



Know thyself to know the truth: Fighting disinformation in the age of artificial intelligence through foundational, psychological and emotional literacy

European View
2025, Vol. 24(2) 224–237
© The Author(s) 2025
DOI: 10.1177/17816858251390921
journals.sagepub.com/home/euv
S Sage

Giulia Maria Galli

Interinstitutional Relations Unit of the European Commission, Belgium

Abstract

In the age of artificial intelligence–driven manipulation, deepfakes and algorithmic distortion, the crisis of truth is not merely technological but epistemological. Traditional digital literacy efforts are insufficient if they do not address deeper vulnerabilities in human cognition and communication. In this article, a radical, long-term strategy is proposed, one centred on foundational literacy (in terms of etymology, linguistics and translation), psychological literacy and emotional literacy. By empowering individuals to decode not only information but also their emotional and cognitive responses to it, such education can counter both polarisation and corrosive scepticism. This three-tiered approach draws on ancient wisdom—‘*gnōthi seauton*’, or ‘know thyself’—to reframe resilience to disinformation as a journey to self-knowledge. This article presents a call to action for policymakers to invest in education that fosters discernment, self-awareness and social cohesion, building not only critical thinkers but reflective citizens capable of safeguarding democratic societies.

Keywords

Information, Disinformation, Misinformation, Malinformation, Artificial intelligence, Deep-fakes, Fake news, Digital literacy, Functional illiteracy, Emotional illiteracy, Psychological illiteracy

Corresponding author:

Giulia Maria Galli, Interinstitutional Relations Unit of the European Commission, 1049 Brussels, Belgium.
Email: giuliamaria.galli@outlook.it



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).



Introduction

If, after seeing this screenshot¹ and reading its content, you googled this headline and hoped, even for a second, to find out more, the next 5–10 minutes you spend reading this article could be very useful. Do not feel discouraged; nobody is immune to the power of disinformation, and mystification can infiltrate every environment, not only the digital one.

In this article, ideas on how to empower individuals and improve their discernment will be presented. These ideas will focus on the crux of the problem: the ability to master communication and one's emotions. The main argument of this article is that in order for digital literacy programmes to be effective, they also need to address functional, psychological and emotional illiteracy.

Throughout the article, it will become increasingly clear that information disorders, as described below, thrive on the inability of individuals to critically analyse both messages and their psychological responses to their exposure to such messages. This inability is particularly pronounced among younger generations. As will be discussed later, it is precisely younger cohorts who prove the most vulnerable to challenges such as functional and psychological illiteracy. In what has been defined as a post-truth society,² where emotions and beliefs outweigh objectivity in shaping public opinion, foundational literacy can allow individuals to recognise their own feelings and analyse them before they take over from their discernment skills. When truth becomes fragmented and perception is shaped by emotional narratives, traditional markers of credibility erode. Therefore, the inability to distinguish between true and false is not only an informational crisis, but also an epistemological one, challenging the traditional markers of credibility, knowledge, truth and authority.

The link between information disorders and the erosion of trust in institutions is very strong. Therefore, in this article, a call to action is presented. The problem can be addressed by institutions, but their initiatives must mirror its multilayered and complex nature, which cannot be tackled only by regulations or digital literacy campaigns. In the first section, the historical and conceptual framework of information disorders will be outlined, tracing their evolution from propaganda to the artificial intelligence-driven era of deepfakes. In the second section, the psychological and sociological mechanisms underlying susceptibility to disinformation will be examined, highlighting how these

mechanisms provoke two main societal reactions: polarisation and scepticism. In the third section, the unintended effects of certain countermeasures will be analysed, showing how some educational tools can inadvertently fuel distrust. In the final section, existing regulatory measures and innovative proposals will be presented, culminating in the central contribution: that reinforcing literacy in its broadest sense will provide the essential foundation for more effective digital literacy initiatives. The conclusion will stress that only a long-term, foundational approach can safeguard meaning, truth and social cohesion in an evolving information environment.

Understanding the spectrum of information manipulation: a historical and conceptual framework for information disorder

The use of the term disinformation is fairly recent. It first appeared in English dictionaries towards the end of the 1980s. Its origin dates back to the Russian neologism ‘дезинформация’ (dezinformatzija), which was coined in 1923 when the vice-president of the State Political Directorate (the body that preceded the KGB) called for the establishment of a special disinformation office to conduct tactical intelligence operations. From this point onwards, disinformation became a tactic used in Soviet psychological warfare (Pacepa and Rychlak 2013).

After this brief introduction, one might assume that information manipulation is a Russian invention. However, this is not the case. Distorted information is a deeply human phenomenon and, as such, has existed for as long as humankind itself. This may explain why the terms disinformation, misinformation and malinformation are so frequently used interchangeably. To grasp just how ancient information manipulation techniques are, one need only consider what is often cited as one of the earliest documented disinformation campaigns in history: the smear campaign orchestrated by Octavian against Mark Antony, in the pre-Christian era. Octavian portrayed Mark Antony as a traitor to Rome, suggesting that his loyalty lay more with his lover Cleopatra and the interests of Egypt than with his own country.

While information manipulation itself has a long history, academic research on the topic is new, as are attempts to assign each term a clear and distinct meaning. We can define the ensemble of distorted information as ‘information disorder’, where the different actions distorting information each have a univocal meaning, denoting the intentions of the sender of a message.

Within information disorder, there are three main categories:

1. *Malinformation*: genuine information that is shared with the intent to cause harm.
2. *Misinformation*: information that is simply false or incorrect, for whatever reason.

3. *Disinformation*: fake or inaccurate information that is intentionally spread to mislead and/or deceive. It is misinformation coupled with a psychological intention to deceive or harm others.

To these categories we can add the following subcategories:

1. *Propaganda*: when disinformation is deployed in the service of a political agenda.
2. *Fake news*: news articles that are intentionally and verifiably false and could mislead readers (Shu et al. 2020; Van der Linden 2023).
3. *Deepfake*: the fake video output of a deep-learning-based technology in which the face of a person is swapped with the face of another person. This term originated in late 2017 after a Reddit user named ‘deepfakes’ claimed to have developed a machine-learning algorithm that helped him to transpose celebrity faces onto pornographic videos (Tolosana et al. 2020).

Moreover, information disorders have a dynamic nature and can overlap. Disinformation can, for example, turn into misinformation. Let us consider a disinformation campaign spread on social media with malicious intent. The user who sees this content may not be aware of its falsity and might share it with his or her communities or friends. In this case, the user’s action cannot be classified as disinformation; rather it is misinformation because the intent was not malicious (Shu et al. 2020). An example of this is the health crisis caused by hydrogen peroxide and alcohol poisoning that occurred in Pakistan during the Covid-19 pandemic. At the time, a disinformation campaign claimed that ingesting hydrogen peroxide or pure alcohol could eliminate the virus. Communities then shared this false advice, believing it to be helpful. As a result, many consumed these toxic substances, triggering a public health emergency (Van der Linden 2023).

From propaganda to deepfakes: three phases of information disorder

Three stages of information disorder can be identified in modern history. The first occurred during the twentieth century. Throughout the century, the practice of spreading propaganda became institutionalised, most notably during the First and Second World Wars, when state actors systematically harnessed media to mobilise support and vilify opponents.

The second phase stretches from the beginning of the twenty-first century to 2022, the year in which ChatGPT was released. During these 23 years, the scale and velocity of the growth of the digital environment transformed the spread of information disorder. The speed of information dissemination in digital environments started to outpace the capacity to fact-check. Moreover, because algorithmic curation on social media platforms tends to amplify emotionally charged or sensational content, the visibility and reach of false or misleading information increased. The Covid-19 pandemic represented a wake-up call in this regard.

We are now living in the third phase of information disorder, where artificial intelligence (AI) has democratised the production and dissemination of false information, enabling virtually anyone to act as both creator and distributor. Unlike traditional centralised and state-driven propaganda, contemporary disinformation often emerges from a diverse array of actors, including individuals. Driven by the recent technological advancements in generative AI tools, the proliferation of deepfake content on social media platforms grew by 550% between 2019 and 2023. This exponential increase has triggered concerns about the scale of the phenomenon and led the World Economic Forum to point to deepfakes and disinformation as one of the key global challenges in 2024 (Deloitte 2025). This trend is particularly alarming given that individuals aged 18 to 25 spend an average of three hours per day on social media, with younger cohorts exceeding four hours daily—a figure that continues to rise. At the same time, these platforms are increasingly relied upon as primary sources of news and information. The capacity to evaluate authenticity is therefore vital for younger generations, who often lack the habit of actively seeking a plurality of sources.

But what is at stake that makes information disorder so concerning? Apart from the economic and political impacts, the most concerning risk is existential: the inability to discern what is true and what is false, and the consequent two reactions: strong polarisation or widespread unconditioned scepticism. Individuals who are not aware of or careful about the information they consume might become further polarised; whereas those who are aware might become increasingly alive to the fact that malicious actors are deliberately and systematically producing false content with considerable speed, frequency and intensity. As a result, the latter may adopt a more cynical and distrustful attitude, questioning the legitimacy of any informational source. Although these are contrasting dynamics, both progressively erode trust, not only in the information ecosystem itself but also in institutions, organisations and all forms of authoritative knowledge.

Behind these two opposing reactions lie deeper psychological and social dynamics. These will be explored in the following sections.

The comfort of falsehoods: how disinformation satisfies psychological needs and fuels polarisation

What psychological mechanisms drive individuals to believe in distorted narratives and conspiracy theories in the first place? This section offers a list of factors that lead individuals to embrace distorted information:

1. Perceptual bias: beliefs, emotions and desires influence how we interpret reality, leading people to perceive what they wish to be true rather than what is factual. This factor is also called *truthiness*—the idea of conferring truth to what your gut feelings perceive as such (Van der Linden 2023).
2. First-narrative bias: in moments of uncertainty or in an information vacuum, the first narrative to emerge, regardless of its accuracy, tends to dominate and shape public opinion, highlighting the importance of timely and accurate communication.

3. Cognitive biases, which fall into two subcategories:
 - availability bias: prioritising information that comes to mind easily;
 - anchoring bias: fixating on initial information, which distorts people's understanding.
4. Confirmation bias: people tend to favour information that aligns with their pre-existing views.
5. Trusted-source bias and the familiarity heuristic: information received from personal contacts is deemed more accurate than information coming from verified sources.
6. Conformity bias: the tendency to change one's beliefs or behavior to fit in with others.

The other two mechanisms that our minds are sensitive to are the *illusory truth effect*, which leads us to perceive repetition as truth, and *fluency*. This latter can be defined as the ease with which information is processed by the brain. Something is fluent when it is easy to understand. This can influence the brain's judgement ability, leading it to make fast and intuitive judgements, rather than assessing the facts analytically (Van der Linden 2023). Moreover, *social recognition* is also crucial. Disinformation in particular exploits psychological domains such as an individual's social identity and sense of group belonging. Judgements on the veracity of information can be based on one's ability to build a social consensus, find evidence to support the latter, build its internal coherence and establish its credibility (Pinhanez et al. 2023). The influence of social motivation in the information environment is so powerful that it can overrule *accuracy*—the innate desire of humans to seek the truth. This leads to another bias: the *worldview backfire effect*. This occurs when corrective information, challenging an individual's beliefs or ideology, strengthens his or her original beliefs instead of undermining them (Van der Linden 2023).

The most sensitive to these mechanisms are, once again, young people. Youth, and adolescents in particular, are especially susceptible to the need for social recognition, which translates into a strong desire to feel accepted by their in-group. For this reason, they are more vulnerable to confirmation bias and conformity bias, both of which provide fertile ground for disinformation to thrive.

The above illustrates how disinformation fills psychological gaps users might not be aware of. Being unaware of the existence of these psychological mechanisms and unable to analyse the emotions that they trigger provide fertile ground for polarisation. Moreover, faced with a choice between comforting certainty and the unsettling pursuit of truth, many avoid risk, allowing disinformation to flourish and deepen polarisation both online and offline. In the digital world, polarisation is represented by digital environments called echo chambers, filter bubbles and rabbit holes:

- Echo chambers are social environments in which individuals are primarily exposed to information that confirms their existing beliefs. Dissenting views are excluded or discredited, reinforcing groupthink and amplifying misinformation.

Echo chambers are created by users themselves (e.g. a WhatsApp group with like-minded people).

- Filter bubbles are algorithm-driven information environments that tailor content to a user's past behaviour. This passive curation shields users from opposing perspectives, deepening cognitive biases without their awareness. They are technologically created (e.g. an Instagram feed).
- Rabbit holes are a specific kind of filter bubble, where algorithmic viewing recommendations encourage users to watch more and more extreme content over time, on a path of escalating exposure to increasingly extreme or misleading content, often driven by algorithmic recommendations (Van der Linden 2023).

The definitions of these various polarising environments highlight the distinction between spaces that are created socially and those created technologically. The question arises: is social media to blame for steering users into echo chambers and ultimately leading them down rabbit holes? Or does social media merely reflect offline polarising dynamics? Extensive research is being conducted to analyse the spread of disinformation offline. One study (Brown and Enos 2021) geolocated 180 million registered voters in the US and found that even in the same neighbourhoods, Republicans and Democrats cluster away from each other. This finding raises the distinct possibility that online echo chambers are not a product of social media but are instead induced by the offline environment.

Although social media has not created these phenomena from scratch, it has reshaped the speed at which information is shared, its scale and the medium through which this sharing takes place (Van der Linden 2023). For this reason, awareness campaigns have been established and policies introduced to tackle the issue, leading to the extreme opposite result: diffused scepticism. This will be the topic of the following section.

The cynicism trap: when fighting disinformation backfires

Research shows that higher levels of education reduce the likelihood of believing fake news and conspiracy theories. Inspired by medical science, the research community has experimented with 'inoculation' methods to protect against the 'disease' of disinformation. The idea mirrors the way in which antibiotics work: expose individuals to weakened 'strains' of misinformation so that they can better recognise and reject them.

Among these methods are interactive games, such as the University of Cambridge's 'Bad News' and 'Go Viral', which present users with both true and false statements to refine their discernment skills. After playing, participants were more likely to flag fake news as fake but were also more likely to label genuine news as false. The ability to distinguish truth from falsehood did not improve; instead, players became more cynical about all information (*Financial Times* 2024). This outcome reveals a dangerous side effect of the 'antibiotic' approach: in medicine, antibiotics kill pathogens reliably, but in the information realm, they can create widespread scepticism. While the World Health

Organization declared a worldwide ‘infodemic’ in 2020, the analogy between medical treatment and information defence ultimately falls short.

If fake news alters how people interpret real news, deepfakes directly reshape what counts as ‘real’ in the first place. These AI-generated videos, audio recordings or images can depict public figures making statements or engaging in actions that have never happened. Their strength lies in perfect mimicry: voices, gestures, facial expressions and physiognomy remain intact, but the content is fabricated to be indistinguishable from authentic material. This precision makes deepfakes powerful tools for manipulating public opinion.

Whether spread through fake news or deepfakes, disinformation undermines trust—the cornerstone of social contracts, economic exchanges and democratic life (Pinhanez et al. 2022). Repeated exposure to fabricated stories fosters cynicism, leaving people unsure of whom or what to believe. This distrust weakens institutional effectiveness by discouraging public cooperation and engagement. The damage extends to democracy itself. Fake news and deepfakes distort the flow of information that citizens rely on for informed decision-making. This was seen in Donald Trump’s false claims about the 2020 US election, which even among a small, misinformed minority had the ability to inflict significant harm, as seen in the Capitol Hill attack (Van der Linden 2023). Moreover, the exhausting task of discerning truth from falsehood can discourage participation altogether, fuelling abstentionism and weakening democratic legitimacy.

Finally, disinformation threatens social cohesion. Targeted falsehoods can incite mistrust, hatred and even violence between groups. In this sense, polarisation is not the only danger; the paralysing scepticism generated by constant exposure to both fake news and deepfakes can be equally corrosive to the democratic fabric.

So how can one draw a line between true and false, real and fake? The next and final section of this article will dive into the policy measures that have been taken to address both polarisation and scepticism, and some new solutions, aimed at tackling the very heart of the problem, will be put forward.

From regulation to education: reinforcing literacy as the foundation of information resilience

Academic and institutional actors have long been committed to developing strategies to both prevent and mitigate the effects of information disorders. Two main approaches can be identified: educational and regulatory. The educational approach seeks to empower individuals by equipping them with the skills, knowledge and critical tools necessary to approach a state of ‘immunity’ to misinformation and to recover from exposure to distorted content. The regulatory approach, by contrast, targets the structural level, focusing on digital platforms and addressing the systemic factors that facilitate the production and dissemination of manipulated information.

Let us start by delving into the regulatory approach. The EU has taken proactive steps to regulate deepfakes. The EU AI Act, for instance, requires creators of generative AI to label their content as such, making it clear that the media has been manipulated. Anticipating the potential for election interference, in 2024 the EU published the Digital Services Act Election Guidelines for Very Large Online Platforms and Very Large Online Search Engines, which outlined potential mitigation techniques to address the risks of deepfakes and disinformation (Deloitte 2025). The Digital Services Act itself is a broader policy that builds on the 2000 E-Commerce Directive. It seeks to respond to three core problems: first that citizens are exposed to increasing risks online, particularly on very large online platforms; second, that the supervision of online platforms is largely uncoordinated and ineffective in the EU; and finally, that national-level regulations risk leading to increasing barriers in the internal market and reinforcing competitive advantages for established very large platforms and digital services (OECD 2024). The EU has even gone a step further and regulated deepfakes through the Strengthened Code of Practice Against Disinformation and the Code of Conduct Countering Illegal Hate Speech. Moreover, through the above-mentioned AI Act, the EU is the first supranational institution to have regulated AI. Under the Act, the opportunity for disinformation is limited through the compliance requirements for developers of general-purpose AI models. These include the conducting of systemic risk assessments, which started in August 2025 (Deloitte 2025).

Alongside the EU's initiatives, OECD member governments are adapting their policy frameworks to counter the threats posed by disinformation. In Germany, for instance, efforts to address disinformation have been integrated into the national security strategy. Indeed, disinformation is part of the broader concept of hybrid threats, which can be defined as

actions conducted by state or non-state actors to undermine or harm democratic governments by influencing decision-making. These threats combine military and non-military, as well as covert and overt means, including disinformation, cyber-attacks, economic pressure, migration, deployment of irregular armed groups, and use of regular forces. Such actions are coordinated and synchronised, using a variety of means and designed to remain below the level of detection and attribution. (NATO 2024)

Due to the growing use of these hybrid threats, governments and institutions have increased their focus on disinformation to such an extent that the language surrounding it has become warlike, with institutional figures using expressions such as 'the fight against disinformation'. This shift has also elevated the issue to a priority on political agendas, leading to a proliferation of regulations aimed at protecting not only individuals but also national interests.

However, legislation does not address the core issue of human judgement and decision-making (Van der Linden 2023). For this reason, the final part of this section will focus on the educational approach. The main product of this approach is the concept of digital literacy (here this term is used interchangeably with the term media and

information literacy). UNESCO defines media and information literacy as ‘empower[ing] people to think critically about information and use of digital tools. It helps people make informed choices about how they participate in peace building, equality, freedom of expression, dialogue, access to information, and sustainable development’ (UNESCO 2023). Digital literacy focuses on the competences needed to live and work in a society where communication and access to information increasingly occur through the use of digital technologies (OECD 2022).

We can therefore think of digital literacy as a set of skills to develop and prescriptions to follow. The University of Cambridge has concentrated its research on the creation of ‘antigens’, maintaining the parallel with the medical realm. Here are the most relevant:

- The truth has to be more fluent. As explained in the second section of this article, one of the main reasons why the public falls victim to distorted information, especially fake news, is that it is more accessible, more quickly understood and catchier. Although difficult, because science-based content, history and legal texts take time to explain, it is necessary to make these subjects more appealing and accessible, and hence, more fluent. It is also necessary to raise awareness of the fact that lies are often presented as simple and short statements that reduce the complexity of any issue.
- The politicisation of content must be prevented. This would resolve the tension between people’s desire for accuracy and their need to be liked and accepted by members of their own social network.
- Echo chambers must be avoided and platforms’ algorithms left unfed. Despite the existence of offline echo chambers, studies show that social media networks play a critical role in their spread because the algorithmic filtering of content acts as a funnel, nudging users towards further polarisation and extremism (Van der Linden 2023).

Other researchers are focusing on the development of a new science of disinformation intended to enable the labelling of information disorder as a cybersecurity problem. This approach aims to protect hardware devices, with one idea being to create a shield to protect against fake media through software installed on a user’s device that can check the media the user is being exposed to, reacting not only by warning the user of possible issues with that content, but by providing methods to question and verify its veracity (Pinhanez et al. 2022). Such a tool, combined with educational programmes, could enhance the effectiveness of digital literacy campaigns. Users should not have to rely solely on their own judgement, which can be fallible and biased, especially when they lack the means to verify the true origin of the content. While individuals can assess the final message they see, they cannot retrace the full digital history of every piece of media. Having access to a tool that performs this preliminary verification, flagging false content before it spreads, would allow users to focus on reliable messages from authoritative sources. They could then move to the next stage of discernment: evaluating the quality

of the information and deciding for themselves whether it is sufficiently neutral to represent the facts accurately or is excessively framed to promote a particular perspective.

The proposals analysed so far are undoubtedly relevant. However, they do not address the heart of the problem. Because, as previously mentioned, people can also be vulnerable to information disorders offline, it is necessary to target the root causes of this: that is, the ability of individuals to understand the messages they receive, process them, discern their meaning and formulate a well-thought-through opinion about them, while being aware of the psychological effects that such content may have on them.

Let us start by considering the ability of individuals to understand messages. Before the advent of AI disinformation campaigns, a widely discussed topic in public debate was *functional illiteracy*. This refers to a condition whereby a person is unable to understand, evaluate, use and engage with written texts in order to actively participate in society, achieve their goals, and develop their knowledge and potential (Schlechty 2004). According to the PISA³ studies carried out by the OECD, the number of young people underachieving in reading is on the rise in the EU.

Let us linger for a moment on this phenomenon. If a person is unable to *understand, evaluate, use and engage with* a written text, how can they be expected to *understand, evaluate, use and engage with* manipulated content? This is why the main argument of this article is that the only way to tackle the problem of disinformation and to prevent phenomena such as polarisation or extreme scepticism is to return to the very roots of literacy. Individuals must be provided with the tools necessary to fully grasp the meaning of any message, not only those potentially distorted with malicious intent. Because communication today is increasingly complex and multilayered, providing invigorating text analysis and interpretation in compulsory education could make all the difference.

Some European countries have already begun promoting educational initiatives aimed at developing students' discernment and improving the ways in which they engage with the world of information. For example, Finland has introduced a National Media Education Policy (2019) that promotes media literacy across sectors, involving schools, libraries, non-governmental organisations, universities and civil society. Students study misinformation, disinformation, propaganda and misleading statistics, creating their own media for peer review (Salomaa and Palsa 2019).

While this example represents a promising approach, it still requires a foundational layer of literacy to serve as the basis for digital literacy initiatives. This is the first proposal of this article, to introduce foundational literacy initiatives. These could include the following:

- Etymology courses: understanding the origins of terms and their evolution over time enables individuals to be more precise in formulating and interpreting

messages, bringing them closer to the ideal circumstance in which each chosen word has an unambiguous meaning, reducing the space for miscommunication.

- Linguistics courses: learning the structure of language increases the sensitivity of information consumers, improving their ability to spot manipulation. An intrinsic feature of human language, absent from AI language, is the link between signifier, signified and referent—the triad described by Ferdinand de Saussure (de Saussure 2011). The signifier refers to the arbitrary sequence of sounds making up a word, the signified covers its semantic field, and the referent indicates the object or idea represented. AI cannot engage with referents rooted in physical reality, which often makes AI-generated messages feel inhuman. A shared awareness of these concepts would help people detect such messages.
- Translation courses: one principle of translation is that there is no perfect translation—something is always lost: a nuance, a sound device, a cultural reference. The same occurs when conveying information: each narrator emphasises different aspects of an event, and unconscious mental re-elaboration can alter details without malicious intent. Understanding this would encourage users to consult multiple sources to form the broadest possible view and to develop opinions that are truly their own, rather than mediated.

The strands mentioned above could form the three main pillars underpinning the literacy programme initiative.

The second proposal is to invest in psychological and emotional literacy. The second section of this article outlined the psychological mechanisms exploited by disinformation campaigns. Psychological literacy could address this problem. Psychological literacy refers to the ability to apply psychological knowledge to solve real-world problems, understand behaviour and communicate effectively. A lack of psychological literacy results in difficulties in understanding our own and others' mental processes, and in the inability to critically evaluate emotional responses or contextualising actions. A related, more established concept is emotional illiteracy. This refers to the difficulty of recognising, labelling and managing the emotions that psychological processes trigger. If programmes of psychological and emotional literacy were established, the aforementioned mechanisms (confirmation bias, illusory truth, cognitive biases etc.) would be widely recognised and more easily defeated.

Finally, a short-term measure is to raise awareness among producers of viral content regarding their responsibility for shaping the information ecosystem, both online and offline. In 2024 a Danish study found that 52% of those belonging to Gen Z had purchased a product recommended by an influencer they follow on social media (Opeep1 2024). If influencers can shape consumer choices, their potential to influence opinions and beliefs is even greater, especially among young people. It is therefore essential that this role be treated with the seriousness it deserves, with influencers adhering to clear standards and ensuring the accuracy of their messages before disseminating them.

Through this foundational scheme, we can build an educational framework that transcends time, generations and technological change, replacing short-sighted policies reacting to the latest developments with a lasting, adaptable approach to safeguarding meaning and truth.

Conclusion

Disinformation has always been part of human communication, but new technologies, such as AI and social networks, have dramatically expanded its scope. Today, the two main societal reactions to information disorders—polarisation and excessive scepticism—are often perceived unequally, with the former seen as more dangerous. Such a view is short-sighted: widespread scepticism can be just as corrosive to trust, democracy and social cohesion. While regulatory frameworks and digital literacy programmes are crucial, they remain incomplete without addressing the basic problem: individuals' lack of capacity to understand, evaluate, use and engage with information. The proposals presented here, grounded in etymology, linguistics and translation studies, aim to strengthen this foundation so that educational efforts can endure across time, generations and technological change. In the long run, only an education that goes beyond reacting to the latest technological threat can build the resilience needed to protect both truth and the public's trust in it.

Notes

1. Generated by the author with AI.
2. This term was popularised in the 2010s thanks to its appearance as the Oxford Dictionaries' 2016 'Word of the Year'.
3. Programme for International Student Assessment.

References

- Brown, J. R. & Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-021-01066-z>. Accessed 2 September 2025.
- Deloitte. (2025). *The rise of deepfakes: What digital platforms and technology organizations should know*. January. <https://www.deloitte.com/content/dam/assets-zone2/uk/en/docs/industries/technology-media-telecommunications/2025/deloitte-the-rise-of-deepfakes-january.pdf>. Accessed 2 September 2025.
- de Saussure, F. (2011). *Course in general linguistics*. Trans. W. Baskin. New York: Columbia University Press.
- Harford, T. (2024). 'It's only a matter of time before disinformation leads to disaster'. *Financial Times*, 24 January. <https://www.ft.com/content/0afb2e58-c7e2-4194-a6e0-927afe0c3555>. Accessed 2 September 2025.
- NATO. (2024). *Countering hybrid threats*. 7 May. https://www.nato.int/cps/en/natohq/topics_156338.htm. Accessed 2 September 2025.
- OECD. (2024). *Facts not fakes: Tackling disinformation, strengthening information integrity*. Paris.
- Opeepl. (2024). *Gen Z social media trends: How males and females differ in platforms and influencer impact*. September. <https://www.opeepl.com/blog/how-gen-z-males-and-females-differ-in-their-social-media-usage>. Accessed 2 September 2025.

- Pacepa, I. M., & Rychlak, R. J. (2013). *Disinformation: Former spy chief reveals secret strategies for undermining freedom, attacking religion, and promoting terrorism*. Washington, DC: WND Books.
- Pinhanez, C. S., Flores, G. H., Vasconcelos, M. A., Qiao, M., Linck, N., De Paula, R., & Ong, Y. J. (2022). *Towards a new science of disinformation*. New York: Cornell University Library.
- Salomaa, S., & Palsa, L. (2019). *Media literacy in Finland: National Media Education Policy*. Finland, Ministry of Education and Culture. Helsinki.
- Schick, N. (2020). *Deep fakes and the infocalypse: What you urgently need to know*. New York: Hachette.
- Schlechty, P. C. (2001). *Shaking up the schoolhouse: How to support and sustain educational innovation*. San Francisco: Jossey-Bass.
- Shu, K., Wang, S., Lee, D., & Liu, H. (2020). *Mining disinformation and fake news: Concepts, methods, and recent advancements*. Cham: Springer.
- Smith, R., & Liddy, B. (2024). The EU's approach to disinformation. *Deloitte*, 18 July. <https://www.deloitte.com/uk/en/services/audit-assurance/blogs/2024/eu-approach-to-disinformation.html>. Accessed 2 September 2025.
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Aythami, M., & Ortega-Garcia, J. (2020). *Deepfakes and beyond: A survey of face manipulation and fake detection*. Amsterdam: Elsevier.
- UNESCO. (2023). Empowering youth for sustainable development: The role of media and information literacy in promoting green skills. 15 August. <https://www.unesco.org/en/articles/empowering-youth-sustainable-development>. Accessed 24 September 2025.
- Van der Linden, S. (2023). *Foolproof: Why we fall for misinformation and how to build immunity*. London: Harper Collins.

Author biography



Giulia Maria Galli, currently at the Secretariat-General Interinstitutional Relations Unit of the European Commission, has worked in institutional affairs at Pirelli and holds a Master's in International Relations and a Harvard merit scholarship. She also contributes pro bono to Villa Vigoni's cultural diplomacy initiatives and is part of Young European Leadership, a youth organisation that selects EU youth delegates for international summits.