

## Challenges raised by AI regarding the right to information

This document provides an indicative and non-exhaustive view of challenges raised by AI concerning the right to information. These challenges are distributed among the significant stages of media work (information gathering, processing, dissemination, as well as strategic positioning).

### 1. Information gathering

*This phase encompasses journalistic research, investigation, and gathering raw materials such as pictures or videos. It also involves the tools and methods that information professionals rely on to select news and topics to cover.*

#### ***Fabricated contents***

As generative AI (GAI) democratises, some estimate that synthetic content may soon overwhelm human-made content. A growing concern is the ability to fabricate evidence, such as HD videos or audio recordings. While image manipulation has historical precedents, GAI now allows for the near-instantaneous, cheap creation of synthetic content. The challenge for information integrity lies in discerning fake from real, and in the potential for genuine content to be mistaken as synthetic.

While some place hope in the coming progress of AI detection tools, [many argue](#) that GAI's ability to bypass these tests will progress at a similar pace. Synthetic content - text, image, or sounds - are already [close to undetectable](#).

Some advocate for the automatic marking of synthetic content at the creation stage. This technique, known as "watermarking", embeds patterns like combinations of [pixels](#) or [words](#) in artificial content. [Others](#) promote adopting content authenticity standards, which involve attaching secured metadata (author, date, etc.) at the creation stage to ensure the content's provenance stays verifiable. However, GAI tools that don't apply such marks and standards would still exist and be used by anyone who feels they have good reasons to.

#### ***Influence of AI on media coverage***

Social media and news aggregators' AI-driven recommender systems increasingly influence media coverage. While most journalists [tend to think](#) social media adversely affects journalism, a majority of them [admit](#) to relying on them to identify which topics and stories to cover, among other use cases.

This strongly [impacts](#) media coverage strategies and makes journalism vulnerable to platform leaders' arbitrary [internal decisions](#) and [external influence](#) strategies. Further, time spent on social media is mainly driven by attention-seeking design and recommendation systems, which risks skewing information gathering in favour of sensationalist, divisive, or misleading content. (More on this in Part 3).

In the near future, media outlets may be incentivised to use AI predictive tools for picking topics to cover, aligning media coverage strategy with their audience to reach targets, rather than with editorial choices or journalistic integrity. Such tools may be coupled with search engine bots for gathering content on various topics, diminishing the role of ethics and human judgement.

## ***Unreliability of language models***

Conversational search engines are rapidly emerging players that will increasingly shape how we find and access information. Although their outputs may appear human-like, Large Language Models (LLMs) merely simulate an understanding of human concepts such as truth, honesty, fairness, or harm. LLMs do not have an underlying model of reality and do not aim for truth. They're engineered to [generate plausible text](#), that is, text that is likely to follow from the input they're given, based on the patterns they've learned in their training data set and on the instructions they're given during the supervised learning stage. As a result, language models often generate convincing falsehoods, sometimes supporting them with [false quotes](#) and [false sources](#). Because LLMs are specifically trained to produce plausible texts, these falsehoods share the appearance of logic and truth.

## ***Privacy***

AI-driven tools are increasingly being used for data collection and surveillance. While these tools can be invaluable for gathering vast amounts of data quickly, the indiscriminate data collection can infringe on individual rights to privacy and intellectual property. (More on this in Part 3). AI systems may also [leak personal information](#), which has implications for the types of data journalists should or should not share with AI tools.

## **2. Information processing**

*This phase involves analysing, organising and contextualising gathered facts and content.*

### ***Independence, impartiality, and “no-harm principle” concerns***

As algorithms can analyse data and generate content in seconds, news agencies and media outlets are increasingly using them [to investigate](#) complex issues or to automate the creation of content based on data and statistics. These include financial reports, sports results, or weather forecasts.

While some newsrooms have quickly [integrated generative AI](#) tools like ChatGPT into their operations, as of 2023, only a minority have established guidelines. Many journalists use chatbots and GAI for story inspiration, contextualisation, or illustration purposes. Generated content ranges from templates and outlines to news stories, or even sometimes to [entire news and information websites](#) operating with little to no human oversight. The content produced is heavily influenced by an opaque set of training data, undisclosed pre-prompts, and reinforcement learning protocols determined by AI system providers.

GAI tool providers often [sideline ethics concerns](#) and do not provide robust guarantees towards journalistic ethics and standards. While they claim to train, fine-tune, and pre-prompt their models to prevent them from serving dangerous goals, their systems can easily be [jailbroken](#) and be used for [harmful purposes](#). While there might be little incentive to intentionally prompt a model to provide misleading content from a journalistic perspective, there's a significant risk of [inadvertently receiving](#) such content.

## ***Biases***

Biases can be more insidious than factual inaccuracies, because they subtly distort perceptions and reinforce systemic prejudices in much less perceptible ways which are

harder to verify. Research has [found](#) that in a given text, even the subtlest representation of a metaphor can have a powerful influence over how readers perceive a social problem and attempt to solve it.

LLMs and GAI tools [have shown biases](#) in various contexts, from gender and race to [political](#) and moral views. Like the [issue](#) of unreliability, this appears to be a structural issue. LLMs tend to reproduce or even [amplify](#) biases contained in their training dataset.

### ***Audience reach-optimisation***

GAI brings a wide range of new techniques to adapt content to the implicit incentives set by social media and news aggregators. LLMs can be coupled with basic methods such as A/B testing for titles and thumbnails, or be used to [draft entire articles](#) optimised for Search Engine Optimization (SEO) and recommender systems. There's a growing concern that AI-generated content, designed to maximise engagement, may inherently favour extreme, sensationalist, and divisive narratives, mirroring the [trends observed in social media](#) news feeds.

### ***Transparency, explainability, and openness challenges***

AI system providers offer varying levels of transparency concerning the design (code, parameters, training dataset, and methodology), performance (reliability, robustness, bias, etc.), and actual use-cases (e.g., recommendation systems, military or medical usage) of their products.

Organisations like the [UN](#) and [several governments](#) have advocated for increased transparency and explainability in AI, raising both technical and ethical questions.

On the technical side, while requesting any level of transparency regarding the design, performance, and actual usage of AI systems is feasible, this doesn't necessarily make them *explainable*. Neural networks, particularly Large Language Models (LLMs), are often described as black boxes. Their design [inherently prevents](#) the simple explanation of their functioning, making it virtually impossible to clarify how they arrive at specific conclusions.

On the ethical side, while high levels of transparency seem broadly desirable for both performance and usage, complete transparency in design presents risks and complex dilemmas. Requiring providers to release models' source code, as Meta did with [LLaMA](#), could make these models accessible to reckless or malevolent actors.

### ***Hyper-personalisation***

GAI makes it economically and technically feasible to hyper-personalise information in its format and content, potentially skewing public understanding of critical issues.

GAI tools are being used to produce synthetic [news anchors](#). In a media landscape competing for attention, such anchors may be modified to match viewers' preferences in voice and appearance and may even be designed to foster emotional bonds with humans.

### **3. Information dissemination**

*This phase encompasses the selection of the appropriate medium – be it print, broadcast, digital, or social media – and the actual distribution or publication of the content.*

#### ***Social media recommender systems***

As Herbert A. Simon famously put it in 1971, “a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

AI-driven recommendation systems have been social media’s primary response to this challenge. These systems, which organise billions of newsfeeds daily and autoplay hundreds of billions of videos on platforms such as Instagram and TikTok, have been successful from a certain viewpoint: 5 billion people use social media, and a typical user spends around [2 hours and 30 minutes](#) a day on it, predominantly engaging with AI-recommended content.

However, “efficiently” allocating attention has very different meanings depending on one’s goal and perspective. Social media’s recommendation systems mainly optimise for user engagement and ad-profits, which explains their economic success. As many studies have shown, this tends to [incentivize both coverage and dissemination](#) of sensationalist, dismissive, polarising news, possibly at the expense of trustworthy, impartial, in-depth journalism.

AI recommender systems are, at present, the most impactful technology concerning global information dissemination. They exert a major influence on trending topics online and on social media, significantly shaping the topics and perspectives journalists and media outlets give importance to.

#### ***Search engines and news aggregators***

Search engines also influence media coverage and content dissemination. For all geographical regions, Google Search and Google News account for between [40% and 75%](#) of the total externally driven traffic to publishers’ websites. In general, recommendation criteria are kept private, and may depart significantly from journalistic and media ethics.

Media outlets utilise various internal AI-based solutions to disseminate information, such as news feeds apps, search bars, push notifications, and newsletters, all of which can be personalised and optimised for audience reach objectives. An international survey conducted for Reuters Institute’s [Journalism, media, and technology trends and predictions 2023](#) indicates that two-thirds of publishers are experimenting with AI to “drive story selection/recommendations on [their] website and app”.

#### ***Conversational search engines***

Conversational search engines are emerging players in the information dissemination landscape. These platforms use natural language to respond to user queries, offering a personalised and interactive search experience. This raises numerous concerns about the criteria these tools employ in prioritising or excluding content according to their interests without checks and balances.

#### **4. Strategic positioning of media in the AI era**

*For safeguarding the right to information, it's important that the prevailing economic and legal incentive structures favour information organisations and professionals who genuinely uphold ethical values.*

##### ***Threats to ethics' economic incentives***

Media traffic has become increasingly reliant on gatekeepers competing for short-term attention. As the marginal cost of producing artificial articles and news content approaches zero, a strategy based on publishing a limited number of quality articles may face increasing competition from one based on generating a large volume of reach-optimised ones.

Media outlets might automate daily editorial decisions via AI systems to determine which topics are trending online, how to maximise user engagement based on the outlet's specific audience, and which angle and format to adopt. Media outlets cautious about generative AI might struggle against competitors who prioritise reach and cost-efficiency over informational integrity. This trend would gradually lead to a situation where the editorial process includes little to no human decision-making and results solely from the automatic aggregation of short-term private incentives.

##### ***Media outlets' relations towards AI providers and information gatekeepers***

Several publishers and media organisations [have accused](#) GAI providers of using their content to train their models without neither fair compensation nor explicit agreement. GAI tools produce content and build value based on the data they're trained on, often resulting in text that mirrors specific sentences or images that mimic patterns, leading to unoriginal and copied content. Low-quality news websites also use AI to [deceptively rewrite content](#) from mainstream news outlets. (In the USA, early court decisions have [prevented copyrighting](#) visual content generated by AI.) Further, foundational models [have been accused](#) of absorbing a vast amount of publicly available information and content and walling it off into proprietary products.

Conversational agents may divert media outlets' traffic while at the same time using their data, leading to advertising and subscription revenue losses. If not compensated for, this could threaten journalism's economic sustainability and give increasing importance to [anti-competitive](#), [web-destroying](#), but nevertheless compelling AI-chatbots.

Media outlets are also confronting a growing asymmetrical power dynamic with social media companies. This disparity is evident in several areas, including advertising revenues and editorial decisions, and amplifies as advertisers increasingly shift from traditional media to social media platforms.

##### ***Journalism automation***

GAI suggests a potential for significant [staff reductions](#) in media outlets. Some (poor quality) news sites are already [publishing articles written largely or entirely by bots](#). Others may soon [follow suit](#), as well as [tech companies](#).

Because GAI systems do not directly collect facts by conducting interviews, taking pictures, or attending meetings, they cannot update their view unless they are explicitly trained on new data to take into account a new piece of information. If a substantial portion of media revenues is diverted from information gathering, fewer original material may be collected.