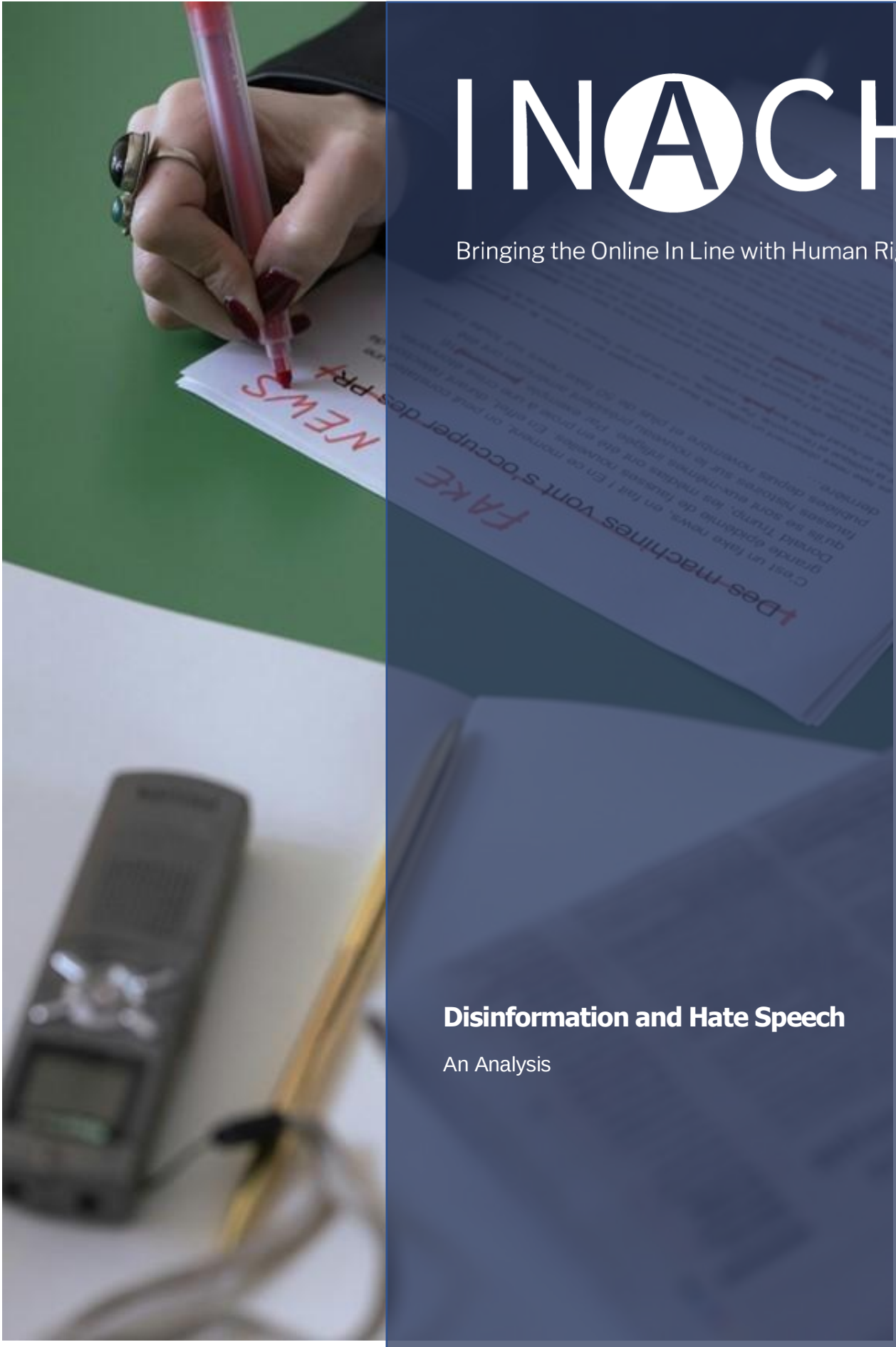

INACH

Bringing the Online In Line with Human Rights

Disinformation and Hate Speech

An Analysis

Compiled by
Adinde Schoorl, Cosima Hofacker,
Nikol Pardava & Catharina Chulick

2025

TABLE OF CONTENTS

International Network Against Cyber Hate – INACH

INACH was founded in 2002 to use intervention and other preventive strategies against cyber hate. The member organisations are united in a systematic fight against cyber hate, for example as complaints offices, monitoring offices or online help desks. In their respective countries, they provide important contacts for politicians, internet providers, educational institutions, and users.

Funding for INACH is provided by its members, the European Commission and other donors. The International Network Against Cyber Hate (INACH) unites multiple organizations from the EU, Africa, Israel, North Macedonia, Russia and South America. While starting as a network of online complaints offices, INACH today pursues a multi-dimensional approach of educational and preventive strategies.

This publication has been produced with the financial support of the Citizens, Equality, Rights and Values (CERV) Programme of the European Union. The contents of this publication are the sole responsibility of the International Network Against Cyber Hate and can in no way be taken to reflect the views of the European Commission.

Supported by the Citizens, Equality, Rights and Values (CERV) Programme of the European Union

Introduction

With nearly 25 years of experience monitoring online hate speech, INACH has gained deep insights into its evolving nature and societal impact. One of the most significant trends observed in recent years is the increasing convergence of hate speech with disinformation and conspiracy narratives. These hybrid threats are not only disseminated organically but are also strategically weaponized by political actors, governments, and organized movements to advance ideological or geopolitical objectives.

This fusion of hate and falsehoods has transformed hate speech into a more visible, complex, and transnational challenge—one that is deeply embedded in today's digital ecosystems. It now poses serious risks to democratic discourse, social cohesion, and public trust.

Recognizing these risks, the European Union has taken substantial steps to shape a coherent policy and regulatory response. Key initiatives include the Digital Services Act (DSA), the Code of Conduct on Illegal Hate Speech Online, and the Code of Conduct on Disinformation. These frameworks serve as critical instruments for upholding fundamental rights online while fostering accountability among digital platforms. Of particular interest to policymakers is the Code of Conduct on Disinformation, a soft law instrument developed through co-regulation. Unlike traditional legislation, such Codes of Conduct operate on the basis of voluntary commitments by online platforms, informed by shared standards and guided by principles of risk prevention, transparency, and civic responsibility.

As Borz, De Francesco, Montgomerie & Bellis (2023) note, such codes represent "mechanisms within the public and private spheres with the goals of managing potential risks while promoting development." The European Commission has employed this

model in diverse policy areas — including arms trade, food regulation, and emerging technologies — and is now adapting it to the digital domain.

This report contributes to ongoing policy debates by offering a critical examination of the EU's regulatory response to disinformation, with a specific focus on its intersection with hate speech. Given the increasing convergence of these phenomena, the Code of Conduct on Disinformation is positioned as a key instrument for mitigating harm in the online space.

The report proceeds in three parts:

1. **Conceptual Analysis**: It explores how disinformation operates in relation to hate speech, identifying key dynamics and risks.
2. **Case Studies**: It presents insights from three EU Member States—the Netherlands, Slovakia, and Germany—to highlight national-level implications and variations in exposure and response.
3. **Policy Evaluation and Recommendations**: It assesses the current scope and limitations of the Code of Practice on Disinformation and offers actionable recommendations for policymakers. These focus on strengthening accountability mechanisms, closing regulatory gaps, and enhancing coordination between EU institutions, Member States, and digital platforms.

At a time when democratic societies face heightened polarization and growing threats from coordinated online harms, it is essential for EU and national policymakers to adopt forward-looking, evidence-based approaches. This report aims to support that effort by identifying both the regulatory progress achieved and the urgent policy needs that remain.

Chapter 1: Why disinformation?

In recent years, disinformation has emerged as one of the most urgent and complex challenges in the digital age. The proliferation of alternative information channels — particularly through social media — and the global rise of populist rhetoric have contributed to a climate in which objective truth is increasingly contested. Terms like *"fake news"* have become deeply embedded in public discourse, often weaponized to undermine trust in traditional media and journalistic institutions. A notable example is U.S. President Donald Trump's 2018 *"Fake News Awards"*, which sought to discredit mainstream media outlets and elevate partisan narratives (Politico, 2018). Despite its prevalence in both political language and public debate, the concept of disinformation remains widely misunderstood and inconsistently defined, often leading to confusion about its boundaries, characteristics, and consequences.

1.1 Definitions

The terms *disinformation*, *misinformation*, and *malinformation* are frequently used interchangeably in public discourse, yet each refers to a distinct phenomenon. Accurate terminology is crucial for understanding and addressing the different forms of false or harmful information in both policy and practice. However, as new contexts and digital formats emerge, the boundaries between these categories can become blurred.

- **Misinformation** refers to *false or misleading information shared without intent to cause harm*. It often arises when individuals or organizations unknowingly disseminate inaccurate content. For example, during unfolding news events, early reports may contain errors that are shared before facts are confirmed. Misinformation also occurs when individuals unknowingly circulate inaccuracies, believing them to be true.

- **Malinformation**, by contrast, involves *the use of genuine, truthful information with the intent to cause harm*. This type of content is strategically weaponized to damage reputations or incite harassment. Common examples include doxing—the public release of private personal information—and Non Consensual Intimate Images. Although the content itself is not false, its deliberate misuse is harmful and often driven by hateful intent.

- **Disinformation** is *false information deliberately created and disseminated to deceive or manipulate*, often for political, ideological, or economic gain. Unlike misinformation, disinformation is intentionally deceptive. Even if subsequent sharers do not realize the content is false, its origin lies in a conscious effort to mislead. Disinformation frequently manifests in the form of conspiracy theories, propaganda, or manipulated multimedia content such as images, videos, or audio recordings (Britannica 2025).

To summarize:

- **Misinformation**: False information shared without harmful intent.

- **Disinformation**: False information shared *with* harmful intent.

- **Malinformation**: True information shared *to* cause harm, often by violating privacy or context (Wardle & Derakhshan, 2017).

This report will focus primarily on disinformation and its relationship with online hate speech.

In order to fully grasp the issue at hand, it is also necessary to define *hate speech*. While no universally binding legal definition exists under international human rights law, several prominent organizations have provided widely recognized interpretations.

The Council of Europe defines hate speech as:

"*All types of expression that incite, promote, spread or justify violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status such as 'race', colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation*." (Council of Europe 2022)

The United Nations similarly defines it as:

"*Any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are… based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor*." (UN Strategy and Plan of Action on Hate Speech 2025)

The definition of Hate Speech according to INACH is:

'Intentional or unintentional public discriminatory and/or defamatory statements; intentional incitement to hatred and/or violence and/or segregation based on a person's or a group's real or perceived race, ethnicity, language, nationality, skin colour, religious beliefs or lack thereof, gender, gender identity, sex, sexual orientation, political beliefs, social status, property, birth, age, mental health, disability disease (INACH).'

While closely aligned, these definitions underscore the fact that hate speech remains a contested legal concept, especially in the context of freedom of expression. The lack of a universally accepted definition poses challenges for enforcement and regulation, particularly in digital spaces. Although many forms of hate speech are criminalized under both national and international law, online content that falls into so-called "borderline" territory often escapes legal accountability. This includes content that may not meet the legal threshold for illegality but still spreads hate or disinformation—frequently described as *"awful but lawful."* This grey area is problematic for several

reasons. Most notably, lawful yet harmful content can act as a catalyst for illegal behaviour, especially in user-generated comment sections. For instance, a post that skirts the boundaries of hate speech law may provoke clearly unlawful responses, such as direct threats or incitement to violence. In essence, *legal hate breeds illegal hate*. This dynamic poses significant regulatory and enforcement challenges, which will be explored further in the next chapter through the lens of disinformation's strategic function, dissemination methods, and societal consequences.

1.2. The risks of disinformation

Both offline and online hate speech present serious threats to human rights, social cohesion, and democratic governance. However, online hate speech poses unique challenges due to the speed, scale, and anonymity afforded by digital platforms. The ease of anonymous sharing enables harmful content to circulate rapidly and reach large audiences with minimal accountability. In recent years, the increasing accessibility of advanced artificial intelligence (AI) tools has further accelerated the creation and dissemination of hateful content, lowering the barrier to entry for malicious actors.

The same dynamics apply to disinformation, which spreads uncontrollably across websites, forums, social media platforms, and encrypted messaging apps. AI-enhanced content manipulation, including deepfakes and sophisticated image or video alterations, has made visual media increasingly unreliable as evidence, raising new challenges for journalism, justice systems, and democratic debate.

Disinformation is closely intertwined with hate speech. False or misleading content is often deliberately crafted and shared to incite division, spread prejudice, or target specific groups or individuals, frequently accompanied by dehumanizing or hateful visual elements. As Wardle (2024) notes, this is particularly dangerous in conflict-affected or high-risk areas, where longstanding religious, ethnic, or cultural tensions can

be inflamed by conspiracy theories, disinformation campaigns, or the strategic deployment of myths.

While the spread of false information is not a novel phenomenon, the internet—especially social media—has transformed its scale, speed, and impact. Automated bots, coordinated campaigns, and recommendation algorithms significantly amplify harmful content, exacerbating polarization and eroding trust in shared facts (Vasist, Chatterjee & Krishnan, 2023).

According to the EU Media and News Survey (2023), social media has become one of the primary news sources for many Europeans, next to TV and radio. The European Parliament Youth Survey (2024) reveals that for 42% of respondents aged 16–30, social media is their main source of political and social information, surpassing television (39%). While 70% of young people express confidence in recognizing disinformation, 76% believe they have already encountered it.

If disinformation is not adequately addressed, the long-term consequences are profound. As the Financial Times (2023) reports, unchecked disinformation can lead to deepening cynicism, declining trust in democratic institutions, and fragmentation of the public sphere. Over time, scepticism morphs into nihilism, and even the most credible sources of information are viewed with suspicion.

The COVID-19 pandemic showcased the real-world risks of disinformation. It saw an unprecedented surge in false narratives, some propagated by foreign actors seeking to influence domestic debates within the EU (European Commission 2025). These campaigns exploited uncertainty and fear, undermining trust in health authorities, governments, and science—with effects still visible in public opinion today (Li, Zhang & Niu, 2021).

The goal of disinformation is not always for people to believe it, but to saturate the public sphere with so many different, contradicting narratives that people do not know

who to believe anymore. This undermines trust in public institutions and democratic processes and shared reality.

Disinformation empowers actors to construct "alternative realities", giving them significant influence over public narratives. This strategy is frequently used by populist parties and political extremists to delegitimize mainstream institutions and foster mistrust. Disinformation and hate speech are both key drivers of polarization, especially online, where platforms often reward sensational or divisive content (Vasist, Chatterjee & Krishnan, 2023).

With the rise of AI and algorithmic tools, policymakers face increasing challenges in regulating the multimodal creation and transnational spread of disinformation. The manipulation of digital audiences now includes tactics such as:

- AI generated images
- Mischaracterizing events through false narratives
- Inauthentic amplification by bots and fake accounts
- Microtargeted disinformation using advertising tools
- Harassment and abuse of journalists, especially those with opposing or critical viewpoints (World Economic Forum 2022)

One concrete example is the 2025 German elections, during which Meta and X reportedly authorized hate-driven political ads in violation of electoral integrity principles (Euractiv 2025).

While disinformation often serves political ends, it is also increasingly used to target minorities and marginalized groups. An EU study found that Roma communities are frequent victims of domestic disinformation, while Kremlin-backed campaigns have targeted Jewish communities. Women, especially those in politics, are disproportionately affected: research by the Center for Countering Digital Hate (CCDH)

found that Instagram failed to act on 93% of abusive comments aimed at prominent U.S. female politicians, including those containing death and rape threats (CCDH 2024).

These examples underscore the interconnected nature of disinformation and hate speech. Both are mutually reinforcing, deeply damaging, and strategically employed to undermine trust, exclude communities, and destabilize democratic societies.

Chapter 2: Borderline content and disinformation

Political actors and movements have become skilled at staying within the bounds of hate speech laws. While illegal content can be reported and removed, a growing challenge lies in borderline content—material that is harmful but does not explicitly break the law. This type of content often involves disinformation that fuels hatred, division, and polarization. For instance, false narratives about immigrants may not be illegal, but they incite hostility toward refugees, Muslims, and other marginalized groups. This is what we refer to as “awful but lawful”—content that spreads harm without violating legal standards. Because platforms cannot remove such content based on existing laws, new strategies are needed. These include stronger platform guidelines, improved digital literacy, fact-checking initiatives, and co-regulatory frameworks to address harmful but legal online speech. At the same time, often do these false narratives actually violate the community guidelines of the platforms and could be removed based on these, yet they often do not remove it. In other words, commitment from the social media platforms is essential in dealing with disinformation.

2.1. The effects of disinformation

Here are a few short case studies showing the effects of disinformation and hate speech.

2.1.1 The Netherlands

It is a well-established tactic for political actors to strategically frame certain issues — such as immigration — in order to gain public support. Across Europe and the United States, this often involves the deliberate spread of disinformation intertwined with political messaging, particularly targeting immigrants, refugees, and ethnic minorities.

In the Netherlands, monitoring of online hate content has revealed that right-wing politicians frequently share videos involving people of colour, such as public disturbances or scenes at asylum centres, often without context. These posts are typically accompanied by captions claiming that foreign populations are destroying Dutch society, promoting the narrative that immigration is the root of national decline. This tactic, while not illegal, constitutes borderline content—what is often called "awful but lawful". Though it doesn't breach hate speech laws, it spreads xenophobic disinformation and fuels hate online.

These posts also open the door to hate speech in user comments, where individuals—sometimes bots—engage in coded or indirect hate to evade content moderation. This method has proven effective; the far-right Freedom Party (PVV), led by Geert Wilders, gained a landslide victory in the 2023 Dutch elections. There were many reasons for his victory, however, undoubtedly Wilders' online strategy also contributed to it (NOS 2023).

Following the formation of a coalition government that included the PVV and other populist parties until June 2025, anti-immigration rhetoric has been institutionalized. Despite the lack of evidence, the government continues to speak of an "immigration crisis" and falsely links it to issues like the housing shortage. These narratives ignore the reality that refugees and migrants are themselves affected by housing insecurity (Pointer 2023). The numbers on immigrants coming to the Netherlands show a stable line over the last 30 years with a peak in 2015 due to the war in Syria. However, due to closing of asylum seeking centres, a lack of housing for foreigners who have received a permit to stay and a staff shortage at the national immigration institution (IND) lead to full asylum centres and disturbances in the villages around those centres. In other words, the Dutch are dealing with a 'shelter crisis' instead of an immigration crisis (RTL Nieuws 2024).

More concerning is the mainstreaming of conspiracy theories within Dutch politics. Senior PVV members, including the former Minister for Asylum and Migration, Marjolein

Faber, and former Minister for Foreign Trade and Development, Reinette Klever, have publicly expressed belief in the Great Replacement theory—a racist conspiracy that claims immigrants are systematically replacing native populations. Faber later distanced herself from the term but maintained her concerns about "demographic change," signalling continued alignment with its core ideas (NOS 2024).

In 2025, Faber publicly described Ukrainian President Zelensky as a dictator, repeating Russian disinformation narratives—a clear indication of the influence foreign actors can exert over domestic politics (NOS 2025).

The Children's Book Week of 2023 offers a stark example of how disinformation can rapidly escalate into hate campaigns. Dutch author Pim Lammers, known for his inclusive children's books featuring LGBTQ+ themes, was selected to read the opening poem. After it emerged that Lammers had also written an adult story depicting child abuse (from the victim's perspective), a wave of online hate began, driven largely by so-called *"momfluencers"*—celebrities using their social platforms to express outrage and accuse him of "glorifying paedophilia."

As the hate campaign intensified, Lammers received serious threats, forcing him to withdraw from the event and go into hiding. Journalistic investigations later traced the campaign's origins to an extreme-right politician, who had authored a column denouncing Lammers for undermining "traditional family values." The hate narrative jumped from a fringe blog to mainstream platforms, showing how online disinformation and hate speech can rapidly scale with political help.

The issue became even more politicized when another right-wing politician raised it in Dutch Parliament, questioning educational policies and reinforcing anti-LGBTQ+ sentiments at the national level (Het Parool 2025).

This evolving pattern of weaponizing disinformation and hatred for political gain—whether through anti-immigration rhetoric or attacks on LGBTQ+ rights—presents a serious threat to social cohesion, democratic values, and public trust. It highlights the

urgent need for comprehensive strategies that address both illegal hate speech and the more insidious, lawful-but-harmful content that fuels discrimination and division.

2.1.2 Slovakia

The disinformation landscape in Slovakia has become increasingly complex, with critical turning points around the 2020 parliamentary elections and the onset of the COVID-19 pandemic. The pandemic triggered an unprecedented wave of false information that not only distorted public understanding but also began to influence policymaking. As Disinfo.EU aptly noted:

"The year 2020 was a wake-up call for Slovakia – the COVID-19 pandemic and consequent infodemic, which continues to this day in connection with the war in Ukraine."

This period marked the Slovak government's first coordinated efforts to address disinformation, including the revision of key policy documents and the development of strategic action plans.

Slovakia remains particularly susceptible to disinformation due to low media literacy and limited critical thinking skills among large segments of the population. According to the 2022 GLOBSEC Trends survey, Slovaks are more inclined toward conspiracy theories than citizens in other Central and Eastern European countries. Additionally, a 2023 Reuters Institute study found that only 27% of Slovaks trust the media, creating an environment where falsehoods are more likely to gain traction and spread rapidly.

The war in Ukraine has amplified disinformation efforts, particularly those promoting pro-Russian narratives. These campaigns—rooted in messaging that began after the annexation of Crimea—seek to portray Russia as a victim and defender of traditional values, while minimizing or outright denying Russian war crimes in Ukraine.

Far-right political figures, such as MEP Milan Uhrík of the Republika party, play an active role in this manipulation by labelling critics as *Russophobes*, thus reframing legitimate

concerns as prejudice. The impact is evident: in 2024, only 40% of Slovaks believe Russia is responsible for the invasion of Ukraine, a sharp decline from 51% the previous year (Euronews, 2024; GLOBSEC, 2022).

Today, Slovakia continues to struggle with disinformation on multiple fronts. While alternative media outlets and Facebook remain primary sources of false narratives, political actors are increasingly complicit in their spread. As reported by Disinfo.eu, these campaigns are well-coordinated, often originating on fringe platforms before gaining momentum via social media and endorsements from political figures. This disinformation ecosystem makes it especially difficult to promote fact-based discourse and critical thinking, undermining public trust and democratic resilience.

The Slovak case illustrates how disinformation evolves through a complex interplay of digital manipulation, political opportunism, and societal vulnerabilities. As Slovakia enters another critical period, it becomes clear that traditional counter-disinformation measures are no longer sufficient. Policymakers must adopt robust, flexible, and long-term strategies, including:

- Enhancing media literacy and civic education,
- Promoting trusted and independent journalism,
- Increasing transparency and accountability for political communication,
- And investing in cross-sector cooperation between government, civil society, and tech platforms.

The Slovak experience serves as a warning—and a call to action—for democracies across Europe.

2.1.3 Germany

Elections are prime targets for disinformation, and the 2025 German federal elections on 23 February are no exception. As seen during the European Parliament elections,

disinformation campaigns have intensified across topics including the war in Ukraine, climate change, immigration, and the electoral process itself.

According to ongoing investigations by renowned fact-checking outlet CORRECTIV, false and misleading claims are already circulating widely. One example includes a manipulated video where FDP politician Marcus Faber is falsely labelled a Russian double agent by former party colleague Christian Blume, a claim promoted by the fringe outlet *Andere Meinung*. CORRECTIV traced this to a broader Russian disinformation campaign targeting German politicians.

Another case involves AfD lead candidate Alice Weidel, who, in a January 2025 livestream with Elon Musk, falsely claimed that *"Hitler was a communist who saw himself as a socialist"*. Such narratives aim to distort historical facts and stir ideological confusion.

As Wardle (2024) notes, disinformation often exploits existing societal tensions. In Germany, this is particularly evident in immigration debates. In December 2024, pro-Russian actors circulated a fake article claiming Germany planned to "import 1.9 million workers from Kenya." CORRECTIV's investigation revealed that this was part of a coordinated disinformation campaign, relying on AI-generated websites and videos designed to amplify Russian propaganda.

Similar tactics were seen in August 2024, when a fake website targeted Foreign Minister Annalena Baerbock. In total, at least five major fake domains and narratives have been linked to such campaigns in the lead-up to the elections.

Visual disinformation also plays a central role. A recent viral video appeared to show a Syrian asylum seeker abusing an elderly man in a German retirement home. In reality, the footage was from Detroit in 2020, and the abuser was neither a caregiver nor

connected to Germany. The video was repurposed with false context to stoke anti-immigrant sentiment.

Public anxiety around disinformation is high. According to the Bertelsmann Foundation's 2024 "Insecure Public" study,

- 84% of Germans view deliberate online disinformation as a major or very serious societal problem.
- 81% believe it poses a direct threat to democracy and social cohesion, especially around sensitive topics like immigration, health, war, and climate change.

The German government has adopted a multi-agency, strategic approach to counter disinformation:

- Individual ministries respond to content that affects their remit, especially if it targets government action or officials.
- The Federal Foreign Office (AA) focuses on foreign disinformation efforts.
- The Federal Ministry of the Interior (BMI) coordinates responses to hybrid threats, including influence campaigns.
- The Federal Office for the Protection of the Constitution (BfV) monitors threats to democratic order, including foreign propaganda and cyber operations.

Alongside enforcement, citizen education and awareness remain a central priority, as Germany continues to strengthen its defences against both domestic and foreign disinformation.

2.2. Lessons learnt

An analysis of the Netherlands, Slovakia, and Germany reveals several urgent and interrelated trends regarding the spread and impact of disinformation and hate speech in democratic societies:

1. Disinformation has fully entered the political mainstream

Disinformation and hate-driven narratives are no longer confined to the political fringes—they have penetrated all levels of politics, from electoral campaigns to policy justification. Politicians and even governing parties openly use disinformation to shape public opinion, polarize society, and consolidate power. In the Netherlands, for instance, Thierry Baudet, the leader of once one of the biggest national political parties – the FVD – claimed in an interview that the world is ruled by 'evil lizards' and supported Putin's war in Ukraine (NOS 2022). Members of the Dutch Farmer's party – the BBB – which was part of the coalition government until June 2025 have spread doubts continuously doubts about scientific studies related to alternative energies, the climate crisis and other related topics. They consistently claim other alternative studies prove the opposite, creating distrust in scientific research in general (Tubantia 2024).

2. Low media literacy and misinformation vulnerability

A widespread lack of media literacy creates fertile ground for disinformation to take root. Many citizens lack a clear understanding of freedom of expression, are unfamiliar with how to verify sources, and increasingly rely on social media as their primary news source. In some countries, this issue is compounded by state-controlled or politically compromised media, further blurring the line between fact and manipulation.

3. Hate is Embedded in Disinformation—Often Unnoticed

Disinformation frequently carries coded or indirect hateful narratives. In the Dutch Children's Book Week case, for example, many influencers unintentionally fuelled anti-LGBT+ hate while believing they were protecting children from paedophilia. This underscores how emotive manipulation, and viral online storms obscure the origins of disinformation and redirect public outrage toward marginalized groups.

4. Disinformation undermines trust in democratic institutions

A consistent pattern across all three countries is the erosion of trust in democratic institutions. Disinformation—particularly when amplified by political figures—encourages citizens to turn to alternative sources of information, where they are fed narratives suggesting that governments, media, and experts cannot be trusted. This dynamic weakens democratic resilience and makes societies more vulnerable to extremism and foreign influence.

5. Borderline content fuels polarization and targeted hate

So-called “borderline content”—legal but harmful material—plays a significant role in escalating online hate. Disinformation that scapegoats immigrants, refugees, or LGBTQ+ communities for societal issues generates real-world consequences, opening the door to explicit hate speech, both legal and illegal. The amplification of such content online intensifies polarization and fosters a climate of hostility.

6. A multi-layered response is essential

Given the complex, evolving, and interconnected nature of disinformation and hate speech, addressing them requires a multi-dimensional strategy. It must include:

- Improved digital and media literacy education
- Stronger platform accountability and content moderation
- Support for independent journalism and fact-checking
- Robust cross-sector and cross-border cooperation
- And targeted policy measures addressing both domestic and foreign disinformation actors

Concluding, disinformation is not just a communications problem—it is also a democratic one. To safeguard democratic values and social cohesion, a coordinated, sustained, and systemic approach is urgently needed.

Chapter 3: Regulations

The European Union's tradition of soft power is reflected in its preference for Codes of Conduct as tools of governance. Rather than imposing rigid legislation—which can be slow and prone to legal challenges—the EU often opts for soft regulation to foster voluntary cooperation, especially in dynamic fields like digital policy.

The Code of Conduct on Disinformation (formerly the Code of Practice) exemplifies this approach. Initiated by the European Commission, it encourages digital platforms to take voluntary action against disinformation, allowing for flexibility in how commitments are met. This aligns with broader EU strategies of self- and co-regulation, where regulation is built on collaboration rather than enforcement.

As Heldt (2019) highlights, signatories maintain significant discretion, making the Code a pragmatic tool where regulatory and industry goals align.

3.1. Code of Conduct on Disinformation

The European Commission has developed several key initiatives to counter disinformation, beginning with the 2018 Communication on 'Tackling Online Disinformation: a European Approach'—a toolbox of measures aimed at safeguarding EU values. This was followed by the European Democracy Action Plan, which introduced guidelines on the responsibilities and accountability of online platforms. That same year, the Code of Practice on Disinformation was launched—the first global instance of industry-wide voluntary self-regulation against disinformation.

The COVID-19 disinformation monitoring program, led by signatories of the Code, served as a transparency initiative to ensure platforms took proactive steps. However, a 2021 Commission assessment revealed significant gaps, prompting the adoption of a

Strengthened Code of Practice on June 16, 2022. This revised version now includes 34 signatories and a broader range of voluntary commitments. To ensure long-term effectiveness, a Permanent Task-Force was created, along with a Transparency Centre to inform the public about policy implementation.

As of February 2025, the Code has been integrated into the Digital Services Act (DSA) and renamed the Code of Conduct on Disinformation. It now outlines 44 commitments and 128 specific measures focused on areas such as election integrity, ad transparency, platform accountability, user empowerment, and crisis response.

Key commitments and measures:

- **Ad placement scrutiny**: Signatories must ensure that entities spreading disinformation do not benefit from advertising revenue and must block disinformation from being monetized or disseminated via ads.
- **Political advertising transparency**: Political content must be clearly labelled so users can easily identify it.
- **Service integrity**: Platforms are expected to limit manipulative tactics like fake accounts, bot amplification, impersonation, and malicious deepfakes. Stronger cooperation among signatories is encouraged.
- **User empowerment**: Platforms must provide tools to flag disinformation, promote media literacy, and offer access to authoritative sources. They are also expected to make recommender systems more transparent and resistant to abuse.
- **Monitoring and reporting**: Platforms must regularly review disinformation tactics (TTPs) and report their responses per EU member state. VLOPs are required to report every six months.

During elections, signatories can activate a Rapid Response System (RRS), facilitating swift coordination among stakeholders to identify and address sensitive information. This system has already been used during the 2024 European and Romanian elections.

Fact-checking is central to the EU's anti-disinformation strategy. Platforms work with fact-checkers to label false content and ensure consistent application across all EU countries and languages. The European Digital Media Observatory (EDMO) plays a key role by connecting fact-checkers, researchers, journalists, and educators. In 2024, EDMO launched a task force providing daily updates and weekly trend reports to flag emerging disinformation.

Complementing EDMO, the European Fact-Checking Standards Network (EFCSN) promotes rigorous standards of independence, transparency, and methodological integrity. It supports cross-European collaboration and helps develop professional norms for the fact-checking field.

The EU DisinfoLab—a coalition of six European fact-checkers and research groups—has been instrumental in drafting the Code of Professional Integrity for Independent European Fact-Checking, with support from the Commission. Members include AFP (France), Correctiv (Germany), Demagog (Poland), Pagella Politica/Facta (Italy), EU DisinfoLab (Belgium), and Fundación Maldita.es (Spain). This initiative aims to establish consensus-based standards, enhancing both legitimacy and practical impact in the fight against disinformation across the EU.

3.2. Challenges

A critical weakness in the current European framework for countering disinformation lies in the exclusion of certain fringe platforms from the scope of the Code of Conduct on Disinformation. These platforms often act as incubators for harmful narratives, which are later amplified on mainstream platforms. Ignoring the interconnected nature of the digital ecosystem undermines efforts to contain disinformation at its source. Moreover, many of these fringe platforms lack meaningful moderation, and in some

cases, they are explicitly designed to host illegal content, further complicating enforcement and cooperation.

The impact of disinformation extends far beyond its immediate spread. Once circulated, disinformation can leave a lasting digital footprint, even if removed later. The erosion of public trust in media and democratic institutions is a long-term consequence that no amount of fact-checking can fully reverse. Effective countermeasures must therefore be part of a multi-faceted, preventative approach that addresses both systemic vulnerabilities and root causes.

Despite significant investment in fact-checking and debunking efforts, questions remain about their actual impact. While disinformation has been linked to various harmful outcomes, there is limited empirical evidence of direct causation between disinformation and specific political events. Moreover, users who distrust mainstream media are unlikely to engage with or believe debunking content—limiting its reach and effectiveness.

Recent developments signal potential setbacks: Meta has withdrawn from partnerships with fact-checkers, and similar moves in the EU may follow. Both Meta and X (formerly Twitter) have publicly pushed back against the Digital Services Act (DSA), portraying fact-checking as a threat to "free speech"—despite the fact that such rhetoric often serves to legitimize and amplify hate speech and violent content. The DSA, with its enforcement powers and financial penalties, remains a crucial counterbalance to these trends.

Another significant blind spot in the regulatory conversation is the working conditions of content moderators. Often employed through third-party firms in the Global South, moderators face long hours, minimal pay, and insufficient training. They are expected to make complex decisions—often within just 7 seconds—on whether content is harmful, illegal, or should be protected as free expression. Exposure to graphic content without access to mental health support takes a severe psychological toll, yet accountability is

scarce. Because moderators are contracted through local companies, pursuing legal remedies against major platforms is nearly impossible. It is unrealistic to expect these moderation teams to clear the digital space from all available disinformation in these conditions or in any conditions at all.

Content moderation also suffers from linguistic inequity. Platforms prioritize widely spoken languages like English, leaving lesser-used or regional languages—where disinformation can be just as dangerous—under-moderated. A stark example is the case of Facebook in Myanmar (2017), where limited moderation in Burmese allowed calls for violence against the Rohingya to spread unchecked, contributing to a genocide. This failure highlights the need for locally informed, linguistically diverse moderation strategies.

The Code does not address what to do about the political strategy of using disinformation to attract votes, to create hateful narratives and generate emotions that are unleashed on scapegoats. It means that these hateful narratives dominate platforms without the tools to address them.

The Code of Conduct on Disinformation, even in its strengthened form, fails to address many systemic issues. It does not mandate concrete objectives or require platforms to disclose measurable indicators like the number of bots removed, flagged posts, or user reports addressed. While some transparency is offered at the national level, granular, local-level data is lacking, which is especially problematic during events such as national elections, where disinformation thrives in specific contexts.

The Code's vagueness regarding required platform actions—paired with the absence of enforcement mechanisms—has hindered consistent implementation. A recent study (Mundges, 2024) found that reporting was often incomplete or lacked robustness, with qualitative assessments frequently missing relevant detail.

The absorption of the Code into the Digital Services Act marks a shift from voluntary self-regulation to a co-regulatory model. With this transition, platforms will face legal accountability for failing to uphold their commitments. Systematic monitoring and the possibility of sanctions may finally bring the needed pressure to ensure meaningful compliance.

Chapter 4: Recommendations

The previous chapters outlined the definition and risks of disinformation and explored its harmful effects across three national political contexts. They also traced the development and scope of the Code of Conduct on Disinformation. This chapter presents recommendations for strengthening the fight against disinformation. While the Code is a significant initiative, addressing the complexity of disinformation demands measures that extend beyond its current commitments.

4.1 Strengthening the Code of Conduct on Disinformation

One of the Code's greatest strengths lies in its voluntary cooperation between platforms, European institutions, and civil society. However, its recent integration under the Digital Services Act (DSA) provides an opportunity to add enforcement mechanisms that enhance accountability and ensure more robust compliance.

Currently, the Code outlines broad commitments but lacks clear, detailed obligations. It asks platforms to have policies in place but does not define what these should include or how they should be implemented. This lack of specificity limits its effectiveness.

To address this, the following improvements are recommended:

- **Context-specific disinformation policies**: Platforms should tailor their policies to national and local realities, as disinformation strategies vary by region. Local moderation teams fluent in native languages and culturally aware of their context should work in close coordination with civil society organizations to detect and respond to emerging threats.
- **Permanent readiness**: Monitoring efforts should not be limited to elections or crises. A continuous, proactive approach is needed to counter how disinformation is used as a long-term tool of influence.
- **Improved moderation standards**:

- Enhance discoverability controls to prevent disinformation from surfacing easily via common search terms.
- Implement consistent labelling systems for flagged content. Not only regarding incorrect information or AI generated content, but also for repurposed content by adding the obligation to show the source of the original content.
- Integrate visible fact-checking links to credible sources.
- Adjust engagement metrics to reduce the algorithmic amplification of false content.

4.2 Political Responsibility

As discussed in Chapter 3, disinformation is increasingly deployed as a political strategy. Politicians and parties of various ideologies—though some more than others—use disinformation in campaigns, social media messaging, and even political platforms to manipulate public perception and gain votes.

Political actors must lead by example and commit to preserving fair, fact-based democratic debate. Media outlets also carry a significant responsibility in this effort. They must engage in pre-bunking and debunking practices to prevent disinformation from dominating the public space and skewing political discourse.

4.3 Education and Media Literacy

Given that users often never encounter debunking messages after engaging with false content, education remains the most effective long-term defence against disinformation. Citizens must be equipped to identify falsehoods and distinguish between credible and untrustworthy information sources, especially in the lead-up to elections.

Educational efforts should focus on the following:

- **Understanding freedom of speech**: Many users believe that freedom of speech allows for unrestricted expression. However, the concept also entails responsibilities and limitations—particularly the obligation to protect inclusive, respectful public discourse. True freedom of speech safeguards minority voices and promotes a balanced, democratic exchange.
- **Recognizing trustworthy information sources**: Mistrust in mainstream media has driven many citizens to rely on social media platforms for news. While mainstream outlets must work to regain public trust, platforms also have a role in curbing the spread of harmful content by influencers. The notion that influencers are inherently "independent" is misleading—many are financially backed by movements or campaigns, often spreading curated narratives. This relationship must be clearly communicated to all segments of society to promote critical engagement with online content.

Bibliography

- Borz, Gabriela, De Francesco, Fabrizio, Montgomerie, Thomas L. & Bellis, Michael Peter, 'The EU soft regulation of digital campaigning: regulatory effectiveness through platform compliance to the code of practice on disinformation', Policy Studies, Volume 45, 2024 – Issue 5, p 147
- Britannica, Palfrey, John, 'Misinformation and Disinformation – Overview, differences, how it is spread, free expression & AI', 28 March 2025, link
- CEPA, Odarchenko, Kateryna, 'Slovak vote shows need for NATO action on Russian disinformation', 31 January 2024, link
- Council of Europe, 'Combating hate speech', 2022, link
- Center for Countering Digital Hate, 'Abusing women in politics, how Instagram is failing women and public officials', 14 August 2024, link
- Correctiv, 'Bundestagswahl 2025: diese Falschbehauptungen, Fakes und Gerüchte kursieren', 2 December 2024, link
- Correctiv, Thust, Sarah, Bernhard, Max & Hock, Alexej, ‚Russische Einflussoperation verbreitet Fake-Artikel zu Migrationsabkommen mit Kenia', 23 January 2025, link
- Correctiv, Timmermann, Sophie, ‚'Video von Angriff in einem Pflegeheim stammt aus Detroit – nicht aus Münster', 16 January 2025, link
- Disinfo.eu., Duboczi, Peter, Ruzickova, Michaela, 'The disinformation landscape in Slovakia', 19 September 2023, link
- Euractiv, Datta, Anupriya, 'Hate speech failures by Meta and X undermine German election', 22 February 2025, link
- Euronews, Carter, Brian, 'Slovakia's disinformation history serves as a cautionary tale for the EU', 27 May 2024, link
- European Commission, 'Tackling coronavirus disinformation', link, entered on 18 April 2025
- European Parliament News, 'TV still main source for news but social media gaining ground', 17 November 2023, link

- European Parliament Barometer 2024, Youth Survey, link
- Financial Times, Higgins, Eliot, 'What to do about disinformation', 16 December, link
- Globsec, 'Globsec trends, CEE amid the war in Ukraine', 2022, link
- Het Parool, Dielman, Tom, 'Extreem-christelijke stichting Civitas Christiana moet alle beschuldigingen aan adres van kinderboekenschrijver Pim Lammers verwijderen', 16 april 2025, link
- Internet Policy Review, 13(3)., Mündges, Stephan & Park, Kirsty, 'But did they really? Platforms' compliance with the Code of Practice on Disinformation in review', 2024, link
- KRO-NCRV Pointer, 'PVV overschat aantal woningen dat naar migranten gaat', 15 November 2023, link
- NOS, 'Politieke aardverschuiving: PVV veruit de grootste, coalitie afgestraft', 23 November 2023, link
- NOS, 'Faber begrijpt commotie, maar blijft bij uitspraken over omvolking en islam', 18 Juni 2024, link
- NOS, 'Faber (PVV) neemt afstand van term 'omvolking', maar ziet wel zorgelijke demografische ontwikkeling', 24 Juni 2024, link
- NOS, 'Minister Faber noemt Zelensky 'niet democratisch gekozen', neemt woorden terug', 21 Februari 2025, link
- NOS, 'Grote zorgen in de Tweede Kamer over complottheorieën Baudet', 18 Oktober 2024, link
- Politico, Schwartz, Jason, 'Trump gives out 'Fake News Awards' to CNN, N.Y. Times, Wash Post', 17 January 2018, link, entered 18 April 2025
- Pubmed, Vasist, Pramukh, Chatterjee, Debashis & Krishnan, Satish, 'The polarizing impact of political disinformation and hate speech: a cross country configural narrative', 17 April 2023, link
- Reuters Institute. (2023). Digital News Report. Retrieved from Reuters Institute

- RTL Nieuws, Bremer, Floor, ‘Hebben we nu een asielcrisis of niet? Feiten en cijfers op een rij’, 15 October 2024, link
- RTV Drenthe, ‘Haatberichten over schrijver Pim Lammers worden verwijderd’, 16 April 2025, link
- Science Direct, Li, Jianbiao, Zhang, Yanan, Niu, Xiaofei, ‘The Covid-19 Pandemic reduces trust behavior’, February 2021, link
- Tubantia, van Raaij, Leo, ‘BBB kamerlid beticht van verspreiden desinformatie, zelf ziet hij dat anders: ‘Dat maak jij ervan’, 11 December 2024, link
- University of Victoria, LibGuides, ‘How fake news spreads – fake news’’, 31 January 20025, link
- Warsaw Institute, Rogalewicz, Mikolai, ‘Russian disinformation vs. parliamentary elections in Slovakia’, 22 November 2023, link
- Wardle, Claire, ‘A conceptual analysis of the overlaps and differences between hate speech, misinformation and disinformation’, June 2024, link
- World Economic Forum, Feingold, Spencer, ‘The four key ways disinformation is spread online’, 9 August 2022, link