algorithm: A large-scale comparative field study of misinformation interventions*. 2023. URL: http://knightcolumbia.org/content/its-the-algorithm-a-large-scale-comparative-field-study-of-misinformation-interventions(visited on 08/22/2025).
[215] J. Xu. “Does the medium matter? A meta-analysis on using social media vs. traditional media in crisis communication”. In: *Public Relations Review* 46.4 (2020), p. 101947. ISSN: 0363-8111. DOI: 10.1016/j.pubrev.2020.101947.
[216] Y. Jin, B. F. Liu, and L. L. Austin. “Examining the Role of Social Media in Effective Crisis Management: The Effects of Crisis Origin, Information Form, and Source on Publics’ Crisis Responses”. In: *Communication Research* 41.1 (2014), pp. 74–94. ISSN: 0093-6502. DOI: 10.1177/0093650211423918.
[217] B. T. Truong et al. *Delayed takedown of illegal content on social media makes moderation ineffective*. 2025. DOI: 10.48550/arXiv.2502.08841.
[218] Y. Huang and M. DiStaso. “Responding to a Health Crisis on Facebook: The Effects of Response Timing and Message Appeal”. In: *Public Relations Review* 46.3 (2020), p 101909. ISSN: 0363-8111. DOI:10.1016/j.pubrev.2020.101909.
[219] A. Kozyreva, P. Lorenz-Spreen, S. M. Herzog, et al. “Toolbox of individual-level interventions against online misinformation”. In: *Nature Human Behaviour* (2024). DOI: 10.1038/s41562-024-01881-0.
[220] G. Pennycook, Z. Epstein, et al. “Shifting attention to accuracy can reduce misinformation online”. In: *Nature* 592 (2021), pp. 590–595. DOI: 10.1038/s41586-021-03344-2.
[221] J. Garland et al. “The case against efficiency: friction in social media”. In: *npj Complexity* 3 (2026). DOI: 10.1038/s44260-025-00061-z.
[222] C. A. Kelly and T. Sharot. “Web-browsing patterns reflect and shape mood and mental health”. In: *Nature Human Behaviour* (2024). DOI: 10.1038/s41562-024-02065-6.
[223] J. Roozenbeek et al. “Psychological inoculation improves resilience against misinformation on social media”. In: *Science Advances* 8 (2022), eabo6254. DOI: 10.1126/sciadv.abo6254.
[224] S. Wineburg et al. “Lateral Reading on the Open Internet: A District-Wide Field Study in High School Government Classes”. In: *Journal of Educational Psychology* (2022).
[225] E. Porter and T. J. Wood. “The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom”. In: *Proceedings of the National Academy of Sciences* 118.37 (2021). ISSN: 1091-6490. DOI: 10.1073/pnas.2104235118.
[226] S. Lewandowsky, J. Cook, et al. *The Debunking Handbook 2020*. 2020. URL: https://sks.to/db2020.
[227] K. Aslett et al. “News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions”. In: *Science Advances* 8 (2022), eabl3844. DOI: 10.1126/sciadv.abl3844.
[228] L. Fazio et al. “Combating misinformation: A megastudy of nine interventions designed to reduce the sharing of and belief in false and misleading headlines”. In: *Center for Open Science* (2024). DOI: 10.31234/osf.io/uyjha.
[229] A. Simchon, T. Zipori, et al. “A Signal Detection Theory Meta-Analysis of Psychological Inoculation Against Misinformation”. In: *Current Opinion in Psychology* 67 (2025), p. 102194. DOI: 10.1016/j.copsyc.2025.102194.
[230] A. Kozyreva, S. Wineburg, et al. “Critical Ignoring as a Core Competence for Digital Citizens”. In: *Current Directions in Psychological Science* 32 (2023), pp. 81–88. DOI: 10.1177/09637214221121570.
[231] J. Brailovskaia et al. *Soziale Medien und die psychische Gesundheit von Kindern und Jugendlichen*. Halle (Saale): Deutsche Akademie der Naturforscher Leopoldina e. V., 2025.
[232] N. Walter, J. Cohen, et al. “Fact-Checking: A Meta-Analysis of What Works and for Whom”. In: *Political Communication* 37.3 (2020), pp. 350–375. DOI: 10.1080/10584609.2019.1668894.
[233] I. Bachmann and S. Valenzuela. “Studying the Downstream Effects of Fact-Checking on Social Media: Experiments on Correction Formats, Belief Accuracy, and Media Trust”. In: *Social Media + Society* 9.2 (2023), p. 20563051231179694. DOI: 10.1177/20563051231179694.
[234] N. Walter, J. J. Brooks, et al. “Evaluating the Impact of Attempts to Correct Health Misinformation on Social Media: A Meta-Analysis”. In: *Health Communication* 36.13 (10, 2021), pp. 1776–1784. ISSN: 1041-0236, 1532-7027. DOI: 10.1080/10410236.2020.1794553.
[235] N. Aruguete et al. “Truth be told: How “true” and “false” labels influence user engagement with fact-checks”. In: *New Media & Society* 27.3 (2025), pp. 1443–1464. ISSN: 1461-4448, 1461-7315. DOI: 10.1177/14614448231193709.
[236] B. Nyhan and J. Reifler. “The Effect of Fact-Checking on Elites: A Field Experiment on US State Legislators”. In: *American Journal of Political Science* 59 (2015), pp. 628–640. DOI: 10.1111/ajps.12162.

[237] A. Herasimenka et al. *Strategies for Improving the Global Information Environment: Results from a Systematic Review and Meta-Analysis*. International Panel on the Information Environment (IPIE), 2023. DOI: 10.61452/ljqp5793.
[238] J. E. Uscinski and R. W. Butler. "The Epistemology of Fact Checking". In: *Critical Review* 25 (2013), pp. 162–180.
[239] M. A. Amazeen. "Revisiting the Epistemology of Fact-Checking". In: *Critical Review* 27 (2015), pp. 1–22. DOI: 10.1080/08913811.2014.993890.
[240] M. Mendoza et al. "A Study on Information Disorders on Social Networks during the Chilean Social Outbreak and COVID-19 Pandemic". In: *Applied Sciences* 13.9 (2023), p. 5347. ISSN: 2076-3417. DOI: 10.3390/app13095347.
[241] S. Vosoughi, D. Roy, and S. Aral. "The spread of true and false news online". In: *Science* 359 (2018), pp. 1146–1151. DOI: 10.1126/science.aap9559.
[242] B. Nyhan, E. Porter, et al. "Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability". In: *Political Behavior* (2019). DOI: 10.1007/s11109-019-09528-x.
[243] E. Porter, Y. Velez, and T. J. Wood. "Correcting COVID-19 vaccine misinformation in 10 countries". In: *Royal Society Open Science* 10 (2023). DOI: 10.1098/rsos.221097.
[244] E. Porter and T. J. Wood. "Factual corrections: Concerns and current evidence". In: *Current Opinion in Psychology* 55 (2024), p. 101715.
[245] B. Swire et al. "Processing political misinformation: comprehending the Trump phenomenon". In: *Royal Society Open Science* 4 (2017), p. 160802. DOI: 10.1098/rsos.160802.
[246] R. C. Moore, R. Dahlke, and J. T. Hancock. "Exposure to untrustworthy websites in the 2020 US election". In: *Nature Human Behaviour* (2023), pp. 1–10. DOI: 10.1038/s41562-023-01564-2.
[247] M. Lavigne et al. "Inattention and Differential Exposure: How Media Questioning of Election Fraud Misinformation Often Fails to Reach the Public". *Unpublished manuscript* (2025). URL: https://business.columbia.edu/sites/default/files-efs/imce-uploads/nyhan-fraud-exposure-online.pdf(visited on 07/10/2025).
[248] G. Pennycook and D. G. Rand. "Fighting misinformation on social media using crowdsourced judgments of news source quality". In: *Proceedings of the National Academy of Sciences* 116.7 (2019), pp. 2521–2526. DOI: 10.1073/pnas.1806781116.
[249] T. Renault, D. R. Amariles, and A. Troussel. *Collaboratively adding context to social media posts reduces the sharing of false news*. 2024. DOI: 10.48550/arXiv.2404.02803.
[250] Y. Chuai et al. "Did the Roll-Out of Community Notes Reduce Engagement With Misinformation on X/Twitter?" In: *Proceedings of the ACM on Human-Computer Interaction* 8 (2024), pp. 1–52. ISSN: 2573-0142. DOI: 10.1145/3686967.
[251] L. Dierickx et al. "Striking the Balance in Using LLMs for Fact-Checking: A Narrative Literature Review". In: *Disinformation in Open Online Media*. Ed. by M. Preuss et al. Vol. 15175. Cham: Springer Nature Switzerland, 2024, pp. 1–15. ISBN: 978-3-031-71209-8 978-3-031-71210-4. DOI: 10.1007/978-3-031-71210-4_1.
[252] K. Hackenburg et al. *The Levers of Political Persuasion with Conversational AI*. 2025. DOI: 10.48550/arXiv.2507.13919..
[253] L. Mason. *Uncivil agreement: How politics became our identity*. University of Chicago Press, 2018. ISBN: 0-226-52468-X.
[254] C. C. Berning and C. Ziller. "Social trust and radical right-wing populist party preferences". In: *Acta Politica* 52.2 (2017). ISBN: 0001-6810 Publisher: Springer, pp. 198–217.
[255] T. Boeri et al. "Populism and civil society". In: *Economica* 88.352 (2021). ISBN: 0013-0427 Publisher: Wiley Online Library, pp. 863–895.
[256] H. Coffé, B. Heyndels, and J. Vermeir. "Fertile grounds for extreme right-wing parties: Explaining the Vlaams Blok's electoral success". In: *Electoral Studies* 26.1 (2007). ISBN: 0261-3794 Publisher: Elsevier, pp. 142–155.
[257] F. Berlingieri, B. d'Hombres, and M. Kovacic. *Disentangling Loneliness and Trust in Populist Voting Behaviour in Europe*. SSRN Scholarly Paper. Rochester, NY, 2025. DOI: 10.2139/ssrn.5368679.
[258] A.-P. Quaglia et al. *Democracy 'as it happens': how public spaces enable democratic resilience and citizen participation*. Luxembourg: Publications Office of the European Union.
[259] J. Qi, S. Mazumdar, and A. C. Vasconcelos. "Understanding the Relationship between Urban Public Space and Social Cohesion: A Systematic Review". In: *International Journal of Community Well-Being* 7.2 (2024), pp. 155–212. ISSN: 2524-5309. DOI: 10.1007/s42413-024-00204-5.
[260] A. Kozyreva, P. Lorenz-Spreen, R. Hertwig, et al. "Public attitudes towards algorithmic personalization and use of personal data online: evidence from Germany, Great Britain, and the United States". In: *Humanities and Social Sciences Communications* 8 (2021). DOI: 10.1057/s41599-021-00787-w.
[261] International Panel on the Information Environment (IPIE) et al. *Expert Survey on the Global Information Environment 2024: Searching for Solutions*. International Panel on the Information Environment (IPIE), 2024. DOI: 10.61452/IQVC8768.
[262] S. Zuboff. "Surveillance Capitalism and the Challenge of Collective Action". In: *New Labor Forum* 28 (2019), pp. 10–29. DOI: 10.1177/1095796018819461.

[263] T. Wu. *The attention merchants*. London, U.K.: Atlantic Books, 2017.
[264] J. A. Frimer et al. "Incivility Is Rising Among American Politicians on Twitter". In: *Social Psychological and Personality Science* 14 (2023), pp. 259–269. DOI: 10.1177/19485506221083811.
[265] M. Feinberg and J. A. Frimer. "Incivility Diminishes Interest in What Politicians Have to Say". In: *Social Psychological and Personality Science* 14.7 (2023), pp. 787–795. DOI: https://doi.org/10.1177/19485506221136182.
[266] S. Ketonen-Oksi, J. J. Jussila, and H. Kärkkäinen. "Social media based value creation and business models". In: *Industrial Management & Data Systems* 116.8 (2016), pp. 1820–1838.
[267] S. C. Briand et al. "Infodemics: A new challenge for public health". In: *Cell* 184.25 (2021), pp. 6010–6014.
[268] M. Cinelli, W. Quattrociocchi, et al. "The COVID-19 social media infodemic". In: *Scientific Reports* 10 (2020). DOI: 10.1038/s41598-020-73510-5.
[269] G. Beknazar-Yuzbashev et al. *Toxic Content and User Engagement on Social Media: Evidence from a Field Experiment*. Working Paper 11644. CESifo Working Paper, 2025. URL: https://www.econstor.eu/handle/10419/314683(visited on 11/07/2025).
[270] M. Van Alstyne. "Free Speech vs. Free Ride: Navigating the Supreme Court's Social Media Paradox". In: *Communications of the ACM* 67.11 (2024), pp. 29–31.
[271] G. Parker and M. Van Alstyne. "Internetwork externalities and free information goods". In: *Proceedings of the 2nd ACM conference on Electronic commerce*. EC00: The 2nd ACM Conference on Electronic Commerce. Minneapolis Minnesota USA: ACM, 2000, pp. 107–116. ISBN: 978-1-58113-272-4. DOI: 10.1145/352871.352883.
[272] J.-C. Rochet and J. Tirole. "Platform competition in two-sided markets". In: *Journal of the european economic association* 1.4 (2003), pp. 990–1029.
[273] C. C. Miller. "Another Try by Google to Take on Facebook". In: *New York Times* (2011). URL: https://www.nytimes.com/2011/06/29/technology/29google.html.
[274] G. Parker, M. Van Alstyne, and S. Choudary. *Platform Revolution: How Networked Markets Are Transforming the Economy and How to Make Them Work for You.* WW Norton & Company, 2016. ISBN: 0-393-24912-3.
[275] M. Djourelova, R. Durante, and G. J. Martin. "The Impact of Online Competition on Local Newspapers: Evidence from the Introduction of Craigslist". In: *The Review of Economic Studies* 92.3 (2025), pp. 1738–1772. ISSN: 0034-6527. DOI: 10.1093/restud/rdae049.
[276] European Commission. *Commission publishes guidelines on the protection of minors | Shaping Europe's digital future*. 2025. URL: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-protection-minors(visited on 08/19/2025).
[277] A. Mathur et al. "Dark Patterns at Scale". In: *Proceedings of the ACM on Human-Computer Interaction* 3 (2019), pp. 1–32. DOI: 10.1145/3359183.
[278] Directorate-General for Justice and Consumers (European Commission) et al. *Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation : final report*. Publications Office of the European Union, 2022. ISBN: 978-92-76-52316-1. URL: https://data.europa.eu/doi/10.2838/859030(visited on 08/10/2025).
[279] Dark commercial patterns. OECD Digital Economy Papers 336. 2022. DOI: 10.1787/44f5e846-en.
[280] C. Entrena-Serrano. "Watch, Scroll, Repeat: How Interface Design Shapes Consumptive Curation Affordances on TikTok". In: *Social Media + Society* 11 (2025). DOI: 10.1177/20563051251358529.
[281] S. Livingstone et al. "Can platform literacy protect vulnerable young people against the risky affordances of social media platforms?" In: *Information, Communication & Society 29* (2026). pp. 455-472. DOI: 10.1080/1369118X.2025.2518254.
[282] F. Votta et al. "Who Does (n't) Target You? Mapping the Worldwide Usage of Online Political Microtargeting". In: *Journal of Quantitative Description: Digital Media* 4 (2024).
[283] S. C. Boyle et al. "Different digital paths to the keg? How exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students". In: *Addictive Behaviors* 57 (2016), pp. 21–29. ISSN: 0306-4603. DOI: 10.1016/j.addbeh.2016.01.011.
[284] H. Hendriks et al. "Causal Effects of Alcohol-Related Facebook Posts on Drinking Behavior: Longitudinal Experimental Study". In: *Journal of Medical Internet Research* 23.11 (2021). Journal of Medical Internet Research. DOI: 10.2196/28237.
[285] A. Vannucci et al. "Social media use and risky behaviors in adolescents: A meta-analysis". In: *Journal of Adolescence* 79.1 (2020). pp. 258–274. DOI: 10.1016/j.adolescence.2020.01.014.
[286] Council of Canadian Academies. *Fault Lines: Expert Panel on the Socioeconomic Impacts of Science and Health Misinformation*. Ottawa, ON: Council of Canadian Academies, 2023.
[287] T. Assenza et al. "From buzz to bust: How fake news shapes the business cycle". 2024. URL: https://hal.science/hal-04958375v1/file/wp_tse_1516.pdf(visited on 02/16/0206).
[288] C. R. Sunstein. "Do social media platforms increase well-being? Three unresolved puzzles". In: *Theory and Society* 54.2 (2025), pp. 331–341.

[289] L. Bursztyn, B. R. Handel, et al. *When product markets become collective traps: The case of social media*. National Bureau of Economic Research, 2023.
[290] H. Allcott, M. Gentzkow, and L. Song. "Digital Addiction". In: *American Economic Review* 112.7 (2022), pp. 2424–2463. DOI: 10.1257/aer.20210867.
[291] I. A. Anderson and W. Wood. "Overestimates of social media addiction are common but costly". In: *Scientific Reports* 15 (2025). DOI: 10.1038/s41598-025-27053-2.
[292] E. Brynjolfsson, A. Collis, et al. *The Consumer Welfare Effects of Online Ads: Evidence from a 9-Year Experiment*. 2024. DOI: 10.3386/w32846. URL: https://www.nber.org/papers/w32846(visited on 08/22/2025).
[293] E. Brynjolfsson, S. T. Kim, and J. H. Oh. "The attention economy: measuring the value of free goods on the internet". In: *Information Systems Research* 35.3 (2024), pp. 978–991.
[294] G. Aridor. "Measuring Substitution Patterns in the Attention Economy: An Experimental Approach". In: *The RAND Journal of Economics* (2025). DOI: 10.1111/1756-2171.70001.
[295] C. Montag et al. "Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories". In: *International Journal of Environmental Research and Public Health* 16.14 (2019), p. 2612. ISSN: 1660-4601. DOI: 10.3390/ijerph16142612.
[296] M. Flayelle et al. "A taxonomy of technology design features that promote potentially addictive online behaviours". In: *Nature Reviews Psychology* 2.3 (2023), pp. 136–150. ISSN: 2731-0574. DOI: 10.1038/s44159-023-00153-4.
[297] A. Domin et al. "Smartphone-Based Interventions for Physical Activity Promotion: Scoping Review of the Evidence Over the Last 10 Years". In: *JMIR mHealth and uHealth* 9.7 (2021), DOI: 10.2196/24308.
[298] L. Laranjo et al. "Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression". In: *British Journal of Sports Medicine* 55.8 (2021), pp. 422–432. 1473-0480. DOI: 10.1136/bjsports-2020-102892.
[299] M. Beluri et al. "Exploration of the Dynamics of Buy and Sale of Social Media Accounts". In: *Proceedings of the 2025 ACM Internet Measurement Conference*. IMC '25. New York, NY, USA: Association for Computing Machinery, 2025, pp. 32–47. DOI: 10.1145/3730567.3732927.
[300] G. Bergmanis-Korāts and T. Haiduchyk. *Social Media Manipulation for Sale: Experiment on Platform Capabilities to Detect and Counter Inauthentic Social Media Engagement*. Riga, Latvia: NATO Strategic Communications Centre of Excellence. URL: https://stratcomcoe.org/publications/social-media-manipulation-for-sale-experiment-on-platform-capabilities-to-detect-and-counter-inauthentic-social-media-engagement/311(visited on 07/11/2025).
[301] H. Chung. *Gen Z men and women most divided on gender equality, global study shows*. 2025. URL: https://www.kcl.ac.uk/news/gen-z-men-and-women-most-divided-on-gender-equality-global-study-shows(visited on 11/10/2025).
[302] R. Nennstiel and A. Hudde. "Is there a growing gender divide among young adults in regard to ideological left–right self-placement? Evidence from 32 European countries". In: *European Sociological Review* (2025), DOI: 10.1093/esr/jcaf021.
[303] E. I. for Gender Equality. *Gender Equality Index 2024: Tackling violence against women, tackling gender inequalities | European Institute for Gender Equality*. 2024. URL: https://eige.europa.eu/publications-resources/publications/gender-equality-index-2024-tackling-violence-against-women-tackling-gender-inequalities?language_content_entity=en(visited on 12/08/2025).
[304] A. Krendel. "The men and women, guys and girls of the 'manosphere': A corpus-assisted discourse approach". In: *Discourse & Society* 31.6 (2020). SAGE Publications Ltd, pp. 607–630. DOI: 10.1177/0957926520939690.
[305] R. Mamié, M. Horta Ribeiro, and R. West. "Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube". In: *Proceedings of the 13th ACM Web Science Conference 2021*. WebSci '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 139–147. DOI: 10.1145/3447535.3462504.
[306] E. A. Renström and H. Bäck. "Manfluencers and Young Men's Misogynistic Attitudes: The Role of Perceived Threats to Men's Status". In: *Sex Roles* 90 (2024), pp. 1787–1806. DOI: 10.1007/s11199-024-01538-2.
[307] E. Bujalka, B. Rich, and S. Bender. "The manosphere as an online protection racket: How the red pill monetizes male need for security in modern society". In: *Fast capitalism* 19.1 (2022), pp. 1–16.
[308] L. Schlegel and R. Kowert. *Gaming and extremism: The radicalization of digital playgrounds*. Taylor & Francis, 2024.
[309] Calvert SL, Appelbaum M, Dodge KA, Graham S, Nagayama Hall GC, Hamby S, Fasig-Caldwell LG, Citkowicz M, Galloway DP, Hedges LV. The American Psychological Association Task Force assessment of violent video games: Science in the service of public interest. Am Psychol. 2017 ;72(2):126-143. doi: 10.1037/a0040413.
[310] G. Lamphere-Englund and J. White. "The Online Gaming Ecosystem: Assessing digital socialisation, extremism risks and harms mitigation efforts". In: (2023). Publisher: Global Network on Extremism and Technology.
[311] A. Tielemans. "A Survey of Violent Extremist and Terrorist Activities Across the Gaming Environment". In: *blog post, Global Network on Terrorism and Technology,* (2021).
[312] F. Scott-Morton and M. Kades. "Interoperability as a competition remedy for digital networks". In: *Available at SSRN 3808372* (2021). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3808372(visited on 09/08/2025).

[313] F. Scott-Morton, G. S. Crawford, et al. “Equitable interoperability: the” supertool” of digital platform governance”. In: *Yale J. on Reg.* 40 (2023), p. 1013.
[314] M. W. Van Alstyne et al. “‘In situ’data rights”. In: *Communications of the ACM* 64.12 (2021), pp. 34–35. ISSN: 0001-0782.
[315] D. Acemoglu and S. Johnson. “The urgent need to tax digital advertising”. In: *Network Law Review, Spring* 5 (2024).
[316] P. Romer. “A Tax That Could Fix Big Tech”. In: *New York Times* (2019), ISSN: 2269-9740.
[317] M. Van Alstyne. “Free Speech and the Fake News Problem”. In: *Available at SSRN 4414261* (2023).
[318] K. Müller and C. Schwarz. “From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment”. In: *SSRN* (2018). DOI: 10.2139/ssrn.3149103.
[319] A. P. Kwan, S. A. Yang, and A. H. Zhang. “Crowd-Judging on Two-Sided Platforms: An Analysis of In-Group Bias”. In: *Management Science* 70.4 (2024), pp. 2459–2476. DOI: 10.1287/mnsc.2023.4818.
[320] A. H. Zhang. *High Wire: How China Regulates Big Tech and Governs Its Economy*. Oxford University Press, 2024. ISBN: 0-19-768225-1.
[321] J. Allen, A. A. Arechar, et al. “Scaling up fact-checking using the wisdom of crowds”. In: *Science Advances* 7 (2021). DOI: 10.1126/sciadv.abf4393.
[322] C. Martel et al. “Crowds Can Effectively Identify Misinformation at Scale”. In: *Perspectives on Psychological Science* 19 (2024), pp. 477–488. DOI: 10.1177/17456916231190388.
[323] R. DiResta. “Free speech is not the same as free reach”. In: *Wired* (Aug. 30, 2018). URL: https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/.
[324] A. D. Nichols. “Truth warrants: A market-based approach to combat misinformation”. PhD thesis. Boston University, 2025.
[325] S. Altay, A.-S. Hacquin, and H. Mercier. “Why do so few people share fake news? It hurts their reputation”. In: *New Media & Society* (2020). DOI: 10.1177/1461444820969893.
[326] S. Mehta et al. “Market Design Interventions for Safer Agentic AI”. In: *ICIS 2025 Proceedings*. Vol. 30. 2025.
[327] R. A. Dahl. *Polyarchy: Participation and opposition*. Yale university press, 1971. ISBN: 0-300-15357-0.
[328] G. O’Donnell. “The quality of democracy: Why the rule of law matters”. In: *Journal of democracy* 15.4 (2004). ISBN: 1086-3214, pp. 32–46.
[329] S. D. Hyde. “Democracy’s backsliding in the international environment”. In: *Science* 369.6508 (2020), pp. 1192–1196. DOI: 10.1126/science.abb2434.
[330] M. Almeida and S. Vasilopoulou. *Scoping Report: Future Challenges to Democracy*. 2025.
[331] C. Boix. “Democracy, development, and the international system”. In: *American political science review* 105.4 (2011). ISBN: 1537-5943, pp. 809–828.
[332] A. Tarkowski and P. Keller. “Digital Public Space–A Missing Policy Frame for Shaping Europe’s Digital Future”. In: *European Public Spheres, Digitisation and Public Welfare Orientation, 2021*.
[333] T. Christakis. ““European Digital Sovereignty”: Successfully Navigating Between the “BBrussels Effect” and Europe’s Quest for Strategic Autonomy”. In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3748098.
[334] D. Di Marco, S. Thabit Gonzalez, et al. *Open but Not Powerless: Towards a Common Understanding of EU Digital Sovereignty*. 2025. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC144908(visited on 01/16/2026).
[335] S. Fratini. “The sociotechnical politics of digital sovereignty: Frictional infrastructures and the alignment of privacy and geopolitics”. In: *Big Data & Society* 12.4 (2025).
[336] R. Rommen. *JD Vance suggested the US’s support for NATO could be pulled if Europe tries to regulate Elon Musk’s X as free speech debate rumbles on*. Publisher: Business Insider. 2024. URL: https://www.businessinsider.com/jd-vance-nato-support-eu-regulation-x-musk-free-speech-2024-9.
[337] S. Lewandowsky. “Internet platforms must be held accountable for their actions”. In: *Science* 391 (2026), eaee9835. DOI: 10.1126/science.aee9835.
[338] B. Özturan et al. “Declining information quality under new platform governance”. In: *Harvard Kennedy School Misinformation Review* (2025). DOI: 10.37016/mr-2020-176.
[339] S. Rathje, C. Pretus, et al. “Unfollowing hyperpartisan social media influencers durably reduces out-party animosity”. In: *Center for Open Science* (2025). DOI: 10.31234/osf.io/acbwg_v2.
[340] S. Talaga et al. *Changes to the Facebook Algorithm Decreased News Visibility Between 2021-2024*. 2025. DOI: 10.48550/ARXIV.2507.19373.
[341] P. Davies. *Elon Musk’s X says French probe into algorithm is ‘politically motivated’*. Publisher: Euronews. 2025. URL: https://www.euronews.com/next/2025/07/21/elon-musks-x-says-french-probe-into-algorithm-is-politically-motivated (visited on 01/05/2026).
[342] R. Collard. *Germany accuses Elon Musk of trying to interfere in its national elections*. Publisher: NPR. 2024. URL: https://www.npr.org/2024/12/31/nx-s1-5243166/germany-accuses-elon-musk-of-trying-to-interfere-in-its-national-elections (visited on 01/05/2026).
[343] R. Verwiebe et al. *Digitalisiert, politisiert, polarisiert? Eine Analyse von Social-Media-Feeds junger Menschen zur Bundestagswahl 2025 auf TikTok, YouTube, Instagram und X*. Bertelsmann Stiftung, 2025. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/digitalisiert-politisiert-polarisiert(visited on 01/05/2026).

[344] T. Enarsson. “Navigating hate speech and content moderation under the DSA: insights from ECtHR case law”. In: *Information & Communications Technology Law* 33 (2024), pp. 384–401. DOI: 10.1080/13600834.2024.2395579.
[345] C. St. Aubin and J. Liedke. *Most Americans favor restrictions on false information, violent content online*. 2023. URL: https://www.pewresearch.org/short-reads/2023/07/20/most-americans-favor-restrictions-on-false-information-violent-content-online.
[346] Douek, Evelyn. “Governing Online Speech: From “Posts-as-Trumps” to Proportionality and Probability’(2021).” *Columbia Law Review* 121: 759.
[347] W. Henricksen and B. Betz. “The Stolen Election Lie and the Freedom of Speech”. In: *Penn State Law Review* (2023). DOI: 10.2139/ssrn.4354211.
[348] U. Ecker, J. Roozenbeek, et al. “Misinformation poses a bigger threat to democracy than you might think”. In: *Nature* 630 (2024), pp. 29–32. DOI: 10.1038/d41586-024-01587-3.
[349] R., Tony, and M. Oosterom. “Digital authoritarianism: A systematic literature review.” Information Technology for Development 31.4 (2025): 860-884.
[350] R. Deibert et al. *Access controlled: The shaping of power, rights, and rule in cyberspace*. the MIT Press, 2010.
[351] B. Tau. *Means of Control: How the Hidden Alliance of Tech and Government Is Creating a New American Surveillance State*. New York: Crown, 2024. 400 pp. ISBN: 978-0-593-44322-4.
[352] P. Cullen et al. *The landscape of hybrid threats – A conceptual model – Public version, Giannopoulos, G.(editor), Smith, H.(editor) and Theocharidou, M.(editor), Publications Office of the European Union, 2021*,https://data.europa.eu/doi/10.2760/44985.
[353] R. Jungwirth et al. *Hybrid Threats: A Comprehensive Resilience Ecosystem*. 2023. DOI: 10.2760/37899.
[354] M. Burgess and N. Bernal. *Chatbots Are Pushing Sanctioned Russian Propaganda*. Publisher: Wired. 2025. URL: https://www.wired.com/story/chatbots-are-pushing-sanctioned-russian-propaganda/.
[355] A. Bargăoanu, C. Buzoianu, and C. Nastasiu. *Weaponized Nostalgia – A Key Driver in Hybrid Operations Targeting the Cognitive Domain*. 2025. DOI: 10.3233/NICSP250032.
[356] Institute for Strategic Dialogue. *Country Report: Assessment of Foreign Information Manipulation and Interference (FIMI) in the 2025 German Federal Election*. Institute for Strategic Dialogue, 2025. URL: www.isdglobal.org.
[357] D. T. Schroeder et al. *How Malicious AI Swarms Can Threaten Democracy*. June 10, 2025. DOI: 10.48550/arXiv.2506.06299.
[358] N. Di Marco, S. Brunetti, et al. “Post-hoc Evaluation of Nodes Influence in Information Cascades: The Case of Coordinated Accounts”. In: *ACM Transactions on the Web* (2024). ISSN: 1559-114X. DOI: 10.1145/3700644.
[359] F. Cinus et al. “Exposing Cross-Platform Coordinated Inauthentic Activity in the Run-Up to the 2024 US Election”. In: *Proceedings of the ACM on Web Conference 2025*. WWW ’25. ACM, 2025, pp. 541–559. DOI: 10.1145/3696410.3714698.
[360] L. Luettgau et al. *Conversational AI increases political knowledge as effectively as self-directed internet search*. 2025. DOI: 10.48550/arXiv.2509.05219.
[361] J. B. Taylor and S. Richey. “AI chatbots and political learning”. In: *Journal of Information Technology & Politics* 23.1 (2024). pp. 108-118. DOI: 10.1080/19331681.2024.2422929.
[362] A. Androutsopoulou et al. “Transforming the communication between citizens and government through AI-guided chatbots”. In: *Government Information Quarterly* 36.2 (2019), pp. 358–367. DOI:10.1016/j.giq.2018.10.001.
[363] M. Haman and M. Školník. *Who Would Chatbots Vote For? Political Preferences of ChatGPT and Gemini in the 2024 European Union Elections*. arXiv:2409.00721. 2024. DOI: 10.48550/arXiv.2409.00721.
[364] M. O’Brien. *Musk’s latest Grok chatbot searches for billionaire mogul’s views before answering questions*. Publisher: AP News. 2025. URL: https://apnews.com/article/grok-4-elon-musk-xai-colossus-14d575fb490c2b679ed3111a1c83f857.
[365] M. Jeyaretnam. *How Elon Musk’s Wikipedia-Alternative ‘Grokipedia’ Describes Elon Musk*. Publisher: TIME. 2025. URL: https://time.com/7328846/elon-musk-grokipedia-wikipedia-differences-grok-xai-ai-ideological-bias/.
[366] R. Rogers. *Elon Musk’s Grokipedia Pushes Far-Right Talking Points*. Publisher: Wired. 2025. URL: https://www.wired.com/story/elon-musk-launches-grokipedia-wikipedia-competitor/.
[367] T. Renault, M. Mosleh, and D. Rand. “@Grok Is This True? LLM-Powered Fact-Checking on Social Media”. In: *OSF* (2026). DOI: 10.31234/osf.io/85quw_v2.
[368] H. Farrell and A. L. Newman. “Weaponized interdependence: How global economic networks shape state coercion”. In: *International security* 44.1 (2019). pp. 42–79.
[369] H. Farrell and A. Newman. *Underground empire: How America weaponized the world economy*. Random House, 2023. ISBN: 1-80206-208-4.
[370] M. Rutherford. “The CLOUD Act”. In: *Berkeley Technology Law Journal* 34 (2019), pp. 1177–1204.
[371] M. Rojszczak. “CLOUD act agreements from an EU perspective”. In: *Computer Law & Security Review* 38 (2020), p. 105442. DOI: 10.1016/j.clsr.2020.105442.

[372] D. G. Widder, S. West, and M. Whittaker. *Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI*. SSRN Scholarly Paper. Rochester, NY, 2023. DOI: 10.2139/ssrn.4543807.
[373] M. Micheli et al. *Data sovereignty for local governments. Considerations and enablers*. European Commission, 2024. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC138657.
[374] L. Smillie and M. Scharfbillig. *Trustworthy Public Communications*. JRC Publications Repository. ISBN: 9789268179338 ISSN: 1831-9424. 2024. DOI: 10.2760/695605.
[375] L. Hogg et al. *Shaping the Future of Social Media with Middleware*. 2024. DOI: 10.48550/arXiv.2412.10283. arXiv: 2412.10283[cs].
[376] S. Lai et al. *New Paradigms in Trust and Safety: Navigating Defederation on Decentralized Social Media Platforms*. Carnegie Endowment for International Peace. URL: https://carnegieendowment.org/research/2025/03/fediverse-social-media-internet-defederation?lang=en(visited on 10/20/2025).
[377] F. Foos. "The Use of AI by Election Campaigns". In: *LSE Public Policy Review* 3.3 (2024). DOI: 10.31389/lseppr.112.
[378] I. P. on the Information Environment (IPIE). *The Role of Generative AI Use in 2024 Elections Worldwide*. 2025. URL: https://www.ipie.info/research/tp2025-2(visited on 07/17/2025).
[379] A. McIntyre, L. Conover, and F. Russo. "A Network Approach to Public Trust in Generative AI". In: *Philosophy & Technology* 38.4 (2025), p. 137. DOI: 10.1007/s13347-025-00974-6.
[380] D. K. Abendroth et al. *From bytes to business: Strategic EU ownership and dependencies in virtual worlds*. JRC Publications Repository. 2025. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC142876(visited on 08/20/2025).
[381] C. Mougan et al. "The science and practice of proportionality in AI risk evaluations". In: *Science* 391.6787 (2026), pp. 769–771. DOI: 10.1126/science.aea3835.
[382] O. Guest et al. "Against the Uncritical Adoption of 'AI' Technologies in Academia". In: *Zenodo (preprint)* (2025). DOI: 10.5281/ZENODO.17065099.
[383] C. M. Abels et al. "The governance & behavioral challenges of generative artificial intelligence's hypercustomization capabilities". In: *Behavioral Science & Policy* (2025). DOI: 10.1177/23794607251347020.
[384] L. Floridi. "The End of an Era: from Self-Regulation to Hard Law for the Digital Industry". In: *Philosophy & Technology* 34 (2021), pp. 619–622. DOI: 10.1007/s13347-021-00493-0.
[385] A. Hernández-Cano et al. "Apertus: Democratizing Open and Compliant LLMs for Global Language Environments". In: *arXiv* (2025). DOI: 10.48550/ARXIV.2509.14233.
[386] E. Lopez-Lopez et al. "Generative artificial intelligence–mediated confirmation bias in health information seeking". In: *Annals of the New York Academy of Sciences* 1550 (2025), pp. 23–36. DOI: 10.1111/nyas.15413.
[387] M. Eriksson et al. "Can we trust ai benchmarks? an interdisciplinary review of current issues in ai evaluation". In: *arXiv preprint arXiv:2502.06559* (2025).
[388] A. Bradford. "The Brussels Effect: How the European Union Rules the World". In: *Faculty Books* (2020). DOI: https://doi.org/10.1093/oso/9780190088583.001.0001.
[389] S. Nenno et al. "Content-based detection of misinformation expands its scope across politicians and platforms". In: *OSF* (2025). DOI: 10.31235/osf.io/p6yh9_v1.
[390] SIMODS. *Measuring the State of Online Disinformation in Europe on Very Large Online Platforms*. SIMODS project, 2025.
[391] A. Bovet and H. A. Makse. "Influence of fake news in Twitter during the 2016 US presidential election". In: *Nature Communications* 10 (2019). DOI: 10.1038/s41467-018-07761-2.
[392] N. Grinberg et al. "Fake news on Twitter during the 2016 US presidential election". In: *Science* 363 (2019), pp. 374–378. DOI: 10.1126/science.aau2706.
[393] J. Allen, B. Howland, et al. "Evaluating the fake news problem at the scale of the information ecosystem". In: *Science Advances* 6 (2020), eaay3539. DOI: 10.1126/sciadv.aay3539.
[394] T., Dunlop, Kozarev, V., Guimarães Pereira, Â., Rosa, P. Citizen engagement impact: Making a difference to policy, institutions and society. Forthcoming.

LIST OF ABBREVIATIONS

AfD	Alternative für Deutschland
AI	Artifical Intelligence
API	Application Programming Interface
CDU	Christlich Demokratische Union Deutschlands
CoFE	Conference on the Future of Europe
COVID	COronaVIrus Disease
CSU	Christlich-Soziale Union in Bayern
DC-EDIC	Digital Commons European Digital Infrastructure Consortium
DMA	Digital Markets Act
DSA	Digital Services Act
EDS	European Democracy Shield
EU	Europeana Union
FIMI	Foreign Information Manipulation and Interference
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
ICC	International Criminal Court
IPIE	International Panel on Information Environments
JRC	Joint Research Centre
LLM	Large Language Model
US	United States of America
US CLOUD act	US Clarifying Lawful Overseas Use of Data Act
VLOP	Very Large Online Platform
VLOSE	Very Large Online Search Engine
WMD	Weapons of Mass Destruction

LIST OF FIGURES

Getting in touch with the EU

In person

- All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union.
You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

Scan the QR code to visit:
Joint Research Centre
https://joint-research-centre.ec.europa.eu

Publications Office
of the European Union