

Misinformation



Enes Altuncu ¹, Shujun Li ² and
Jason R. C. Nurse ²

¹Istanbul University-Cerrahpaşa, Istanbul,
Türkiye

²University of Kent, Canterbury, UK

Keywords

Misinformation · Disinformation · False
information · Information disorder ·
Information integrity · Fake news

Definition

‘*Misinformation*’ refers to false or misleading information shared without intent to deceive. It differs from ‘*disinformation*’, which is intentionally deceptive. The spread of misinformation has intensified with the advent of digital technologies, particularly online social media and artificial intelligence (AI), enabling the rapid and large-scale dissemination of information. Belief in misinformation is influenced by cognitive biases, social identity and emotional responses, whilst sharing behaviour may occur even without belief. Efforts to address misinformation include detection (e.g. fact-checking) and prevention (e.g. prebunking), often combining human and automated systems. Regulatory frameworks in regions such as the EU and the UK aim to increase

platform accountability. Terminological inconsistencies, insufficient consideration of human aspects and the dual role of AI tools, such as large language models (LLMs), present ongoing challenges.

Introduction

From early urban legends to coordinated disinformation campaigns, misinformation has long influenced society, often with harmful consequences. Nevertheless, social media and the advancements in digital technologies, including artificial intelligence (AI), have dramatically accelerated and facilitated the generation and dissemination of misinformation. The 2025 Global Risks Report published by the World Economic Forum identified misinformation as one of the most significant emerging threats to society (Elsner et al., 2025). Whilst a more popular term for misinformation has been ‘*fake news*’, especially since the 2016 US presidential election, its usage has often been criticised in academic discourse due to its inadequacy to describe the complex nature of the phenomenon (Wardle & Derakhshan, 2017). However, there is still no consensus on the terminology, and various terms, such as information disorder, information pollution and infodemic, are also used inconsistently across disciplines and public discourse.

This entry aims to provide an overview of online misinformation by considering different

aspects of the phenomenon and the efforts of multiple research disciplines. It begins with a historical background, followed by psychological explanations for why people believe and share misinformation. The entry also outlines some key approaches to the detection and prevention of misinformation online, including current regulatory efforts. Ultimately, it concludes by highlighting some open issues that merit further research.

Historical Context

One of the earliest documented instances of misinformation shaping political outcomes dates to Ancient Rome, when Octavian launched a successful propaganda campaign against Mark Antony. This effort significantly influenced public opinion and ultimately enabled Octavian to become the first Roman Emperor. The invention of the printing press in 1440 marked a turning point in the dissemination of information, significantly amplifying the reach and speed of misinformation. This helped the generation of the first widespread hoax, *The Great Moon Hoax* of 1835, a series of articles published in New York's *The Sun* newspaper on the discovery of life on the moon, featuring illustrations of imaginary creatures (Posetti & Matthews, 2018).

In the modern era, digital technologies, including the Internet, smart devices, online social media platforms and AI, have further intensified the spread of misinformation. These tools have not only widened the audience for such content but also complicated its detection and mitigation. For example, fabricated news stories influenced public discourses during several national elections worldwide, including the USA, Brazil and India (Reis et al., 2020). Similarly, during the COVID-19 pandemic, widespread conspiracy theories fuelled vaccine hesitancy and led to preventable deaths (Islam et al., 2020).

Psychology of Misinformation

People believe misinformation not only because of their lack of awareness of or access to accurate information but also due to several cognitive, social and emotional factors (Ecker et al., 2022). Cognitively, people tend to rely on intuition, familiarity and repetition rather than analytical thinking, making repeated misinformation seem more believable (i.e. *the illusory truth effect*). Even those with prior knowledge or access to accurate sources can be exposed to the illusory truth effect regardless of their cognitive ability. Socially, individuals are more likely to believe messages from sources they perceive as credible, like themselves or part of their in-group, even if those sources are unreliable. Experts and political elites can heavily influence public beliefs, especially when they share false claims. Moreover, people often forget or ignore source information, making it harder to evaluate content critically. Emotionally, messages that evoke strong emotions (e.g. fear, anger, happiness) are more persuasive and widely shared, often distracting people from their assessment of some significant cues of misinformation, such as source credibility.

Although a common assumption is that believing misinformation leads to its propagation, researchers have found that belief and sharing behaviour are generally disconnected (Pennycook & Rand, 2021). People can identify false claims yet still prefer to share them, especially in political or emotionally charged contexts. Existing research explains this belief-sharing gap in three ways. Initially, some people genuinely make mistakes and indeliberately share misinformation as true. On the contrary, some others spread false information on purpose to express political identity, provoke or entertain, even if they doubt its accuracy, therefore moving misinformation towards disinformation. Finally, some people simply lack reflective thinking despite their strong preference for sharing only accurate information. The fast-paced, distraction-rich environment of social media causes users to ignore accuracy when deciding what to share.

Detection and Prevention

Efforts to combat misinformation can be broadly divided into two strategic approaches: detection and prevention. The detection strategy focuses on identifying misinformation after its generation to prevent its further dissemination. This can be done in multiple ways, including content moderation and fact-checking tools. In contrast, the prevention strategy aims to act proactively by stopping the generation, propagation and/or acceptance of misinformation. Preventive efforts include legal deterrents, educational programs to enhance critical thinking and media literacy, issuing general warnings or using *inoculation games* that expose users to misinformation in advance (known as *prebunking*) to build psychological resistance. Whilst presented as two different approaches, detection and prevention have a close relationship; effective detection can trigger and enhance preventive measures.

Detecting misinformation usually involves an investigative process known as *fact-checking*, which consists of several steps: identifying checkworthy claims, gathering evidence, assessing accuracy and providing perspective on claims. Depending on the level of human involvement, fact-checking can be carried out manually, semiautomatically or fully automatically. Manual fact-checking relies on traditional investigative methods, such as contacting sources or consulting experts, and has been the standard practice amongst journalists before the rise of digital tools. Whilst they can provide higher accuracy, manual methods lack scalability due to the huge amount of misinformation circulating online.

Digital technologies can aid or automate fact-checking tasks. Automatic fact-checking uses tools such as AI and natural language processing (NLP) to identify the veracity of checkworthy claims without human intervention. Despite its scalability and technological advancements, automatic fact-checking still faces considerable limitations, including model accuracy, training data availability, dataset imbalances and domain coverage. Constructing high-quality training data remains difficult and often still requires human

input. Furthermore, large-scale datasets covering a wide timespan and diverse fact-checking sources are quite rare (Altuncu et al., 2025). Other challenges include dealing with multiple languages and media types, ambiguous claims, algorithmic bias and insufficient coverage of different domains. From a psychological standpoint, fully automated and black-box systems may fail to persuade users or even backfire, particularly if they offer verdicts without explanations. Users, including both professional fact-checkers and the general public, often express scepticism about fully automatic fact-checking (Juneja & Mitra, 2022). Given the limitations of both manual and fully automated approaches, hybrid systems have emerged to combine the strengths of human intelligence and automation, where tasks are shared between people and digital tools for more effective results. Human actors (e.g. professional fact-checkers, domain experts or crowdsourcing workers) can be engaged at various points in the process, with digital systems assisting in tasks like claim determination or evidence retrieval (Altuncu, 2025).

Regulation

Besides the technical and psychological approaches, regulations play a key role in addressing misinformation. Recent regulations introduced in various countries and regions aim to hold platforms accountable, promote transparency and protect users from the harmful effects of misinformation. In the following, we briefly introduce some of them.

One of the earliest national efforts was Germany's *Network Enforcement Act (NetzDG)*, introduced in 2017. NetzDG requires social media platforms with over two million users to remove 'clearly illegal' content, including disinformation, within 24 hours and all illegal content within 7 days of it being posted.

At the European level, the *EU Digital Services Act (DSA)*, adopted in 2022, represents a comprehensive regulatory framework for digital platforms. It imposes risk management obligations on all online platforms, including the mitigation

of systemic risks such as disinformation. Platforms must conduct annual risk assessments, provide algorithmic transparency and offer meaningful access to data for vetted researchers. The DSA empowers regulators to enforce compliance through fines and audits, significantly raising the bar for platform accountability in the EU.

Complementing the DSA is the *EU Code of Practice on Disinformation*, a self-regulatory initiative launched in 2018. It brings together major technology platforms, advertising companies and civil society organisations to combat disinformation through voluntary commitments. Signatories commit to disrupting monetisation of false content, improving political ad transparency, empowering users with tools to recognise disinformation and collaborating with fact-checkers. In 2025, the Code was formally integrated into the DSA framework and adopted as an official *Code of Conduct*, strengthening its legal and regulatory significance.

In the UK, the *Online Safety Act*, passed in 2023, introduces a new regulation to protect users from online harms, including misinformation. Under the Act, major platforms must assess and address the risks posed by both illegal and legal-but-harmful content, especially for children. The Act also emphasises transparency reporting, user empowerment tools and age-appropriate content controls.

Open Issues

Misinformation Terminology

Terminology around misinformation varies significantly across academic disciplines, posing a major challenge to interdisciplinary collaboration and coherent knowledge development (Altay et al., 2023). Different fields often adopt conflicting definitions of core concepts, which makes the formation of a shared conceptual framework more challenging. Additionally, no consensus exists on a comprehensive umbrella term that adequately captures the full scope of the phenomenon. This lack of definitional clarity has practical and scientific consequences (Altay et al., 2023). Ambiguous or overly broad definitions can be exploited by policymakers to justify censorship under the guise of combating

misinformation. From a research perspective, semantic inconsistencies may contribute to apparent contradictions in empirical findings, e.g. in assessing the impact or prevalence of misinformation. To address these challenges, misinformation researchers and practitioners should clearly define the concepts that they use and be transparent about their underlying assumptions. Such clarity can help reduce interdisciplinary misunderstandings and support effective research, interventions, collaboration, responsible policy design, awareness and educational efforts.

Human-Centred Challenges in Fact-Checking

Fact-checking is not merely a technical task but a socio-technical process that involves complex human factors. However, current solutions often overlook these aspects, leading to several unresolved challenges:

Trust Deficit: Users tend to distrust both automated systems and human fact-checkers when they act alone. A major open issue is how to build trust, both in the credibility of sources and in the stakeholders involved. Systems that indiscriminately retrieve evidence from widely available but not necessarily credible sources (e.g. generic web search or Wikipedia) can compromise trust. Human-AI collaboration, where each checks and balances the other, is a promising but under-explored direction.

Usability Barriers: Many existing systems require technical expertise or impose a high cognitive load, limiting accessibility. Designing intuitive, interactive and cognitively light interfaces remains an open challenge. Clear and concise outputs, along with thoughtful content presentation and visualisation, are essential for broader adoption.

Lack of Explainability: Users need to understand how a system reaches its verdict to trust and effectively use it. Most current systems fail to provide clear, unbiased and ethically sound explanations. Whilst explainable AI (XAI) offers promising techniques, very few fact-checking systems implement them. Moreover, exposing model internals can increase vulnerability to adversarial attacks. Thus, developing safer, user-friendly

explanation formats, such as example-based or natural language justifications, remains a key research gap.

Large Language Models (LLMs)

The emergence of large language models (LLMs) has introduced both opportunities and challenges for misinformation research. LLMs can generate false or misleading content indeliberately, known as *hallucinations*, which can appear highly fluent and convincing. These outputs often lack source attribution, making them difficult to verify. Moreover, LLMs can be exploited for large-scale disinformation, including generating a wide range of fake content, impersonating individuals and bypassing content moderation systems.

Nevertheless, LLMs also offer potential support for fact-checking. They have shown moderate success in tasks like claim matching, evidence retrieval and veracity classification. When combined in modular pipelines or paired with human oversight, LLMs can enhance fact-checking efficiency. However, current limitations in accuracy imply that using them as assistive tools rather than replacements for human fact-checkers can be a better practice (Augenstein et al., 2024).

Cross-References

- [Deep Fakes](#)
- [Online Deception](#)
- [Safety by Design: Preventing Online Harms](#)
- [The UK Online Safety Act in the Context of Global Regulation](#)

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

- Altay, S., Berriche, M., Heuer, H., Farkas, J., & Rathje, S. (2023). A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. *Harvard Kennedy School Misinformation Review*, 4(4), 1–34. <https://doi.org/10.37016/mr-2020-119>
- Altuncu, E. (2025). *Fact-checking ecosystem: From modelling to semi-automated stakeholder-based detection of false information* [Doctoral dissertation, University of Kent]. <https://doi.org/10.22024/UniKent/01.02.108725>
- Altuncu, E., Baskent, C., Bhattacharjee, S., Li, S., & Roy, D. (2025). FACTors: A new dataset for studying the fact-checking ecosystem. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 3530–3539). <https://doi.org/10.1145/3726302.373033>
- Augenstein, I., Baldwin, T., Cha, M., Chakraborty, T., Ciampaglia, G. L., Corney, D., ... & Zagni, G. (2024). Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence*, 6(8), 852–863. <https://doi.org/10.1038/s42256-024-00881-z>
- Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Elsner, M., Atkinson, G., & Zahidi, S. (2025). *The global risks report 2025* (20th ed., Technical Report). World Economic Forum, <https://www.weforum.org/publications/global-risks-report-2025/>
- Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. H. M., Hasan, S. M., Kabir, A., ... & Seale, H. (2020). COVID-19–related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, 103(4), 1621. <https://doi.org/10.4269/ajtmh.20-0812>
- Juneja, P., & Mitra, T. (2022). Human and technological infrastructures of fact-checking. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1–36. <https://doi.org/10.1145/3555143>
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Posetti, J., & Matthews, A. (2018). A short guide to the history of ‘fake news’ and disinformation. *International Center for Journalists*, 7, 2018-07. https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf
- Reis, J. C. S., Melo, P., Garimella, K., Almeida, J. M., Eckles, D., & Benevenuto, F. (2020). A dataset of fact-checked images shared on WhatsApp during the Brazilian and Indian elections. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1), 903–908. <https://doi.org/10.1609/icwsm.v14i1.7356>
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27, pp. 1–107). Council of Europe. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>