



# FAKE NEWS FIGHT BACK

*Andreea Plesea reveals why trusted editorial ecosystems hold the key to the future of information integrity*

**D**isinformation has become one of the most persistent and complex challenges of the digital age. The sheer scale and rapid spread of misinformation can threaten democratic institutions, public safety and institutional trust. While not a new phenomenon, the pace and sophistication of disinformation in the digital age have accelerated beyond the reach of traditional countermeasures. In an era where manipulated videos can reach millions in hours, simply correcting falsehoods after the fact is no longer enough. The stakes are clear: unless we can build information systems that are intelligent, transparent and capable of operating at the

same speed and scale as disinformation, the public's access to factual, contextualised news will continue to erode.

In the past decade alone, we have seen a proliferation of fact-checking websites, digital literacy programmes and social platform moderation policies. While valuable, these tools are fundamentally reactive. They rely on identifying falsehoods after they have already reached an audience - by which point the damage is often done. Compounding this is the way we now consume news. Many people no longer use verified, trusted "mainstream" sources of news. Instead, they encounter headlines and snippets within algorithm-driven feeds on social media platforms, stripped of original context and promoted by their favourite influencer whose

**The way we now consume news allows misinformation to flourish, while correct information struggles to get through**

priorities may not align with accuracy. This allows misinformation to flourish in echo chambers, while corrective information struggles to reach them.

Even search engines and publicly available large language models (LLMs) can unintentionally add to the problem. A query on a controversial topic might surface a muddle of credible journalism, opinion blogs and outright fabricated content - leaving the individual to work out what's true and trustworthy, and what's false or inflammatory. Generative AI platforms trained on vast swathes of the internet make matters worse by blending unverified material into fluent, confident-sounding responses that present falsehoods with the same weight and authority as facts. The result is a constant fog of information, where certainty is elusive and the burden of verification rests entirely on the reader.

## THE SCALE AND SPREAD OF MISINFORMATION CAN THREATEN INSTITUTIONAL TRUST

If you can't beat them, as the saying goes, join them - but do it better. This is where agentic AI comes in: a class of artificial intelligence that goes far beyond basic retrieval or summarisation. Unlike traditional AI, which may serve up whatever answer is statistically most likely, agentic AI is built to act with intent. It can interpret the nuance behind a question, understand the context in which it's being asked and retrieve information only from trusted, pre-approved sources.

Every fact it shares is anchored to a transparent citation, giving the user an instant audit trail. Instead of treating each question in isolation, it maintains the thread of a conversation, adapting as the discussion develops. If a request is ambiguous, it won't guess - it will probe, clarify and guide the user towards a precise and accurate answer. This ability to combine autonomous reasoning, contextual understanding and verifiable output makes agentic AI a powerful weapon against disinformation. Rather than passively repeating whatever the internet says, it actively works to ensure that only credible, validated information is delivered.

However, even the smartest AI is only as good as the information it's trained on. This is where trusted editorial ecosystems come in. In this model, the AI is not connected to the entire open web. Instead, it operates within a curated, journalist-governed knowledge base; a dataset exclusively made up of verified content from reputable news organisations and subject-matter experts. This approach brings several advantages to combat fake news:

- **Accuracy:** Content is vetted by professional journalists before entering the dataset.
- **Transparency:** Every answer is linked to its original source, giving users a clear audit trail.
- **Independence:** Editorial governance ensures that no single interest group can manipulate the dataset.
- **Security:** By excluding unverified sources, the risk of AI-generated 'hallucinations' is vastly reduced.

One real-world example is the multilingual European news AI agent ChatEurope, which is built and governed by a consortium of major European media

organisations. Rather than trawling the internet for answers, it draws exclusively from its partners' verified news archives and delivers responses in multiple languages - each backed by a direct citation.

## MULTILINGUAL ACCESS AS A RESILIENCE TOOL

Language barriers have long been a powerful and influential factor for those intent on spreading disinformation. False narratives can take root in one language, circulating freely for days or even weeks before credible counter narratives are translated and reach the same audience. By that point, the damage is often already done. A trusted editorial ecosystem powered by agentic AI has the potential to break this cycle.

Such systems can deliver the same verified information simultaneously across multiple languages, meaning there is a significantly smaller gap in timing that is exploited by bad actors. At the same time, high-quality multilingual translation preserves the cultural and contextual nuances that are often lost in masses of machine-generated translations, ensuring the meaning and intent of the original reporting remain intact. This approach also extends the reach of factual content to communities that are traditionally underserved, whether because of geography, limited access to local journalism or lower digital infrastructure. The result is a more level information playing field, on which access to truth is no longer dictated by location, language or literacy, but by the quality and integrity of the sources themselves.

Unlike static fact-checking pages or scripted chatbots, agentic AI systems can hold layered conversations with users. This means they can adapt to the flow of a discussion, ask clarifying questions and build on previous exchanges to refine the accuracy of information delivered. For example, a user might start with a question such as: "What's the EU's current stance on agricultural subsidies?" The AI responds with a sourced, concise answer, serving up the sources it has used for its response. Over time, persistent context allows the AI to counter layered misinformation, where false claims evolve or branch off from a single core narrative, without forcing the user back to a search bar.

By design, a trusted editorial ecosystem that feeds agentic AI sidesteps many of the pitfalls that plague open-internet AI tools. No popularity-driven rankings are skewing the visibility of information; relevance and verification take precedence over what happens to be trending. The knowledge base is deliberately closed to unverified or malicious sources, so the system is never in a position to repeat content that has not passed through a rigorous editorial filter. And because the infrastructure is open-source, the way answers are generated is transparent, which democratises the scrutiny process and eradicates reliance on a 'black box' of algorithms.

Perhaps most importantly, the AI remains accountable to the editorial standards of the organisations that govern it. That means its automated outputs are held to the same principles of fairness, attribution and evidence-based reporting that underpin professional journalism. The technology may be advanced, but it is still anchored