Cybersecurity and Law

2026 Nr 1 (15)

DOI: 10.35467/cal/219881

# Wayback Democracy: Empirical Traces of Disinformation and Polarization on X

**Urszula Soler**
Warsaw University of Life Sciences
ORCID: 0000-0001-7868-8261
E-mail: urszula_soler@sggw.edu.pl

**Stefano Lovi**
University of International Studies of Rome UNINT
ORCID: 0009-0009-9978-8689
E-mail: lovistefano1997@gmail.com

## Abstract

This study aims to identify the structural, psychological, and technological forces that enable the rapid diffusion of digital disinformation and to assess their cumulative impact on democratic stability. It focuses on how platform governance, cognitive vulnerabilities, and emergent authority structures (particularly parasocial leaders) interact to reshape public opinion formation in online environments.

Adopting an interdisciplinary approach grounded in communication theory, cognitive psychology, and platform studies, the research examines algorithmic architectures, echo-chamber patterns, and bias-driven engagement dynamics. The Wayback Machine is used as an empirical tool to track the growth trajectories of selected disinformation-driven accounts on X after the platform's deregulation, allowing longitudinal comparison of visibility, reach, and interaction patterns.

Findings indicate that reduced moderation accelerates the bandwagon effect and intensifies self-reinforcing echo chambers, enabling manipulative narratives to gain disproportionate visibility. Parasocial leaders emerge as pivotal actors: their perceived authenticity and charisma help bypass traditional epistemic filters, making audiences more susceptible to emotionally charged or misleading content. The study shows that disinformation spreads not only because of false claims but because platform design and psychological predispositions jointly amplify persuasive distortions.

The article concludes that disinformation represents a systemic challenge rooted in technological infrastructures, cognitive dynamics, and shifting forms of authority. Addressing requires coordinated regulatory reforms, enhanced digital literacy, transparent platform governance, and collaborative technological interventions. These multilayered strategies are essential for strengthening democratic resilience amid increasingly opaque and algorithmically shaped information ecosystems.

## *Keywords*

*digital disinformation, echo chambers, algorithmic governance, parasocial leaders, democratic resilience*

# Introduction

The digital transformation of the public sphere has profoundly altered the ways in which information is produced, disseminated, and validated. Social media platforms, designed around algorithmic visibility, personalization, and virality, have become central infrastructures of contemporary political communication. While these platforms have enabled unprecedented access to diverse sources of information, they have also created fertile conditions for the circulation of disinformation, the formation of echo chambers, and the emergence of new forms of charismatic authority such as parasocial leaders. Within this environment, traditional mechanisms of epistemic filtering (editorial oversight, professional journalism, institutional expertise) have been weakened, giving rise to an increasingly fragmented and polarized information ecosystem. Understanding these transformations is essential for assessing the current challenges facing democratic resilience in an era defined by algorithmic governance and post-truth dynamics.

The choice to focus on digital disinformation and the platform X (formerly Twitter) is justified by the convergence of several factors. First, disinformation has evolved from a marginal communicative disturbance to a systemic threat with direct consequences for institutional legitimacy, electoral integrity, and social cohesion. Second, recent governance changes at X, including reduced moderation, weakened verification mechanisms, and ideological shifts toward deregulation, provide a unique real-time laboratory for observing how platform policies shape the amplification of manipulative narratives. Third, empirical investigation into disinformation actors remains limited by the scarcity of transparent longitudinal data. The Wayback Machine offers an innovative methodological response, enabling the reconstruction of growth trajectories otherwise unavailable through conventional platform analytics. Finally, examining the interplay between cognitive biases, algorithmic infrastructures, and parasocial leadership fills a gap in existing interdisciplinary scholarship, which often treats these dimensions in isolation.

Current literature on digital disinformation emphasizes several interlinked dimensions. Research in communication studies and political science highlights the erosion of epistemic trust and the rise of polarized information ecosystems. Cognitive psychology points to mental shortcuts, such as confirmation bias, availability bias, and the bandwagon effect, that make individuals particularly vulnerable to manipulative content. Platform studies demonstrate how algorithmic governance, optimized for engagement rather than accuracy, systematically amplifies emotionally charged and polarizing material.

At the same time, the literature on influencers and parasocial interaction reveals how digital charisma substitutes institutional expertise with perceived authenticity, enabling “parasocial leaders” to mobilize large audiences with minimal scrutiny. Studies of computational propaganda document the role of bots, coordinated networks, and platform incentives in accelerating disinformation flows. More recent work on deepfakes and synthetic media underscores how technological sophistication further destabilizes the traditional boundaries between truth and deception.

Despite this rich body of research, empirical analyses of platform deregulation and longitudinal visibility patterns of disinformation actors remain scarce. The present study contributes to this gap by combining theoretical

literature with historical web-archive data to map how deregulation on X correlates with the accelerated growth of disinformation profiles.

This paper aims to integrate these perspectives, aiming to provide a unified theoretical and empirical framework of the conditions that favor the spread of disinformation and its democratic consequences. The article aims to clarify the structural, cognitive, and technological mechanisms that support the viral spread of manipulative content; to empirically analyze the effects of X deregulation on the growth trajectories of disinformation profiles; and to discuss the implications of these processes for institutional trust, polarization, and the quality of democratic deliberation.

These objectives lead to several central research questions: how do algorithmic architectures interact with cognitive biases to facilitate the circulation of manipulative narratives? How has the change in platform X's governance practices affected the visibility of disinformation actors? What consequences do these dynamics have on the structure of the digital public sphere and the stability of democratic institutions? The working hypotheses suggest that reduced moderation has contributed to accelerating the growth of disinformation profiles; that increased visibility produces a strengthening of the bandwagon effect and, consequently, greater perceived legitimacy of these actors; and that these processes fuel the strengthening of echo chambers and the centrality of parasocial leadership figures.

Methodologically, the article combines an interdisciplinary theoretical analysis with an empirical investigation based on historical data extracted from the Wayback Machine. This approach allows us to retrospectively observe the evolution of selected profiles, measuring variations in their growth during periods characterized by changes in platform governance. Qualitative and interpretative data analysis also allows us to connect the empirical dynamics to the theoretical mechanisms discussed in literature.

The article's structure reflects this approach: after a historical contextualization of the disinformation phenomenon, it examines the contemporary actors and techniques, the cognitive and algorithmic dynamics that favor the formation of echo chambers, the role of parasocial leaders, and the democratic implications of the phenomenon. This is followed by a presentation of the empirical case study of Platform X, before discussing the main countermeasures and outlining general conclusions.

## Definition and Historical Context of Disinformation

Disinformation and propaganda have historically served as strategic tools for political and economic interests, from forged documents in antiquity to military deception in World War II. However, the contemporary digital environment represents a far-reaching paradigm shift in both the speed and personalization of deceptive narratives. From the "*acta diurna*" of ancient Rome to the propaganda campaigns of World War II, leaders and governments have consistently understood the power of storytelling in shaping public consciousness[1]. However, the contemporary digital environment represents a far-reaching paradigm shift, both in the speed and personalization of disinformation.

What distinguishes today's disinformation from its past manifestations is not only the content, but above all the infrastructure that enables its spread. The

[1] G.S. Jowett, V. O'Donnell. Propaganda & Persuasion, Thousand Oaks 2018.

emergence of social media platforms such as Facebook, Twitter (formerly Twitter), Instagram, and TikTok has created a decentralized information ecosystem in which users are both producers and consumers of content. The erosion of editorial gatekeeping and the triumph of the logic of virality over accuracy have dramatically changed the way information is validated and considered reliable[2].

Digital disinformation differs from traditional propaganda primarily in who is practicing it, and consequently the tools used. While propaganda was typically state-driven and one-way, and therefore implemented through official or establishment channels, today disinformation can be crowd-sourced, bottom-up, and much more participatory, blurring the lines between propagandist and user.

The shift from print to the internet, and subsequently to mobile communication and algorithmic filtering, has profoundly affected the way societies communicate. The digital revolution has democratized content creation, but it has also introduced deep vulnerabilities into the public sphere. The collapse of the traditional information ecosystem, combined with the rise of influencers and alternative media sources, has created, as we have already said, fertile ground for the spread of false or misleading information[3].

This participatory culture allows anyone with access to the internet to contribute to the public narrative. The hashtag launched by Elon Musk #youarethemedianow has become emblematic in this regard. While on the one hand this opens new opportunities for marginalized voices, on the other it facilitates the circulation of hoaxes, conspiracy theories and manipulative content. Furthermore, the enormous abundance of information leads to an “information overload”, which weakens the ability of individuals to identify reliable sources[4].

Technological innovations such as bots, artificial intelligence, and deep learning have further intensified the threat of disinformation. These tools enable the mass production of false content that may appear authentic at first glance, which is particularly dangerous in contexts such as elections, public health, or international scenarios.

## Actors and Strategies of Contemporary Disinformation

Disinformation campaigns are orchestrated by state actors (e.g., the Internet Research Agency)[5], political groups, and individual users seeking geopolitical or economic advantage. These actors frequently employ troll farms, bots, and automated networks to simulate grassroots consensus, a tactic known as “astroturfing”[6]. Modern techniques include:

1) fake news production: the deliberate creation and dissemination of false news, constructed to appear credible and attract attention. They can be published on sites specifically created for this purpose or spread through apparently authentic social profiles. During the 2016 US presidential elections, false

[2] D.M. Lazer, et al., The science of fake news, “Science” 2018, 359(6380), p. 1094-1096.
[3] L. McIntyre, Post-Truth, “European Journal of Communication” 2018, 33(5), p. 574-575.
[4] D. Bawden, L. Robinson, The dark side of information: overload, anxiety and other paradoxes and pathologies, “Journal of Information Science” 2009, 35(2), p. 180-191.
[5] R. DiResta, K. Shaffer, B. Ruppel, D. Sullivan, R. Matney, R. Fox, J. Albright, B. Johnson, The Tactics & Tropes of the Internet Research Agency, Stanford Internet Observatory 2019.
[6] P.N. Howard, B. Kollanyi, Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum, Social Science Research Network 2016.

articles declaring the death of celebrities or unfounded accusations against political candidates were spread, generating millions of shares and influencing public opinion. Usually, as in the case of the latest European elections, these are supported using *spindoctors*, i.e. accounts created specifically to send links and materials containing false news, and then quickly delete the profiles themselves once the information has entered the circulation of the platforms. In this way, they manage to hide the origin of the fake news or sources of disinformation[7];

2) decontextualization: this technique consists of taking authentic content, such as an image or video, and presenting it in a false context to alter its meaning. A common tactic involves "CheapFakes", such as misrepresenting unrelated older footage (like a climate protest in Vienna[8]) as evidence of staged civilian casualties during the conflict in Ukraine, thereby generating unfounded outrage;
3) memetic warfare: memes, images or short videos accompanied by ironic or satirical texts, have become powerful tools of disinformation. They can convey ideological, political or cultural messages in a simplified and viral way, often eluding normal critical filters. During election campaigns or current events, memes that ridicule one candidate or praise another can spread rapidly, influencing the public's judgment without providing in-depth analysis;
4) use of Deepfake technology: these are audiovisual contents manipulated through artificial intelligence, which allow for the extremely realistic simulation of the faces and voices of real people. These tools make it increasingly difficult to distinguish real contents from manipulated ones, raising serious concerns about the erosion of information credibility. Deepfakes have already been used in political campaigns and in revenge porn, but their potential extends to the discrediting of public figures, the manipulation of financial markets and the incitement to violence[9].

Deepfakes pose a twofold challenge: they act as a powerful means of active disinformation while simultaneously generating the "liar's dividend" [10]. This cognitive loophole allows individuals to deny the authenticity of genuine, compromising evidence by claiming it was artificially generated (a tactic already observed in recent political scandals[11]). Conversely, synthetic media has been weaponized in conflict zones, such as the widely debunked 2022 deepfake video urging Ukrainian surrender[12], demonstrating the technology's unprecedented potential to disrupt real-time geopolitical crises[13].

[7] R. DiResta, The Spin Doctors Are In. Guest: Renée DiResta, Center for Human Technology, 2020.
[8] F. Saul, Bewegung in Leichensäcken – dieser Fake-Clip wird seit Kriegsbeginn immer wieder verbreitet, Stern 2023.
[9] R. Chesney, D.K. Citron, Deepfakes and the New Disinformation War, "Foreign Affairs" 2011, 98(1), p. 147-155.
[10] D. Fallis, The Epistemic Threat of Deepfakes, "Philos Technol" 2021, 34(4), p. 623-643.
[11] D. Victor, "Access Hollywood" Reminds Trump: "The Tape Is Very Real", The New York Times 2017.
[12] T. Simonite, A Zelensky Deepfake Was Quickly Defeated, The Next One Might Not Be, WIRED 2022.
[13] D.L. Byman, C. Gao, C. Meserole, V.S. Subrahmanian, Deepfakes and International Conflict, Foreign Politicy at Brookings 2023.

## The Structural Link Between Echo Chambers and Cognitive Biases

The dynamics of online disinformation cannot be understood without a careful analysis of the interaction between algorithmic architectures and human cognitive mechanisms. The information manipulation techniques described in the previous chapter thrive in digital ecosystems governed by the attention economy, in which metrics such as likes, shares, retweets and views define the visibility and circulation of information.

This logic of attention, rooted in precise economic objectives, namely increasing the time spent on the platform and, consequently, advertising revenues, favours content that is emotionally charged, polarizing and easily digestible. In this context, the problem lies not only in the way algorithms work, but also in the way they are optimized: not for their information value, but for the emotional reactivity they can generate[14]. Such optimization leads to the systematic overexposure of sensational content to the detriment of accurate and contextualized information.

Not only algorithmic codes contribute to this dynamic, but also user behaviour. The use of digital information is often characterized by short attention spans and preferences for short, visual, and emotionally impactful content. This makes disinformative content, often constructed with manipulative purposes and therefore designed to strike quickly, more effective in terms of social metrics than traditional journalistic content, which requires critical reading and contextualization.

In this environment, algorithmic selection acts as a filter that enhances various cognitive biases already present in the user's mind. For example, confirmation bias pushes users to favour information that confirms their pre-existing beliefs, ignoring or devaluing dissonant ones[15]. This behaviour, often implemented unconsciously, is further reinforced by social media algorithms, which tend to propose content like the preferences previously expressed by the user, favouring a unilateral worldview where contradictory and constructive debate find little space.

The result is the formation of so-called echo chambers and filter bubbles, or digital environments in which the user is exposed almost exclusively to perspectives that reflect or reinforce his or her opinions. Such environments not only reduce exposure to dissent but also limit the ability to develop critical thinking and to question one's cognitive schemas[16]. Contrary information is ignored, discredited or perceived as a threat, reinforcing the dichotomy of an "Us" versus a "Them" that fuels polarization.

The implications of these dynamics are profound and affect both the cognitive and sociopolitical spheres. From a psychological point of view, repeated exposure to ideologically coherent content strengthens group identity and increases trust in the narratives shared by one's ingroup. This identity strengthening is a powerful driver of polarization, which can intensify to the point of leading to ideological radicalization. One of the most cited examples in

[14] Z. Tufekci, Facebook Said Its Algorithms Do Help Form Echo Chambers, and the Tech Press Missed It, "New Perspectives Quarterly" 2015, 32(3), p. 9-12.
[15] B.J. Casad, J.E. Luebering, Confirmation bias, Britannica 2025.
[16] D. Spohr, Fake news and ideological polarization: Filter bubbles and selective exposure on social media, "Business Information Review" 2017, 34(3), p. 150-160.

literature concerns the functioning of YouTube's recommendation engine, which according to several studies can push users towards increasingly extreme content over time, through a sequence of progressive suggestions. In these cases, the logic of engagement fuels a path of algorithmic radicalization, in which the user is gradually led from mainstream content to conspiracy or extremist narratives.

Algorithmic metrics (likes, shares) artificially amplify the bandwagon effect, acting as social signals that implicitly legitimize content. During the COVID-19 pandemic, conspiracy theories gained traction largely through such amplification, creating a self-referential circuit of social validation[17]. Platforms commercially exploit this by boosting popular content to maximize engagement, making it exceedingly difficult to halt viral disinformation once it reaches a critical visibility threshold.

## The Role of Parasocial Leaders

The transformation of the public sphere in the digital age has given rise to a new type of opinion leader: the social media influencer. Unlike traditional opinion leaders, who were typically pundits, intellectuals, or journalists whose authority derived from institutional roles or recognized expertise, influencers often draw their power from visibility and charisma rather than knowledge or analytical expertise. This shift coincides with the emergence of what Horton and Wohl famously called the "parasocial relationship" way back in 1956, a one-sided psychological bond that media consumers form with mediated figures who are unaware of their existence.

In the contemporary digital ecosystem, *parasociality* has been dramatically amplified by constant interactivity and algorithmic personalization. Influencers present themselves in intimate and accessible ways, sharing personal stories, behind-the-scenes content, and daily routines, thereby encouraging audiences to develop feelings of trust, familiarity, and emotional closeness. These pseudo-relationships strengthen the persuasive power of influencers, who are often perceived as more authentic and trustworthy than traditional media figures[18]. In the contemporary digital ecosystem, this type of relationship has not only become normalized, but has been amplified and made more sophisticated.

The result is the creation of an "algorithmic intimacy"[19], in which emotional engagement is mediated but perceived as authentic. Users feel connected to influencers as true friends or confidants, which creates fertile ground for the suspension of critical thinking. Unlike traditional media, where authority was linked to journalistic standards and verification mechanisms, parasocial communication is based on an affective and performative dynamic, which rewards narrative coherence, the *spectacularization* of personal experience, and charisma.

Building online authority depends less on institutional credentials and more on branding strategies, narrative coherence, and audience engagement. Influencers cultivate a loyal following by creating a distinctive persona and

[17] R.B. Cialdini, N.J. Goldstein, Social influence: Compliance and conformity, "Annual Review of Psychology" 2004, no. 55, p. 591–621.
[18] C. Abidin, Visibility labour: Engaging with Influencers' fashion brands and #OOTD advertorial campaigns on Instagram, "Media International Australia" 2016, 161(1), p. 86-100.
[19] S. Bishop, Influencer Management Tools: Algorithmic Cultures, Brand Safety, and Bias, "Social Media + Society" 2021, 7(1).

maintaining ongoing, responsive interactions with their audiences. Their power does not necessarily lie in offering factual or well-reasoned commentary, but in their ability to capture attention, resonate emotionally, and cultivate perceived intimacy. This results in a communication environment where emotional connection and entertainment value often outweigh rational reflection[20].

A case in point are wellness or alternative health influencers, some of whom have played a significant role in spreading conspiracy theories during the COVID-19 pandemic, promoting pseudoscientific information about vaccines, unapproved therapies, and alleged global conspiracies. The bond of trust established with followers, combined with an emotional and "anti-establishment" narrative, has made these actors incredibly effective in conveying disinformation messages, often with concrete effects on public health[21].

These dynamics have profound implications for the formation of public opinion. The rise of parasocial leaders facilitates the viral spread of unverified and emotionally charged content that reinforces tribal divisions rather than promotes democratic discourse. When influencers intervene in political or ideological debates, their messages often bypass traditional mechanisms of scrutiny and validation. Their followers, emotionally invested in the influencer's personality, are more likely to accept their claims uncritically, contributing to the normalization and amplification of misinformation or polarizing narratives[22].

## Democratic Consequences of Disinformation

One of the most alarming effects of disinformation in the digital age is the erosion of trust in democratic institutions. By polluting the information environment with false or misleading narratives, disinformation undermines trust in the media, the scientific community, and public governance. This creates what scholars call an "epistemic crisis", in which the very idea of shared truth becomes suspect and consensual reality shatters into partisan echo chambers[23].

As users increasingly struggle to distinguish between facts and fiction, the legitimacy of democratic institutions suffers. Trust in electoral processes, judicial decisions and public policies becomes fragile, especially when malicious actors intentionally sow doubts about the integrity of these systems. This is particularly evident in moments of crisis, such as the COVID-19 pandemic or the recent highly contentious elections in Romania, when disinformation can fuel conspiracy theories, public confusion and institutional paralysis.

Disinformation also plays a central role in intensifying social polarization and conflict. By promoting divisive and emotionally provocative narratives, disinformation deepens ideological fractures and consolidates group identities. Digital platforms, designed to maximize engagement through emotionally resonant content, tend to privilege outrage, fear and resentment: emotions that

[20] A. Marwick, R. Lewis, Media manipulation and disinformation online, "Data & Society" 2017.
[21] M. Cinelli, et al., The COVID-19 social media infodemic, Scientific Reports 2020, no. 10, p. 16598.
[22] E.C. Tandoc, D. Lim, R. Ling, Diffusion of disinformation: How social media users respond to fake news and why, "Journalism" 2019, 21(3), p. 381-398.
[23] S. Lewandowsky, U.K.H. Ecker, J. Cook. Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era, "Journal of Applied Research in Memory and Cognition" 2017, 6(4), p. 353-369.

are highly effective in driving user interaction but toxic to deliberative democracy[24].

This polarized environment fuels the rise of extremism and populist rhetoric, often making compromise or dialogue impossible. Public forums, both online and offline, become battlegrounds between opposing realities, where trust and empathy are replaced by hostility and suspicion. These conditions are ripe for manipulation by actors seeking to destabilize social cohesion or exploit divisions for political gain.

Perhaps most significantly, disinformation has been shown to have a direct impact on electoral processes. The 2016 US presidential election and the Brexit referendum are stark examples of how digital disinformation campaigns, often coordinated by foreign entities, can distort public opinion, limit voter turnout, or distort public debate. These interventions, amplified by micro-targeted ads and bot-generated content, can subtly influence outcomes without leaving obvious traces, thus evading accountability[25].

Democracies are increasingly vulnerable to these systemic disruptions. The recent highly contentious elections in Romania are emblematic: the institutional invalidation of a candidate sparked violent protests and provided propaganda victories for foreign adversaries, illustrating how disinformation exploits institutional fragilities to undermine Western democratic legitimacy.

## Case Studies: Disinformation on X

Analysing disinformation mechanisms requires not only a theoretical framework but also the empirical observation of concrete cases that allow the discussed phenomena to be translated into real-world data and dynamics. From this perspective, platform X represents a particularly relevant field of investigation. Its communication architecture, combined with recent transformations introduced by the new corporate governance, has made it a privileged space for the circulation of conspiracy narratives and manipulative content. The decision to drastically reduce content moderation and verification systems, in the name of a more “libertarian” approach to communication, has had measurable effects on the amplification of disinformation actors and their ability to attract new audiences.

To concretely observe this process, we used the Wayback Machine, a digital archive that preserves historical copies of websites and social media pages, thus allowing us to retrospectively analyse the evolution of selected profiles. The tool, despite its limitations (the discontinuity of snapshots and the lack of complete data on interactions and networks), lends itself to an empirical documentation function that is difficult to replicate through other open sources. Starting from monitoring Telegram channels dedicated to disinformation and conspiracy theories, such as “BioClandestine”[26] or “ULTRA Pepe Lives Matter”[27],we analysed the major X accounts that were relaunched in the dissemination of such narratives. The historical observation of these profiles

[24] C.A. Bail, et al., Exposure to opposing views on social media can increase political polarization, “Proceedings of the National Academy of Sciences” 2018, 115 (37), p. 9216-9221.
[25] W.L. Bennett, S. Livingston, The disinformation order: Disruptive communication and the decline of democratic institutions, “European Journal of Communication” 2018, 33(2), p. 122-139.
[26] Link to the channel: https://t.me/bioclandestine [access: 16.12.2025].
[27] Link to the channel: https://t.me/PepeMatter [access: 16.12.2025].

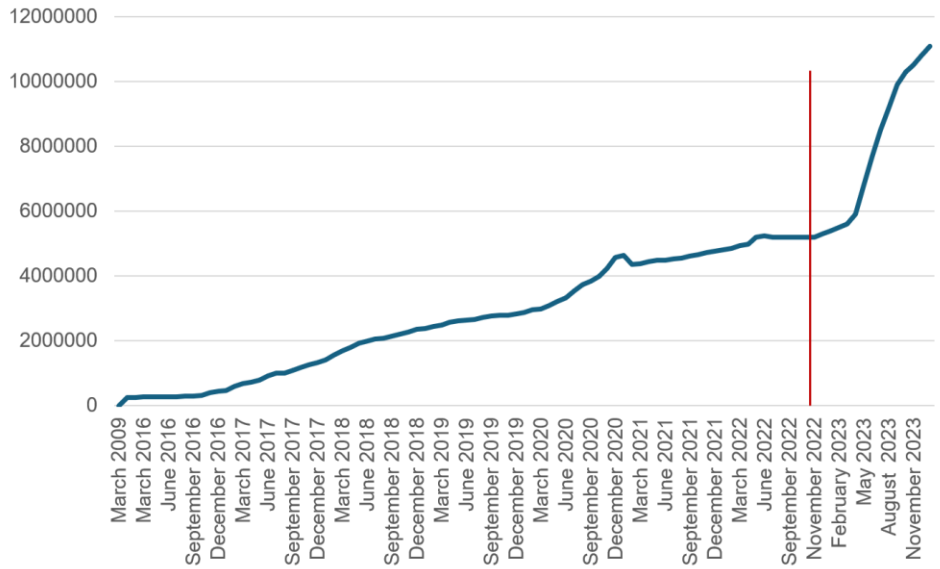
shows significant and, in some cases, sudden growth in the number of followers starting in the period following the relaxation of moderation policies.

The increase should not be interpreted simply as a quantitative fact, but as a sign of the psychological and social mechanisms already discussed in this article. The perception of a growing community fuels the bandwagon effect, encouraging new users to embrace narratives that appear shared by a growing majority. At the same time, the growing audience strengthens the formation of echo chambers, as content produced and reposted by these accounts more easily reaches an already predisposed audience, further reducing exposure to opposing voices. Furthermore, the expansion of their follower base helps legitimize the image of some digital influencers as parasocial leaders, whose authority derives not from certified expertise, but from the symbolic capital represented by their numbers and perceived intimacy with their followers.

Using the Wayback Machine allows us to visualize this process as a temporal trajectory: before deregulation, accounts showed slow or stable growth, while subsequently significant leaps were observed, often coinciding with political events or international crises. Algorithmic deregulation has therefore not only reduced the filters for potentially false or manipulative content but has also created the conditions for its active amplification, fostering visibility that translates into social legitimacy.

The evidence analysed, which we make verifiable by publishing the datasets and graphs in a publicly accessible online archive[28], strengthens the hypothesis that the deregulation of moderation has had a direct impact on the amplification of disinformation narratives and their centrality in the digital information ecosystem.

**Fig. 1.** *Evolution of the number of followers of account no. 1. The account in question was opened in 2009. It currently has 16.6 million followers. From November 2022 to November 2023 alone, it saw a 102% increase.*




[28] Link to the online dataset: https://docs.google.com/spreadsheets/d/1DhUiX4MzYV75hhH Baea HW-W2001_ANmW/edit?usp=sharing&ouid=100775684869765758764&rtpof=true&sd=true [access: 16.12.2025].

**Fig. 2.** *Evolution of the number of followers of account no. 4. The account in question was opened in 2021. It currently has 1.3 million followers. From November 2022 to November 2023 alone, it saw a 1363% increase.*

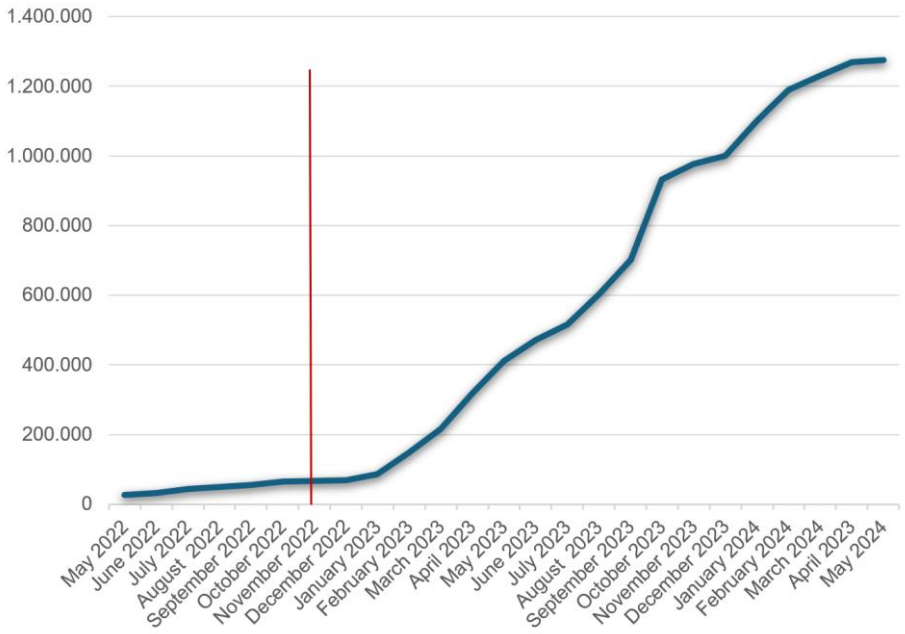
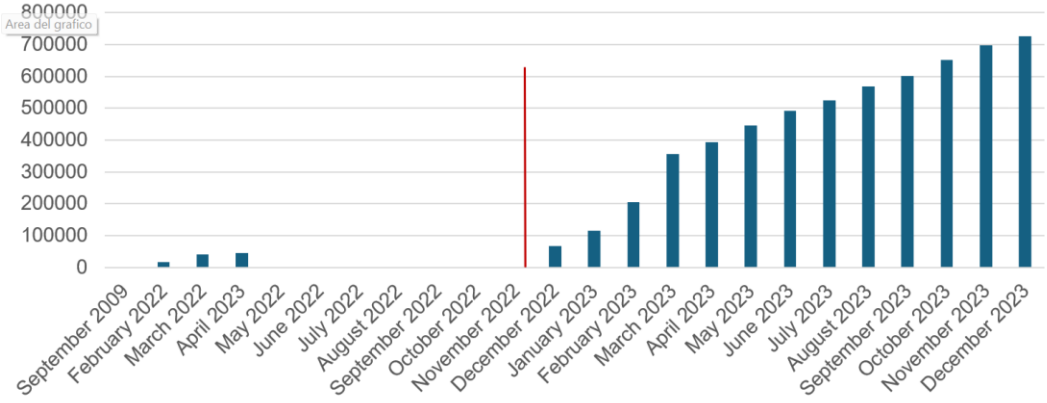



**Fig. 3.** *Evolution of the number of followers of account no. 18. The account in question was opened in 2009. It currently has 1.9 million followers. From December 2022 to December 2023 alone, it saw a 972% increase. From May 2022 to December 2022, the account was banned.*




This case study demonstrates in concrete terms how platform governance choices can directly impact the health of the digital public space. Far from being a limited phenomenon, the growth of disinformation actors on X reflects a structural trend: when the burden of distinguishing between authentic and manipulative content is entirely delegated to individual users, who lack the necessary cognitive and informational tools, disinformation gains a competitive advantage in terms of attention, dissemination, and political impact.

The empirical analysis conducted through the Wayback Machine aligns with the theoretical framework developed in the previous chapters. It demonstrates that the dynamics of the bandwagon effect, the strengthening of echo chambers, and the legitimation of parasocial leaders are not conceptual abstractions but are measurably reflected in the concrete transformations of

digital platforms. The case of X suggests that disinformation thrives not simply thanks to sophisticated communication strategies, but also by virtue of corporate choices that, by modifying the rules of the algorithmic game, reshape the structural conditions of the contemporary public sphere.

## Strategies to Counter Disinformation

Countering disinformation in all its forms is one of the most urgent and complex challenges for contemporary democracies today. The polymorphic and adaptive nature of the phenomenon, capable of evolving extremely rapidly in response to new technologies, global events and changes in user behaviour, requires a global, multilayered and deeply coordinated approach. There is no single solution: however, several researchers argue that effective responses must intertwine regulation, media literacy, technological innovation, interdisciplinary research and civic engagement[29].

One of the most debated areas of intervention in recent years concerns the regulatory supervision of digital platforms. Long self-regulated according to market logic, big tech is now at the center of public and political debate for the role they play in the spread of harmful content, including fake news, hate speech and coordinated manipulation campaigns. Governments, supranational authorities and advocacy bodies are calling for greater transparency in the management of algorithms, content moderation and use of personal data. A prime example is the European Union's Digital Services Act (DSA), which came into force in 2023, which imposes stringent obligations on large platforms to assess and mitigate the "systemic risks" associated with their services, such as electoral disinformation, radicalization or manipulation through bots and automated accounts[30].

However, regulatory regulation alone is not enough. Laws can set external limits and establish oversight mechanisms, but they do not act directly on the cognitive and emotional vulnerabilities of individuals, which are one of the main vectors of disinformation dissemination. For this reason, an effective counteraction strategy must necessarily include interventions aimed at strengthening media and digital literacy. Educating citizens from school age to understand the logic of the information system, to recognize reliable sources, to distinguish between opinion and verified data, and to develop critical thinking towards online content is an essential investment for democratic resilience in the long term. In this context, innovative educational initiatives such as prebunking programs have shown promising results.

Research by various authors has shown that brief interventions, such as videos or interactive games that expose users to rhetorical techniques common in disinformation, such as scapegoating, exaggerated emotionality or the use of false experts, can significantly increase individuals' critical capacity[31]. Unlike traditional debunking, which seeks to correct false information after it has been internalized, pre-bunking acts ex ante, strengthening "cognitive defences" against future attempts at manipulation.

---

[29] B. Rød, C. Pursiainen, N. Eklund, Combatting Disinformation – How Do We Create Resilient Societies? Literature Review and Analytical Framework, "Eur J Secur Res" 2025.

[30] R. Gorwa, What is platform governance?, "Information, Communication & Society" 2019, 22(6), p. 854.

[31] S. van der Linden, J. Roozenbeek, J. Compton, Inoculating Against Fake News About COVID-19, "Front Psychol" 2020, no. 23, p. 11:566790.

At the same time, a collective and multilateral effort involving a broad spectrum of social actors is needed. Countermeasures against disinformation cannot be entrusted exclusively to states or technology companies. Structured collaboration between public institutions, digital platforms, university researchers, journalists, educators and civil society organisations is needed. Only through an integrated and collaborative approach is it possible to address the complexity of the current information challenge. Initiatives such as international fact-checking consortia, such as the International Fact-Checking Network, interdisciplinary teams for the analysis of electoral data, or task forces against foreign influence, such as the Center for Strategic Communication of the European Commission, represent models of virtuous cooperation that can be replicated and strengthened.

Finally, it is essential to invest in the development of responsible technologies, capable of detecting and containing disinformation in real time, without falling back into mechanisms of censorship or arbitrary control. Artificial intelligence can play a positive role in this context, for example in the automatic detection of deepfakes, in the flagging of suspicious content or in the promotion of reliable content. However, it is essential that these technologies are designed with transparency and human oversight criteria, to prevent them from becoming tools of potential abuse.

However, this balance is severely compromised when major platforms reduce moderation and verification systems. The recent shift towards a more "libertarian"[32] approach by tech leaders leaves the field open to manipulative content and complicates the enforcement of regulations like the Digital Services Act. Consequently, the burden of media literacy falls entirely on individual users, fostering a climate of generalized distrust.

## Summary

Digital disinformation constitutes a structural challenge for contemporary democracies, requiring systemic, interdisciplinary responses. This study demonstrates how manipulative narratives thrive in disintermediated environments where algorithmic logics prioritize engagement over accuracy, fostering echo chambers and empowering parasocial leaders as new post-truth authorities. Empirical evidence from X highlights that algorithmic deregulation actively accelerates the visibility and influence of disinformation actors. Addressing this crisis demands moving beyond simple content removal; it requires redesigning algorithms for transparency, enforcing regulatory frameworks like the DSA, and investing in critical digital literacy. Only through collective commitment can we build a resilient digital citizenry capable of sustaining a participatory democracy.

## Bibliography

Abidin, C., Visibility labour: Engaging with Influencers' fashion brands and #OOTD advertorial campaigns on Instagram, "Media International Australia" 2016, 161(1), https://doi.org/10.1177/1329878X16665177.

[32] L. McMahon, Z. Kleinman, C. Subramanian, Facebook and Instagram get rid of fact checkers, BBC 2025.

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Fallin Hunzaker, M.B., Lee, J., Mann, M., Merhout, F., Volfovsky A., Exposure to opposing views on social media can increase political polarization, "Proceedings of the National Academy of Sciences" 2018, 115 (37), https://doi.org/10.1073/pnas.1804840115.

Bawden, D., Robinson, L., The dark side of information: overload, anxiety and other paradoxes and pathologies, "Journal of Information Science" 2009, 35(2),, https://doi.org/10.1177/0165551508095781.

Bennett, W.L., Livingston, S., The disinformation order: Disruptive communication and the decline of democratic institutions, "European Journal of Communication" 2018, 33(2), https://doi.org/10.1177/0267323118760317.

Bishop, S., Influencer Management Tools: Algorithmic Cultures, Brand Safety, and Bias. Social Media + Society, 2021, 7(1), https://doi.org/10.1177/20563051211003066.

Byman, D.L., Gao, C., Meserole, C., Subrahmanian, V.S., Deepfakes and International Conflict. Foreign Politicy at Brookings, 2023.

Casad, B.J. and Luebering, J.E., Confirmation bias. Britannica 2025.

Chesney, R., Citron, D.K., Deepfakes and the New Disinformation War. "Foreign Affairs" 2019, 98(1).

Cialdini, R.B., Goldstein, N.J., Social influence: Compliance and conformity, "Annual Review of Psychology" 2004, 55, https://doi.org/10.1146/ annurev.psych.55. 090902.142015.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F. Scala, A., The COVID-19 social media infodemic. Scientific Reports 2020, 10, 16598. https://doi.org/10.1038/s41598-020-73510-5.

DiResta, R., The Spin Doctors Are In. Guest: Renée DiResta. Center for Human Technology 2020.

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. Johnson, B., The Tactics, Tropes of the Internet Research Agency, Stanford Internet Observatory 2019.

Howard, P.N., Kollanyi, B., Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum. Social Science Research Network 2016.

Downing, T., D-Day deception Operation Fortitude: The World War Two army that didn't exist. BBC 2024. https://www.bbc.com/culture/article/20240531-d-day-deception-operation-fortitude-the-world-war-two-army-that-didnt-exist.

Fallis, D., The Epistemic Threat of Deepfakes. "Philos Technol" 2021, 34(4), doi: 10.1007/s13347-020-00419-2.

Gorwa, R., What is platform governance?. Information, Communication & Society, 2019, 22(6), p. 854–871. https://doi.org/10.1080/1369118X.2019.1573914.

Horton, D., Wohl, R.R., Mass communication and para-social interaction, "Psychiatry: Journal for the Study of Interpersonal Processes" 1956, 19, https://doi.org/10.1080/00332747.1956.11023049.

Jowett, G.S., O'Donnell, V., Propaganda & Persuasion, Thousand Oaks, 2018.

Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Zittrain, J.L., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., The science of fake news, "Science" 2018, 359(6380), https://doi.org/10.1126/science.aao2998.

Lewandowsky, S., Ecker, U.K.H., Cook, J., Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era, "Journal of Applied Research in Memory and Cognition" 2017, 6(4), https://doi.org/10.1016/j.jarmac.2017.07.008.

Marwick, A., Lewis, R., Media manipulation and disinformation online, "Data & Society", 2017.

McDougall, J., Media Literacy versus Fake News: Critical Thinking, Resilience and Civic Engagement, "Media Studies" 2019, 10(19).

McIntyre, L., Post-Truth, "European Journal of Communication" 2018, 33(5), https://doi.org/10.1177/0267323118799184.

McMahon, L., Kleinman, Z., Subramanian, C., Facebook and Instagram get rid of fact checkers, BBC 2025, https://www.bbc.com/news/articles/cly74mpy8klo.

Rød, B., Pursiainen, C., Eklund, N., Combatting Disinformation – How Do We Create Resilient Societies? Literature Review and Analytical Framework, "Eur J Secur Res" 2025, https://doi.org/10.1007/s41125-025-00105-4.

Saul, F., Bewegung in Leichensäcken – dieser Fake-Clip wird seit Kriegsbeginn immer wieder verbreitet, Stern 2023.

Simonite, T., A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be. WIRED, 2022. https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/.

Spohr, D., Fake news and ideological polarization: Filter bubbles and selective exposure on social media, "Business Information Review" 2017, 34(3), https://doi.org/10.1177/0266382117722446.

Tandoc, E.C., Lim, D. Ling, R., Diffusion of disinformation: How social media users respond to fake news and why, "Journalism" 2019, 21(3), https://doi.org/10.1177/1464884919868325.

Tikkanen, A., Donation of Constantine, Britannica 2002.

Tufekci, Z., Facebook Said Its Algorithms Do Help Form Echo Chambers, and the Tech Press Missed It, "New Perspectives Quarterly" 2015, 32(3), https://doi.org/10.1111/npqu.11519.

van der Linden, S., Roozenbeek, J., Compton, J., Inoculating Against Fake News About COVID-19, "Front Psychol" 2020, 23:11:566790. https://doi.org/10.3389/fpsyg.2020.566790.

Victor, D., "Access Hollywood" Reminds Trump: "The Tape Is Very Real", The New York Times 2017. https://www.nytimes.com/2017/11/28/us/politics/donald-trump-tape.html.

Wardle, C., Derakhshan, H., Information Disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe, 2017. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html.