

ARTICLE

# It's not just the Web, It's the Spider as well: Prioritising Education over Structural Changes

Jelena Pavličić Cerović<sup>1\*</sup>  and Adam Nedeljković<sup>2</sup> 

<sup>1</sup>Institute for Philosophy, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia and <sup>2</sup>Department of Philosophy, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia

\*Corresponding author: Email: [jelena.pavlicic@f.bg.ac.rs](mailto:jelena.pavlicic@f.bg.ac.rs)

(Received 12 September 2024; revised 15 February 2025; accepted 5 April 2025)

## Abstract

The responsible and reliable use of web-based technologies is unquestionably crucial in the era of information abundance. Strategies aimed at mitigating false and misleading content online can be divided into structuralist and educational approaches. In this paper, we aim to demonstrate that most proponents of these strategies overlook the active role of users in web-based information and communication channels. They generally tend to view internet users as passive recipients of information who require external regulation and guidance to discern between reliable and unreliable online information in making well-informed judgments. What we aim to show is that responsible and meaningful engagement with information online encompasses much more than the existing proposals within demarcation approaches. It entails developing abilities that can be characterised as ‘third-order assessments’, which can help users not only to minimise the negative factors undermining the reliable use of information technology but also in ensuring reliable and responsible use of its communicative aspects. The paper concludes by offering suggestions on cultivating these skills and proposes their integration into school curricula.

**Keywords:** Epistemic assessment; online practice; mediated online content; norms of sharing

## 1. Introduction

Long before the widespread adoption of the internet, the issue of endurance and proliferation of misinformation such as fake news – ‘misleading information intentionally published and presented as news’ – (Anderau 2021, 210) has posed a challenge to our knowledge-seeking practices (cf. O'Connor and Weatherall 2019, 1–2; cf. Black and Fullerton 2020, 73; Manfra and Holmes 2020, 129–30).<sup>1</sup> As noted by Allcott and Gentzkow (2017, 214), falsehoods have historically permeated public discourse

<sup>1</sup>Here, we use the term ‘misinformation’ as a blanket term to encompass all types of misleading content (disinformation, fake news, propaganda, and the like) and faulty information, regardless of the intention behind its creation or transmission. However, we fully understand how adopting our suggestions for further research could serve as a basis for insisting on making distinction based on the intentions of content creators. Approaches that adopt this distinction can be found, for example, in Ramezani et al. (2019) and Black and Fullerton (2020).

through various channels within popular media. However, the emergence and utilisation of new information and communication technologies, especially online platforms such as social media (Rini 2017, 43; Marin 2021, 363; Millar 2019, 525–26), are seen as catalysts for the unparalleled spread of misleading content to a significantly broader audience. Without delving into the question of how acute and serious the problems of misinformation proliferation and fake news consumption are in the digital space (for more on this, see: Lazer et al. 2018; but also consider: Altay et al. 2023), it is undeniable that the internet possesses certain characteristics that facilitate the spread of false, misleading, and dubious content. Such ‘epistemically toxic content’ (a phrase borrowed from Record and Miller 2022a) can be mediated, both in terms of perpetuation and diffusion, not only through technology but also by the agents themselves.

By technologically mediated digital content, we mean the content that is suggested or delivered to users through algorithms, such as those employed by search engines or social media platforms. With these mechanisms, the following phenomena are associated: (a) personalised searches (Simpson 2013; Mößner and Kitcher 2017; Black and Fullerton 2020), (b) page rank (Höchstötter and Lewandowski 2009), (c) autocomplete technology (Miller and Record 2017), (d) sponsored content (de Villiers-Botha 2022), and (e) information or filter bubble (Nguyen 2018). For example, the results we see and click on during our Google searches are technology-mediated, as they are selected according to the PageRank algorithm, which lacks transparency (see Brin and Page 1998, 109ff.; de Villiers-Botha 2022, 331–32) and can be influenced by sponsored content (de Villiers-Botha 2022, 330). Similarly, the autocomplete technology (Miller and Record 2017) nudges users to search for specific results that are also personalized (Simpson 2013, 427; Black and Fullerton 2020, 78). Personalization refers to the curation of content that users are exposed to, based on profiles constructed by algorithms. This carries the risk of sheltering users from diverse information and potentially exposing them to unreliable content simply because it aligns with their past searches (see Mößner and Kitcher 2017). In this way, it is argued that the interaction of algorithms can create the epistemic bubble in which the user is situated (Nguyen 2018, 4).

Agent-mediated content refers to the content that arises because of certain user activities that may lead to the generation of epistemically toxic content. Some of these activities can stem from selective exposure of users to a specific circle of sources, perhaps those who share their core beliefs, leading to so-called echo-chambers (Nguyen 2018), while others may result from unregulated sharing practice (Rini 2017). While echo chambers actively isolate users from typical sources of dissenting opinions, unregulated sharing norms allow them to evade responsibility for their sharing behaviour.

Given the negative impact that these types of mediation have on the reliability of online content, it is unsurprising that in recent years, there has been a growing tendency to provide a deeper understanding of trustworthiness of the various agents and technologies engaged in online information dissemination along with the formulation of strategies for dependable and responsible utilisation of information technologies in seeking advice grounded in sufficient evidence (e.g. Black and Fullerton 2020). If we follow the classification proposed by Michel Croce and Piazza (2021), these endeavours can be categorised into two approaches: the educational and structural interventions (distinction also made in Lazer et al. 2018, 1095). Advocates of the educational approach (e.g. Polizzi and Taylor 2019; Croce and Piazza 2021) argue that the lack of resources and strategies for evaluating the quality of online content (sometimes described as digital media literacy [Guess et al. 2020] or information/media literacy [Black and Fullerton 2020, 85; Lichtenberg 2021, 19]) renders users susceptible to social media misinformation. Refining these skills, perhaps through engaging in fact-checking (Nygren et al. 2021; Eisemann and Pimmer 2020), broadening sources of information

(Croce and Piazza 2021) or developing strategies to identify trustworthy (Polizzi and Taylor 2019) or reliable information online sources (Wiley et al. 2009; Herrero-Diz and López-Rufino 2021), is believed to be crucial in improving individuals' knowledge seeking online activities.

On the other hand, proponents of the structural camp believe that the inherent structures and specific features of web-based technology, along with the dynamics of users' interaction with it, place them in a fairly disadvantaged epistemic position that cannot be significantly improved by altering their individual epistemic performances (deçVilliers-Botha 2022; Millar 2019, 2021). In their view, effective interventions aimed at enhancing current information-seeking practices should entail structural changes within the information environment itself (Rini 2017) so that it promotes truth (Lazer et al. 2018) and well-validated claims (cf. Levy 2024). In the following two sections (1.1 and 1.2.), we will look more closely at these proposals and provide an evaluation of the structuralist interpretation of the internet as an epistemic practice, including their understanding (or lack thereof) of the roles and positions of internet users within it. We will argue that substantial improvements in the reliable utilisation of web-based informational and communicative technologies stem not from reforms at the technological level, but rather from improving users' epistemic performance.

However, as we will see further, proposed educational interventions aimed at this goal mostly focus on developing strategies for users to discern reliable information and its corresponding sources. In section 1.3, we aim to show that while such demarcation strategies can be beneficial, they are only part of the solution. It is necessary to understand the factors that mediate their usage and diminish their effectiveness to unlock their full potential. This, as we will see in sections 1.4 and 1.5., necessitates establishing different assessment dimensions for users to evaluate the circumstances under which they acquire beliefs they deem as knowledge, as well as their own influence on the epistemic climate. In other words, we will illustrate the necessity for users to comprehend both their passive and active roles, as well as the relationship between their intentions to share and their ability to assess the accuracy of content. Understanding their own epistemic limitations and the external factors that hinder their cognitive efforts is crucial for improving their epistemic position in such a context. Finally, we recommend that social studies teachers and other practitioners actively discuss and integrate this approach as the foundation of a more comprehensive education program aimed at promoting the more reliable use of web technologies in knowledge-seeking practices.

### **1.1. The motivations behind a structural approach**

As mentioned in the previous section, some proponents of the structural camp believe that users' reliance on web-based information and communication channels cannot be effectively improved by educational interventions. For example, Boyd Millar (2021, 9) believes that potential enhancements through certain educational programs would be notably incremental – so minimal as to render almost any educational intervention scarcely justifiable in terms of both cost and exertion. As he argues, the inherent nature of web-based technologies exploits certain human epistemic defects such as uncritical acceptance of information congruent with prior beliefs (confirmation bias) and susceptibility to (illusory) truth effect. Truth effect refers to the fact that people are more willing to believe a piece of information if they encounter it repeatedly (see Hassan and Barber 2021). This applies to falsehoods as well, which are ubiquitous on social media (Millar 2019, 528). In this way, social media traps individuals in a bubble that shields them from contradictory evidence that could otherwise challenge their beliefs (Millar

2019, 529–30), making it unrealistic to expect the average user to overcome these limitations on their own. Therefore, Millar proposes structural changes, which could involve ‘some combination of laws, regulations, and incentives to create an information environment that ordinary human beings can navigate without regularly forming outlandish false beliefs’ (Millar 2019, 534). In a similar vein, Tanya de Villiers-Botha argues that users are not epistemically blameworthy for adopting false beliefs via Google searches (de Villiers-Botha 2022, 336–37). Since Google lacks transparency about the reliability markers it uses to recommend results, and users have no insight into the parameters behind it (de Villiers-Botha 2022, 333, 335–36), they are mostly blameless for acquiring false beliefs through the search engine (de Villiers-Botha 2022, 336–37). Similar reasoning can be ascribed to Regina Rini, who asserts that social media users cannot be criticised for believing in falsehoods if their belief results from assigning greater credibility to a testifier who is a member of the same partisan network as they are (Rini 2017, 54). The reason for this is, as she argues, that relying on the testimonies of like-minded people is expected and individually reasonable.

For these reasons, proponents of the structuralist approaches believe that the primary focus in ensuring safer and epistemically optimal navigation through online spaces should be on systematic, (infra)structural, or institutional changes. These changes could involve implementing various strategies, including deliberate regulation (Levy 2024) and legislative frameworks (O’Connor and Weatherall 2019, 182–83), labelling (Vosoughi et al. 2018, Ramezani et al. 2019) and removing false news (Bronstein and Vinogradov 2021; Lotto et al. 2023), greater transparency from companies regarding the functioning of their algorithms that underlie personalization and data usage (de Villiers-Botha 2022; Lotto et al. 2023), modifications to the platform itself such as nudging (Thornhill et al. 2019), diversifying news feeds (Huber 2020) and website content (Sunstein 2007, 208–9, 193–95 in Anderson 2011, 158) and calculating and displaying publicly a Reputation Score for individual users (Rini 2017). In the following, we will attempt to show that while some of these proposals suffer from serious shortcomings, others can achieve their purpose only in combination with carefully designed and compatible educational strategies.

## 1.2. *Evaluating key strategies of the structural approach*

As we indicated in the previous section, proponents of the structural approach view the internet as an environment that presents significant challenges and offers limited technological possibilities for users in terms of navigation and mastery.<sup>2</sup> For example, Millar says that ‘the human brain was designed by natural selection to operate best in a specific range of circumstances, and our hardwired cognitive tendencies make it especially difficult to avoid false beliefs in the current information environment’ (Millar 2019, 528). Similarly, when it comes to social media activities, there are authors who believe that they possess characteristics that diminish users’ inclination to engage in critical thinking or fact-checking procedures (Rini 2017: 43). Moreover, given the limitations on users’ resources (epistemic, motivational, practical, etc.) and the inherently unfavourable conditions fostered by social networks, it is unreasonable to expect users to track and verify or every single claim and its corresponding sources (Rini 2017: 54; de Villiers-Botha 2022, 337; Millar 2019, 531).

It is undeniable that structuralists are correct in asserting that expecting internet users to combat misinformation alone by tracking down and investigating the trustworthiness

<sup>2</sup>According to Isaac Record (2013), an action is considered technologically possible for an individual if they have both the material and conceptual means to carry it out.

and reliability of all sources is highly unrealistic, given constraints of time, energy, motivation, attention, competence, skill, and education. However, we are sceptical that fact-checking techniques and myth-busting interventions would be any more effective if entrusted to third-party initiatives. In the following discussion, we will outline several reasons why fact-checking, regardless of its form, is inadequate for ensuring the reliable use of web-based technologies in information-seeking practice.

(a) Problems with fact-checking

There are different methodologies for the fact-checking process. One, illustrated by Steensen and colleagues (2023) refers to the process which addresses the need for speed in the verification process by ensuring the accuracy of claims made during live political debates or discussions. In this process, most of the material fact-checkers use is prepared in advance, often sourced from official channels. With pre-approved sources in place, any claim deviating from them is likely to be judged as false. It's easy to see how this methodology, if widely adopted, could be used to 'silence marginal voices' (Record and Miller 2022a). Since in this approach fact-checkers rely on claims already widely accepted as true and on the existing 'hegemonic view of what constitutes important and reliable information' (Steensen et al. 2023, 15) it may push them toward a confirmatory epistemology (as noted by Steensen et al. 2023). This is undesirable because it can promote a single, uniform perspective – a kind of 'hivemind' epistemic conformity – that runs counter to the fundamental human desire for curiosity and exploration. We will explore the drawbacks of this further in the paper.

A different type of fact-checking procedure is illustrated by Lucas Graves (2017), representing traditional fact-checking where most of the effort takes place after a claim has been identified as newsworthy. The process begins with selecting a claim based on its newsworthiness or significance, followed by contacting the source (if possible). The claim is then traced using a search engine or news database, experts are consulted, and finally, the work is reviewed by editors who vote on it (Graves 2017, 524–28). However, given the abundance of misinformation on social media and ubiquity and the fast pace of transmission of false content (Vosoughi et al. 2018), it is not realistic to expect that every or even most of the potentially dangerous and false claims can be given such a thorough examination and due process as described by Graves (2017). This kind of fact-checking may have to compromise between epistemic rigour (being thorough and critical) and being 'efficient' in terms of the number of fact-checked claims, while still dealing with the unclear methodology and potential latent bias.

Chloe Lim (2018, 6) found that fact-checkers do not agree on selection criteria and that they tend to disagree with each other on the truth value of the claim if the claim is ambiguous. And when they do agree that a claim is false, it is not clear what they find objectionable about the claim, the context, or a certain detail (Uscinski 2015, 3, 4). Researchers such as Uscinski and Butler (2013) and Uscinski (2015) further highlight this issue, pointing out the lack of selection criteria for determining which claims should undergo fact-checking (Uscinski and Butler 2013, 164ff.), leading to biases prevailing in the fact-checking process. Even though fact-checkers are entrusted with a very serious task, we still need to be aware that they are human beings, who are not always error-free and may be biased (see also Markowitz et al. 2023, 13).<sup>3</sup> Ideology can significantly influence how fact-checkers perceive the world and what they consider to be significant, credible, false, or true, regardless of their professionalism or background (Uscinski 2015, 5). They also highlight fact-checkers' tendencies to fact-check statements that are inherently difficult to verify, such as those about the future or

<sup>3</sup>To consider a response to some of the arguments against fact-checking presented here see Amazeen (2015) and Graves (2017, 520–523).

causality (Uscinski and Butler 2013, 170), and to analyse statements in a manner that strips them of context (Uscinski 2015, 2). Therefore, effectively monitoring many claims made on social media and deciding what to delete (as proposed by Bronstein and Vinogradov [2021]), constitutes challenging work that requires time. By the time the relevant claim is adequately checked, the errors and falsehoods may have already become irrelevant or second nature. Because of these and similar reasons that make fact-checking technologically unfeasible for individual users (more on this in Mattioni 2024), some propose modifying the epistemic environment itself – such as assigning the demanding task of fact-checking, or at least part of it, to algorithms (for an overview, see Graves 2018).

However, there are similar problems with automated fact-checking as well. As Vosoughi and colleagues (2018) argued, false claims ought to be labelled as false to discourage users from sharing, and in the case of automated fact-checking, AI fact-checking systems have to decide, based on evidence, that a claim is false, and false in a significant sense. But to succeed in that, reliable fact-checking requires understanding things like meaning and context (Graves 2018; Uscinski and Butler 2013, 167–68; for the importance of context see Record and Miller 2022b). Understanding context is extremely important in cases of ‘malinformation’, which involves presenting a true statement in a misleading way. For example, in an attempt to criticize a politician, stating that ‘he increased the deficit in 2022’ may be technically true, but it lacks crucial context. The deficit primarily increased to support the population against soaring energy prices – a policy any reasonable leader might have pursued to some extent. It is difficult to determine definitively whether a statement – especially a political one – is entirely true or false, or strictly black or white, as it may fall within shades of grey (Graves 2018). Furthermore, some claims do not lend themselves to precise and distinct criteria for evaluating their accuracy or truthfulness. For example, metaphorical descriptions, due to their suggestive and open-ended nature, are challenging to verify using specific, clearly defined conditions (Williams 2001, 92). AI may confuse sarcastic statements or irony for fake news or misinformation since it is not developed enough at this point to reliably recognize sarcasm and satire (Santos 2023, 683; Zhang et al. 2024). We can express sarcastic attitudes with a mere shift in inflection and ‘sarcasm detection is often considered a holistic and non-rational cognitive process that does not conform to step-by-step logical reasoning’ (Zhang et al. 2024, 10; see also Wikipedia Contributors 2025a). Although progress has been made in developing algorithms that can detect sarcasm in speech, much more work is needed to develop effective multilingual sarcasm detection. It is important to note that AI systems are typically limited to English and a few other languages, whereas misinformation is not confined to any particular language – potentially compromising their effectiveness in fact-checking (Bontridder and Pouillet 2021, e32–11; Quelle and Bovet 2024, 12). We also need to acknowledge the use of the so-called ‘algospeak’, which refers to the use of language in such a way as to avoid triggering social media algorithms (Klug et al. 2023; Steen et al. 2023). For example, certain words that could potentially trigger AI’s reaction are intentionally spelled incorrectly or a rhyming word is used as a substitute for an intended word, or even a picture of a rhyming word. Slang can evolve too quickly for AI to keep up with.

Wondering whether something is a fact can be out of place because we do not always communicate facts or care about being factual. Whether some piece of information is factual or not can be beside the point in some cases, like when pointing out what someone else has said (Marin 2021, 365) or when exchanging jokes. Consider, for example, the joke: ‘You ever wonder how trains eat? They choo, of course’. Now, imagine we shared this as a caption for a photo of us eating a sandwich on a train. Then, imagine AI fact-checking our post by stating that trains do not eat because they are machines, not



living beings, and are powered by steam, electricity, or diesel. We would consider such an action to be way off the mark. Although this is a simplified example of a joke that an AI-driven humour detector would easily recognize, failing to detect more sophisticated jokes could lead to over-labelling them as misinformation. Given that a significant portion of structural approaches heavily rely on the simple false/true distinction, concerns about the effectiveness of fact-checking on social media in mitigating the spread of misleading content (Wasike 2023) are not surprising.

Additionally, AIs are always trained on certain datasets, which may contain biased (Papageorgiou et al. 2024, 23) unreliable, misleading, or outdated information – a phenomenon known in the literature as data poisoning (Alber et al. 2025). Even a small percentage of misinformation can affect an algorithm's reliability, and it's important to remember that even the most sophisticated databases may contain misinformation or outdated information (Alber et al. 2025). The situation seems even darker if AI used unfiltered information to train. Faulty information can also appear in the training data as a result of a cyberattack (Eddy 2024). But even with training based entirely on accurate data, generative AI models can still produce hallucinations, generating potentially inaccurate content by unpredictably combining existing patterns (Weise and Metz 2023). The question is how willing we are to entrust such a critical task – assessing the reliability of the content we rely on – to systems that are prone to hallucination.

Also, AI can exhibit political bias, as demonstrated by Rozado (2023), due to the inherent biases of the humans who train it, which can, inadvertently influence the algorithm. Moreover, as the world changes information about the world changes too, but most training data are static and not updated constantly, which makes AI unreliable because it cannot keep up with the ever-changing information landscape (Papageorgiou et al. 2024, 23–24). Additionally, if we were to permit AI trained on predetermined authoritative data sets, we would be putting in danger dissenting voices or even just plurality of voices and opinions (Marsden and Meyer 2019, 45 in Bontridder and Pouillet 2021, e32-11) which were, as history shows us, very important in the pursuit of truth.

Finally, it should be pointed out that AI can, in effect, play a negative role in shaping our information environment, as it can be easily and cheaply used to create misinformation by manipulating various forms of digital media content, such as audio and video (e.g. AI-generated fake content or 'deepfakes') (Bontridder and Pouillet 2021, Section 2.1). This is quite concerning, as large generative models, utilizing pre-trained networks with diffusion or transformer frameworks and vast datasets, can perform multi-modal tasks and generate high-quality, convincing content (Xu et al. 2023: 9292). When it comes to the transparency of content regulation, we should keep in mind that the inner workings of AI systems are, to a great extent, epistemically opaque to us. If AI is to be used as a tool to differentiate truth from fiction, it must be understood by its users (Hsieh et al. 2024, 11). AI is sometimes described as a 'black box' due to its obscurity and high complexity. Models are extremely intricate, with a tremendous number of parameters (Hsieh et al. 2024, 12), which may compromise users' trust in them (Mishima and Yamana 2022, 1249). Although efforts to improve AI explainability are underway, this field of research is still in its infancy, raising doubts about how well we can understand these processes and how capable AI is of explaining itself to us. Some authors argue that the inner workings of AI algorithms, which operate independently, may be beyond our comprehension (Matthias 2004; Humphreys 2009; Mittelstadt et al. 2016; Koskinen 2024). Thus, even if many of the AI performance limitations we have highlighted were to be overcome, we would still be left with the issue of interpretability, as users trust in a model largely depends on how effectively they understand the reasoning process and its decisions (cf. Mishima and Yamana 2022).

(b) Transparency and diversification

As indicated previously, the term ‘structural approach’ is an umbrella term that encompasses various ideas with the information environment at its core focus. Structurally changing the environment is not necessarily equivalent nor conducive to censorship. That does not mean that there are no researchers calling for censorship of social media. For example, Bronstein and Vinogradov (2021, 1) call for the ‘deletion of false and dangerous information’ and hail censorship as a valuable tool in fighting online health misinformation.

But if the overarching institution censoring dissenting voices is the same one propagating misinformation, who will keep that institution in check (cf. Niemiec 2020)? What constitutes false and dangerous content? And who makes the call on this? And, given the extensive presence of epistemically toxic content online, how are we supposed to manage their overwhelming numbers? It’s essential to recognize that entrusting truth determination to government or private companies grants immense power to already powerful entities, which may not prioritise truth when it is inconvenient (cf. Record and Miller 2022a). In such cases, these entities may vaguely define what they deem dangerous or misleading content (Niemiec 2020) and misuse it to suppress criticism. It can also happen that media platforms treat facts as secondary, spreading misinformation without regard for the truth of what they communicate (i.e. bullshitting) due to a special collaborative relationship between the state and the media, where parties work together in mutually beneficial partnerships (Gibbons 2024). Additionally, this attitude may be driven by state pressure or financial incentives from corporate sponsors, financial partners, and other private entities. As Adam Gibbons (2024, Section 2.2) notes in his interesting incentives-based analysis of bullshit in politics, ‘as long as there are sufficiently many who will reward the media even when they bullshit (whether by tuning in, paying for subscriptions, or what have you), bullshitting will remain rational’. By the same token, the economic structures that support key infrastructure elements like search engines, databases, and social networks can silence certain voices or even overshadow them entirely (Mößner and Kitcher 2017, 30). As Mößner and Kitcher emphasize, while it is true that the internet, as an epistemic environment, allows all perspectives to be expressed, it cannot be said that all perspectives are equally heard. But dissenting voices are crucial for advancing our epistemic practice (de Melo-Martín and Intemann 2012). The world is complex and uncertain, and just as previously accepted justifications can be undermined by new evidence (Williams 2001, 161) what is believed wrong today may prove true tomorrow.

Taking these considerations into account, the question arises: what stance can we adopt regarding structuralists’ proposals, such as diversifying users’ news feeds (Huber 2020) and employing nudging techniques on the systematic level like presenting original posts alongside alternative views (Thornhill et al. 2019)? Although interventions of this kind sound promising at first glance, the question is how effective such measures would truly be. As pointed out by Duncan Pritchard (2013, 237), access to information is futile if individuals lack the cognitive abilities to interpret it effectively. Without these abilities, it is far from obvious whether increased access to information sources, and consequently exposure to unreliable content that could otherwise be avoided, would render users better informed.

Granted, the question arises whether these interventions can be solely entrusted to media algorithms and Big Tech companies. Just as in the case of human fact-checking procedures, relying on the government, corporations, or technology itself to regulate information flow through diversification is not only questionable in terms of whether implemented mechanisms will contribute to greater knowledge production but also too risky as these entities may not want or be able to prioritise truth. However, what structuralists propose and what we believe can truly help users achieve greater



technological possibility for epistemic vigilance (Mattioni 2024) is greater transparency of companies regarding their operations and data usage (cf. Lotto et al. 2023; de Villiers-Botha 2022). Such changes are always welcome, not only because they expand users' knowledge horizons and control, but also because we believe that, in the context of social media, these strategies can help restore users' already damaged trust in certain platforms (Huber 2020, 41; Spence 2021, 3). On the other hand, one must be realistic about the extent to which demands for transparency can be met, if for no other reason than the following: (a) companies have financial interests in protecting their business models from competitors; (b) explanations of technical processes – such as algorithmic filtering procedures, targeting systems, and advertising practices – are often lengthy and filled with technical jargon, making them difficult for laypeople to understand; and (c) ultimately, the complexity and rapid evolution of modern machine learning models make it challenging even for their designers and companies to develop standardized evaluation measures that can reliably assess their performance and decision-making processes.

So even structuralist proposals advocating for technology-level source diversification and transparency should be approached with caution and limited expectations, as none of them is comprehensive. One of the reasons why this is the case lies in the fact that most proponents of the structural approach focus solely on the inherent characteristics and qualities of digital infrastructures, neglecting the influence of user interactions on the online epistemic climate. In other words, most of the structuralists have an overly simplistic view of the human user, social media, language games played there and the internet in general. They view it as a single-level field churned up with both misinformation and genuine information, through which a subject with several orientation-related imperfections attempts to navigate, often stumbling and getting hurt along the way. As the subject wanders through the treacherous landscape, obstacles need to be either removed or, in some cases, established so that they travel in a 'correct' path. The only viable way to help it is to shape the environment around it. Thus, one accusation that can be levelled against structuralist interventions is the denial of the (epistemic) agency of the user (cf. Lazer et al. 2018, 1096), which may be attributed to the current methodology of researching the user's online activities, usually accomplished by accumulating and analysing vast amounts of data. The consequence of this is the *datafication of audiences* (Livingstone 2019). When we reduce users of media to mere data (and there is a lot of data about their online activities), they look like undifferentiated abstract objects and not individuals. Consequently, we lose a great deal of information about their motivations, interpretations, and concerns (Couldry and Kallinikos 2017 in Livingstone 2019).

But, when we remove ourselves from this intellectual oversimplification, we can see that the internet resembles a maze of vast size and complexity with different segments (or rooms) in which different games are played and in which different rules are followed, and at least one source of these rules is the human user himself (see also Marin 2020, 2021). It is the active and creative role of the user that some structuralists fail to see. Even structuralists, like Rini, who acknowledge the active role of social media users in shaping digital epistemic landscape, exonerate the need for individual-level alterations in social media sharing practice by focusing on technology-level reforms as a sufficient solution. Namely, as an effective strategy in reducing propagation of epistemically toxic content online Rini (2017) proposes that Facebook calculate a Reputation Score for individual users based on the frequency with which each user chooses to share disputed stories. Setting aside the myriad problems associated with verifying the accuracy of shared content that we discussed earlier, this proposal only addresses symptoms rather than the underlying cause of the issue. It still leaves room for users to manipulate the content; for

instance, profiles with previously attained high Reputation Scores could be used to engage in selective and misleading presentations of well-supported facts and the spread of *malinformation*.

So, in tailoring any kind of reforms or interventions, it's crucial to bear in mind that users are often the ones who significantly contribute to the propagation of epistemically toxic content through reposting or sharing (cf. Marin 2021, 363). That users play a crucial role as both witnesses and sources of factual reports is an aspect acknowledged by some proponents of the educational approach (Black and Fullerton 2020). Yet, even for them, the main focus remains on users perceived as receivers of testimony, raising pressing questions about what kinds of information sources to consult, what evidence-gathering procedures to consider, and how to devise practical and easy-to-use guidelines for discerning reliable/legitimate information/voices from unreliable/illegitimate ones. In the next section, we aim to show that not only are these strategies insufficient for identifying information firmly grounded in the best available evidence but also for understanding and facilitating users' meaningful and epistemically useful interaction with online content.

### 1.3. Educational approach

Given the enormous influence that technology has on the ways in which knowledge is acquired in today's everyday practices, numerous research studies have emphasised the importance of developing media literacy/digital skills and introducing 'critical thinking' courses into school programs. The current work reflects a fluid terminology around the question of what constitutes literacy and critical thinking in the digital space. However, it seems that most authors imply by these terms the practice of something that can be described as *epistemic vigilance*. Epistemic vigilance refers to a set of mechanisms that can be readily deployed to continuously calibrate trust (cf. Mercier and Sperber 2017).

Despite considerable debate regarding the precise nature of trust, there is widespread recognition that in deciding whom to trust and identifying trustworthy information, both epistemological and ethical considerations are equally important (Intemann 2023, Section 1). That is to say, the abilities required to evaluate the credibility of various information sources involve two key aspects. First, the ability to evaluate the strength of the given evidence and the validity of relevant claims. Second, it involves assessing whether the information provider is benevolent and honest and has moral integrity (Anderson 2011, 145; de Melo-Martín and Intemann 2018, 90), and whether they maintain an appropriate attitude toward the potential epistemic consequences of their work (Intemann 2023, Section 1). However, achieving the former requires a significant level of competence in terms of skill and a genuine understanding of the relevant issue from the trustor themselves, something they often lack in many domains. For this reason, philosophers have proposed criteria that focus on the credibility of the testifier (i.e. potential trustee) rather than the trustworthiness of their assertions (Levy 2024). Acknowledging the challenges that lay people face when it comes to evaluating evidence and the merits of scientific research, Elizabeth Anderson (2011, 145) argues that laypersons have 'second-order capacity' to judge trustworthiness of information sources in making well-informed decisions, emphasising that individuals decide what to believe by evaluating whom to believe. This decision depends on three types of assessment. The first is an assessment of the expertise or competence of the agents who create, endorse, and distribute the relevant claim. The second is an assessment of their honesty, and the third is their responsiveness to counterarguments to their belief. The guidelines provided by Anderson can certainly be useful, but in cases of persistent substantive disagreements between apparently equally honest parties, they may be of limited significance. As Neil

Levy (2024) pointed out, traditional criteria for evaluating expertise – such as credentials and institutional affiliations – can be useful, but they often fall short in persistent disputes. In many cases, both sides of a debate can present impressive credentials, leaving laypeople struggling to identify the more trustworthy source. Levy cites figures like Frederick Seitz, a former President of the National Academy of Sciences, who lent his support to the ‘merchants of doubt’ on issues such as tobacco and climate change. Similarly, Robert Malone, a figure embraced by critics of COVID-19 vaccination, promotes himself as the inventor of mRNA technology, giving his position credibility despite the controversies surrounding it (Levy 2024). In addition to experts who may intentionally promote false narratives, some unintentionally overstep the boundaries of their competence when offering advice – boundaries that laypeople often cannot clearly discern. Furthermore, online content found on blogs, social media, and forums, frequently relied upon by internet users, may come from anonymous sources, making it impossible to verify their credibility (Heersmink 2018). In reality, when two experts disagree, only in rare high-stakes situations might we expect the average internet user to thoroughly examine their track records, assess the merits of their academic achievements, or search for signs of dishonesty in their research and advising history. However, even then, if they are presented with a convincingly clear or familiar explanation that dismisses the opposing side as corrupt, most people will not realize they’ve been misled.

When it comes to the second criterion, honesty, it is important to remember that testimony recipients tend to view testifiers who align with their own views – whether those are pre-existing beliefs, political affiliations, ideologies, or values – as honest, well-intentioned, and reliable (cf. Rini 2017; Braman et al. 2005). In that regard, it could be said that individuals not only decide what to believe by evaluating whom to trust, as Anderson emphasizes, but also that ‘who you believe depends on what you believe’ (Herzog 2006, 105 in Hardoş 2018, 276). For example, an individual’s perspective on existential risks associated with AI – such as concerns about its potential to pose an existential threat to humanity – may reduce their confidence in the competence of AI researchers to discuss the associated risks and benefits. Specifically, individuals with a strong fear of AI may distrust experts who present testimony that challenges their readiness to view advancements in AI as safe. A person’s previous beliefs play a significant role in determining how they attribute epistemic authority to others (Hardoş 2018, 276), while the way we perceive those we trust is often emotionally charged (Furman 2020). Given that individuals naturally consider their own judgement as truthful, they tend to see themselves as following the standard of relying on reliable sources when they, in fact, only interact with information that mirrors their existing beliefs (Worsnip 2019, Section 2). Having these tendencies in mind, it becomes clearer why the dishonesty behind public scientific testimony is difficult for laypersons to uncover. Here, we cannot expect much help from the third line of assessment introduced by Anderson, which is calling on the testifier’s responsiveness to counterevidence and counterarguments. Of course, in situations where a testifier fails to respond rationally to counterarguments and merely repeats their claim, laypeople can recognize that the testifier does not adhere to the standards of dialogic rationality and dismiss them as untrustworthy. But we can expect the parties in a debate to arm themselves with much more sophisticated arguments and persuasive tactics, some of which offer a subtly manipulative yet unified and easy-to-accept explanation for the disagreement. If these debates are filled with technical terms and epistemically opaque jargon, testimony recipients are left with few options in forming their judgment. This judgment will either be shaped by their prior beliefs, by what they can somewhat recognize as aligning with their value system, or by whatever evokes in them what C. Thi Nguyen (2021) describes as a ‘sense of clarity’. It is precisely the allure of clarity, as Nguyen argues, that can

influence our judgments of expertise, leading us to accept an explanation and view its proponents and creators as more credible.

So, it is questionable whether an effective assessment of the reliability of information and its corresponding sources can be based solely on credentials and assessment of trustworthiness of information sources. Keeping this in mind, opponents of the educational approach might argue that second-order assessments of the trustworthiness of online sources are precisely the ones that pose problems in our quest for knowledge as they are either not readily available for implementation or are ineffective in cases in which previous beliefs and predispositions significantly influence one's decision on whom to trust. Moreover, second-order assessment of honesty can reinforce practices that prioritise sources with a partisan orientation, perpetuating epistemic bubbles.<sup>4</sup> As mentioned earlier, individuals within these bubbles lack exposure to diverse information and arguments, hindering their ability to verify accuracy or validity (Nguyen 2018), which is precisely what second-order assessments are intended to ensure.

Furthermore, there's a concern that these demarcation guidelines disregard practice of the contribution of lay expertise to proper understanding of different phenomena (more on such contribution in: Wynne 1998) and, related to that, the successful involvement of public opinion in science (more on this: Bedessem and Ruphy 2020). Like the fact-checking procedure, as Noortje Marres (2018, 428–29) effectively summarises, demarcation strategies risk reinforcing stereotypical divisions between the discerning and the non-discerning, exacerbating the dichotomy between knowledge-capable individuals and others. Nevertheless, in a democratic society, citizens have a legitimate right to critique scientific findings and recommendations presented to them if they perceive them as lacking certain values or incorporating detrimental ones (de Melo-Martín and Intemann 2018, 126–27), and as emphasised by Pavličić and colleagues (2023), there are certain epistemic benefits in terms of scientific discoveries and the provision of reliable testimony when expertise boundaries are transcended.

Does that mean that we should abandon the pursuit of a satisfactory resolution to the epistemological challenges confronting reliable usage of web-based technology? Some might infer from our argument against the fact-checking procedure that we should discard the distinction between claims supported by solid evidence and those that lack such support, while our argument against credibility indicators suggests that we should disregard expertise and not designate any voice as inherently illegitimate. But that is not the case. Science stands as our most reliable source of empirical knowledge, and the concept of equal epistemic status – suggesting all individuals and internet sources are equally reliable in terms of knowledge and information – is a myth (Mößner and Kitcher 2017, 4). Most of the points outlined above have been discussed in the context of our everyday fallible, complex, and ever-evolving epistemic practices, where there is a need to leave room for reforms that recognize the importance of marginalised voices. Like any other practice – whether everyday, scientific, legal, etc. – online interaction practices are characterised by specific characteristics and limitations that can lead to epistemic harms. Following some proposals we aim to emphasise the importance of understanding these limitations (Lichtenberg 2021, 17: see also: van Dijck 2010). Gaining insight into these mechanisms and understanding the nature of the underlying processes behind our online activities – as well as the various political, social, and psychological factors involved in the production, distribution, and acquisition of online knowledge – is crucial for improving our epistemic performance and engaging more effectively in the exchange

<sup>4</sup>We are grateful to our colleague Aleksandar Prica for raising this point during a roundtable held at the Faculty of Philosophy, University of Belgrade, in 2023, where we sought to present some of the redundant of the themes and questions addressed in this article.

of ideas. How this can be achieved will be explored, to some extent, in the next section of our paper.

#### **1.4. Toward more comprehensive education approach**

As previously highlighted, people today generally have access to a much wider range of information than in pre-digital times. However, the question remains: does this greater access make them better informed? One may still find it challenging to integrate conflicting viewpoints into a cohesive perspective and develop an adequate understanding. So, although exposure to a variety of information is necessary, it alone is simply not enough. Students should understand why diversifying – considering alternative viewpoints and widening information sources – is important, what the limitations are, and accordingly, what can be achieved over time. Without this ability, it is difficult to see how diversification alone could ‘ensure that social media users get better at distinguishing good from bad news’ (Croce and Piazza 2021, 10). One of the epistemically unfavourable consequences of exposure to more varied information sources and even-handedly reporting may result in users’ wrong perception that epistemically unequal justified claims are equally justified (cf. Anderson 2011). And, when people step out of their epistemic bubbles, they may still encounter false claims, perhaps even a greater number.

On the other hand, we believe that mindful engagement with a variety of information, accompanied by training to discuss controversial political issues (Kahne and Bowyer 2017), can be a valuable tool for students to make well-informed decisions based on the best available evidence. But what do we mean by mindful engagement with information? One cannot be able to assess their own responsibilities and competencies until they reach a certain level of willingness and ability to critically evaluate their own beliefs (cf. Pongiglione 2021, 72).<sup>5</sup> Building on Elizabeth Anderson’s (2011) approach, this broader perspective – which includes an understanding of one’s own psychological tendencies and cognitive limitations (cf. Russo and Schoemaker 1992, 8) as well as their exploitation by algorithms and business models (Niemiec 2020; Maréchal and Biddle 2020; de Villiers-Botha 2022) – can be formulated as a ‘third-order assessment’. Understanding the underlying mechanisms of information-gathering procedures, both internal and external, psychological and algorithmic, is the necessary step toward a deeper understanding of our own epistemic position from a proverbial bird’s-eye view. Many proposed strategies to mitigate the proliferation of misleading content overlook these mechanisms, as they often focus solely on evaluating the reliability of online information sources – essentially, on what happens *on the screen*.

##### **1.4.1. On the screen**

Some researchers argue that educational interventions about fake news detection ought to take place as soon as children start to use media (Eisemann and Pimmer 2020). Certainly, teaching users of recognizing credibility markers is one of the initial steps in addressing the problem highlighted by the study conducted by the Stanford History Education Group. This study found that young people struggle to assess the reliability of

<sup>5</sup>For example, just as philosophers are asked sometimes to formulate an argument against their position, students can exercise practices of asking for and providing reasons by looking for source that they disagree with or the one from which their source of information diverges (akin to formulating a counterargument see Russo and Schoemaker 1992 and this kind of practice in Williams 2001).

information presented to them online (Wineburg et al. 2016, 4; see also Phippen et al. 2021, 40; and Polizzi 2020, Section 1). However, as previously discussed, we should not rely too heavily on such strategies alone for ensuring the reliable use of web-based technology. Our ability to do so also depends on understanding how capable we, as users *in front of the screen*, are in assessing the reliability of others, as well as recognizing the extent to which social, political, financial, and other factors operating *behind the screen* shape our judgments of reliability and, more broadly, our knowledge acquisition and dissemination in web-based practices.

#### 1.4.2. *Behind the screen*

The piece of information delivered to a social media user results from various unseen and opaque processes, which are a complex amalgamation of algorithmic, financial, social, and psychological factors (Niemiec 2020). With that in mind, it is important to understand, to a reasonable extent, how these external factors and content-shaping algorithms mediate our knowledge-seeking activities and influence our judgments about whom to trust and consider reliable, potentially compromising our ability to make well-informed decisions. The digital ecosystem is populated by malicious stakeholders, whether individuals, political parties, military leaders, policymakers, or governments, who manage troll farms (fake accounts and websites), information operations (gathering of strategic intelligence on an opponent combined with the spread of propaganda), social bots (programs that automatically like, share posts, and send messages), and similar entities to spread deceptive information and manipulate relevant narratives. However, while AI techniques for combating misinformation can partially identify false, inaccurate, or misleading content, they cannot determine the intent of the sharer (Bontridder and Poulet 2021, e32-10). Since detecting such intentions is crucial for avoiding reliance on misleading information designed to manipulate, students should develop the ability to accurately determine whether a speech act is intended to inform, mock, manipulate, or serve another purpose (Mattioni 2024). They should also be familiar with sophisticated forms of deception, the fabrication process, and other possible malicious behaviours on social media.

Another reason highlighting the limitations of education interventions solely focused on identifying accurate information and reliable sources stems from the recognition that not all information shared on social media serves the purpose of informing or misinforming, nor is it consistently taken seriously or understood as factual or testimonial (cf. Marin 2021, 367). The primary goal of social media is to maximize profit by exploiting our attention rather than providing high-quality information. It achieves this through an ad-driven business model that keeps users engaged, captivated, and fixated, leveraging micro-targeting to deliver highly specific advertisements (Maréchal and Biddle 2020). We believe that educating new generations about the 'backend' of online platforms and the extent to which they make it technologically possible for users to evaluate the reliability of their own sources – at least as far as our knowledge allows – is just as crucial as helping them understand the environment and society they live in (see also Mattioni 2024, Section 3).

#### 1.4.3. *In front of the screen*

Just as it is important to understand the extent to which technological affordances contribute to the formation of truth-conducive beliefs, it is equally important to recognize our own role in shaping an epistemically toxic environment. First, it must be



acknowledged that we can never hope to become ultra-reliable fact-checkers ourselves, as the very challenges that undermine formal fact-checking efforts also apply to individual users. The vast array of information encountered on social media will often exceed our own areas of expertise – if we possess any relevant expertise at all – and the techniques required for verification may lie beyond our cognitive and technological capacities (see Mattioni 2024, Section 3.2). This does not mean that any investigation of a claim is futile. But it means that we need to take a step back and keep ourselves in check. Our moods and emotions play a crucial role in shaping false beliefs and driving the spread of misinformation (Xu et al. 2023; Horner et al. 2021). At the same time, our perception of the world is influenced by our existing beliefs and societal factors. Confirmation bias, a well-documented and robust psychological phenomenon, significantly affects how we process information. Moreover, we are more likely to believe and share false information when we see others accepting or spreading it (Xu et al. 2023).

The awareness of the above mechanisms and psychological features will also assist us in determining the appropriateness of our sharing activities. We can be very careful about what information we adopt as a belief, but we can be, on the other hand, extremely careless when sharing. We can care about truth deeply and be perfectly well-meaning yet unreliable (Lichtenberg 2021, 18) by, let's say, believing that we are fully competent in some subject matter while unfoundedly exceeding our true competences. It is often the case that people, who make judgments about some field they have no expertise in, do so unwittingly. Moreover, in our interconnected and interdisciplinary world, even experts can struggle to determine the boundaries of their expertise and where someone else begins (Gerken 2018; Pavličić et al. 2023).

And, as we suggested above, we are not always interested in the facts or the truth (Record and Miller 2022a). As some studies suggest, the veracity of headlines has minimal influence on sharing intentions (Pennycook et al. 2021). The reason for this probably lies in social media platforms encouraging individuals to prioritise other factors, such as whether sharing will draw attention and approval from followers and friends (see also: Bronstein and Vinogradov 2021). One can share some piece of information that they are very sceptical of, or that we know is not true, to draw attention or point and say 'Look how silly this is!' (cf. Marin 2021, 365). Also, one may want to provoke, agitate, and annoy (troll) people in their friend group (Record and Miller 2022b). And while non-linguistic signals like emoticons, reaction options, and typographical conventions on social media can shape the interpretation of posts, they are far less useful than facial expressions or tone of voice in assessing the trustworthiness or sincerity of testimonies (Mattioni 2024, Section 3.1).

For example, let's say we regularly post flat-earth content (even though we don't believe in it) just to tease our close friend, who is an astrophysicist, and enjoy his reaction because it's amusing to us. This is simply banter or trolling, not an attempt to misinform anyone. However, let's now imagine that our friend is equally entertained by pretending to find the arguments in those posts quite plausible, and we continue this playful exchange on my public profile. To people inclined to believe in this theory, our playful conversation could lend credibility to it. Situations with unforeseen consequences of our posting and sharing practice can particularly turn dark if something known as the collapse of context occurs (Record and Miller 2022b). Context collapse refers to situations where multiple (different) audiences are flattened into a single context (Brandtzaeg and Lüders 2018). Due to its nature, social media presents the same content (for example, a post we shared) to different audiences, including people who do not know us, which can lead to significant misunderstandings (Marwick and Boyd 2011, in Record and Miller 2022b). We can act very responsibly as a receiver of information, yet

act in a misleading manner as a disseminator, because we are unaware of the epistemic constraints and contextual nature of social media, and do not consider more local norms that exist on different social media platforms (like private groups).<sup>6</sup> Therefore, the awareness of context and multiple perspectives online (Marin 2021; Record and Miller 2022b) is something that must be thoroughly researched and taught when it comes to responsible sharing and engaging with information. We believe that taking into account these considerations in order to standardise the practice of sharing will not only contribute to mitigating epistemically toxic content but also prevent the abandonment of public posting that is currently occurring, specifically the transition of users to closed groups, leaving open networks behind (The Economist 2024).

That's a pity, since technology is an integral part of our everyday epistemic practice in which internet and social media by extension, as a new learning environment, can contribute immensely to the dissemination and consumption of valuable information. As Paul Thagard (1997) argued some time ago, although seeking information on the World Wide Web is not always entirely reliable, it's important to recognize that in the hands of careful users, especially when guided by scientists, posting information can increase its reliability. We believe that Thagard's remark still applies today. As it has likely become more evident by now, knowing where to look and whom to listen to is simply not enough. The literature offers various frameworks for evaluating online posts, some of which are presented as acronyms for mnemonic purposes.<sup>7</sup> However, assessing the credibility or truthfulness of online content is only part of the responsibility of an informed internet user. Equally important is the mindful and responsible sharing of information. Drawing inspiration from these existing frameworks, we propose our own model designed to encourage more conscientious sharing practices. Below, we outline a potential structure for such a framework, along with its corresponding acronym.

### 1.5. ACCEDE

**A – Audience:** One of the first questions that we should ask ourselves before posting some material is: Who is the intended audience for this post? Answers to this question can help us be mindful of the tone and language; inappropriate or offensive language can alienate the audience.

**C – Collapse of the context:** After determining who the intended audience is, we should ask ourselves: Is the post likely to spread virally? In other words, does the post relate to any current events, trends, or viral topics that could increase its visibility and shareability, thus opening the possibility of a collapse of context?

**C – Contextualizing:** Can we prevent the potential collapse of context by providing context to the post that we are about to share? As we have already discussed, context (to a certain extent) helps prevent individuals interacting with our posts from making assumptions or jumping to conclusions based on incomplete information.

<sup>6</sup>As emphasised by Marin (2021, 366), there is no one context in which we exchange information. Each social media platform differs (e.g., LinkedIn vs. Instagram), and different spaces within social media (e.g., friends chat group, private group) have distinct rules (see also Record and Miller 2022b). Failing to acknowledge this variation risks leading users to treat non-serious content seriously, fact-check jokes and memes unnecessarily, and demand overly cautious and epistemically responsible behaviour even in informal settings such as private group chats with friends.

<sup>7</sup>See, for example, Mandalios (2013), Wiley et al. (2009), and Blakeslee (2004), whose proposals have likewise been examined in our previous analyses (see Pavličić et al. 2023; Nedeljković and Pavličić 2023, 2024).

**E – Explaining away:** Despite all our estimations and precautions, if our post still becomes subject to misinterpretation, misuse, or misrepresentation of the intended message, can we clarify and rationalize its content to explain it away?

**D – Detriments:** Apart from the undesirable consequences for our own reputation that we aim to prevent when things go wrong online, we should try to anticipate, understand, and acknowledge the real-world consequences of our online activities.

**E – Evaluation:** As mentioned earlier, we are not always solely concerned with being factual. However, when we do aim for factual accuracy, it's essential to be aware of our evidence-seeking duties and epistemic limitations. This necessitates an examination of the specific topic we want to comment on, accompanied by metaknowledge.

Perhaps this outlook may not be the most optimal, possibly far from it; however, developing a framework within which users could navigate the online space more securely and efficiently, comprehending it better, and finding ways in which they can responsibly express their ideas are of paramount importance. We require a free market for our ideas, and the withdrawal of individuals into closed groups threatens to close it off.

### 1.6. Concluding remarks

At the beginning of the paper, we introduced the distinction between agent-mediated and technology-mediated digital content to highlight that human users, due to insufficiently informed and reflective or malevolent sharing of information, are responsible for creating and disseminating misleading content (cf. Marin 2021, 363; Lichtenberg 2021, 17). Our goal was to demonstrate that neither proponents of structural approaches nor advocates of educational strategies for building a more reliable web-based epistemic landscape fully recognize the proactive (both creative and destructive) role that users play across various web-based information and communication channels. In fact, they tend to view internet users as passive recipients of information who need external protection against misinformation and clear guidance to distinguish between reliable and unreliable online information for making well-informed judgments. However, responsible and reliable use of web-based technology in knowledge-seeking practices involves more than understanding and improving testimonial knowledge practices, where information is received from others. A significant portion of unreliable information on the web arises precisely because of an average internet user's creation, propagation, and dissemination of content which seems to be the unregulated social media sharing practices. However, we expressed doubt that these practices can be regulated institutionally from the top down without integrating knowledge from research on user tendencies online into educational programs. To the best of our knowledge, there is limited research on sharing behaviour and the consequences of lacking clear sharing norms (Rini 2017; Arielli 2018; Record and Miller 2022b; Marsili 2021; Marin 2021) and only a few studies have explored the relationship between the ability to discern the truthfulness of content and sharing intentions (e.g. Pennycook et al. 2021). So, what we need is more research on these complex and interconnected questions as the foundation for building a comprehensive education approach for responsible and meaningful engagement with information and communication technology. It may seem like a gargantuan task, but only an educational approach aligned with thoughtful engagement with research evidence (Mouthaan and Révai 2023) can significantly contribute to reducing epistemically toxic content on various informational and communicative channels and mastering their reliable use.

## References

- Anderau G. (2021). 'Defining Fake News.' *KRITERION – Journal of Philosophy* 35(3), 197–215. <https://doi.org/10.1515/krt-2021-0019>
- Alber D., Alber D.A., Yang Z., Alyakin A., Yang E., Rai S., Valliani A.A., Zhang J., Rosenbaum G.R., Amend-Thomas A.K., Kurland D.B., and Kremer C.M. (2025). 'Medical Large Language Models are Vulnerable to Data-Poisoning Attacks.' *Nature Medicine* 1–9. <https://doi.org/10.1038/s41591-024-03445-1>
- Allcott H. and Gentzkow M. (2017). 'Social Media and Fake News in the 2016 Election.' *The Journal of Economic Perspectives* 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Altay S., Berriche M., and Acerbi A. (2023). 'Misinformation on Misinformation: Conceptual and Methodological Challenges.' *Social Media + Society* 9(1). <https://doi.org/10.1177/20563051221150412>
- Amazeen M. (2015). 'Revisiting the Epistemology of Fact-Checking.' *Critical Review* 27(1), 1–22. <https://doi.org/10.1080/08913811.2014.993890>
- Anderson E. (2011). 'Democracy, Public Policy, and Lay Assessments of Scientific Testimony.' *Episteme* 8(2), 144–164.
- Arielli E. (2018). 'Sharing as Speech Act.' *Versus* 47(2), 243–258.
- Bedessem B. and Ruphy S. (2020). 'Citizen Science and Scientific Objectivity: Mapping Out Epistemic Risks and Benefits.' *Perspectives on Science* 28(5), 630–654. [https://doi.org/10.1162/posc\\_a\\_00353](https://doi.org/10.1162/posc_a_00353)
- Black J. and Fullerton C. (2020). 'Digital Deceit: Fake News, Artificial Intelligence, and Censorship in Educational Research.' *Open Journal of Social Sciences* 8(7), 71–88. <https://doi.org/10.4236/jss.2020.87007>
- Blakeslee S. (2004). 'The CRAAP Test.' *LOEX/ Quarterly* 31(3). <https://commons.emich.edu/cgi/viewcontent.cgi?article=1009&context=loexquarterly>. Accessed 1st of March 2024.
- Bontridder N. and Pouillet Y. (2021). 'The Role of Artificial Intelligence in Disinformation.' *Data & Policy* 3, e32. <https://doi.org/10.1017/dap.2021.20>
- Braman D., Grimmelmann J., and Kahan D.M. (2005). 'Modeling Cultural Cognition.' *Social Justice Research* 18(3). Available at SSRN: <https://ssrn.com/abstract=1000449>
- Brandtzaeg P.B., and Lüders M. (2018). 'Time Collapse in Social Media: Extending the Context Collapse.' *Social Media + Society*. <https://doi.org/10.1177/2056305118763349>
- Brin S. and Page L. (1998). 'The Anatomy of a Large-Scale Hypertextual Web Search Engine.' *Computer Networks and ISDN Systems* 30, 107–117. [http://dx.doi.org/10.1016/S0169-7552\(98\)00110-X](http://dx.doi.org/10.1016/S0169-7552(98)00110-X)
- Bronstein M. and Vinogradov S. (2021). 'Education Alone is Insufficient to Combat Online Medical Misinformation.' *EMBO reports* 22(3). <https://doi.org/10.15252/embr.202052282>
- Couldry N. and Kallinikos J. (2017). 'Ontology.' In J. Burgess, A. Marwick and T. Poell (eds), *The SAGE Handbook of Social Media*, pp. 146–159. London: Sage Publications.
- Croce M. and Piazza T. (2021). 'Consuming Fake News: Can We Do Any Better?' *Social Epistemology* 1–10. <https://doi.org/10.1080/02691728.2021.1949643>
- de Villiers-Botha T. (2022) 'Reassessing Google as an Epistemic Tool' SACAIR2022 Online Conference, the 3rd Southern African Conference for Artificial Intelligence Research, South Africa. <https://philarchive.org/go.pl?id=DEVREGA&proxyId=&u=https%3A%2F%2Fphilpapers.org%2Farchive%2FDEVREGA.pdf>. Accessed January 12th, 2024.
- de Melo-Martín I. and Intemann K. (2018). *The Fight against Doubt: How to Bridge the Gap between Scientists and the Public*. New York: Oxford University Press.
- de Melo-Martín I. and Intemann K. (2012). 'Scientific Dissent and Public Policy. Is Targeting Dissent a Reasonable Way to Protect Sound Policy Decisions?' *EMBO Rep* 14(3), 231–235. <https://doi.org/10.1038/embor.2013.8>
- Eddy N. (2024). 'What is Data Poisoning, and How Can You Prevent It?' Technology Solutions That Drive Business. 2024, December 20. <https://biztechmagazine.com/article/2024/12/what-is-data-poisoning-pe rfcon>. Accessed on 23 of January, 2025.
- Eisemann C. and Pimmer C. (2020). 'Educational Approaches to Address Fake News: Preliminary Insights from a Systematic Review'. 17th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2020). Lisbon, Portugal: IADIS Press.
- Furman K. (2020). 'Emotions and Distrust in Science.' *International Journal of Philosophical Studies* 28(5), 713–730.
- Gerken M. (2018). 'Expert Trespassing Testimony and the Ethics of Science Communication.' *Journal for General Philosophy of Science* 49(3), 299–318. <https://doi.org/10.1007/s10838-018-9416-1>

- Gibbons A.F. (2024). 'Bullshit in Politics Pays.' *Episteme* 21(3), 1002–1022. <https://doi.org/10.1017/epi.2023.3>
- Graves L. (2017). 'Anatomy of a Fact Check: Objective Practice and the Contested Epistemology of Fact Checking.' *Communication, Culture & Critique* 10(3), 518–537.
- Graves L. (2018). 'Understanding the Promise and Limits of Automated Fact-Checking. Reuters Institute, February 2018. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves\\_factsheet\\_180226%20FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf); <https://doi.org/10.60625/risj-nqnx-bg89>. Accessed 24th of January 2025.
- Guess AM., Lerner M., Lyons B., Montgomery JM., Nyhan B., Reifler J., and Sircar N. (2020) 'A Digital Media Literacy Intervention Increases Discernment between Mainstream and False News in the United States and India'. *Proc Natl Acad Sci USA* 117(27), 15536–15545. <https://doi.org/10.1073/pnas.1920498117>
- Hardoš P. (2018). 'Who Exactly is an Expert? On the Problem of Defining and Recognizing Expertise.' *Sociológia-Slovak Sociological Review* 50(3), 268–288.
- Hassan A. and Barber S.J. (2021). 'The Effects of Repetition Frequency on the Illusory Truth Effect.' *Cognitive Research Principles and Implications* 6(1), 38. <https://doi.org/10.1186/s41235-021-00301-5>
- Heersmink R. (2018). 'A Virtue Epistemology of the Internet: Search Engines, Intellectual Virtues and Education.' *Social Epistemology* 32(1), 1–12. <https://doi.org/10.1080/02691728.2017.1383530>
- Herrero-Diz P. and López-Rufino C. (2021). 'Libraries Fight Disinformation: An Analysis of Online Practices to Help Users' Generations in Spotting Fake News.' *Societies* 11(4), 133. <https://doi.org/10.3390/soc11040133>
- Herzog D. (2006). *Cunning*. Princeton: Princeton University Press.
- Höchstötter N. and Lewandowski D. (2009). 'What Users See - Structures in Search Engine Results Pages.' *Information Sciences* 179, 1796–1812. <https://doi.org/10.1016/j.ins.2009.01.028>
- Horner C.G., Galletta D., Crawford J., and Shirsat A. (2021). 'Emotions: The Unexplored Fuel of Fake News on Social Media.' *Journal of Management Information Systems* 38(4), 1039–1066. <https://doi.org/10.1080/07421222.2021.1990610>
- Hsieh W., et al. (2024). A Comprehensive Guide to Explainable AI: From Classical Models to LLMs (No. arXiv:2412.00800). arXiv. <https://doi.org/10.48550/arXiv.2412.00800>
- Huber J. (2020). 'Fighting Fake News Successfully: Why Rini's Display of Reputation Scores Fails and How to Remedy Her Account.' *Studies in Philosophy, Politics and Economics* 2(1), 39–44.
- Humphreys P. (2009). 'The Philosophical Novelty of Computer Simulation Methods.' *Synthese* 169(3), 615–626.
- Intemann K. (2023). 'Science Communication and Public Trust in Science.' *Interdisciplinary Science Reviews* 48(2), 350–365.
- Kahne J. and Bowyer B. (2017). 'Educating for Democracy in a Partisan Age: Confronting the Challenges of Motivated Reasoning and Misinformation.' *American Educational Research Journal* 54(1), 3–34. <https://doi.org/10.3102/0002831216679817>
- Klug D., Steen E., and Yurechko K. (2023). How Algorithm Awareness Impacts Algospeak Use on TikTok. Companion Proceedings of the ACM Web Conference 2023, 234–237. <https://doi.org/10.1145/3543873.3587355>
- Koskinen I. (2024). 'We Have No Satisfactory Social Epistemology of AI-Based Science.' *Social Epistemology* 38(4), 458–475.
- Lazer D.M.J., Baum M.A., Benkler Y., Berinsky A.J., Greenhill K.M., Menczer F., Metzger M.J., Nyhan B., Pennycook G., Rothschild D., Schudson M., Sloman S.A., Sunstein C.R., Thorson E.A., Watts D.J. and Zittrain J.L. (2018). 'The Science of Fake News.' *Science* 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Levy N. (2024). 'No Trespassing! Abandoning the Novice/Expert problem.' *Erkenntnis* 1–18. <https://doi.org/10.1007/s10670-024-00794-8>
- Lichtenberg J. (2021). 'Lies, Bullshit, False Beliefs, Ignorance, Skepticism: Some Epistemic Fallout of Our Current Political Times.' *Political Epistemology* 1, 13–21.
- Lim C. (2018). 'Checking How Fact-Checkers Check.' *Research & Politics* 5(3). <https://doi.org/10.1177/2053168018786848>
- Livingstone S. (2019). 'Audiences in an Age of Datafication: Critical Questions for Media Research.' *Television & New Media* 20(2), 170–183. <https://doi.org/10.1177/1527476418811118>
- Lotto M., Hanjahanja-Phiri T., Padalko H., Oetomo A., Butt Z.A., Boger J., Millar J., Cruvinel T., and Morita P.P. (2023). 'Ethical Principles for Infodemiology and Infveillance Studies Concerning Infodemic Management on Social Media.' *Frontiers in Public Health* 11, 1130079. <https://doi.org/10.3389/fpubh.2023.1130079>

- Mandalios J.** (2013). 'RADAR: An Approach for Helping Students Evaluate Internet Sources.' *Journal of Information Science* 39(4), 470–478. <https://doi.org/10.1177/0165551513478889>
- Manfra M. and Holmes C.** (2020). 'Integrating Media Literacy in Social Studies Teacher Education.' *Contemporary Issues in Technology and Teacher Education* 20(1), 121–141.
- Maréchal N. and Biddle E.R.** (2020). 'It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge.' *New America*, March 17th, 2020. <https://www.newamerica.org/oti/reports/its-not-just-content-its-business-model/>. Accessed 9th of January, 2025.
- Marin L.** (2020). 'Three Contextual Dimensions of Information on Social Media: Lessons Learned from the COVID-19 Infodemic.' *Ethics and Information Technology* 23(S1), 79–86. <https://doi.org/10.1007/s10676-020-09550-2>
- Marin L.** (2021). 'Sharing (mis) Information on Social Networking Sites. An Exploration of the Norms for Distributing Content Authored by Others.' *Ethics and Information Technology* 23(3), 363–372. <https://doi.org/10.1007/s10676-021-09578-y>
- Markowitz D.M., Levine T.R., Serota K.B., and Moore A.D.** (2023). 'Cross-Checking Journalistic Fact-Checkers: The Role of Sampling and Scaling in Interpreting False and Misleading Statements.' *PLoS ONE* 18(7), e0289004. <https://doi.org/10.1371/journal.pone.0289004>
- Marres N.** (2018). 'Why We Can't Have Our Facts Back.' *Engaging Science, Technology, and Society* 4, 1–21.
- Marsden C. and Meyer T.** (2019). 'Regulating Disinformation with Artificial Intelligence: Effect of Disinformation Initiatives on Freedom of Expression and Media Pluralism'. European Parliamentary Research Service (EPRS) – Scientific Foresight Unit (STOA), March 2019, PE 624.279. <https://op.europa.eu/en/publication-detail/-/publication/b8722bec-81be-11e9-9f05-01aa75ed71a1>. Accessed 1st of March 2024.
- Marsili N.** (2021). 'Retweeting: its Linguistic and Epistemic Value.' *Synthese* 198(11), 10457–10483.
- Marwick A.E. and Boyd D.** (2011). 'I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience.' *New Media & Society* 13(1), 114–133.
- Matthias A.** (2004). 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata.' *Ethics and Information Technology* 6(3), 175–183.
- Mattioni M.** (2024). 'Is Epistemic Autonomy Technologically Possible Within Social Media? A Socio-Epistemological Investigation of the Epistemic Opacity of Social Media Platforms.' *Topoi* 43, 1503–1516.
- Mercier H. and Sperber D.** (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/9780674977860>
- Millar B.** (2019). 'The Information Environment and Blameworthy Beliefs.' *Social Epistemology* 33(6), 525–537. <https://doi.org/10.1080/02691728.2019.1676840>
- Millar B.** (2021). 'Misinformation and the Limits of Individual Responsibility.' *Social Epistemology Review and Reply Collective* 10(12), 8–21. <https://wp.me/p1Bfg0-6kD>
- Miller B. and Record I.** (2017). 'Responsible Epistemic Technologies: A Social-Epistemological Analysis of Autocompleted Web Search.' *New Media & Society* 19(12), 1945–1963. <https://doi.org/10.1177/1461444816644805>
- Mishima K. and Yamana H.** (2022). 'A Survey on Explainable Fake News Detection.' *IEICE Transactions on Information and Systems* E105.D(7), 1249–1257. <https://doi.org/10.1587/transinf.2021EDR0003>
- Mittelstadt B.D., Allo P., Taddeo M., Wachter S., and Floridi L.** (2016). 'The Ethics of Algorithms: Mapping the Debate.' *Big Data & Society* 3(2). <https://doi.org/10.1177/2053951716679679>
- Mößner N. and Kitcher P.** (2017). 'Knowledge, Democracy, and the Internet.' *Minerva* 55, 1–24. <https://doi.org/10.1007/s11024-016-9310-0>
- Mouthaan M. and N. Révai** (2023) *Building a Culture of Research Engagement in Education, in Who Really Cares about Using Education Research in Policy and Practice?: Developing a Culture of Research Engagement*. Paris: OECD Publishing. <https://doi.org/10.1787/c8ebdafc-en>
- Nedeljković A. and Pavličić J.** (2023). 'Odgovorno ponašanje u saznavnim praksama koje se oslanjaju na upotrebu internet pretraživača.' *Antropologija* 23, 54–73.
- Nedeljković A. and Pavličić J.** (2024). 'The Moral Dimension of Epistemic Practices in Online Environments'. *Theoria* 67(4), 167–176 <https://doi.org/10.2298/THEO2404167N>
- Nguyen C.T.** (2018). 'Echo Chambers and Epistemic Bubbles.' *Episteme* 17(2), 141–161. <https://doi.org/10.1017/epi.2018.32>
- Nguyen C.T.** (2021). 'The Seductions of Clarity.' *Royal Institute of Philosophy Supplement* 89, 227–255. <https://doi.org/10.1017/S1358246121000035>
- Niemiec E.** (2020). 'COVID-19 and Misinformation: Is Censorship of Social Media a Remedy to the Spread of Medical Misinformation?' *Science & Society EMBO Rep* 21, e51420. <https://doi.org/10.15252/embr.202051420>



- Nygren T., Guath M., Axelsson C.W., and Frau-Meigs D. (2021). 'Combating Visual Fake News with a Professional Fact-Checking Tool in Education in France, Romania, Spain and Sweden.' *Information* 12(5), 201. <https://doi.org/10.3390/info12050201>
- O'Connor C. and Weatherall J.O. (2019). *The Misinformation Age: How False Beliefs Spread*. New Haven, CT: Yale University Press.
- Papageorgiou E., Chronis C., Varlamis I., and Himeur Y. (2024). 'A Survey on the Use of Large Language Models (LLMs) in Fake News.' *Future Internet* 16(8), 298. <https://doi.org/10.3390/fi16080298>
- Pavličić J., Dimitrijević J., Vučković A., Đorđević S., Nedeljković A., and Tešić Ž. (2023). 'Friend or Foe? Rethinking Epistemic Trespassing.' *Social Epistemology* 38(2), 249–266. <https://doi.org/10.1080/02691728.2023.2286255>
- Pennycook G., Epstein Z., Mosleh M., Arechar A.A., Eckles D., and Rand D.G. (2021). 'Shifting Attention to Accuracy Can Reduce Misinformation Online.' *Nature* 592(7855), 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
- Phippen A., Bond E., and Buck E. (2021). '3- Effective Strategies for Information Literacy Education: Combatting 'Fake News' and Empowering Critical Thinking.' *Chandos Digital Information Review* 39-53. <https://doi.org/10.1016/B978-0-12-822144-0.00003-3>
- Polizzi G. (2020). 'Digital Literacy and the National Curriculum for England: Learning from How the Experts Engage with and Evaluate Online Content.' *Computers & Education* 152, 103859.
- Polizzi G. and Taylor R. (2019). 'Misinformation, Digital Literacy and the School Curriculum.' Media Policy Briefs 22. London: Media Policy Project, London School of Economics and Political Science.
- Pongiglione F. (2021). 'Information, Ignorance, and Obligations to Know.' *Biblioteca della libertà* 56, 55–76.
- Pritchard D. (2013). 'Epistemic Virtue and the Epistemology of Education.' *Journal of Philosophy of Education* 47(2), 236–247.
- Quelle D. and Bovet A. (2024). 'The Perils and Promises of Fact-Checking with Large Language Models.' *Frontiers in Artificial Intelligence* 7, 1341697. <https://doi.org/10.3389/frai.2024.1341697>
- Ramezani M., Rafiei M., Omranpour S., and Rabiee H. (2019). 'News Labeling as Early as Possible: Real or Fake?' Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. <https://doi.org/10.1145/3341161.3342957>
- Record I. and Miller B. (2022a). 'Wrong on the Internet: Why Some Common Prescriptions for Addressing the Spread of Misinformation Online Don't Work.' *Communique* 105, 22–27.
- Record I. and Miller B. (2022b). 'People, Posts, and Platforms: Reducing the Spread of Online Toxicity by Contextualizing Content and Setting Norms.' *Asian Journal of Philosophy* 1(41), 41. <https://doi.org/10.1007/s44204-022-00042-2>
- Record I. (2013). 'Technology and Epistemic Possibility.' *Journal for General Philosophy of Science* 44(2), 319–336. <https://doi.org/10.1007/s10838-013-9230-8>
- Rini R. (2017). 'Fake News and Partisan Epistemology.' *Kennedy Institute of Ethics Journal* 27(2), E-43-E-64. <https://doi.org/10.1353/ken.2017.0025>
- Rozado D. (2023). 'The Political Biases of ChatGPT.' *Social Sciences* 12(3), 148. <https://doi.org/10.3390/sscsci12030148>
- Russo J. and Schoemaker P. (1992). 'Managing Overconfidence.' *Sloan Management Review* 33: 7–17.
- Santos F.C.C. (2023). 'Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis.' *Journalism and Media*, 4(2), 679–687. <https://doi.org/10.3390/journalmedia4020043>
- Simpson T.W. (2013). 'Evaluating Google as an Epistemic Tool.' In H. Halpin and A. Monnin (eds), *Philosophical Engineering*, pp. 89–104. Chichester, UK: Wiley-Blackwell. <https://doi.org/10.1002/9781118700143.ch7>
- Spence E. (2021). *Stoic Philosophy and the Control Problem of AI Technology: Caught in the Web*. United Kingdom: Rowman & Littlefield International.
- Steensen S., Kalsnes B., and Westlund O. (2023). 'The Limits of Live Fact-Checking: Epistemological Consequences of Introducing a Breaking News Logic to Political Fact-Checking.' *New Media & Society* 26(11), 6347–6365. <https://doi.org/10.1177/14614448231151436>
- Steen E., Yurechko K., and Klug D. (2023) 'You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok'. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231194586>
- Sunstein C. (2007). *Republic.com 2.0*. Princeton: Princeton University Press.
- Thagard P. (1997). Internet Epistemology: Contributions of New Information Technologies to Scientific Research. Unpublished Manuscript, University of Waterloo. <https://web.archive.southampton.ac.uk/cogprints.org/674/1/Epistemology.html#anchor08>. Accessed 10th of March 2024.

- The Economist.** (2024). The End of the Social Network. The Economist. 2024, February 1. <https://www.economist.com/leaders/2024/02/01/the-end-of-the-social-network>. Accessed January 23, 2025.
- Thornhill C., Meus Q., Peperkamp J., and Berendt B.** (2019). A Digital Nudge to Counter Confirmation Bias. *Frontiers in Big Data* 2, 11. <https://doi.org/10.3389/fdata.2019.00011>
- Uscinski J.E. and Butler R.W.** (2013). The epistemology of fact checking. *Critical Review*, 25(2), 162–180. <https://doi.org/10.1080/08913811.2013.843872>
- Uscinski J.E.** (2015). The Epistemology of Fact Checking (Is Still Naive): Rejoinder to Amazeen. *Critical Review* 27(2), 243–252. <https://doi.org/10.1080/08913811.2015.1055892>
- van Dijck J.** (2010). Search Engines and the Production of Academic Knowledge. *International Journal of Cultural Studies* 13(6), 574–592. <https://doi.org/10.1177/1367877910376582>
- Vosoughi S., Roy D., and Aral S.** (2018). 'The Spread of True and False News Online.' *Science* 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Wasike B.** (2023). You've Been Fact-Checked! Examining the Effectiveness of Social Media Fact-Checking against the Spread of Misinformation. *Telematics and Informatics Reports* 11. <https://doi.org/10.1016/j.tel.2023.100090>
- Wikipedia Contributors.** (2025a). Sarcasm. Wikipedia. 2025a, January 7. <https://en.wikipedia.org/wiki/Sarcasm>. Accessed January 23, 2025.
- Wiley J., Goldman S., Graesser A., Sanchez C., Ivan A.I., and Hemmerich J.** (2009) 'Source Evaluation, Comprehension, and Learning in Internet Science Inquiry Tasks'. *American Educational Research Journal* 46(4), 1060–1106. <https://doi.org/10.3102/0002831209333183>
- Williams M.** (2001). *Problems of Knowledge: A Critical Introduction to Epistemology*. Oxford: University Press Oxford.
- Wineburg S., McGrew S., Breakstone J., and Ortega T.** (2016). Evaluating Information: The Cornerstone of Civic Online Reasoning. Stanford Digital Repository. <http://purl.stanford.edu/fv751yt5934>. Accessed 1st of March 2024.
- Worsnip A.** (2019). 'The Obligation to Diversify One's Sources: Against Epistemic Partisanship in the Consumption of News Media.' In C. Fox and J. Saunders (eds), *Media Ethics: Free Speech and the Requirements of Democracy*, pp. 240–264. London: Routledge.
- Weise K., and Metz C.** (2023). When A.I. Chatbots Hallucinate. The New York Times. 2023, May 1. <https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html>. Accessed 12th of December 2024.
- Wynne B.** (1998). 'May the Sheep Safely Graze? A Reflexive View of the Expert-Lay Knowledge Divide.' In S. Lash, B. Szerzynski, B. Wynn (eds), *Risk, Environment and Modernity: Towards a New Ecology*, pp. 44–84. London: Sage Publication Ltd. <https://doi.org/10.4135/9781446221983.n3>
- Xu D., Fan S., and Kankanhalli MS.** (2023). 'Combating Misinformation in the Era of Generative AI Models'. Proceedings of the 31st ACM International Conference on Multimedia. Association for Computing Machinery, New York, NY, USA.
- Zhang Y., Zou C., Lian Z., Tiwari P., and Qin J.** (2024). SarcasmBench: Towards Evaluating Large Language Models on Sarcasm Understanding (No. arXiv:2408.11319). arXiv. <https://doi.org/10.48550/arXiv.2408.11319>

Jelena Pavličić Cerović works as Research Associate at the Institute of Philosophy, Faculty of Philosophy, University of Belgrade. She has received an MSc and PhD in Epistemology from the University of Belgrade, where she taught courses and seminars on General Methodology of Science, Foundations of Philosophy and Methodology of Science, General Methodology, and History of Philosophy (19th-century philosophy). The main focus of her research lies in Epistemology (ancient and modern), Social Epistemology, and the Social Epistemology of Science. [jelena.pavlicic@f.bg.ac.rs](mailto:jelena.pavlicic@f.bg.ac.rs) <http://orcid.org/0000-0002-5572-3826>

Adam Nedeljković works as a Research Associate at the University of Belgrade, Faculty of Philosophy, Department of Philosophy. The main fields of interest are formal and social epistemology [nedamned@hotmail.com](mailto:nedamned@hotmail.com). <https://orcid.org/0000-0003-1842-1981>

---

**Cite this article:** Pavličić Cerović J. and A. Nedeljković (2025). "It's not just the Web, It's the Spider as well: Prioritising Education over Structural Changes." *Episteme* 1–22. <https://doi.org/10.1017/epi.2025.27>