

Weaponising Uncertainty:

Disinformation, Hybrid Threats
and Democratic Governance in the
Baltic Sea Region





Weaponising Uncertainty:

Disinformation, Hybrid Threats and Democratic Governance in the Baltic Sea Region

Working Group on Digital Literacy and Democracy, Baltic Sea Region Youth Forum

Published: 29.04.2026

1. Introduction

In September 2022, explosions tore through **the Nord Stream pipelines in the Baltic Sea**. Before investigations could establish facts, the information space was already saturated with competing narratives: Western sabotage, Russian self-destruction, covert operations and elaborate conspiracies. These claims spread rapidly across social media, state-affiliated outlets and alternative platforms, often contradicting one another yet reinforcing a shared outcome - confusion, distrust and political tension. The incident was not only a geopolitical flashpoint; it was a clear demonstration of how disinformation now operates in real time, forming public perception as events are still unfolding.

This reality is no longer exceptional. Disinformation has become a defining feature of the contemporary information environment. Digital platforms allow content to circulate instantly across borders, while algorithmic systems amplify material that provokes strong emotional reactions, regardless of its accuracy. At the same time, advances in artificial intelligence have lowered the cost and increased the sophistication of producing false or manipulated content. As a result, disinformation today is faster, more scalable and more difficult to detect than ever before.

Young people find themselves at the centre of this challenge in a way that is often misunderstood. The assumption that growing up online confers resilience to manipulation does not hold up to scrutiny. Digital fluency is not the same as critical media literacy and the very platforms that young people rely on for news, political information and social connection are structurally designed to reward engagement over accuracy. At the same time, this generation is not simply passive. Young people are increasingly active participants in countering disinformation, as fact-checkers, civic educators and advocates for platform accountability. Their role is both a vulnerability and a source of resilience.

This policy paper brings together analytical perspectives on the nature, spread and impact of disinformation, with a particular focus on democratic governance and youth



agency. The paper proceeds in four main parts. **Section 2** establishes a conceptual foundation, defining disinformation and distinguishing it from related forms of information disorder, before tracing its historical evolution from pre-digital propaganda through to today's AI-generated synthetic media. **Section 3** examines the strategic intentions behind disinformation campaigns and the mechanisms through which they operate and the particular vulnerabilities of the Baltic Sea region as a high-stakes information environment. **Section 4** turns to the implications for democracy, exploring how disinformation corrodes electoral integrity, undermines media ecosystems, fuels political cynicism and deepens societal polarisation. **Section 5** surveys the landscape of responses, from EU regulatory frameworks and platform governance to civil society initiatives and youth-driven media literacy programmes and assesses what a genuinely effective, coordinated response might look like.

The urgency of this work should not be understated. Disinformation is not a communication problem that better technology will eventually solve. It is a structural challenge to the conditions that make democratic life possible, shared facts, institutional trust and an informed public. Meeting that challenge requires understanding it clearly, which is what this paper sets out to do.

2. What Is Disinformation?

When we hear the word 'Disinformation' today we usually think of false or misleading information that was deliberately created and spread online or in the news. Within policy and research communities, intentionality is the defining feature because here, disinformation is not simply incorrect content, but a strategic tool used to influence public perception, disrupt decision-making, or undermine institutional trust (Wardle & Derakhshan, 2017).

To avoid conceptual ambiguity, it is important to distinguish disinformation from related forms of information disorder. Often these two words are thrown around in the public and by the media, *Misinformation* and *Malinformation*. Misinformation refers to false information shared without intent to mislead, often circulating organically through social networks or interpersonal communication (Lewandowsky et al., 2017). Malinformation, on the other hand, involves the strategic use of genuine information to cause harm. This could for example happen by releasing private data or presenting truthful content in a distorted or decontextualized manner (Bennett & Livingston, 2020).

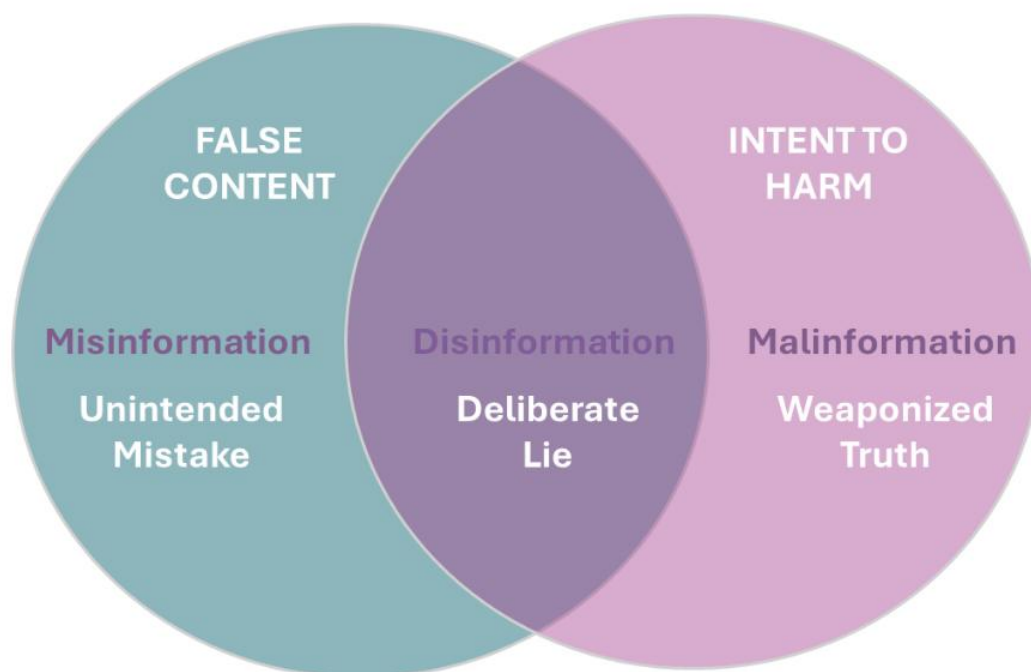


Figure 1, (BSRYF Working Group Digital literacy and Democracy, 2025)

For the purposes of this paper, we define disinformation as intentionally false or manipulated information produced and distributed to mislead audiences and shape perceptions or behaviour. This definition provides a background for looking into how disinformation has evolved throughout different eras and generations and how it functions within contemporary information environments. We will also take a look at GenZ's perspective and strategies on how to deal or combat disinformation in their communities.

2.1 Forms and Types of Disinformation

Today we see that disinformation appears as fake stories that imitate legitimate news, manipulated visuals such as edited photos or videos, imposter content that mimics credible institutions, or misleading framing that presents accurate facts without necessary context. We have all seen examples of these just by turning on our smartphones whether we see something shared on Meta by our grandparent's generation or something shared on an Instagram story by a friend. These formats vary in quality but share the common goal of distorting public understanding.

How these are spread relies on different types of distribution mechanisms. Some campaigns are coordinated efforts by state or non-state actors, while others rely on bot networks that artificially amplify content. Disinformation may also be disseminated through influencer or proxy accounts, which lend credibility to false narratives, or through anonymous forums that enable untraceable and rapid circulation.



However, throughout the years disinformation has been used for different strategic serves purposes. We have seen it being used to influence political processes, to create financial profit through scams or, promote ideological agendas during elections, or deepen social polarization. These ways of influence often overlap, but we will take a look into what defined disinformation in the different eras in history next.

Over time, the evolution of disinformation has reached a rapid growth as the form of digital communication and social media platforms has transformed how information is perceived. The current scope of technology has enabled the wildfire spread of disinformation as it can be produced, shared, and consumed anywhere at any time. In this context, several types of disinformation have emerged, highlighting their problematic nature over the past few years.

2.1.1 False Content

The inaccurate information that is designed to deceive through rapidly spreading on social media is known as False Content. This refers to information that is entirely incorrect but presented as factual and is fabricated to represent the most extreme form of information distortion by the creation of entirely altered narratives, data, and visual assets.

False content denotes information that is completely inaccurate yet portrayed as true. This information can come from mistakes, misinterpretations, or deliberate efforts to disseminate falsehoods. For younger audiences who often depend on social media as a main news source, the likelihood of encountering false information is notably elevated. Research indicates that younger individuals are more prone to discovering news via peer networks instead of conventional media outlets, which heightens the risk of facing false information (Allcott & Gentzkow, 2017).

2.1.2 Misleading Content

Information that contains factual elements but omits key context and often exaggerates details and data to be framed in a biased manner. According to Wardle (2017), this is by far the most difficult kind of fake news to uncover since there are many misrepresented facts and news articles that can easily be accessed, especially on social media. This then leads to media outlets distorting reality with selective reporting that can shape public perceptions in relation to complex issues.

Wardle and Derakhshan (2017) highlight that deceptive content is especially influential as it alters perception instead of fabricating information. For younger audiences still honing their media literacy abilities, telling apart accurate and misleading narratives can prove difficult. Influencers, bloggers, and online figures—many of whom hold considerable sway over young audiences—might unknowingly propagate inaccurate understandings of political or social matters.



2.1.3 Manipulated Content

Manipulated content describes authentic information that has been changed to deceive viewers. This might consist of altered images, edited videos, or modified audio recordings. Technological progress like artificial intelligence has facilitated the creation of highly realistic altered media, including deepfake videos.

Visual media is pivotal in youth culture, especially on platforms such as TikTok and Instagram, where visuals and videos lead communication. Since visual content is frequently seen as more trustworthy than text, altered images or videos can significantly affect young people's views on events.

From a young person's viewpoint, altered content offers both dangers and possibilities. Although young users might initially have confidence in visual media, they are becoming more adept at utilizing digital tools to assess and confirm online information. Numerous young individuals engage in online groups that recognize altered media and reveal deceptive visual content

2.1.4 Satire and Parody

Satire and parody denote creations that deliberately amplify or misrepresent reality for comedic effect or societal critique. In contrast to other types of misinformation, satire aims not to mislead audiences. Nonetheless, it can occasionally be misunderstood when communicated beyond its initial context.

Young viewers often interact with satirical material via memes, parody clips, and humorous observations on the internet. Satirical media can act as a means for young people to evaluate political leaders, organizations, and societal standards. Platforms such as TikTok and X (formerly Twitter) have emerged as significant venues for the distribution of satirical political commentary among younger audiences.

The swift dissemination of online material can obscure the lines between satire and accurate journalism. When satirical content is circulated without context, some younger users might perceive it as genuine news. Consequently, media literacy education is essential for assisting young people in differentiating humor, commentary, and factual content

2.1.5 Propaganda

Propaganda denotes information aimed at swaying public perception and advocating for a specific political ideology or agenda. In contrast to misinformation that can disseminate accidentally, propaganda is typically systematic and deliberate. Governments, political entities, and advocacy groups frequently deploy propaganda to influence narratives and rally public backing.

Young individuals are frequently primary targets of propaganda efforts as they constitute a large segment of the digital community and are crucial in influencing future political environments. Digital platforms allow political figures to customize propaganda messages for younger audiences using targeted ads, viral content, and collaborations with influencers.



Jowett and O'Donnell (2019) contend that contemporary propaganda functions more through digital communication channels, enabling messages to disseminate quickly within peer groups. From a young person's viewpoint, propaganda can shape political views, voting habits, and interpretations of global occurrences. Nonetheless, numerous young individuals are increasingly recognizing propaganda methods and are proactively challenging politically driven narratives on the internet

2.2 Historical Evolution of Disinformation

2.2.1 Pre-Digital Foundations: Propaganda and Information Control

Disinformation predates the internet and has long been used as a strategic political tool. Long before the digital era, disinformation functioned as a tool of statecraft. In pre-digital contexts, information manipulation was characterized by centralized production, elite control, and limited distribution channels. Governments used controlled media environments to shape public opinion, suppress dissent, and mobilize populations.

Wartime propaganda during World War II illustrates this dynamic. Posters, radio broadcasts, and newsreels were deployed to influence morale, demonize adversaries, and maintain public support for military objectives (Burke, 2016). Although slower and more constrained than modern disinformation, these efforts demonstrate the longstanding strategic value of information manipulation.

2.2.2 Early Internet Era (1990s–mid-2000s): Decentralized Information

The emergence of the internet fundamentally altered the information landscape. As institutional gatekeeping declined, individuals gained the ability to publish content globally with minimal oversight. This shift created new vulnerabilities, as verification mechanisms were limited and early online communities became fertile ground for unverified claims.

During this period, chain emails circulated fabricated warnings about computer viruses or public health threats, while early conspiracy websites promoted claims that major events were staged or manipulated. Hoax stories spread across message boards and forums, demonstrating how decentralized digital communication enabled false narratives to reach large audiences quickly.

2.2.3 Social Media Era (mid-2000s–late 2010s): Viral Disinformation

The rise of social media transformed disinformation from isolated falsehoods into rapidly spreading network phenomena. Platform algorithms prioritized engagement over accuracy, enabling sensational or polarizing content to gain disproportionate visibility.

This era saw the emergence of coordinated troll campaigns, such as those linked to the Internet Research Agency, which targeted political discourse across multiple countries (Bradshaw & Howard, 2019). The Cambridge Analytica scandal revealed how personal data could be used to micro-target political messaging. Meanwhile,



false health claims circulated widely during public crises, illustrating how viral disinformation could undermine public safety and institutional trust.

2.2.4 Contemporary Era (2020s–): Synthetic and Hybrid Disinformation

Recent technological developments have ushered in a new phase of disinformation characterized by synthetic media, AI-generated content, and hybrid influence operations. These tools enable actors to blend authentic and fabricated material in ways that are increasingly difficult to detect.

Deepfake videos depicting public figures making fabricated statements, AI-generated news sites designed to resemble legitimate outlets, and synthetic voice scams imitating real individuals illustrate the sophistication of contemporary disinformation (Ferrara, 2020). These developments have expanded the operational capabilities of both state and non-state actors, creating new challenges for policymakers and information security practitioners.

2.3 Summary on cross-regional narratives

Beyond this typography, we can understand disinformation through their narrative framing. In our region of the Baltic Sea States, cross-regional narratives that cut across national contexts while adapting to local conditions can be observed. These narratives are not isolated claims but form interconnected discursive frameworks that reinforce one another. The following narratives have been identified through the country factsheets from Baltic Sea States compiled by the EU Disinfo Lab as well as through the EU vs. Disinfo Database, which chronicles instances of disinformation.

A central cluster of narratives relates to security, particularly in relation to Russia and NATO. One dominant claim portrays NATO as an aggressive and occupying force that undermines national sovereignty, with member states depicted as subordinated to United States interests. This narrative is frequently accompanied by assertions that public support for NATO is artificially manufactured, that neutrality would provide greater security, and that sanctions against Russia are self-defeating. Closely linked is the countervailing narrative that Russia acts purely defensively, responding to Western provocation and NATO expansion. Within this framing, responsibility for conflict, especially in Ukraine, is shifted toward Western actors, while Ukraine itself is delegitimized as corrupt or externally controlled. Together, these narratives invert conventional security discourses and seek to erode public support for Euro-Atlantic institutions.

A second thematic grouping concerns governance and institutional trust. Here, disinformation frequently targets mainstream media and democratic institutions. Established media outlets are framed as instruments of propaganda, subject to elite or foreign control, while alternative platforms are elevated as the sole sources of “truth.” This narrative often intersects with election-related disinformation, public health misinformation (notably during the COVID-19 pandemic), and claims that dissenting or pro-Russian viewpoints are systematically censored in European societies. In parallel, democratic institutions including governments, courts, and electoral systems are depicted as corrupt, illegitimate, or manipulated. Elections are framed as rigged, and public authorities as acting against the interests of citizens.



These claims are commonly reinforced by populist rhetoric, allegations of systemic corruption, and broader anti-democratic sentiment.

A third set of narratives revolves around culture and identity, particularly in relation to migration. Migrants and refugees are frequently portrayed as security threats, sources of crime, or agents of cultural erosion. Governments and supranational actors such as the EU are accused of deliberately engineering migration flows against the will of the population. These narratives often draw on conspiracy theories such as the “great replacement” and are amplified through fabricated or misleading accounts of criminal activity.

Economic dimensions of disinformation intersect with governance and identity concerns, particularly in the narrative that certain states, especially those with historical ties to the former Soviet Union, are “failed” or artificial. Such claims emphasize economic decline, demographic crisis, and dependence on external actors, sometimes extending to the delegitimization of statehood itself. These narratives are frequently coupled with anti-EU sentiment, nostalgia for the Soviet period, and criticism of contemporary policy initiatives, including energy transitions and infrastructure projects, which may also connect to climate-related disinformation.

Across these thematic areas, an overarching meta-narrative of conspiratorial thinking plays a unifying role. This worldview posits that political leaders, media organizations, international institutions, and economic actors form a coordinated “hidden elite” that acts against the interests of ordinary citizens. Major societal developments ranging from war and migration to public health measures and economic challenges are interpreted as deliberate and interconnected outcomes of elite manipulation. Rather than appearing as a single explicit claim, this conspiratorial logic operates as a connective framework that links otherwise distinct narratives into a coherent belief system. It is often sustained by populist messaging, antisemitic tropes, and narratives about the moral decline of modern societies.

Several key observations emerge from this analysis. First, while individual narratives can be analytically distinguished, they are highly interlinked and mutually reinforcing. Second, they can be broadly categorized into the thematic domains of security, governance and institutional trust, economy, and culture/identity, yet they frequently overlap in practice. Third, despite the presence of cross-regional patterns, significant national variation persists: for example, stronger emphasis on state legitimacy narratives in post-Soviet contexts, comparatively lower levels of election-related disinformation in countries such as Sweden, and a pronounced focus on health-related disinformation in Germany. Fourth, disinformation in the Baltic Sea region is fundamentally narrative-driven rather than purely event-driven, relying on durable storylines that are reactivated and adapted to specific circumstances. Finally, both academic research and civil society initiatives in the region have predominantly focused on Russian influence and disinformation campaigns, shaping the analytical framing of the phenomenon itself.



3. The Strategic Logic and Operational Architecture of Disinformation

Intentions, Mechanisms, and the Architecture of Manipulation in the Baltic Sea Region

The digital domain has become a primary arena of public and civic life; a space of connection, knowledge exchange, and political deliberation. It is also, increasingly, a battlefield.

Younger generations across the Baltic Sea Region and beyond rely on social media as a principal source of information and political orientation (Laaninen & Bol, 2025). Yet the very centrality of the digital environment to democratic life has made it a target: malign state and non-state actors systematically deploy disinformation to confuse, polarise, and manipulate target populations, exploiting it as a hybrid instrument of considerable strategic potency.

The weaponisation of the information space has become a defining feature of contemporary international conflict, reflecting intensifying confrontation among major geopolitical powers. Disinformation forms part of a broader hybrid arsenal, beyond a simply a communication challenge but as a structural dimension of modern security competition with tangible political consequences.

Disinformation campaigns are evolving as rapidly as the technologies that enable them. While their disruptive potential is well-documented, understanding of how such operations are strategically designed and operationalised remains insufficient. Coordinated policy responses and durable societal resilience must be built on a clear analytical foundation that addresses the strategic intentions of information operations and the technical and social infrastructure through which they propagate. The implications of these dynamics for democratic systems and target populations are examined in the following chapter.

3.1 Strategic Intentions Behind Disinformation

Disinformation is never deployed without purpose. Its use reflects a set of discrete but interrelated strategic objectives, each targeting a different dimension of democratic society, typically pursued in parallel and mutually reinforcing.

3.1.1 Societal Destabilisation and Polarisation

Among the most consistently documented objectives is the deliberate exploitation of pre-existing social fault lines. Disinformation does not manufacture division from nothing; it amplifies, sharpens, and weaponises cleavages — ethnic, political, economic, or historical — that already exist within target societies. By feeding narratives that reinforce extremist positions and deepen mutual distrust, coordinated campaigns erode the social cohesion on which democratic governance depends (Rød et al., 2025). In the Baltic Sea region, this has manifested in content targeting minority communities, national identity anxieties, and unresolved historical grievances.

3.1.2 Erosion of Institutional Trust

A related but analytically distinct objective is the systematic degradation of public confidence in democratic institutions (EEAS, 2026). Electoral systems, public health agencies, judicial bodies, and governments are targeted not primarily to advance specific policy outcomes, but to cultivate a generalised sense that institutions are corrupt, captured, or incompetent. The media sector has been a particular focus: disinformation campaigns against journalists and news organisations deepen existing distrust, producing an accelerating disorder in the information environment that democratic publics depend on for a shared sense of reality. A population that has been conditioned to distrust its institutions is poorly positioned to defend them.

3.1.3 Electoral and Political Interference

Disinformation campaigns are rarely diffuse. They are calibrated to specific high-stakes events — electoral cycles, public health emergencies, civil unrest, or breaking news — that provide fertile conditions for information manipulation. Emotionally charged, uncertain moments are treated by threat actors as structural vulnerabilities: windows during which cognitive biases are most readily exploited and audiences are most susceptible to influence (EEAS, 2026).

Documented patterns of Russian Foreign Information Manipulation and Interference (FIMI) in recent European elections reveal a consistent three-phase operational logic:

Phase	Timing	Key Tactics
1	Months before election	Consolidate presence; discredit pro-EU or mainstream candidates through allegations of corruption, health concerns, or foreign subservience.
2	Weeks before election	Expand targeting to state institutions; exploit domestic fault lines — immigration, anti-refugee sentiment, censorship allegations; fabricate imminent-conflict narratives.
3	Final days	Undermine electoral integrity directly: fearmongering about security threats, abstention campaigns, and pre-emptive seeding of doubt about result validity.

Figure 2, (EEAS, 2026)

This playbook, while adapted to local context, is structurally repetitive — and increasingly anticipatable. Its predictability is itself a policy asset: patterns identified across elections in Germany, Poland, and Moldova suggest a common operational template that can be mapped, monitored, and pre-empted through coordinated early-warning mechanisms (EEAS, 2026).

3.1.4 Gray-Zone Strategic Advantage



Perhaps the most analytically significant framing for disinformation in the Baltic Sea security context is its function within gray-zone operations — the repertoire of coercive activities deployed below the threshold of overt military or economic aggression. Disinformation is ideally suited to this environment: it is inexpensive to produce, difficult to attribute, and capable of generating confusion and decision-paralysis that constitute strategic gain without triggering formal retaliation thresholds.

Democracies face a structural asymmetry here. They cannot easily counter information operations without risking violations of the open information norms they are committed to upholding. That asymmetry is not incidental; it is deliberately exploited (Bachmann et al., 2023).

3.2 Mechanisms of Dissemination: Infrastructure and Amplification

Understanding why disinformation spreads requires attention not only to its content, but to the technical and social infrastructures through which it travels; infrastructures that are, in important respects, structurally biased toward amplification.

Two dynamics are central: the cognitive vulnerabilities that make individuals susceptible to false content, and the platform architectures that convert individual susceptibility into systemic spread. Confirmation bias and perceived believability are among the primary mechanisms through which disinformation is consumed and propagated (Kim & Dennis, 2019; Moravec et al., 2020). In highly polarised environments, individuals are more inclined to consume and credit information from sources aligned with their pre-existing beliefs or political identity (Weismueller et al., 2024). The intentionality of spread is thus closely linked to believability: those who find content plausible are likely to pass it on, compounding its reach with each iteration.

The architecture of a disinformation campaign involves at minimum two actors: the sender and the receiver. Creators intentionally produce and distribute false content in pursuit of a strategic goal. The reach and impact of that content, however, depends heavily on receivers, users who, whether knowingly or not, amplify it through engagement. On social media, liking, commenting, and sharing function as amplification mechanisms regardless of intent: false information has been demonstrated to spread faster and further than accurate information (Vosoughi et al., 2018). Campaigns are frequently designed to exploit this dynamic, deploying bots and paid participants to seed disinformation into information networks, where unwitting users carry it forward.

This distinction between intentional and unintentional spread is consequential for policy design. Those who recognise the falsity of content and share it deliberately, deceivers, represent a different challenge from those who spread it in good faith (i.e. misleaders), or from those who engage with content reflexively without evaluating its veracity at all. French, Storey, and Wallace (2024) term this third category mindless redistribution: engagement that lacks deceptive intent but produces deceptive effect. While eliminating the production of disinformation may be impossible, reducing its unintentional amplification by unwitting participants is both feasible and consequential.

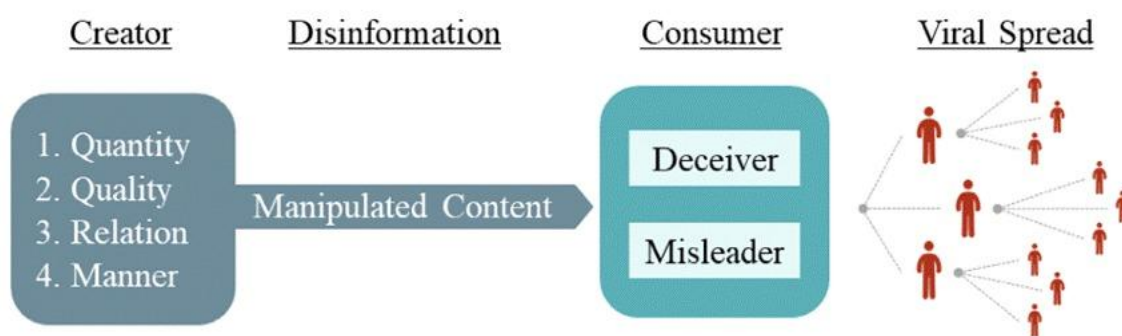


Figure 3, (French, Storey, and Wallace, 2024)

3.2.1 Algorithmic Amplification and Platform Architecture

Social media platforms optimised for engagement systematically privilege content that triggers strong emotional responses, outrage, fear, disgust, regardless of its accuracy. Recommendation systems designed to maximise attention metrics elevate emotionally charged posts, while monetisation models reward divisive content that sustains user activity. Research consistently finds that attention-driven architectures promote polarising and misleading material, frequently elevating it above verified information (Tufekci, 2015; Vosoughi et al., 2018). Disinformation is therefore not an anomaly within these systems; it is, in many respects, a predictable output of their design logic.

This structural dynamic is compounded by the opacity of platform governance. Content moderation remains inconsistent, underfunded, and unevenly applied across languages and regions, a particular vulnerability in the linguistically diverse Baltic Sea context. Limited transparency over algorithmic design, moderation decisions, and political advertising creates conditions in which coordinated inauthentic behaviour can operate with limited detection risk.

3.2.2 AI-Enabled Disinformation at Scale

Artificial intelligence has fundamentally transformed the scalability of disinformation operations, rendering them cheaper, faster, and substantially harder to detect. Threat actors now deploy AI tools across the full production and distribution pipeline. Written content is generated at scale using near-identical templates, then translated and adapted across multiple languages simultaneously. AI-generated articles are disseminated through fabricated news sites; synthetic posts and comments are propagated by Coordinated Inauthentic Behaviour (CIB) networks.

The use of synthetic audio and video has moved beyond experimentation to become a standard operational technique. Advanced voice-cloning enables the credible impersonation of political figures. AI-generated imagery was used extensively during Germany's 2025 legislative elections to portray apocalyptic visions of German cities. These are not edge cases; they represent an operational norm that is rapidly diffusing.

Large Language Models are themselves becoming targets of manipulation. The Russian FIMI infrastructure Portal Kombat is suspected of conducting what analysts



term LLM grooming, flooding the information space with low-quality multilingual content designed to corrupt AI training data and inject false claims into the outputs of widely used AI tools. As citizens and institutions increasingly rely on AI-generated synthesis for information, the integrity of those systems becomes a first-order information security concern. The arms race between disinformation production and detection is accelerating and detection is currently losing ground.

3.3. The Baltic Sea Region: A High-Stakes Information Environment

The Baltic Sea region presents a distinctive configuration of vulnerabilities that amplifies the strategic relevance of disinformation — and the urgency of a coordinated regional response.

The region's exceptionally high digital connectivity expands the exposure surface available to information operations while simultaneously enabling rapid cross-border amplification of disinformation content. Linguistic diversity — combined with the presence of Russian-speaking minority communities in Estonia, Latvia, and Lithuania — has historically enabled threat actors to tailor campaigns to specific audiences with a precision that generic content moderation frameworks struggle to address. These are not incidental vulnerabilities; they are structural features of the regional information environment that adversaries have demonstrated both the willingness and the capability to exploit.

Russia's doctrine of information warfare is well-documented, and its operations in the Baltic Sea region have been the subject of extensive governmental and academic analysis. Disinformation in this context cannot be treated as a generic governance challenge: it is simultaneously a security challenge requiring inter-agency coordination, and a democratic challenge requiring responses calibrated to protect civil liberties and open information norms. These two imperatives are not always aligned, and navigating the tension between them is itself a governance task of considerable complexity.

Challenges related to the governance of global digital platforms — their inconsistent moderation practices, limited transparency, and commercial incentive structures — create structural vulnerabilities that national regulation alone cannot adequately address. The transnational character of both the platforms and the threat actors means that effective response requires regional coordination and frameworks. The question is whether that institutional potential can be mobilised with sufficient coherence and speed to match the pace of an evolving threat landscape.

4. Democratic Implications and Systemic Effects of Disinformation

While Chapter 3 examined the strategic intentions and operational mechanisms of disinformation, this chapter turns to its implications for democratic systems.

The effects of disinformation on democratic processes are widely discussed, yet remain difficult to precisely identify and measure. Establishing clear empirical causal relationships between disinformation and specific democratic outcomes presents a



significant methodological challenge, particularly given the complex, multi-causal nature of political behavior and institutional dynamics.

Despite these limitations, it is possible to theorize the implications of disinformation for democratic systems. A range of plausible effects can be identified that are supported, at least in part, by emerging empirical evidence. This chapter adopts a theoretically informed approach, focusing on how disinformation reshapes democratic processes across epistemic, behavioural, institutional, and societal dimensions.

4.1 Epistemic Effects: Distrust and Uncertainty

Disinformation poses a fundamental challenge to the way we generate and evaluate political information and communication itself. At its core, democracy depends on a foundation of shared narratives that enables citizens to form opinions and make informed political decisions. As Colomina et al. (2021) noted, “information and shared narratives are a precondition for good quality democratic public discourse.” The erosion of such shared narratives creates an environment in which competing and often incompatible versions of reality coexist, making consensus and deliberation increasingly difficult. As Mansbridge et al. (2012) define it, the epistemic function of deliberative systems “is to produce preferences, opinions, and decisions that are appropriately informed by facts and logic and are the outcome of substantive and meaningful consideration of relevant reasons.” (p.11). When disinformation distorts the informational environment, this function is compromised, limiting the capacity of democratic systems to generate informed and reasoned outcomes. Importantly, the effects of disinformation are not limited to those who fully believe false claims. Disinformation beliefs can linger and impact individuals’ attitudes even when they recognize them as false or have been exposed to a debunking or fact-checking (Arcos et al., 2022). In addition, pre-existing beliefs are vital in understanding the effect of disinformation narratives on individuals, as such beliefs shape how information is received, interpreted, and integrated into existing worldviews (ibid.).

Taken together, these dynamics contribute to a broader epistemic breakdown in which trust in political information and communication is diminished and certainty about political facts becomes increasingly elusive.

4.2. Effects on Core Democratic Domains

4.2.1 Electoral Processes

The electoral process represents a central site in which the potential harmful effects of disinformation are discussed. As the primary mechanism through which democratic will is expressed, elections depend on the ability of citizens to make informed decisions. In this sense, the electoral arena is not only a target of disinformation, but also a key domain in which its consequences unfold.

Disinformation can undermine this capacity by distorting the informational environment in which voters form their preferences (Johnson and Favour-Benson, 2024). One of the most significant concerns in this regard is election interference, understood not only as technical manipulation but also as the strategic shaping of



public opinion. Through manipulation of the voters' will and opinions, disinformation campaigns in elections can influence these elections; the discourse on disinformation and the evidence of these campaigns furthermore sow distrust in the legitimacy of the electoral process and its capability to represent a freely formed will of the citizenry (Colomina et al., 2021). Large scale Russian disinformation campaigns during the 2016 US presidential election, have brought election interference through disinformation into the mainstream discourse, showing the potential dangers of such campaigns on electoral decision making.

4.2.2. Media Ecosystem

The media ecosystem constitutes a central arena where the epistemic function of democratic systems is performed. As an intermediary between political institutions and the public sphere, it plays a crucial role in structuring how information is produced, disseminated, and interpreted. As discussed in Chapter 3.1.2 the media ecosystem therefore serves as a primary target to erode institutional trust. McKay and Tenove (2021) view deliberative democracy and disinformation through the lens of media ecosystems mediating discourse between the public sphere and the institutional realm. Within this framework, the integrity of democratic deliberation depends heavily on the structures and dynamics of the media environment.

Social media is part of this media ecosystem, and its growing prominence has fundamentally altered how information circulates. As outlined in Chapter 3.2, platform architectures and algorithmic structure of social media platforms amplify vulnerability to disinformation and enable the spread of fringe disinformation campaigns into broader online discourse, blurring the boundaries between marginal and mainstream political discourse.

At the same time, key features of online communication further complicate users' ability to evaluate information. The anonymity and pseudonymity of social media postings obfuscate consumers' ability to assess source credibility, as different types of content are presented in similar formats (ibid.). This contributes to what can be understood as a form of epistemic challenge, in which traditional cues for judging reliability are weakened or disappear altogether.

Finally, existing cognitive bases interact with disinformation and compound its effects on the media landscape. The "hostile media phenomenon" can lead individuals to perceive even balanced or independent journalism as biased against their own views, driving them toward alternative sources that reinforce their positions. This contributes to increased polarization and segmentation within society (Arcos et al., 2022). Moreover, this dynamic can be strategically exploited by foreign or coordinated actors to portray the media environment as systematically biased, a perception that can extend to debunkers and fact-checkers, who may themselves be dismissed as politically partial (ibid.). In this way, disinformation can undermine the credibility of traditional media, weakening its role as a trusted source of information within a democratic society (Colomina et al., 2021).

Taken together, these dynamics illustrate how disinformation not only introduces false content into the media ecosystem, but also transforms the conditions under which information is evaluated. In doing so, it deepens epistemic uncertainty and further erodes the foundations necessary for democratic discourse.

4.2.3 Behavioural Effects: Political Cynicism and Disengagement

Disinformation not only distorts information environments but also reshapes how individuals relate to knowledge and political communication more broadly. Beyond the risk of believing false claims, epistemic issues arise through the deliberate targeting and discrediting of trusted sources such as professional journalism, scientific institutions, and government bodies (McKay and Tenove, 2021). This contributes to the emergence of political cynicism, in which individuals come to view the entire political process as biased, strategically motivated, and ultimately unreliable (ibid.).

In such an environment, the distinction between credible and non-credible information becomes increasingly blurred. Rather than encouraging more critical engagement, this can lead to withdrawal from political information altogether. Empirical findings suggest that exposure to large amounts of conflicting information can reduce news consumption (Mitchell et al., 2019), indicating a disengagement from the informational basis of democratic participation.

These developments can be further understood through fundamental dynamics of human communication. Disinformation campaigns are engineered to exploit well-documented features of human cognition. Confirmation bias — the tendency to evaluate information more favourably when it aligns with pre-existing beliefs — is the most consistently targeted vulnerability. Disinformation packages false claims within narratives that resonate with the ideological priors of specific audiences, making critical evaluation less likely precisely when it matters most.

The illusory truth effect compounds this vulnerability: repeated exposure to a claim increases its perceived credibility, regardless of its accuracy. In a high-volume information environment characterised by constant content recirculation, this mechanism becomes operationally significant. Disinformation need not persuade immediately; sustained presence may suffice to shift credibility thresholds incrementally over time. Furthermore, Truth-Default Theory posits that individuals generally assume others are truthful unless credibility is actively undermined (Arcos et al., 2022). Attempts to reduce this default assumption and increase vigilance can in turn lead to generalized and chronic distrust, extending even to legitimate sources of epistemic authority such as government institutions and professional journalism (ibid.). The broader discourse surrounding disinformation plays a critical role in this process. Heightened awareness of disinformation may deepen selective trust rather than foster open evaluation. In this way, disinformation discourse can amplify epistemic cynicism, transforming healthy skepticism into a more pervasive reflexive distrust.

This dynamic has significant implications for democratic deliberation. If citizens no longer assume that communication is generally truthful, but instead approach all information with suspicion, they may disengage from the political process altogether.

4.3 Institutional Effects: Democratic Legitimacy

The dynamics outlined above have broader implications for democratic legitimacy. As trust and epistemic stability are progressively undermined across multiple domains, the perceived legitimacy of democratic systems as a whole comes into



question. The cumulative effects of distorted information environments, weakened media structures, and increasing political cynicism extend beyond isolated issues and instead affect the foundational relationship between citizens and democratic institutions.

Disinformation creates an environment of eroding trust in public institutions, including governments, parliaments, courts, their respective processes, public figures, journalists, and the free media (Colomina et al., 2021; Johnson and Favour-Benson, 2024). As confidence in these actors declines, so too does the willingness of citizens to accept their authority, decisions, and outputs as legitimate. This erosion of trust poses a fundamental challenge to democratic governance, which relies not only on formal procedures but also on the belief that these procedures are fair, credible, and representative.

4.4 Societal Effects: Polarization and Social Fragmentation

The loss of a shared narrative foundation and “truth” as well as an increased distrust in the legitimacy of democratic processes to articulate the actual will of its citizenry is inherently linked to societal polarization. Polarization can be a consequence of distorted information environments and drive democratic instability.

A central dimension of polarization concerns the erosion of social cohesion and shared narratives. As discussed in the previous section, democratic discourse depends on at least a minimal level of shared understanding. When this common ground deteriorates, citizens are increasingly divided not only in their opinions, but in their perceptions of reality itself. This fragmentation creates conditions in which collective deliberation becomes more difficult and political conflict more entrenched. Malign actors purposefully exploit existing cleavages (see Chapter 3.1.1) and use disinformation to actively reinforce and deepen societal divides. In effect, disinformation is closely linked to hate speech, incitement of on- and offline violence, and discrimination (Johnson and Favour-Benson, 2024; Jahr; Colomina et al., 2021). Disinformation campaigns often target ethnic or social groups, such as migrants or certain religious groups. This has consequences beyond political discourse and has led to documented cases of ethnic violence (ibid.). Disinformation also contributes to broader patterns of political radicalization, which in turn can positively impact heightened political violence and domestic terrorism (Arcos et al., 2022).

Beyond these direct effects, disinformation harms the ethical functions of democracy by weakening mutual respect among citizens and contributing to “affective polarization” between social groups (McKay and Tenove, 2021). In this context, political opponents are increasingly perceived not simply as adversaries, but as fundamentally opposed or even illegitimate (Johnson and Favour-Benson, 2024). The relationship between these dynamics and digital communication is particularly important: The interaction between social media behavior and affective polarization has been described as “techno-affective polarization” (McKay and Tenove, 2021: 709), emphasizing how technological infrastructures amplify emotionally charged divisions.

These processes ultimately feed back into core democratic mechanisms. By polarizing the citizenry of democratic societies, disinformation undermines the



electoral process (Colomina et al., 2021), not only by shaping preferences but by eroding the conditions necessary for fair and widely accepted political competition. In this sense, polarization represents both a consequence of disinformation and a pathway through which its broader democratic harms are realized.

4.5 Disinformation as Systemic Risk

The evidence for dis- and misinformation's systemic reach is concrete and global. Internationally, viral rumours contributed materially to the rapid collapse of Silicon Valley Bank in 2023. Election disinformation has fuelled institutional delegitimation and physical unrest — from the January 6th attack on the US Capitol to post-election violence in Jakarta in 2019. In the Baltic Sea region, coordinated information operations have demonstrably influenced electoral outcomes, inflamed minority tensions, and complicated the regional security environment.

These are not isolated incidents. They are observable instances of a recurrent dynamic in which disinformation exploits existing fragilities — political distrust, financial volatility, institutional stress — and converts points of vulnerability into wider destabilisation. Disinformation functions not as a one-off disruption, but as a recursive, self-reinforcing systemic risk: it undermines fragile information infrastructures, destabilises institutions, fractures social trust, and weakens governance capacity. These breakdowns, in turn, amplify the production and circulation of further disinformation, creating feedback loops that demand structural — not episodic — responses.

What makes this framing consequential for policy is its implication for response design. Systemic risks do not yield to episodic fixes. The defining feature of systemic risk is not its scale, but its relational dynamics: its capacity to propagate through networks, interact with latent vulnerabilities, and push stable systems toward tipping points. Addressing disinformation as systemic risk demands structural responses — to platform architecture, regulatory frameworks, and cross-border coordination — rather than reactive, content-by-content interventions.

Behavioural interventions remain valuable but cannot serve as the primary response. They are too reactive, too slow, and insufficiently attentive to the infrastructures that reward virality over accuracy. Media literacy programmes do not alter the fact that content moderation remains inconsistent, underfunded, and politicised. Information integrity should be treated as a public good rather than an individual responsibility — requiring regulatory frameworks that address design incentives, algorithmic transparency, and platform accountability at the systemic level.

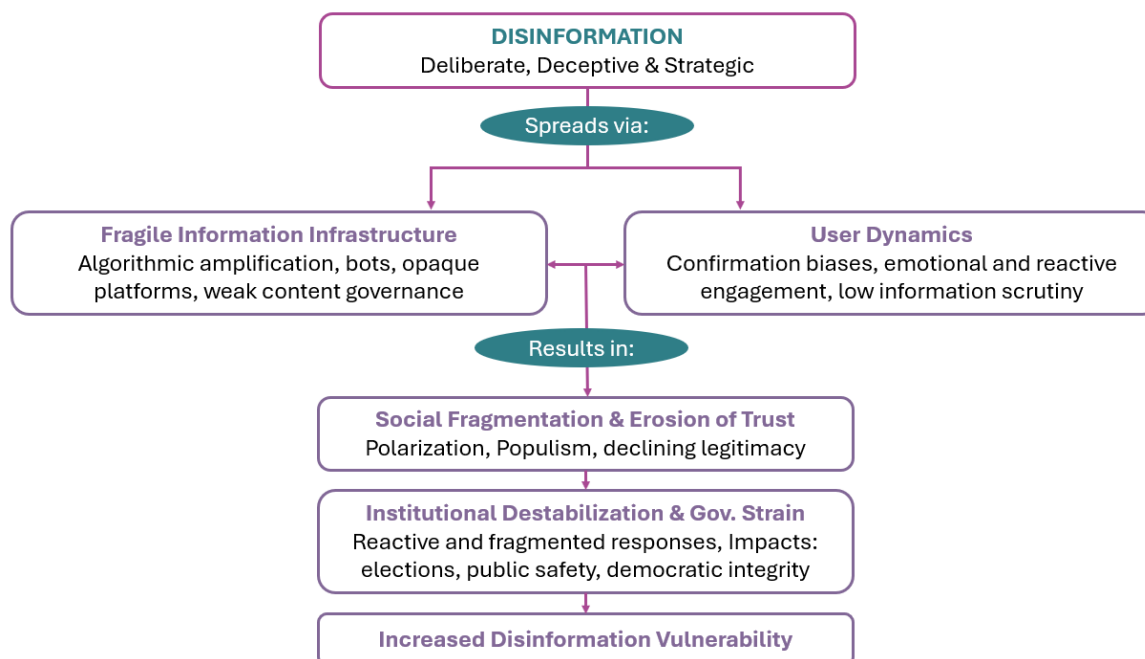


Figure 4, (BSRYF Working Group Digital literacy and Democracy, 2026)

4.6 Particular Vulnerabilities Among Young People

The assumption that digital nativity confers resilience to disinformation is empirically unwarranted. Technical fluency in navigating platforms does not translate automatically into critical information literacy. Young audiences may, in fact, face distinctive structural vulnerabilities. The social dynamics of peer influence operate powerfully on platforms designed around validation metrics — likes, shares, follower counts — that are structurally unrelated to content accuracy. Viral content achieves its reach through engagement, not epistemic merit, and the social desirability of sharing trending material can override critical evaluation.

The dominant formats of contemporary social media — short-form video, image-based posts, ephemeral stories — favour emotional impact over evidential substance, and AI-generated content is specifically engineered to maximise affective resonance within these formats. Well-crafted disinformation can systematically outcompete accurate but less compelling content for audience attention. This is not an argument for restricting young people's access to digital spaces. It is a structural observation that underscores why digital and media literacy must be treated as a policy priority — a complement to, and not a substitute for, platform-level and regulatory intervention.

4.7 Countermeasures and Democratic Trade-offs

While disinformation poses significant risks to democratic processes, efforts to counter it are not without their own challenges. Measures aimed at limiting the spread and impact of disinformation can themselves produce unintended consequences that affect democratic norms and practices. At the core of this issue



lies a fundamental tension between competing democratic principles. Accountability for online content exists between the regulation of mass media by state actors with associated risks such as state capture and violations of free speech and the responsibility of the state to protect citizens' ability to participate in a deliberative public discourse (McKay and Tenove, 2021). This tension reflects a broader challenge: safeguarding the integrity of democratic communication without undermining the freedoms that enable it.

In practice, countermeasures often involve difficult trade-offs. They can infringe upon democratic processes through mechanisms such as interference with internet services, state censorship, restrictions on online speech, and increased barriers for the free media (Colomina et al., 2021). Additional concerns arise with regard to surveillance. Concerns about the harmful effects of disinformation may themselves be exploited by state actors to further encroach on civil liberties (ibid.).

Taken together, these dynamics highlight that responses to disinformation are themselves embedded within the same normative and political tensions that characterize democratic governance more broadly. Addressing disinformation is therefore not simply a technical challenge, but a fundamentally political one, requiring careful balancing between competing values and priorities.

5. Solutions, Best Practices and Youth-Driven Responses

5.1 Introduction: The structural difficulty of responding to disinformation

Taken together, the dynamics outlined above point to the broader systemic implications of disinformation for democratic governance. While the precise causal effects remain difficult to isolate, the cumulative impact on trust, deliberation and institutional legitimacy underscores the extent to which disinformation reshapes the conditions under which democratic processes operate.

In light of these challenges, responses to disinformation must be situated within a broader framework that accounts for the structural, political and societal dimensions. The following section therefore examines a range of policy responses, best practices and youth-driven initiatives directed at strengthening resilience to disinformation.

Developing effective responses to disinformation is difficult because the phenomenon sits at the intersection of technological governance, democratic politics, economic incentives and media culture. Disinformation is not simply false content circulating online; it reflects structural transformations in the public sphere driven by digital platforms, algorithmic amplification and geopolitical competition. The emergence of social media has fundamentally reshaped political communication and public debate, empowering both legitimate actors and anti-democratic forces while disrupting traditional media systems[1]

This structural dimension explains why regulatory responses remain contested. EU debates over platform governance, for instance, involve competing conceptions of democracy, free speech and market regulation rather than a straightforward



technical policy problem. Decisions about disinformation regulation effectively determine “*the sort of public sphere and democracy we ought to strive for*”[2]

The scale and institutional seriousness of the disinformation challenge is reflected in the European Union’s rapid policy mobilisation beginning in 2018. By that year, the European Commission had launched a Eurobarometer public opinion survey, conducted a public consultation process, established a High-Level Expert Group on fake news and disinformation and later introduced a voluntary Code of Practice on Disinformation signed by major technology companies. These steps illustrate that disinformation had moved from being a media concern to becoming a formalised governance issue within the EU policy framework[3]

The framing of disinformation has increasingly shifted from a media quality issue to a security concern. European policymakers have linked disinformation to hybrid threats, geopolitical conflicts and democratic resilience, especially following events such as the COVID-19 pandemic and the war in Ukraine[4]. This securitization has accelerated policy responses but has also introduced tensions between freedom of expression, surveillance risks and platform accountability.

At the same time, economic incentives complicate solutions. Large digital platforms operate on engagement-driven business models that can inadvertently reward emotionally charged or misleading content. Calls for stronger regulation must therefore confront powerful corporate interests and lobbying pressures. As a result, responses to disinformation require coordinated action among governments, civil society, educators, technology companies and citizens, particularly young people, who are both primary digital media users and potential drivers of resilience initiatives.

5.2 Responses from civil society, governments, big tech and educators

5.2.1 Government regulation and public policy responses

National governments and supranational organizations such as the European Union remain central actors in combating disinformation. The EU has emerged as a key regulatory arena due to the transnational nature of digital platforms and its capacity to influence global standards. Lobbying by technology companies increasingly targets EU institutions because policy decisions made there form the broader digital governance environment[5]

Recent regulatory initiatives, including the Digital Services Act (DSA) and Digital Markets Act (DMA), aim primarily to increase platform transparency, require risk assessments and impose accountability obligations rather than directly determining the truthfulness of specific content. These measures reflect a “co-regulatory framework” combining public oversight with platform self-regulation[6].

However, regulatory diversity across EU member states remains significant. Some governments emphasize rapid content removal and legal sanctions against intentional disinformation, while others prioritize media literacy or strategic communication. Bouza García and Oleart identify several regulatory approaches, including criminalization of deliberate disinformation, technological solutionism



relying on platform tools and efforts to rebuild institutional trust through education and fact-checking[7].

This variation illustrates that regulation alone cannot solve the problem. Over-reliance on legal responses risks censorship concerns, while insufficient regulation leaves platform dynamics largely unchecked.

The institutionalisation of disinformation policy within the EU further reflects the emergence of co-regulatory governance. Following recommendations from the High-Level Expert Group, major digital platforms committed under the Code of Practice on Disinformation to increase transparency in political advertising, close fake accounts and collaborate with independent fact-checkers. These developments illustrate how governments and platforms increasingly share responsibility for mitigating online manipulation[8]

5.2.2 The role of big technology companies

Technology platforms play a decisive role because they control the infrastructure through which information circulates. Their algorithmic systems prioritize engagement, shaping visibility and amplification patterns that can inadvertently favor sensational or misleading content.

At the same time, large technology companies actively participate in shaping regulatory frameworks. Their economic and communicative power allows them to influence policy discussions, sometimes pre-empting stricter regulatory proposals[9].

Transparency has become a major point of contention. Research into access to platform data has become increasingly restricted, limiting independent monitoring of disinformation dynamics. The removal or restriction of tools previously used by researchers, such as social media analytics platforms (e.g. Crowdtangle), illustrates this trend and raises questions about accountability and democratic oversight.

5.2.3 Civil society and youth organizations

Civil society organizations have emerged as crucial actors in building societal resilience against disinformation. Their activities include awareness campaigns, fact-checking initiatives, research collaborations and educational workshops.

In 2025, the European Commission allocated nearly €5 million[10] to support fact-checking, media literacy and youth-focused programs that combine workshops, online courses and gamified learning. Projects like *RESIST – Strengthening Societal Resilience to Disinformation in Europe* [11] show how civil society, educators and authorities collaborate to equip young people with critical thinking and digital literacy skills.

European civil society networks continuously monitor disinformation trends, organize conferences and webinars, share analytical resources and produce policy recommendations aimed at strengthening democratic information ecosystems[12]. These initiatives help bridge the gap between academic research, policymaking and grassroots activism.



The **European Youth Forum** and other youth-led organizations play a central role in building media literacy and civic skills by organizing workshops, campaigns and training that empower young people to critically assess information, engage in democratic processes and resist online manipulation.

5.2.4 Educators and media literacy initiatives

Educational institutions also play a long-term preventive role by fostering critical media consumption habits. Research consistently demonstrates that media literacy training improves individuals' ability to distinguish between accurate and misleading information.

Empirical evidence highlights both the scale of youth engagement with social media and the persistence of critical literacy gaps. For example, a 2023 *Eurostat* survey[13] found that only **36 % of young people aged 16–29 in the European Union reported trying to verify the truthfulness of online content from other sources**, underscoring that most do not critically check information encountered online.

These findings underscore the need for holistic training approaches that combine strategic planning, targeted interventions and operational activities to build lasting critical evaluation skills rather than relying on one-off awareness campaigns.[14].

For instance, the EU supports programs such as *How to Use Critical Thinking to Enable Media Literacy*[15] and *Media Literacy and Fighting Fake News*[16], which train educators across member states in analytical skills and classroom strategies for tackling misinformation. Other initiatives like the *Erasmus+-funded "QYourself" project*[17] and the *Media Masters* program [18] develop teaching materials and youth-focused interventions designed to embed critical media evaluation into everyday learning, illustrating how sustained educational efforts can build resilience far beyond one-off campaigns.

These initiatives include:

- Training educators and youth workers,
- Developing digital literacy curricula,
- Partnering with schools and universities,
- Creating online educational platforms and
- Involving volunteers in community outreach.

Such programs highlight that resilience against disinformation depends not only on factual correction but also on cognitive skills, institutional trust and democratic participation.

5.3 Regulation, media literacy, platform governance, civic engagement and digital resilience

5.3.1 Competing regulatory logics



European disinformation policy reflects competing logics. One approach treats disinformation primarily as a security threat requiring strong intervention, including content removal and monitoring. Another emphasizes platform self-regulation and voluntary cooperation with limited government intervention[19].

This dual logic creates policy tension. Excessive securitization may threaten civil liberties, while excessive reliance on self-regulation risks ineffective enforcement.

5.3.2 Media literacy as long-term resilience

Media literacy is widely recognized as a cornerstone of sustainable responses. Evidence shows that exposure to digital literacy training improves individuals' ability to evaluate online information and recognize misleading narratives[20].

Research also indicates that structured media literacy interventions can improve discernment between credible and misleading information and foster greater scepticism toward online content. These findings suggest that educational approaches represent not only preventive tools but also measurable resilience-building strategies against disinformation[21].

Importantly, media literacy does not simply involve fact-checking skills. Effective programs address:

- Source evaluation,
- Algorithm awareness,
- Emotional manipulation recognition,
- Digital citizenship responsibilities and
- Critical engagement with political communication.

Youth-focused initiatives emphasize participatory learning methods, including collaborative analysis of case studies, group discussions and simulation exercises.

5.3.3 Civic engagement and democratic resilience

Strengthening democratic participation is another essential dimension. Disinformation thrives in contexts of institutional distrust, political polarization and media fragmentation. Initiatives promoting civic engagement, such as workshops on democratic values, volunteering programs and community dialogue projects, contribute indirectly to disinformation resilience.

European research reinforces this assessment. A comparative study of national information resilience in **Austria, Czechia, Finland, and Spain** found clear differences in how societies cope with disinformation: **Finland and Austria**, with higher societal trust, diverse media ecosystems, and stronger institutional support, exhibited **greater resilience** to online false content, while **Czechia and Spain**, characterized by more polarized political environments and fragmented media landscapes, were linked to **lower resilience**. The analysis showed that national



context matters, that generic strategies do not uniformly strengthen resilience and that targeted efforts to build media literacy and trust are essential for enhancing democratic participation[22].

Educational organizations highlight that training should enhance not only technical skills but also autonomy, social awareness and active citizenship[23]. This broader approach recognizes that disinformation is ultimately a societal issue rather than solely a technological one.

5.3.4 Platform governance and transparency

Transparency obligations for platforms remain central to EU policy. The DSA framework emphasizes algorithmic transparency, risk assessments and accountability mechanisms rather than direct censorship[24]

These measures aim to:

- Clarify how content is prioritised,
- Reduce algorithmic amplification of harmful content,
- Improve data access for researchers, and
- Enhance platform responsibility.

However, implementation challenges remain, particularly regarding enforcement consistency and cross-border coordination.

5.4 Detection tools and response strategies

5.4.1 Technological detection tools

Technological tools form an important but insufficient component of disinformation responses. A range of verification tools is used by researchers, journalists and activists, including:

- Reverse image search tools (e.g. Google Images),
- Fact-checking databases (e.g. Google Fact Check Tools),
- Bot detection systems (e.g. Botometer X),
- Social network analysis software (e.g. OSoMeNet),
- Digital archives such (e.g. the Wayback Machine).

However, tool accessibility has become more limited. Some platforms have restricted API access or discontinued transparency tools, complicating independent monitoring (For example Crowdtangle by META)



Policy debates increasingly warn against excessive reliance on purely technological solutions. While detection tools are essential, “technological solutionism” risks transferring regulatory authority to private platforms without sufficient democratic oversight. Effective responses therefore require combining technological verification with institutional transparency, civic education and human judgement[25].

5.4.2 Emerging AI-based detection technologies

Research projects increasingly explore artificial intelligence-based detection methods. The DISSIMILAR project, for instance, uses digital watermarking algorithms and machine learning to identify manipulated multimedia content and verify original sources[26].

Such technologies are particularly important given the rapid development of deepfake technologies and automated content generation. Nevertheless, technological detection cannot replace human judgement, media literacy or institutional trust.

5.4.3 Response strategies beyond detection

Effective responses must extend beyond detection to include:

- Rapid fact-checking dissemination,
- Coordinated communication strategies,
- Collaboration between journalists and researchers,
- Public awareness campaigns,
- Educational workshops.

Educational workshops analyzing misinformation incidents, for example, encourage participants to examine structural causes, evaluate journalistic practices and propose preventive measures[27].

These participatory approaches help translate abstract media literacy concepts into practical skills.

5.5 Conclusion

Efforts to combat disinformation have expanded significantly over the past decade, yet no comprehensive solution has emerged. Regulatory frameworks such as the EU's digital platform governance initiatives represent important progress, but they operate within complex economic, political and technological ecosystems.

Three key trends are likely to form future responses:

First, regulatory consolidation.

European digital governance initiatives will continue to influence global standards, particularly regarding platform transparency and accountability.



Second, technological evolution.

Artificial intelligence will simultaneously generate new disinformation risks and provide new detection capabilities, requiring continuous adaptation.

Third, societal resilience.

Media literacy, civic engagement and youth participation will remain central to long-term solutions. Educational programs that combine technical skills with democratic values and critical thinking appear particularly promising.

Persistent gaps between high social media engagement and relatively low critical evaluation skills, such as findings that fewer than 20% of students critically assess online image sources despite widespread social media reliance, highlight the continued importance of combining regulatory, educational and civic approaches to address disinformation sustainably[28].

Ultimately, the most effective strategy is likely to be a hybrid one combining regulation, technological innovation, education and civic engagement. Disinformation is not simply a communication problem but a structural challenge for contemporary democracies. Addressing it therefore requires sustained cooperation between governments, platforms, educators, civil society organizations and citizens — especially younger generations who will form the future digital public sphere.

6. Conclusion

Disinformation is not a new fact, but its contemporary form represents a qualitative shift in both scale and impact. What began historically as centralized propaganda has evolved into a decentralized, algorithmically amplified and increasingly AI-driven system of information manipulation. In today's digital environment, disinformation operates as a structural feature of the information ecosystem, one that exploits technological infrastructures, cognitive vulnerabilities and political divisions simultaneously.

This paper has demonstrated that disinformation in the Baltic Sea region and more broadly in democratic societies, must be understood as a **strategic and systemic phenomenon**. It is deployed deliberately to achieve geopolitical and domestic objectives: **destabilizing societies, eroding trust in institutions, interfering in political processes and securing advantage in the gray zone between peace and conflict**. Its effectiveness lies in its ability to interact with existing societal fractures and digital architectures that reward engagement over accuracy.

The consequences for democracy are substantial. Disinformation undermines the epistemic foundations on which democratic systems depend, weakens the integrity of electoral processes, distorts media ecosystems and contributes to political cynicism and disengagement. At the same time, it accelerates polarization, erodes social cohesion and challenges the legitimacy of democratic institutions. Crucially, the response to disinformation is itself politically contested, involving trade-offs between security, freedom of expression and state intervention.



The analysis of **the Baltic Sea region** highlights that these challenges are intensified in highly digital, geopolitically exposed and linguistically diverse environments. Here, disinformation is not merely a communication issue but a **security concern embedded in broader patterns of hybrid conflict**, requiring coordinated regional and international responses.

Policy responses have expanded significantly, particularly within the European Union, yet no single solution is sufficient. Regulatory frameworks such as **platform accountability measures** represent necessary progress, but they do not address the full scope of the problem. **Technological detection tools** are essential but inherently reactive. **Media literacy initiatives** and **youth engagement** offer long-term resilience but require sustained investment and integration into broader policy strategies.

The central finding of this paper is that **disinformation cannot be effectively addressed through isolated interventions**. It is a systemic risk that demands a systemic response. This includes:

- Structural reforms to platform governance and algorithmic transparency
- Coordinated regulatory approaches across national and supranational levels
- Investment in media literacy and critical thinking as core democratic competencies
- Strengthening institutional trust and civic engagement
- Cross-sector cooperation between governments, civil society, technology companies and citizens

Ultimately, it forces a fundamental question: **how to preserve open, pluralistic information environments while protecting them from manipulation and abuse**. The answer will not come from a single policy tool, but from a sustained, balanced and adaptive approach that recognizes disinformation for what it is, not just false content, but a defining structural challenge of the digital age.