



Co-funded by
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.



Reading to Break Free

FUTURE CHRONICLES

Handbook to Counter Gendered Disinformation



Co-funded by
the European Union

Reading to Break Free

The Future Chronicles Handbook to Counter Gendered
Disinformation

Written by: Sebastian Arrosamena Mellgren (Traces&Dreams), Amalia Ranieri (Società Cooperativa Sociale Sinergie A.R.L.), Roya Baicu (Asociația In Varietate Concordia), Andrei Cornea (Asociația In Varietate Concordia), Ali Honaramiz (Europsky Dialog), Monika Kmetová (Europsky Dialog), Sara Mesa (Jóvenes hacia la Solidaridad y el Desarrollo – Jovesolidés), Beatriz Muñoz (Jóvenes hacia la Solidaridad y el Desarrollo – Jovesolidés), Aleksandra Radevska (Journalists for Human Rights), Natasha Dokovska (Journalists for Human Rights)

Edited by: Sebastian Arrosamena Mellgren

Layout by: Benedetta Beatrice

Graphics by: Giorgia D'Elia



Co-funded by
the European Union

Reading to Break Free - The Future Chronicles Handbook to Counter Gendered Disinformation © 2026 by Sebastian Arrosamena Mellgren (Traces&Dreams), Amalia Ranieri (Società Cooperativa Sociale Sinergie A.R.L.), Roya Baicu (Asociația În Varietate Concordia), Andrei Cornea (Asociația În Varietate Concordia), Ali Honaramiz (Europsky Dialog), Monika Kmeťová (Europsky Dialog), Sara Mesa (Jóvenes hacia la Solidaridad y el Desarrollo – Jovesolides), Beatriz Muñoz (Jóvenes hacia la Solidaridad y el Desarrollo – Jovesolides), Aleksandra Radevska (Journalists for Human Rights), Natasha Dokovska (Journalists for Human Rights). is licensed under Creative Commons Attribution Non Commercial 4.0 International. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc/4.0/>

Isbn: 978-91-531-8186-6



Co-funded by
the European Union

Contents

Prewords	6
Why This Book Matters	6
What This Book Is	6
Understanding the Landscape	7
If You Or Someone You Know Experiences Online Abuse	8
Final Words	9
1. Communication, Media, and Disinformation	11
1.1 Introduction	11
1.2 What is Communication?	12
1.3 Mass Communication	13
1.4 What are Media?.....	16
1.5 Disinformation.....	19
1.6 Fake News and Post-Truth.....	21
1.7 Gendered Disinformation	23
1.8 Conclusion	25
1.9 References	27
2. Comprehensive Guide on the Psychology of Communication	29
2.1 Introduction	29
2.2 Cognitive Biases and Perception in Communication	29
2.3 Networks and Algorithms	32
2.4 Coordination and Influence.....	34
2.5 Ethical Debunking	36
2.6 Conclusion	38
2.7 References.....	38
3. New Media and Social Platforms	40
3.1 Introduction: The Pocket Power of TikTok.....	40
3.2 Beyond TikTok - Reddit, Discord, and Snapchat in Youth Culture	42
3.3 Discord: From Gaming Chat to Virtual Hangouts and Activism.....	43
3.4 The Other Stage - Twitch, Clubhouse, and Telegram in Youth Social Lives	45
3.5 How Platforms Shape Misinformation and Misogyny.....	47
3.6 Conclusion: Navigating the Intersection of Gender, Disinformation, and Digital Spaces	49
3.7 References.....	50



4. Media Literacy - Techniques for critical thinking in the digital age	51
4.1 Introduction to Media Literacy	51
4.2 Dimensions of Media Literacy	52
4.3 The Importance of Critical Thinking in the Digital Age	56
4.4 Techniques and Strategies for Critical Thinking	58
4.5 Practical Skills in Media Literacy	61
4.6 References	63
5. Identifying disinformation patterns	64
5.1. Introduction to Disinformation Patterns	64
5.2. Information Disorder - Defining the Landscape	64
5.3: Key Actors Behind Disinformation	72
5.4: Risks and Alternatives to Debunking	75
5.5. Conclusion	76
5.6. References	76
6. Gendered Disinformation: Language and Representations	79
6.1 Introduction: The Gendered Weapon	79
6.2 Language, Gender, and Power	80
6.3 Conclusion: Gendered Disinformation and Paths to Resistance	83
6.4 References	85
7. Visual (Dis)Information	87
7.1 Why Images Matter More Than You Might Think.....	87
7.2 Visuals are Not Neutral Reflections of Reality	88
7.3 How Images Can Mislead and Manipulate	88
7.4 Verifying Visuals: Tools and Their Limits.....	90
7.5 Deconstructing Visual Narratives	92
7.6 Tactics for a Fairer Internet	93
7.7 Ethical use of Visuals	94
7.8 Putting it All Together.....	96
7.9 References	96
8. AI and the Media System	98
8.1 Introduction	98
8.2 Hypnocracy	99
8.3 The Fragility of Truth	100
8.4 Disinformation 2.0.....	101



Co-funded by
the European Union

8.5 AI, Journalism, and News Production.....	103
8.6 Global Inequalities and Algorithmic Colonialism	104
8.7 Conclusion	105
8.8 References	106
9 Fact-Checking and Source Verification	109
9.1 Introduction: The Art and Ethics of Verification	109
9.2 Truth as a Moving Target	109
9.3 Typologies of Fact-Checking Resources	111
9.4 The Craft of Checking the Truth	112
9.5 Automation, AI, and the Future of Fact-Checking.....	114
9.6 References	117



Co-funded by
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.



Co-funded by
the European Union

Prewords

Why This Book Matters

Disinformation is another word for lies. Gender-based disinformation means lies targeting people because of their gender. When false or manipulated information intentionally targets people based on gender, using stereotypes, sexualised imagery, and coordinated attacks, it damages personal well-being, silences voices, restricts participation, and reinforces unequal power dynamics that threaten our democratic values.

If you are reading this, you likely recognise these patterns. You may have seen coordinated campaigns targeting women in public roles, encountered deepfake images meant to humiliate, experienced harassment yourself, or are seeking to understand what friends, family, or community members face online. You are not alone. Millions encounter these harms daily. Your experiences, concerns, and voices are important, and this book exists because your safety and participation in digital spaces matter.

What This Book Is

This textbook is part of the Future Chronicles Project (<https://future-chronicles.eu/>), an Erasmus+ initiative co-funded by the European Union, involving six organisations from different countries. Each chapter is written by a partner organisation, offering a unique perspective on the topic. Future Chronicles targets young people aged 18–30. Our main aim is to provide you with tools and knowledge to counter gender-based disinformation online.

We hope that this book may serve as an inspiring guide. The book combines research with experiences, practical strategies, and thought-provoking questions to help you understand gendered disinformation, recognise its patterns, protect yourself and your community, and build resilience against these attacks. While the focus is mainly on the European context shaped by EU law, most principles can be applied globally.

This book addresses serious issues such as online abuse, non-consensual sharing of images, hate speech, coordinated harassment campaigns, and the gendered disinformation that exploits these tactics. We aim to present these topics honestly, showing both the harms and the ways, people are resisting and fighting back. The content is approached in a trauma-informed manner. However, some sections may still be upsetting, especially if you have personal experience with online abuse. Please engage with the material at your own pace. Feel free to take breaks, skip sections, step away, and return when you're ready.

If you need support, the resources listed below are staffed by trained professionals who offer confidential, non-judgmental assistance. Asking for support is a sign of strength.



Co-funded by
the European Union

Understanding the Landscape

What Is Gendered Disinformation?

Gendered disinformation is false or misleading information deliberately targeting individuals or groups based on their gender. It combines intentionally fabricated or manipulated content with gender-based bias, weaponising gendered narratives, sexualised images, and moral framing to discredit women, gender-diverse people, and movements advocating for gender equality.

Research indicates that women are frequently the targets of disinformation campaigns (Burke et al., 2023; Lind, 2023). Non-binary, LGBTQ+ individuals, refugees, and migrants are also commonly targeted (Sobieraj, 2020; ILGA-Europe, 2023). For instance, Black and Muslim women often face racist, sexist, and Islamophobic abuse through fake or real images (Amnesty International, 2018; Farkas & Bene, 2021). This form of gendered harm aims to hurt individuals based on their gender or gender identity, often to shame, stereotype, silence, or undermine their credibility. The situation is especially difficult for women involved in politics, journalism, and activism.

Studies across the EU highlight the scale of this problem. In 2024, a report from the Council of European Municipalities and Regions (CEMR) found that 70% of women in European politics have endured abuse and harassment, with countries like Germany and Ireland reporting even higher rates. Globally, female politicians are frequently targeted with abuse. A 2016 study across 39 countries showed that 80% of female parliamentarians faced psychological violence; 65% received sexually demeaning comments, and 44% experienced threats of violence, including death, rape, assault, or kidnapping. The situation has further accelerated because of accessible AI tools. AI tools for deepfake video and image creation are being used for non-consensual explicit content, i.e., pornography. A study from 2023 estimates that approximately 98% of online deepfake videos were pornographic, with about 99% featuring women (Valeski, 2024). These figures are not just statistics; they symbolise silenced voices, diminished participation, and the systematic exclusion of individuals whose perspectives are crucial to our democracies.

What Is Online Abuse?

Online abuse encompasses harmful online content like harassment, threats, or the misuse of intimate images. These kinds of technology-enabled violence are serious violations impacting millions of young people. The content types listed below are illegal in Europe:

- **Non-consensual Image Sharing** ("Revenge Porn"): Distributing private or intimate images without permission, with the intent to harm, humiliate, or control.
- **Deepfake Pornography**: Creating or distributing fake sexual images or videos using AI or other technology without consent.
- **Doxxing**: Sharing someone's private information online, such as their home address or phone number, often to harass, threaten, or intimidate them.



Co-funded by
the European Union

- **Child Sexual Abuse Material:** Sharing images or videos of child sexual abuse is strictly illegal across the EU, with dedicated enforcement agencies and hotlines monitoring and removing such material.
- **Hate Images and Threatening Content:** Posting messages that promote hate, violence, or harassment based on gender, race, religion, sexuality, disability, or other identity factors.

These acts are crimes. They cause measurable, documented harm: psychological trauma, social isolation, economic damage, and in severe cases, escalation to physical violence or self-harm.

The Current Legal and Social Landscape

The European Union has made significant legislative strides to combat online abuse. In May 2024, the EU adopted its first-ever **Directive on Combating Violence Against Women and Domestic Violence (Directive 2024/ 1385)**, a landmark piece of legislation that explicitly criminalises various forms of cyber violence. What makes this directive remarkable is that it explicitly recognises that violence against women stems from "historically unequal power relations" and frames it as a **structural issue requiring systemic change**, not an individual problem, but a societal one that demands collective action.

The **Digital Services Act (DSA)**, which became fully applicable across the EU in February 2024, provides additional protections for minors online. It requires digital platforms to ensure high levels of privacy, safety, and security. It provides mechanisms for reporting harmful content and holding platforms accountable for their responsibilities in creating safer digital environments.

These legal frameworks affirm a core truth: safety and dignity in digital spaces are essential. You have online rights similar to offline ones. Perpetrators can face accountability. Platforms are responsible. The EU Directive and DSA are in place because people need protection, and this effort is crucial for safeguarding democracy.

If You Or Someone You Know Experiences Online Abuse

If you are targeted by online abuse, know that you are not alone, and help is available.

If it feels safe to do so, consider:

1. Taking screenshots or saving evidence of the harmful content
2. Reporting it to the platform using their flagging features
3. Blocking or muting the person or account targeting you
4. Reaching out to a trusted person, a friend, family member, mentor, or professional who can help you seek help

Remember: **You can, but do not have to, report what has happened to the police.** You can also seek advice from support organisations; they can help you explore all your options,



Co-funded by
the European Union

legal, practical, and emotional, at your own pace and according to your own needs and circumstances.

Getting Support

The following resources are staffed by trained professionals who believe survivors and provide confidential, non-judgmental support:

- **Victim Support Europe:** <https://victim-support.eu/help-for-victims/info-on-specific-types-of-victims/sexual-violence-victims/>
- **Take It Down:** <https://cyberbullying.org/take-it-down>
- **StopNCII.org:** <https://stopncii.org/?lang=en-gb>
- **Right To Be!:** <https://righttobe.org/feel-support/>
- **End Violence Against Women:** <https://www.endviolenceagainstwomen.org.uk/new-campaign-survivors-demand-law-to-stop-image-based-abuse/>

If you are in crisis or immediate danger, please contact emergency services in your country or a local crisis line.

If You Encounter Abuse Targeting Someone You Know

If you encounter harmful content online, a good first step is to report it to the platform using the flagging feature. If the content explicitly targets someone, you can contact them privately in a direct message to offer support or suggest they seek legal advice.

A Note on Responding to Harmful Content

Many online safety guides recommend not reacting or commenting publicly on harmful posts. This is important: **any response, even disagreement, can help harmful content spread more quickly.** This happens because the algorithms used by (most) social media platforms tend to promote content that generates engagement, regardless of its accuracy. **It is often more effective to amplify positive content than to amplify hateful or false content by responding to it directly.**

Instead of publicly challenging harmful posts, consider:

- Reporting the content to the platform
- Privately supporting those targeted, offering solidarity, resources and information
- Make a post about hate and disinformation, why it is harmful and encourage others to report it.
- Sharing positive content about the group being targeted.

By amplifying what's genuine and compassionate rather than what's false and harmful, you help shift the digital landscape towards healthier information ecosystems.

Final Words

Engaging with this material requires bravery. If you have faced online abuse, your resilience is genuinely admirable. If you are reading this to support others, your solidarity makes a



Co-funded by
the European Union

significant difference. If you believe in a fairer internet and a more equal society, this work is meant for you.

Together, survivors, allies, educators, researchers, policymakers, and individuals from around the world, we can create digital spaces where everyone, regardless of gender, race, sexuality, ability, or other identities, can engage safely, fully, and without fear.

Change is possible. When we do it together.

Let's begin.

In Solidarity,
Sebastian Arrosament Mellgren



Co-funded by
the European Union

1. Communication, Media, and Disinformation

Written by Amalia Ranieri, Sinergie, 2025

1.1 Introduction

The ability to communicate is one of the most fundamental capacities of human life. It is through communication that we form relationships, build communities, transmit knowledge, and give meaning to our existence. However, communication is never neutral: it is always embedded in cultural, technological, and political contexts that shape not only what is said, but also how it is said, who gets to speak, and who is heard.

This chapter explores the layered relationship between **communication, media, and disinformation**, providing both a theoretical foundation and practical tools for critical engagement. We begin by asking a seemingly simple question: What is communication? From there, we move into the world of **mass communication**, which introduces organisational structures, technologies, and audiences far larger than interpersonal exchanges. The discussion then turns to the **definition of media**, a concept that resists easy categorisation but remains central for understanding how symbolic forms circulate in society.

The second half of the chapter addresses the darker side of communication: **disinformation**. Starting from historical cases such as Orson Welles' War of the Worlds broadcast, we trace how falsehoods spread and why they are believed. We then explore the emergence of **fake news** and the age of **post-truth**, analysing the psychological, social, and technological dynamics that make disinformation so pervasive in our digital ecosystems. Finally, we address **gendered disinformation**, a particularly harmful form of manipulation that disproportionately targets women in politics, journalism, and public life.

Throughout the chapter, **theoretical insights** are combined with **real-world stories, critical reflection questions, and practical exercises**. The aim is not only to provide knowledge but also to nurture essential media literacy, i.e., the ability to interpret, question, and resist manipulative communication.

By the end of this chapter, you will be able to:

- Distinguish between different models of communication.
- Understand the main features of mass communication and media systems;
- Recognise how disinformation operates, both historically and today;
- Analyse the role of fake news in the post-truth environment.
- Identify and reflect on the specific dynamics of gendered disinformation.
- Apply critical tools to your own media environment.



Co-funded by
the European Union

This is not a purely academic exercise. In an age of algorithmic platforms, viral falsehoods, and weaponised disinformation campaigns, communication is no longer only about expression, but it is also about power. To study communication today means to examine the contested terrain where democracy, identity, and truth are negotiated.

1.2 What is Communication?

At first glance, communication seems so evident that it hardly requires definition: we talk, we listen, we write, we post on social media – surely that is communication. However, scholars have debated the meaning of the term for more than sixty years, producing a wide variety of models and frameworks. Understanding these models is essential because each one highlights different aspects of how messages circulate and how meaning is made. A first distinction must be made between **communication** and **information**. The two words are often used interchangeably, but their roots reveal subtle differences. The Latin *in-formo* – to give form – suggests that information is an act of shaping, of imposing structure upon raw material. Communication, by contrast, implies *communicare*: to share, to make common. In this sense, information is often one-directional, while communication is relational. A journalist who reports the news is primarily informing, but when citizens discuss the news with each other, interpret it, and negotiate its meaning, they are communicating.

This difference also appears in social hierarchies. Consider the classroom: the teacher is positioned on a raised platform, literally elevated above the students, symbolising the asymmetry of knowledge. The teacher "informs", but genuine communication emerges only when students engage in dialogue, ask questions, and reshape the message in their own terms.

Over time, scholars have proposed a variety of **conceptual models** of communication. Each reflects the intellectual climate of its era and emphasises specific dimensions while obscuring others. Here we explore five influential perspectives.

1. Communication as Contact:

One of the oldest understandings of communication views it primarily as contact. To communicate is to establish a link, a connection, or a bridge between people. Infrastructures such as railways or postal systems were once described as forms of communication precisely because they enabled contact across distance. In this sense, communication is less about meaning and more about connection itself.

2. Communication as Transfer of Resources and Influence

Another model, rooted in behaviourism, conceives communication as a process of transferring resources, energy, or influence from one agent to another. This view reduces communication to a **stimulus–response mechanism**: A emits a signal, B reacts. Early theories, such as the "magic bullet" or "hypodermic needle" model, portrayed media



Co-funded by
the European Union

messages as injections directly shaping passive audiences. Although later research rejected such simplifications, the model remains a valuable reminder of communication's capacity to influence behaviour.

3. Communication as Transmission of Information

After the Second World War, technological metaphors became dominant. The Shannon and Weaver model (1949) described communication as the transmission of information from a source to a receiver, through a channel, with possible "noise" interfering. This model was groundbreaking for engineering and cybernetics, but it also reinforced the supremacy of the source, which "informs" and thus controls the shaping of meaning.

Researchers who study signs and meanings like Umberto Eco nuanced this perspective by introducing the concepts of the "Model Author" and "Model Reader", reminding us that texts anticipate interpretive pathways but cannot determine them fully.

4. Communication as Sharing

A more humanistic approach highlights communication as a form of sharing. From the Indo-European root *no-* ("to make common"), this perspective emphasises participation, reciprocity, and co-construction of meaning. Barnett Pearce (1989) argued that the true significance of communication lies not in the transmission of information, but in the social processes by which people negotiate reality together. For example, when communities deliberate on local issues, communication is less about passing data and more about generating intersubjective agreement.

5. Communication as Inference

Finally, another approach conceives communication as inference: the construction of meaning through clues, implicatures, and contextual knowledge. Consider overhearing two passengers on a train:

- "They even print it on napkins now."
- "But they never give you good Tuscan wine."
- "The quality has improved compared to the past."

The conversation makes sense not because of explicit information, but because listeners use shared knowledge and situational context to generate meaning. Communication is less about transmission and more about collaborative interpretation.

1.3 Mass Communication

If communication is the fabric of human interaction, mass communication is the loom on which modern societies have been woven. Unlike interpersonal exchanges, which unfold face-to-face, mass communication relies on technical and organisational apparatuses to reach vast, dispersed audiences. At first glance, the difference might seem purely quantitative: one-to-one becomes one-to-many. However, the implications are far more profound.

The Problem of Feedback



Co-funded by
the European Union

One of the central distinctions between interpersonal and mass communication lies in the feedback they provide. In a classroom, a lecturer can read the room: nodding heads and attentive gazes encourage continuation, while restless fidgeting or distracted expressions prompt adjustments. This improper feedback – feedback that occurs during the act of speaking itself – allows for dynamic adaptation.

In mass communication, however, such immediacy is impossible. A television anchor does not see the millions of viewers behind the camera lens. A newspaper editor does not know, in the moment of writing, how readers will interpret tomorrow's headlines. Feedback becomes deductive, gathered indirectly through sales figures, ratings, surveys, or, today, digital metrics such as clicks and likes. However, even these measures come after the fact, offering only partial glimpses into audience response.

The arrival of digital technologies has blurred this distinction somewhat. Interactivity – comment sections, live polls, real-time messaging – promises to restore immediacy. However, such feedback remains asymmetrical: platforms capture data far more efficiently than audiences shape content. Despite technological shifts, the imbalance persists.

This imbalance leads to another defining trait: the **asymmetry** between sender and receiver. In interpersonal communication, the relationship can be relatively symmetrical, even if not perfectly equal. Friends in conversation, for instance, take turns speaking and listening, negotiating meaning together. In mass communication, the sender is typically an organisation with structured power (newsrooms, film studios, broadcasting corporations) while the audience is anonymous, dispersed, and often imagined. The relationship is therefore asymmetrical: one speaks, many listen. To be sure, audiences are not passive.

Letters to the editor, call-in radio shows, or, more recently, hashtags and viral memes demonstrate that receivers can respond and even exert pressure. Still, these responses remain mediated and constrained within structures broadly defined by producers.

John B. Thompson's insight remains crucial: mass communication is not a monologue, but neither is it an open dialogue. It is a **structured exchange**, institutionalised and generalised, in which symbolic forms circulate across social space.

Defining "Mass"

The very term "mass" complicates the picture. Does "mass communication" require a numerically large audience? Not necessarily. A niche cable channel with a modest but dispersed viewership still qualifies as mass communication, because its messages are produced by organisations for broad distribution, not for face-to-face reciprocity. What matters is not size alone, but **structure**: the separation between production and reception and the reliance on distribution technologies.

This is why buying a newspaper is not the same as engaging in interpersonal dialogue. The purchase is both a selection and a preliminary form of feedback, but the newspaper remains a symbolic product, designed for replication and broad circulation.

Broadcasting as Technique and Mode

The concept of **broadcasting** helps to crystallise these ideas. Literally meaning "to scatter seeds", broadcasting refers to one-to-many transmission, where a single sender disseminates the same message simultaneously to many receivers. Unlike a speech in a



Co-funded by
the European Union

public square, broadcasting does not depend on physical co-presence but on technological mediation: radio waves, television signals, or digital streams.

For decades, broadcasting was the dominant paradigm of mass communication. A family gathered around the radio or television set was not merely consuming information; they were participating in a cultural ritual, synchronised with millions of others across space. The authority of broadcasting lay in its simultaneity and reach. Nevertheless, broadcasting also created vulnerabilities. Its asymmetry concentrated power in the hands of those controlling the means of transmission, whether states, corporations, or political movements. From Nazi propaganda in the 1930s to Cold War media battles, broadcasting demonstrated both the promise and peril of centralised communication.

Technologies of Mass Communication

To understand how mass communication operates, it is helpful to distinguish three kinds of technologies (Colombo, 1994):

1. **Transmission technologies** eliminate or reduce spatial distance. Telegraphs, telephones, radio, and the internet enable real-time transmission of messages across vast distances.
2. **Representation technologies** provide partial representations of reality – photography, cinema, television – allowing events to be seen and heard by those far removed from them.
3. **Reproduction technologies** enable infinite replication: from the printing press to vinyl records to digital copies, these technologies allow cultural products to circulate widely and repeatedly.

Each has reshaped not only communication but also society itself. The gramophone, for instance, did not merely reproduce music; it transformed the economics of cultural production, enabling the rise of recording stars like Enrico Caruso. His 1904 recording of Leoncavallo's *Mattinata* was not simply an artistic performance, but an artefact designed for a technical medium. The technology shaped the art, and in turn, the art shaped social demand.

Commodification of Symbolic Forms

Another defining feature of mass communication is the **commodification of symbolic goods**. Books, newspapers, films, and broadcasts are not only cultural artefacts but also commodities circulating in markets. Their value lies not only in meaning but also in exchange: the ability to sell copies, attract advertising, or capture subscriptions.

This commodification has consequences. It encourages repetition and formulaic content, since what sells tends to be reproduced. It blurs the boundary between information and entertainment, as commercial imperatives reward what captures attention rather than what informs. Furthermore, it raises ethical and political questions about who benefits from the circulation of symbolic forms.

Separation of Production and Reception

Mass communication also entails a **structural separation** between the contexts of production and reception. A television show is produced in a studio with lights, cameras,



Co-funded by
the European Union

and scripts, but consumed in private living rooms scattered across cities and nations. This separation is both spatial and social: producers and receivers inhabit different worlds, with various interests and perspectives.

Digital technologies complicate this picture, enabling audiences to become producers through blogging, vlogging, and tweeting. Nevertheless, even here, the infrastructure is controlled by platforms whose algorithms, policies, and business models shape what circulates. The separation remains, even if blurred.

Audiences: From Mass to Segments

One of the most persistent myths of mass communication is the "undifferentiated audience": the idea of a homogeneous mass of receivers. In reality, audiences have always been diverse, segmented by class, gender, culture, and geography. With the advent of digital media, segmentation has intensified. Algorithms tailor feeds to individual preferences, creating personalised "mass communication" that paradoxically fragments the public sphere.

This raises a paradox. On one hand, audiences are more empowered: they choose, click, and comment. On the other hand, they are more predictable: algorithms anticipate preferences and reinforce them, creating echo chambers. The "mass" has not disappeared, but it has mutated into **networked publics**, overlapping yet fragmented, visible yet opaque.

1.4 What are Media?

So far, we have distinguished between interpersonal and mass communication, tracing how technologies and organisational structures expand the reach of messages. However, an essential question remains: what exactly are media? The word is used so often – in politics, journalism, and everyday speech – that it risks becoming self-evident. Nevertheless, when we try to define it precisely, complexities emerge. Are books media? Is the telegraph? A smartphone? A video game? Even philosophical systems, argued Marshall McLuhan, can be understood as media.

In the 1960s, Canadian scholar **Marshall McLuhan** famously declared that "the medium is the message". For McLuhan, media were not just channels for transmitting information, but extensions of human senses and capacities: a wheel is a medium of the foot, the electric light a medium of vision. In this expansive sense, media included both material technologies (television, print, cinema) and cultural forms (poetry, scientific paradigms). This sweeping definition was both liberating and problematic. On the one hand, it broke with technological determinism and encouraged scholars to view media as cultural environments that shape perception and society. On the other hand, its breadth risked diluting the concept: if everything is media, then the term explains little.

To overcome this indeterminacy, Italian sociologist **Fausto Colombo** proposed a more precise yet flexible definition: media are socio-technical apparatuses that mediate communication between subjects.

This definition highlights several key elements:



Co-funded by
the European Union

- Media are **technologies**, but never merely technologies. They are embedded in social, economic, and cultural systems.
- Media involve **actors** (individuals, groups, institutions) who use, regulate, and interpret them.
- Media operate within **networks**, linking production, distribution, and reception across space and time.
- Media are simultaneously **technological and cultural circuits** that produce meanings and shape identities.

By conceiving media as socio-technical systems, Colombo moves us beyond simplistic binaries, technology versus culture, structure versus agency, and invites us to see media as relational processes.

Central to Colombo's definition is the notion of **mediation**. To mediate is to intervene between, to translate, to transform. Communication is never immediate: it always passes through mediators, whether words, images, devices, or institutions.

British scholar **Roger Silverstone** compared mediation to translation in literature: never complete, always partial, sometimes contested, often creative. Mediation is not a transparent window on reality but a process that shapes what is visible, thinkable, and sayable. This is why media are powerful: they do not simply carry messages, they transform them.

Please think about how the same event (a protest, for instance) looks different depending on whether it is represented in a newspaper, a livestream, or a TikTok video. Each medium frames, filters, and amplifies certain aspects while silencing others. Mediation is always political, because it involves choices about representation, visibility, and voice.

Media as Socio-Technical Apparatuses

Understanding media as socio-technical apparatuses opens several lines of inquiry:

1. **Media and control** – Who owns and regulates media systems? How do power relations shape what circulates? For instance, state television in authoritarian regimes serves very different functions than public service broadcasters in democratic contexts.
2. **Media and representation** – How do media depict social actors, issues, or events? Representation is never neutral: choices of framing, imagery, and narrative carry ideological weight.
3. **Media and space** – Media reshape how we experience space and time. Television's "remote vision" allowed viewers to witness distant events in real time, collapsing geographical boundaries. The internet goes further, producing what Pierre Lévy called a "space of knowledge" beyond physical geography.
4. **Media as cultural circuits** – Media are not isolated channels but part of larger cultural industries. Films, music, fashion, advertising, and social media interact in complex circuits, reinforcing or contesting dominant narratives.

Another reason why defining media is difficult is their **multi-coded nature**: media combine different sign systems (visual, auditory, textual) often within the same artefact. A film



Co-funded by
the European Union

blends images, dialogue, and sound; a website combines text, video, and interactive elements. This hybridity complicates classification and underlines the creative power of convergence.

The rise of digital platforms further accelerates this complexity. A smartphone is simultaneously a telephone, a camera, a gaming device, a newspaper, and a marketplace. Such multifunctionality challenges older distinctions between media types and demands new analytical frameworks.

Media, Mediation, and Power

Because mediation is never neutral, the study of media always entails questions of **power**. Which voices are amplified, and which are silenced? Who gets to frame the narrative? Whose stories become visible, and whose remain invisible?

Consider television's role in politics. For much of the 20th century, it gave leaders unprecedented visibility, allowing them to enter citizens' living rooms. However, visibility can also be dangerous; leaders who are overexposed risk losing control of their image and becoming vulnerable to scandal or parody. Media visibility is a double-edged sword: a source of power, but also a site of fragility.

Digital platforms intensify this paradox. Social media empowers users to become content producers, but platform algorithms privilege certain kinds of visibility, those that generate engagement, often through outrage or sensationalism. Mediation thus becomes entangled with commercial logics, reinforcing inequalities of voice and attention.

One way to synthesise these insights is through a **mediological perspective**: rather than treating media as neutral tools, mediology examines the ways they structure social life, mediate cultural memory, and shape political possibility. French theorist Régis Debray argued that media are central to the transmission of culture across generations: they are how societies remember, imagine, and project themselves into the future.

Real-Life Example: Television as Remote Vision

In the 1990s, John B. Thompson analysed television as a medium of **remote vision**: the ability to "see the world with one's own eyes" across distance. For the first time, citizens could witness events unfold in faraway places: wars, elections, and natural disasters. This transformed political accountability, as leaders could no longer hide behind distance. At the same time, it created new vulnerabilities: images taken out of context or manipulated could shape public perception more strongly than facts.

Today, livestreaming and social media extend this principle. A smartphone in a protester's hand can transmit images worldwide in seconds. The potential for empowerment is immense, but so is the potential for distortion and manipulation.

Reflection Activity

Choose a media artefact you engage with daily, for example, a news app, a social media platform, or a podcast. Analyse it using Colombo's definition of media as a socio-technical



Co-funded by
the European Union

apparatus. Who produces it? Through which technologies? How does it mediate communication between subjects? What power dynamics does it reproduce or contest?

1.5 Disinformation

If communication is about sharing meaning and building understanding, disinformation is about breaking that trust. It is the deliberate creation and spread of false or misleading information with the intention of deceiving audiences, shaping perception, or manipulating behaviour. While misinformation may arise from error or confusion, **disinformation is intentional**: a strategy deployed to influence public opinion, damage reputations, or advance political agendas.

Though the term feels contemporary, the phenomenon is not new. Rumours, propaganda, and forgeries have accompanied communication since the earliest political struggles. What changes are brought about by the **technologies, scales, and contexts** through which disinformation circulates?

The War of the Worlds

On the evening of October 30, 1938, American broadcaster CBS aired a radio adaptation of H. G. Wells' novel *The War of the Worlds*. Directed and narrated by Orson Welles, the programme simulated a live news report of a Martian invasion. Despite multiple announcements clarifying that the broadcast was fictional, thousands of listeners panicked. Some fled their homes, others flooded police stations and newspapers with calls, and highways were clogged with cars.

Why did so many mistake a radio play for reality? The sociologist **Hadley Cantril** sought to answer this question in his study *The Invasion from Mars* (1940). He identified several factors that made the broadcast particularly convincing:

- The **realistic tone** of alternating news bulletins and eyewitness accounts.
- The **trust placed in radio** as a medium of authoritative information.
- The use of **experts and officials** in the script lends credibility.
- References to **real locations**, such as New Jersey and Manhattan, give the story a sense of geographic plausibility.
- The fact that many tuned in late, missing the opening disclaimer.

Cantril also analysed the audience's responses, categorising listeners into four groups:

1. **Critical listeners**, who recognised the story as fiction;
2. **External checkers**, who verified its accuracy through other sources;
3. **Uncertain verifiers**, who looked for confirmation but misinterpreted what they saw;
4. **Unquestioning believers** who accepted the broadcast at face value.

This typology introduced the concept of **critical ability**: the capacity to evaluate a stimulus, compare it with existing knowledge, and draw accurate conclusions. Education, personality



Co-funded by
the European Union

traits, and even religious outlook influenced whether listeners interpreted the programme as fiction or reality.

Though remembered today as a curious media anecdote, *The War of the Worlds* illustrates enduring dynamics: the authority of media, the psychology of panic, and the diversity of audience responses. It foreshadowed contemporary debates about why some people fall prey to disinformation while others resist.

Disinformation, of course, is not limited to fictional accidents. During the 20th century, authoritarian regimes mastered the art of propaganda, using mass media to fabricate realities. Nazi Germany, for instance, deployed radio and film to cultivate myths of national destiny and demonise enemies. The Cold War witnessed a proliferation of disinformation campaigns, with both the Soviet Union and the United States planting forged documents, doctored images, and fabricated stories to discredit rivals.

These historical precedents highlight that disinformation is not merely about lies. It is, again, about **power**: who controls narratives, whose voices are amplified, and how symbolic forms shape collective behaviour.

Fast-forward to the present, and the dynamics of disinformation have shifted: today's digital ecosystem makes it easier than ever to produce, circulate, and consume false content. The barriers to entry are low: anyone with a smartphone can generate convincing images, videos, or posts. At the same time, platforms reward virality, privileging content that provokes strong emotional reactions, often outrage or fear. Unlike the relatively centralised broadcasting of the past, disinformation now thrives in **networked environments**. Memes, hashtags, and viral videos spread horizontally, not just vertically. This gives disinformation the speed of contagion, moving through social networks like wildfire. Moreover, algorithmic systems amplify such content, creating self-reinforcing loops.

Visibility, Vitrinisation, and the Spectacle

Understanding contemporary disinformation also requires looking at how media transform visibility: sociologist **Giovanni Codeluppi** coined the term vitrinisation to describe the process by which social life is turned into spectacle, like objects displayed in a shop window. In the age of social media, identity itself becomes a performance staged for constant exposure. This logic of spectacle creates fertile ground for disinformation. False stories, sensational images, and conspiratorial claims thrive precisely because they are **spectacular**: they grab attention, invite emotional investment, and demand sharing. Visibility becomes both currency and vulnerability: to be seen is to risk distortion.

John B. Thompson's work on the changing boundaries between public and private further illuminates this shift. Where once public figures could maintain controlled personas, today hypervisibility makes them constantly exposed to scrutiny, manipulation, and attack. Disinformation exploits this exposure, weaponising images and narratives to undermine reputations and credibility.



Co-funded by
the European Union

Disinformation and Democracy

The consequences extend far beyond individual reputations. Disinformation erodes **trust** – in media, institutions, and even in the possibility of shared truth. When falsehoods circulate unchecked, citizens lose confidence in their ability to distinguish fact from fiction. This breeds cynicism, disengagement, or, worse, radicalisation.

In democratic societies, disinformation undermines informed deliberation, replacing reasoned debate with emotional polarisation. In authoritarian contexts, it becomes a tool for repression, legitimising crackdowns under the guise of combating "fake news". Either way, disinformation corrodes the conditions of democratic life.

Real-Life Example: COVID-19 and Infodemic

The global pandemic of 2020 offered a stark example of the dangers of disinformation. Alongside the spread of the virus came an **infodemic**: a flood of false information about cures, vaccines, and conspiracy theories. From baseless claims that 5G towers spread the virus to rumours of microchips in vaccines, disinformation undermined public health measures and fuelled distrust in science. The World Health Organisation recognised that managing the infodemic was as crucial as managing the pandemic itself. This illustrates how disinformation can directly endanger lives, not only political systems.

1.6 Fake News and Post-Truth

In recent years, few terms have travelled as quickly across academic debate, political discourse, and everyday conversation as **fake news**. Once used informally to describe dubious journalistic practices, it is now a central concept in the struggle to understand how truth is challenged in a digital society. Nevertheless, the term itself is problematic: vague, contested, and often weaponised. To analyse it critically, we must unpack its multiple dimensions – epistemic, intentional, sociological, and formal – while situating it within the broader condition of **post-truth**.

The phenomenon is not new. False stories, satirical hoaxes, and propaganda campaigns have always accompanied media history. What distinguishes the present is the **speed, scale, and reach** enabled by digital platforms.

In 2016, the Oxford Dictionary declared "post-truth" the word of the year, defining it as a situation in which "objective facts are less influential in shaping public opinion than appeals to emotion and personal belief". The following year, the Collins Dictionary named "fake news" its word of the year, noting its explosive use in political rhetoric. This was no coincidence: the Brexit referendum in the United Kingdom and the election of Donald Trump in the United States had both been marked by an unprecedented surge of online disinformation.

The Brexit campaign, for example, was infused with the now infamous claim that the UK was sending £350 million a week to the EU, money that could instead fund the National



Co-funded by
the European Union

Health Service. Although quickly debunked, the slogan on a red bus proved remarkably powerful. Similarly, during the U.S. election, the so-called "Pizzagate" conspiracy alleged that Hillary Clinton was running a child-trafficking ring out of a Washington pizzeria. Absurd though it was, the story spread widely on social media and even inspired an armed attack on the restaurant. These events revealed that fake news was not a marginal curiosity but a **structural force**, capable of shaping political outcomes and destabilising trust in democratic institutions.

Defining Fake News: Four Properties

Scholars propose that fake news can be analysed through four dimensions:

1. **Epistemic (Knowledge-related)** – the relation of a truth statement. Fake news often involves outright falsehoods, but not always. Sometimes it is literally accurate yet misleading by implication. For example, reporting that "47 robberies have occurred since the arrival of refugees" suggests causality where none exists. The statement is factually correct, but the implication is false.
2. **Intentional** – the motivations of the creator. Does fake news require intent to deceive? Some argue yes: a claim is fake only if its author knows it is false. Others emphasise that dissemination matters regardless of intention. A person may share phoney news in good faith, genuinely believing it to be true. Intention, therefore, must be considered at the level of **producers** rather than **propagators**.
3. **Sociological** – the scale of dissemination. Fake news is not simply a false statement; it is falsehood amplified across networks. Some argue it becomes phoney news only when widely shared, while others stress that the **aim to reach mass audiences** is sufficient, even if actual virality is limited.
4. **Formal** – the appearance of credibility. Fake news mimics the form of legitimate journalism – articles, headlines, or broadcasts – so that audiences mistake it for genuine reporting. This imitation gives it persuasive power.

Taken together, these properties explain why fake news is so insidious: it blurs boundaries between true and false, fact and opinion, satire and reporting, sincerity and manipulation. Understanding fake news also requires identifying the different **actors** involved. We can distinguish three roles:

Producers – individuals or organisations who create fake news. They may be driven by ideology, profit (through clicks and advertising revenue), or political strategy.

Receivers – those who encounter fake news. Not all receivers believe what they read; some may recognise it as false or consume it as entertainment. The most vulnerable are those who accept it uncritically.

Propagators – users who share fake news with their networks, regardless of whether they believe it. Propagators are crucial: by retweeting, reposting, or forwarding, they extend the reach of disinformation beyond its original audience.

Asingle person may occupy all three roles at different times. For instance, a receiver who believes a false story can become a propagator by sharing it, inadvertently amplifying disinformation.



The extraordinary virality of fake news cannot be explained by deception alone. **To understand why false stories circulate so widely, we need to examine the interplay among psychology, cognition, and social dynamics.** The psychological dynamics at play are explored in the next chapter. Theoretically, we continue to hear about the **social structures** of communication. Social media networks often function as epistemic bubbles: spaces where users are only exposed to like-minded voices. In such environments, alternative perspectives are simply absent, excluded by algorithmic design and the composition of one's social connections. More insidious still are echo chambers, where dissenting sources are not only lacking but actively discredited. Inside an echo chamber, members reinforce each other's beliefs while dismissing outside information as propaganda or lies. Trust becomes concentrated within the group, and suspicion is directed outward.

A further mechanism, known as group polarisation, intensifies the problem. When people deliberate in homogeneous groups, they often adopt more extreme positions than they initially held. In online spaces where like-minded individuals cluster, fake news can push groups toward increasingly radical interpretations, fuelling outrage and deepening divisions.

Taken together, these cognitive tendencies, psychological shortcuts, and social dynamics reveal why fake news spreads so effectively. It is not simply a matter of malicious producers deceiving passive audiences. Instead, it reflects the vulnerabilities of human cognition, the ways our memories and desires work, and the structures of our digital communication environment. Fake news succeeds because it resonates with how people think, feel, and belong.

Post-Truth Politics

The prominence of fake news has been amplified by the cultural condition known as **post-truth**. In post-truth politics, the line between fact and opinion blurs, and emotional resonance trumps empirical evidence. Politicians exploit this climate by making statements that "feel true" to their supporters, even when demonstrably false.

This does not mean truth has disappeared. Instead, truth competes with other logics (emotion, identity, affect) in shaping public discourse. Fake news thrives because it aligns with these logics, offering emotionally satisfying narratives in place of complex realities.

1.7 Gendered Disinformation

Not all disinformation spreads in the same way, nor does it harm all public figures equally. Women in positions of visibility are often confronted with a particular kind of hostility, one that does not simply question their ideas but undermines their legitimacy as political and professional actors. This systematic undermining is what we can call **gendered disinformation**: a phenomenon that combines manipulation of information with the enduring structures of sexism and misogyny. It is a form of what Pierre Bourdieu once described as symbolic violence: subtle yet pervasive attacks that discredit women not



Co-funded by
the European Union

based on their policies but on the fact that they are women. For a more analysis of gendered disinformation, see Chapter 6.

Unlike men, who are more often criticised for their political or ideological stances, women are singled out for their bodies, their private lives, or their supposed character flaws. They are sexualised or ridiculed, turned into memes or mocked with infantilising nicknames, and often saddled with fabricated quotes designed to spark outrage. Social media algorithms ensure that these attacks travel quickly and widely, magnifying their reach far beyond the initial insult. Moreover, the consequences rarely remain online: threats, intimidation, and reputational harm spill over into the offline world, discouraging women from public participation.

The story of **Maria Ressa**, journalist and Nobel Peace Prize laureate from the Philippines, is a striking example. From the moment she began exposing corruption and authoritarianism in her country, she became the target of relentless online harassment. At the height of these campaigns, she received up to ninety abusive messages per hour, many of them misogynistic in nature. She was called a "witch" or dismissed as "not a woman", her image twisted in demeaning memes that circulated widely on Facebook. The online assault paved the way for something even more dangerous: the criminalisation of her journalism. The rhetoric of disinformation and misogyny created an environment in which she could be prosecuted under "cyberlibel" laws, a chilling reminder that digital abuse and state repression are not separate phenomena but deeply connected.

In Italy, **Laura Boldrini**, former President of the Chamber of Deputies, faced a form of disinformation that was different but equally revealing. Over the years, she was repeatedly associated with statements she never made (claims that Italians should host migrants in their homes, or that the Italian juridical system should not punish crimes committed by migrants). These so-called "phantom quotations" were paired with images of urban decay or unrelated photos depicting migrants, weaving a narrative of threat and betrayal. Even when fact-checkers debunked these stories, the damage had already been done; the falsehoods had entered the public imagination. At the same time, Boldrini was targeted with pornographic photomontages and violent fantasies. One local politician suggested publicly that she be locked in a house with migrants, an act so egregious that it led to a court conviction. Her case demonstrates not only the personal cost of gendered disinformation but also the difficulty of repairing reputational harm once a lie has been allowed to spread.

In Canada, **Catherine McKenna**, Minister of the Environment, encountered another recurring strategy: infantilisation. She was consistently referred to as "Climate Barbie", a nickname that trivialised her expertise and undermined her authority as a policymaker. The label was repeated so widely that it became shorthand for dismissing her, making it easier for fabricated stories about her environmental policies to gain traction. Claims that she wanted to ban fireplaces or impose taxes on family holidays circulated alongside the nickname, the sexist frame giving them extra credibility. The hostility she faced did not remain confined to the digital sphere: her constituency office was vandalised with



Co-funded by
the European Union

misogynistic graffiti, she received death threats, and eventually she withdrew from frontline politics. Her experience shows how disinformation, when shaped by sexism, can force women out of public roles altogether.

Taken together, these stories make it clear that gendered disinformation is not an unfortunate series of isolated incidents but a **structural phenomenon**. It works by combining lies with sexism, by using digital platforms to magnify harassment, and by blurring the line between symbolic attacks and material consequences. Ressa's case highlights how online violence legitimises institutional repression; Boldrini's illustrates the enduring power of fabricated quotations tied to xenophobic narratives; McKenna's shows how sexist stereotypes can reduce complex expertise to caricature.

The broader implications for democracy are profound. A society where women are systematically attacked, ridiculed, and delegitimised is a society where democratic debate is impoverished. Gendered disinformation narrows the field of participation, discouraging women from seeking leadership roles and creating toxic environments that stifle diverse voices. It reinforces patriarchal power by normalising misogyny and presenting it as part of everyday political contestation. Over time, this not only silences individuals but weakens the inclusiveness and quality of democratic life itself.

However, these stories also reveal resilience and resistance. Maria Ressa has used her international platform to denounce online violence and call for global solidarity in defending press freedom. Laura Boldrini has transformed her experience into public advocacy for stronger protections against online hate. Catherine McKenna continues to work on environmental issues, refusing to be silenced, even if from a less exposed position. Each, in her own way, demonstrates that women subjected to disinformation are not simply victims but also agents who contest and expose these attacks.

Reflection Activity

Think about the three cases described. What struck you most: the scale of the harassment, the inventiveness of the lies, or the consequences for democratic participation? How might you design a campaign to counter gendered disinformation? Would you focus on fact-checking, creating supportive networks, legal reform, or building empowering counter-narratives?

1.8 Conclusion

Over the course of this chapter, we have moved from the foundations of communication to the darker realities of disinformation, tracing how meaning is shared, mediated, manipulated, and resisted. What began as a simple question, What does it mean to communicate? , opened into a journey through the asymmetries of mass communication, the complexities of media as socio-technical systems, and the vulnerabilities that arise when truth competes with spectacle, ideology, and manipulation.



Co-funded by
the European Union

We saw that communication is never neutral. It can be contact, transmission, sharing, or inference; it can inform, but it can also mislead. The classroom, the newspaper, and the radio broadcast each reveal the subtle but significant differences between information and genuine dialogue. We then considered the leap from interpersonal to mass communication, recognising how the separation between production and reception, the commodification of symbolic goods, and the authority of broadcasting reshaped modern societies. Our exploration of media reminded us that technologies are never mere tools. They are mediators, shaping not just what we see but how we see it, not just what we say but what can be said. Mediation, as Silverstone observed, is like translation: always incomplete, always political. This awareness prepared us to face the challenges of disinformation, both historical and contemporary. From Orson Welles' War of the Worlds to the COVID-19 infodemic, from fabricated headlines to algorithmically amplified falsehoods, we have seen how manipulation exploits our trust in the media and our own cognitive and social vulnerabilities.

Fake news, we learned, spreads not only because someone intends to deceive but because human beings are wired to seek coherence, to trust what is familiar, and to remain loyal to the groups to which we belong. Disinformation thrives in the fertile soil of confirmation bias, repetition, bubbles, and echo chambers. In the age of post-truth, it is not that truth has disappeared, but that it must compete with narratives that are emotionally satisfying, spectacular, and aligned with identities.

Perhaps most striking was the discussion of gendered disinformation. Here, the manipulative power of disinformation intersects with patriarchal structures, producing attacks that do not simply mislead but actively delegitimise women as public subjects. The stories of Maria Ressa, Laura Boldrini, and Catherine McKenna remind us that the consequences are not abstract: they involve threats, reputational harm, and the silencing of voices essential to democracy.

What, then, can we take away? First, that critical media literacy is not an optional skill but a democratic necessity. To live in the digital age requires more than consuming information; it requires questioning, verifying, comparing, and reflecting. It means recognising the difference between communication and propaganda, between dialogue and manipulation. Second, the fight against disinformation cannot be left to individuals alone. Platforms, institutions, educators, and citizens must all share responsibility in building environments where truth is valued and diverse voices can flourish. Third, that disinformation is not only about lies but about power: who controls visibility, whose voices are amplified, and who is pushed into silence.

As you reflect on this chapter, think about your own media practices. How do you respond when you encounter dubious stories online? Which voices do you trust, and why? How do algorithms shape the information you see, and what might be missing from your feed? Most importantly, how can you use your awareness not only to protect yourself but also to support others, to create a digital world that is fairer, more inclusive, and more resilient to manipulation?



Co-funded by
the European Union

1.9 References

Appadurai, A. (1996). *Modernity at large: Cultural dimensions of globalisation*. Minneapolis: University of Minnesota Press.

Barnett Pearce, W. (1989). *Communication and the human condition*. Carbondale: Southern Illinois University Press.

Bourdieu, P. (1998). *La domination masculine*. Paris: Éditions du Seuil.

Cantril, H. (1940). *The invasion from Mars: A study in the psychology of panic*. Princeton: Princeton University Press.

Codeluppi, V. (2007). *La vetrinizzazione sociale. Il processo di spettacolarizzazione degli individui e della società*. Torino: Bollati Boringhieri.

Colombo, F. (2003). *La comunicazione sociale*. Bologna: Il Mulino.

Eco, U. (1979). *Lector in fabula: la cooperazione interpretativa nei testi narrativi*. Milano: Bompiani.

Lévy, P. (1997). *Collective intelligence: Mankind's emerging world in cyberspace*. New York: Plenum.

Lowery, S. A., & DeFleur, M. L. (1995). *Milestones in mass communication research: Media effects*. New York: Longman.

McLuhan, M. (1992). *The Gutenberg galaxy: The making of typographic man*. Toronto: University of Toronto Press. (Original work published 1962)

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. New York: Farrar, Straus and Giroux.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Silverstone, R. (1999). *Why study the media?* London: Sage.

Thompson, J. B. (1995). *The media and modernity: A social theory of the media*. Cambridge: Polity Press.

Volli, U. (2000). *Manuale di semiotica*. Roma-Bari: Laterza.



Co-funded by
the European Union

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe.



Co-funded by
the European Union

2. Comprehensive Guide on the Psychology of Communication

Written by Andrei Cornea, from In Varietate Concordia (IVC), 2025

2.1 Introduction

The psychology of communication underpins how humans create, interpret and respond to messages across contexts. This textbook chapter explores the cognitive, social, emotional, and ethical dimensions of communication, offering practical exercises, real-life case studies, and comparative insights between Romania and other European Union countries, including Sweden, Spain, Italy, Germany, and France. By integrating theoretical frameworks with applied examples, the text emphasises how psychological principles shape perception and interpersonal dynamics in both personal and public spheres. Gender shapes communication, and social norms and stereotypes affect how messages are perceived, who is trusted, and how misinformation spreads. In this textbook, we will also examine how gender interacts with bias, algorithms, and coordinated messaging across Europe. The textbook mirrors a comprehensive training course and is organised into five submodules. Each submodule presents theoretical foundations, real-world examples, exercises, critical analysis and a conclusion to reinforce learning outcomes.

2.2 Cognitive Biases and Perception in Communication

Understanding Cognitive Biases

Cognitive biases are systematic deviations from rational judgment, which influence how people perceive, interpret, and respond to messages. These biases often stem from the brain's need to simplify complex information for rapid decision-making (Kahneman, 2011). Key biases include:

Confirmation bias: Preferring information that aligns with pre-existing beliefs while discounting contradictory evidence.

Information bias: Occurs when we perceive or share information unevenly. Gendered content, such as posts criticising women leaders, is often amplified or ignored based on stereotypes rather than accuracy.

Anchoring bias: Being overly influenced by initial information when forming judgments.

Availability heuristic: Overestimating the likelihood of events based on their vividness or recency (Tversky & Kahneman, 1974).



Why these matters in gender are essential is because gender affects communication in visible and subtle ways; women and minorities often face disproportionate attacks online. Bias happens when our brains take shortcuts to make quick decisions. These shortcuts can make us see things in a certain way without realising it, sometimes leading to unfair or inaccurate conclusions. Before exploring communication psychology, it is crucial to understand the concept of bias.

Bias means a systematic deviation from rational or balanced judgment. In simple terms, bias occurs when our brains take shortcuts, leading us to see things in a particular way, often without realising it. These shortcuts help us make fast decisions, but can lead to unfair, distorted or inaccurate conclusions. When it comes to gender bias, people are more likely to believe information that confirms stereotypes. For example, a social media post claiming that women are “too emotional to lead” may be accepted uncritically because it aligns with societal stereotypes. Confirmation bias is when a person might trust a news story that supports their opinion and ignore others that do not. Biases operate in the brain’s decision-making regions, such as the prefrontal cortex, where decision-making and logic occur, and the amygdala, which houses emotional reactions, such as fear, anger, and pride. Emotional responses can override logical reasoning, making us more likely to believe and share messages that confirm our pre-existing beliefs. For example, posts attacking women politicians often exploit emotional reactions, reinforcing gender stereotypes. A stereotype threat occurs when women internalise societal expectations, doubting their abilities in STEM fields, for example, which affects how they respond to messages or share information.

Heuristics and Mental Shortcuts

Heuristics are mental shortcuts that facilitate quick decisions but can produce systematic errors. Common examples include:

Representativeness heuristic: Judging probability based on similarity rather than statistical logic.

Affect heuristic: Relying on emotional reactions rather than rational analysis.

Social proof: Conforming to perceived majority opinions, even when they are inaccurate (Cialdini, 2001).

In practice, heuristics explain why emotionally charged or socially validated messages spread rapidly, independent of factual accuracy.

Source Credibility, Trust and Memory Effects

Perceived credibility, determined by expertise, trustworthiness, and likability, strongly affects message acceptance. Memory effects, such as the illusory truth effect, show that repeated exposure increases perceived accuracy, even for false information (Fazio, 2020). Effective communication strategies must consider these cognitive tendencies to avoid inadvertently promoting misinformation. Prevention strategies beyond awareness include pausing before sharing something online. Check sources before forwarding a message. Diversify your



Co-funded by
the European Union

information, follow media outlets with different perspectives. Identify emotional triggers and recognise when anger, fear, or pride is influencing your judgment. It is also good to link such strategies to gender and see if stereotypes about women or minorities are influencing your reaction somehow.

Real Life Examples: Romania vs Sweden

Romania: During the 2020 parliamentary elections, social media analysis revealed the rapid viral spread of emotionally charged political memes. Many posts exploited confirmation bias and the availability heuristic, emphasising corruption scandals and nationalist narratives. Fact-checking efforts struggled because repeated exposure had already cemented false beliefs in public memory (Popescu, 2021). Also, in Romania, online posts repeatedly framed women politicians as “too emotional,” exploiting stereotypes and confirmation biases.

Sweden: Swedish campaigns emphasise transparency, fact-based messaging, and source credibility. Citizens demonstrate lower susceptibility to viral misinformation due to higher institutional trust and education that teaches the recognition of cognitive biases (Nygren & Svensson, 2020). Women politicians receive coverage focused on competence rather than gendered appearance. Swedish citizens show lower susceptibility to viral misinformation due to institutional trust and education on cognitive biases.

Exercises and Tasks

Reflection Question: Collect social media posts about male vs. female politicians. Identify which biases (confirmation bias, stereotype threat) influenced belief or sharing. Discuss patterns.

Group Discussion: Compare Romania and Sweden. How do societal trust and media literacy influence susceptibility to gendered disinformation?

Applied Task: Conduct an experiment testing how credibility cues (source expertise, gender) influence message acceptance among peers.

Critical Analysis / Discussion

Cognitive biases are unavoidable but manageable. Gendered disinformation exploits emotional triggers, stereotype reinforcement, and repeated exposure. Comparing Romania and Sweden highlights:

- How societal trust and education modulate susceptibility
- How algorithmically amplified gendered posts exploit common cognitive shortcuts
- The importance of integrating gender awareness into media literacy programs.

Conclusion

Cognitive biases and heuristics fundamentally shape perception, attention, and memory. Awareness of these mechanisms enables communicators to design ethical messages and allows audiences to evaluate information critically. Gendered disinformation requires specific attention to stereotype threat, confirmation bias, and algorithmic amplification.



Co-funded by
the European Union

Comparative insights between Romania and Sweden illustrate the importance of societal context in shaping communication outcomes.

2.3 Networks and Algorithms

Introduction and Overview

Digital platforms do not simply host communication; they actively shape how information spreads, which messages reach audiences, and who becomes most visible. Understanding these dynamics is central to the psychology of communication because they reveal the interplay between human cognitive biases, emotional engagement, and systemic design. Platforms are addictive because of dopamine rewards, social validation and ease of processing. They are designed to maximise engagement through features such as autoplay, infinite scroll, and likes. These features exploit cognitive shortcuts and emotional responses, increasing the likelihood that content will spread. Gendered posts often trigger stronger reactions, which, in turn, amplify algorithmic amplification.

This submodule examines algorithmic amplification, cross-platform behaviour, and algorithmic biases, with concrete examples from Spain. Readers will explore how platform features exploit psychological tendencies such as social proof, cognitive fluency and emotional salience. Exercises, reflection questions, and a critical discussion encourage practical understanding.

Attention Traps

Attention traps are platform features designed to capture and hold users' focus, often leveraging emotional salience: (Brady et al., 2017; Berger & Milkman, 2012)

Autoplay videos and infinite scroll: Keep users engaged passively.

Trending panels and notifications: Highlight content likely to generate reactions.

One-click sharing: Facilitates rapid, low-effort information propagation.

Case Example – Spain: In municipal elections, WhatsApp and Facebook posts targeted women with content suggesting that local policy changes endangered family care roles. Emotional messaging, such as narratives of “overburdened mothers,” was amplified via one-click sharing and group forwarding, exploiting cognitive shortcuts like social proof and affect heuristics.

Gender implication: Women-centric narratives often trigger empathy or outrage, increasing algorithmic amplification. Men-focused posts may leverage authority, responsibility, or economic fear.

Algorithm Boost & Bias

Algorithms reward engagement signals such as likes, shares, and comments. Early engagement can trigger platform ranking systems, thereby increasing visibility and producing cascade effects (Vosoughi, Roy, & Aral, 2018). Algorithmic bias happens when digital



Co-funded by
the European Union

platforms show unfair results because their systems learn from biased human data. Platforms prioritise posts that generate strong emotional responses. Content attacking women leaders or portraying minority women negatively spreads faster because it triggers anger or fear. Algorithms unintentionally reinforce gender stereotypes.

When linked to gendered disinformation, algorithmic bias can:

- Amplify sexist or emotional posts, such as attacks on women politicians or influencers.
- Hide credible voices, since educational or neutral content gets less engagement.
- Target people differently by gender, for example, showing fearful health content mainly to women.

Case Example – Spain: During the 2021 COVID-19 pandemic, a series of posts depicting mothers refusing vaccines as “endangering their families” went viral, reaching thousands more than posts from male doctors delivering factual guidance. Emotional storytelling and caregiving narratives engaged audiences and were preferentially boosted.

The psychological mechanism exploited is social proof, as repeated gendered stories create a perceived consensus. Confirmation bias occurs when users accept narratives that align with traditional gender roles.

Digital Pathways

Messages rarely remain on a single platform. Content may originate on fringe forums, gain traction on Twitter, and then migrate to Instagram and TikTok for broader exposure. This cross-platform movement ensures that narratives persist and reinforce each other.

Case Example – Spain: Climate-sceptic narratives targeting women as “family protectors” began on niche blogs, migrated to Twitter via hashtags, and then appeared as Instagram memes and short videos. Each stage reinforced gendered stereotypes, making correction difficult.

Algorithmic Effects on Politics

Algorithmic systems are not neutral; they prioritise content likely to generate engagement. Political and emotionally charged content often receives disproportionate amplification, while subtle or nuanced messages remain less visible (Huszár et al., 2022).

Case Example – Spain: During local elections, posts emphasising women’s caregiving responsibilities and men’s economic roles received higher engagement and algorithmic boosting. Repetition across Facebook, Instagram, and Telegram normalised these narratives, shaping public perception even when factual corrections were issued.

The psychological explanation is that emotional salience activates the amygdala, repetition strengthens memory (hippocampus), and social proof encourages conformity, making gendered misinformation particularly resistant to correction.

Critical Analysis / Discussion



Co-funded by
the European Union

Digital platforms amplify existing cognitive vulnerabilities. Gendered disinformation benefits from:

- Emotional triggers that increase shares
- An algorithmic ranking that disproportionately favours engaging content
- Cross-platform migration that reinforces narratives.
- Awareness of these dynamics enables audiences to assess content critically. It empowers policymakers to design interventions targeting algorithmic bias, particularly in gendered narratives.

Conclusion

Algorithmic design interacts with human psychology to shape message spread and influence. Gendered content, particularly narratives targeting women or reinforcing stereotypes, is amplified through emotional triggers, social proof, and cross-platform movement. Understanding these mechanisms allows communicators, educators, and policymakers to mitigate harm while promoting ethical information consumption.

2.4 Coordination and Influence

Overview

Coordinated communication campaigns use timing, repetition, and consistent framing to shape perceptions, influence behaviour, and reinforce social norms. Gendered disinformation leverages these strategies to target women, men, or specific minority groups, exploiting stereotypes and societal expectations. This submodule explores how coordinated messaging spreads, why it is persuasive, and how audiences can critically evaluate it. Examples focus on Romania, Spain, and Italy, illustrating how gender-specific narratives are amplified and resisted.

Signals of Coordination

Coordination is often subtle, but psychological cues can reveal orchestrated activity. Coordinated campaigns frequently target women with content reinforcing traditional roles (“women should care for family, not politics”). Posts are repeated across multiple platforms to strengthen stereotypes and normalise biased perceptions. Some key patterns include:

- **Synchronised timing** – Messages posted at similar times to maximise attention.
- **Repetition and templates:** Similar phrasing or slogans that reinforce ideas and create familiarity.
- **Role differentiation** – Influencers, amplifiers, and “seed” accounts distributing messages in stages.
- **Cross-channel activity:** The same message appears across multiple platforms to increase reach.



Co-funded by
the European Union

Gendered examples are that women are targeted with messages emphasising emotionality, caregiving or vulnerability. On the other hand, men may receive messages emphasising authority, financial responsibility or protection of family/ nation. Minority women may face intersectional targeting, combining gender and ethnic stereotypes.

Example: Romania and Italy - In Romania, social media campaigns repeatedly highlighted women politicians' appearance and emotional expressions, undermining credibility. Hashtags, memes, and posts from community groups reinforced these stereotypes.

In Italy, immigrant women were portrayed as threats to family stability, with repeated hashtags, viral memes, and emotionally charged stories circulated across multiple platforms.

Patterns in Messaging

Emotion-based triggers, such as anger, fear, or pride, make messages more memorable and shareable.

Framing: Simplified "us vs. them" narratives activate cognitive shortcuts and biases.

Priming: Exposure to repeated themes primes audiences to interpret new information consistently.

Example: Romania and Spain: In Spain, social media posts depicted women as overburdened mothers facing unsafe family conditions, suggesting they were blamed for societal problems. In Romania, similar narratives spread in diaspora communities, emphasising women's responsibility for emotional labour and family cohesion. Both cases show how repetition, emotional framing and gendered stereotypes influence belief and sharing behaviour, even without factual support.

In these cases, the psychological mechanisms exploited were confirmation bias, as audiences accept information confirming gendered stereotypes, and social proof, as repeated across groups to create a perceived consensus.

Exercises

Identify Coordinated Messaging: Track posts targeting women in two countries: note timing, repetition, cross-platform presence, and emotional framing.

Psychological Analysis: Analyse how cognitive biases (confirmation bias, stereotype threat, availability heuristic) are exploited in gendered campaigns.

Counter-Messaging Design: Create a sample social media post to counter a gendered misinformation narrative. Apply principles like repetition, empathy, and credible sources.

Analysis

Coordination in communication is often invisible but psychologically powerful. Recognising the interplay between timing, repetition, and framing allows citizens to resist manipulation. Ethically, analysts must balance identifying patterns with avoiding false claims about intent. In Romania, Spain, and Italy, coordinated campaigns show how messages can shift local narratives about women, influencing political, social, and public policy attitudes.



Conclusion

Coordination strategies exploit psychological principles, such as repetition, emotional triggers, and social cues, to shape public opinion. Gendered narratives are particularly susceptible to amplification due to:

- Societal stereotypes about women and men
- Cross-platform repetition reinforces norms
- Emotionally charged messaging that engages intuitive decision-making

By identifying timing patterns, repeated messages, and role differentiation, audiences can resist gendered disinformation and strengthen media literacy. Cross-country comparisons reveal that while coordination strategies are universal, cultural and societal context shapes their effectiveness.

2.5 Ethical Debunking

Overview

Debunking misinformation isn't just about setting the record straight; it's also about understanding people. When it comes to gendered disinformation, the challenge becomes both psychological and ethical. False stories often target women, minorities or people at the intersections of different identities. They play on our emotions, biases and social norms. Correcting them effectively means more than dropping facts; it requires timing, empathy and careful framing. This section looks at how ethical debunking works in practice, using real examples from France, Spain, and Germany.

Why Ethics Matter

Correcting false or misleading information carries both risks and responsibilities:

- **Amplification:** Highlighting false gendered claims may unintentionally increase visibility
- **Backfire effect:** Corrections can entrench pre-existing gender biases.
- **Legitimisation:** Addressing fringe claims may confer legitimacy.

Ethical debunking requires attention to values, emotions, and social context, avoiding harm while promoting accurate information. Debunking Strategies:

Prebunking/ Inoculation - Introduce audiences to common gendered misinformation tactics before they encounter them. Builds cognitive resilience and reduces susceptibility.

Example: In France, female journalists produced storytelling videos that exposed myths about women leaders' "emotional instability." Videos highlighted the psychological tricks behind gendered stereotypes, enabling viewers to recognise manipulative narratives before encountering them online. The psychological mechanisms addressed were confirmation bias, stereotype threat, and social proof.

Narrative Redirection - Promotes accurate information without unnecessarily repeating false claims. Uses stories, analogies, and emotionally resonant content.



Co-funded by
the European Union

Example: Spain - During misinformation campaigns about women's roles in environmental activism, NGOs highlighted successful female climate leaders and community projects. Messaging focused on competence, agency, and empowerment, rather than repeating false claims about women being irresponsible or uninformed, using calm, relatable storytelling instead of confrontation. Empathy increased trust. The psychological mechanisms addressed were emotional framing and the illusory truth effect.

Direct Correction - Clear, evidence-based responses to false claims. Must balance visibility with risk of amplifying misinformation.

Example: Germany - Fact-checking organisations countered viral posts claiming female candidates were “too emotional to lead.” Corrections included:

- Visual comparisons of male vs. female candidates' speeches
- References to verified public statements and performance data
- Concise, empathetic explanations emphasising competence.

The psychological mechanisms addressed were confirmation bias and, in this last case, empathy and trust, because messaging delivered by female professionals increased credibility.

Psychological Foundations - Effective debunking of gendered disinformation requires awareness of cognitive tendencies:

- **Continued Influence Effect:** False gendered claims persist even after correction.
- **Illusory Truth Effect:** Repetition strengthens perceived truth.
- **Confirmation Bias:** Users favour information that confirms existing gender stereotypes.
- **Stereotype Threat:** Women may doubt their abilities due to societal expectations.
- **Emotional Engagement:** Fear, shame, or pride can override factual reasoning.

Debunking strategies must address these mechanisms through timing, framing, repeated reinforcement, and sensitivity to identity.

Real World Case Studies

Election Disinformation (France, 2017): Viral posts suggested female candidates were “too emotional” to lead. Debunking involved female journalists and politicians sharing personal experiences, emphasising competence and experience. Results showed increased public confidence and reduced stereotype influence.

Climate Change Misinformation (Spain, 2021). False narratives framed women activists as “overly emotional” or uninformed. NGOs countered with storytelling highlighting successful female-led projects, reframing the narrative around expertise and empowerment.

Public Health Misinformation (Germany, 2020) Female doctors addressed false claims portraying women as irresponsible caregivers during COVID-19 health campaigns. Empathetic messaging emphasised informed choices, caregiving competence, and trust in professional guidance.

Exercises



Co-funded by
the European Union

Ethical Reflection: Design a prebunking post targeting a specific gendered misinformation case. Explain which psychological biases it addresses.

Comparative Analysis: Compare a gendered misinformation case from France and Spain. Analyse cultural context, psychological triggers, and debunking effectiveness.

Debunking Role-Play: Students practice delivering corrections to peers, emphasising empathy, credibility, and stereotype reduction.

Analysis

Debunking gendered misinformation requires balancing:

- **Ethics:** Avoid harm, respect autonomy, prevent stigmatisation
- **Psychology:** Target cognitive biases, emotional reactions, and stereotype threats.
- **Practicality:** Ensure reach, visibility, and platform effectiveness.

Prebunking reduces susceptibility before exposure. Narrative redirection avoids reinforcing false stereotypes. Direct correction works when delivered with credibility, empathy, and psychological insight. Cross-country examples demonstrate that context-specific strategies, combined with psychological awareness, produce the most effective outcomes in countering gendered disinformation.

2.6 Conclusion

Key takeaways:

Debunking is a psychological and ethical practice, not just a matter of factual correction. Gendered disinformation requires tailored approaches addressing cognitive biases, emotional triggers, and stereotypes. Strategies should combine prebunking, narrative redirection, and selective direct correction. Effective interventions strengthen public trust, promote informed decision-making, and challenge harmful gendered narratives.

2.7 References

Cialdini, R. B. (2001). *Influence: Science and practice*. Allyn & Bacon.

ECDC. (2020). *Infodemic management: A practical guide*. European Centre for Disease Prevention and Control.

Fazio, L. K. (2020). Repetition increases perceived truth even for known falsehoods. *Current Opinion in Psychology*, 38, 1–6.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.



Co-funded by
the European Union

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Psychological Science in the Public Interest*, 18(1), 1–53.

Minici, M., et al. (2024). Cross-platform traces of information operations: Dataset overview.

Nygren, T., & Svensson, L. (2020). Trust in Swedish institutions and susceptibility to misinformation. *Nordic Journal of Communication*, 10(2), 45–62.

Nyhan, B., & Reifler, J. (2010). When corrections fail. *Political Behaviour*, 32(2), 303–330.

Pacheco, D., Hui, P.-M., Torres-Lugo, C., Truong, B. T., Flammini, A., & Menczer, F. (2020). Uncovering coordinated networks on social media. Preprint.

Popescu, G. (2021). Misinformation dynamics in Romania: Cognitive and cultural dimensions. *Eastern European Journal of Communication*, 8(1), 23–39.

Starbird, K., Arif, A., & Wilson, T. (2019). Disinformation as collaborative work. *Proceedings of CSCW*, 127–145.

Swire-Thompson, B., & Lazer, D. (2020). Public health and online misinformation. *Annual Review of Public Health*, 41, 433–451.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.

Zannettou, S., et al. (2019). Cross-platform meme and narrative flows. *Graphika/ Academic Reporting*.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralised content in social networks. *PNAS*, 114(28), 7313–7318.

Huszár, F., Ktena, S. I., O’Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *PNAS*, 119(1).

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9(1), 4787.



Co-funded by
the European Union

3. New Media and Social Platforms

Written by Ali Honaramiz and Monika Kmet'ová, *Europsky Dialog* 2025

3.1 Introduction: The Pocket Power of TikTok

Imagine starting your day with a device that fits in your hand but holds the whole world's creativity, conversations, and conflicts within its glowing screen. This is the reality for millions of young people across Europe and beyond who use TikTok not just for entertainment, but as a primary window into culture, identity, news, and social connection. Unlike the TV shows or newspapers that shaped previous generations, TikTok is a fast-flowing river of short videos - music, humour, drama, education, activism - all mixed by algorithms designed to surprise, delight, and sometimes unsettle its audiences within seconds (Pew Research Centre, 2025). For youth like those in Slovakia, TikTok has revolutionised how they experience community, voice, and agency in a digital age.

TikTok's platform is built on participation and remix culture. What once was the domain of celebrities is now daily life for users who can film, edit, and share content from their bedrooms. A dance created by a 14-year-old in Atlanta can become a global phenomenon, with millions copying the steps and creating their own versions. For many youth, TikTok is not only a place to witness culture but to invent it, making the previously passive act of watching into an active, creative experience (World Economic Forum, 2025).

The Cultural Explosion: A Musical, A Challenge, and a Movement

One of the most extraordinary cultural phenomena birthed by TikTok is the "Ratatouille: The Musical" project. What started as a simple, humorous video by a young New York teacher in 2020 turned into an international sensation. Emily Jacobsen's playful homage to a beloved Pixar character became a viral creative collaboration that brought thousands of strangers together to write songs, choreograph dances, and build costumes - all within the TikTok ecosystem. Eventually, the project blossomed into a professional Broadway-style live-stream event raising over two million dollars for charity, attended by fans and celebrities worldwide. This story shows TikTok's capacity to turn isolated moments of joy into collective artistic experiences that transcend geography and traditional entertainment gatekeepers (Pew Research Centre, 2023).

On the flip side, TikTok's viral momentum can sometimes lead to more troubling outcomes. The "Devious Licks" trend in 2021 is a prime example. Directed and amplified by young users, this phenomenon saw thousands of students post videos of themselves vandalising school property, from stealing soap dispensers to damaging washrooms. While many participants framed this as rebellious fun, the consequences were serious: schools faced damage costs, disruptions, and expulsions. TikTok responded by banning related hashtags and increasing moderation. Still, the trend left a mark demonstrating how digital virality can spill into real-world harm within days (Ingram, 2021).



Co-funded by
the European Union

Finding Belonging and Identity in the Digital Crowd

For many teenagers, TikTok offers more than laughs and distractions; it provides a vital sense of belonging. Unlike more superficial platforms, TikTok's algorithm quickly groups users into "interest tribes", whether that's readers recreating literary debates, young activists sharing climate change tips, or neurodivergent youth finding community in explaining autism. One 17-year-old from Trnava explained that TikTok was the only place where she saw people "like me" who understood the complexities of being both Slovak and part of global online culture, describing it as a "cool, supportive older cousin who always knows what's going on" (Palmer, 2025).

This inclusiveness extends beyond Slovakia, reflecting the global youth's experience of intersectional identities and creative expression. Viral dance challenges like the "Renegade," created in her bedroom by 14-year-old Jalaiah Harmon, highlight how youth voices and creativity shape popular culture today. These challenges build social capital, foster peer connection, and sometimes launch creative careers, demonstrating a tangible pathway from online participation to offline opportunities (BBC Bitesize, 2023).

Dismantling the Myths: Anxiety Behind the Smile

Despite TikTok's contagious energy, it also fosters pressures that many youth struggle to articulate. The constant need to produce content, chase visibility, and maintain a curated persona online generates what mental health experts call "performance anxiety." The "always on" culture of TikTok has led 39% of surveyed teens to report feelings of overwhelm. At the same time, many young creators hide insecurities behind smiles perfected for the camera. It's become common to hear teenagers speak of deleting apps for mental health breaks, only to return, drawn back by FOMO - the fear of missing out on social connection or viral moments (Pew Research Centre, 2025).

Moreover, TikTok's algorithm sometimes elevates risky content. A dark example includes the "Benadryl Challenge," where young users took dangerous doses of allergy medication to capture hallucination videos, causing hospitalisations and raising urgent questions about platform responsibility. Despite increased warnings and content removal efforts, the challenge highlighted the fine line TikTok navigates between viral excitement and user safety (WHO, 2024).

More than Entertainment: TikTok as a News Source

One of the most surprising phenomena of the TikTok era is its role in shaping youth perceptions of current events. More and more teenagers regard TikTok as their primary news source, preferring "user-generated" live accounts to traditional media broadcasts. During protests, climate disasters, or political upheavals, on-the-ground videos and commentary often appear first on TikTok feeds. Young users appreciate the raw authenticity and immediacy, but sometimes struggle to distinguish verified facts from speculation or misinformation (Palmer, 2025).



Co-funded by
the European Union

This shift challenges the entire news ecosystem. Governments and NGOs now work to collaborate with "influencers" and content creators to disseminate accurate, timely information. Yet the risk of misinformation spreading rapidly on TikTok remains a critical challenge, demanding enhanced digital literacy education and fact-checking innovations. TikTok by the Numbers (2024 Snapshot)

To give a clearer picture, here are some illuminating facts about TikTok's reach and influence today:

TikTok boasts over 1.8 billion users worldwide, with approximately 30% aged under 18 (Palmer, 2025).

More than half of the viral songs on Europe's music charts got their boost from TikTok dance challenges or memes (CNN, 2023).

In surveys, 72% of Gen Z respondents identify TikTok as a primary source of news and cultural trends (Pew Research Centre, 2025).

Educational TikTok channels have increased exam scores in pilot studies, illustrating the efficacy of "microlearning" through social video (Pew Research Centre, 2025).

Voices of the Teens

To conclude this section, it's illuminating to listen to what the youth themselves say about TikTok:

Martin, 16, Bratislava: "I never watched TV news - too slow and boring. Now I know about protests or floods minutes after they start, because someone's there, filming and sharing, no filters."

Kamila, 19, London: "We used TikTok to raise awareness about Sudan. My little video reached kids not just here but in Berlin and Toronto, and connected us like never before. I even spoke at a youth conference!"

Jakub, 17, Poprad: "TikTok helps me find people like me, but it's stressful too. The drama, bullying - it's everywhere. I had to take a break and talk with my parents. Still, I don't want to leave; the good things are amazing."

Reflection and Activities

Recall a TikTok trend or challenge. How did it affect your community or school? Was it positive, negative, or both?

Track a day's worth of news or cultural updates. How many did you get first from TikTok? Did you verify their accuracy?

Try creating a TikTok video about something you care about - science, art, or local news. Observe how your community responds.

3.2 Beyond TikTok - Reddit, Discord, and Snapchat in Youth Culture

Reddit: The Internet's Town Square for Seeking Truth and Community

Imagine entering a vast virtual library and community centre rolled into one, with thousands of themed rooms where people from all over the globe discuss everything from



Co-funded by
the European Union

gardening tips to the latest Marvel movie theories, from mental health support to conspiracy theories. This is Reddit, often described as the "front page of the internet," an expansive platform of forums called "subreddits," each dedicated to a specific topic, idea, or interest.

For young people, Reddit offers a refuge from TikTok's fast-paced, visual bombardment. It is a place to dive deeply into niche interests, seek advice anonymously, and participate in intense discussions. Slovak teens and young adults frequently visit subreddits related to language learning, video games, academic help, and even personal stories. A well-known subreddit is r/Teenagers, where young people candidly share their struggles with school, friendships, mental health, and relationships. The anonymity Reddit allows can be both a blessing and a curse: it creates a space for candid sharing without fear of judgment, yet it can also shield trolls or misinformation (Hasan, 2024).

One compelling example comes from a Slovak teenager who created a post in r/Slovakia about dealing with anxiety during exam season. The post garnered hundreds of empathetic responses and tips, connecting a geographically scattered community in a shared experience. This kind of peer support contrasts with the more public, performative spaces of Instagram or TikTok, highlighting Reddit's role in fostering intimacy and detailed conversation (Hasan, 2024).

However, the platform has also been a breeding ground for polarising theories. During the COVID-19 pandemic, several subreddits became hubs for anti-vaccine sentiment and conspiracy theories, spreading fears that sometimes leaked into offline protests. This reality drives ongoing debates about moderation, community responsibility, and freedom of speech - issues that young Redditors often navigate critically, aware of the information's power and risks.

3.3 Discord: From Gaming Chat to Virtual Hangouts and Activism

Discord began in 2015 as a communication tool for gamers, combining voice, video, and text chat into a single app. Over time, it morphed into a robust social hub, hosting millions of communities not just for play but for friendships, activism, education, and fandoms. For many young people growing up in post-pandemic isolation, Discord became their central "third place" - neither home nor school, but a social space where they could be themselves.

What makes Discord unique is its server-based architecture. Unlike the public streaming and sharing model of TikTok or Instagram, Discord servers are invitation-only or private communities shielded from general browsers. These servers can be small and intimate or large with thousands of members. They offer structured discussions, real-time project collaboration, or just casual chat.



Co-funded by
the European Union

A vivid example from Slovakia is the emergence of local Discord servers dedicated to environmental activism. Youth groups used these servers to organise clean-up days, discuss climate policy proposals, and coordinate digital campaigns such as the hashtag "EcoTok." In one case, a server called "GreenTalk Bratislava" served as a launchpad for a regional youth sustainability summit in 2024, attended by local politicians and NGOs (Palmer, 2025).

Discord's emphasis on voice communications and less algorithmic visibility also means that young people often feel more in control and less distracted by endless scrolling. Yet, challenges exist. Some servers became hotspots for inappropriate content or recruitment into harmful groups, prompting calls for improved moderation and education on safe server use (SedonaSky, 2023).

Snapchat: The Ephemeral Window into Youth Culture and Social Drama

Snapchat, launched in 2011, revolutionised social media by making content ephemeral. Instead of posts visible forever, Snapchat introduced messages, photos, and videos that disappear seconds after being viewed. This fleeting nature made Snapchat a favourite among younger teens, creating a space for more intimate and authentic communication free from permanent digital footprints.

For Slovak and European youth, Snapchat remains a popular private channel for maintaining close friend groups, sharing moments spontaneously, and escaping the pressure of exposure on more public platforms. The "Stories" feature, which aggregates posts visible for 24 hours, also bridges private and semi-public sharing, blending storytelling, humour, and real-time event reporting.

However, with ephemerality came unique risks. The platform's design made it attractive not just for harmless fun but also for risky behaviours like sexting or sharing sensitive images, sometimes without a complete understanding of consent or repercussions. Alarming cases surfaced where criminal networks used Snapchat to lure or exploit teenagers, hidden beneath the guise of disappearing messages (BBC, 2023; SedonaSky, 2023).

A tragic Slovak case that made headlines in 2025 illustrated these dangers: a teenager's Snapchat messages were used as evidence in a criminal investigation into cyber exploitation, sparking national conversations on digital safety education and parental involvement. This underscored the need for balanced digital literacy - leveraging Snapchat's creative potential while educating young users about privacy, consent, and reporting mechanisms.

Yet Snapchat also innovates with positive features: its augmented reality (AR) filters encourage creative self-expression; its "Snap Map" connects friends in real time, helping teenagers feel less isolated; and its "Spotlight" feature brings TikTok-style virality into the Snapchat environment. The app remains a vibrant part of youth culture, especially for spontaneous and private social interaction.



Co-funded by
the European Union

Summary and Interactive Reflection

Reddit offers deep-dive anonymity and community, but requires critical thinking to navigate misinformation.

Discord builds intimate communities for friendship and activism, but it also highlights the importance of moderation.

Snapchat provides ephemeral, private communication essential for youth connection, but presents unique safety and privacy challenges.

Activities for Exploration:

Visit a chosen subreddit related to a hobby or social issue and report back on the type of content and discussions you find. How is it different from TikTok or Instagram?

Create a mini Discord server with friends or classmates to organise a school event or project. Discuss what made communication easier or more complicated in this space.

Reflect on the advantages and risks of ephemeral messaging like Snapchat. What rules would you create for yourself and friends to stay safe and respectful?

3.4 The Other Stage - Twitch, Clubhouse, and Telegram in Youth Social Lives

Twitch: The Live-Streaming Stage Where Fans and Creators Connect in Real-Time

If TikTok is the fast scroll, Twitch is the live theatre. Launched in 2011, Twitch began as a gaming-focused platform where users could broadcast their gameplay in real-time and interact with viewers via chat. What started as a niche corner for gamers quickly morphed into a complex ecosystem that youth worldwide now use for entertainment, education, creative expression, and community building.

The magic of Twitch is in its immediacy and interactivity. Unlike prerecorded content, Twitch streams function like instant social events - viewers watch and chat live, influencing streamers' actions, sharing jokes, and collaborating through loyalty points and rewards. For many young people, Twitch is less about passive consumption and more about forming a sense of belonging in the moment through shared rituals, memes, and crowd-driven entertainment.

In Slovakia, as elsewhere, Twitch helped fuel a wave of local content creators who built followings by mixing game streams with chat commentary, music sessions, or even virtual study halls. One popular channel, "StudyWithMeSK," hosts online study groups where teens and young adults work together silently or whisper, helping recreate a library feeling, distant schooling stripped away. The streamer's chat becomes a space of accountability, encouragement, and shared focus, demonstrating Twitch's potential beyond gaming (Palmer, 2025).

However, Twitch also faces challenges. The public, live format makes moderation complex: harmful language can spread rapidly, and young users have reported exposure to inappropriate content or harassment. Twitch has invested in AI moderation tools and



Co-funded by
the European Union

community activists to prevent abuse, while local groups in Slovakia promote "safe streaming" workshops educating youth on respectful conduct (SedonaSky, 2023).

Clubhouse: The Audio-Only Wave of Conversation and Connection

Clubhouse burst onto the scene in 2020 as an audio-only social app where users drop into virtual "rooms" to listen or speak on a wide variety of topics - from podcasts and book clubs to political debates and behind-the-scenes celebrity interviews. For youth who grew up navigating visual and textual overload, Clubhouse offers a refreshing, intimate experience focused entirely on voice and presence, without the distractions of images or video.

Though Clubhouse has seen slower growth in Slovakia than in Western markets, it has nonetheless found enthusiastic pockets of young users who appreciate the platform's egalitarian, spontaneous debate culture. College groups use it to organise study sessions; activist circles discuss ongoing protests; multilingual conversations flourish, offering language learners a live platform to practice skills and meet peers (Palmer, 2025).

Clubhouse's real-time audio format encourages vulnerability and authenticity. Participants must listen actively and respond instantly, recreating the dynamics of a face-to-face chat. This has led to rich, complex discussions that build empathy and knowledge. However, some users also report feeling left out when conversations move too fast or skew towards insider groups.

The platform's relatively open-access nature has raised concerns about privacy and misinformation, prompting developers to introduce recording limits and transparent moderation policies. Nonetheless, its unique format has prompted educators and youth leaders to explore audio storytelling and debate formats for classroom and community use, expanding digital literacy beyond screens (Hasan, 2024).

Telegram: The Versatile Messenger Serving Private and Group Communication

Telegram occupies a different niche. Launched in 2013 and popular in Central Europe, it is a messaging app that combines privacy, encryption, and flexible communication options, including private messages, large groups, channels, and bots. For many young people in Slovakia and beyond, Telegram serves as a trusted hub for social coordination, sharing educational resources, and even activism networking.

Unlike mainstream platforms optimised for public broadcasting, Telegram thrives in creating semi-private and confidential communities. It is popular among students for exchanging notes, for groups to organise protests or events, and for hobbyist channels distributing digital art or fan fiction. The encrypted, secure environment appeals to youth concerned with privacy in an increasingly surveilled world (Palmer, 2025).

Telegram's channel feature allows influencers, NGOs, and youth activists to broadcast updates to thousands, facilitating grassroots mobilisation with a high degree of control over content and audience. During the 2024 environmental protests in Slovakia, Telegram



Co-funded by
the European Union

channels played a critical role in disseminating real-time information, consent protocols, and media guidelines, circumventing censorship pressures (Palmer, 2025).

Yet, this privacy also renders Telegram a potential haven for toxic disinformation and extremist groups, as authorities struggle to monitor private channels effectively. Educators emphasise critical digital literacy, encouraging youth to verify source credibility and recognise the limits of encrypted spaces in grappling with harmful content (Hasan, 2024).

Reflection and Activities

Join a Twitch stream or watch clips from "StudyWithMe" channels. How does live interaction change the feeling compared to TikTok or Instagram videos?

Explore a Clubhouse conversation on a topic you care about. Notice how voice-only communication affects your listening and participation.

Find a public Telegram channel related to your interests or activism. Reflect on the differences in privacy, trust, and content control between this and more public platforms.

3.5 How Platforms Shape Misinformation and Misogyny

The Gendered Landscape of Online Disinformation

Social media platforms have revolutionised communication, activism, and community-building. Yet, alongside these positives lies a darker reality - gendered disinformation. This refers to false or manipulative content targeting individuals or groups based on their gender, often designed to undermine, silence, or discredit women, LGBTQIA+ persons, and activists. Unlike generic misinformation, gendered disinformation weaponises deep-rooted sexism, stereotypes, and cultural biases, exploiting digital spaces to inflict real harm (Gehrke, 2025).

Digital platforms do not merely host this content; their design and algorithms amplify it in ways that disproportionately affect women and marginalised persons. This amplification, sometimes labelled "algorithmic violence," is a structural form of harm embedded in how content is recommended, monetised, and moderated across networks (EPLU Committee, 2025). Understanding the gendered dynamics of misinformation requires a platform-by-platform exploration.

TikTok: Creativity Meets Polarisation

TikTok's algorithm prioritises virality and emotional engagement, enabling the rapid spread of both empowering and harmful content. Feminist creators and LGBTQIA+ advocates have found a vibrant stage here, where underrepresented narratives can go global in minutes. Campaigns like #MeToo and viral body positive movements gained momentum through TikTok's reach (Pew Research Centre, 2023).

However, the same platform is also fertile ground for gendered disinformation. Videos disparaging feminists or spreading "misandrist" conspiracy theories circulate alongside fitspiration content promoting unrealistic body standards, a phenomenon that can reinforce harmful gender stereotypes (Reddit Feminism community, 2024). Deepfake videos



Co-funded by
the European Union

and fake "exposed" videos targeting female politicians and activists have emerged, exploiting TikTok's remix culture (Roos, 2025).

Instagram: The Image of Perfection and the Weaponisation of Femininity

Instagram's visual-centric platform has been criticised for intensifying gendered pressures on bodies and appearance, particularly for girls and young women. Algorithms reward idealised, often unrealistic images, fostering environments where users aspire to narrow beauty standards (EU Parliament, 2023).

Gendered misinformation on Instagram often manifests as "toxic femininity" trends, where hostility between genders is amplified under the guise of feminist discourse. Public comment sections frequently devolve into gender wars, with misogyny and misandry fueling polarised online interactions (Reddit CMV, 2025). Bots and coordinated campaigns spread pseudo-scientific claims designed to discredit women in positions of influence, damaging public trust in female leaders and journalists (Roos, 2025).

Conversely, Instagram has been a powerful tool for positive campaigns, such as youth-led environmental activism and feminist art movements, highlighting a complex duality in its gendered content ecosystem.

Reddit: The Home of Extremes and Echo Chambers

Reddit houses thousands of communities with diverse perspectives, including feminist safe spaces and harshly misogynistic "manosphere" subreddits. This platform's anonymity encourages raw and unfiltered discourse, but also allows organised campaigns of harassment and the spreading of gendered disinformation.

Research has documented Reddit as a breeding ground for far-right misogyny, conspiracy theories, and anti-feminist rhetoric, often monetised and algorithmically amplified. Yet, it remains a critical space for feminist intersectional activism and discussions on gender justice, illustrating the broad spectrum of discourse possible on the same platform (Hasan, 2024; Reddit Feminism, 2024).

Discord and Telegram: Private Channels with Public Consequences

The privacy and encryption features of Discord and Telegram create safe havens for marginalised groups and youth activists. However, they also enable the covert spread of gendered disinformation and harassment.

Encrypted Telegram channels have been sites for sharing manipulated images, doxing of female politicians, and coordinated smear campaigns. These private, hard-to-monitor spaces pose significant challenges to content moderation and victim protection (Europarl, 2025). Similarly, Discord servers sometimes facilitate misogynistic communities or the distribution of gendered hate content despite efforts toward moderation (SedonaSky, 2023). The dual nature of privacy-driven platforms requires nuanced digital literacy interventions that emphasise safety, verification skills, and community accountability.

Platform Policies and the Influence of Tech Power



Co-funded by
the European Union

Social media platforms are not neutral spaces but systems shaped by corporate power, policy decisions, and profit-driven algorithms. Their designs determine which voices are amplified and which are silenced, influencing public discourse on gender and equality. The moderation approaches of major platforms - such as Meta, X (formerly Twitter), and TikTok - often prioritise engagement and free-speech narratives over user safety. When moderation is loosened, gendered disinformation and harassment rise, disproportionately targeting women and LGBTQIA+ users (Roos, 2025). These shifts reveal how platform owners' ideological choices directly affect the prevalence of online gender harms. Algorithms also shape visibility. Content that provokes emotional reactions tends to be amplified, allowing sensationalist, polarising, or sexist material to spread faster than balanced or educational posts. Meanwhile, feminist and equity-focused content can face suppression or lower reach (EPLU Committee, 2025). Such algorithmic biases reinforce existing social stereotypes, a phenomenon described as algorithmic patriarchy. Lastly, data-driven advertising further embeds gender inequality into digital systems. Targeted marketing frequently reproduces traditional gender roles - showing technical job ads to men and beauty products to women - illustrating how the commercial logic of platforms intersects with cultural bias.

Addressing gendered disinformation thus requires more than removing harmful content. It demands systemic accountability, transparent algorithms, fair moderation, and digital literacy that helps users critically understand how tech power influences social norms.

Reflective Questions

In what ways can "algorithmic violence" be understood as a continuation of offline gender biases?

How might your personal experiences on platforms like TikTok, Instagram, or Reddit reflect patterns described in this chapter?

Should social media companies be held accountable for algorithmic amplification of misogynistic content? Why or why not?

How do privacy-focused platforms like Discord and Telegram both protect and endanger marginalised groups?

What role does profit play in how platforms handle or ignore gendered disinformation?

3.6 Conclusion: Navigating the Intersection of Gender, Disinformation, and Digital Spaces

The platforms youth inhabit wield enormous influence in shaping digital culture and political discourse, carrying both promises and perils regarding gender equality. On one side, young creators and activists use TikTok, Instagram, Reddit, Discord, and Telegram to amplify marginalised voices, mobilise for rights, and reshape narratives. On the other hand, deeply embedded misogyny, algorithmic biases, and coordinated gendered disinformation campaigns continue to threaten public participation, mental health, and community safety.



Co-funded by
the European Union

Building resilient digital environments requires comprehensive approaches: critical media literacy grounded in gender awareness, transparent platform policies, inclusive content moderation, and a broader societal commitment to challenging gendered stereotypes and violence - both online and offline.

Youth, educators, policymakers, and platform developers must work in tandem to ensure new media fulfil their democratic potential without becoming accelerators of gendered harm.

3.7 References

European Parliament Committee (2025). Image-based sexual violence in the context of AI and social media.

Gehrke, M. (2025). Gendered disinformation as violence: A new analytical agenda. *Harvard Misinformation Review*.

Hasan, S. (2024). Social media and its impact on youth: A sociological study. *Society and Culture Development in India*, 4(2).

Ingram, N. (2021). *Thrivers: The Surprising Reason Why Some Kids Struggle, and Others Shine*. London: Penguin.

Palmer, J. (2025). How TikTok has become a key source of information for young Gen Z. *Berlin School of Business and Innovation*.

Pew Research Centre (2023). *Teens and social media: Key findings*.

Pew Research Centre (2025). *Teens, Social Media and Mental Health*.

Reddit Feminism community (2024). *Online discussions on gender and misinformation*.

Roos, C. (2025). *Gendered Disinformation as Infrastructure*. Tech Policy Press.

SedonaSky (2023). *Algorithmic violence: How social media amplifies gendered disinformation*.

WHO - Regional Office for Europe (2024). *Teens, screens and mental health*. WHO Europe, 25 September.

World Economic Forum (2025). *I went viral on TikTok. Here's what I learned*.



Co-funded by
the European Union

4. Media Literacy - Techniques for critical thinking in the digital age

Written by Sara Mesa and Beatriz Muñoz, Jovesolidés, 2025

4.1 Introduction to Media Literacy

Media literacy is a concept that has evolved significantly over the past decades. It initially focused on the need to critically understand messages from traditional media, newspapers, radio, and television, with a strong emphasis on reading advertising and entertainment content critically.

With the expansion of the Internet and social media, however, this term now has a much broader meaning. Today, it involves not only analysing messages but also accessing information, evaluating it, producing content, and participating actively and ethically in digital environments (Livingstone, 2004).

In today's context, marked by rapid digitalisation and the omnipresence of online information, media literacy becomes an essential skill. News no longer comes only from traditional outlets but also circulates widely on digital platforms, where verified content coexists with rumours, fake news, and hate speech. Information overload and the difficulty of distinguishing between facts, opinions, and manipulations make media literacy a basic requirement for exercising critical, conscious citizenship.

The European Union has identified media literacy as a fundamental strategic priority to strengthen active citizenship and social cohesion in the digital era. Both the Audiovisual Media Services Directive and recent digital education plans stress that Member States should implement national policies promoting media literacy from early education to adult learning, integrating it into curricula and continuous training programmes.

These initiatives aim not only to provide technical skills but also to develop critical, ethical, and social skills that allow individuals to evaluate information responsibly and participate consciously in the public sphere.

In this sense, media literacy is conceived as a 21st-century civic right: just as literacy was essential for democratic participation in the past, the ability to understand, analyse, and manage digital communication is crucial today. This competence allows citizens to identify reliable information against misinformation, recognise biases and media manipulation, and exercise critical judgment that strengthens democracy and social resilience.

It also promotes more inclusive and equitable citizenship, enabling people to interact with confidence and security in a complex digital environment where information flows rapidly in multiple formats.



Co-funded by
the European Union

In short, media literacy is a pillar for ensuring equal opportunities in the information society. It empowers citizens to navigate complex media ecosystems, recognise attempts at manipulation, and participate actively in public debate. More than an isolated skill, it is a transversal and enduring competence that supports democratic cohesion and strengthens social resilience against misinformation (Livingstone, 2004).

Next, we will move on to 4.2 Dimensions of Media Literacy, exploring the different dimensions that make up media literacy: access, analysis, creation, and participation. This comprehensive perspective will help readers understand media literacy not just as an individual skill, but as a set of interrelated abilities essential for navigating the information society critically and safely.

4.2 Dimensions of Media Literacy

Media literacy is not just about passively consuming information; it involves a set of interrelated skills that allow people to understand, evaluate, and participate actively in the communication environment.

These skills are organised into dimensions ranging from accessing information to creating content and participating ethically online. Understanding these dimensions is essential in a world where media and digital platforms play a central role in daily life, influencing public opinion, education, and decision-making.

Breaking media literacy into dimensions shows how each contributes to forming individuals capable of interacting with information critically and responsibly. It is not only about protecting against misinformation but also fostering active and ethical participation that reinforces democracy and social cohesion.

This chapter focuses on four key dimensions, access, analysis, creation, and participation, that form the core of competent and aware digital citizenship.

Access: finding and using information across different media

The first dimension, access, refers to the ability to locate, select, and use information effectively across different formats and platforms.

This skill is essential because digital information is dispersed across multiple channels, including traditional media and news portals, social media, blogs, podcasts, and video platforms. Accessing information properly requires knowing where to search and understanding which sources are relevant and reliable in context (Livingstone, 2004).

Access goes beyond the technical ability to use search engines or digital tools; it also requires strategy and critical judgment.



Co-funded by
the European Union

For instance, when researching a current issue, a media-literate person compares information across different outlets, checks publication dates, assesses the reputation of each source, and recognises possible bias or editorial agendas.

A practical example could be searching for information about vaccines online: someone with strong access skills will not rely on the first result on Google but will instead consult official institutions (such as the World Health Organisation or national health ministries), review scientific articles, and assess the reliability of what they encounter on social media.

In a diverse European context, where multiple languages and cultures coexist, the Access dimension also includes the ability to navigate multilingual and multi-format information, using translation tools, filters, and verification systems that help citizens access high-quality details efficiently. This skill is key to reducing exposure to misinformation and promoting an informed and engaged public.

Access, therefore, forms the foundation upon which all other media literacy dimensions are built. Without the ability to locate and select trustworthy information, critical analysis, responsible creation, and ethical participation become limited.

Developing effective access strategies is thus the first essential step toward becoming a competent and resilient digital citizen in a complex information landscape.

Analysis: evaluating messages, formats, and sources

The analysis dimension focuses on the ability to critically examine information, evaluating content, form, purpose, and reliability.

Analysis involves questioning messages, identifying intentions, and recognising elements that may influence perception, such as tone, images, headlines, and rhetorical devices.

This skill is essential for distinguishing between facts and opinions, detecting bias, and making informed decisions. For example, when reading a news article about an international conflict, a media-literate individual does not simply accept the presented information. Instead, they consider who published the article, its reputation, whether it cites credible sources, whether it includes multiple perspectives, and how its language may provoke specific emotions, such as fear, anger, or sympathy.

Through this process, analysis helps identify disinformation patterns, sensationalist headlines, or cultural bias that may distort understanding.

Analysis also applies to multimodal content, such as videos, memes, or infographics. A viral video on social media, for instance, may combine shocking visuals with dramatic music and suggestive text; a critical reading requires evaluating each element and its communicative intent.



Co-funded by
the European Union

Analytical skills not only protect against manipulation but also enhance deeper comprehension and promote a more mindful approach to media consumption.

Ultimately, analysis is the cornerstone of media literacy: it transforms access to information into critical knowledge. Without this ability, digital citizens risk falling prey to propaganda, disinformation, or one-sided narratives, weakening their capacity for informed and responsible participation in public life.

Creation: producing responsible and verified content

The creation dimension involves moving from passive consumption to conscious, ethical content production. Creating responsible content means cross-checking sources, verifying data, and respecting copyright before sharing. It also requires reflecting on social impact: how our posts affect others, what values they convey, and whether they contribute to a fair and inclusive digital environment.

As Frau-Meigs (2021) points out, media literacy in the algorithmic age requires not only technical skills but also a communication ethic grounded in truthfulness, empathy, and collective responsibility.

Producing verified content not only improves the quality of public discourse but also strengthens users' trust and combats disinformation through everyday practice.

In summary, media creation is not merely a technical competence but an act of digital citizenship: it means using one's voice critically, respectfully, and constructively in Europe's shared digital public space.

Table 1. A practical guide for responsible digital content creation

Stage	Key Questions	Good Practices	Practical Examples
VERIFY	Is the information I'm about to share reliable and up to date?	Cross-check with at least two verified sources (trusted media, official databases, EU institutions, gender-focused NGOs).	Before posting about health, consult the European Medicines Agency (EMA) data.



Co-funded by
the European Union

ASSESS IMPACT	Could this content negatively affect someone or reinforce gender stereotypes?	Avoid discriminatory language or gender clichés; check the emotional tone.	Make sure a meme does not reinforce stereotypes about women at work or motherhood.
CREATE WITH PURPOSE	What is the aim of my message: to inform, raise awareness, or educate?	Define a clear intention and use inclusive, accessible language.	When discussing media literacy campaigns, highlight the inclusion of women and minorities in education.
ATTRIBUTION AND COPYRIGHT	Am I respecting authorship and citing sources properly?	Use copyright-free materials (Creative Commons, institutional image banks) and always mention data sources.	When sharing statistics on digital gender violence, include the source and publication date.
ETHICAL SHARING	Am I ready to engage in dialogue about what I publish?	Encourage constructive conversation, reply respectfully, and correct mistakes if needed.	If someone points out an incorrect fact about women's participation in tech, update your post and explain the verified data.

Source: Author's elaboration

Participation: active citizenship and digital ethics

Finally, the participation dimension focuses on how young people can act as responsible citizens in the digital environment, not only by consuming and creating information, but also by engaging ethically and critically in online spaces. In today's Europe, where social media and digital platforms are central to cultural, educational, and political exchange, digital participation is a key tool for inclusion, democracy, and the promotion of social values such as equality, respect, and diversity.

Ethical participation means engaging in debates, collaborating in collective projects, and sharing verified information without reproducing stereotypes, hate speech, or misinformation. For example, a group of students could run a blog or a video channel promoting educational content on gender equality, media literacy, or sustainability, ensuring



Co-funded by
the European Union

that their data sources are reliable, their images reflect cultural and gender diversity, and their tone fosters inclusion and reflection.

Participation also includes civic tech initiatives and digital campaigns across Europe, such as youth forums, online consultations, and virtual volunteering platforms. These allow young people to voice their opinions on community and European issues, promoting shared responsibility and collective action. A concrete example is the European Youth Portal, where young Europeans can access information, participate in discussions, and collaborate on projects that make a difference in their communities.

Moreover, ethical digital citizenship involves understanding how algorithms and personalised content shape the information we see. Young people should be aware that platforms may prioritise certain opinions, amplify polarisation, or spread misinformation, and they can counter these effects by sharing responsibly, verifying information, and promoting peer education.

In essence, active digital participation not only strengthens civic competencies but also helps build a more critical, inclusive, and resilient European community that is better equipped to counter disinformation and discrimination. Promoting this dimension through media literacy empowers young people to become agents of change, capable of shaping their digital environments with awareness and integrity.

Practical examples:

- Launching social media campaigns on gender equality using verified data from Eurostat or UN Women.
- Participating in online educational debates, questioning stereotypes, and sharing trustworthy resources.
- Organising online workshops or webinars to teach peers how to identify fake news or viral hoaxes.

4.3 The Importance of Critical Thinking in the Digital Age

In the digital age, being informed does not always mean truly understanding what is happening. This is why critical thinking becomes an essential skill for anyone navigating social media, websites, blogs, or video platforms. Critical thinking can be defined as the ability to analyse, evaluate, and question information, separating facts from opinions, identifying potential biases, and recognising when a source is trustworthy. Its core principles include curiosity, reasonable doubt, logical coherence in analysis, and a willingness to revise one's own ideas in light of new evidence.

Often, people confuse simply being informed with genuinely understanding a topic. Being informed might mean just reading a headline, watching a viral video, or sharing an article without pausing to consider whether the information is accurate or complete. Understanding, however, requires digging deeper into the content, checking the reliability of



Co-funded by
the European Union

sources, connecting data to the broader context, and reflecting on the possible consequences of the information.

For example, reading a news article about the gender pay gap is not enough; critical thinking allows one to analyse the source, compare the data with official statistics, understand the factors behind the information, and recognise whether the message manipulates emotions or reinforces stereotypes.

Critical thinking also acts as an antidote to misinformation. Fake news, viral rumours, and misleading messages often appeal to strong emotions or present data out of context. A person with well-developed critical thinking skills does not react automatically to what they see or read; instead, they ask, "Who published this?" Why? What evidence supports it? Before sharing or accepting content as accurate, these questions help prevent the spread of misinformation.

For instance, before forwarding a meme about a social issue, a critical student will check whether the source is reliable, whether verifiable data exists, and whether the message could reinforce prejudice or misunderstandings.

In short, developing critical thinking enables individuals to become more conscious consumers and creators of information, capable of questioning, reflecting, and acting ethically in a digital environment saturated with data, opinions, and diverse content. It is not just about avoiding misinformation; it is about participating responsibly and making informed choices in the digital world.

Table 2. Critical Thinking Checklist for Young People: Gender and Media Literacy



Co-funded by
the European Union

Question	What to Look For / Do	Practical Example
Who is publishing this?	Identify the source: is it a reliable media outlet, a gender-focused NGO, a researcher, or an unknown person?	A statistic on digital gender-based violence published by UN Women is more reliable than a viral Facebook post with no author.
Is it information or opinion?	Distinguish verifiable facts from interpretations or sensational claims.	“25% of young women have received abusive messages on social media” is a fact; “social media only makes women’s lives worse” is an opinion.
Is there evidence?	Look for data, studies, or references supporting the information.	Before sharing a meme about the gender pay gap, check whether an official study or an EU report confirms it.
Is it recent and relevant?	Check the publication date and whether the data are still valid.	A 2010 article on women's access to technology may no longer reflect the current situation in Europe.
Could it reinforce stereotypes or prejudice?	Assess whether the content reproduces harmful gender roles or false ideas about minorities.	Avoid memes that mock women in STEM or exaggerate gender differences in abilities.
Is it worth sharing?	Ask yourself if it adds educational value or encourages critical thinking.	Before forwarding a sensationalist article about feminism, verify if it really provides useful and well-documented information.

Source: Author's elaboration based on the "Crítica" project by Jovesólides

4.4 Techniques and Strategies for Critical Thinking

As mentioned earlier, one of the foundations of critical thinking is asking key questions before consuming or sharing information. Knowing who published a message, what their intention is, and which source supports the information allows you to assess the reliability of any online content quickly. These questions not only help filter out false or biased content but also encourage reflection before acting, whether sharing, commenting, or making decisions based on the information.



Co-funded by
the European Union

For example, when reading a post about the gender pay gap, a critical reader does not simply accept the figure. They ask: Who is publishing this? Is it a recognised media outlet, a European institution, or an anonymous user? What is the purpose of the message: to inform, raise awareness, sell a product, or spark debate? Is the source reliable, and does it provide verifiable evidence? This brief preliminary analysis helps detect misinformation and manipulated news, forming the first line of defence against viral hoaxes circulating on social media.

Expanding this approach, asking key questions also helps contextualise information. For instance, an alarming headline about digital violence against women might be generally accurate, but does it refer to a specific country or all of Europe? Is the statistic up to date?

These questions guide the reader toward a more complete and critical understanding of the content, preventing hasty conclusions or misinterpretations.

In short, integrating these key questions into daily digital routines helps people, especially young users, develop more conscious information-consumption habits, laying the groundwork for applying basic verification techniques and more advanced cognitive strategies, which we will cover in the following sections.

Basic Verification Techniques

Once the key questions about the information we consume are asked, the next step is to apply basic verification techniques. These tools ensure that content is reliable before sharing it or making decisions based on it. Two essential methods are triangulation and data corroboration.

Triangulation consists of comparing the same information across multiple independent sources. For example, if a viral video claims that "60% of young Europeans do not trust the media," this can be checked by looking for data from official organisations such as Eurostat, European Commission reports, or recognised media outlets. If multiple reliable sources agree, the information is more likely to be accurate; if the data differ significantly, caution is needed, and further evidence should be sought.

Data corroboration involves verifying specific facts in a news story or social media post. This includes verifying numbers, dates, quotes, and the names of people or institutions mentioned. For instance, a post about gender-based violence might consist of a striking statistic; corroborating it means checking the original report from UN Women or the academic study that supports the figure, rather than relying solely on the summary circulating online.

Applying these techniques not only protects against fake news and misinformation but also helps to analyse any message more critically, even when it comes from seemingly trustworthy sources.



Co-funded by
the European Union

In the European context, these skills are crucial for responsible participation in digital discussions and for preventing the spread of rumours or false data.

As McGrew, Ortega, Breakstone and Wineburg (2018) point out, teaching people to triangulate information and verify facts is among the most effective strategies for strengthening media literacy and combating misinformation in digital spaces.

Cognitive Strategies: Recognising Biases

A central part of critical thinking is learning to recognise bias, both in the information we receive and in our own perceptions.

Everyone has pre-existing ideas, preferences or beliefs that can influence how we interpret messages. Recognising these allows for a more objective evaluation of information and leads to better decision-making.

For example, an article on gender equality might present statistics that confirm our existing beliefs. If we do not question our tendency to accept only information that aligns with our perspective, we risk reinforcing confirmation bias. An effective strategy is to ask: "Am I accepting this information because it's true, or because it confirms what I already believe?" It is also helpful to seek opposing viewpoints or data to balance our perspective and avoid partial conclusions.

Another cognitive strategy is learning to detect bias in the message itself, such as exaggerated emotions, clickbait headlines or selective use of data. For instance, a viral meme about women in tech may highlight only one extreme example, creating a distorted impression. Analysing these elements helps avoid emotional manipulation and encourages a calm, reasoned interpretation of information.

These strategies promote deeper media literacy, enabling responsible participation in online discussions, avoiding the spread of false information, and contributing to a more reflective and ethical digital environment.

Distinguishing Facts, Opinions, and Emotions

A key skill in critical thinking is recognising the difference between facts, opinions, and emotions. This distinction allows for objective analysis of information and prevents sensationalist or biased messages from influencing interpretation.

- **Facts:** verifiable data, such as figures, statistics, dates, or specific events. For example, "According to Eurostat, in 2022, 32% of European women worked in technology sectors" is a fact because official sources can confirm it.
- **Opinions:** judgments, interpretations, or personal viewpoints about a topic. For example, "Women are not interested enough in technology" is an opinion, as it reflects a judgment that cannot be objectively verified.



Co-funded by
the European Union

- **Emotions:** Affective reactions that messages can provoke, such as fear, outrage or surprise. Example: an alarming headline about digital gender violence that exaggerates data to provoke outrage appeals more to emotion than to evidence.

Understanding these differences is especially important when dealing with viral news or memes, which often mix facts, opinions, and emotions in a single message. This distinction enables users to evaluate information more accurately, apply verification methods, and detect bias, as discussed in the previous section on cognitive strategies.

Moreover, the ability to separate facts from opinions and emotions strengthens media literacy and ethical information use. It helps determine which parts of a message constitute solid evidence and which require more critical analysis, thereby preventing the spread of misinformation.

In summary, this skill acts as an extra filter for critical thinking: by recognising facts, opinions and emotions, users can interpret information responsibly, reinforce verification habits and contribute to a more thoughtful and safe digital environment.

4.5 Practical Skills in Media Literacy

This section aims to provide practical tools to navigate the digital environment safely.

Beyond theory, these skills help identify misleading content, evaluate sources, recognise manipulations, and use digital verification tools effectively.

Such abilities are essential for responsible digital citizenship, particularly in Europe, where information sources are diverse and overloaded, requiring constant critical consumption.

1. Identifying Misleading Headlines and Clickbait

Sensationalist headlines aim to grab attention and generate clicks, but often distort reality. Detecting them is key to avoiding the spread of misinformation.

What to do: read beyond the headline, analyse the full content, and check whether the data presented aligns with other sources.

Practical example: a headline like "All women leave technology!" may exaggerate a real situation. Check official statistics and reliable articles before sharing.

Good practice: In Jovesólides'Jovesólides' Crítica project, participants learn to spot sensationalist messages on social media and digital news, developing the ability to critically analyse information before interacting with it.

2. Recognising Visual and Audio Manipulation



Co-funded by
the European Union

Images and videos are powerful tools that can be edited or taken out of context. Recognising visual or audio manipulation helps avoid misinterpretation or emotional manipulation.

What to do: look for suspicious elements, check for inconsistencies, and verify whether the content has been reused or modified.

Practical example: a meme about gender violence that combines shocking images with real statistics may provoke fear or outrage. Using reverse image search tools can verify whether images were used before and in what context.

Good practice: Jovesólides' Jovesólides' Coco project teaches how to recognise hate content and digital manipulation, providing tools to analyse images, videos, and audio online critically.

3. Evaluating the Credibility of an Online Source

Not all information on the Internet has the same quality or reliability. Evaluating authorship, reputation, and traceability is essential for responsible consumption.

What to do: ask who is publishing the information, whether the source is recognised, if verifiable references exist, and if the content can be traced to its origin.

Practical example: before sharing an article on pay equality, check whether it comes from official sources such as Eurostat, European Commission reports, or respected media outlets. Sites without explicit references should be approached with caution.

4. Basic Digital Verification Tools

Using digital tools makes it easier to check the authenticity and context of content:

- Reverse image search: Google Images or TinEye, for example, helps detect manipulated or out-of-context photos.
- Date verification: confirm that information is current to avoid spreading outdated data.
- Metadata analysis: review hidden data in digital files, including authorship, location, and potential modifications.

Practical example: verify a chart on digital violence by consulting the source, checking the date and author before sharing.

Integrating these techniques enables individuals to act responsibly, ethically, and safely in digital spaces, detect misinformation before it goes viral, and promote conscious, critical consumption.



Co-funded by
the European Union

The Crítica and Coco projects are examples of how these skills can be applied in practice, combining content analysis, source verification, and detection of hate or manipulated messages.

Reflective Questions

- How do emotions influence the way you interpret online content? Reflect on how specific images, videos or words make you react, and how that emotional response can affect your judgment.
- Can you recognise an example of gender-based disinformation you have seen recently? How was it presented, and why do you think it was effective or misleading?
- What role do you think young citizens should play in promoting ethical digital communication? Consider how small actions, such as correcting false data, using inclusive language, or verifying visuals, can contribute to a healthier digital environment.

4.6 References

Frau-Meigs, D. (2021). *Media and Information Literacy in the Age of Algorithms: Critical Thinking, Ethics and Responsibility*. UNESCO.

Jovesólides. (2025). CRÍTICA – Proyecto de alfabetización mediática. Retrieved from <https://jovesolid.es/proyectos-emprendedores/critica>

Jovesólides. (2025). CoCo – Recursos contra el odio. Retrieved from <https://jovesolid.es/proyectos-emprendedores/coco-recursos-contr-el-odio>

Livingstone, S. (2004). Media literacy and the challenge of new information and communication technologies. *Communication Review*, 7(1), 3–14.

McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2018). The Challenge That's Bigger Than Fake News: Teaching Youth to Evaluate Digital Information. *American Educator*, Fall 2018.

Redecker, C., & Punie, Y. (2017). *European Framework for the Digital Competence of Educators (DigCompEdu)*. Luxembourg: Publications Office of the European Union.

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.



Co-funded by
the European Union

5. Identifying disinformation patterns

Written by Aleksandra Radevska and Natasha Dokovska, Journalists for Human Rights (JHR), 2025

5.1. Introduction to Disinformation Patterns

Propaganda, rumour, manipulation, and fake news in public communication are well-known terms for the typical information consumer. But analysing the information disorder and how disinformation operates not as isolated falsehoods, but as structured, recurring patterns designed to manipulate perception and behaviour is a more detailed task. It connects the technical, social, and gendered dimensions of disinformation, showing how these patterns exploit both cognitive biases and digital infrastructures.

The goal in this chapter is to move from recognising disinformation to understanding its design and how stories are built, repeated, and amplified across networks. The focus on gendered disinformation highlights how women, girls, and marginalised groups are disproportionately targeted through narratives that draw on sexism, fear, and moral outrage.

Throughout the chapter, be encouraged not only to identify patterns but to question their underlying power structures: Who benefits from these narratives? How are identities and emotions used to manipulate? And how might civic and media literacy challenge these dynamics? We live in a world where information is instant and abundant. This constant flow of posts, stories, and videos shapes your identity, beliefs, and opportunities. Because of this, learning how to handle information is not optional; it is a form of digital self-defence.

5.2. Information Disorder - Defining the Landscape

We live in an age where information is instant, abundant, and emotionally charged. Every scroll, click, and share can shape how we see the world, and how the world sees us. Yet not all the content that flows through our feeds is accurate or even meant to inform. Much of it is designed to persuade, provoke, or manipulate. This is what scholars call information disorder, a term that captures the many ways in which information can mislead, distort, or harm.

When analysing information disorder, it is necessary to distinguish between misinformation, disinformation, and malinformation:

- **Misinformation**: false or inaccurate information shared without intent to deceive (Wardle & Derakhshan, 2017), often spread because individuals believe it to be true or find it emotionally resonant, "because it aligns with their pre-existing beliefs."



Co-funded by
the European Union

(Lewandowsky, Ecker, & Cook, 2017). For example, rumours, outdated facts, or wrongly interpreted data.

- **Disinformation:** False or misleading information deliberately produced and circulated to deceive, manipulate, or advance a strategic goal, often for political, financial, or ideological gain. "Disinformation operates as a strategic instrument of influence, deliberately manipulating facts, emotions, and identities to achieve political or psychological effects." (Baines & Jones 2020).

- **Malinformation:** Genuine information shared out of context or with malicious intent to cause harm, like private messages leaked to shame or silence someone. "The strategic leaking of authentic but private materials illustrates how malinformation can undermine trust and damage democratic discourse." (Bradshaw & Howard, 2019).

These categories matter because they shape how we respond. We can correct misinformation with facts, but disinformation requires exposure, accountability, and critical awareness. Malinformation, meanwhile, challenges us to think ethically about privacy, harm, and responsibility.

While "fake news" remains a popular term, it oversimplifies complex dynamics. Political actors have weaponised it to discredit journalism. (Tandoc, Lim, & Ling, 2018). Scholars now reject it in favour of more precise concepts that focus on intent, orchestration, and harm. Together, these categories reveal that not all harmful content online is "fake." Sometimes, the most damaging narratives are half-truths or real stories told for the wrong reasons. Recognising the difference empowers young people to protect themselves and others from manipulation.

Getting familiar with information disorder helps them make informed choices without being manipulated by hidden agendas, recognise emotional design, how outrage, fear, or humour are used to trigger engagement and distort judgment, spot manipulation early, identifying recurring patterns and tactics before they go viral, and support digital citizenship, promoting transparency, empathy, and accountability in online communities.

Learning to navigate this landscape means more than spotting "fake news." It's about building resilience: understanding how information flows, how it can be distorted, and how to keep your critical mind active in a world that constantly competes for your attention. In short, information literacy is empowerment, a skill that lets young people shape the digital future rather than be shaped by it.

Developing critical digital literacy

Social media platforms play a central role in the debate over fake news. Platforms like Facebook, X(Twitter), TikTok, and YouTube have become the main arenas for news consumption, yet their algorithms prioritise engagement over accuracy.

These systems reward content that evokes strong emotions, outrage, fear, and identity affirmation, thereby amplifying polarising and misleading content. Research on "echo



Co-funded by
the European Union

chambers" (Pariser, 2011) and "filter bubbles" (Sunstein, 2018) shows how algorithmic curation can reinforce pre-existing biases and isolate users from diverse viewpoints. Consequently, the debate over fake news is inseparable from questions of platform responsibility, data ethics, and democratic oversight. Should private companies regulate truth? Who decides what counts as harmful or false? The struggle between freedom of expression and accountability remains unresolved, reflecting deeper tensions between technological innovation and public interest.

Although much of the discourse originates in Western democracies, the fake news phenomenon manifests differently across global contexts. In authoritarian settings, accusations of fake news often serve as tools for repression. In post-conflict or ethnically divided societies, misinformation can inflame existing tensions. In developing media markets, limited journalistic infrastructure and digital literacy make communities more vulnerable to manipulation.

In the Balkan region, for instance, disinformation is often intertwined with nationalist narratives, corruption, and foreign influence. These dynamics illustrate how the fake news crisis is not only about information quality but also about democratic resilience and social cohesion.

Given its politicisation, many researchers advocate abandoning the term "fake news" altogether. Instead, they propose frameworks such as information disorder (Wardle & Derakhshan, 2017) or networked propaganda (Benkler, Faris & Roberts, 2018). These models emphasise: the ecosystemic nature of disinformation; the role of social and technological infrastructures; the motivations and emotional mechanisms behind sharing falsehoods.

This shift moves the discussion from what is fake to why false information thrives, exploring the human, structural, and cultural dimensions of belief and persuasion. This debate is particularly relevant for young people, digital natives, navigating the most complex information environment in history, one where social validation often substitutes for credibility and algorithms shape perceptions of reality.

Becoming informed citizens now requires more than distinguishing true from false; it involves understanding how information operates as power, who creates it, who benefits from it, and how it affects civic participation, gender equality, and social trust.

Developing critical digital literacy is, therefore, essential not only for protecting oneself from manipulation but for actively participating in shaping democratic discourse

5.3. Disinformation Patterns and How to Recognise Them

Disinformation is not random or chaotic; it follows recognisable patterns of design, dissemination, and emotional appeal. These patterns reveal how disinformation functions not only to mislead but to reshape public perception, polarise societies, and erode trust in democratic institutions. It is effective not so much because of the false facts but more because it strategically manipulates truth, context, and emotion.



Co-funded by
the European Union

Understanding these recurring patterns helps us move from individual fact-checking to systemic detection, recognising how disinformation operates across platforms and audiences.

Fabrication and false attribution are a pattern that involves the creation of entirely false information, often presented as coming from a credible or official source. Fabricated stories mimic journalistic conventions (logos, bylines, formatting) to simulate legitimacy. Signs to be recognised include unfamiliar or suspicious domain names (e.g., ".co" instead of ".com"), a lack of verifiable author information or contact details, and claims attributed to institutions or experts that do not exist. Such patterns rely on surface credibility; they look real before they are read critically.

Manipulation of genuine content occurs when authentic photos, videos, or documents are altered, selectively edited, or taken out of context to change their meaning, resulting in "malinformation" that blurs the boundary between truth and deception and makes detection more difficult. Signs to be recognised include image or video inconsistencies (lighting, cropping, background changes), mismatched timestamps or geographic settings, and headlines that contradict the body text or caption. Tools like reverse image search or metadata analysis often reveal when a visual has been repurposed or reframed to serve a new, misleading narrative.

Disinformation spreads through repetition across multiple accounts and channels, a process that creates the illusion of consensus. Coordinated networks (including bots or troll farms) amplify certain narratives until they dominate public discourse. Signs to be recognised are dozens of accounts sharing identical text within minutes, sudden spikes in engagement or trending hashtags without organic buildup, or links repeatedly shared by the same cluster of actors across different platforms. This pattern reflects the logic of networked propaganda (Benkler, Faris & Roberts, 2018), where visibility equals legitimacy.

As disinformation relies heavily on emotional arousal, fear, anger, pride, or outrage, to prompt rapid engagement and sharing, content framed as a moral conflict ("us vs. them," "good vs. evil") bypasses analytical reasoning. Signs to be recognised are overuse of emotionally charged language ("shocking," "horrific," "traitors," "destroying our country"), simplistic explanations for complex issues or appeals to identity ("real patriots," "true believers," "protect our children"). Emotionally manipulative framing signals that a message's goal is not to inform but to mobilise sentiment.

Conspiratorial and polarising narratives are a pattern involving constructing a coherent worldview that explains complex realities through hidden plots or malevolent actors. Conspiracy narratives transform uncertainty into certainty and scepticism into belonging. It can be recognised through claims of secret knowledge or hidden truth ("what the media won't tell you"), a precise enemy figure or scapegoat (e.g., minorities, elites, global institutions) or dismissal of all contradicting evidence as part of the "cover-up." These narratives thrive in environments of low institutional trust, providing emotional clarity at the cost of factual accuracy.



Disinformation frequently involves reusing old or unrelated content to fit new events. This can include photos from previous disasters or videos from different countries. It can be recognised by images or videos that do not match current weather, architecture, or local details; content circulating immediately after breaking news events (before verified reports appear); or content with no credible timestamps or source links. Temporal manipulation works because people tend to assume visual evidence is recent and trustworthy.

Fake experts or selectively cited data to appear evidence-based is also something disinformation campaigns often rely on. This pseudo-authority appeals to those seeking rational justification for pre-existing beliefs. Recognition signs are experts or organisations that lack traceable academic or institutional credentials, misuse statistics (percentages without sample sizes or context), or reference "studies" that cannot be located or verified. This pattern exploits the appearance of scientific credibility while disregarding its actual standards of rigour.

Beyond the content itself, disinformation can be recognised through its psychological and cultural cues. It aligns with what people want to believe and how they identify socially. Cognitive markers are confirmation bias - the tendency to accept information that supports one's worldview, availability bias - emotionally vivid stories feel more "true.", and social validation - repetition and likes create perceived credibility. Cultural markers are narratives that reinforce national, ethnic, or gendered hierarchies; language that evokes nostalgia for a lost past or a moral order; and binary framings that deny nuance ("heroes vs. traitors," "truth vs. lies").

Recognising these cues requires media literacy combined with self-awareness, understanding not only how we consume media but also why specific messages feel persuasive.

Detection Strategies: Your Digital Self-Defence Kit

As we analyse the patterns which disinformation follows, with the intent to reshape public perception, polarise societies, and erode trust in democratic institutions, while using false facts to manipulate truth, context, and emotion strategically, we see there is a need to move from individual fact-checking to systemic detection, recognising how disinformation operates across platforms and audiences. It can be accomplished through effective detection strategies, though they also have limitations.

Detection falls into two domains: content-based approaches, which include analysing textual or stylistic markers (Shu et al., 2017) and source credibility, drawing on curated lists of low-quality domains maintained by fact-checkers and journalists (Lazer et al., 2018), and context-based approaches that examine diffusion patterns, bot activity, and coordinated behaviour (Ferrara et al., 2016).

Each method has limits. Content-based systems can miss subtle manipulations, while context-based approaches often lack the metadata needed to establish intent. Ethical



Co-funded by
the European Union

concerns also arise when automated moderation risks silencing marginalised voices (Gorwa, Binns, & Katzenbach, 2020).

Toward Pattern Recognition as Civic Literacy

Disinformation thrives where citizens lack trust, context, or critical skills. Detecting it is not a matter of paranoia but of pattern literacy, recognising recurring manipulative forms across issues and contexts.

When individuals learn to identify these patterns, the repetition, emotional charge, false context, and pseudo-expertise, they become less reactive and more reflective. This capacity for pattern recognition is central to democratic resilience: it helps communities resist manipulation and sustain a shared space of reasoned debate.

The detection of disinformation has evolved from simple fact verification to ecosystem analysis. In the early 2010s, journalists and researchers primarily focused on verifying individual claims, the content of disinformation. Today, the field recognises that falsehoods are embedded in networked systems of production and circulation (Starbird, 2019; Benkler et al., 2018).

Modern detection, therefore, operates at multiple levels:

- Micro-level: individual posts, visuals, or claims;
- Meso-level: coordinated networks, influencers, or campaigns;
- Macro-level: long-term narrative ecosystems and information operations.

This shift reframes detection as a process of pattern recognition, network mapping, and intent assessment, rather than merely identifying "fake" versus "true."

Detection is not just about catching lies; it's about strengthening your critical mind. Here are the key skills (known as essential digital literacy) you need:

1. Slow Down Your Thinking

Disinformation relies on fast reactions. Before you click "Like" or "Share," pause and ask yourself: Does this message make me feel very angry or very proud? (If yes, slow down!) Who benefits from my believing this narrative? What action or emotion is it trying to provoke?

2. Read Laterally, Not Down

Instead of focusing solely on the article or post you are reading (reading down the page), professional fact-checkers open new tabs and verify the source across the web (reading laterally). Action: If a claim comes from "The Institute for Truth," open a new tab and search for the Institute. Check its credentials, affiliations, and funding before trusting its claim.

3. Verify the Visuals

Photos and videos are easily manipulated. Action: Use a reverse image search (like Google Lens or TinEye) to see if the photo or video has been used before with a different caption. Look for earlier versions; they reveal whether the image is being repurposed.

4. Check the System Context



Co-funded by
the European Union

Look beyond the content and analyse how it is spreading. Action: Compare the claim across multiple independent, reputable outlets. If only one suspicious source is reporting a significant incident, treat it with high scepticism. Check if the information is being amplified in an inorganic, coordinated way, suggesting manipulation rather than natural sharing. Action: Use Botometer, Hoaxy, Graphika, Gephi, and CrowdTangle to map and visualise content flows. Temporal clustering and language pattern analysis reveal automation or orchestration. These methods shift detection from identifying lies to identifying the architecture of amplification.

5. Use Linguistic and Semantic Detection

Look for identifiable linguistic signatures that disinformation often has. Natural language processing (NLP) research identifies stylistic and semantic cues associated with manipulative intent. Fake or misleading news tends to use: high emotional valence (anger, fear, disgust); frequent superlatives ("amazing," "unbelievable"); imperatives ("share this now," "don't trust them"); simplified causal language ("Xcauses Y"). Disinformation relies on narrative coherence, offering simple, emotionally satisfying explanations for complex issues. Action: Focus on narrative tropes: heroes vs. villains, victims vs. aggressors, "the system vs. the people." Recognising these recurring frames helps anticipate new waves of similar manipulation. Example: "They don't want you to know the truth", a classic disinformation frame invoking secrecy and betrayal.

6. Use Cross-Platform and Temporal Detection

Disinformation rarely stays within one platform. It flows across ecosystems, Facebook, TikTok, Telegram, and YouTube, adapting its form to each audience. Thus, detection requires cross-platform triangulation and timeline analysis. Tracing where and when a narrative first appeared helps distinguish organic discussion from strategic injection.

Chronological mapping often reveals that viral misinformation was seeded days earlier on fringe forums before reaching mainstream social media. Coordinated spikes in engagement, especially at non-local hours, suggest automated amplification. Plotting volume over time can expose orchestrated virality distinct from natural diffusion.

Collaborative and Institutional Detection Strategies

No single actor can address the complexity of disinformation. Detection requires multi-level collaboration among journalists, researchers, platforms, and civil society.

Organisations such as IFCN, EUvsDisinfo, and Bellingcat combine crowdsourced intelligence with expert analysis. Collaborative verification accelerates detection through distributed expertise, local knowledge, linguistic diversity, and specialised technical skills.

Recent initiatives (e.g., Meta's Ad Library, X Twitter's Community Notes) allow partial public oversight of content promotion. However, limited data access remains a significant obstacle; detection at scale depends on platform openness and ethical data sharing.

Detection is most sustainable when embedded in public literacy. "Prebunking", exposing people to weakened versions of misinformation before they encounter it, builds psychological immunity (Roozenbeek & van der Linden, 2019). Interactive games like Bad News and Go Viral! Demonstrate that awareness of manipulation techniques improves future detection accuracy.



Co-funded by
the European Union

Limitations and Ethical Considerations

Even though new tools are getting better at spotting false information, there are still some significant challenges to keep in mind:

- It's all about context: What's seen as "disinformation" in one country might be seen as opinion or humour in another. Culture and politics really shape what counts as "false."
- Biased tech: AI tools can sometimes get it wrong, flagging jokes, memes, or minority voices as "fake," while missing real problems.
- Too much transparency? If platforms share exactly how their detection systems work, bad actors can learn to trick them.
- Info overload: Seeing too many fact-checks and warnings can make people tune out completely, or think everything is fake.

That's why it's essential to strike a balance: use solid, evidence-based methods, while also staying aware of context, intent, and people's rights to express themselves and access information.

Detection as a Civic Skill

Spotting disinformation isn't just about using innovative tools; it's a democratic life skill.

To really detect and resist false information, we need to be able to:

- Think critically about the sources of our information.
- Recognise manipulation, who's trying to shape our opinions and why;
- See the power behind messages, and how influence can make something look like "the truth."

In this sense, disinformation detection isn't just personal, it's collective resilience. It's about helping our communities stay informed, discuss openly, and make decisions based on facts, even when the information space feels chaotic. As Benkler and colleagues (2018) remind us: "The goal is not just to eliminate falsehoods, but to strengthen the knowledge foundations of democracy."

Reflection Activity

When it comes to fighting disinformation, there are no easy answers, just a lot of tricky trade-offs. Should detection tools focus on speed, flagging content the moment it appears (even if that means some posts get unfairly removed)? Or should they focus on accuracy and transparency, even if that means harmful falsehoods might stay online longer? From a gender perspective, the challenge gets even more complex. If detection systems ignore intersectionality, the way gender, race, language, and culture overlap, they can actually make things worse. That's why we need a "both-and" approach, not choosing between tech and people, or speed and care, but combining them. Use both content and context analysis to understand not just what is said, but why and how it spreads. Balance freedom of expression with the need to protect people from harm and hate. Pair technical tools with civic education so that young people can think critically, question sources, and build online spaces that are fair, inclusive, and informed.



5.3: Key Actors Behind Disinformation

Political Drivers

Political science and communication research highlight that states and political elites use disinformation as a tool of power projection (Howard et al., 2018; Giles, 2016).

Disinformation functions both externally (through geopolitical influence) and internally (to manage domestic legitimacy). Theories of propaganda, from Cold War covert communication to "computational propaganda" (Bradshaw & Howard, 2019), show how states adapt old techniques to digital infrastructures.

Political elites and state actors deploy disinformation to control narratives, delegitimise opponents, and sustain authority. Externally, states use disinformation for geopolitical projection, to weaken rivals, influence elections, or spread ideological influence abroad. For example, Russian disinformation networks targeting Western Balkans media (Serbia, Montenegro, North Macedonia) spread narratives aligned with Kremlin interests, anti-EU sentiment, vaccine scepticism, and traditionalist gender norms. This extends the "information warfare" logic outlined by Giles (2016) and Howard et al. (2018).

Internally, ruling parties or leaders deploy disinformation to manufacture legitimacy or discredit dissent. For example, in North Macedonia before 2017, partisan portals circulated stories glorifying government projects while labelling opposition or journalists as "foreign agents." These campaigns blurred the line between state PR and propaganda.

Computational propaganda (Bradshaw & Howard, 2019) modernises Cold War-style narrative control through digital means, bots, troll farms, and algorithmic amplification, allowing power to be exercised through visibility manipulation rather than censorship.

Looking into the gender dimension, female politicians, journalists, and activists often face hybrid disinformation: both political (to silence criticism) and gendered (to discredit through misogyny). For example, in Serbia and Bosnia, women journalists critical of nationalist or clerical power structures were attacked as "immoral" or "foreign-funded," blending political defamation with patriarchal tropes.

Financial Drivers

The disinformation economy is not sustained only by ideology; it's profitable (Silverman & Alexander, 2016; Hughes & Waismel-Manor, 2021). Concepts like the "attention economy" explain why sensationalism and controversy thrive on ad-driven platforms, and economic theories of "clickbait capitalism" situate disinformation within broader structural incentives created by global advertising markets.

The "attention economy" rewards content that triggers emotional engagement, outrage, fear, or moral disgust. Algorithms prioritise velocity over veracity. For example, clickbait portals in the Western Balkans (e.g., Macedonian "fake news farms" during the 2016 US.



Co-funded by
the European Union

election) generated thousands of euros in ad revenue by producing sensationalist headlines about Hillary Clinton or migrants.

"Clickbait capitalism" (Hughes & Waismel-Manor, 2021) frames this as an economic structure: disinformation is a rational market behaviour within ad-driven ecosystems. Publishers optimise content for virality because digital ads pay per view, not per truth. Major platforms benefit indirectly: every viral post, even false or toxic, increases engagement metrics and thus advertising revenue. Platform affordances (autoplay, recommendations, trending lists) monetise outrage and reward polarisation. Gendered and sensational stories ("female politician's scandal," "immoral influencer," etc.) generate high engagement. Misogynistic clickbait performs better because it exploits emotional and moral reactions, shame, envy, disgust, and converts sexism into profit.

Ideological Drivers

Sociological and psychological models show how ideological worldviews – nationalism, extremism, conspiracy thinking – fuel the production and spread of disinformation (Fielitz & Marcks, 2019; Argentino, 2020). Cognitive frameworks such as belief perseverance and motivated reasoning explain why conspiracy communities resist factual correction (Pennycook & Rand, 2019).

Disinformation thrives where ideology and identity are emotionally charged. Nationalism and populism frame truth as partisan, "our truth" versus "their lies." For example, Balkan populist leaders often equate critical journalism or feminist movements with "Western agendas," merging anti-globalist and anti-gender rhetoric.

Conspiracy worldviews provide simple explanations for complex social change ("global elites control everything"). According to Fielitz & Marcks (2019), these movements sustain alternative epistemic communities that reject institutional truth.

Psychological mechanisms such as motivated reasoning and belief perseverance (Pennycook & Rand, 2019) show how people selectively accept information that confirms their identity and dismiss contradictory facts. Once individuals integrate a belief into their group identity, correction feels like betrayal.

Moral panic dynamics, around migration, feminism, or LGBTQ+ rights, turn ideological disinformation into a tool for cultural control. For example, in 2019–2021, "anti-gender" disinformation campaigns in Eastern Europe framed the Istanbul Convention as an attack on "traditional family values," blending conspiracy, nationalism, and patriarchal ideology. Ideological disinformation reinforces gender hierarchies by portraying women's rights as foreign, unnatural, or dangerous to social cohesion, legitimising political backlash and public hostility toward feminists, LGBTQ+ activists, and women journalists.

Gendered case studies

Sanna Marin "Party Video" Disinformation (Finland, 2022)



Co-funded by
the European Union

In 2022, a private video of Finnish Prime Minister Sanna Marin dancing at a private event was leaked and went viral. What began as a non-political moment quickly turned into a global gendered disinformation event.

The mechanisms of attack that were used were malinformation and misrepresentation - real footage was stripped of context and circulated with captions implying drug use, irresponsibility, and moral failure; amplification dynamics - memes and edited clips spread rapidly across TikTok, Twitter, and Telegram, turning a personal moment into a moral scandal; and gendered narrative - commentators framed her as "unserious," "immature," and "unfit to lead", tropes rarely used against male politicians in similar situations.

The event revealed the double standards of political credibility for women leaders: personal behaviour is moralised and politicised. Marin took a voluntary drug test (negative), highlighting the pressure of defending against gendered insinuations rather than factual claims.

A mix of malinformation (accurate content used to harm) and gender bias fueled this case, showing how disinformation ecosystems exploit gender norms to question women's competence and morality.

Ana Brnabić and Gendered Disinformation (Serbia, 2020–2022)

Serbian Prime Minister Ana Brnabić, as the first openly gay head of government in the Balkans, became the target of gendered and homophobic disinformation campaigns. Mechanisms of attack were political and Ideological blend - nationalist and conservative media spread conspiracy narratives claiming her "Western agenda" was aimed at "destroying Serbian values"; memetic framing - fake images and demeaning memes portrayed her as "controlled by Brussels," or mocked her sexual orientation to question her patriotism; as well as platform amplification - Facebook groups aligned with extremist and clerical actors circulated these memes through coordinated shares, giving the illusion of widespread sentiment.

The campaigns served a dual function: mobilising conservative voters through moral outrage and discrediting liberal or pro-EU positions. This demonstrates how gender and sexuality are instrumentalised in hybrid information operations, not only to attack individuals but to shape political identity narratives. Brnabić's case exemplifies gendered disinformation as ideological weaponry, a fusion of sexism, nationalism, and populism that mobilises resentment and fear to maintain power hierarchies.

Reflection Activity

Disinformation rarely fits into neat categories. Instead, hybridisation blurs lines between state, commercial, and ideological actors. Applying a gender lens adds further complexity: political and ideological actors often target women journalists, activists, and politicians with gendered disinformation that weaponises sexism and stereotypes. Recognising the actor–incentive nexus is essential to effectively countering such gendered harms. Who benefits from this narrative being amplified? Is this story trying to inform, persuade, or



provoke me? What kinds of content keep me scrolling, and why do I think that is? When I share something quickly, am I amplifying someone's message or helping it go viral in ways I don't intend? How are women and gender minorities represented in this content? Are stereotypes, sexuality, or morality being used to attack credibility rather than debate ideas? Is this "outrage post" being monetised through clicks, shares, or donations? How can I respond constructively when I spot manipulative or false narratives? What role can young people play in creating a healthier information space, as digital citizens, not just consumers? What kind of internet culture do I want to help build?

5.4: Risks and Alternatives to Debunking

Debunking, the act of correcting false or misleading information, seems like the obvious solution to misinformation. But it's not always that simple. If done carelessly, debunking can unintentionally reinforce the very narratives it aims to dismantle.

The main risks include:

- **Amplification: Giving visibility to obscure claims (Tandoc, Lim, & Ling, 2018).** When journalists, fact-checkers, or influencers respond to a false claim, they make it more visible. For example, a conspiracy circulating in small online groups can suddenly reach millions if a major outlet "debunks" it. This is called the "Streisand effect; Attempts to suppress or disprove a claim can increase curiosity and attention. Why it matters: Algorithms don't know what's true; they just reward engagement. So, even critical posts can help spread falsehoods further. Before responding, ask: Who has seen this claim so far? Is it viral or still marginal? If it's obscure, silence or "prebunking" may be wiser.
- **Legitimisation: Denials can validate narratives as "forbidden truths" (Sunstein & Vermeule, 2009).** When we engage with false narratives too seriously, we risk making them sound credible, as if they deserve equal weight in public debate. For example, when scientists "debate" climate change deniers on TV, it creates the false impression that there's still uncertainty about basic climate science. Denials can backfire by framing misinformation as "the thing they don't want you to know", feeding into the "forbidden truth" mindset common in conspiracy communities. Avoid repeating false claims verbatim. Focus instead on explaining the tactic, not the rumour (e.g., "This post uses fear to manipulate emotion" instead of "No, vaccines don't cause infertility").

-Backfire and Polarisation: While rare, corrections can entrench belief or deepen divides (Nyhan & Reifler, 2010).

Sometimes, correcting misinformation can make people cling more tightly to their original beliefs, especially if the correction threatens their identity or worldview. For example, studies by Nyhan & Reifler (2010) found that strong partisans often doubled down on their beliefs after being corrected about political facts. This is known as the backfire effect, though it's less common than once thought, it still occurs when people feel attacked or shamed. In online spaces, corrections can also trigger polarised comment wars, reinforcing



Co-funded by
the European Union

"us vs. them" dynamics. Approach with empathy and respect. Instead of saying "You're wrong," try "That's a common claim, here's what researchers actually found."

Ethical Takeaways:

- Avoid repeating false claims unnecessarily , focus on correcting the tactic, not the myth itself.
- Consider audience identity; corrections may need to be framed differently depending on values, cultural context, and prior beliefs.
- Protect vulnerable targets , especially women, minorities, and activists.
- Combine strategies , debunking, prebunking, and narrative redirection to minimise ethical risks.
- Maintain transparency , explain why a claim is false, how it spreads, and who is responsible for verification.

5.5. Conclusion

Disinformation isn't just about a few fake stories; it's a system designed to shape how we think, feel, and act. It works by repeating patterns, using emotional language, featuring fake experts, editing visuals, and coordinating sharing, which make false narratives feel real. These tactics are powerful because they use both our emotions and the way digital platforms are built.

Significantly, disinformation doesn't affect everyone in the same way. Women, girls, and marginalised groups are often targeted with sexist and harmful messages meant to silence or discredit them. This shows that disinformation isn't just about facts; it's also about power and control.

Learning to spot these patterns is a key life skill. It helps us pause before reacting, question what we see online, and understand who might benefit from spreading a particular message. We can all use simple strategies: checking sources, verifying images, looking for emotional triggers, and thinking about the bigger picture.

Fighting disinformation isn't the job of governments or platforms alone; it's something we can do together. By becoming more aware, we protect not just ourselves but also our communities and democracy. In the end, recognising and resisting manipulation is a form of empowerment.

5.6. References

Argentino, M. (2020). QAnon and the storm of the U.S. Capitol: The offline effect of online conspiracy movements. *The Conversation*.

Baines, P., & Jones, N. (2020). Influence operations and the psychological dimensions of



Co-funded by
the European Union

disinformation. *Defence Strategic Communications*, 8, 63–92.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalisation in American Politics*. Oxford University Press.

Bradshaw & Howard (2019). "The Global Organisation of Social Media Disinformation Campaigns," *Journal of International Affairs*, 71(1.5), 23–32.)

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>

Fielitz, M., & Marcks, S. (2019). *Digital fascism: Challenges for the open society in times of disinformation*. Amadeu Antonio Stiftung.

Giles, K. (2016). *The handbook of Russian information warfare*. NATO Defence College

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. <https://doi.org/10.1177/2053951719897945>

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, social media and political polarisation in the United States, 2012–2018*. Oxford Internet Institute.

Howard, P. N. (2020). *Lie machines: How to save democracy from troll armies, deceitful robots, junk news operations, and political operatives*. Yale University Press.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>

Lewandowsky, Ecker, & Cook, 2017. Beyond Misinformation: Understanding and Coping with the 'Post-Truth' Era, *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behaviour*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.

Pennycook & Rand (2019). *PNAS*, 116(7), 2521–2526.

Roozenbeek, J., & van der Linden, S. (2019). The Fake News Game: Actively Inoculating Against the Risk of Misinformation. *Journal of Risk Research*, 22(5).



Co-funded by
the European Union

Silverman, C., & Alexander, L. (2016). How teens in the Balkans are duping Trump supporters with fake news. BuzzFeed News.

Starbird, K. (2019). Disinformation's Spread: Ecosystem or Epidemic? *Nature Human Behaviour*, 3, 451–452.

Sunstein, C. (2018). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

Sunstein, C. R., & Vermeule, A. (2009). Conspiracy theories: Causes and cures. *Journal of Political Philosophy*, 17(2), 202–227. <https://doi.org/10.1111/j.1467-9760.2008.00325.x>

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news." *Digital Journalism*, 6(2), 137–153. <https://doi.org/10.1080/21670811.2017.1360143>

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe.



Co-funded by
the European Union

6. Gendered Disinformation: Language and Representations

Written by Roya Baicu, In Varietate Concordia (IVC), 2025

6.1 Introduction: The Gendered Weapon

After seeing the title of the module, you might ask yourselves why **gendered disinformation** is an important topic, as many of you probably didn't get the chance to learn about it until now. Well, gendered disinformation has become one of the most potent forms of **manipulation** in contemporary politics and social media. **Women** in public life, politicians, journalists, activists, entertainers, and other visible figures are disproportionately **targeted** with attacks that undermine their credibility, authority, and participation. Unlike general criticism, which may focus on ideas, policies, or competence, **gendered disinformation targets identity, appearance, or morality** (Butler, 1990). Politicians are portrayed as unfit or emotional; activists are accused of being radical, unnatural, or controlled by external forces; journalists are sexualised, insulted, or framed as untrustworthy; public figures such as actresses, TV hosts, and musicians are harassed for their looks, behaviour, or private lives. The common thread is clear: the focus shifts from what women say or do to what they are presumed to be, silencing debate and weakening legitimacy.

This pattern is visible everywhere, including Romania. In **Romania**, politicians like Oana Țoiu and Laura Codruța Kövesi faced intense **online harassment** that focused on their personal traits, morality, or looks rather than their policy work. We've seen manipulated images (deepfakes) circulated widely, and leaders like Gabriela Firea were targeted with memes about her voice, demeanour, and appearance. Beyond individual cases, complex movements, often tied to political or religious networks, weaponise the vague narrative of "**gender ideology**" to mobilise fear and resistance against equality movements. Public discourse frequently emphasises their physical appearance, private behaviour, or moral character over professional achievements, illustrating that gendered disinformation affects women across all sectors of public life.

Globally, the dynamics are strikingly similar. **New Zealand's** former Prime Minister **Jacinda Ardern** was described as "too emotional" or "weak". U.S. Vice President Kamala Harris faces accusations of being "incompetent" and having a "weird laugh", and viral campaigns about her private life targeted **Finland's Sanna Marin**. In every instance, gender is used as a weapon, shifting the focus away from political substance.

The **weaponisation** of gender in disinformation campaigns operates through several interrelated strategies. Language frames women as irrational, morally suspect, or disloyal; media representation marginalises or erases their contributions; intersectional vulnerabilities, such as race, class, or ethnicity, amplify attacks, and emotional manipulation provokes fear, anger, or disgust, increasing the virality of harmful narratives.



Co-funded by
the European Union

Let's start with a big, urgent truth: **disinformation is gendered**, and gender is being weaponised in the service of **exclusion**. Understanding gendered disinformation requires looking beyond isolated incidents to the broader structures of power and inequality that they reproduce. When women in politics, media, entertainment, or activism are targeted, the harm is personal but also deeply political. These campaigns send a warning to others: this is what happens when women occupy public roles. The chilling effect on **democratic participation is significant**.

This chapter begins with a simple but urgent claim: disinformation is gendered, and gender is weaponised in the service of exclusion. In the sections that follow, we will explore theories of language and gender, examine media representation and symbolic erasure, analyse intersectional dynamics, and investigate Romanian and global case studies. We will also reflect on tools available for resistance, from critical reading strategies to community-based responses.

Ultimately, the goal is not only to understand how gendered disinformation works but also to consider how we, as readers, citizens, and allies, can **resist** it. By paying attention to the words we use, the stories we amplify, and the voices we silence, we can begin to disarm the gendered weapon and reclaim the power of representation for democratic and inclusive purposes.

6.2 Language, Gender, and Power

Here's the **key idea**: language and media are never truly **neutral**. They actively shape how we think, what we perceive, and who holds power. Gendered disinformation cleverly uses words, images, and even silence to distort women's credibility, undermine their expertise, and normalise stereotypes. This submodule will explore exactly how language becomes a strategic instrument in these harmful campaigns.

Theoretical Foundations

Feminist linguistics has long shown that language both reflects and reproduces **social hierarchies**. Robin Lakoff (1975) argued that women's speech is conditioned to appear tentative or deferential, a perception rooted in social norms rather than biology. Dale Spender (1980) highlighted how linguistic structures encode **male experience as the default**, marginalising women. Deborah Tannen (1990) demonstrated how power dynamics shape different communicative styles, while Judith Butler (1990) reframed gender itself as performative, showing how repeated discourse produces norms that can be mobilised against women.

This framework helps us see how disinformation campaigns exploit **familiar tropes**: the "hysterical woman," the "seductive manipulator," or the "foreign agent." These are not new; they are cultural stereotypes embedded in language and recycled in contemporary digital media. Intersectionality sharpens this dynamic: Roma women activists, for example, are labelled not only as "emotional" but also as "outsiders," reinforcing both gendered and ethnic hierarchies.



Co-funded by
the European Union

Language as a Weapon

Disinformation rarely attacks policies or ideas directly. Instead, it shifts focus to personal identity. Repeated linguistic framing paints women as "weak," "incompetent," "immoral," or "unfit," redirecting public debate away from substantive issues. Over time, audiences internalise these cues, diminishing trust and credibility (Kahneman, 2011).

Romanian Examples:

- Viorica Dăncilă was mocked through memes and commentary emphasising her accent, clothing, and manner of speaking. Instead of being seen as a policy critic, she was framed as "simple" or "unprepared," reinforcing the stereotype that women in politics lack competence.
- Laura Codruța Kövesi, Romania's former European Chief Prosecutor, was targeted with campaigns labelling her "arrogant," "politically biased," or a "foreign puppet." Attacks focused on demeanour and appearance rather than her legal record, discrediting her authority through gendered tropes.
- Oana Țoiu, when appointed to political office, faced a wave of deepfake images and viral posts mocking her looks. The attacks trivialised her role and spread faster than official statements, illustrating how disinformation exploits visual and linguistic stereotypes simultaneously.

Global Examples:

- Jacinda Ardern was described as "too emotional" despite her decisive leadership during crises.
- Hillary Clinton was framed as either "cold" or "unlikeable," exemplifying the double bind women leaders face.
- Sanna Marin was discredited through viral content sexualising her private life, overshadowing her political role.

Media Representation and Symbolic Erasure

Media plays a **central role** in amplifying these dynamics. Representation is not just about **visibility** but also about how people are shown and framed. Symbolic erasure occurs when women are either absent from coverage, reduced to stereotypes, or presented only through trivial or distorted lenses.

In Romania, women politicians are cited less often than men and are typically questioned on "soft" topics such as family and education rather than on policy or security (Centrul FILIA, 2023). This pattern reflects a broader global imbalance in media visibility. Across Europe, less than one in five expert sources cited in news stories are women, particularly in political, economic, and scientific domains, where male voices continue to dominate (European Parliament, 2018). In the United Kingdom, for instance, nearly 77% of quoted experts are men (King's College London, 2023), while global monitoring shows that women account for only about 18% of expert or commentator roles in mainstream media coverage (Global Media Monitoring Project, 2015; UNDP, 2023). Symbolic erasure thus works through both omission



Co-funded by
the European Union

and distortion, narrowing who is seen as an authority and reducing women's contributions to footnotes, curiosities, or aestheticised narratives. Angela Merkel's wardrobe or **Kamala Harris's** tone of voice often became subjects of public scrutiny, overshadowing substantive discussions of their policies and reinforcing the persistent gendered framing of expertise and leadership.

Visual media strengthens these effects. Men are often shown engaged in strategic decisions, while women are photographed in moments of emotional expression. Memes magnify such cues: in Romania, doctored images of Gabriela Firea and other women politicians portrayed them as incompetent or manipulative, spreading stereotypes under the guise of humour. Internationally, Hillary Clinton's emails became meme material not only for political critique but also to question her morality and emotional stability.

Gatekeeping and Agenda-Setting

Representation is shaped by editorial and institutional decisions. Research shows that **women's achievements in male-dominated fields receive less attention.** At the same time, stories highlighting their personal lives or controversies are amplified (Baker & McEnery, 2015). Disinformation campaigns exploit these editorial tendencies: they push stereotypes that resonate with existing biases while silencing or overshadowing narratives of competence and expertise.

Intersectionality in Representation

The effects of symbolic erasure and linguistic stereotyping are intensified for minority women. Roma women activists in Romania are not only portrayed as "too emotional" but also as threats to cultural norms (E-Romnja, 2023). Globally, women of colour in politics and journalism are targeted with a double burden of sexism and racism. These layered attacks illustrate how disinformation leverages multiple axes of vulnerability to exclude entire groups from public legitimacy.

Consequences for Public Discourse

The cumulative impact is profound. **Gendered language and symbolic erasure** reduce women's legitimacy, discourage their participation in public life, and distort democratic debate. When women's voices are consistently trivialised or silenced, audiences internalise narrow definitions of authority and expertise. Disinformation campaigns are not merely exploiting this weakness; they are actively reinforcing it, normalising discriminatory narratives as common sense.

Why Does This Matter?

Resisting gendered disinformation requires attention to both language and media structures. Journalists and editors must commit to equitable coverage. Audiences must cultivate critical awareness of framing and representation. Policy frameworks and advocacy organisations can challenge bias and demand accountability from platforms that profit from disinformation. By recognising how language and symbolic erasure work together, we can begin to dismantle the stereotypes that undermine democracy and inclusion.



Co-funded by
the European Union

Suggested Activities

1. **Critical reflection:** How does language shape perceptions of authority? Can you identify examples of intersectional attacks in Romanian or international media?
2. **Media analysis task:** Select a meme or article about a woman in public life. Identify linguistic and visual strategies used. Discuss how these framings shape audience perceptions.
3. **Comparative discussion:** Contrast Romanian and global cases. What cultural similarities and differences emerge in how women are represented?

6.3 Conclusion: Gendered Disinformation and Paths to Resistance

Introduction

Throughout this module, we have traced how gender operates as a weapon in the architecture of disinformation. We have examined how language frames women as irrational, emotional, or morally suspect; how media representation distorts visibility; how digital platforms amplify attacks; and how psychological biases and social dynamics make audiences particularly susceptible. Concrete Romanian and international cases brought these abstract mechanisms to life, showing the very real consequences that gendered disinformation has on women's legitimacy, safety, and participation in public life.

This final submodule draws the threads together. It offers a synthesis of the insights gained, reflects on the implications for democracy and public debate, and outlines strategies, both personal and systemic, to counter the chilling effects of disinformation.

Synthesis of Key Insights

One of the clearest lessons is that gendered disinformation does not occur in isolation. It emerges from overlapping layers: the symbolic and the digital, the psychological and the structural. At the linguistic level, labels and stereotypes continue to shape how women are perceived. Emotional tropes, too soft, too ambitious, too unstable, work precisely because they resonate with long-standing cultural scripts about gender roles. At the media level, women are often underrepresented or framed through appearance, personality, or morality, rather than competence or achievement.

Digital environments have intensified these dynamics. Algorithms prioritise sensationalism, memes condense stereotypes into easily shareable formats, and coordinated networks of trolls and bots give harassment the appearance of mass consensus. Psychological mechanisms, such as confirmation bias, further entrench these narratives: audiences are more likely to believe disinformation when it confirms what they already suspect about women in leadership.

Case studies from Romania demonstrate how these forces converge in practice. Laura Codruța Kövesi was persistently portrayed as arrogant or as under foreign control. At the



Co-funded by
the European Union

same time, Gabriela Firea and Viorica Dăncilă were mocked for their appearance, accents, or femininity rather than their political actions. Roma women activists faced intersectional attacks that combined sexism and racism, branding them as outsiders threatening cultural norms (E-Romnja, 2023). Globally, Hillary Clinton, Jacinda Ardern, and Sanna Marin were all subject to similar strategies, showing that while the details vary, the gendered logics of disinformation are remarkably universal.

Implications for Democracy and Public Life

The impact of gendered disinformation is not limited to the individuals targeted; it reverberates across entire societies. By undermining women's credibility and silencing diverse voices, disinformation distorts the conditions of public debate. The chilling effect is evident in the reluctance of women journalists, activists, or politicians to post freely online, to pursue controversial issues, or to remain visible in public spaces. When women are portrayed as exceptions or as unfit for leadership, audiences internalise these cues about who "belongs" in politics, the media, or civic leadership.

This has profound implications for democracy. Disinformation is not only about truth and falsehood; it is fundamentally about power. Gendered disinformation, in particular, reinforces patriarchal hierarchies by weaponising existing social biases and exploiting digital infrastructures. It tells women, implicitly and explicitly, that their authority is conditional, precarious, and always open to attack. The consequence is a narrowing of public life, with fewer voices heard, fewer perspectives represented, and accountability weakened.

Pathways to Resistance

Yet resistance is possible, and it is already underway. At the individual level, critical media literacy and awareness of cognitive biases can make audiences less vulnerable to manipulation. **Learning to question framing techniques**, identify emotional manipulation, and evaluate credibility across multiple sources equips citizens with tools to resist disinformation's appeal.

At the organisational level, NGOs and advocacy groups play a central role. In Romania, Centrul FILIA documents gender bias in the media. At the same time, E-Romnja focuses on the intersectional harassment of Roma women activists. Both organisations produce counter-narratives that reframe women's voices and contributions, ensuring they remain visible despite attempts to erase them. Globally, women leaders, journalists, and activists have built peer networks and alternative platforms, blogs, podcasts, and independent news sites that circumvent traditional gatekeepers and foster solidarity.

At the structural level, platforms and institutions must take responsibility. Social media companies need to address the algorithmic amplification of harmful content. At the same time, governments and regulators can push for stronger accountability mechanisms. Media



Co-funded by
the European Union

organisations, too, must commit to equitable representation, resisting the tendency to reproduce stereotypes or focus disproportionately on women's personal lives.

Resistance must also be intersectional. Strategies that work for white, urban, professional women may not address the compounded vulnerabilities of minority, rural, or LGBTQ+ women. Tailored approaches grounded in empathy, inclusivity, and collaboration are essential for building resilience across communities.

Conclusion

The story of gendered disinformation is, ultimately, the story of who gets to participate in public life. By understanding the mechanisms that make disinformation effective, from emotional biases to algorithmic amplification, we see clearly that these attacks are not accidental but systematic. They seek to discourage participation, delegitimise authority, and preserve existing hierarchies.

But as much as disinformation thrives on repetition and amplification, resistance thrives on solidarity and knowledge. By equipping individuals with critical skills, supporting organisations that defend targeted voices, and demanding structural accountability from platforms and media, we can shift the balance. The path to resilience is not easy, but it is possible.

Gendered disinformation tells us who we are not supposed to be. Resistance, in all its forms, is how women, minorities, and allies rewrite that script, claiming space, visibility, and authority in the public life of democratic societies.

Reflection and Critical Thinking

1. Critical reflection questions

- How do language, media representation, and digital dynamics intersect to amplify gendered disinformation?
- What strategies can be implemented at personal, organisational, and societal levels to resist gendered attacks?

2. Small group discussion

- Discuss one Romanian and one international case. How could interventions have been applied to reduce harm or reclaim visibility?

3. Media analysis task

- Choose a viral narrative or post targeting a woman in public life. Map how language, visuals, and digital amplification contribute to disinformation. Propose alternative framings that promote agency and inclusion.

6.4 References

Baker, P., & McEnery, T. (2015). *Corpus approaches to discourse: A critical review*. Routledge.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.



Co-funded by
the European Union

Centrul FILIA. (2023). Gender and media analysis report: Representation in Romanian media. [https:// www.centrulfilia.ro](https://www.centrulfilia.ro)

E-Romnja. (2023). Online harassment and disinformation targeting Roma women activists in Romania. Bucharest Press.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. [https:// doi.org/ 10.1145/ 2818717](https://doi.org/10.1145/2818717)

Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., & Moran, S. (2018). Falling for fake news: Investigating the consumption of news via social media. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–10. [https:// doi.org/ 10.1145/ 3173574.3173950](https://doi.org/10.1145/3173574.3173950)

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Lakoff, R. (1975). *Language and women's place*. Harper & Row.

Marincea, A. (2024). *Investigative journalism in the age of online harassment*. Bucharest Press.

Pennycook, G., & Rand, D. (2018). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. [https:// doi.org/ 10.1016/ j.cognition.2018.06.011](https://doi.org/10.1016/j.cognition.2018.06.011)

Spender, D. (1980). *Artificial language*. Routledge & Kegan Paul.

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. William Morrow.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. [https:// doi.org/ 10.1126/ science.aap9559](https://doi.org/10.1126/science.aap9559)



Co-funded by
the European Union

7. Visual (Dis)Information

Gendered Harms and Creative Resistance in the Age of Visual Disinformation

Written by Sebastian Arrosamena Mellgren, Traces&Dreams, 2025

7.1 Why Images Matter More Than You Might Think

Every day, we see thousands of images and videos online, from social media to news sites. These visuals may seem harmless, but they have a decisive influence on what we think, believe, and act upon. Sometimes images tell the truth, other times they mislead or harm. That's why learning to read visual content critically, visual literacy, is essential. This chapter will guide you step by step in engaging with visual information. It covers the importance of visuals in information and disinformation with a focus on gendered harms. The chapter will teach you how to stay critical, verify images, and deconstruct visual stories. Furthermore, it will explore tactics for building a fairer Internet and for ethical visual storytelling. Finally, all is put together into simple steps you can implement in your everyday life.

Visuals, such as photographs, memes, or videos, dominate how we consume information today. Visuals have the power to shape beliefs more powerfully than words alone. When scrolling through news or social media, images reach the brain faster than text and shape what we believe and act upon. Even if an image provides no objective evidence for a claim, we tend to think the claim more when it is paired with an image. For example, viewers were more likely to believe a headline about a celebrity being alive (or dead) when a photo of that celebrity was shown alongside the claim. This well-documented "truthiness effect" means that images often make information seem trustworthy just by association and familiarity (Bonn Institute, 2025). Our brains use visuals to fill gaps in understanding, which can rapidly change beliefs, and both educators and disinformers exploit this mechanism.

Images can provoke strong emotions. For instance, a photo of a smiling child or a scene of devastation can elicit sympathy, concern, or fear, which, in turn, can sway opinions and behaviours. One famous example shows a girl named Phan Thi Kim Phuc running naked after being hit by napalm in the Vietnam War (you can search for her name if you wish to see the image). This image had an essential role in shifting the public opinion about the war in the US and Europe, as it showed the horrors of the war. Emotional responses to visuals are immediate; they can bypass rational filters and push viewers to believe or act quickly, often without further questioning. This makes visual disinformation particularly important and dangerous. You can read more about the power of images at the Bonn Institute (2025) <https://www.bonn-institute.org/en/news/psychology-in-journalism-5>



Co-funded by
the European Union

7.2 Visuals are Not Neutral Reflections of Reality

It is important to remember that photographs and videos are not neutral reflections of reality; they are choices about what to show and what to hide. For example, a picture of a protest focused on angry crowds tells a different story from one showing peaceful signs. Both may be real, but each shapes public perception in distinct ways. This is called visual framing, and we will get back to it later.

Over time, repeated exposure to particular images forms what feels "normal" or "true". For instance, Instagram has widely affected beauty ideals worldwide. The Cultivation Theory explains this process. Cultivation theory holds that long-term exposure to repeated imagery constructs a "common-sense map" of social reality. Stereotypes and ideals in social media, news, entertainment, and advertising are not created overnight; they are sold to us image by image, solidifying social norms and expectations (Croteau & Hoynes, 2023). This means that gender identities are constructed and maintained through the contemporary media landscape (Gauntlett, 2023).

The choice of how to frame an image and which images to show has always been political. However, today's media environment intensifies this dynamic, as images can be easily edited or entirely fabricated. Cheapfakes (simple edits or misattributions) and deepfakes (AI-generated synthetic photos or videos) make it increasingly difficult to sort fact from fiction. The consequence is not only that more visuals are faked, but also that authentic images can be accused of being fake. This creates what is called the "liar's dividend", where wrongdoers can claim that any evidence used against them is faked. This means that we now live in a time where the foundations of evidence-based debate are being challenged (McIntyre, 2018).

Next time a photo or video grabs your attention, pause. Ask: "Am I reacting or thinking critically? Who created and shared this, and why?" Recognising that images influence our feelings is the first step in developing visual literacy, the capacity to question, contextualise, and evaluate the purpose, framing, and reality of images and videos. We will continue developing these skills in the textbook.

7.3 How Images Can Mislead and Manipulate

Visual disinformation thrives not just because images are powerful, but also because today's media landscape has made them easy to twist, stage, and misuse. This submodule explores three primary tactics used in visual disinformation: staging and misrepresentation, meme-ification, and outright fakery. Note that other disinformation tactics, such as algorithmic amplification, suppression, or censorship, are not covered here but elsewhere in the textbook.

Staging, Editing, and Misrepresentation

Images shape who is seen, how they're portrayed, and who is left invisible. Decades of research reveal that mainstream media often under-represent women and minority groups,



Co-funded by
the European Union

or show them in limited, stereotypical ways (Lind, 2017; Entman, 1994; Burke et al., 2022). For example, black men are more often shown as criminals and Latin American women as "sexualised". Image "staging" means deliberately arranging, cropping, or selecting visuals to steer a viewer's interpretation.

As we said before, repeated exposure to visuals shapes audience expectations and beliefs about how groups are and act. Games and films with hyper-sexualised female avatars, makeover shows tying women's success to their looks, as well as male "action heroes", all feed into narrow scripts about gender roles. These scripts are normalising some identities while making others less thinkable. David Gauntlett's work traces how popular media channels us towards some versions of femininity and masculinity. You can read more about it here: https://www.redbrick.me/wp-content/uploads/2025/04/10.4324_9780203930014_previewpdf.pdf.

Memes and Visual Persuasion

Memes (combinations of images and text) are a modern force for persuasion that often simplify complex issues into catchy, emotionally resonant visuals, making them highly shareable (Alafnan, 2025). Memes can draw people into political conversations or, worryingly, deepen polarisation and spread stereotypes. In extreme cases, coordinated "memetic warfare" campaigns deliberately use memes to dehumanise or radicalise targets (Mishra & Karumbaya, 2024). For example, far-right groups have created racist, AI-generated memes designed to inject hate into mainstream platforms. Burke et al. (2022) analyse how right-wing populist actors mobilise memes, video clips, and livestreams that blend racism, misogyny, and nationalism. This visual content depicts feminists or migrants as 'threats' without using facts or rational arguments. However, memes can also be used to spread information in an accessible way or to counter harmful narratives. We will get back to this later in the chapter.

Questions:

- Is it helpful to respond to hatred, such as the examples above, with 'counter-mems', or is it simply a continuation of the memetic warfare?
- If it is justified to respond, what sort of responses are ok and what would not be ethically justifiable?
- And, what sort of responses are efficient? Is humour the best strategy?

Cheapfakes and Deepfakes

Although this may shift rapidly in the coming years, most visual disinformation still uses low-tech methods, so-called "cheapfakes". Cheapfakes include old photos passed off as current, basic editing (cropping, slowing down video, altering colour, etc.), or changing captions. This simplicity is part of their danger: we see, believe, and share before critical filters kick in. An example from 2019 was the infamous video of US politician Nancy Pelosi, which was slowed down to make her appear drunk. Meanwhile, advances in generative AI have enabled anyone to fabricate synthetic images and videos that appear indistinguishable from reality. By 2025, these tools will be available to most Internet users, and "deepfakes" will become increasingly common. Recent studies have shown that



Co-funded by
the European Union

viewers can no longer consistently distinguish between authentic photos or videos and AI-generated versions (Roca et al., 2025). In the near future, most visuals found online may be AI-generated, or at least we will not be able to tell.

7.4 Verifying Visuals: Tools and Their Limits

In an environment full of deepfakes, viral hoaxes, and recycled images, verifying photos and videos is more important than ever. This submodule introduces practical verification tools and reflects on the challenges of assessing authenticity in the age of synthetic media. Please remember that, as AI rapidly evolves, any technical tool that is useful today might be useless tomorrow. Therefore, the most effective way to protect yourself from visual disinformation is to combine critical thinking with verification techniques.

The Verification Toolbox

A growing arsenal of investigative tools now allows everyday users to scrutinise the images and videos that circulate online.

- **Reverse Image Search:** Plugging an image into services like Google Lens, Bing, or TinEye can reveal where else it has appeared online. This is particularly useful to indicate when a photo is "recycled" or posted out of context. Try searching for parts of an image separately, such as faces or logos, for better results.
- **Video Frame Analysis (Keyframes):** Tools such as InVID-WeVerify let users split a video into still images, enabling them to check whether parts have circulated before or been altered.
- **Metadata:** Digital files often include metadata, such as the date and location where a photo was taken. Analytical plugins can expose if a photo's timestamp or location supports the story (or not). However, social media platforms often erase metadata from images, so it may not be available. The person who uploaded the image may also have deliberately erased or faked the metadata.
- **Forensic filters** are digital tools that examine minute details, such as shadows, lighting, and pixels we can't see with the naked eye. They help find signs of editing or AI manipulation in images.
- **Geolocation** is the process of determining where a photo or video was taken by identifying landmarks, street signs, or landscape features. Sometimes you can even use the sun's position, combined with the time and date, to estimate where a photo was taken. You can use Google Maps or Google Earth to verify what the place actually looks like. There are also new AI tools that can do this for you, though their reliability may vary.
- **Chronolocation** estimates: when an image was taken by looking at clues like shadows, weather, or seasons. For example, if the trees don't match the claimed season, that's a red flag.

Remember, no single tool is perfect. If you need to be sure, use as many methods as possible to verify information. You might still not be 100% sure, and that in itself is a



Co-funded by
the European Union

valuable realisation. If you want to learn more about verification tools, check out "Exposing the Invisible - The Kit" at: <https://kit.exposingtheinvisible.org/en/>

Verification tools to check out:

- **Forensically:** A free, web-based suite for digital image forensics that functions as a "magnifying glass" for images. It includes clone detection (highlighting copied regions), error level analysis (ELA), metadata extraction, noise analysis, magnification with contrast enhancement, and C2PA content authenticity verification. Particularly useful for identifying manipulation via cloning tools, though it requires experience to interpret results correctly and should not be relied upon as the sole verification method.
- **FotoForensics:** An online platform offering error level analysis through heat-map-style visualisations that highlight areas of different compression levels in JPEG images. The colour-coded output helps identify potentially manipulated regions where the compression signature differs from the rest of the image, making alterations more visible.
- **InVID-WeVerify:** A browser extension and platform designed for journalists and fact-checkers to verify images and videos. It combines multiple verification techniques, including reverse image search, metadata analysis, forensic filters, and fragmentation detection, streamlining the verification workflow for time-sensitive investigations.
- **ExifTool:** A robust command-line application that reads, writes, and edits metadata across nearly any file format. It extracts EXIF data, including camera settings, GPS coordinates, timestamps, software used, and device information, crucial for establishing provenance and detecting inconsistencies in digital evidence chains.
- **Google Lens:** Google's visual search tool that performs reverse image searches, identifies objects, landmarks, text, and products within images. Essential for verification work to trace image origins, find earlier instances online, and identify locations or subjects.
- **SunCalc:** A web-based calculator that shows sun position, sunlight phases, and shadow angles for any location and date. Invaluable for verifying whether shadows and lighting in photographs match the claimed time and location, helping detect temporal inconsistencies.
- **ShadeMap:** An interactive mapping tool visualising sunlight and shadow patterns across urban landscapes throughout the day and year. Used to verify photographic authenticity by cross-referencing shadow directions and lengths with specific times and locations, particularly useful for architectural and outdoor scene analysis.

Questions:

- How much effort can you reasonably put into finding out if something is real or not?
- What are the consequences for society if trust in visuals collapses entirely? Who gains, and who loses from such a situation?



7.5 Deconstructing Visual Narratives

As we have emphasised, even if a visual is real, it might still try to trick you or convey a specific idea. That's why it's essential to learn to deconstruct visual narratives.

Deconstructing means looking beyond what you first see to understand the story an image or video is telling.

Every picture and video tells a story. It can be a selfie, a news report, or a meme. These stories shape how we think and feel about people and issues. For example, a photo of a leader surrounded by happy students may tell a story of success and hope. But a photo of that same leader alone in an empty room might make them look lonely or unpopular. Both images can be real, but they give very different messages.

Deconstruction Toolkit

This toolbox is grounded in a blend of critical media theory, semiotics, and intersectionality, drawing on cultural studies, media literacy, and visual sociology. You do not need to use every tool every time you analyse an image; instead, you can choose the ones that best fit your situation.

- **Source and intention tracking:** Use verification techniques to find the source, creation date, and creator. What ideological affinities does the creator and distributor have? For example, is the newspaper that published the article conservative or progressive?
- **Surface Scan:** Look carefully at everything in the picture. Who or what is shown? What symbols, clothes, gestures, or places do you see?
- **Image arrangement:** Notice where the main person or object is placed and what colours are used in the image. Is the camera looking up or down? These choices affect how strong, weak, or trustworthy the subject appears. Generally, a person shot from below looks powerful, while soft colours and a distant gaze may signal vulnerability. Are the people in the picture framed as victims, heroes, villains or something else?
- **Analyse story symbolism:** Who and what appear? What symbols, gestures, or settings stand out? For example, a politician in a factory setting with rolled-up sleeves may want to communicate "man of the people", "hard working", and "relatable". Ask, what story is being told?
- **Emotional trigger:** What emotions is the image intended to trigger? Whose bodies or issues are linked to these feelings?
- **Intersectional audit:** Ask, which identities are present, missing, or stereotyped? How do race, gender, class, and sexuality intersect in the image?
- **Storyline switch:** Ask: What other stories could be told using a similar image? Try swapping hero/ victim/ villain roles, or reframing the emotional signals. This trains "narrative flexibility" and helps spot constructed meaning.
- **Cultural Lens Check:** How would this image be understood in a different location, historical period, or by a different group? This tool exposes ethnocentrism or time-



bound assumptions. If the message only works for a specific group, you can assume that group is the intended target.

- **Counter-narrative:** If you wanted to communicate the opposite of the intended story, how would you change the image?
- **Personal Impact Reflection:** After analysis, reflect on how your opinions or feelings were affected by the visual. What did the image initially make you feel, and how are you feeling now?

Combining verification with deconstruction

The best way to understand images is to use both verification and deconstruction. Sometimes people think that "if an image is real, it shows the truth" or that "deconstruction is enough since truth does not matter." Both of these positions are simplifications that can leave us open to new ways of being tricked by visuals. To really protect ourselves from disinformation, we need to use two skills together.

7.6 Tactics for a Fairer Internet

This submodule explores practical ways to resist, challenge, and change harmful visual narratives online. By building resilience, working together, and demanding responsibility, individuals and communities can protect themselves and help create fairer digital spaces. This chapter is intended for people who wish to work towards a more inclusive Internet. For information about what to do in case of ongoing hate or abuse, see the book's pre-words.

Developing Visual Counter Narratives

Counter-narratives are stories, images, or messages created to challenge and oppose harmful or false ideas. They can change how people think by offering new, positive, or truthful perspectives rather than hurtful stereotypes or lies. In the digital world, counter-narratives often use the same tools as harmful messages, such as memes, social media posts, and videos.

To develop a counter-narrative:

1. Start by understanding the harmful message or stereotype you want to challenge. Who does it affect? What lies or fears does it spread?
2. Do you belong to the affected group? If not, ask members of that group how they would like to be supported.
3. Find out what others are already doing about the problem and potentially join existing groups.
4. Use images, stories, or humour to offer your view. This can mean sharing real experiences, making jokes to expose absurd ideas, or showing pride and strength. Be careful not to reinforce stereotypes in doing so.
5. Share your messages where they will reach the affected community or a broad audience of people who can agree with you. It is generally good to try to shift the



Co-funded by
the European Union

opinions of the “silent majority”, not try to engage with the people who are actively against you.

6. Invite others to join, share their stories, or remix the messages. Collective voices make counter-narratives stronger.

Examples of Feminist Counter Narratives

- The satirical Xaccount @NoToFeminism
- Feminists have co-opted the hashtag #FeministsAreUgly to post selfies, critique unfair beauty standards, and mock sexist stereotypes.
- Feminist memes: <https://www.boredpanda.com/feminist-memes/>
- You can make your own memes using Imgflip or similar pages: <https://imgflip.com/>.

Demand Platform Accountability

Calling out harmful content and supporting each other online is very important. But to create lasting change, we also need to fix the broader system that allows such content to spread. That means holding social media platforms responsible for how they manage content and protect their users.

Under the EU’s Digital Services Act (DSA), which started applying to major platforms in 2023 and more broadly in 2024, platforms must follow strict new rules. They have to take more decisive action to remove illegal or harmful content, including image-based abuse and hate speech. They must provide easy ways for users to report harmful content and explain clearly how they handle these reports.

The DSA also requires platforms to be transparent about their content moderation. This includes publishing regular reports on how much harmful content they remove and how they manage risks from algorithms that might spread toxic material. Huge platforms such as YouTube, Instagram, and TikTok must assess and reduce systemic risks, meaning they must actively work to prevent their services from harming users or society.

In 2023, young people’s advocacy in the EU helped update the Digital Services Act with stricter rules to fight image-based abuse. This shows that when we work together, we can push for better laws and hold platforms accountable to create safer, fairer online spaces for everyone.

Reflection questions

- Which examples of creative resistance resonate the most with you?
- Which laws would you like to see to protect people online?

7.7 Ethical use of Visuals

Ethical use of images means carefully considering consent, trauma, respect for those depicted, copyright, and the well-being of everyone involved. In this submodule, we look into key issues related to the ethical sharing of visuals. It may help you decide what content to share and what not to.



Co-funded by
the European Union

When to share shocking images

Images or videos showing violence, abuse, disasters, or tragedies can be important news. Still, they can also profoundly affect people's emotions and mental health. Seeing such graphic content repeatedly is linked to stress, anxiety, and even trauma symptoms in viewers.

Journalists and educators often face difficult choices:

Sometimes, powerful images, like the famous open-casket photo of Emmett Till (a young black boy lynched in the USA in the 1950s), have awakened public awareness or forced social change. On the other hand, shocking images can re-traumatise victims, traumatise viewers and/or disrespect the people depicted or their families.

Best practice involves adding content warnings before showing graphic images, blurring disturbing details, and always asking: What is the purpose of showing this? If the image showed someone you love, would you want it widely shared?

Consent and Dignity

Respecting the dignity and choices of those depicted in images is a core ethical principle, especially for victims of violence or disasters.

Consent matters: Always try to get permission from people in the photo or their families. When this isn't possible, use extra caution: avoid close-ups or sharing unnecessary details.

Protect the vulnerable: Hide identities or blur faces when sharing images of people affected by sexual violence, hate crimes, or sensitive situations.

Avoid exploitation: Groups like Médecins Sans Frontières (MSF) insist images should show people with power over their own story, avoiding "poverty porn" that exploits suffering. Try comparing positive examples (where families or subjects consented) with cases where photos went viral without their consent, causing harm. Intend not to harm.

Copyright and Legal Boundaries

Ethical visual use means respecting copyrights and intellectual property laws: In the EU, photographers and creators have strong legal rights over their work. Only use images if you have permission, a proper license, or if they are openly licensed (Creative Commons) or in the public domain. News and education have some exceptions, but sharing images for disinformation is legally risky. When sharing user-made content (like tweets), always credit the creator and, if possible, ask for permission. Model good practice in your work by constantly checking image sources, using licensed images, and teaching others how to do the same.

Reflection Questions

How do you balance the right of the public to know with the dignity of those shown?
What differences do you notice in ethical rules across countries or cultures?



Co-funded by
the European Union

What new questions do AI-generated images or deepfakes raise for you, educators, or journalists?

7.8 Putting it All Together

You've learned a great deal in this chapter! Now, it's time to bring everything together and see how to use visual literacy skills in your daily life. These skills will help you protect yourself, support others, and contribute towards building a fairer digital world.

What You've Learned

Visuals are powerful. They influence our beliefs and feelings faster than words. Visual disinformation is becoming an increasing problem. It deceives us through staging, cropping, editing, and even deepfakes. Certain groups, such as women, LGBTQ+ individuals, and racial minorities, often experience more harm online. Tools like reverse image search and metadata analysis help verify the authenticity of images. Analysing images reveals hidden stories, biases, and power dynamics behind pictures. We can learn to both confirm and examine visuals. We can work towards a fairer Internet by supporting each other, sharing or creating counter-narratives, and advocating for better laws and regulations. Avoid commenting on fake or harmful content; instead, report it. It is essential to respect consent, dignity, and copyright when sharing visuals.

How to Use These Skills in Everyday Life

Try these steps when you encounter an image online:

1. **Pause Before Sharing:** Take a moment. Ask yourself, where did this image come from? Is it genuine? What's the story behind it?
2. **Question the Emotions:** Notice what feelings the image tries to evoke, fear, anger, sympathy, and consider why.
3. **Check Multiple Sources:** See if trusted sources also share this story or image.
4. **Support Inclusive Visuals:** Share images that showcase diverse voices and challenge stereotypes.
5. **Be Kind and Responsible:** Consider who might be affected before posting or commenting.
6. **By practising this, you help slow the spread of false information and foster healthier online spaces.** You do not need to be an expert or well-known to make a difference. Small actions accumulate. Teach friends how to verify images or invite them to read this guide.

7.9 References

Aleta Valente. (n.d.). In Wikipedia. Retrieved 5 November 2025.

Croteau, D. R., Hoynes, W. D., & Childress, C. (2021). *Media/ society: Technology, industries, content, and users* (7th ed.). SAGE.

Digital Services Act. (n.d.). In Wikipedia. Retrieved 5 November 2025.



Co-funded by
the European Union

Gauntlett, D. (2023). *Creativity: Seven keys to unlock your creative self*. Polity.

Khosravi Ooryad, S. (2024). Memeing back at misogyny: Emerging meme-feminism, visual tactics, and aesthetic world-building on Iranian social media. *Feminist Media Studies*, 24(5).

Lawrence, E., & Ringrose, J. (2018). @Notofeminism, #Feministsareugly, and misandry memes: How social media feminist humour is calling out antifeminism. In J. Keller & M. E. Ryan (Eds.), *Emergent feminisms: Complicating a postfeminist media culture* (pp. 211–232). Routledge.

McIntyre, L. (2018). *Post-truth*. MIT Press.

Mina, A. X. (2019). *Memes to movements: How the world's most viral media is changing social protest and power*. Beacon Press.

NotAllMen. (n.d.). In *Wikipedia*. Retrieved 5 November 2025.

Se vuoi, posso uniformare tutto in APA 7 rigoroso, oppure adattarlo a un deliverable UE



Co-funded by
the European Union

8. AI and the Media System

Written by Amalia Ranieri for Cooperativa Sociale "Sinergie", 2025

8.1 Introduction

Imagine scrolling through your phone late at night. A video surfaces: the president of a country resigning in anger; another clip follows, showing a beloved actor confessing to crimes. A voice message pings in your inbox, apparently from your mayor, urging you not to vote tomorrow. All of it feels urgent, believable, and real, but none of it is. Welcome to the age of synthetic truth, where the line between fact and fabrication dissolves faster than we can process it.

The history of disinformation is long. Roman emperors manipulated the *acta diurna*; Cold War propaganda deployed forged documents and planted news. But what we face today is unprecedented in scale and speed. Machines now write, paint, and speak for us, producing plausible realities at an industrial pace. What once required studios, editors, and specialists can now be achieved by anyone with an internet connection. Yet this is not merely a story of technological novelty; it is also a story of perception, power, and fragility. Truth, as philosophers from Hannah Arendt to Michel Foucault have argued, has never been a fixed object; it has always been contested, mediated, and fragile. What AI does is turn that fragility into a daily condition of life. We are bombarded with streams of information that blur the line between authentic and artificial, reliable and manipulated. And while the technologies are dazzling, the real stakes are human: democracy, trust, and the ways we understand one another.

This chapter is written as a journey through this new terrain. Each section is a stop along the way:

- We begin with a publishing case study, which will allow us to start a critical reflection on the use of artificial intelligence.
- We then move to **Generative AI**, exploring how machines fabricate coherent but unstable truths, raising questions about hallucinations, corporate control, and the politics of knowledge.
- From there, we enter the world of **Disinformation 2.0**, where deepfakes and synthetic voices destabilise elections, weaponise gendered harassment, and reshape security concerns.
- We then step inside the newsroom to examine how AI unsettles **journalism**, challenging traditions of authorship, verification, and trust.
- The next stop takes us to the Global South, where **algorithmic colonialism** reveals how inequalities of language, labour, and sovereignty shape AI's hidden hierarchies.
- Finally, we reflect on **creativity and authorship**, asking what becomes of art and journalism when machines participate in cultural production.

Throughout, we weave in stories, scandals, and case studies: from a fake resignation video in Moldova to lawsuits by artists against AI companies, from Kenyan content moderators filtering toxic data for Silicon Valley to the grassroots projects fighting to preserve endangered languages in the digital age.



Co-funded by
the European Union

8.2 Hypnocracy

In January 2025, a book appeared on Italian shelves that promised to change the way we think about politics, media, and the fragile line between truth and illusion. Its author was presented as Jianwei Xun, a Chinese philosopher, and the book carried the intriguing title *Hypnocracy: Trump, Musk, and the New Architecture of Reality*. Reviews poured in across cultural outlets – *Il Fatto Quotidiano*, *Doppiozero*, *Le Grand Continent* in France – praising its sharp analysis. The text dissected how figures like Donald Trump and Elon Musk act not merely as politicians or entrepreneurs, but as world-makers, architects of narratives that capture imaginations on a global scale.

Why did so many intelligent readers believe in him? Perhaps because the figure of the mysterious Eastern sage fit our expectations, or because the rhythm of reviews, headlines, and conversations produced the feeling of credibility. But the more profound lesson was unsettling: truth had not been discovered or verified; it had been performed into existence. This is the essence of what Hypnocracy named: a new form of power based less on force and more on trance. In this landscape, leaders like Donald Trump or Elon Musk operate not primarily as politicians or entrepreneurs, but as hypnotists of collective attention. Their tools are not arguments or policies, but spectacle, repetition, and suggestion. A late-night tweet, a staged controversy, a viral meme: each acts like a spark in the theatre of attention, setting off cycles of outrage and fascination that keep them at the centre of public imagination. The metaphor of hypnosis is apt: a hypnotist does not impose reality; they guide attention until the subject begins to accept suggestions as self-evident. Similarly, the hypnocratic system does not silence dissent; it absorbs it. Satirical memes mocking Musk's eccentricity or Trump's blunders may seem oppositional, yet they amplify visibility, feeding the trance. In hypocrisy, ridicule and resistance are not threats but the very nutrients for performance.

This is not entirely unprecedented. Guy Debord, in *The Society of the Spectacle* (1967), argued that modern life was increasingly mediated by images that transform social relations into appearances. But what Hypnocracy exposed is the next stage: one in which algorithmic virality turns spectacle into rhythm, a constant pulse that guides collective emotion. Information overload does not liberate us with options; it exhausts us into compliance. The scroll never ends, and with it, the trance deepens.

Digital platforms are the stage and the machinery of this hypnosis. Trending hashtags, viral videos, and algorithmic feeds are all designed to maximise engagement. A video's popularity becomes evidence of its significance; a story repeated enough times feels real. This "architecture of influence" blurs the line between voluntary attention and algorithmic capture. We choose what we see, but the trance is orchestrated behind the scenes. The hoax of Jianwei Xun becomes, in this sense, more than a literary prank. It is a mirror, showing how much of our reality is already constructed in this hypnocratic way: trust built not on verification but on circulation, credibility performed through repetition rather than grounded in evidence. The disappointment readers felt upon discovering the author's nonexistence was, in fact, the same disappointment that haunts much of our media life:



the discovery that what we believed was real may have been no more than a well-staged suggestion.

How, then, do we resist? Hypnocracy offered no heroic awakening, no promise of stepping outside the trance once and for all. Instead, it suggested cultivating a fragile margin of awareness within it: noticing the rhythms, questioning the repetitions, pausing when the scroll demands constant movement. Resistance is not a total escape but a practice of micro-attention, small acts of slowing down that disrupt the spell. For us, as readers, this is a critical lesson. The battle against disinformation is not only about fact-checking or detecting deepfakes; it is also about recognising the hypnotic architectures that shape perception itself.

8.3 The Fragility of Truth

When you ask ChatGPT the question "What is truth?", the reply arrives instantly, dressed in authority: it lists Aristotle's correspondence theory, William James's pragmatism, and postmodern critiques, before closing with a balanced reflection: truth has many dimensions, no single definition. The words flow smoothly, persuasive in tone, but beneath the surface, something strange is happening. The machine is not "thinking" about philosophy. It predicts the most probable sequence of words based on patterns it has absorbed from billions of documents. What looks like knowledge is, at its core, statistical guesswork.

This is the paradox of generative AI. It does not lie in the human sense, for it has no intention to deceive; nor does it tell the truth, since it has no access to reality. It produces plausibility. The result is a new epistemic fragility: the feeling of coherence without the substance of verification (Floridi, 2020).

Developers call the most notorious symptom of this fragility hallucination. Ask a system to summarise an obscure court case or produce a medical reference, and it may fabricate details with total confidence. In 2023, lawyers in New York submitted a legal brief generated by ChatGPT, only to discover that the cited cases were entirely invented (Katz et al., 2023). In scientific settings, chatbots have produced bibliographies containing articles that never existed (Lee et al., 2023). These are not rare bugs: studies show hallucination rates ranging from 15 to 27 per cent depending on task complexity (Maynez et al., 2020; OpenAI, 2023). The problem is not only technical but perceptual. Human beings tend to trust fluent language, a phenomenon that psychologists call the illusory truth effect: the more familiar or smooth a statement sounds, the more likely we are to believe it (Fazio et al., 2015). Generative AI industrialises this effect. It can produce endless streams of fluent but unreliable text, saturating the information environment with uncertainty.

To make matters more complex, AI systems are not neutral mirrors; they are filtered through the corporations that design them. Google's Gemini avoids sensitive U.S. political content; Chinese models refuse to discuss dissent; OpenAI has imposed limits on copyrighted material and specific geopolitical terms. These guardrails are rarely transparent, yet they shape what the machine can and cannot say. The apparent voice of



Co-funded by
the European Union

the chatbot is, in fact, the synthesis of statistical prediction and corporate discretion. As Foucault would remind us, truth is always embedded in regimes of power (Foucault, 1977). This also means that many perspectives, languages, and ways of knowing are missing from these models. Some have never been included in the data; others may have been filtered out on purpose. In practice, this produces what Abeba Birhane (2021) calls epistemic coloniality: a situation in which certain forms of knowledge (usually those from Western, English-speaking, or dominant cultures) are given priority, while others remain invisible.

Defining AI Hallucinations

An AI hallucination occurs when an artificial intelligence system, especially a generative model, produces information that is **plausible but false or unsupported by fundamental data**. These hallucinations happen because the model generates outputs based on patterns in training data rather than factual verification. As a result, the AI may invent facts, sources, or images that sound coherent but lack grounding in reality.

Between Misinformation and Disinformation

Hallucinations also blur categories we once treated as distinct. Misinformation refers to error without intent; disinformation to deliberate falsehood. But when a chatbot fabricates a convincing story, where does it fall? It lacks intent, but its impact can be identical to that of disinformation: erosion of trust, amplification of conspiracy theories, confusion about what is real. In this sense, hallucinations function like "ambient disinformation," diffused not by malice but by design. Hannah Arendt (1971) warned that when citizens can no longer distinguish between fact and fiction, they retreat into cynicism, a condition ripe for authoritarianism. Generative AI accelerates this danger by overwhelming verification systems with streams of fluent fabrications. Yet it also forces us to confront an uncomfortable insight: truth has always been fragile. It has always depended on practices of verification, trust, and collective dialogue.

8.4 Disinformation 2.0

In late September 2023, voters in Slovakia were preparing to cast their ballots in a tightly contested parliamentary election. Just days before polls opened, a leaked audio recording began to circulate online. It appeared to capture a liberal party leader plotting to rig the vote and, in a surreal twist, discussing plans to raise the price of beer. The clip was fake, generated with artificial intelligence, but it spread faster than any correction. By the time fact-checkers intervened, the damage was done: people were already angry, confused, and suspicious. This is the terrain of what many call "**Disinformation 2.0**": a new stage in the manipulation of truth, powered by generative AI. Unlike past propaganda, which required expertise, resources, and coordinated infrastructures, AI-driven disinformation is cheap, fast, and scalable. What once took a team of forgers or an editing studio can now be produced by a teenager with a laptop.

The manipulation of information has always been part of political life: Roman emperors used the *acta diurna* to bolster imperial narratives. At the same time, Cold War intelligence



Co-funded by
the European Union

agencies planted stories and forged documents to discredit rivals (Rid, 2020). But the digital environment changes the equation: speed, scale, and automation. Generative AI systems such as GPT models, MidJourney, and ElevenLabs can produce credible text, images, and voices in minutes (Wardle & Derakhshan, 2017). The barriers to entry have collapsed.

Deepfakes and Gendered Disinformation

Deepfakes have become the most visible symbol of this change: in Moldova's 2023 local elections, a fabricated video showed President Maia Sandu resigning and endorsing a pro-Russian party; in the United States, voters in New Hampshire received robocalls in January 2024 featuring a cloned voice of President Joe Biden, telling them not to vote in a primary. The goal was simple: not to change votes, but to suppress turnout by undermining trust in the process. These examples highlight a disturbing reality: disinformation now targets not only opinions but the infrastructure of democracy itself. Ballot boxes remain untouched; it is the perception that is hacked.

The new tools also intensify older forms of harassment, particularly against women. In Bangladesh, a female opposition MP was attacked with a deepfake that placed her face onto a bikini-clad body. In a conservative society, this was enough to tarnish her reputation, illustrating what researchers describe as gendered disinformation: falsehoods weaponised to intimidate women and silence their political voices (Hosseini, 2023). The effect is not only reputational harm but the erosion of democratic participation.

If deepfakes represent intentional deception, large language models add another layer: hallucinations. These are confident but false outputs generated without malicious intent (Ji et al., 2023). Yet the impact can be equally corrosive. A fabricated legal precedent, an invented medical citation, or a false biographical detail, each may spread unchecked once repeated. The line between misinformation and disinformation blurs, creating a constant stream of what Luciano Floridi (2020) calls "epistemic opacity", that is, a situation where it becomes difficult to tell what is true and what is not, even when information sounds perfectly convincing. The fusion of deliberate deepfakes and unintentional hallucinations has led policymakers to treat disinformation as more than a communication issue. Increasingly, it is framed as a cybersecurity threat. Just as ransomware exploits vulnerabilities in digital infrastructure, disinformation exploits vulnerabilities in information systems: social media algorithms, trust in news, and the virality of messaging apps (Bradshaw & Howard, 2019). The consequences are wide-ranging. In 2017, French construction giant Vinci lost billions in valuation after a fake press release alleged accounting fraud. In 2013, a hacked Associated Press tweet claiming explosions at the White House erased \$136 billion in U.S. markets within minutes (DHS, 2014). With generative AI, such forgeries can be created faster, spread wider, and look more credible than ever.

National security is equally at stake. During moments of crisis (such as terrorist attacks, natural disasters, pandemics), synthetic disinformation can paralyse response systems, sow panic, and fracture public trust. Democracies, already fragile, may find themselves destabilised not by bombs or tanks but by pixels and sound waves.



Co-funded by
the European Union

8.5 AI, Journalism, and News Production

Newsrooms have always been shaped by technology: the telegraph collapsed distances, the radio brought voices into kitchens, and television made politics a matter of images as much as words. Each innovation changed not only how news travelled, but also how journalists imagined their work. Artificial intelligence continues this story, but with a twist because, unlike earlier tools, AI does not simply transmit or store information. It can generate it. And that changes everything.

The first experiments were modest. In the 2010s, outlets such as the Associated Press began using algorithms to draft thousands of quarterly earnings reports. The technology came from Automated Insights, a company specialising in natural language generation. Using structured data – stock prices, profit margins, batting averages – machines could produce formulaic stories in seconds. Journalists initially feared layoffs, but AP argued the opposite: automation freed reporters from repetitive drudgery, allowing them to pursue investigations and analysis (Marconi & Siegman, 2017). The Washington Post followed with Heliograf, an AI system that covered the 2016 U.S. elections by generating hundreds of short updates. These early systems were essentially advanced template engines: fast, accurate, but devoid of nuance. Few saw them as threats to the core of journalism. Then came GPT-2, GPT-3, and finally ChatGPT. Suddenly, machines could produce not only routine earnings summaries but also op-eds, profiles, and even features. By 2023, audiences were reading AI-generated travel blogs, sports commentaries, and political analyses, sometimes without realising it. What was once peripheral now presses at the heart of the profession.

Journalism was already fragile: advertising revenues were shrinking, local outlets were closing, and labour was precarious. AI entered this precarious world as both promise and threat. For managers, it was an opportunity to cut costs. For unions, it was a looming danger. The National Union of Journalists in the UK warned in 2023 that unregulated AI use risked lowering wages, eroding editorial standards, and turning reporters into "curators of machine output" rather than investigators (NUJ, 2023). And yet, new niches emerged: data journalism, algorithmic accountability reporting, and AI literacy became prized skills. The BBC experimented with AI-powered translation to broaden its linguistic reach. The New York Times invested in AI to scan datasets for investigative leads. For some journalists, AI became a partner, extending rather than diminishing their craft. Still, for others, it was a reminder of how precarious their profession had become.

At the core lies the question of authorship. If an article is partly written by an algorithm, who is the author? Should the AI be credited, or is authorship inseparable from human accountability? Most scholars agree that only humans can bear responsibility (Diakopoulos, 2019). Yet practices vary. The Guardian published an opinion piece explicitly labelled as written by GPT-3, inviting readers to reflect on the experiment. Smaller local outlets in the United States quietly integrated AI without disclosure. Transparency remains inconsistent, and trust suffers as a result. Surveys confirm this ambivalence. A Knight Foundation poll in 2023 found that audiences accepted AI for routine tasks such as data summaries but strongly opposed it for investigative or opinion writing. The message was clear: readers



Co-funded by
the European Union

value human judgment when values and interpretation are at stake. Journalism is not only about relaying facts; it is about authority, credibility, and trust.

Verification, long the backbone of journalism, faces a new challenge. Generative AI is prone to hallucinations, fabricated quotes, nonexistent sources, and invented statistics (Ji et al., 2023). A fluent but false line in an AI-written article may be harder to detect than a clumsy template error. Fact-checkers now spend more time cross-verifying not only human sources but also machine-generated text. Ironically, AI does not eliminate labour but redistributes it, from writing to verification. UNESCO (2023) has warned that fact-checking will become more burdensome and more urgent in an era of synthetic content.

Barbie Zelizer (2019) reminds us that journalism is not only about information but about authority: who has the right to narrate reality. If audiences begin to suspect that news articles are machine-written without oversight, trust may collapse. Yet there is also an opportunity. If readers become accustomed to asking, "Did a machine write this?", they may also learn to question the credibility of human writing. Scepticism could strengthen democratic vigilance, provided it does not curdle into cynicism, where all claims are dismissed as equally unreliable. Journalism stands at a crossroads. Rejecting AI altogether risks irrelevance in an industry that demands speed and efficiency. Embracing it without transparency risks losing credibility. The future will depend on crafting new norms: clear disclosure of AI use, robust verification practices, and renewed attention to the symbolic meaning of authorship. Machines can generate prose, but only humans can be held accountable for what enters the public sphere.

8.6 Global Inequalities and Algorithmic Colonialism

Behind the illusion of AI as a universal tool lies a stark asymmetry, as most large language models (LLMs) are trained on data overwhelmingly drawn from English, supplemented by a handful of dominant languages such as Chinese, Spanish, or French. According to the Common Crawl dataset, one of the most significant open sources used for training, over 60% of its content is in English, even though only around 17% of the world's population speaks it. By contrast, languages spoken by hundreds of millions of people, such as Hausa, Bengali, and Quechua, barely register. This imbalance shapes what machines "know" and, more importantly, whose knowledge counts as valuable. When an AI system can generate pages on Shakespeare but struggles with Wole Soyinka, when it recalls Silicon Valley start-ups but falters on indigenous cosmologies in the Amazon, it reproduces a hierarchy of epistemologies. English becomes not only a medium of communication but the epistemic foundation of machine intelligence. To access AI fluently is, implicitly, to accept English as the default language of knowledge.

Language carries culture, memory, and worldview. When minoritised or indigenous languages are absent from digital corpora, their associated knowledge systems risk erasure. Concepts such as Ubuntu in Southern Africa or *sumak kawsay* ("good living") in the Andes embody ethical frameworks that do not translate neatly into English. When AI ignores these linguistic worlds, it narrows the horizon of what counts as knowledge. Abeba



Co-funded by
the European Union

Birhane (2021) calls this epistemic coloniality: the continuation of colonial patterns of domination in the digital realm. The consequences are global. In Latin America, Quechua and Aymara speakers struggle to find representation in digital tools dominated by Spanish and Portuguese. In Asia, Hindi and Bengali, which are spoken by vast populations, remain marginal compared to English or Mandarin. UNESCO warns that half of the world's 7,000 languages could disappear by the end of the century; AI, rather than preserving them, may accelerate their decline (UNESCO, 2021).

But the problem extends beyond language: AI development depends on massive datasets harvested from around the globe. Social media posts, news articles, images, and voices flow into corporate repositories, where they are processed into training material. Yet the infrastructure that stores and monetises this data (cloud servers, research labs, AI companies) is concentrated in the Global North. The benefits accumulate in Silicon Valley or Shenzhen, not in the communities whose digital traces fuel the system (Couldry & Mejías, 2019). The asymmetry also shapes labour. In 2022, Time reported that Kenyan workers were paid less than \$2 an hour to filter traumatic content – graphic violence, child abuse, hate speech – for OpenAI's training pipelines (Perrigo, 2023). Their invisible work cleansed the datasets so that Western users could interact with "safe" chatbots. This echoes older colonial patterns: dangerous, demeaning tasks outsourced to marginalised populations while profits flow elsewhere.

Michael Kwet (2020) calls this digital colonialism: the Global South's dependence on infrastructures controlled by foreign corporations. Sovereignty is compromised not only economically but also technologically. Nations may regulate AI domestically, but if they lack the infrastructure to develop alternatives, they remain dependent on external powers. Even the research landscape mirrors these inequalities. A 2021 study found that over 80% of the most cited AI papers were authored by scholars based in North America, Europe, or China (Birhane, 2021). African, Latin American, and indigenous voices are underrepresented in shaping the ethical frameworks and technical priorities of AI. As a result, what counts as "responsible AI" often reflects Western anxieties (privacy, copyright, transparency) while overlooking concerns more urgent elsewhere: surveillance, political manipulation, ecological costs. Europe's position is ambivalent. The AI Act aims to set global standards for ethical AI. Yet its infrastructures remain deeply tied to U.S. tech giants. If Europe feels dependent, the Global South feels doubly so: bound to both U.S. technological dominance and European regulatory frameworks.

Yet across the globe, communities are building alternative models: for example, the Masakhane project unites African researchers to develop natural language processing tools for African languages; Mozilla's Common Voice crowdsources voice data to diversify speech recognition; indigenous groups in New Zealand and Canada are digitising their languages, designing AI that reflects oral traditions and communal values.

8.7 Conclusion

To live in the age of artificial intelligence is to live within an ecosystem where truth is constantly negotiated. What defines this moment is not merely the presence of



Co-funded by
the European Union

disinformation, but its transformation into a systemic condition: produced at scale, automated by machines, and woven into the infrastructures of everyday communication.

Throughout this chapter, we have travelled across different landscapes of this new reality. We began with **Hypnocracy**, where power operates not through repression but through trance, saturating our attention until resistance is absorbed into performance. We then examined how **generative AI** produces fluency rather than truth, with hallucinations and corporate filters shaping what knowledge becomes visible. The **Disinformation 2.0** stage revealed how deepfakes, synthetic voices, and hallucinated texts destabilise elections, markets, and public trust. Inside the newsroom, we saw how **AI unsettles journalism**, challenging traditions of authorship, verification, and authority. Finally, the view from the Global South reminded us that **algorithmic colonialism** continues older hierarchies of power, privileging specific languages and knowledges while silencing others.

Taken together, these perspectives show that AI is not neutral, but it crystallises choices about language, visibility, whose labour counts, and whose knowledge matters. It amplifies long-standing vulnerabilities in democracy and journalism, while also making visible the fragility of truth itself. Yet fragility need not mean defeat. Resilience is possible. It requires regulatory transparency (European Union, 2024), grassroots projects that reclaim linguistic and epistemic diversity, and renewed journalistic practices grounded in verification and disclosure. Most of all, it requires citizens capable of critical literacy: recognising that fluency is not the same as credibility, pausing before sharing, and asking not only what is said but who benefits from its circulation.

Reflective Questions

1. When you read a fluent piece of text online, how do you decide whether it was written by a human or a machine, and does the answer matter?
2. Consider a language you speak that is underrepresented online. What cultural concepts might risk invisibility if AI systems cannot process that language?
3. If journalism increasingly relies on AI, what responsibilities should news organisations have to disclose, verify, and protect the integrity of their work?

8.8 References

Arendt, H. (1971). Lying in politics. *New York Review of Books*.

Barthes, R. (1967). The death of the author. *Aspen*, 5–6.

Birhane, A. (2021). Algorithmic colonisation of Africa. *Big Data & Society*, 8(2), 1–12.
<https://doi.org/10.1177/2053951720988254>

Bradshaw, S., & Howard, P. N. (2019). The global disinformation order: 2019 global inventory of organised social media manipulation. Oxford Internet Institute.



Co-funded by
the European Union

Couldry, N., & Mejías, U. A. (2019). *The costs of connection: How data is colonising human life and appropriating it for capitalism*. Stanford University Press.

Debord, G. (1967). *The society of the spectacle*. Buchet-Chastel.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

European Union. (2024). *Artificial Intelligence Act*. Official Journal of the European Union.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993–1002.
<https://doi.org/10.1037/xge0000098>

Floridi, L. (2020). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Pantheon Books.

Foucault, M. (1969). What is an author? *Bulletin de la Société française de philosophie*, 63(3), 73–104.

Hosseini, M. (2023). Gendered disinformation: How digital manipulation targets women in politics. *Journal of Information Technology & Politics*, 20(3), 289–308.
<https://doi.org/10.1080/19331681.2023.2173421>

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.
<https://doi.org/10.1145/3571730>

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293.
<https://doi.org/10.18653/v1/2020.acl-main.560>

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, A. (2023). GPT-4 passes the bar exam. SSRN. <https://doi.org/10.2139/ssrn.4385465>

Kwet, M. (2020). Digital colonialism: U.S. empire and the new imperialism in the Global South. *Race & Class*, 62(1), 3–26. <https://doi.org/10.1177/0306396820919294>

Lee, J., Kim, J., & Lee, K. (2023). Hallucination in large language models: A bibliometric review. *Information Processing & Management*, 60(6), 103375.
<https://doi.org/10.1016/j.ipm.2023.103375>



Co-funded by
the European Union

- Marconi, F., & Siegman, A. (2017). *The future of augmented journalism: A guide for newsrooms in the age of intelligent machines*. Associated Press.
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarisation. *Proceedings of ACL 2020*, 1906–1919.
<https://doi.org/10.18653/v1/2020.acl-main.173>
- NewsGuard. (2023). *Tracking AI-generated news websites*. NewsGuard Technologies.
- NUJ. (2023). *Artificial intelligence in journalism: Opportunities and threats*. National Union of Journalists (UK).
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- OpenAI. (2023). GPT-4 technical report. arXiv. <https://arxiv.org/abs/2303.08774>
- Perrigo, B. (2023, January 18). Exclusive: OpenAI used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic. *Time Magazine*.
- Reuters Institute. (2023). *Digital news report 2023*. Reuters Institute for the Study of Journalism.
- Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux.
- UNESCO. (2021). *Artificial intelligence and freedom of expression*. UNESCO Publishing.
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe Report.
- WIPO. (2021). *Artificial intelligence and intellectual property policy*. World Intellectual Property Organisation.
- Zelizer, B. (2019). Why journalism is about more than digital technology. *Digital Journalism*, 7(3), 343–350. <https://doi.org/10.1080/21670811.2019.1571932>



Co-funded by
the European Union

9 Fact-Checking and Source Verification

Written by Aleksandra Radevska and Natasha Dokovska, Journalists for Human Rights (JHR), 2025

9.1 Introduction: The Art and Ethics of Verification

Over the past twenty years, fact-checking has become a crucial aspect of how we handle information online. It began as something journalists occasionally did, but it has evolved into a distinct field of its own that combines journalism, science, and even law. At its heart, fact-checking is about protecting truth in a world where anyone can publish anything, and where facts, opinions, and lies mix faster than ever. Modern fact-checking started as a niche corner of political reporting and gradually became something much broader. In his landmark study, Lucas Graves (2016) called it “a new knowledge practice” because it does more than debunk falsehoods. Indeed, it borrows the journalist’s speed, the scientist’s precision, and the lawyer’s insistence on evidence. It reframes accuracy as a public service, not just for readers, but for democracy itself.

But fact-checking is not mechanical truth-telling. Every check involves choices: which claims matter most, which sources count as credible, and how to translate complex data into something ordinary people can understand. Each of those choices carries values and politics, however invisible they may seem. What story do you tell when you say something is “false”? Which voices get amplified, and whose are quietly dismissed?

Scholars call this process “epistemic boundary work”: fact-checkers, knowingly or not, draw lines between credible and non-credible, legitimate and fringe, truth and noise. In a perfect world, those boundaries would be clear. In reality, they’re constantly under pressure from populist rhetoric, commercial incentives, and algorithmic feeds that reward outrage over accuracy. Merely insisting that truth exists has become a radical act.

9.2 Truth as a Moving Target

Behind every fact-check lies an old philosophical question: What is truth? Fact-checkers inherit three classic answers, and juggle all of them at once.

- The correspondence view says truth mirrors reality: a statement is true if it matches what actually happened. That’s what we use when verifying numbers, documents, or quotes.
- The coherence view says truth lives in the pattern, which makes sense in the wider web of knowledge. This comes into play when reporters cross-check sources or scientific consensus.
- The pragmatic view argues that truth is what works, the version of reality that helps people act more wisely. Public-health fact-checks often rely on this when explaining risks or safety.

These lenses remind us that not everything can be fact-checked in the same way. A statistic can be verified; a moral claim cannot. Drawing that boundary—what can versus what should be checked—is part of what gives the field its credibility and its humility.

Credibility as a Process, Not a Brand



Co-funded by
the European Union

For a long time, newspapers could rely on their mastheads to carry authority. But in the fact-checking era, credibility has to be earned, step by transparent step. The International Fact-Checking Network's Code of Principles (2016) formalised this idea: trust is procedural. It depends on clarity about sources, funding, methodology, and corrections; proof that errors are admitted, not buried.

Visual rating tools like the Truth-O-Meter or Pinocchios are sometimes criticised as gimmicks, but their value lies in what they represent: a visible, public promise that the reasoning behind a verdict can be retraced. A credible fact-check should be portable: its logic should stand whether it's read on a newspaper site, cited in a classroom, or stored years later in an archive.

That's why preservation tools like the Wayback Machine, Perma.cc, and Archive. Today, matters as much as clever writing. Every permanent link is more than a citation; it's a safeguard against internet amnesia.

Of course, even the most meticulous fact-check can fail if people don't accept it. Research in psychology shows that our brains cling to familiar stories even after they're disproved, a phenomenon called the continued-influence effect (Lewandowsky et al., 2012). We tend to believe repeated claims —the illusory truth effect —no matter how false they are. Yet the good news is that corrections, when well-crafted, do work across political lines (Wood & Porter, 2019). This science of persuasion gives fact-checking a practical blueprint: lead with the truth, provide a clear alternative explanation, and keep it visible long after the myth fades from the headlines.

Deciding What to Check

Not all lies cause equal damage; that's why fact-checkers must balance urgency with potential harm, especially when the falsehood affects people's health, safety, or democratic participation. A rumour about a polling location may matter more than a viral exaggeration about taxes. Speed is important, but so is precision. The goal is not just to debunk quickly but to interrupt harm. Tools like Community Notes or live debunk blogs help close the time gap between rumour and correction, but these systems work only when combined with human judgment: What needs verifying now, and what can wait?

The struggle for truth is never neutral. As we know now, gendered disinformation shows how bias operates not only in claims but in credibility itself. Women and marginalised voices are often doubted, mocked, or digitally silenced through sexualized hoaxes. Fact-checkers therefore face a dual responsibility: to correct falsehoods and to repair visibility. Practices like harm-first triage, blurred reproductions of abusive content, and chain-of-custody documentation protect both accuracy and dignity.

Keeping Truth from Vanishing

Finally, every verified claim must be preserved. Misinformation thrives on the instability of the web (deleted tweets, edited pages, expired links). Good fact-checking, therefore, depends on time-proof evidence. Archiving ensures that even when the internet forgets, the memory of verification endures. Fact-checking sits at the crossroads of technology, ethics, and psychology: a discipline still inventing itself as the digital world keeps changing the rules. It's both an act of journalism and a statement of belief —that truth, though fragile and contested, is still worth building systems to protect.



Co-funded by
the European Union

9.3 Typologies of Fact-Checking Resources

Institutional Fact-Checkers and Rating Systems

One big part of the fact-checking world is made up of professional organisations that verify claims every day. Some are part of major newsrooms (like The Washington Post's **Fact Checker**) while others are independent non-profits such as **Full Fact** in the UK, or regional hubs like **Africa Check** and **Chequeado**. These groups publish their verdicts in a transparent way and often use simple rating systems that make their work easier to understand: for example, Pinocchio's for The Washington Post or the Truth-O-Meter used by **PolitiFact**. These scales help readers quickly grasp the verdict and ensure that journalists inside the newsroom follow the same standards when assessing a claim. On the websites of PolitiFact and The Washington Post Fact Checker, you can find public methodology pages that explain how they evaluate evidence and handle corrections (PolitiFact; Washington Post Fact Checker). Fact-checking has also become more structured at a global level: The **International Fact-Checking Network (IFCN)** was one of the first to set international standards through its Code of Principles, while in Europe, the **European Fact-Checking Standards Network (EFCSN)** now certifies fact-checkers that respect a detailed Code focused on transparency, ethics, and methodology, for example, by requiring them to publish how they are funded, who runs the organization, and how they correct mistakes (EFCSN Code of Standards). These codes act as "meta-resources" because they define what good fact-checking should look like and how sources must be documented.

If you want to explore who's active in this field, the **Duke Reporters' Lab** keeps the most complete global database of fact-checking projects. Its annual report not only lists hundreds of initiatives (443 in 2025!) but also analyses trends and challenges in the field, such as how platform policies can affect fact-checking visibility.

Structured Data: ClaimReview, MediaReview, and Discovery Tools

Another important area deals with how fact-checks can be found online. The **ClaimReview** system (Schema.org) allows publishers to add special metadata (like the text of the claim, the verdict, and sources) so that platforms like Google can index and display them consistently. The **Google Fact Check Explorer** and its API use this data, making it possible to search verified claims by keyword, person, or even image. The Google News Initiative offers training on how to use these tools effectively. For images and videos, there's also **MediaReview**, a related standard that helps describe whether a piece of visual content has been edited or manipulated. This tool was created through collaboration between Duke's Reporters' Lab, fact-checking organisations, and tech platforms (Reporters' Lab; ClaimReview Project, MediaReview hub). However, there's been a recent change: as of **June 2025**, Google no longer displays ClaimReview data directly in Search results. It still supports it in Fact Check Explorer, though, so it's crucial for fact-checkers to keep tagging their work for visibility through Explorer and the API (Google Search developer docs, 12 June 2025).

Verification Handbooks and Training Resources

If you want to learn how to verify information, the **Verification Handbook** series by the **European Journalism Centre** is a must-read. The first handbook (2014) taught journalists how to check user-generated content (UGC), while the latest edition (**Verification Handbook 3: For Disinformation and Media Manipulation**) focuses on detecting bots, coordinated manipulation, and deepfakes, with plenty of real examples. These books are free and



Co-funded by
the European Union

widely used in newsroom trainings and online courses (EJC). Other great resources include **Amnesty International's Citizen Evidence Lab**, which provides guides and tools for verifying online material related to human-rights issues, such as methods to check when and where a video was uploaded.

OSINT and Forensic Toolkits for Image and Video Verification

To verify images and videos, journalists often use **open-source intelligence (OSINT)** tools: browser-based applications that help them identify where and when a photo or video was taken. One of the best-known tools is the **InVID-WeVerify plugin**, developed within the EU's **vera.ai** program. It includes features like reverse image search, metadata analysis, and forensic filters. Guides by **Bellingcat** and other organisations show how to use it in real investigations. **Amnesty's YouTube Data Viewer** is another essential tool: it reveals the exact upload time of a video and generates thumbnails that can be reverse-searched, which helps uncover original uploads and spot reused or miscaptioned footage. For location and time estimation, journalists use **SunCalc** (to analyse shadows and sun position) and **MapChecking** (to estimate crowd sizes in images). Finally, **EXIF data viewers** such as Exif. Tools can provide hidden metadata from images, though this data is often removed, so it's always good to double-check with other sources.

Transparency Pages and Standards

Most professional fact-checkers maintain **transparency pages** that explain their rating scales, sources, and correction policies. These are valuable resources for learning how journalistic accountability works in practice. Comparing, for instance, **PolitiFact's Truth-O-Meter**, **Reuters/ AFP's categories**, and **BBC Reality Check's approach** can show how different outlets balance clarity, nuance, and transparency.

9.4 The Craft of Checking the Truth

Fact-checking has often been called the “craft of doubt”, but reducing it to mere scepticism misses the point. Over the past decade, journalists, researchers, and online investigators have worked to turn fact-checking into a discipline with clear methods, open processes, and professional standards. It grew from necessity. In a noisy digital world full of claims, memes, and hot takes, people began to ask not just “Is this true?” but “How do you know?”

Craig Silverman's **Verification Handbook** (2014; updated 2021) helped transform how verification was taught. Instead of treating it as a kind of detective's intuition, Silverman showed how fact-checking could be systematic. Drawing from investigative journalism and open-source intelligence (often shortened to OSINT), he included practical checklists for verifying sources, analysing metadata, and using digital forensics. His main message was clear: verification isn't a magic instinct; it's a method. Every step must be transparent, repeatable, and clear enough for others to follow (Silverman, 2021).

This scientific mindset echoes philosopher Karl Popper's idea of falsifiability (Popper, 1963): the notion that for something to be scientific, it must be testable and open to being proven wrong. In the same way, good fact-checking invites challenge. When a fact-checker links to primary sources, adds screenshots, or explains how an image's origin was verified, they're practising what scholar Lucas Graves calls public epistemology (Graves, 2016): showing the public how knowledge itself is constructed.



But fact-checking is not just about nailing down accuracy. It's also about drawing boundaries —deciding what counts as reliable knowledge and what doesn't (Gieryn, 1983). That boundary work can be political. When fact-checkers challenge a government's statistics, they are also questioning authority. When they validate activists or citizens as credible sources, they expand who gets to speak with legitimacy. Philosopher Miranda Fricker (2007) calls this dynamic epistemic justice: recognising how power influences whose knowledge is taken seriously. Fact-checking that's aware of this doesn't just test facts; it tests fairness.

The Three Commitments of Verification

Modern verification stands on three commitments:

1. Procedural transparency: showing not just results but the steps behind them.
2. Epistemic humility: remembering that even verified claims are provisional and can change.
3. Boundary work with justice: defining trustworthy information in a way that's fair and inclusive.

These values guide all verification work, whether it's investigating a viral quote, decoding a video, or testing a political statement.

Tracking Down Textual Truths

One of a fact-checker's classic missions is verifying what someone actually said. Words, after all, travel fast and often get twisted. Even in an age of deepfakes, old-fashioned misquotations remain a powerful form of misinformation.

The golden rule is simple: go to the primary source. Yet that's usually the hardest step. A post might quote "a UN report" without a link, or a politician might cite "a new study" without any details. Fact-checkers dig through repositories like Hansard in the UK, the U.S. Congressional Record, or the EU's EUR-Lex database until they find the original document. As Silverman warns (2021), relying on second-hand reporting can spread errors like a chain reaction.

Context is just as important, because partial quotes can hijack meaning. During the Brexit referendum, viral posts claimed that Jean-Claude Juncker had said "Britain will be punished", but looking back at the full press conference revealed that he'd said something far less dramatic: "Leaving has consequences" (Full Fact, 2016). A few missing words can turn moderation into a menace. Another favourite online mischief is the fake quote from a famous person. Psychologists Allport and Postman (1947) showed that people naturally trust statements linked to authority figures. That's why a line often credited to Albert Einstein – "Insanity is doing the same thing over and over and expecting different results" – has spread so widely, even though it first appeared in a Narcotics Anonymous pamphlet decades after his death (O'Toole, 2017). Fact-checkers use digital libraries, quotation guides, and linguistic analysis to trace such quotes to their true origins.

The Speed Trap

Fact-checkers face a unique challenge: the race against time. False information runs wild long before truth catches up. During fast-moving crises (shootings, protests, or natural disasters), misinformation can multiply within minutes. The Las Vegas shooting of 2017 saw hundreds of false claims about attackers and motives, all shared millions of times within



Co-funded by
the European Union

hours. To cope, many newsrooms use triage systems: they focus first on the most dangerous or viral claims, publish what's known so far, and update as evidence emerges. They call this live verification (Graves & Cherubini, 2016). It's a delicate balance: move too slowly, and the rumour wins; move too fast, and you risk being wrong.

Keeping Truth Alive

Verification isn't finished when the fact-check is published: every claim must be archived, because online evidence disappears quickly; a problem known as "link rot". Studies show that even court decisions lose nearly half their linked evidence over time (Koehler, 2004). Tools like the Wayback Machine or Perma.cc preserve online sources, ensuring that every debunk comes with a lasting trail of proof.

For sensitive stories, from human rights abuses to gender-based harassment, keeping that digital trail secure is crucial. Organisations like Amnesty International's Citizen Evidence Lab teach ways to record where content came from, who captured it, and how it was edited. This "chain of custody" makes digital findings more trustworthy and, in some cases, legally admissible (Amnesty, 2021).

Increasingly, fact-checkers also "show their work" publicly, through shared verification notebooks or interactive threads (like those from Bellingcat). Graves (2016) calls this "making verification visible". It turns credibility from something earned once into something demonstrated continuously. But transparency has limits too: revealing too much can also cause harm. The European Fact-Checking Standards Network (2022) recognises this balance: methods should be visible whenever possible, but people's safety always comes first.

The Human Side of Verification

Even the best systems have flaws. False stories spread fast because they're emotional, novel, or dramatic. Truth often arrives more slowly, weighed down by the need for evidence and care. By the time a correction appears, the lie has already shaped people's opinions, which is a frustrating reality that researchers call the continued influence effect (Lewandowsky et al., 2012). People remember the story, not the correction. That's why good debunks don't just say "This is wrong"; instead, they replace false stories with new, vivid explanations that make sense. Instead of simply stating "vaccines don't cause autism", effective fact-checkers explain why the myth started and highlight the overwhelming scientific consensus that proves vaccines are safe (Ecker et al., 2022).

Verification is a craft built on patience, empathy, and reflection. It asks hard questions: Could this debunk harm someone? Who benefits from exposing this? Am I reinforcing bias by repeating it? It's not just a technical skill but an ethical one, part of what philosopher Sandra Harding (1991) might call the pursuit of responsible knowledge. Fact-checkers, then, are not only defenders of accuracy. They are builders of trust, teaching the public how to think clearly in a world that never stops talking.

9.5 Automation, AI, and the Future of Fact-Checking

Not long ago, fact-checkers worked almost entirely with their own eyes, ears, and notebooks. They sifted through speeches and social media posts, verified quotes,



Co-funded by
the European Union

compared charts, and hunted down sources by hand. But as the online world exploded, millions of posts per minute, it became clear that human patience alone couldn't keep pace. Enter automation and artificial intelligence: powerful partners that promised speed, but also brought their own uncertainties.

The first wave of experiments looked a lot like digital treasure hunts. One early system, **ClaimBuster** (Hassan et al., 2017), was trained to scan debates and detect “check-worthy” statements: claims that sounded factual enough to verify. Imagine a robot watching a political debate, highlighting lines that sound like: “Unemployment is at its lowest in decades”. It could flag such moments almost instantly, helping journalists focus on the quotes that mattered most. But machines quickly showed their blind spots, because they confused empty slogans with real data points and couldn't grasp sarcasm, irony, or the subtle jabs that often carry gendered disinformation. A meme mocking a woman politician as “too emotional” doesn't read as a “factual claim” to an algorithm; it slips right past.

As projects like **FEVER** (Thorne et al., 2018) pushed automation further, tools began retrieving evidence automatically, scanning Wikipedia for supporting passages. Organisations such as Full Fact in the UK soon used similar systems to pair politicians' statements with existing fact-checks within seconds. The speed was dazzling, but partial. Computers fetched documents about a topic, not necessarily about the claim. And because online archives are patchy, automation risks echoing societal bias: there's plenty of data about powerful institutions, but little that represents marginalised groups (D'Ignazio & Klein, 2020). If AI systems can't “see” those stories, they also can't fact-check them.

The Machines Learn to See

While text-based claims dominate headlines, modern misinformation often comes wrapped in pictures and videos. Enter the era of deepfakes and cheapfakes: the visual wild west of misinformation. The EU's InVID-We Verify project taught machines to reverse-search images, check metadata, and grab keyframes to track a video's origin. Later, deep-learning models learned to spot digital tampering by studying lighting inconsistencies or unnatural facial motion (Hwang et al., 2020). Yet this too became a cat-and-mouse game. As detectors improved, manipulators adapted. Like spam evolving to outwit filters, new fakes hide their tracks better every month. Sometimes algorithms even misfire, labelling genuine footage as false. And many viral deceptions don't involve editing pixels at all; a photo from 2018 reposted during a 2025 protest can still mislead millions. Machines can catch shadows and seams, but they can't tell why a photo is being misused. Only human reasoning can do that.

When Language Models Join the Team

Then came large language models, GPT-style systems that could write, summarise, and sound eerily human. Suddenly, automation wasn't just catching errors; it was also explaining them. These tools could digest multilingual news feeds, detect recurring falsehoods, and draft tentative fact-checks for human review (Zeng et al., 2023). But they come with a catch: confidence without accountability. LLMs sometimes hallucinate sources or misstate facts entirely (Maynez et al., 2020). Their reasoning is a black box: no one, not even their creators, can always explain why they say what they say. Studies (Lucy & Bamman, 2021) also show that these systems mirror social bias, often repeating gendered or cultural stereotypes found in their training data. That's why the European



Co-funded by
the European Union

Fact-Checking Standards Network (EFCSN, 2022) insists that final verdicts must remain in human hands. Machines can help investigate; humans must decide.

In practice, the best setups blend strengths. At Spain's Newtral, AI listens to live political debates, flagging potentially checkable claims; editors then decide what truly matters. In France, AFP Factual uses AI to detect trending images but still relies on journalists for provenance checks. These hybrid approaches reflect a simple truth: automation can map the territory, but only humans can read its meaning. This partnership raises new ethical puzzles, though. We all know that algorithms feel neutral but aren't: they reflect their data and design – who built them, what they measured, and what they missed. Meanwhile, large-scale social-media monitoring invites privacy concerns, since detecting falsehoods often means scanning private or semi-private conversations. The rule of thumb is now clear: automation should assist verification, not outsource it.

Unequal Data, Unequal Detection

Many automated tools were trained on English or other high-resource languages. That means Polish, Romanian, Welsh, or Basque disinformation may go unseen by global detectors. The same gap appears in gendered contexts: abusive or sexualized attacks against women often slip through because models don't recognise local slang or cultural nuance (Gorwa et al., 2020). In this sense, automation can accidentally extend the invisibility it was meant to challenge.

Also, AI itself can be fooled. Researchers found that subtle image tweaks, imperceptible to people, can make detectors completely misclassify a fake (Goodfellow et al., 2015). Disinformation actors exploit this, designing content to fly under automatic radar using altered fonts or deliberate distortions. Verification is thus an arms race, a digital chess match where each side learns from the other's last move.

Despite its futuristic image, automation rests on very human shoulders. Training data must be labelled, often by low-paid workers in the Global South who view and tag content all day so AI can "learn" (Grey & Suri, 2019). Even within newsrooms, maintaining automated pipelines demands constant technical care. Many fact-checking projects survive on temporary grants, while misinformation remains cheap and viral. Automation saves labour only if that labour is first invested, ethically and sustainably.

Looking Ahead

Automation isn't replacing fact-checkers; it's reshaping them. Machines handle the volume, spotting anomalies, gathering evidence, but humans interpret, contextualise, and own accountability. The next frontier lies in explainable AI, systems that don't just give answers but show their reasoning (Doshi-Velez & Kim, 2017). Others are tackling cross-lingual tools to connect fact-checks across Europe, or citizen evidence labs that invite the public to help archive truth online. The goal is simple but profound: a partnership where humans and machines strengthen, rather than erode, trust. AI can sharpen our eyes, but our judgment still comes from the oldest source of all, a questioning mind.



9.6 References

- Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2019). Protecting world leaders against deep fakes. In Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).
- Allport, G. W., & Postman, L. (1947). The psychology of rumour. Henry Holt.
- Amnesty International. (2021). Citizen Evidence Lab: Digital verification for human rights. Amnesty International.
- Babakar, M., & Moy, W. (2016). The state of automated fact-checking. Full Fact.
- Bellingcat. (2015). MH17 Investigation: Tracking the Buk missile launcher. Bellingcat.
- Bellingcat. (2022). Geolocation and chronolocation in Ukraine war reporting. Bellingcat.
- Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. New Riders.
- DiFonzo, N., & Bordia, P. (2007). Rumour psychology: Social and organisational approaches. American Psychological Association.
- D'Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT Press.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- Drake Law Library. (2022). Perma.cc: Preserving web sources for legal research. Drake University Law Library.
- Dries, M., van der Bles, A. M., & van der Linden, S. (2025). Communicating uncertainty in public statistics: Effects on trust and comprehension. [Journal article, forthcoming].
- EFCSN (European Fact-Checking Standards Network). (2022). Code of Standards. European Fact-Checking Standards Network.
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., et al. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29.
- European Commission. (2022). The Digital Services Act package. European Commission.
- EUR-Lex. (2013). Directive 2013/ 33/ EU of the European Parliament and of the Council of 26 June 2013 laying down standards for the reception of applicants for international protection. Official Journal of the European Union.
- Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993–1002.
- Full Fact. (2016). Fact checks on Brexit referendum claims. Full Fact.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53–96.



- Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, 48(6), 781–795.
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In the International Conference on Learning Representations (ICLR).
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15.
- Graves, L. (2016). *Deciding what's true: The rise of political fact-checking in American journalism*. Columbia University Press.
- Graves, L. (2018). Boundaries not drawn: Mapping the institutional roots of the global fact-checking movement. *Journalism Studies*, 19(5), 613–631.
- Graves, L., & Cherubini, F. (2016). The rise of fact-checking sites in Europe. Reuters Institute for the Study of Journalism.
- Grey, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.
- Gun Violence Archive. (2019). Analysis of crime statistics. Gun Violence Archive.
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8(1), 1–51.
- Harding, S. (1991). *Whose science? Whose knowledge? Thinking about women's lives*. Cornell University Press.
- Hassan, N., Li, C., Arslan, F., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by Claimbuster. In Proceedings of the 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).
- Hwang, Y., Kim, J., & Lee, H. (2020). Detecting manipulated media: Deepfake and cheapfake forensics. *IEEE Signal Processing Magazine*, 37(1), 118–132.
- International Fact-Checking Network (IFCN). (2016). Code of Principles. Poynter Institute.
- InVID-WeVerify Consortium. (2020). Verification case studies. WeVerify.eu.
- Koehler, W. (2004). A longitudinal study of Web pages continued: A consideration of document persistence. *Information Research*, 9(2), paper 174.
- Lewandowsky, S., Ecker, U. K. H., Seifert, C., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., et al. (2020). *The Debunking Handbook 2020*. Cognition and Climate.
- Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation through inoculation. *Nature Human Behaviour*, 5(10), 1231–1233.



Co-funded by
the European Union

- Lucy, L., & Bamman, D. (2021). Gender and representation bias in GPT language models. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP).
- Markowitz, D. M., Amazeen, M. A., & Graves, L. (2023). Comparing fact-checking verdicts across outlets: Agreement, scales, and selection. *Journalism*, 24(5), 987–1005.
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarisation. In Proceedings of ACL 2020, 1906–1919.
- Nakov, P., Da San Martino, G., Barrón-Cedeño, A., Papotti, P., Shaar, S., & Alam, F. (2021). Automated fact-checking for assisting human fact-checkers. *Communications of the ACM*, 64(6), 88–98.
- Newton, C. (2023). Meta is shutting down CrowdTangle. Platformer.
- Nyhan, B., & Reifler, J. (2015). Displacing misinformation about events: An experimental test of causal corrections. *Journal of Experimental Political Science*, 2(1), 81–93.
- O'Toole, G. (2017). Insanity is doing the same thing over and over and expecting different results. Quote Investigator.
- Onder, G., Rezza, G., & Brusaferrò, S. (2020). Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA*, 323(18), 1775–1776.
- Pinch, T. J. (2019). Epistemic boundary work in a post-truth era. In *Routledge Handbook of Post-Truth* (pp. 45–58). Routledge.
- Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge & Kegan Paul.
- Porter, E., & Wood, T. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in 12 countries. *Proceedings of the National Academy of Sciences*, 118(37), e2104235118.
- Silverman, C. (Ed.). (2021). *Verification handbook for disinformation and media manipulation*. European Journalism Centre.
- Society of Professional Journalists (SPJ). (2014). *SPJ Code of Ethics*. Society of Professional Journalists.
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(3), 538–556.
- Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. In Proceedings of NAACL-HLT 2018 (pp. 809–819).
- UNESCO. (2020). *Balancing act: Countering digital disinformation while respecting freedom of expression*. UNESCO.
- UNM Law Library. (2025). *Perma.cc for legal scholarship: A user guide*. University of New Mexico School of Law.
- UN Special Rapporteur on Freedom of Opinion and Expression. (2023). *Gendered disinformation and freedom of expression (A/ 78/ 288)*. United Nations.



Co-funded by
the European Union

Uscinski, J. E., & Butler, R. W. (2013). The epistemology of fact-checking. *Critical Review*, 25(2), 162–180.

Van der Bles, A. M., van der Linden, S., Freeman, A. L. J., Mitchell, J., Galvao, A. B., Zaval, L., et al. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *PNAS*, 117(14), 7672–7683.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe.

WeVerify Consortium. (2020). Verification case studies. WeVerify.eu.

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behaviour*, 41(1), 135–163.

Zeng, J., Chan, C., & Fu, K. (2023). AI and disinformation: Opportunities and risks. *Journal of Information Technology & Politics*, 20(2), 163–181.