

AI and the Right to Information

*RSF's 7 recommendations to protect
an AI-dominated information space*

To prevent Artificial Intelligence (AI), including generative (GenAI), from becoming a weapon of massive disinformation, RSF is calling for comprehensive regulation targeting AI and the right to reliable information, strong public policies supporting reliable AI systems in the information ecosystem and reinforcing global democratic AI governance.

These recommendations stem from various projects and initiatives by RSF targeting AI and information : the "[Paris Charter on AI and Journalism](#)", which establishes the first global ethical framework guiding media in the AI era, the [Spinoza project](#), a language model developed by and for the media and the report "[AI as a Public Good: Ensuring Democratic Control of AI in the Information Space](#)" by the Forum on Information and Democracy with RSF's input.

I. REGULATION

Recommendation 1: Require AI systems providing access to news and information to amplify diverse and reliable sources.

- Classify AI systems in news dissemination as high-risk and mandate independent evaluations.
- Require these systems to enhance the visibility of reliable information sources.
- Mandate these systems to uphold fundamental journalistic values and support pluralism.

Recommendation 2: Establish a clear accountability framework to prevent the creation and dissemination of harmful deepfakes.

- Implement a comprehensive accountability regime with criminal sanctions for the intentional publication of harmful deepfakes.
- Mandate AI providers to prevent generation of harmful, non-consensual deepfakes.
- Encourage the adoption of robust authenticity standards for recording equipment, media, and digital platforms.

Recommendation 3: Establish enforceable rights for media outlets and journalists in the face of AI.

- Ensure transparency and the right for media and journalists to opt-out when their content is used to train AI or generate AI content.
- Secure fair compensation for content usage through independent and transparent distribution.

II. PUBLIC POLICIES

Recommendation 4: Support and fund AI safety and independent evaluations.

- Mandate proportional funding for AI safety.
- Support and fund independent evaluation of AI systems in news and information.
- Maintain a European platform documenting AI systems and their impact studies.
- Implement special liability regimes for AI developers accountable for harms in the information space.

Recommendation 5: Recognize news and information as a priority domain for the development of trustworthy AI systems.

- Invest in public interest AI systems and digital platforms.
- Fund the development of a large language model grounded on European media data.
- Maintain a public repository of all legally accessible data for training and grounding AI systems.

III. GOVERNANCE

Recommendation 6: End the AI arms race and promote international cooperation for AI governance.

- Strengthen international cooperation within the Partnership on Information and Democracy.
- Establish global standards and regulations for AI systems in the information space.
- Involve media and journalists in international AI governance.

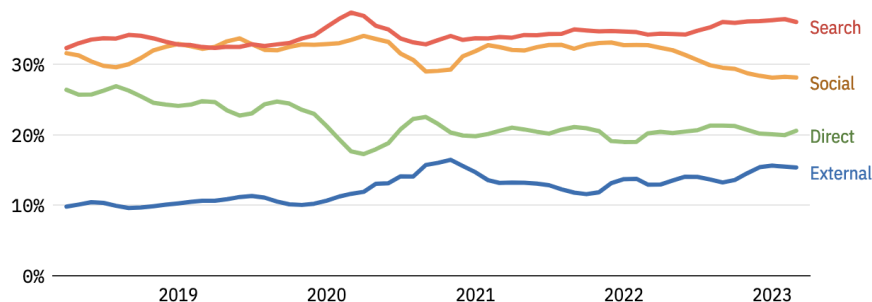
Recommendation 7: Ensure that regulatory and oversight capabilities for AI consistently match the industry's influence and power.

- Allocate resources proportional to the economic power of the AI sector for regulation and oversight.
- Regulate the pace of AI development to ensure alignment with safety measures and governance frameworks.
- Ensure AI Office consultations primarily involve independent research institutions and stakeholders who advocate for the public interest.

I. Regulating AI to prevent harms to the right to reliable information

Recommendation 1: Require AI systems providing access to news and information to amplify diverse and reliable sources.

Search engines and social media are the primary sources of traffic for online news websites, accounting for about [two-thirds of their reach](#). Media's heavy reliance on these platforms to reach their audience threatens the independence and sustainability of journalism. Specifically, two companies, Google and Meta, hold a globally dominant position in the distribution of news and information: **five billion people access information daily through their platforms and services**.



Source: Press Gazette analysis of Chartbeat data

PressGazette

[Share of traffic](#) of 546 UK and US news websites by origin. Red: search engines (~95 % from Google); Orange: social media (~80% from Facebook); Green: direct landing on media web pages; Blue: other external web pages.

Half of the world's population spends [two and a half hours](#) daily on social media, primarily consuming content recommended by AI. Recommender systems prioritise advertising revenues and user engagement over quality, amplifying [polarising, harmful, and sensationalist content](#). The leaders of these platforms have significant discretionary power and lack strong checks and balances, enabling them to shape the dissemination of information according to their interests or ideology.

Examples of harms

- Following a unilateral decision by their leaders, between the end of 2022 and the end of 2023, Facebook and X respectively [slashed the visibility of journalistic news sites](#) on their platforms by 52% and 27%.
- On Tiktok, Dylan Page, a young news influencer from the UK, receives more regular video views than the BBC or New York Times combined, even on important stories like Gaza.
- During the first four months of the Covid-19 pandemic, 34 disinformation-hosting websites [gathered](#) 80 million interactions on Facebook, which is more than ten times the traffic that the WHO received during the same period. Analysts [estimate](#) that hundreds of thousands of deaths could have been avoided with a safer information environment.

- Social media platforms significantly amplify mis- and disinformation, often with devastating consequences, notably by [triggering violence in Tigray, Ethiopia](#), and [fueling hate against Rohingyas in Myanmar](#).
- During the 2016 US presidential election, YouTube algorithm was [six times more likely](#) to recommend videos that favoured Trump over Hillary Clinton, presumably playing a significant role in the result of the election.
- During the 2023 Swiss and German elections, Microsoft Copilot was [found](#) to make mistakes in 30% of the cases when answering factual election-related questions. Incorrect answers were often wrongly attributed to media sources.

Current legal framework

The Digital Services Act (DSA) mandates very large online platforms (VLOPs) and very large online search engines (VLOSEs) to:

- disclose criteria being used by their recommender systems and options to modify them (art. 27)
- offer at least one version of their recommender system that does not rely on profiling (art. 38) (N.B.: experience shows that only a very small fraction of users select an alternative option when the default mode of the recommender system maximises engagement, particularly when the option is buried in the settings.)
- assess and mitigate the “systemic risks” raised by their services (art. 34 and 35)
- be subject to regular external audits aimed to assess whether the identified systemic risks are being properly monitored and mitigated by the platform (art. 37)
- provide vetted researchers access to their data for the same purpose (art. 40)

Under the AI Act :

- AI systems used in critical infrastructure, public services or employment are identified as high-risk and must undergo self-evaluation, adversarial testing, and risk mitigation actions;
- Providers of General Purpose AI (GPAI) models posing systemic risks must conduct self-evaluation, adversarial testing, and mitigate risks;
- Developers of GPAI models must provide a detailed summary of the training data and follow EU copyright law;
- **AI systems used in news and information are not classified as high-risk.**

Detailed recommendation

AI systems that play a structural role in the dissemination of news and information, such as VLOPs, VLOSEs, or conversational agents answering news and information questions, should:

- Be considered as high-risk.
- Be required to enhance the visibility of reliable information sources, identified as such by self-regulation standards like the [Journalism Trust Initiative](#) (JTI).
- Provide robust assurances demonstrating adherence to fundamental journalistic values (accuracy of reported facts, impartiality, non-harm, non-discrimination, accountability for mistakes) as well as pluralism when they reuse or reprocess journalistic information. Such standards must be designed and implemented by the European AI Office in collaboration with representatives from the journalism and media industry.

- Undergo independent and comprehensive evaluations to demonstrate compliance with the aforementioned obligations.

Recommendation 2: Establish a clear accountability framework to prevent the creation and dissemination of harmful deepfakes.

Within just one year after generative AI tools became publicly available, the volume of images created by AI [surpassed](#) the total number of photographs taken by humanity in the first 150 years of photography.

This surge in realistic synthetic content poses a dual threat: on the one hand, it risks misleading public opinion by fabricating facts and events; on the other hand, it leads to systematic doubt about the authenticity of all content, including genuine one.

The AI Act requires AI providers to ensure that AI-generated content is marked in a machine-readable format to indicate its artificial nature. AI deployers must disclose the artificial origin of deepfakes. However, transparency alone is insufficient: deepfakes, even when identified as such, may still harm people, create confusion, and pollute factual reporting. Additionally, malicious users can easily circumvent transparency obligations by using open-source tools. Moreover, detection tools [are not reliable](#).

Examples of harms

- Deepfakes are [frequently used](#) during elections. The technology can deceive and confuse voters, casting doubt on what is real.
- Thousands of false advertisement campaigns published on Meta platforms featured [deepfake videos of well-known Danish journalists](#). Most of the fake content was not removed by Meta despite being reported as fraudulent.
- ElevenLabs permits [replicating anyone's voice](#) based on a few seconds of audio, with safety guardrails easily [bypassed](#).
- Meta [said](#) that it relied on a team of a hundred independent fact-checkers to identify deepfakes. For a platform counting 3 billion regular users, that makes 1 moderator for every 30 million users, highlighting the inadequacy of current measures to effectively address the scale of the problem.

Current legal framework

Under the AI Act :

- providers of AI systems are required to ensure that all AI-generated content is labelled in a machine-readable way;
- creators of deepfakes must transparently display the synthetic nature of the content.

Under the DSA, the commission has published [guidelines](#) for the mitigation of systemic risks online for elections. In particular, VLOPs and VLOSEs are asked to assess and mitigate specific risks linked to GenAI, for example by clearly labelling deepfakes.

[France](#) and [Italy](#) have established criminal penalties for the creation and dissemination of

deepfakes, respectively, when they are made without the consent of the targeted individuals and when they cause harm to someone.

Detailed recommendation

- A clear and comprehensive accountability regime should be established for the deliberate creation and dissemination of harmful deepfakes, including proportionate criminal sanctions for the intentional publication of deepfakes that cause harm or are created without the consent of the individuals involved.
- AI providers should be required to robustly prevent the generation of harmful and non-consensual deepfakes. Everyone has a right to image privacy and should not have their appearance stolen.
- The EU should encourage the development of robust authenticity standards for recording equipment manufacturers (cameras, etc.), and the rapid adoption of these standards by media and digital platforms.

Recommendation 3: Establish enforceable rights for media outlets and journalists.

The growing role of intermediaries between media and their audiences has led to a significant loss of advertising revenue for media outlets, as a substantial portion of these revenues is captured by search engines and social media using asymmetric bargaining power. Recently, most journalistic content available digitally worldwide has been scraped by AI companies to train their models, with no consideration for copyright laws or content creators [content creators](#). Since conversational agents need to be grounded in up-to-date content to improve their accuracy, AI companies continuously seek journalistic content. In response, a growing number of news publishers (~ 50%) [have blocked AI companies' crawlers](#) from accessing their content, while AI companies have started licensing content from a few news publishers.

Chatbots using media content pose two threats to the sustainability of journalism. First, they risk diverting media outlets' traffic by using their data while providing insufficient compensation. Second, they risk exerting opaque, arbitrary power in choosing which media content is licensed and which remains invisible on their systems.

Example of harms

- The news industry undergoes [massive layoffs](#) while AI companies experience [their fastest growth](#) in history, partly due to value built on copyright data they have not paid for.
- OpenAI has signed deals with a few publishers, such as Axel Springer and Le Monde, risking inequitable access to ChatGPT's 100M+ regular users for large outlets (including tabloids like Bild) while further decreasing the audience for smaller outlets.

Current legal framework

The AI Act mandates general-purpose AI (GPAI) systems, such as certain chatbots, to meet transparency requirements, including compliance with EU copyright law and publishing detailed summaries of the content used for training. However, the legal landscape is marked by a lack of

clarity in copyright laws relevant to AI, such as the American 'fair use doctrine' and the European Text and Data Mining exception.

The current legal framework does not guarantee fair remuneration or satisfactory bargaining conditions for content creators and media outlets.

Detailed recommendation

Enforceable rights for media organisations, journalists, and content creators should be established regarding the use of their content in AI training and applications. These rights should include. These rights should include:

- The right to transparency: they must be informed when their content is used to train AI or generate AI content.
- The right to opt-out: they should be able to decide whether their content can be used by AI systems.
- The right to fair compensation: they must be fairly compensated whenever their content is used to train and ground an AI system. This could be facilitated through a revenue-based tax on AI companies or by defining unauthorized use as a copyright infringement.

These measures should be implemented through collective negotiations with publishers and journalists' representatives. Resources collected through these mechanisms should be distributed independently and transparently, supporting editorial independence and ensuring the sustainability of diverse, public interest media outlets.

II. Public policies supporting trustworthy AI systems in the information ecosystem

Recommendation 4: Support and fund AI safety and independent evaluation.

A vast majority of the efforts to build and evaluate AI systems come from the private sector, leading to conflicts of interest, neglected or underestimated risks, [profit-driven](#) rather than public-interest research efforts, [threats to independent evaluation](#), and disproportionate lobbying power. [Only about 2%](#) of scientific research publications in AI globally address safety issues. The imbalance regarding funding is even worse: there is a thousand-to-one gap between the resources dedicated to enhancing AI capabilities versus those allocated to improving AI safety. AI companies also [drain talent from universities](#) and quickly [expand their influence](#) in academia.

Examples of harms

- Today, roughly [70% of individuals holding a PhD](#) in artificial intelligence get jobs in private industry, compared with 20% two decades ago. The industry now predominantly drives AI research, significantly influencing the purposes for which AI is used, as well as its evaluation and regulation.
- A [2023 survey](#) revealed that 70% of top AI scientists thought AI safety research should be prioritised more than it currently is.
- Big technology companies frequently [fire](#) or [trim down](#) their ethics teams, especially when [they raise alarm](#) about risks or harmful impacts.

Current legislation

The AI Office was established as a European Commission agency with the aim to be the central hub of AI expertise and the cornerstone for a unified European AI governance system. It aims to foster trustworthy AI across the internal market in collaboration with relevant public and private actors. The EU AI Office will play a key role in implementing the AI Act and preparing delegated acts, in particular by developing tools and methodologies for evaluating the capabilities and reach of general-purpose AI models, classifying models with systemic risks, and drawing up state-of-the-art codes of practice, and investigating possible rule violations.

Detailed recommendation

- Require proportional funding for safety: for every dollar invested in AI capabilities, AI developers should be required to allocate a proportionate amount to AI safety measures and governance frameworks.
- Impose strong, provably safe requirements on AI models to ensure they comply with recognized technical, ethical, and legal standards before their release.
- The EU should ensure public access to a comprehensive European platform that documents AI systems, including detailed information on vulnerabilities, independent impact studies, evaluations, and recommendations.
- Adopt special liability regimes holding AI developers accountable for the harms caused by their systems. Collaborate with relevant stakeholders, including journalists and the media industry, to determine in which cases of fundamental rights violations a strict

liability regime should apply to developers and deployers of AI systems used for output generation or personalization.

Recommendation 5: Recognize news and information as a priority domain for the development of safe and trustworthy AI systems.

The current state of the AI industry, characterised by a few private companies developing and distributing general-purpose AI models to thousands of enterprises in the information sector, raises concerns for the independence of journalism, the technological sovereignty of the media, the protection of confidential data, and the respect for intellectual property. Technology companies already hold an oligopolistic position in the dissemination of information. Generative AI could enable them to extend their dominance to information production, posing a significant risk to the independence, pluralism, and sustainability of journalism.

Examples of harms

- Until 2014, the most significant machine-learning models were released by academia. Since then, the industry has taken over. Today, investment in AI and the development of AI models is [largely dominated](#) by American big tech companies.
- “European” AI start-ups have extensive [transatlantic ties](#) and are in fact largely influenced by American big tech companies.

Current legislation

In January 2024, the European Commission launched an [AI innovation package](#) to support small-to-medium enterprises in developing trustworthy AI that aligns with EU values and regulations. The 'GenAI4EU' initiative, part of this package, aims to foster the development of novel use cases and emerging applications across Europe's 14 industrial sectors and the public sector. These applications include robotics, health, biotech, manufacturing, mobility, climate, and virtual worlds, yet **they notably exclude the news and information sector.**

Detailed recommendation

- GenAI4EU should recognize news and information as a priority domain that requires funding to develop public interest AI tools.
- In particular, the EU should fund the development of a large language model grounded on European media data, integrating high standards of pluralism and reliability of information, such as [the Spinoza project](#).
- The EU should support the open data movement and maintain a public repository of all public and private data lawfully accessible to train and ground AI systems, specifying conditions under which it is accessible to AI developers and deployers.

III. Reinforcing global democratic AI governance

Recommendation 6: End the AI arms race and promote international cooperation for AI governance.

Competition among big technology companies is mirrored at the state level, where national governments, aware of the strategic importance of these technologies for maintaining geopolitical leadership, are incentivized to promote less stringent regulations within their borders. [Dominant AI narratives](#) in the EU often make use of age-old international relations concepts such as great power competition, geopolitics, and arms races. Whether true or false, these narratives undermine regulation efforts and create self-fulfilling prophecies. The perceived risk of the EU being surpassed by the US, or Western countries being overtaken by China, is often used as justification for deregulating the sector. However, poorly regulated technologies in the information space can have—and have had—devastating impacts in the very countries where they are developed, as demonstrated by unregulated social media. Even in domains where innovation may lead to short-term geopolitical advantage (like [AI-weaponry](#)), an AI arms race is a collectively losing strategy associated with existential threats to human societies.

Current legislation

In March 2024, United Nations General Assembly adopted by consensus a resolution on [“Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development”](#)—the first-ever to establish a global consensus approach to AI governance, encouraging Member States and multi-stakeholders (**including the media**), “to develop and support regulatory and governance approaches and frameworks related to safe, secure and trustworthy artificial intelligence systems [...], recognizing that effective partnership and cooperation between Governments and multi-stakeholders is necessary in developing such approaches and frameworks”.

In April 2024, France and China signed a [Joint Declaration on Artificial Intelligence and Global Governance](#), acknowledging the risks associated with this technology and calling for “strengthening global AI governance with the aim of promoting development and use for the common good.” Specifically, *“Both parties will also strive to enhance international cooperation and interoperability between AI governance frameworks and initiatives, building on the work carried out at the United Nations level”*.

Detailed recommendation

Only global, multilateral public governance can ensure the long-term public interest, the preservation of human rights, and innovation for the benefit of all, over the pursuit of short-term profits and particular national interests.

- Democratic governance of the information and communication space is a prerequisite for democracy itself. The EU must ensure that democratic principles are the basis of international AI governance by strengthening cooperation within the [Partnership on Information and Democracy](#).
- The EU must work to establish global standards, principles, and rules as recommended above, to be taken into account during the development, deployment, and use of AI systems in the information space.

- As essential guardians of the right to information, journalists, media outlets, and journalism support groups should play an active role in the international governance of AI systems.

Recommendation 7: Ensure that the regulatory and oversight capabilities for AI consistently match the industry's level of influence and power.

There is a growing mismatch between exponential technological innovation and lagging normative and regulatory frameworks, particularly in protecting democratic values and human rights principles. In March 2023, over 1,000 industry experts signed an open letter urging a pause in AI development to assess and mitigate the dangers of systems like GPT-4. The development of AI is predominantly driven by private companies seeking a dominant market position, often prioritising speed over safety (see section 2.1). Their products and services sometimes reach hundreds of millions of people without any prior independent control or evaluation. Furthermore, AI developers currently face little to no consequences for the downstream harms caused by their models, leading to rapid and potentially reckless deployment.

Examples of harms

- AI companies exert [major lobbying power](#) at the EU. In 2023, [78 percent of meetings](#) on AI of high-level EU commission officials were with corporate representatives, influencing regulation to align with their interests (despite big tech officials publicly [calling for](#) global AI regulation).
- A [2023 survey](#) shows that an overwhelming majority of American voters are concerned about AI risks with 72% wanting to slow down its development compared to just 8% who prefer speeding it up. The poll indicates a lack of trust in tech executives to self-regulate AI, with 82% favouring federal oversight.

Detailed recommendation

- Allocate resources for evaluating, legislating, and enforcing AI regulations proportional to the economic power of the sector. This requires providing sufficient funding and manpower for regulatory bodies by allocating 10% of AI companies' annual budgets to international AI governance and oversight.
- Control the pace of AI development and deployment. The speed of AI development should be constrained by our ability to control these systems and implement robust safety measures and governance frameworks.
- Ensure AI Office consultations primarily involve stakeholders advocating for the public interest, such as civil society organisations and independent public research institutions.