The Data Tank

Media, Democracy and Generative AI: A Critical Juncture

A literature review of the current context and ways forward

May 2026

Anna Colom, Elena Murray, Marta Poblet

Acknowledgements

This literature review is part of the 'Strengthening Media in the age of Generative AI' action-research and dialogue by the non-profit organisation The Data Tank, a think-and-do tank based in Belgium. We are grateful to the Adessium Foundation for the funding and to the following reviewers of this report for their time and expert engagement:

Stefaan Verhulst (internal reviewer)
Co-Founder, The GovLab and The Data Tank

Marius Dragomir (external reviewer)
Director, Media and Journalism Research Center

Paul Keller (external reviewer)
Director of Policy, Open Future

Sameer Padania (external reviewer)
Expert consultant on independent journalism and information ecosystems

May 2026

Recommended citation:

Colom, A.; Murray, E.; Poblet, M (2026) Media, Democracy and Generative AI: A Critical Juncture. A literature review of the current context and ways forward. Belgium: The Data Tank.

The Data Tank

ADESSIUM FOUNDATION

Table of contents

Executive summary

The irruption of Generative Artificial Intelligence (GenAI) represents a watershed moment for public interest media in Europe and globally. This report presents the main insights from a literature review drawing on academic, industry, and policy sources to map the challenges that Generative AI poses to the media ecosystem – particularly small and medium-sized organisations – as well as the solutions being proposed to address them.

The literature review includes a total of 221 sources resulting from a two-pronged data collection process: (i) a systematic approach based on seven search strings from the year 2023 onwards, and (ii) a non-systematic discovery based on the authors' professional networks and related snowballing and ecosystem mapping. The analysis followed a thematic approach and the themes identified were categorised based on whether they related to the diagnosis of the current context or to the potential ways forward.

On the diagnosis side, our review identifies multiple, interconnected challenges. Namely: Generative AI providers scrape journalistic content at massive scale while returning negligible traffic or compensation. AI bot traffic also imposes disproportionate infrastructure strain, which is even harder for smaller organisations to cope with. Meanwhile, although exact figures on declining traffic vary, 'zero-click' searches are diverting audiences away from publishers. Together, these trends endanger the business models of news organisations, already walking on thin ice following the platformisation initiated by social media companies. Likewise, opacity in Generative AI models and outputs, inherent errors in both content accuracy and attribution, the limitations of Generative AI to summarise journalistic content in context-specific and nuanced ways, and the bypassing of original sources and editorial gatekeepers risk undermining the integrity of information as a key pillar for democracy. The growing concentration of Generative AI power in a few hands, as well as concentration in ways information is generated (with English-centric and large national media favoured over more local or diverse sources), further erodes pluralism, equity, and diversity.

Taken together, these challenges impact media business models, information integrity and democracy, raising questions about the adequacy of existing regulatory frameworks. Bilateral licensing deals leave smaller media outlets behind, and do not address issues of

information integrity and pluralism. They also have an impact on brand and trust.

Ultimately, the changes driven by Generative AI are seen to affect the whole political economy of the media and information ecosystem, in a way that cannot be fixed by technical solutions, literacy and skills, or bilateral licensing deals alone. This report, then, covers potential solutions identified in the literature, which are emerging at multiple levels:

- Technically, tools such as robots.txt (the file in web pages signalling preferences on crawling) and other protocols, pay-to-crawl systems, and attribution standards give publishers some control over content use, even if enforcement remains an issue.

- Licensing arrangements —direct, collective, and statutory— are also developing, though their adequacy for smaller organisations remains equally contested.

- Regulatory proposals include copyright reform, competition regulation, and levy mechanisms to redirect AI revenues back to public interest media actors.

- In parallel, there are calls for new audience-focused business models, increased public and philanthropic investment, data commons, and collective action by media coalitions. Media, data and AI literacy for both media and audience remain necessary, but not sufficient on its own.

While these solutions differ in depth and breadth, the literature agrees on the need for a holistic, coordinated response beyond single interventions. There is a sense that, for the media ecosystem to keep fulfilling its role in a democratic society, rebalancing power asymmetries between media organisations and large Generative AI providers will require collective action as well as institutional imagination.

1. Introduction

The emergence of Generative AI represents a watershed moment for public interest media in Europe and globally. As Large Language Models (LLMs) and AI-powered search engines increasingly mediate how audiences access and consume information, news publishers, public service broadcasters, and small and medium-size independent media face new challenges to their sustainability, relevance, and societal role. This literature review synthesises recent scholarship and media industry analysis to map both the landscape of problems confronting public interest media and the emerging solutions being proposed and implemented across regulatory, technical, economic, and social domains.

The aim is to offer a comprehensive snapshot of both the challenges facing media, information integrity and democracy, as well as the existing or potential solutions and pathways to protect independent and public interest media and the role it plays in societies.

2. Methodology note

This literature review is guided by the following research questions:

1. What are the key challenges documented until now on the bargaining power and sustainability of media organisations in Europe vis-a-vis large Generative AI providers?

2. What approaches and ways forward (regulatory, technical, collective action, tools and literacy, etc.) have been attempted and documented until now for media organisations to leverage their power?

3. How do these questions relate, in particular, to small and regional media organisations and networks in Europe?

The literature search employed two data retrieval mechanisms: a systematic review of literature and a non-systematic discovery based on the authors' professional networks and related stakeholders' mapping and snowballing.

The systematic search was done using seven Boolean search strings and results were filtered using custom date ranges to include publications from January 2023 onwards (see appendix for details). This time constraint was chosen due to the rapidly evolving and changing AI media landscape, to avoid outdated or irrelevant information. While a number of foundational arguments relevant to the topic (such as platformisation, news-media bargaining, and copyright reform) pre-date 2023, these were selected as part of the analysis and argument development from some of the recent sources themselves, as well as sources from the non-systematic part of the search, which includes documents prior to 2023. The results were then screened in two phases to check for duplicates and relevance, as detailed in the appendix.

The second strand of search is based on the authors' networks in addition to a stakeholder mapping conducted by the authors. It is therefore relevant to mention the authors' specialisms and positionality, which encompass the fields of media and journalism, data and AI governance, law and innovation, open-knowledge, civic tech, and democracy. A total of 221 sources were included in the final analysis. The sources collected came from both academic and grey literature to ensure the review encompassed the latest and most relevant initiatives

and reflections on the topic from the media industry and relevant stakeholders in addition to academia. While the 221 sources have been included in the analysis and played a role in the identification of themes, not all are referenced in this report.

Thematic analysis was conducted to analyse and annotate key insights from the articles included in the final review. This analysis followed established methodological approaches, including close reading of the literature, extraction of relevant data using quotes, and annotation and coding into themes and sub-themes developed by the researchers. Airtable, a relational database platform, was used to facilitate data annotation and labelling. Each article was analysed and coded individually by one of the authors, and predominant themes and patterns were identified by labelling and annotating data points with a main theme and a sub-theme, which were developed throughout the coding process, in close consultation and discussion between the principal researchers. This inductive approach resulted in an overarching binary categorisation based on diagnosis-oriented themes on the one hand, and solutions-oriented themes on the other. This frame is used to structure the report but the limitations of this simplification are probed and untangled in each sub-section where, for example, what are presented as potential solutions may be amplifying some of the asymmetries presented in the diagnosis section, such as bilateral licensing deals between large media publishers and AI companies. The literature identified through this two-pronged process is biased towards sources from Western and English-speaking contexts; an additional limitation which highlights the need to rebalance power in access to academic and policy publications.

3. Defining public interest media

The findings of the literature review highlight a number of wide-ranging implications of Generative AI technologies on the broader journalism ecosystem, affecting a range of actors from commercial newsrooms to major public broadcasters. However, these effects are more likely to be particularly significant for smaller and independent media organisations. Given the importance of epistemic integrity and the maintenance of a healthy public sphere in discussing media futures, our analysis places a particular focus on the risks and consequences of Generative AI for public-interest media, and therefore commences with how public-interest media are defined and understood.

The literature reviewed offers several complementary conceptualisations of both public interest media and public service media that provide important framing for understanding the challenges posed by Generative AI. Public interest media (PIM) and public service media (PSM) are often used as synonyms or overlap considerably, and sometimes differ by placing the focus on different institutional and functional aspects.

An often cited definition of public service media refers to 'the provision of media services, across devices and platforms, that contributes to the democratic, cultural and social well-being of society' (Picone and Donders, 2020 in Campos-Rueda & Goyanes, 2023). This definition is intentionally broad, encompassing diverse operational models while emphasising the societal purpose that distinguishes public interest media from purely commercial enterprises. Yet, it is also noted that 'in Europe, public media services are almost exclusively entrusted to public broadcasting entities' (Campos-Rueda and Goyanes, 2023).

Similarly, Dragomir et al. (2024) observe that 'the EU has accorded a high priority to PSM within its media policy framework', while also noting that 'the onus of regulating PSM entities rests with national governments', underscoring the institutional and jurisdictional specificity of PSM.

PSM are also often considered to have a tripartite foundation of public funding, public control, and public service mandate, which serve as a structural function that distinguishes PSM from commercial media and in principle protects them from market forces (Thomass, 2024).

Social purpose as a defining trait

The concept of 'public interest media' appears in the examined literature as a broader, more functional category defined by purpose and orientation rather than by its institutional form. It includes organisations that serve the public interest in information provision regardless of their specific ownership, funding, or governance structures.

For example, quoting an expert in the field, Simon (2023) reports the definition of public interest media as 'free and independent' outlets whose role is to 'inform people on the issues that shape their lives, in ways that serve the public's rather any political, commercial or factional interest, to enable public debate and dialogue across society, and to hold those in power to account on behalf of the public interest' (James Deane in Simon, 2023).

Nel and Rymajdo (2024) identify a few characteristics of highly relevant public-interest news: (i) quality: covering 'emphasised accuracy, engagement, and informativeness, avoiding sensationalism, and maintaining trust and balance' as well as 'robust fact-checking mechanisms'; (ii) value: addressing 'pressing public health concerns, contribute to community well-being, and empower individuals for active civic participation'; (iii) enabling decision-making: 'informing citizens about society, civic processes, and local issues, fostering participation and awareness'; and (iv) holding power (government and institutions) accountable, 'promoting transparency and accountability within society' (2024).

A democracy pillar

Van den Bulck et al. (2025) highlight the distinction between serving citizens rather than consumers and elaborate on normative principles underlying PIM and PSM like universality (ensuring content 'reaches every citizen the organization is expected to address') and distinctiveness (being 'distinct by virtue of the functions it performs and the value it brings to society'). Dutkiewicz quotes McGonagle to refer to this distinctiveness and universality: 'in its essence, public interest content is content – information, ideas, opinions and data – that relates to matters of interest to society. Such content helps individuals to make informed opinions and decisions, which in turn helps them to participate in public debate and in public affairs more generally' (McGonagle, cited in Dutkiewicz, 2024).

This social and public service purpose is connected to democratic functions. Thus, Van den Bulck et al. (2025) introduce the concept of 'epistemic welfare' as a framework for understanding this democratic function, defining it as 'concerned with creating and maintaining conditions and capabilities for epistemic agency of citizens in the public sphere'. This framework positions public interest and public service media as institutions that should empower citizens 'to reach—what social epistemology calls—epistemically valuable states that encompasses (i) having true beliefs, (ii) avoiding errors, (iii) having justified beliefs, (iv) having rational beliefs (or partial beliefs), and (v) having knowledge' (Goldman et al., 2011).

This essential democratic role is threatened by the way technological, economic, and political pressures undermine the integrity of information systems around the world (D'Arma et al., 2025). For many, this makes public service media a critical infrastructural asset:

> [A]s the functions carried out by PSM are reshaped by the digital environment, the role it plays as an aspect of the complex, interdependent systems of communication needed to support democracy and national security is such that PSM is now akin to other infrastructure sectors of national importance including transport, energy or healthcare. (Doyle et al., 2025)

Ultimately, it has been a common defining trait of PSM and PIM that it is service and public-oriented, relying not on the market but on 'its core role to preserve and promote democracy by providing access to unbiased information and facts on which to build common ground and encourage a fruitful debate (Sehl, 2020 and Donders, 2021 in Campos-Rueda & Goyanes, 2023).

From these different definitions, we can synthesise a working concept of public interest media as institutions characterised by: diverse funding sources; governance structures that protect editorial independence; a mission to serve their audiences as members of a democratic polity rather than as consumers; commitment to universality and accessibility; production of content that serves democratic, cultural, and social functions; and a role in maintaining epistemic standards and the integrity of the public sphere.

4. Diagnosis of the current context

The analysis of the literature identifies a stream of themes referring to the diagnosis of the current situation. The impacts identified and developed in this section refer to: (i) extractive content practices; (ii) infrastructure strain; (iii) declining traffic and audience displacement; (iv) sustainability of business models; (v) threats to information integrity and democracy; (vi) inadequate legal frameworks; and (vii) attitudes towards Generative AI by both media professionals and audiences.

4.1 Extractive content practices

A prominent theme in the literature is the extractive relationship between Generative AI providers and newsrooms and media professionals, although a clear distinction between the use of content for the purpose of training models and the use of content as input into trained models (and the different mechanisms involved in these two phases) is not always made explicit in the sources.

Intensive web scraping illustrates extractive behaviour with measurable figures. Radsch (2025) shares recent data from Cloudflare, a content delivery network and edge-security provider that observes a substantial share of global web traffic: 'Cloudflare's data paints a stark picture of how AI crawlers operate at indefensibly extractive levels: OpenAI's scraping-to-referral ratio is 1,700:1 and Anthropic is at 73,000:1.'

This extractive dynamic also extends to crawlers that do not belong to AI companies. This is the case of Common Crawl, a not for profit organisation that has been crawling the web and providing free datasets of web pages since 2007. The issue, as raised by Reisner, is that Common Crawl does not log in to websites but it does not execute the paywall code that checks subscriber status 'so it gets the full articles'. Reisner estimates that 'the foundation's archives contain millions of articles from news organizations around the world' (Reisner, 2025).

The implications of the 'scraping-to-referral ratio', or the 'crawl-to-refer ratio'— pages a tool crawls vs. times it drives users to a website – as Tome (2025) puts it, shows 'the imbalance between how much content AI systems consume and how little traffic they return. For publishers, it can feel like giving away the raw material for free.' As he continues:

> If training-related crawling continues to dominate while referrals stay flat, creators face a paradox: feeding AI systems without gaining traffic in return. Many want their content to appear in chatbot answers, but without monetization or cooperation, the incentive to produce quality work declines. [...] The Web now stands at a fork in the road. Either a new balance emerges — one where the new AI era helps sustain publishers and creators — or AI turns the open web into a one-way training set, extracting value with little flowing back. (Tome, 2025)

Others have defined this relationship as a 'vampiring' rather than a 'symbiotic' one, and therefore a 'fundamentally different economic and technical proposition' from traditional crawling that was based on more semi-reciprocal benefits (I & Cath, 2025).

4.2 Infrastructure strain and technical burdens

Beyond economic extraction, AI bots impose significant burdens on the technical infrastructure of media content providers. AI data scrapers (harvesting data from the web for training purposes) and AI crawlers (gathering specific information about web pages and their links) now make up a substantial part of Internet traffic.

Data from Cloudflare Radar show that, for the news and publication industry, the most prominent AI bots scanning public websites to collect data used in search engines, AI model training, and other data processing tasks are Googlebot (41.2%) Bingbot (14.7%) ChatGPT-User (8.4%) GPTBot (7.5%) and Meta-ExternalAgent (7.4%) (Cloudflare Radar, May 2026).

The activity of AI bots affects not only media publishers but also cultural heritage institutions, open access repositories, and nonprofit information providers that are critical components of the public knowledge ecosystem. Wikipedia, a foundational resource for AI training, exemplifies these pressures. Tremayne-Pengelly reports Wikipedia sources observing bot traffic 'at an all-time high':

> In the early days of LLMs, tech companies told Wikipedia they only needed to train their models on its data once a year. "That's how it started," Lane Becker, senior director of earned revenue at the Wikimedia Foundation, told Observer. "Now, it is rapacious," he added. As Wikipedia becomes more central to the infrastructure A.I., the

> organization is grappling with rising bot traffic, the need for attribution and how to sustain its ecosystem in the face of powerful new users. (Tremayne-Pengelly, 2025)

Hinchliffe describes a similar impact on open-access online collections, comparing bots to swarms which overload servers:

> We see large numbers of bots visiting collections in a short period of time. A bot will land on a page, download everything it can, and then follow every link on the page in search of new data. Unlike people who tend to focus on predictable parts of a collection, the bots just want everything. (…) The swarms tend to overload servers. That results in slower response times for normal users. Eventually, the volume will knock the server offline entirely. (Hinchliffe, 2025)

The impact of infrastructure strain on small organisations

The literature consistently reports that AI bot traffic falls disproportionately on small organisations. These costs include direct expenses (e.g. hosting, bandwidth, and security) as well as a degraded user experience. As noted by I and Cath (2025), AI bots slow down and add cost to websites from organisations with already limited infrastructure , like independent news organisations or activist web pages.

These smaller organisations also have reduced capacity to cope. This means that smaller organisations are not able to effectively monitor, respond or mitigate these impacts even if they are fully aware of the issue (C. Radsch, 2025). For public interest media specifically, infrastructure costs compound the content extraction documented in the previous section, creating an additional burden that threatens organisational sustainability.

Ultimately, the impacts on infrastructure may also have a chilling effect on the future creators of open content, making them reluctant to create new open access content, as argued by Hinchliffe, who also adds: ‘That would be a bad outcome for everyone involved, including the entities deploying the bots in the first place (…) We need to find a way that everyone can operate sustainably’ (Hinchliffe, 2025).

4.3 Declining traffic and audience displacement

Overall, the literature documents significant declines from referral traffic from search engines and digital platforms to news publishers and content creators, driven by the growing integration of Generative AI in web search. There is broad consensus in the literature that this traffic decline represents a direct threat to the economic sustainability of public interest media and journalism in general, as it undermines the advertising revenue and subscription conversion pathways that have historically supported digital publishing.

AI Overviews and the transformation of search

Radsch (2025) explains the mechanism leading to declining traffic succinctly: 'traditional search engines—particularly the market-dominant Google—still drive traffic, but increasingly, initial answers come from AI models that summarize content without redirecting traffic.' This shift toward providing answers directly within search queries, rather than directing users to source websites, is now commonly known as 'zero-click search', a trend that as argued by Radsch elsewhere has been noticeable since 2019, but which AI is increasingly exacerbating (2024). This is supported by the findings of Zhao & Berman (2026) who find a consistent and moderate decline in traffic to news publishers after August 2024.

Google's AI Overviews (present in over 200 countries and more than 40 languages) represents the most recent manifestation of the zero-click phenomenon due to Google's dominant market position in search. While still difficult to find consistent statistics and figures, different sources report that the launch of AI Overviews in May 2024 (in the US) and its subsequent rollout in other countries (March 2025 for Europe), alongside with other AI assistants, is deepening the trend of driving users away from original sources of information (BBC & EBU, 2025; Brown, Peter D. & Jaźwińska, Klaudia, 2025; Chapekis & Lieb, 2025; Macpherson, 2025a; Pattison et al., 2025). In February 2026, the European Publishers Council filed a formal antitrust complaint against Google over AI Overviews and AI Mode. As Chairman of the European Publishers Council Christian Van Thillo put it:

> This complaint is not about resisting innovation or artificial intelligence. It is about stopping a dominant gatekeeper from using its market power to take publishers' content without consent, without fair compensation, and without giving publishers any realistic way to protect their journalism. AI Overviews and AI Mode fundamentally undermine the economic compact that has sustained the open web. (European Publishers Council, 2026)

Evidence of the effects on publisher traffic is mounting. DMG Media, owner of MailOnline and Metro, reported to the UK Competition and Markets Authority that 'AI Overviews resulted in a fall in click-through-rates by as much as 89%' (reported in Bearne, 2025). David Higgerson, chief digital publisher at Reach, is also reported articulating the core problem:

> Publishers provide the accurate, timely, trustworthy content that basically fuels Google, and in return we get a click... that hopefully we can monetise to our subscription service. Now with Google Overviews it's reducing the need for somebody to click through to us in the first place, but for no financial benefit for the publisher. It's another example of the distributor of information not being the creator of information but taking all the financial reward for it. (Bearne, 2025)

Pew Research Center data confirms these patterns at scale. In the US, users who encounter an AI summary in a Google search result are less likely to click on a traditional search link: users who encountered an AI summary clicked on a traditional search result link in 8% of all visits, whereas those who did not encounter it clicked on a search result nearly twice as often (15% of visits). In addition, users who encountered an AI summary in their Google search results clicked on a link in the summary itself very rarely – in just 1% of cases (Chapekis & Lieb, 2025). This is notably different from self-reported click-through rates from AI summaries in search answers, with 33% of respondents in a global survey reporting they always clicking through to links in AI overviews, and 37% reporting that they sometimes do; suggesting that perceived click-through rates or intentions to do so may be higher than what is actually observed (Simon et al., 2025).

While these statistics are not specific to searches for news topics, they highlight a broader phenomenon of reduced click-through rates to web pages where AI summaries are present. This is of particular

significance given that, as of March 2025, approximately one-in-five Google searches produced an AI summary (Chapekis & Lieb, 2025).

More recent data from the Reuters Institute at Oxford University show that the expectations of publishers about traffic decline are also consistent, even if the impact may be quantified differently: 'Publishers expect traffic from search engines to decline by more than 40% over the next three years – not quite 'Google Zero' but a substantial impact none the less' (Newman, 2026). The authors add that the proportion of 'zero-click searches' increases significantly with the appearance of Google's AI Overviews, and that the addition of a separate 'AI mode' tab with multimodal and multilingual capabilities adds more reasons for the user to switch to these rather than clicking through to original sources. The extent to which these capabilities will consolidate is still unclear: 'Expectations vary, with a few optimists believing the negative impact of these changes will be less than 20%. At the other end of the scale, around a fifth of respondents expect a loss of more than 75% of their company's search traffic' (Newman, 2026).

The launch of web browsers blending browsing with contextual, personalised AI-generated summaries – such as ChatGPT Atlas and Perplexity Comet – will most likely further exacerbate these effects with larger user bases. The effects may be even more pronounced for smaller and regional publishers, though less well documented due to limited research capacity. While studying platform aggregators, rather than AI specifically, Meyer et al. (2024) find that 'smaller and narrower outlets are indeed the ones who are harmed most by aggregators'. According to the authors this reflects a perception in the industry that small outlets are the most impacted by ongoing digital disruption.

Qualitative effects of traffic decline

Beyond aggregate traffic decline, the examined literature also expresses concerns about qualitative effects on audiences. One of them is potential longer-term changes in information-seeking behaviour and the quality of information that is accessible, with the implications this has for media plurality, the quality of knowledge that the public has and their ability to critically interpret the world. As Pattison et al. argue, AI-generated summaries do not allow for '[s]ystem 2 thinking—deliberative, effortful reasoning that fortifies epistemic resilience—' and instead create 'a frictionless sequence of query, receipt, and tacit acceptance'. For the

authors, this pattern can result in a lower capacity to critically evaluate the integrity of information (Pattison et al., 2025).

We develop these impacts on democracy further below in Section 4.5 and move next to discuss the impacts on the business models and sustainability of media.

4.4 Sustainability of business models

The sustainability of public interest media business models constitutes perhaps the most fundamental challenge facing the sector in the age of Generative AI. In general, the economic foundations of news media have been deteriorating for years (Brogi & Sjøvaag, 2024; Carlini et al., 2026), but recent evidence suggests the crisis has deepened considerably.

The debate now extends to strategies for addressing the challenges posed by Generative AI. In many ways, these challenges can be seen as the latest embodiment of an ongoing struggle for adaptation and durability in a digital media environment characterised by monopolistic platforms that draw on news media content as part of a strategy to keep users engaged. Yet, the Generative AI era is marked by distinct changes and uncharted developments for the sustainability of news journalism. Unlike their Big Tech predecessors, AI companies currently derive most of their revenue from the sale of services rather than advertising, though this is widely expected to change (Carlini et al., 2026), and tech companies themselves are racing to change their products in a bid to compete in an increasingly AI-dominated field, which has 'knock-on impacts for an increasingly delicate news ecosystem' (Dobrescu & Durach, 2025).

Media organisations are therefore at a turning point: the unprecedented challenges posed by Generative AI necessitate the reimagining of new and innovative business models.

> We are at a critical juncture: just as publishers had begun to establish successful strategies toward broad dissemination of their content in a digital space dominated by tech giants like Meta and Google, the game has taken yet another turn. (Exner, 2025)

AI-mediated search and AI chatbots compound these financial pressures, to the point that some analysts consider the present challenges as an existential crisis for journalism, with some media

representatives stating that '[J]ournalism is falling in relevance; it lacks sustainable models and no longer has the same agenda-setting role' (Philea, 2026).

The consequences are dire for the profession. According to media representatives quoted in Brown, Peter D. & Jaźwińska, Klaudia (2025): 'There is a hope for journalism, but it's also a sad hope: that we're just low-paid fact-checkers for Big Tech.'

A structural reconfiguration of the sector

The literature also points towards the all-encompassing changes affecting the whole supply chain, which results in a reconfiguration of the entire ownership model, potentially becoming divided between production, distribution and trust towards modular business models (Eeman, 2026). The need for more holistic approaches entails that business models alone will struggle to face the magnitude of the challenges without appropriate regulation.

> The impact of AI on the news, thus, occurs at the infrastructural level, from the material structures that distribute the news, through the protocols that manage data, to the platform level where content is ultimately consumed (Hesmondhalgh et al. 2023). AI, thus, not only changes the economics of news – but also the political economy of news as an industry. (Sjøvaag, 2024)

In 2025, media outlets in 160 out of the 180 countries assessed (89%) by the economic indicator of RSF's World Press Freedom Index achieved financial stability 'with difficulty', or 'not at all'. This is the lowest level ever recorded according to their survey (Scott, 2025).

Disintermediation without compensation

AI chatbots cause a 'disintermediation' process, delivering content that satisfies informational needs without driving traffic, with immediate impact on monetisation:

> ChatGPT, ClaudeAI, and Gemini are becoming, for many, the first port of call in the search for information, yet they take no responsibility for content, and they compensate no-one for the work done to gather and collate it. How does ChatGPT "know" what happened in Lebanon yesterday? That question can only be answered by correspondents

> from international broadcasters and news agencies. Yet, who will pay for the (expensive) work they do on the ground if the results are free via chatbot? (Exner, 2025)

A report commissioned by the European Parliament has identified the absence of compensation mechanisms as a critical gap. Articles 3 and 4 of the Directive on Copyright in the Digital Single Market (CDSM Directive) allow some uses of protected content without compensation and this is used by AI products to extract content for free.

> Authors' guilds and collecting societies have called for the introduction of new compensation mechanisms, such as equitable remuneration schemes, statutory levies, or collectively managed licences. The lack of such mechanism's [sic] risks replicating the systemic imbalances observed in other digital markets—where creators provide the raw materials, but intermediaries capture the economic value. (Lucchi, 2025)

We cover the recent literature on remuneration, transparency and regulation in Sections 4.6 and 5.3 of this report.

The vulnerability of small media in asymmetric ecosystem

As noted earlier, smaller publishers bear the consequences of disruption more deeply. In a global survey led by the JournalismAI Project at the London School of Economics, some respondents shared this concern:

> There are concerns that AI will exacerbate sustainability challenges facing less resourced newsrooms which are still finding their feet, in a highly digitised world and an increasingly AI-powered industry. (Beckett & Yaseen, 2023)

The 2024 Media Pluralism Report notes that for local media 'innovative practices and models are still not enough to counterbalance the crisis of the traditional models' (Bleyer-Simon et al., 2024). For example, media representatives quoted in Brown and Jaźwińska (2025) point at the consequences of being left out of the bargaining table when it comes to licensing agreements:

> Everyone can suddenly see there's a huge amount of money on the table. Journalism's in a really bad way. So you just take the money now and don't ask questions. But really, long term, you need

> something to protect the reputation of the publication and also the value of that content. [quoted in Brown and Jaźwińska (2025)].

> When the final deal is signed, the lion's share of that money is going to have gone to the incumbents and it's not going to go to small digital local publishers. [quoted in Brown and Jaźwińska (2025)].

Overall, a majority of voices in the literature consider the vulnerability of local media as a systemic issue, which persists despite attempts at innovation and adaptation. The consequences of such a vulnerable position vis-a-vis Generative AI providers – and large media conglomerates that have been able to secure licensing agreements – extend beyond individual organisations to threaten media pluralism, democratic accountability, and the availability of high-quality, locally relevant information.

Beyond the logistics of securing funding and diversified revenues streams, these issues can be seen as a manifestation of deeper issues of power and influence that dominant platforms can exert over journalism (Colussi et al., 2025). This is significant, given that healthy democracies and information integrity depend on well-funded and secure independent media: 'We face a growing worry about the future of local news in our country, which is very important to democratic health at the local level' (Philea, 2026). Through an analysis of public media systems with differing levels of independence and diverse funding sources, Neff and Pickard (2024) find that 'high levels of secure funding for public media systems and strong structural protections for the political and economic independence of those systems are consistently and positively correlated with healthy democracies'.

The economic implications, then, go hand-in-hand with epistemic impacts and, fundamentally, with threats to information and democratic integrity, which are covered in the next section.

4.5 Threats to information integrity and democracy

> We have to start asking ourselves how we can preserve journalism as a force that opposes authoritarian movements, even if the business model completely disappears. (Meredith Whittaker quoted in Exner, 2025)

A central theme in the literature and underpinning the overall diagnosis of the current context is the impact of Generative AI on the integrity of the information ecosystem as an epistemic pillar of democracy. This is an issue that relates to: (i) the technical features and affordances of Generative AI as a threat to epistemic values and (ii) structural aspects of how power is concentrated among a small number of dominant companies behind leading AI models. **Structural aspects also include the narratives that underpin these power inequities, such as hype and promises of efficiency, and the limited bargaining power of media, high quality information providers and civic institutions more broadly.**

Such power concentration also has implications for what potential solutions are within reach of whom. This trend is an exacerbation of the challenges related to the 'platformisation of the internet' that we have seen in the last two decades (Iosifidis, 2025).

These levels of power and concentration are such that AI and Tech companies are considered to be operating with the scale and influence of some sovereign actors (Khanal et al., 2025), but without the same democratic accountability.

> Right now, a handful of companies control much of the nascent AI industry, as well as the technologies upon and through which AI is developed and commercialized. Moreover, these companies are attempting to direct the trajectory of the industry in a way that primarily benefits their private interests, while also entrenching and further expanding their market power. This potentially globally disruptive technology is too important to be left wholly at the whims of today's tech monopolists, free of democratic accountability and control. (von Thun, M., & Hanley, D., 2024)

As stated by Mansell et al., the current threats to information integrity relate to both the reduction and diversity in news content as well as the sustainability of news media as their content is scraped, altered and concentrated by a few large Generative AI models that are at the same time driving audiences away: 'In some countries news media concentration is reducing the diversity and plurality of news content; in others, financial instability is leading to news deserts' (2025).

Generative AI's role in the media sector is therefore deepening the need to align technological innovation with ethical and public interest journalism.

Error and opacity as currently inherent to Generative AI

A key argument in the literature is that the affordances of Generative AI are in themselves contrary to epistemic welfare or journalistic integrity.

Generative AI is based on probabilistic models – data is used to predict which text or content will follow another in response to prompts by the user. Errors – also referred to as hallucinations – are therefore an intrinsic part of these models, rather than exceptions. In addition to generating deepfakes such as video or audio clips that appear to be real, Generative AI assistants can make up news stories and their sources (Dutkiewicz et al., 2025; Hoffmann, 2025; Schiffrin et al., 2025). These errors can occur even when the answer provided by an AI assistant mentions the source of the content, which has implications for trust in journalism. Independent audits provide an insight into the extent of these errors. A study by NewsGuard from 2025 found that 28% of responses from 11 leading AI assistants contained false information when prompted with known misinformation claims (Steinacker-Olsztyn et al., 2025) and a study by the European Broadcasting Union (EBU) and the BBC that same year found that '45% of responses contained at least one significant issue of any type'. Source errors are particularly significant (31%) and concerning as stated in the study by the BBC and EBU: 'Gemini had a particularly high error rate for sourcing in the latest multi-market study: 72% of its responses had a significant sourcing issue. All other assistants were below 25%' (BBC & EBU, 2025).

As stated in this study, **sourcing errors represent a reputational risk for news outlets who are cited alongside misinformation, even when the error is made by the Generative AI assistant alone.** The same study concludes that 'AI assistants are still not a reliable way to access

and consume news' (BBC & EBU, 2025). This research also finds that, in addition to sourcing, AI assistants struggle with developing news stories: 'They often combine information from articles published at different times, creating potentially misleading narratives for readers. This chronological confusion is especially problematic when reporting on evolving situations' (2025). In addition, the confidence with which these assistants present the information and answer questions provide a false sense of quality and make it difficult for citizens to question them or spot errors.

As noted in Section 4.3, the way Generative AI creates and presents content has also been found to contribute to shallow knowledge. A study by Melumad and Yun (2025) argues that this shallower knowledge 'accrues from an inherent feature of LLMs—the presentation of results as summaries of vast arrays of information rather than individual search links—', which the authors argue inhibits users from actively discovering and synthesising information sources themselves.

While there are attempts to develop attribution-based technologies and smaller models that can better control what data AI assistants draw on, opacity is an inherent part of how Generative AI works. Model opacity entails a lack of control over how information will appear and lack of options to correct mistakes or seek redress and accountability (Pattison et al., 2025):

> That opacity blocks the traditional checks on informational power—click data, outside audits, public scrutiny—and turns the search provider into an unaccountable (and unelected) gatekeeper of knowledge. In effect, a search engine moves from a contestable relevance-ranking system to an "answer authority," widening the gap between those who set the algorithmic dial and everyone subject to its judgments. (Pattison et al., 2025)

If model opacity makes it difficult to steward the integrity of what and how content appears on AI assistants and summaries, institutional opacity enables this bypass effect and further erodes accountability: 'In short, model opacity obscures the internal logic behind each answer, while institutional opacity withholds the data and documentation needed to probe that logic from the outside' (Pattison et al., 2025).

It has therefore been argued that **the affordances of Generative AI technologies are contrary to the mechanisms that make vital democratic institutions like public service media work, including**

transparency, accountability, cooperation or interpersonal relationships. As Hartzog and Silbey state: 'Unfortunately, the affordances of AI systems extinguish these institutional features at every turn. In this essay, we make one simple point: AI systems are built to function in ways that degrade and are likely to destroy our crucial civic institutions' (2025).

The 'bypass effect'

Related to this institutional opacity, the asymmetric positioning of AI companies in comparison to media organisations in an inadequately regulated context means that dominant Generative AI corporations have the power to bypass the institutions, norms and gatekeepers 'ranging from local media elites, e.g., the local newspaper, to public intellectuals', which 'played a crucial role in shaping public discourse, setting standards for acceptable speech, and managing the flow of information' (Abiri, 2024). As Abiri argues, this leaves the networked public sphere without 'trusted and trustworthy' intermediaries who can signal and steward the integrity of the knowledge available to society. Generative AI bypasses the 'arbitrors' (Thomson et al., 2024) and this can ultimately turn societies into 'a rhetorical battlefield marked by tribalism and comfortable beliefs, undermining foundational free speech values' (Abiri, 2024). The gatekeeping or bypass effect refers not only to what is known but also how information is framed, therefore affecting epistemic foundations or how we see and interpret the world (Böyük, 2025; Khanal et al., 2025).

Recommender algorithms and engagement metrics are another way in which social media and AI platforms bypass editorial judgement, individual and collective agency, and civic gatekeepers. Recommender algorithms present content tailored to individuals based on predictions, the weight of which is decided by programmers or decision-makers driven by profit rather than journalists following public interest principles (Pérez-Seijo & Nuno Vicente, 2024). As Voinea states: 'The decision to prioritize one metric over another is analogous to an editor's decision to prioritize one news value over another. Thus, gatekeeping power shifts upstream to those who design and configure AI systems' (2025). Algorithmic decisions might be enjoyable or convenient but 'societally irrelevant' or might worsen the effect of echo-chambers showing content that confirms a user's beliefs rather than challenging them by showing different perspectives.

This threatens the foundations of public service media. Algorithmic predictions 'can result in audiences being approached merely as consumers rather than citizens', and they 'often fail to serve smaller, for example, minority language, audiences' (Van den Bulck et al., 2025). AI assistants, like social media platforms, are currently not led by public interest values but by private profit-based incentives (Dutkiewicz et al., 2025).

Analysis on the effectiveness of European regulation like the DMA and the DSA argues that this regulation fails to protect content because it does not tackle the profit-based incentives of big platforms: 'current digital infrastructures perform the "marketplace of ideas" rather successfully—albeit the ideas that are "sold" tend to be those that are most profitable or ideologically desirable for social media companies' (Oleart & Rone, 2025a). It has been argued that AI makes these incentives more prominent: 'AI puts these perverse incentives on steroids. While AI has the potential to lower information processing and dissemination costs, it undermines the incentives for private producers to acquire and process accurate, timely, and reliable content' (Stiglitz & Ventura-Bolet, 2025). In other words, 'AI reduces the cost of producing untruthful information relative to the cost of producing truthful information, effectively making it very cheap to produce lies' (Stiglitz & Ventura-Bolet, 2025).

If Generative AI oligopolies and systems are bypassing public interest principle and processes, media corporations risk leaning into this process, pushed by competitive pressures and the need to lower costs in a difficult economic environment. Yet, as pointed in the literature, irresponsible incorporation of AI systems into content and editorial process can further undermine journalism and, with it, the information ecosystem and democracy: 'As important as it will be for journalism to adapt to and integrate AI, newsrooms that replace journalists will hasten its demise, with profound ramifications for democracy' (Radsch, 2024).

Concentration undermines equity, pluralism and diversity

The concentration of big Generative AI models in the hands of a few players, coupled with the affordances of Generative AI, also impacts the quality and diversity of content as well as principles of plurality and universality, ultimately impacting democracy (Balasingham et al., 2025). A study by Brantner et al., which analysed Microsoft Copilot's sourcing behaviors across languages, found 'a tendency toward homogenization of news sources, reliance on UK and US-based media,

and inconsistencies in source transparency' (2025). The authors warned that platforms borrow 'authority and legitimacy from credible sources to create a false sense of trust, potentially undermining the democratic process' (Brantner et al., 2025). Other research has shown that AI assistants disproportionately concentrate citations from a small number of news outlets (Yang, 2025), which highlights their impact on content nuance and diversity, and their role as gatekeepers both in terms of location as well as language of sources.

If the use of Generative AI tools for news access continues to grow among audiences, this may further contribute to what Sevignani et al. term 'media environment capture', in which 'Big Tech influences entire information environments, consequently shaping our lifeworld while privatizing the public sphere' (Sevignani et al., 2025). This concern is reflected in the results of the most recent Media Pluralism Monitor report, which showed market plurality indicators as the only area with an average 'high-risk' level across EU countries. The plurality of digital markets indicator, which registers risks related to excessive market power – including for AI companies as digital intermediaries of information – was also at a very-high risk level (European University Institute., 2025).

> We are in a very grim period when it comes to trust and integrity in the media and social media sphere. A powerful nexus of private companies and authoritarian politicians garbed as 'free speech' advocates are profiting from hate. (Banaji, 2025)

There is also a growing concern that this market concentration, combined with an inadequate legal framework, is resulting in the development of business models in which public interest data is commodified to protect it from unconsented scraping. While this means that some organisations can receive financial compensation from sharing their content with Generative AI models, it is also generating a two-tiered system (Barnett, 2025; von Thun & Hanley, 2024). As Barnett states: 'the market risks bifurcating into uncurated low-quality AI-distributed content for nonpaying users and curated higher-quality AI-distributed content for businesses, higher-income individuals, and institutional users (2025)'. **While this is not a new concern, Generative AI is amplifying both the need for income for media outlets as well as the value of high quality information.**

Under this extractive relationship, some scholars like Stiglitz and Ventura-Bolet warn that this may lead to 'informational collapse'. In their theoretical model, current errors inherent in Generative AI models (hallucinations and inaccuracies) slow down the substitution of primary sources by AI models. However, if reliability and accuracy improve, the need to check original sources diminishes, eventually leading to informational collapse: 'No one invests in producing accurate information when their work is instantly absorbed and intermediated by an AI that captures all downstream attention and value' (Stiglitz & Ventura-Bolet, 2025).

Extractive dynamics and the risk of informational collapse may also have a deeper impact on small and medium-sized organisations, which will be least equipped to implement protective measures or negotiate compensation, as we develop in this report.

4.6 Inadequacy of existing legal frameworks

The rise of Generative AI has also exposed fundamental inadequacies in existing legal frameworks designed to protect creative works and regulate digital markets. As AI bots increasingly scrape, process, and regurgitate journalistic content, both scholars and policymakers have identified what they consider to be significant gaps between current regulatory instruments and the new challenges of Generative AI for the media industry.

Copyright and the Text and Data Mining (TDM) exceptions

The literature examined scrutinises current inadequacies in the European framework regulating copyright. One key question is whether existing exceptions, particularly those for text and data mining (TDM), adequately address the training of Generative AI models. The debate revolves around the two 'regulatory interfaces' (Schirru & Margoni, 2023) of TDM exceptions in the Copyright in the Digital Single Market Directive (CSMD): Art. 3 (allowing research organisations and cultural heritage institutions to mine content for scientific research) and Art. 4 (allowing TDM when the 'use of works and other subject matter […] has not been expressly reserved [a.k.a. 'opt-out'] by their rightsholder in an appropriate manner' (art. 4(3)) (European Commission, 2019).

The EU copyright framework as it stands has been criticised for not being designed to address the risks posed by AI, and therefore presenting 'opportunistic and misleading interpretations of the text and data mining (TDM) exception' (Carlini et al., 2026). Alonso and Lucchi (2025) expose a mismatch between the current 'regulatory interface' and Generative AI model training. They argue that, while Article 4 of the CDSM Directive currently serves as the main legal basis for commercial AI training activities, these do not fall under the TDM exceptions of Articles 3 and 4 of the CDSM Directive because Generative AI training resembles 'a process of creative reproduction' rather than 'a form of analytical extraction'. Alonso and Lucchi add that '[t]he growing scholarly consensus now recognises that we need new legal tools and categories to address this unprecedented challenge, not a forced reinterpretation of provisions designed for another era' (Alonso & Lucchi, 2025).

Lucchi (2025) doubles down on the same argument regarding the absence of a legal foundation for training AI systems:

> Training generative models involves the large-scale reproduction and internalisation of expressive content—not merely the extraction of factual information—and thus likely exceeds the definitional scope and normative intent of the TDM exception under Article 2(2) of the Directive. (Lucchi, 2025)

The inadequacy of opt-out mechanisms adds to these concerns. Dornis and Lucchi (2025) note that 'such conflict cannot be cured by opt-out mechanisms, especially given their limited effectiveness and incompatibility with international copyright standards prohibiting formalities.' This observation points to a deeper structural problem: **the CDSM Directive's reliance on rightsholders actively reserving their rights presumes a level of technical capacity and awareness that many publishers, particularly smaller and medium-sized newsrooms, may lack.**

Moreover, the interaction between the AI Act and existing copyright law has introduced additional layers of uncertainty in this domain. Lucchi (2025) identifies that while 'Article 53 and Recital 105 reflect a policy assumption that AI developers will rely on existing copyright exceptions', these provisions 'do not settle the legal question of whether those exceptions, as currently drafted, are applicable to Generative AI training'. This regulatory gap means that 'the AI Act presupposes legal clarity rather than establishing it—placing responsibility back on EU copyright

law to determine whether and to what extent Generative AI training can be lawfully conducted without express authorisation or remuneration' (Lucchi, 2025). For a different argument, Quintais (2025a), considers that 'the extent to which copyright exceptions apply to AI-generated outputs' is still a 'notable and evolving issue', although he also considers that, within EU law 'it is now mostly settled that this definition [TDM exception in the copyright directive] covers lato senso pre-training and training activities needed to develop an AI model, including generative AI models' (Quintais, 2025b). Quintais also adds another layer of complexity, which is the intersection of copyright law with private contractual agreements, such as the terms and conditions of AI providers, which also affect issues of authorship and enforcement. Ultimately, Quintais (2025a) also forecasts that the current regulatory framework will fall short of addressing the economic issues for content creators because of this complexity, including the interplay between the AI act and copyright law in Europe, as well as issues of extraterritoriality that remain unresolved.

At the time of writing, the European Parliament has adopted the Report on Copyright and Generative AI Intelligence, where it reaffirms that EU copyright rules apply to any AI system operating in the EU market. However, clarification on how current copyright rules apply is not yet clear. As Keller (2026) notes:

> The most significant change between the June draft and the text the Parliament adopted is a welcome, if hedged, acknowledgement that AI training is covered by the TDM exceptions under the CDSM Directive. Where the draft spent considerable energy arguing that Article 4 did not apply to Generative AI training—only to propose a replacement mechanism with essentially the same structure—the adopted text accepts the applicability of the existing framework while calling for "swift clarification on its application and implementation." This is not a ringing endorsement of the framework, but it is a meaningful retreat from the position that the existing law is simply inadequate. (Keller, 2026)

In parallel, a Resolution from the Assembly of the Council of Europe has asked member States to 'clarify in their national legislation that copyright exceptions such as the text and data mining exceptions introduced by the European Union's Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market are not applicable to the training of AI systems' (Council of Europe, 2026).

Balasingham et al. (2025) draw from empirical insights from the application of Art. 15 CDSM – establishing the 'press publisher's right' to demand remuneration for reuse of their content – to highlight the shortcomings of copyright law:

> However, due to the lack of clarifications [in Art. 15], particularly regarding the level of remuneration, Google, taking advantage of its substantial bargaining power, managed to circumvent the press publishers' right. In response to this circumvention of the law and exploitation of press publishers in France, the French competition authority found that Google abused its dominant position. This case showed that copyright policy by itself is inadequate to protect press publishers. (Balasingham et al., 2025)

In their analysis of three different EU markets (French-speaking Belgium, France, and Spain), Lefèvre et al. (2025) find that there is fragmentation in regulatory frameworks and a lack of clear guidelines that are sector-specific. This 'leads to inconsistencies in AI implementation and fails to adequately address AI's ethical and operational challenges in journalism'. The regulatory questions, in sum, are far from being settled.

The weakening of copyright as a global issue

The inadequacies of the regulatory framework of copyright in the wake of Generative AI are not limited to EU law.

Carlini et al. (2026) argue that the concept of ´fair use' in the United States is 'far too broad' and that 'AI and ability to copy, disseminate and profit from other people's intellectual property at scale has made a mockery of traditional justifications of fair use'. In addition, as argued by Stiglitz and Ventura-Bolet (2025), constructs like 'fair use' do not address damage to the quality of information.

In their comparative analysis of the US, China, and the EU copyright regimes, Kuai (2024) highlights three different policy attitudes – silence in the US, premature laws in the EU, instrumentalisation of laws and journalism in China – to conclude that, despite these differences, 'current regulatory frameworks in all cases have led to a weakening of the institution of copyright, which, in turn, has contributed to the deinstitutionalisation of journalism and the institutionalisation of algorithms'.

Broader regulatory issues

Beyond copyright-specific concerns, scholars have identified broader regulatory arenas that are currently being contested in courts. Jones et al. (2025) have summarised the main arguments in major lawsuits as follows: (i) copyright infringement (scraping of content without permission); (ii) breach of terms of use (process of scraping violates the terms of use of websites and bypasses restrictions); (iii) trademark infringement (Generative AI tools present scraped data as their own diluting the credibility and value of media brands); (iv) lack of attribution (failure to attribute or incorrect attribution); (v) harm to business models (reduction of traffic to news websites results in reduced ad or subscription revenue); and (vi) lack of compensation (AI companies profit from investment and resources by media outlets to produce reliable content).

In December 2025, the European Commission opened an antitrust investigation to assess if Google breached EU competition rules by deploying 'AI Overviews' and 'AI Mode' in a way that imposes 'unfair terms and conditions on publishers and content creators' (European Commission 2025). This follows other actions in Europe led by the Independent Publishers Alliance (IPA) and the Movement for an Open Web, as reported in Diab (2025), claiming that Google is 'abusing its monopoly by forcing participation as a condition of visibility, leaving the publishers no meaningful way to opt out'.

A few voices, though, also express doubts about these concerns. They look skeptically at the moves of media providers as 'lobbying for protectionist AI legislation' (Jarvis, 2025), invoking 'a phantom right to traffic' (Diab, 2025) or, more broadly, calling out the alleged 'free-riding myth' [platforms free riding on 'publishers' investments in news production'] (Colangelo, 2026).

Ultimately, alternative views in the regulatory discussion reflect the critical juncture of the moment, where GenAI-mediated access to news has disrupted an already delicate equilibrium between platforms and media providers – in the absence of a new social contract and in the midst of intense regulatory battles with uncertain outcomes.

In summary, **the existing legal frameworks in Europe governing copyright, data mining, and platform regulation exhibit significant inadequacies when confronted with the challenges posed by Generative AI.** The TDM exception's conceptual mismatch with

generative training, together with the absence of compensation mechanisms, the limitations of technical enforcement tools, and the disparate impact on smaller media organisations collectively suggest the need for legal reform. As Dornis and Lucchi (2025) conclude, 'a new sui generis legal framework would be required to legitimize such uses: one that offers appropriate safeguards for rightsholders and aligns with international copyright obligations'. Until such comprehensive reforms are enacted, public interest media will continue to navigate an uncertain and largely unprotected legal landscape.

4.7 Attitudes toward AI and News

The development of social, technological, and industry norms and practices surrounding the use of journalistic data for AI-generated summaries is inevitably influenced by how these technologies are used and perceived, both by users and news organisations themselves.

Public Attitudes

Public perceptions of this phenomenon are likely to be mediated by personal use of Generative AI technologies. A 2025 Reuters study found that while AI chatbots and interfaces are emerging as a source of news for users, only a small proportion of people surveyed (7%) used AI as a news source each week, with this figure doubling for under-25s (15%) (Newman et al., 2025). This reflects a continuing, broader trend in diversification of news sources in the digital media landscape, where engagement with traditional media sources such as TV, print and websites continues to fall, while social media, video platforms and online aggregators grow. Young people, in particular, are more likely to use social media and AI chatbots to check information (Newman et al., 2025).

People still trust news created by humans more

When audiences are asked to evaluate how they feel about AI-driven or AI-generated news, a nuanced perspective appears. Simon et al. (2025) identify a 'comfort gap' in audiences: that is, people have significantly higher levels of trust in news created by humans as opposed to news made entirely by AI, or with some human oversight. Respondents are overall more comfortable with AI being used for back-end tasks, such as editing for grammar or translation, than for active tasks, such as

rewriting articles for different audiences – which is similar to the process that takes place in creating AI summaries. That these uses create a sense of discomfort among audiences is closely related to trust in news:

> On the ground, reporters describe a more immediate erosion: the chain of authorship snaps when a story is machine-drafted and published under a human name. Trust follows. Experiments at the University of Kansas show that when readers suspect AI involvement, their confidence in a story's credibility drops, even if they can't say how much AI was used. A related KU study found audiences rate human-written releases as more credible than AI ones. (Nogueroles, 2025)

Others have also found that 'audience's expectations overlap with the normative vision of PSM [Public Service Media], where independence and professionalism, understood as commitments to truth-telling, freedom from faction, and accountability, has a central role' (Campos-Rueda & Goyanes, 2023).

Yet, news avoidance is increasing (Mansell et al., 2025), and as discussed in Section 4.3 on audience displacement, users are less likely to click on a source search link when they encounter an AI summary, making audience displacement a real and concerning issue. As reported in the latest World Press Trends Outlook, 'publishers are preparing for "Google Zero," a world where traffic is no longer referred from Google Search to third-party sites' (WAN-IFRA, 2026).

Misinformation has also been identified as a concern when accessing news via non-traditional formats such as AI chatbots and social media (Newman et al., 2025). The same survey by Simon et al. (2025) shows that 'to some extent people think that Generative AI will primarily benefit publishers rather than them as users'. Specifically, 'people are more likely to think that Generative AI will make news cheaper to make (+39) and more up to date (+22), but also less transparent (-8) and, crucially, less trustworthy (-19)' (Simon et al., 2025).

At the same time, users perceive they may benefit from AI-driven technologies even as they remain skeptical of them. A 2023 study found that participants were 'ambivalent about living with algorithm-based news recommendations', and expressed both advantages as well as concerns with missing out on important information and challenging viewpoints by algorithmic recommender systems (Du, 2023).

The use of AI summaries and responses for news is likely to be closely aligned with overall perceptions and trust in AI technologies themselves. A survey from 2023 suggests that news organisations should target users accordingly, as users are more likely to adopt and engage with a technology if they have high levels of positivity and trust, and low levels of skepticism, towards that technology (Lim et al., 2023).

Towards finding spaces of belonging

Another theme in the literature suggests that, by contrast to passive consumption of encountered news or content, newer generations of audiences are also proactively consuming content in niche spaces related to their identity and sense of belonging (Eeman, 2026). This also relates to how trust with audiences is built, including through habit, emotion and identity (Ross Arguedas et al., 2024). As we cover in Section 5.4, this points towards business models that serve niche communities through more engaged and personalised community content.

Journalist Attitudes

AI Integration in Journalistic Workflows

AI technologies are not only influencing how audiences access and perceive news media, but also profoundly altering journalistic workflows and the ways in which news is produced and distributed. This encompasses both journalistic content in AI outputs and the way AI tools are used to produce journalistic output itself. Although the adoption rates and levels of maturity differ significantly among organisations, AI tools are increasingly being incorporated into journalistic practices (TEMS, 2025). Common applications can be found at every stage of the news cycle: from news gathering tasks, such as content aggregation and trend and topic analysis, to news production tasks, including automated writing, editing, translation, transcription, and summarisation, as well as news distribution and moderation tasks, which can involve fact checking, content personalisation, and user engagement analysis (Cools & Diakopoulos, 2024).

Yet, there is a tension between personalisation and the universality and democratic principles of public interest and public service media in particular. Others have pointed out a ‘clear gap in the literature for a more nuanced approach to understanding public views on personalisation based on its social and democratic implications’ and

have recommended the need for more public involvement in decisions and design of personalisation in media (Jay, 2025).

Journalistic Perspectives and Concerns: opportunity and risk

EU survey results show that 'concerns around GDPR, intellectual property and editorial integrity remain significant' among media organisations, which 'reflects both an appetite for AI and ongoing hesitation due to compliance and ethical concerns' (TEMS, 2025). Through an examination of AI guidelines in global media organisations, De-Lima-Santos et al. (2025) find a 'complex interplay of promise and risk in AI-generated content, emphasizing the vital role of human oversight, bias mitigation, and ongoing adaptation to ensure the integrity and accuracy of media content in the AI era'. As with the use of AI technologies in other industries, a significant tension lies in minimising the potential risks, while taking care to not be disadvantaged by failing to capitalise on the potential opportunities afforded by this technology. The interaction and adoption of AI with newsroom practices seems to be defined by this dual mandate: maximising opportunity while minimising risk.

A recent survey found that a majority of UK journalists who took part in the study are pessimistic about the threats AI poses to journalism, and concerned about the potential negative impact of AI on public trust and accuracy in journalism (Thurman et al., 2025). An earlier global survey of news organisations similarly found that respondents were fearful that AI technologies could further commercialise journalism, and worsen the decline in public trust (Beckett & Yaseen, 2023). A study from Romania similarly revealed a common vision among journalists on 'the critical role that the media must play in supporting social equity and cohesion in the AI era', with participants expressing concerns about the proliferation of is misinformation, loss of public trust, polarisation of discourse and need for regulation and ethical responsibility for AI integration in media (Stănescu, 2026).

Concerns about power concentration and opacity

News organisations have also voiced concerns regarding the profit-driven nature of AI technology providers, and their concentration of power and lack of transparency (Beckett, C., & Yaseen, M. 2023). Wasdahl & Srinivasan (2026) highlight this concern in their analysis of how the market incentives of technology companies diverge from those of journalism, and highlight that this difference in logic extends to how audiences are imagined and conceived of by media organisations and

tech companies – as 'a market', or as 'a community' – contrary to the understanding of public interest media as defined in Section 3 of this report.

Concerns on the future of journalism as a profession

These concerns are amplified by the perceived threat that AI poses to journalism and journalists in their professions. Thomson et al. (2024) describe the tension between views that management favour automation for its cost-cutting potential, while journalists worry that their work will be considered superfluous.

This is particularly relevant considering a common narrative on the promises of AI tools to reduce workloads and improve output quality. While a 2023 study found that news headlines produced by Chat-GPT were on average rated as higher quality by participants than headlines written manually , the study also highlighted that some GPT-generated headlines contained significant flaws, and received the lowest quality rankings in certain instances, concluding that 'human intervention remains necessary for editing and final decisions, particularly in high stakes tasks' (Ding et al., 2023). This mirrors other consensus that have been found among journalists that 'human judgement is required to make complex decisions in journalism and that journalistic values should be prioritised in AI tool design' (Gutierrez Lopez et al., 2023).

Interviews with journalists find that such tensions are closely intertwined with journalistic perceptions:

> Across participants, three findings stood out: journalists consistently restricted LLMs to routine, low-risk tasks, organizations grappled with emerging tensions between editorial values and platform-driven incentives, and many respondents described rising audience sensitivity to AI authorship. (Wasdahl & Srinivasan, 2026)

Concerns about redefining journalistic practices and norms touch on a broader process of change and adaptation in the journalistic profession with respect to AI and copyright specifically. A 2023 study conducted in the United States found that journalism institutions generally rejected the idea that copyrighted content available on the internet could be freely repurposed, not only because of the protection of financial investments but also to 'reinforce the boundaries of the journalism profession itself' (Boyles, 2024).

Beyond the diagnosis of extractive data practices, concerns for the sustainability of public interest media and the broader effects of AI-enabled news consumption on the integrity of the media ecosystem, it is equally important to note how **AI technologies are influencing not only audience perceptions and habits of news consumption, but also fundamentally impacting conceptions of what it means to be a journalist in this time, and what value journalistic practices have in our societies that may extend beyond simply publishing facts accurately and efficiently.** Such perceptions and reflections provide a basis for discussion of possible solutions, which we now turn to.

5. Strengthening media's leveraging power to thrive in the age of Generative AI: potential ways forward

Our analysis of the literature shows a significant number of solutions being proposed to rebalance the power asymmetries between media organisations and Generative AI companies. While these solutions overlap, and can be categorised in different ways, we have largely structured them in a micro-meso-macro perspective. Thus, at the micro level we review technical solutions for gatekeeping, attribution and preference signalling that individual organisations can adopt, provided that they have the infrastructure and capabilities to do so; at the meso level we review the growing ecosystem of licenses and the markets around them; at the macro level we provide an account of regulatory proposals in the areas of copyright law and competition policies. We then move on to proposed approaches to business models, funding and investment, and to more holistic avenues that involve collaborative ecosystems of governance and collective action.

Lastly, we cover the need for skills, literacy and, more broadly, social norms on public interest media and infrastructure, as well as covering literature focused on the application of AI systems in newsrooms.

As mentioned in the methodology, this report uses a diagnosis-solutions framework as a guiding structure; however, we acknowledge that this simplifies the complex entanglement between diagnoses and potential solutions. Some of the themes presented as solutions may play a part in the challenges that we cover under the previous diagnosis section. For example, some licensing deals can be a mechanism for legitimising rather than addressing scraping or summarisation by AI systems. We address the limitations of some of the potential ways forward presented in the literature as we cover them in the sub-sections below.

5.1 Technical solutions for gatekeeping content extraction

Technical solutions to manage the access of Generative AI tools to media content are a distinctive domain in the literature and cover protocols, infrastructure and standards through which access can be controlled, signalled, and potentially compensated. While multiple approaches exist, the literature is consistent in noting that technical tools can define and express preferences, but they often cannot enforce them, and they therefore remain largely voluntary.

Robots.txt and its limits

Robots.txt (the Robots Exclusion Protocol) is the main instrument currently available through which websites signal crawling preferences to bots. The existence of this machine-readable text has become particularly significant as an attempt to address the harvesting of huge amounts of data for model training without compensation or attribution: 'As Generative AI systems scale, the stakes have risen: what was once a quiet tool for managing web etiquette is on the precipice of becoming the frontline defense for those trying to assert control over how their content is used in the lucrative AI-driven arms race' (Hingle & Knodel, 2025).

Indeed, the literature suggests that adoption of robots.txt by media publishers has increased rapidly. Brown and Jaźwińska report that 'as of May 2025, 32 percent of the top 50 news websites in the United States were blocking OpenAI's search crawler, 40 percent were blocking its user agent crawler, and 50 percent were blocking the crawler that collects content used to train its Generative AI foundation models. Fifty-six percent were blocking Perplexity's crawler, 58 percent were blocking the crawler behind Google's Gemini, and an average of 60 percent were blocking Anthropic's crawlers' (Brown, Peter D. & Jaźwińska, Klaudia, 2025).

Reputable news websites tend to use robots.txt directives targeting AI crawlers more than websites that contain misinformation do, as found by Steinacker-Olsztyn et al. (2025): 'among websites with a robots.txt file, 60.0% of reputable sites include a DisallowAll directive for at least one AI agent, compared to just 9.1% of misinformation sites. On average, reputable news websites reference 15.5 distinct AI agents, while

misinformation websites reference only 0.77.' Similarly, 'over 50% disallow GPTBot, and 40–50% restrict other major AI crawlers such as CCBot, ClaudeBot, ChatGPT-User, and Google-Extended.'

Despite this uptake, especially by reputable sites, the literature is consistent in identifying robots.txt as a structurally insufficient tool. The core limitation is that it is a request, not a rule. Brown and Jaźwińska (2025) note that the protocol 'is seen as something of a blunt tool and easily circumvented' and cite 2025 Tow Center research finding 'that several major AI companies' crawlers appeared to be accessing content from publishers who had blocked their crawlers'.

There have been a range of proposals to extend and upgrade robots.txt beyond its current functionality, including expanding user agent identification to cover more types of crawlers; adding new properties specifically for AI training to allow more granular content-use control; and the development of a complementary ai.txt file, with versions proposed by both Guardian News & Media and Spawning (Hingle & Knodel, 2025). On the latter, Hingle and Knodel suggest that it would give publishers 'more control over how AI systems use their content, including the ability to limit snippet length, require attribution, and allow or deny use for AI training'. Even so, they also highlight that these tools focus on signalling rather than enforcement: 'Technical tools, standards, and protocols provide mechanisms to express content owners' wishes clearly and machine-readably, but they rely on voluntary compliance from crawlers and AI systems' (Hingle & Knodel, 2025).

There are also strategic costs to blocking that complicate straightforward adoption. Zhao and Berman (2026) find empirically that large publishers who decided to block LLM traffic had less total and human traffic, observing a 7% decline in weekly visits after blocking. Brantner et al. (2025) document a similar situation in the Swedish media ecosystem, where major news publishers opted out of crawling with the result that Microsoft Copilot had barely any content in Swedish and relied mainly on foreign, large and English-language news media. The only Swedish professional news media that Copilot linked to was the public service media (but did so rarely). It also heavily relied on the Swedish Wikipedia.

The visibility costs of opting out add to the already well-documented dilemma mentioned earlier that blocking may also increase the relative weight of misinformation in Generative AI training data: as reputable websites tend to opt-out, whereas misinformation sources tend to

remain open, the overall information ecosystem degrades as 'the same misleading or fabricated information may reappear in new AI-generated content, amplifying its reach and reinforcing its perceived legitimacy over time' (Steinacker-Olsztyn et al., 2025). Radsch (2024) observes that the technical approach to blocking content may represent a short-term solution for individual media organisations, but may also have wider ramifications for the media and knowledge ecosystem, including exacerbating mis- and disinformation, errors ('hallucinations') and impact downstream applications drawing on this content (Radsch, 2024).

Pay-to-crawl systems

Pay-to-crawl systems have emerged as a commercially viable, technical solution to the crawling problem by imposing a financial compensation on Generative AI bot access. Hardinges (2025) defines them as 'emerging technical systems used by websites to automate compensation for when their digital content—such as text, images, and structured data—is accessed by machines.' Although they vary 'in terms of their openness, standardization, and interoperability, as well as the permissiveness of their access controls and payment terms', they tend to typically integrate a combination of the following elements, typical of digital rights management, paywalls or approaches to micropayment:

- Authentication (with cryptographic authentication protocols to verify identities)
- Access control (with granular rules dictating who is permitted, restricted or charged)
- Pricing, contracting and paying (with flexible compensation models and/or standardised licenses and payment mechanisms)
- Content delivery (enabling access to contents in different formats)
- Metering and logging (ensuring that all interactions are registered for auditing and billing purposes)

Hardinges (2025) points to examples such as Pay Per Crawl by Cloudflare, AI RevShare by Valyu, GistAttribution by ProRata, Open Licensing Protocol by RSL and TollBit. Given that these examples are now

part of an emerging attribution and licensing market, we revisit them in Section 5.2.

Nevertheless, it is relevant to highlight here that there is a Creative Commons approach in the literature sharing some concerns with regard to indiscriminate use of pay-to-crawl systems, despite their potential contribution to the sustainability of news content creators, because they restrict the availability of content that should be shared for public interest and legitimate uses, as well as with institutions like non-profits or educators: 'Wide adoption of pay-to-crawl could ultimately represent a shift away from the spirit of the open web towards a more tightly controlled and monetized content ecosystem' (Hardinges, 2025). Eayrs (2025) extends this analysis further: 'pay-to-crawl systems themselves could become new concentrations of power, with the ability to dictate how we experience the web.'

In this regard, Creative Commons proposes principles for responsible deployment, including that pay-to-crawl 'should not become a default setting' deployed automatically by domain hosts or content delivery networks on behalf of all websites, and that 'pay-to-crawl systems should enable choice and nuance, not blanket rules' (Eayrs, 2025).

One of the most acute concerns is the power concentration risk. Radsch (2025) points to it directly when discussing Cloudflare's new system's capacity to technically enforce automatic blocking of AI bots by default. For Radsch, this turns Cloudflare into 'a kind of regulator-by-default, which raises questions about whether Cloudflare's new system will entrench its own market position or set de facto standards for the AI economy'.

The technical enforcement capacity of pay-to-crawl systems is what robots.txt lacks, but it may come at the cost of routing that enforcement through private intermediaries who gain structural power over the terms of the entire AI content economy:

> This is policy in the form of infrastructure: it defines who can enter, under what conditions, and at what cost. In a field where legislative progress has lagged, Cloudflare is stepping into the void—challenging both market and regulatory norms. (Radsch, 2025)

Hinchliffe (2025) notes the arms-race character of the current short-term solutions: 'many collections are turning to third-party anti-bot solutions that operate at the firewall level. Those seem to work

reasonably well, although they are stuck in an arms-race cycle with the bots themselves.'

A further problem is small publisher exclusion, as small and local newsrooms struggle to keep up with the changing landscape of bots and tools to manage or monetise crawling (Radsch, 2025). The technical and operational capacity required to implement and manage pay-to-crawl systems – including bot detection, rate limit configuration, legal contracting, and royalty collection – reproduces the same structural disadvantage that characterises other voluntary technical approaches: the publishers most in need of revenue from their content are least equipped to capture it through market mechanisms that require sophisticated technical infrastructure to deploy.

Attribution technologies

Alongside pay-to-crawl systems, attribution technologies have developed a distinctive niche at the intersection between media publishers and AI companies. A number of sources (Jones et al., 2025; Macpherson, 2025a; C. C. Radsch et al., 2025) highlight ProRata (more specifically, its product Gis) and MisoAI as companies developing attribution products. Industry standards have also been proposed as a self-regulatory mechanism. One example is the Global Media Identifier (GMI) proposed by the German non-profit Global Media Registry. GMI offers a standard to establish source identity 'so that the industry, regulators, audiences and other stakeholders can unambiguously identify the respective source of information, for example, in order to trace back and identify its ultimate beneficial owner' (Global Media Registry, 2026). GMI also includes metadata and interoperability mechanisms with other identifiers such as DOIs, which points to developing persistent links as another key challenge in attribution.

The challenges of attribution persistence

There are challenges to operationalising and maintaining technology-based attribution. As Adeyebi and Adeusi (2025) argue, one technical challenge relates to maintaining attribution 'through content transformation, establishing clear contribution boundaries in collaborative creation, and ensuring system interoperability across platforms'. As they note, the pace of change in Generative AI models requires constant recalibration and operational investment, which 'remains prohibitive for smaller creative organizations'.

There are also challenges in ensuring IP protection techniques simultaneously – e.g. limiting training usability, providing traceability, and protecting creative styles – because they aim for different goals (Šarčević et al., 2024). In general, the adoption of these techniques,'such as content fingerprinting, watermarking, and cryptographic methods', is slow (Amon et al., 2025).

Other options explored include geospatial attribution (Kanza et al., 2024), combining a royalty share method based on cooperative game theory with geospatial tagging to ensure compensation reaches data owners who 'lack means and political power', and exploring the use of location-aware digital rights management to restrict content use geographically. Kanza et al. argue that '[v]erifiable geospatial provenance of data, models, prompts, and content usage, e.g., based on geotagging [...] could help to address local data-ownership laws' (2024).

The lack of robust attribution infrastructure as a technical deficit may also lead to other power asymmetries, as without the ability to trace the use of their content in AI outputs, media organisations cannot establish the empirical basis for compensation claims or verify compliance with licensing terms. It is in this regard that Eeman (2026) situates the attribution challenge within the evolving structural position of publishers more broadly: 'without clear metrics for value, attribution, or substitution in AI-driven environments, negotiating power remains weak. Visibility, integrity, and leverage drift apart. Collective action, shared standards, or regulation is needed to rebalance the equation'.

As Stiglitz and Ventura-Bolet (2025) point out, this attribution failure marks a shift from previous relationships with platforms where 'at least, in the case of search engines and social media, there was typically attribution of the source of information, which the platforms claimed (with little evidence) drove more traffic to the legacy media'. **This regression from imperfect to no attribution is one of the distinctive marks of the Generative AI era relative to previous platform relationships with journalism.**

Preference signalling standards

Another technical response to crawling and attribution issues is the building of preference signalling infrastructure, which 'leverages the traditional norms and reciprocal benefits of an Open Internet'

(Macpherson, 2025b) distinct from both the commercial extraction model and the pay-to-crawl gatekeeping model.

The two relevant initiatives in this space are the ones by the Internet Engineering Task Force (IETF) and Creative Commons (CC). The IETF, through its AI Preferences Working Group, is developing at the time of writing: (i) a standardised vocabulary to express AI-related preferences; (ii) technical methods to attach these preferences with content in already existing formats and protocols, and; (iii) a standard method for reconciling multiple expressions of preferences (Krishna & Nottingham, 2025).

The Creative Commons (CC) Signals project, announced in June 2025, aims to work in close cooperation with the AI Preferences Working Group. CC Signals are 'designed to offer a new way for stewards of large collections of content to indicate their preferences as to how machines (and the humans controlling them) should contribute back to the commons when they reuse and benefit from using the content' (Hardinges et al., 2025). CC Signals are meant to be global tools operating across jurisdictions, so 'applying a CC signal is likely to have a different legal effect depending on who applies it and in what context. Where copyright exists and is applicable, CC signals are intended to leverage the power of copyright without increasing its power'. As the authors put it, CC Signals do not aim to create 'new property rights' but 'to incentivise actions in return' and 'defining manners for machines' (Hardinges et al., 2025). The integration with CC signals to the IETF seeks to reduce complexity for declaring parties and help 'responsible' AI developers comply:

> 'In practice, this means that Declaring Parties will be able to select a category of machine use from this vocabulary to apply their preferences to. These categories, which are still evolving, include — and therefore enable Declaring Parties to distinguish between — TDM, AI Training, and Generative AI Training'. (Hardinges et al., 2025)

Barr (2025) highlights the need for this type of granular signalling by quoting Christopher Flammang Elsevier: 'Today, there is no effective mechanism for those who produce intellectual and creative content to say: "yes, I want to be found through search; no, I do not want to be summarized or rewritten by an AI system". (…) Without that line, we blur two fundamentally different uses: pointing to content (as search has

historically done) and substituting for it (as AI summaries increasingly do).'

Michael Weiberg from NYU, interviewed in Hinchliffe (2025) also makes a case for this standards-based approach:

> The only sustainable way to address this [AI bots] problem is to develop a set of norms and signals that everyone can operate under. That needs to be a conversation that involves collections staff, the teams deploying the bots, and standards setting organizations. I think there is an equilibrium that is technically sustainable for everyone. The only way to find it is to actually talk through everyone's goals and constraints. (Michael Weiberg quoted in Hinchliffe, 2025)

Whether this vision can be institutionalised before the open web's existing norms are fully displaced by infrastructure monopolisation is, the literature suggests, an open and urgent question.

However, **the consistent finding across the literature on technical solutions is that preference expression without enforcement is insufficient to rebalance the power asymmetry between AI companies and content producers.** Brown and Jaźwińska (2025) capture the practitioner's frustration: 'the regulations that underpin how the internet works are being completely challenged by this technology, and these [same tech] companies are in charge of designing how those protocols work'. Hingle and Knodel (2025) are explicit that current technical proposals lack the capacity to enforce standards: 'Technical tools, standards, and protocols provide mechanisms to express content owners' wishes clearly and machine-readably, but they rely on voluntary compliance from crawlers and AI systems'.

5.2 Licensing arrangements

Licensing agreements emerged prominently in the literature as a growing and increasingly significant mechanism for mediating the use of journalistic content in AI technologies. This marketplace has developed in response to the shortcomings of technical solutions and inadequate copyright legal frameworks in protecting the use of journalistic content in LLMs, driving publishers towards two viable strategic choices: 'suing the AI companies for the use of unlicensed content, or negotiating with them to get a fair remuneration' (Carlini et al., 2026). Other authors have

noted that future business models for newsrooms may need to pivot from protecting their content from competitors, and instead towards seeking agreements to reach customers and sustain profits (Wasdahl & Srinivasan, 2026). Due to these impending needs, news organisations are at the forefront of AI content licensing.

Individual, collective, and statutory licensing agreements

Licensing frameworks in the news media and AI context can be classified into three main models: direct voluntary licensing, also referred to as 1:1 licensing, in which organisations voluntarily enter into bilateral agreements with AI companies negotiated on an individual basis; extended collective licensing, which typically involves a Collective Management Organisation (CMO) authorised to license copyrighted works within a defined category on behalf of rightsholders, with terms negotiated between the CMO and users; and statutory licensing, under which AI companies are permitted to use news media content pursuant to terms and remuneration conditions established by law rather than through individual negotiation (Macpherson, 2025b).

Direct licensing agreements

Most deals between AI companies and news organisations represent direct licensing agreements which involve financial compensation for using content, although publishers may also also receive brand attribution, preferential placement, or access to AI-powered technology to use in their own products (Macpherson, 2025b). Recent examples include a March 2026 deal between News Corp and Meta worth up to $50 million annually for at least three years, allowing Meta's AI tools to access News Corp archive and real-time content. Meta has also signed licensing agreements with major publishers including People Inc., CNN, Le Monde Group, and Fox News to support its Meta AI assistant. Other partnerships include The Washington Post and The Guardian with OpenAI, allowing ChatGPT to display summaries, quotations, and links to their journalism; a multi-year agreement between AFP and Mistral giving its Le Chat assistant access to AFP's news output in six languages; and Google's Gemini deals with publishers such as the Associated Press, Estadão, Antara, and Yonhap (Tobitt, 2026).

Licensing agreements are not limited to those involving news content and journalistic data used in the distribution of news stories. The New York Times, for example, signed the first AI licensing agreement with

Amazon making its content more accessible to Amazon customers with direct links to its products (Tobitt, 2026). Licensing deals are also far from confined to news organisations, with Reddit, Axel Springer and Shutterstock among examples of other organisations that have signed public deals for AI content licensing.

Extended collective agreements

Given the difficulty for small and medium-size media organisations to negotiate direct licensing agreements with AI companies, a subset of the literature examines collective licensing as a more equitable framework. The most systematically developed proposal is for extended collective licensing (ECL), under which a collective management organisation (CMO) is authorised to license all copyrighted works within a defined class for specific uses. Macpherson (2025b) explains that under an ECL system, 'all copyright owners in that class would be bound to its terms unless they opt out and choose to negotiate separately. Unlike compulsory licenses, with rates and terms set by the government, the licenses issued under an ECL system are negotiated with users in the free market.' This makes ECL function similarly to voluntary collective licensing, but the government exercises some level of oversight and regulates the overall system.

In this line, the Voss Report recently adopted by the European Parliament calls for the European Commission to establish voluntary collective licensing agreements per sector 'through consultation with collective management societies'. The aim is to establish 'a working licensing market that provides a balanced and efficient framework ensuring the fair remuneration of rights holders while enabling AI providers to access high-quality training data'. The report also 'urges the Commission to ensure that such collective licensing agreements are accessible to all stakeholders, including individual creators and SMEs' (European Parliament, 2026).

A recurrent challenge for collective license in the Generative AI context is the distributional question within the media community. Radsch (2024) argues that 'news organizations should be able to set prices strategically and dynamically and will need to create collectives that can lead negotiations with AI companies'.

Examples of Extended Collective Licensing are discussed further in Section 5.7 in the context of collective action.

Statutory agreements

Voluntary licensing has limitations that we explore further in this section, such as lack of transparency, limited long-term strategy, and exclusion of smaller or niche media outlets.

Licensing deals for AI model training in EU countries are limited, and those that do exist are not fully transparent in their details according to a study by the European University Institute (2025). In this regard, an in-depth analysis commissioned by the European Parliament examines four 'politically realistic' policy options for adapting copyright policy to AI, and finds statutory licensing to be the optimal policy response to maximise welfare, in comparison to approaches such as direct licensing or opt-out exceptions:

> Among realistic options, statutory licensing emerges as a robustly welfare-superior framework under most calibrations. Making the licence compulsory ensures wide access to works, a royalty rate set by a regulatory authority balances interests of rightsholders, AI developers and end users, enabling royalty payments that create incentives for enough new creation to keep AI systems valuable over time. (Peukert, 2025)

The report emphasises the value of statutory licensing as a common and long-term interest based on the importance of continuous flows of new, high-quality data that is needed to sustain the long-term value of AI.

> Aligning incentives through appropriate compensation mechanisms can therefore benefit creators, AI firms and end users in the long run. Ensuring compensation, through a statutory licensing regime, is therefore a common long-term interest, which should simplify enforcement across jurisdictions. (Peukert, 2025)

While a statutory licensing approach may hold promise and has been implemented in the context of news content use through legislation such as the Australian News Media Bargaining Code and Canada's Online News Act, as discussed in Section 5.3, this approach remains, for now, largely theoretical within the EU context for AI licensing.

The emerging marketplace for licensing arrangements

Solutions to facilitate direct licensing agreements are already being developed, leading to the emergence of a distinct licensing marketplace connecting publishers' content supply with AI companies' demand. This works as a way to reduce friction for some publishers and offers a pragmatic approach that can work at scale (Jones et al., 2025).

Radsch et al. (2025) term this the 'AI licensing marketplace':

> They offer publishers a way to control which content can be accessed by LLMs and AI applications, and to receive compensation for access and/or usage. To attract publishers, these intermediaries partner with large AI players, LLMs and AI developers who are mostly seeking to 'ground' their AI-generated output (through retrieval-augmented generation, or RAG), or to obtain data for AI model training. In return, large AI players secure better quality data and reduce their exposure to undermining copyright law. These marketplaces are supported by the intermediary's capability to track AI bots, which enable publishers to block authorized crawling and (potentially) charge for data access. (Radsch et al., 2025)

In addition to avoiding copyright litigation, licensing arrangements may also provide value to AI companies in the form of negotiating access to material that would otherwise not be accessible, obtaining exclusivity or developing new products that require cooperation (Macpherson, 2025b).

Table 1 below provides an overview of some of the companies and organisations in this marketplace that were identified in the literature.

Table 1. Examples in the current licensing marketplace identified in the literature

Company Name	Description [quotes]	Source(s)
Real Simple Licensing (RSL)	'[It] defines specific licensing terms a publisher can set for their content, so AI firms know what data falls under what terms of access. Real Simple Licensing also entails a collective licensing organization that is empowered to negotiate terms and collect royalties. It provides publishers and AI firms "a single point of contact" for paying royalties and provides rightsholders a way to set terms with dozens of potential licensors at once'.	(Macpherson, 2025b)
TollBit	'Through Tollbit's proprietary micropayment system, publishers set their own prices, which can be adjusted over time. This infrastructure allows pay-per-scrape retrieval and manages larger licensing agreements, including some made outside TollBit'.	(Radsch et al., 2025)
	'TollBit lets publishers 'control access, analyze traffic, and prepare for monetization as the agent economy grows'.	(Macpherson, 2025b)
Prorata.ai	'Offers a tool for attribution and monetisation by attributing sources to the output of the language model rather than what is being scraped. They currently take 50% of revenues and are working with e.g. The Financial Times, The Atlantic, Axel Springer, and The Guardian'.	(Jones et al., 2025)
	'ProRata integrates advertising and attribution technologies to enable advertising revenue sharing based on the outputs of LLMs (though so far, only within its own 'ethical' search product, Gist)'.	(Macpherson, 2025b)
ScalePost	'Has launched a private beta of "Pay Per Crawl," a new marketplace where publishers can request compensation from AI companies each time one of their pages is crawled'.	(Macpherson, 2025b)

Company Name	Description [quotes]	Source(s)
Cloudflare	'Has launched a private beta of "Pay Per Crawl," a new marketplace where publishers can request compensation from AI companies each time one of their pages is crawled'.	(Macpherson, 2025b)
	'[It] launched an experimental marketplace that provides web publishers with tools to charge "micropayments" to "AI crawlers" for access to content on their sites. This novel technology apparently provides content owners with the ability to track whether an AI crawler is securing content for training purposes and then to set a price for access, although meaningful implementation may be contingent on cooperation from major AI model developers"'.	(Barnett, 2025)
	'Cloudflare, a Content Delivery Network and cybersecurity company that handles 20 percent of all internet traffic, recently announced that it would block all AI bots by default unless its clients permit access, and it simultaneously entered this burgeoning market with a pilot Pay-Per-Crawl marketplace that lets select publishers charge fees to LLMs and AI applications per accessed webpage'.	(Radsch et al., 2025)
	'Incumbents in web security like Cloudflare were also cited as a promising example of how consent can be managed and other elements of the basic infrastructure of the Internet can be brought into the discussion to try to avoid the fragmentation and bring lots of different actors to the table'.	(Jones et al., 2025)
Created by Humans	'Another startup, Created by Humans, provides a licensing platform through which individual writers and other producers of textual content can license their work to AI model and app developers.'	(Barnett, 2025)
Bria	'A startup, Bria, offers Generative AI developers a platform through which they can acquire licensed visual content for training purposes and other modalities from a portfolio of commercial partners, which avoids liability under intellectual property and privacy laws'.	(Barnett, 2025)

Company Name	Description [quotes]	Source(s)
Copyright Clearance Center	'In March 2024, the Copyright Clearance Center launched a Collective AI License, which provides AI model and app developers the opportunity to license a large pool of content for training purposes'.	(Barnett, 2025)
Perplexity	'In August 2025, Perplexity, a platform that offers an AI-enabled answer and search-engine service, announced the launch of a payments system to make certain payments to publishers when their content is used in their service'.	(Barnett, 2025)
Sphere	'Sphere is a marketplace platform for news outlets and other publishers to charge AI companies for accessing their content. Publishers control which AI companies access their content and set their own prices. For guidance on prices, publishers can also use Sphere's proprietary pricing engine, DataVal, which considers factors such as type of content, recency, engagement, and demand from AI companies, to recommend content valuations less dependent on losses of ad revenues.'	(Radsch et al., 2025)

Community sovereignty-based licensing

Others have also proposed alternative licensing frameworks that better recognise 'community data sovereignty' and adhere to some extent to principles of commons (Chandrasekhar, 2025). For Chandrasekhar, 'frameworks, such as the Nwulite Obodo License, Kaitiakitanga Licenses, the Montreal License, the OpenRAIL Licenses, the Open Data Commons License, and the AI2Impact Licenses hold valuable insights in this regard'.

For example, the Nwulite Obodo license introduces a framework to identify and recognise contributions to content from multiple actors, in this case, of an African languages dataset. It also establishes dual regimes for licensing: 'licensees from developing countries received a royalty-free copyleft license, whereas licensees from countries other than developing countries need to pay for a license to use the dataset' (Chandrasekhar, 2025).

Another example, the Kaitiakitanga License, takes a different approach to ownership by recognising a community as the steward of data, in this case Māori indigenous people. The license also operates with dual regimes: it 'can either prohibit commercial use, or can be modified to

allow commercial use provided that profits from the use of community data is paid back to the community as royalties' (Chandrasekhar, 2025).

Assessing the monetary value of licensing deals

Although voluntary licensing agreements represent a source of income for the news outlets whose data they are drawing on, quantifying a dollar amount for these deals and what this could mean for the news industry remains challenging. In addition, the dominant position of AI companies and information asymmetries has made it difficult for the market to set a sustainable price (Carlini et al., 2026). Responses from a 2026 Reuters report show that publishers have mixed expectations about earnings from AI deals:

> Around a fifth (20%) of publisher respondents – mainly from upmarket news companies – expect future revenues to be substantial, with half (49%) saying that they expect a minor contribution. A further fifth (20%), mostly made up of local publishers, public broadcasters, or those from smaller countries, say they do not expect any income from AI deals. (Newman, 2026)

Holder et al. (2024) estimate the payment that Facebook and Google Search platforms would owe to news publishers for the use of news content, if the proposed Journalism Competition & Preservation Act (JCPA) came into force in the US, and find a compelling case for large payments from platforms to publishers – estimating a dollar amount of between US$11.9 billion and US$13.9 billion a year in the United States. The authors emphasise that more data from AI platforms on user behaviour for news searches and click-through rates is needed to accurately estimate monetary value and loss of revenue from third-party links and summaries.

Regardless of the overall value of licensing deals for publishers, difficulties remain in estimating and assigning monetary value to individual or collective licensing deals. Carlini et al. (2026) discuss a number of potential methods for assessing licensing value, including using CPM rates (cost of advertising to 100 viewers) for comparison, calculating revenues from online advertising received by a platform, and the role of regulation and state intermediaries to mediate value-calculations.

Adding to the challenge of evaluating the value of journalistic data for AI technologies are the differences in data types and uses for AI models. In particular, the differences between data used for model training, model fine-tuning, and RAGs (Retrieval Augmented Generation) could result in different pricing structures. Current practices already reflect 'an emerging industry norm: stakeholders increasingly treat RAG outputs as content uses requiring authorisation' (Dornis & Lucchi, 2025).

Radsch (2024) also emphasises that content isn't one-size-fits-all and that rates and agreements should be set based on careful assessment of the content being provided and following a tiered pricing model: 'Breaking news, thematic verticals and reviews, or local journalism can make real-time searches for information more relevant and accurate. News organizations should strategically set rates that reflect this and consider tiered pricing models that are tailored to the types of content needed for specific use cases.'

Amon et al. (2025) propose that compensation models for AI training data in particular could fall into three main categories: (i) a pay to train model, based on a percentage of training data; (ii) a 'pay to train and inspire', based on contribution to generated outputs, and; (iii) an 'AI Royalties' model, based on a IP negotiated framework.

Concerns and limitations of licensing agreements

Despite the emerging promises of the licensing marketplace, there are a number of concerns and limitations of direct licensing deals that must also be considered.

Exclusion and lack of transparency

Lack of transparency, lack of coordination, and effects on smaller publishers are some of the limitations of voluntary licensing deals. Radsch, for example, argues:

> Bespoke, secretive deals with the largest or most influential news outlets are not a replacement for public policy and will not rescue local news from the precarity created by corporations who skirt the law and enjoy dominant market power. (Radsch, 2024)

According to the 2025 Media Pluralism Monitor results, while 18 countries in the EU did have negotiations or agreements between publishers and

platforms, these were limited to a few media groups, and were in many cases lacking transparency (European University Institute, 2025).

These results point to a critical limitation of voluntary licensing deals – in the absence of regulatory or other enforcement mechanisms, how can licensing the use of journalistic data in AI tools support the broader media ecosystem, beyond funding only the biggest players in the largest markets? In addition, **when the deals that do exist are commonly shrouded in opaque agreements lacking transparency that don't provide clear benchmarks for comparison, it is harder for other players in the media ecosystem with less bargaining power to navigate this emerging market.**

Such a lack of coordination is of particular concern for smaller countries and communities with less widely spoken languages, where AI companies seem unlikely to pay for media content – exacerbating the difficulties already faced by small and diverse media outlets (Carlini et al., 2026). This also has clear implications for media pluralism, particularly in smaller markers and regarding linguistic and cultural representation (Schiffrin et al., 2025).

The lack of transparency of deals is furthermore 'particularly concerning given the public interest in transparency in the information sphere' (Carlini et al., 2026). Transparency in agreements is also important to understand what each license pertains to: whether publishers are providing access to their content to train models, or permission to use their content for generating outputs (Macpherson, 2025b).

There is value in collective negotiation and collaboration between publishers when it comes to licensing agreements, with particular benefits for smaller publishers. On the other hand, if coordination and transparency can be increased in the licensing marketplace and agreements are possible for smaller news outlets, licensing arrangements could potentially have a greater impact for smaller organisations that do not have as many other funding sources and face challenges maintaining their business model (Colussi et al., 2025).

Brand sustainability and disconnection from audiences

There is also a concern that voluntary licensing may be a short-term fix to a long-term problem with many unknowns as to how it will develop in the future. Interviews with news industry stakeholders reveal that in the era of Generative AI, publishers are placing greater importance than ever on cultivating direct relationships with their audiences, with interviewees

from outlets that have strong, direct relationship with their audience feeling more insulated against the threat posed by Generative AI (Brown & Jaźwińska, 2025).

> An executive from a large legacy outlet that has rejected the overtures of various AI companies said, "Clearly, too, what we would think good looks like [in a deal] is being able to maintain direct relationships with users. We're not interested, at the moment at least, in becoming pure suppliers to a platform to no additional end. We are still very much in the direct relationship business." (Brown & Jaźwińska, 2025)

Others have also pointed out that 'as content generated by AI becomes more ubiquitous, it can dilute the value of original journalism and make it harder to identify. This distances the news outlets from their audience, weakening brand' (Park et al., 2026).

This can lead to a dilemma for news publishers, negotiating agreements with AI companies or adopting a counterbalancing strategy, which risks reducing their visibility (Brantner et al., 2025).

Licensing agreements as a short-term fix

The literature review revealed a further concern regarding the long-term effectiveness and sustainability of licensing deals: some authors are skeptical of the promises of licensing agreements, arguing that rather than representing a viable and sustainable solution to the deeper issues affecting the media ecosystem, these agreements represent a strategy by AI players to pre-empt or distract from broader regulatory reforms. This view emphasises that '(d)irect unilateral support to media outlets by tech platforms should not be a replacement for legal regulatory frameworks that seek to create a level playing field and support independent media' (Radsch, 2022), and that such licensing practices may disincentivise large publishers away from litigation or advocacy for legal reform.

Licensing also does not address structural aspects that affect the broader media ecosystem, including increasingly entrenched issues of the monopolistic power of large tech and AI companies (Brogi & Sjøvaag, 2024; Díaz-Noci et al., 2024). Related to this, a further risk is that the funds gained through such direct agreements are used by publishers and shareholders as profits or dividends – rather than as investments in a more sustainable news ecosystem overall, which has further flow-on

effects for smaller publishers who do not receive the same profits from licensing (Flew et al., 2024).

Ultimately, as Brogi & Sjøvaag argue:

> Redistribution of copyright revenues can serve to target the imbalance in bargaining power between news media and digital platforms and can be a source of revenues for the media outlets, but, so far, cannot be considered a practice or a policy that, per se, can resolve the problem of sustainability of the news media sector. Namely, existing redistribution mechanisms are difficult to negotiate, untransparent and tend to favour large, incumbent media organisations, and are thus unlikely to resolve power imbalances between the platform sector and the media sector. (Brogi & Sjøvaag, 2024)

Privatising a public good

Other authors have criticised the basis of licensing agreements, which effectively commodify and privatise the public good that is public interest news media content and public policy: '...policy interventions like the Media Bargaining Code have the effect of privatising notions of the public good by linking them to market-based logics and practices' (Brevini, 2023).

The next section goes beyond licensing approaches to cover more structural solutions grounded in law and policy regulation.

5.3 Law and policy approaches

> (...) journalism cannot be expected to adapt its business models to the AI era without interventions by policymakers to correct market imbalances, enforce intellectual property rights, and require data access and transparency. Industry action must go hand in hand with legislative and regulatory action. (Radsch, 2024)

In response to the regulatory inadequacies reviewed in Section 4.6, the examined literature proposes a diverse array of solutions, from copyright reform, competition policy, or institutional restructuring. Overall, there is an emerging landscape of regulatory and policy approaches designed to address the power asymmetries between AI developers and public

interest media, establish fair compensation mechanisms, and safeguard the sustainability of public interest media in a GenAI-mediated information ecosystem.

Recent debate has also focused on whether copyright law should be updated to account for AI-generated works, including in journalism (so-called 'robojournalism') (Trapova & Mezei, 2022). Although related, this question lies beyond the scope of this report. Rather, the analysis concentrates on the role of regulation and copyright in protecting the value of journalistic work and data used in AI technologies.

The proposed solutions operate across multiple levels of intervention. As Radsch (2022) observes, these policy areas include 'digital taxation, competition policy (also referred to as antitrust), and intellectual property'. This multi-dimensional approach acknowledges that no single mechanism can adequately address the structural challenges facing public interest media. The sections below examine the most frequent themes identified in the literature.

Copyright reform

Copyright is the most extensively considered domain of regulatory intervention in the examined literature. The current debates concentrate on:

- The role of regulation and copyright in protecting the value of journalistic work and data used in AI technologies
- Whether existing IP frameworks adequately address the large-scale use of journalistic and creative content to train AI systems
- The need to reform, reinterpret, or supplement these frameworks

In the EU, the European Parliament's own-initiative draft report (known as the Voss Report) on 'Copyright and generative artificial intelligence – opportunities and challenges' from June 2025 outlines the need for transparency, consent, and fair remuneration for creators whose protected works are used in AI training, generation, dissemination, and distribution (European Parliament, 2025). The study commissioned by the European Parliament emphasises the need for comprehensive reform, calling for 'clear rules on input/output distinctions, harmonised opt-out mechanisms, transparency obligations, and equitable licensing models' (Lucchi, 2025). At the time of writing, nevertheless, the latest

version of the Voss Report (February 2026) has moved away from earlier drafts. It acknowledges that AI training is covered by the Text and Data Mining (TDM) exceptions of the CDSM Directive and therefore lightens the emphasis on opt-out mechanisms to prioritise the importance of remuneration and transparency but, as argued by Keller, maintains a reluctance to endorse decisive approaches other than voluntary licensing mechanisms (Keller, 2026).

Emerging regulatory thinking also focuses on AI-specific frameworks that go beyond traditional copyright. Several authors argue that copyright, regardless of reform, is insufficient on its own and that broader legislative imagination is required. Kuai (2024) cautions against 'the rhetoric of copyright' and calls for 'broader regulatory imagination, and more inclusiveness and prudence during policymaking.' Teilmann-Lock and Savin (2025) propose that the EU's AI Act may offer the 'outline of a conceptual approach' to what they term a 'dataset law' that would address questions of dataset ownership and access beyond copyright, as it opens up the possibility of a 'bespoke, timelimited exclusive right for curated training data sets'. This would make it possible to 'monetise quality data, give authors a collective bargaining fulcrum, and complement the AI Act's transparency duties without rewriting either copyright law or the Database Directive' (Teilmann-Lock and Savin, 2025).

International interoperability is another key aspect to be addressed. To enable this, Hyun et al. (2025) propose a framework to harmonise US and EU approaches to policy in this field. These five principles include: '[T] ransparency through content provenance standards, protected research activities, tiered compensation for training and outputs, evidentiary requirements for creative control, and risk-based compliance'.

To summarise, **there is a view that current copyright laws need to be updated or complemented to respond to the data-based extraction of AI systems, and that transparency and remuneration are key. But there are also views that these measures are not enough to address the asymmetries of power and that more decisive legislation, which we cover in the sections below, is needed to level the playing field.**

Competition policy

A second major regulatory topic, competition policy aims to correct power imbalances between media outlets and Generative AI companies through competition interventions and bargaining frameworks.

Holder et al. (2024) report the early examples of Australia (2021) and Canada (2023), whose regulatory instruments (the News Media Bargaining Code and Online News Act or C-18, respectively) set a model for rebalancing bargaining power between media outlets and digital intermediaries. From the countries mentioned in Holder et al. that followed suit, Indonesia (2024 Presidential Regulation on Publisher Rights for Quality Journalism) and New Zealand (Fair Digital News Bargaining Bill, Second Reading in Parliament) are advanced examples in the same direction.

Examining the Australian News Media Bargaining Code as a potential model, Balasingham et al. (2025) note that its underlying principle is to create a more level playing field considering the dominance of digital platforms. They conclude that in the light of the shortcomings of the EU's CDSM, 'the most adequate protection for press publishers is provided when their right to remuneration is enforceable through a regulatory instrument like a bargaining code, which is accompanied by an efficient and balanced enforcement mechanism'.

Others adopt a more cautious approach towards the Australian code, highlighting that 'it ought to do more to foster the public interest. This could be achieved by tying the code more closely to public interest journalism and by mandating some transparency' (Molitorisz & Attard, 2024). Sandrini and Somogyi (2023) endorse the extension of news media bargaining codes to AI companies, not only digital platforms, so that they cover AI-generated content more explicitly.

In contrast, the most comprehensive critique comes from Colangelo (2026), contending that legislators have 'placed the blame squarely on large online platforms', overlooking that news aggregators and platforms often generate net traffic benefits for publishers. To Brevini (2023), the question lies in the extent to which the Australian Code was able to engage regional and rural publishers in the negotiations.

Beyond the case of Australia, more structural critiques point out that AI industries remain unregulated 'beyond general competition law' and that policy makers need to pay more attention to 'how AI influences states' ability to ensure that their citizens have access to news and

information' (Sjøvaag, 2024). In their critique of EU's policy towards social media platforms, Oleart and Rone (2025) call for public investment in alternative digital infrastructure beyond regulation alone, a topic we cover in Section 5.5.

Platform regulation

In Europe, regulatory instruments such as the Digital Services Act (DSA), the Digital Markets Act (DMA), the AI Act, and the European Media Freedom Act (EMFA) are relevant hard-law pieces for the intersection of Generative AI and journalism.

However, despite hopes that '[a]n effective enforcement of the provision of Art. 53(1)(c) AI Act may empower the rightsholders in the negotiating process, in a market characterized by a steep imbalance of power' (Carlini et al., 2026), recent research shows that, as of February 2026, none of the leading AI developers have published summaries of the data they use to train their models as mandated in Art. 53(1)(d) of the AI Act (Blankvoort et al., 2026). To Abiri (2024), the Digital Services Act (DSA), although imperfect, offers a better framework than the AI Act for regulating Generative AI as it treats it as a systemic media platform rather than a mere technical tool and applies a graduated oversight model where accountability scales alongside a platform's societal reach. By mandating transparency in content moderation, auditing the curation algorithms that drive user engagement, and curbing the data exploitation inherent in 'informational capitalism', the DSA moves beyond narrow risk mitigation to enforce holistic democratic values and proactive accountability: 'The DSA remains an imperfect work-in-progress. However, in contrast to the AI Act's narrow focus on technical risk management, the DSA holds significant advantages for regulating the uniquely media-centric functions and societal impacts of Generative AI models' (Abiri, 2024).

In general, while identifying their potential, the literature considers that the EU's regulatory instruments do not adequately address the specific challenges posed by AI to journalism. Schneiders and Stark (2025) find that the current EU rules do not fully address power imbalances and do not protect local journalism. Brantner et al. (2025) note that although the DSA and AI Act do require platforms to conduct risk assessments and disclose to some extent algorithmic processes, these are insufficient for the specific dynamics of news sourcing in AI-powered search. Likewise, Dragomir et al. (2024) state:

> The EU has addressed global platform power in recent attempts to safeguard its digital future, including the Digital Services Act (DSA), Digital Markets Act (DMA), and the European Media Freedom Act (EMFA). However, these acts do not adequately address PSM's two central and often interconnected problems: funding challenges and political pressures. (Dragomir et al., 2024)

The EMFA is also examined by Longo (2024), who traces its legal basis in Article 114 TFEU and by Carlini (2025), who analyses the Act's 'media plurality test' for digital concentrations, noting that a merger involving a platform providing access to media content (such as between a search engine and an AI company) may come within its scope.

Visibility-related policies are another point identified in the literature. They are considered 'policy — as well as (self-)regulation — which offers the media special treatment that is intended to improve its visibility and accessibility for users on online platforms' (Schiffrin et al., 2025). There is a view that they 'can strengthen individuals' exposure to public interest, diverse, or local journalism' and 'increase the media's financial sustainability and incentivize the media to produce more high-quality content compared with 'click-worthy' content to optimize attention and advertising revenue' (Schiffrin et al., 2025). Schiffrin et al. add that such policies could be implemented by mandating digital platforms to offer special protections to public interest media organisations. Yet, these authors also warn that 'if introduced in restrictive media environments or without safeguards, visibility policies could be captured in ways that undermine independent journalism', citing Russia's 2017 law holding news aggregators liable for 'fake news' as a cautionary example.

Monetisation beyond copyright: levy proposals

A distinctive cluster of proposals centres on financial redistribution (e.g. levies, taxes on digital platforms, or mandatory contributions to journalism funds) as a mechanism for internalising the externalities imposed by AI companies on the information ecosystem. Levy proposals work as mandatory compensation mechanisms, using the revenue captured by Generative AI companies at the aggregation layer to fund the content production layer that makes aggregation possible. Keller (2025) makes the most developed case for a levy-based approach that goes beyond individual copyright enforcement:

> [E]nsuring economic sustainability requires redistributive mechanisms beyond copyright law—specifically, a levy system activated when commercial AI services are deployed, channeling revenue back to all contributors to the information commons. (Keller, 2025)

More specifically, Keller (2025) argues that Generative AI should be regulated through a mandatory remuneration model. In this model, the system would shift the financial burden to AI providers at the moment their service enters the market, proposing two primary funding mechanisms: (i) a levy (fixed fee) offering business predictability and a clear link between consumption and payment, and drawing on established media policy precedents, and; (ii) a tax (percentage of revenue), which is a more progressive option that scales with a company's financial success, capturing generated value (vs. volume of use). Unlike current copyright law, which focuses on specific, protected works, this framework treats the entire body of 'publicly available information' as a digital commons that AI services profit from.

Keller concludes that a levy is preferable to the tax-based approach because it can be specifically earmarked for the information production ecosystem rather than disappearing into a general government budget. By leveraging Collective Management Organizations (CMOs), the system would bypass the 'copyright-only' bottleneck, allowing for the distribution of funds not just to traditional rightsholders, but to the broader network of creators and contributors who sustain the public knowledge base. In Keller's words:

> The system would operate on the principle that commercial AI services benefit from and should contribute to the commons of publicly available information, regardless of the copyright status of individual components within that commons. Collective management organizations would ensure appropriate distribution of levy proceeds to professional creators and other rightholders, but the system would simultaneously recognize and compensate the broader ecosystem of contributors who fall outside traditional copyright frameworks. (Keller, 2025)

Some voices in favour of a levy also come from industry. Mistral CEO Arthur Mensch recently proposed 'a revenue-based levy that would be applied to all commercial providers placing AI models on the market or putting them into service in Europe, reflecting their use of content publicly available online' (Mensch, 2026). The focus in this case is providing legal

certainty to AI companies while creating a 'central European fund' to support cultural sectors in Europe. In Keller's case the levy 'would help ensure the continued vitality of the knowledge commons on which AI systems depend', providing 'a foundation for public AI infrastructure, creating genuine alternatives to purely commercial models' (Keller, 2025).

The next section covers the viability and sustainability of the journalism sector from the point of view of changing business models.

5.4 Business models and sustainability strategies

In addition to regulatory approaches, the literature points to the need for new business models and investment for the media to adapt to a changing ecosystem and reflect new workflows, infrastructure, and audience preferences – even if this area is still emergent, as media organisations are now grappling with challenges, assessing options and testing or piloting solutions.

As covered earlier, there is some consensus that relying on the development of bilateral agreements between publishers and AI companies risks creating a dependence on one income source and diminishing collaboration between news organisations, further threatening the sustainability of the industry.

The current issues of asymmetric power are further accentuated by the complexity of news markets: news producers serve a public good function necessary for democratic processes, often in ways that are at odds with maximising profits, but are 'affected by that market, which is made up of both commercially driven and less commercially driven firms' (Park et al., 2026). For Park et al., this 'complex environment requires all news providers to have sustainable business models if they are to survive and flourish'.

New business models: towards loyalty and diversification

> If the dominant business model in journalism – dissemination funded by advertising – crumbles further, we will need a clear new direction: toward direct relationships with the audience instead of click rates, toward spaces for dialogue instead of a narrow focus on distribution,

> and toward trust garnered by editorials who know the needs of their users, tailoring formats and topics accordingly. (Exner, 2025)

The literature points towards diversification and closer connections with audiences as key business model strategies. Importantly for digital media organisations, a shift away from advertising revenues, prompted by the efficiency of large-scale advertising on algorithm-driven platforms (Park et al., 2026), is expected to persist: 'Advertising is no longer the main support for journalism and there is no one main support to replace it' (Burger et al., 2024). A recent report by the Forum on Information and Democracy also finds the current concentration of power has distorted incentives in the advertising market and that AI assistants will further entrench this dynamic (Atkin, 2026). A 2026 Reuters report, based on a strategic sample of 280 digital leaders from 51 countries and territories, shows that subscription and membership models remain the main revenue focus for publishers (76%), ahead of advertising. Interestingly, online and in-person events are also becoming more important (54%). Following cuts in media support budgets, reliance on philanthropic and foundation support (18%) also declined in 2025 (Newman, 2026).

News organisations in the media landscape need to find nimble and adaptable ways to survive, which involves tapping into a number of different revenue streams, including paid subscriptions, but also other forays such as podcasts, documentaries and a range of services and formats, partnerships, one-off commercial deals, newsletters, membership models, and crowdfunding approaches (Darling, 2024). While subscriptions provide predictable payments and loyalty, and have become important in countries like Norway and Finland which have high rates of news consumption, subscription rates have remained stagnant in many other countries, including the UK and the US, according to Parks et al. (2026). Subscription models have also been criticised for restricting access to high quality news, potentially limiting the democratising potential of journalistic data.

Other funding models are also being explored. A study in Switzerland, for example, found that online local media are testing strategies such as crowdfunding, voluntary contributions, and strategic patrons, such as foundations and NGOs (Burger et al., 2024). In Hungary, Urban et al. (2026) find that news organisations have used community-building strategies to support sustainability, including the organisation of in-person events. Strengthening direct relationships and community-

building with audiences may also alleviate the risk of disconnection between readers and journalists posed by the use of Generative AI tools in receiving and generating news for users.

Changing business models may also influence newsroom strategies on which types of content to produce and invest in, with a Reuters report revealing that publishers plan to scale back service journalism, evergreen content, and general news, which many expect to become commoditised by AI chatbots, and instead plan to invest in original reporting and on the ground reporting, as well as video and audio formats (Newman, 2026).

What represents a successful business model for one organisation may not work for another, with clear differences in funding strategies for small news publishers, and between commercial and not-for-profit providers: 'News organisations thus must find elements in other business models that are more appropriate for their own situations and create models tailored for their purposes and environments' (Park et al., 2026).

Yet, the literature is clear that business models alone will not be enough for the news media ecosystem to address the current structural challenges outlined in this review. The next section covers the need for public and philanthropic investment and priority areas for investment.

5.5 The need for funding and investment

It is clear from both industry and academic literature that investment is key to support the media sector in adapting to new systemic changes. As we have seen, small and medium-size organisations in particular lack the resources and time to invest in new infrastructure or strategise new business models. The literature points out the need for both states and philanthropic actors to invest in independent and public interest media, as well as in digital public interest internet and digital infrastructure that can offer alternatives to BigTech.

Public Investment

Different models of public and philanthropic funding for journalism are currently being explored in Europe. The European Commission announced in 2025 that it would double the budget for media in the new Multiannual Financial Framework (MFF) for 2028-2034 through the

AgoraEU Programme, with the most recent proposal resulting in a €7,6 billion budget, of which €2,8 billion was allocated to media (Georgescu and Hausner, 2026). However, the media ecosystem has warned that public EU funds need to go to 'deserving media' and has asked for an impartial independent mechanism for funds distribution (Money-Kyrle, 2026). The non-profit Media Development Investment Fund (MDIF) has also led an innovation called Plūrālis, a mission-driven investment vehicle for Council of Europe countries. Plūrālis represents a 'structural intervention in independent media economics, assembling a diverse coalition of investors to create a financial governance where editorial independence is the primary asset' (Schneider, 2025).

Public funding can be directed by more local regulation to support local and investigative journalism. As outlined by Goudarzi (2024), there are several examples of this in the United States, such as the California Local News Fellowship program, which provides full-time positions for beginner journalists in local newsrooms of deprived areas for two years, the New Jersey awards grants for local news and information, or the Illinois $25 million tax credit for local news. In Australia, there exist proposals such as a tax offset scheme for journalists' salaries, tax returns for news subscription payments, or reduced GST [VAT] on news products (Park et al., 2026).

Investing in public digital infrastructure

As introduced in earlier sections, building alternative public infrastructure is seen as an important avenue to rebalance power in front of large AI companies and digital platforms.

In the context of the European Union, the literature points out the lack of investment in recent years in alternative forms of technology and social media (Oleart and Rone, 2025). The case for public investment in journalism is bolstered in the absence of alternative, independent digital platforms, which makes the possibility of alternative news media to thrive contingent upon the strategies of the tech giants:

> As such, the conditions for a new form of digital public sphere have not come to fruition. Public support for sustainable news production is a necessary, if not sufficient, condition for the advancement of news that meets the goals of journalism for information, investigation, analysis, social empathy, advocacy, and provision of public spaces for democratic debate. (Flew, 2026)

Therefore, investing in alternative public digital infrastructure can support public interest journalism and address the dependency of media organisations on Big Tech. For example, it could entail funding interoperable open-source ecosystems and commons-based approaches to governing public interest media content that could at the same time facilitate the sharing of resources and technology for smaller and niche media organisations. The Public Service Media and Public Service Internet Manifesto sets out a series of demands in this regard, including '[s]ecuring the existence and funding of PSM as an indispensable component of European democracy and culture' (Thomass, 2024). As Thomass states:

> [T]he basic idea advocated here, that publicly funded media orientated towards the common good can represent a counterweight to media concentration, applies all the more to the internet, which has experienced an unprecedented development of concentration over the past 20 years. (Thomass, 2024)

The contribution of Big Tech to public funding

Investing in a digital public infrastructure would entail sustained funding for public internet services based on approaches such as a public service tax or transnational funding mechanisms. Or, as we have developed in Section 5.3, tax credits or a levy for big platforms and Generative AI providers (Keller, 2025; Schneiders & Stark, 2025).

In addition to a contribution via levies or taxes, there is a section of the literature that also covers commercial partnerships and payments by AI companies to news publishers. While these are not strictly speaking licensing deals, they involve investment to experiment with new formats and ways of displaying information, including attribution and context, and testing ways to drive audience engagement – approaches that also enhance AI assistants with real-time information displayed in different ways, such as the partnerships signed by Google with various national media outlets (Tobitt, 2026).

Philanthropic Investment

The philanthropic sector has been another essential funding avenue, especially for local and investigative media organisations, and part of the business model of these outlets. The importance of this type of funding has increased due to the loss of advertising avenue (Park et al.,

2026) and the multiple challenges facing the media sector. For example, according to a recent report by the Philanthropy European Association, some funders reported 'an increasing focus on tech accountability, the role of platforms, and how AI is impacting the media' (Philea, 2026).

Philanthropic support for news organisations has other benefits, such as supporting investigative journalism that commercial news organisations may not prioritise, and enabling open access of information to readers without pay walls (Darling, 2024), but it is worth noting that receiving money from funders can be problematic in its own right. Philanthropic support can trap media organisations as competing demands are placed on them, like achieving both wide reach (civic impact) and economic sustainability (Benson, 2018). Philanthropic support further risks serving to reinforce issues favoured by philanthropic donors, and favouring to some extent media capture by foundation donors (Benson, 2018).

A study of philanthropic media funding in Hungary supports the notion of philanthropic funding as a partial and transitional solution, finding that philanthropic grants 'enabled survival, professionalization, and audience gains, but rarely produced durable revenue diversity in markets distorted by politicized state advertising, advertiser risk-aversion, and low willingness to pay for news' (Dragomir & Nemeth, 2026).

Building philanthropic coalitions for greater impact

Another theme in the literature points towards the increased need for philanthropic funders to work collaboratively for greater impact. As stated by the funder Civitates, which pools funds for both grantees and funding members: 'Philanthropic funders can have greater impact [on media and democracy] through collaboration' (Civitates, 2025). Other media funders like Limelight foundation are also building a €100M fund 'to secure access to news in every European country' (Limelight Foundation, 2026).

5.6 A commons approach to pooling and governing media data

The complexity of the current challenges and the extent to which media content has become valuable data for Generative AI models has led to

the proliferation of commons-based approaches to pooling resources and governing data as a way to leverage power.

Commons-based approaches are seen as a response to Big Tech's power concentration. They entail collective and democratic forms of governance to protect the public good that is public interest data and media content from extractive capture by a few profit-driven tech companies. Commons can tilt the balance of power in decision-making about how data, and ultimately knowledge, is used and who benefits from it.

> The commons is an approach that allows us to rethink collective ownership and governance regimes and empower more groups in society to have control over and a say in the development of AI and how the benefits of AI can be shared with society at large. Instead of the GPT AI resulting in Big Tech as winners and most of us losers, a commons approach allows more groups in society to be among the winners of AI innovation. (Verdegem, 2024)

Unlike large-scale approaches to collecting data in largely invisible or untraceable ways, 'democratic hypercommons would involve technologies in which the creation, training, and use of Generative AI systems, as well as the benefits of their operation, are distributed in more transparent and equitable ways' (Islam & Greenwood, 2024). Managing a knowledge or data commons can take many forms, but democratic and participatory governance tends to be a key defining feature: 'data cooperatives, collectives, commons and trusts, all (…) are premised on the idea of data as a collective resource conducive to governance models designed for shared benefit' (Macpherson, 2025a).

Data collaboratives for media

Data collaboratives can be seen as a type of institutional arrangement that help media organisations pool resources and decide collectively how the data they create 'serves the public good and contributes to a more equitable and inclusive media landscape ' (Rahman, 2025). In this regard, Rahman proposes that Public Service Media organisations establish 'AI cooperatives or collectives within their communities and professional associations, such as the Public Media Alliance (PMA) and the European Broadcasting Union (EBU) with the objective of pooling AI resources to pursue common goals'.

This may be particularly helpful for small publishers: 'Data collectives may represent a promising solution for small publishers who lack the archives, scale, resources or expertise to negotiate with – or even be found by – AI firms' (Macpherson, 2025a).

European media data spaces as data collaboratives

The European Union is building a large media data space for the European media ecosystem to pool and share resources: the Trusted European Media Data Space (TEMS). It is funded by the European Commission and is 'the flagship European initiative to build a resilient data-driven ecosystem in the media sector'. It aims to enable '[c]ross-border collaboration and data-driven technology for a more competitive European media industry' (TEMS, 2026). Determining how this ecosystem will work in practice and be governed is part of the initial set-up phase. TEMS surveyed the media ecosystem to understand what type of membership relationship would work best for media organisations, and found that a majority (57%) favoured pay-per-use, service-based arrangements, contextual or revenue-sharing models. While some were willing to pay a subscription fee to benefit from the pooled resources and services from the ecosystem, only a small proportion (12.5%) were ready to pay outright, while 17% considered only free options. Overall, most respondents were 'willing to contribute financially to cover technical and governance costs provided the features and pricing are appropriate (72%)' but TEMS concluded with the need to 'test flexible mechanisms on the market, so as to facilitate the participation of organizations of all sizes and capacities' (TEMS, 2025).

Media data spaces have also been pursued at the state level. In Germany, for example, media organisations like ARD, ZDF, RTL Deutschland, ProSiebenSat.1, and the German Press Agency (dpa) have been exploring the creation of national media data infrastructure supported by the Hamburg Ministry of Culture and Media, the co-initiators of the "Beyond Platforms Initiative," and the Fraunhofer Institute for Software and Systems Engineering ISST (Pion, 2025). This ecosystem would help German media have more control over how their data and public interest media content is reused, and develop innovative, collaborative services.

Others, like RadicalXChange or Project Liberty are also exploring frameworks for media to 'exert democratic collective bargaining power over their data' (Macpherson, 2025a). RadicalXChange, for example, is working with OpenMined's remote data science software to develop

a framework in which 'publishers can negotiate for joint decisions controlling the use of their data and appropriate compensation' (2025).

Data governance mechanisms

Key to building and sustaining these data ecosystems is robust data governance, which could draw on existing frameworks for data governance. This includes the FAIR principles – findable, accessible, interoperable, and reusable – and the CARE principles – collective benefit, authority to control, responsibility, and ethics. The literature also points to the value of global Indigenous data sovereignty movements to 'empower marginalised groups to become active contributors and co-creators of a collective future' (Rahman, 2025). As mentioned earlier under the licensing Section 5.2, data governance approaches exist to strengthen community sovereignty over data.

The literature points also to the need for data governance that prioritises democratic values and participation. Mansell et al. argue that inadequate data governance strategies are behind the proliferation of misinformation, as they have not been designed democratically or to serve a public interest purpose:

> These strategies work against information ecosystems that are consistent with democracy; they interfere with political deliberation that is essential for the survival of participatory democracy. Data governance is needed to encourage data and digital infrastructures that enable political communities and allow them to shape the affordances of information ecosystems. (Mansell et al., 2025)

Data governance applies not only to media organisations' use of data and how data collaboratives are governed, but also to the use of data by Generative AI companies. This is why the literature further covers the need for regulation and developing dataset law in particular, as covered in Section 5.3.

5.7 Coalitions and collective action

The need for collective action solutions and collaborative approaches is another key theme identified in the literature in relation to pathways for protecting both content integrity and sustainability.

The level of asymmetry and concentration of power as well as the affordances of Generative AI and their impact on public interest media content and principles mean that the gap in bargaining power is prohibitive. In addition, as seen earlier, bilateral licensing deals do not solve the issue of information integrity, language, context and the needs of smaller media and investigative journalism organisations. As argued by an executive in Brown & Jaźwińska (2025): 'Generative AI is fundamentally different (…) you may need much more collective ecosystem relations. (…) Whether it's a tax or a profit share that goes into a pool that goes wider, there are approaches like that on a collaborative thing. But these one-to-one, confidential deals are not good for the overall ecosystem'. As reiterated by Oleart and Rone (2025): 'collective action needs to be at the centre of any attempt to democratise alternative digital infrastructure'.

Collective action can take many forms. Some have been explored under the sections on licensing and business models; others relate to sharing infrastructure and governance through frameworks such as media data commons, as explored in Section 5.6. As explored earlier, 'Journalism's resilience depends on diversified revenue and shared infrastructure, with cooperation between large and small publishers to reduce duplication and widen reach while protecting editorial independence' (ALT Advisory, 2025). Collective agreement is also deemed necessary in being able to authorise access to media content by AI assistants and in identifying an agreed format for citation and attribution (BBC & EBU, 2025).

What emerges is that 'industry consortia can be instrumental in identifying and promoting best practices for AI in the public interest' and are 'well-positioned to serve as the entity that develops best practices for the whole sector' (Cihon et al., 2021). Instead of each organisation reinventing the wheel, industry coalitions can pool collective experience and action, and then disseminate these to the media ecosystem and individual organisations. This requires distinct categories of roles and organisations to be aware of the entire ecosystem, interactions, and how to come together (Cihon et al., 2021).

There are examples of collective action that involve regulation as well as the set up of collective management institutions, as outlined by Bleyer-Simon et al. (2024). For example, Czech media companies launched a collective association in 2024 to represent them in negotiations with platforms[1], and Denmark's Ministry of Culture approved the Danish Press

1 https://www.slpv.cz/en/

Publications Collective Management Organisation (DPCMO) in 2023, which has the authorisation to reach agreements on behalf of publishers. DPCMO, which represents 99% of the Danish Media Industry, launched legal action against OpenAI for using content from its members to train its models. This is following a failed attempt to negotiate with OpenAI, and failure of a Minister for Culture to intervene (Ronde, 2026). Similarly, a collective management organisation called Xenofon was set up in Greece2 and one called Repropress in Hungary3, whose processes were hampered by changes in regulation (Bleyer-Simon et al., 2024).

Ensuring collective action works for small and niche organisations

A challenge identified in the literature is the extent to which collective licensing management includes and works for small, digital-first or niche organisations. When legacy media have the power to negotiate deals, they can miss opportunities to include smaller or digital-only organisations who may not have the audience build-up or trust of legacy media, nor the resources to strike deals (Colussi et al., 2025). This has an impact on non-dominant or minoritised languages and can create news deserts in specific regions or communities.

As argued by Cook et al. (2023), niche organisations, such as those in exile or facing political pressures, have had to develop strategies and experiment in order to survive. This can provide valuable lessons for other small, niche, or big organisations when it comes to being creative in developing business strategies.

Another approach taken by small media organisations is to pool strength and resources through a cooperative governance approach. For example, the cooperative CoopMèdia was recently set up in Catalonia as a network of fourteen small cooperative outlets with this aim in mind (Directa, 2026).

Surmounting competition and sustaining collaboration

> Solutions to complex problems such as news deserts, news avoidance, mis/disinformation, and algorithmic biases require the engagement of a multitude of actors across levels of governance and

2 https://xenophonrights.gr/

3 https://repropress.hu/

> the implementation of diverse strategies in practice, which explains why PSM is increasingly called upon to play a major role in these collective endeavours. (D'Arma et al., 2025)

Cooperation is deemed necessary to address the complex problems of both democratic integrity and sustainability of the media. Yet, the literature also identifies some skepticism towards cooperation due to economic asymmetries between different media, and due to the limited resources for which especially more disadvantaged outlets compete. This can result in a wary approach to cooperation, which can particularly affect 'news media operating in multiple languages, inconsistent availability of audience data and patchy internet access' (Cook, 2023).

Collaborative engagement therefore requires in-depth and intense dialogue and spaces for reflection and exchange, which require resources, coordinators as well as a representative body to sustain the process and avoid the 'cliff post-event' and ultimately move from ideas exchange to action (Cook, 2023). Some convening and participatory initiatives, such as News Futures 2035 in the UK, concluded with the need to create a long term forum:

> We recommend that everyone who has a stake in the crisis affecting the production, distribution, and consumption of public-interest news should come together to create a News Futures Forum – a long-term initiative, modelled on track two diplomacy, where participants with very different perspectives are able to meet in a high-trust environment to talk about shared challenges and find solutions. (Nel & Rymajdo, 2024)

5.8 The use and governance of AI tools in newsrooms

While not considered a structural approach to gain leveraging power in the current complex context, it is argued in some of the literature that some AI technologies themselves can represent a tool for public interest media to innovate and maintain sustainability in a digital media landscape. As discussed in previous sections, media organisations and journalists are already using and adopting AI tools for a number of different functions. For example, Greek newspaper Makedonia used AI tools to revive regional journalism through personalised community

engagement, and Digitalhaus Franken, a German digital media house, used AI to predict and prevent subscriber churn (LSE - POLIS, 2024). Below we outline the different ways in which AI tools are presented as potentially helpful for the media.

Recommender systems that follow public interest principles

The literature identifies a number of ways in which the use of AI-driven recommender and personalisation systems for news distribution, in particular, can be improved and tailored to support public interest media.

> [T]here is a need for a "public service" algorithm that goes beyond the narrowing down of choice to personal preferences to ensure serendipity, that is, to ensure that personalised services invite users to go beyond their personal interests and to expose themselves to alternative views and interests. (Van den Bulck and Moe 2018, quoted in Pérez-Seijo & Nuno Vicente, 2024).

Van den Bulck et al. (2025) have recently argued that the distinct public-interest mandate of PSM should shape not only the content itself but also how it is distributed through the design of recommender algorithms guided by principles of integrity, noting that 'PSM principles have guided not just what kind of content is offered but how it is offered'. In this regard the authors propose an 'epistemic welfare' framework to operationalise public interest algorithmic recommender systems in a way that advances 'citizens' epistemic agency in the digitalized public sphere' (idem). They propose to do so through the following principles: 'promoting the dissemination of and access to epistemically valuable content (reliability), that answers the greatest number of questions (power), to the broadest and most diverse audiences (fecundity), with minimal time (speed), and limited costs (efficiency)' (Van den Bulck et al., 2025).

This is not only important for media organisations, but also for tech and AI companies that have a role in distributing and displaying news to the public. For example, others have suggested that content from reliable epistemic sources could be highlighted and labelled in newsfeeds and AI news summaries (Schneiders & Stark, 2025). Crucially, this relies on platforms being able to access and responsibly use information from these sources. In this regard, the literature highlights the risk that AI companies could be restricted from drawing on such material if

agreements are not reached, as well as the risk of AI summaries being created by error-prone LLMs (due to hallucinations) and that continue to misattribute content. Alternatively, new platforms oriented to public interest metrics and criteria for access and presentation could be developed as a counterweight to dominant platforms (Schneiders & Stark, 2025).

Building the skills for data stewardship and reuse

Despite a heightened focus on AI-driven tools and algorithms, a prioritisation of data as the foundation of AI is perhaps less evident. A 2025 survey of European journalists found that although most respondents viewed data as very important to their operations, fewer than half had a dedicated data team and 37% lacked a formal data strategy altogether. Although data was widely recognised as important, many organisations were hampered in moving beyond intention by budget, prioritisation and governance constraints (TEMS, 2025). Investments in technical skills, tools, and organisational culture will be required for organisations to fully realise data-driven practices, and training on data stewardship capacities within the news industry could allow media organisations to develop both strategic and ethical data-sharing and safeguarding practices.

This is particularly important given that data collaboratives and data spaces are a potential ecosystem-based solution for media organisations to pool and share resources as we have explored in Section 5.6.

Tailored community engagement and feedback

The use of AI tools has also been explored by news organisations to tailor audience engagement and feedback. In the case of Makedonia, for example, the team is leveraging AI to develop an advanced Personalised Newsletters and Push Notifications (PNN) system based on an analysis of user behaviour and preferences. This results in 'a more meaningful, personalised news experience that builds loyalty and trust over time' (LSE - POLIS, 2024). Another example is SmartNews42, an app that also leverages AI for personalised curation of content from multiple sources based on user behaviour and preferences (Kumar, 2024). Yet, despite this potential, a recent outlook survey found that adoption of AI among surveyed publishers 'appears strongest in newsroom operations, and weakest in areas of monetisation and audience engagement' (WAN-IFRA, 2026).

However, the literature also warns of the risks of depending on Big Tech products for leveraging the potential of these technologies. Previous research, such as by Sørensen et al. (2020) has found that the use of third party services by European media websites to analyse user behaviour resembled the use that commercial players made of these services (Van den Bulck et al., 2025).

Archive analysis and speeding up non-creative tasks

Other potential uses relate to the ability of some AI tools to 'unlock the hidden potential in historical content, drawing out contextual meaning and capitalising on this valuable source of information', or being able to streamline repetitive non-creative tasks that free time and skills for other important aspects like audience engagement or research (LSE - POLIS, 2024).

Costs, time and knowledge are a barrier

Technical training and capacity-building for journalists and organisations was also highlighted in the literature. Thomson et al. (2024) underscore the importance of media organisations and newsrooms in supporting staff in professional development and training to understand the possibilities and pitfalls of Generative AI tools, and how algorithms work. Cools and Diakopuolos (2024) echo this, emphasising the need for 'algorithmic literacy' as well as the need for monitoring and assessments to ensure the responsible use of these tools in journalism.

Data and AI governance frameworks in newsrooms

Integrating AI systems and tools in the newsroom requires establishing robust governance and processes. As found by other authors, economic sustainability and efficiency need to go hand in hand with protecting the quality and fundamental values of journalism, and these fundamental values 'need to be reinterpreted in the context of AI' (Sonni, 2025). As Sonni adds: 'Objectivity, for example, gains a new dimension when algorithms become part of the editorial decision-making process. Transparency is no longer just about the source of information but also about the algorithmic processes that influence news production and distribution' (Sonni, 2025).

For example, a 2025 study in the Greek ecosystem found that 'journalists engage with AI tools both professionally and personally, often without organizational guidance or formal training' which can compromise the quality of journalism (Triantafyllou et al., 2025). However, the reality is

mixed and nuanced as research by de-Lima-Santos et al. (2025) found that all organisations in their study shared a focus on keeping their guidelines up-to-date as technological developments related to AI advance (de-Lima-Santos et al., 2023).

There are several examples of such governance frameworks cited in the literature. The Bavarian broadcaster in Germany, Bayerischer Rundfunk (BR), for example, has devised its own AI ethics guidelines and established the BR AI + Automation Lab and the BR Data lab, which integrates a multidisciplinary team of professionals – including data journalists, machine learning experts, designers, and product developers. BR's AI code of ethics is based on ten core guidelines, which include operating transparently, pursuing '…responsible personalization that does not undermine societal diversity and that avoids the creation of unintended filter bubbles', and integrating 'regular and interdisciplinary ethical reflection—involving journalists, developers, and managers—before investing resources in a project that will entail the use of AI' (Pérez-Seijo & Nuno Vicente, 2024).

The Swiss public media company SRG SSR has also developed guidelines that include transparency and human control over outputs with editorial curation of the use of algorithms to avoid biases and filter bubbles. Their charter on the usage of AI in editorial content production outlines that the company must adopt a human centric approach: 'its goal is to enhance the capabilities of professionals, enabling them to work more efficiently so that they can dedicate more time to high-value-added tasks' (Pérez-Seijo & Nuno Vicente, 2024).

Sweden's independent public service radio broadcaster SR established a company-wide AI Council tasked with three functions: 'first, identifying which AI applications hold the highest strategic value for the radio broadcaster and its audience; second, serving as a reference point for journalistic, legal, and security issues related to AI advancements and proposing guidelines when necessary; and third, initiating and participating in organizationwide learning for all personnel'. SR also established guidelines for the use of Generative AI including the retaining of human editorial control and review, transparency and labelling, and awareness of ethical and safety concerns, such as bias and transfer of sensitive information (Pérez-Seijo S and Nuno Vicente P, 2024).

5.9 Literacy, skills and social norms

Equipping both audiences and media professionals with the right skills as well as establishing social norms on the importance of information integrity in the context of Generative AI, digital public infrastructure and their role in democracy is another key theme in the literature.

Equipping audiences to critically evaluate information

Equipping audiences with information integrity literacy and establishing social norms around the use of Generative AI in the media and its impact on democracy and collective knowledge, is identified in the literature as a way to mitigate these concerns.

At the audience level, heightening news literacy and activating interest in high-quality news can help to increase the demand for accurate, diverse and independently produced news, as well as the diversity and depth of news consumption (Schneiders & Stark, 2025). Pérez-Seijo and Nuno Vicente (2024) put forward that 'the challenge of media literacy arises to advance the use and consumption of information provided by AI in a democratic way' and that this requires access and understanding of these technologies by everyone. Similarly, Kumar (2024) highlights the importance of media and AI literacy for the public to engage critically with news.

Literacy becomes particularly necessary as AI summaries remove context, nuance and source attribution, or incorrectly attribute sources due to inherent hallucination errors, as developed in Section 4.5. A 2025 study found that tracing and evaluating the source and quality of information that AI chatbots provide is in itself challenging, and the authors call for 'enhanced media literacy' not only for citizens but also for media professionals (Brantner et al., 2025).

This extends to AI literacy training for policymakers and educators. To achieve this, experts suggest standardised AI and information literacy conceptual frameworks and methodologies, which could draw on government, private sector, or civil society partnerships to promote training (Mansell et al., 2025).

The BBC and the EBU suggest that although Public Service Media have a key role in providing ongoing AI literacy skills and support to citizens,

AI companies themselves should be doing more to educate users on navigating the benefits and limitations of AI use (BBC & EBU, 2025).

Equipping media to critically use and manage data and AI

As the literature points out, the need for literacy and skills applies to journalists and media professionals more broadly.

One of the key impacts of Generative AI and other types of AI systems is that it is changing workflows in the newsroom as well as in how to implement editorial oversight: 'The growing dependence on automation in news production and distribution has a profound impact on editorial independence as well as on the organizational and business choices of media organizations' (Dutkiewicz et al., 2025).

The bypass effect covered in Section 4.5 also results in less oversight from what would have been editorial institutional gatekeepers in both newsrooms and the public sphere. In this regard, Dutkiewicz et al. (2025) point out that the new relationships between media, AI and related platforms are changing power structures and how public opinion is formed, with big companies concentrating more control over agenda-setting and public opinion. Media professionals need to therefore be attuned to these new dynamics and skilled in understanding how to uphold public interest media principles.

> Poor understanding of GenAI risks a combination of inappropriate use, automation bias, algorithmic aversion and poor organisational decision-making, alongside uncritical reporting. It also hinders the public's ability to navigate the information ecosystem. (Jones, 2023)

Others argue that media professionals also need new skills such as replacing Google-SEO (Search Engine Optimisation) and older social media platforms skills with new optimisation techniques such as Generative Engine Optimization (GEO) or AI SEO (Newman, 2026).

Overall, the use of new AI-driven technologies and tools requires the skills to understand how to adequately integrate them into journalistic workflows, which technologists can struggle to do without an understanding of public interest media principles and newsroom operations (Gutierrez Lopez et al., 2023).

In addition, a 2025 report by the Public Service Alliance found that unofficial use of AI in newsrooms was seen by managers and specialists as a risk to data privacy and security (Wright & Porter, 2025).

Data stewardship is another key area where media professionals require upskilling, particularly as more initiatives to create data collaboratives and media data spaces are emerging alongside a growing consensus to pool data and resources as part of new collective sustainability strategies, as covered in Section 5.6.

Yet, small newsrooms have less time to experiment (Triantafyllou et al., 2025). Supporting smaller media is raised in the literature as particularly important, as they risk being left out of conversations and access to resources because they have less time to engage in critical conversations, as well as because larger public media outlets with greater resources lack an understanding of the challenges faced by smaller outlets. This means that media organisations 'with different, and potentially more challenging, perspectives risk becoming marginalised within PSMs' collective conversations about responsible AI (Wright & Porter, 2025).

Establishing institutional and social norms

In addition to enhanced data, AI and media literacy, and governance frameworks for newsrooms, it is important in such a shifting landscape to reestablish or reinforce the social norms for public interest media principles for society at large.

This can be done through regulatory signallying, for example by funding and building alternative digital public infrastructure, but also by media themselves. As previously mentioned, the effects and consequences of use of journalistic content in AI tools may be in part mediated by the strength of the publisher's relationship with their audience. Publishers could therefore further establish norms of community and interaction with their audiences (Brown & Jaźwińska, 2025).

The case of Wikipedia, which has been identified as a significant source of data for LLM-generated content and summaries, is a good example of collective social norms-setting. Given its relative importance in downstream applications such as LLMs, there remains a need to continue to improve the encyclopedia with accurate and comprehensive information. This requires the establishment of both social norms and information literacy: social norms and initiatives are needed to ensure

that there is a strong Wikipedia editor and contributor community, in particular for underrepresented groups and languages. In addition, information literacy is important to equip users with the critical thinking skills necessary to evaluate AI-generated content and understand sources and their role in the evolving information landscape (Vetter et al., 2025).

The responsibility of Generative AI providers

Other authors highlight the reciprocal relationship between LLM providers and information sources, and the need for norms and literacy to be established not only by audiences, but also by LLM developers:

> I hope that it helps the teams deploying these bots understand that indiscriminately swarming collections has consequences. Anyone building datasets to train AI models has an interest in these collections staying online and continuing to grow. The organizations hosting these collections want them to stay online and grow as well! We need to find a way that everyone can operate sustainably. (Michael Weinberg, quoted in Hinchliffe, 2025)

Both public opinion as well as actions from media organisations and AI companies can influence other companies in following the commonly accepted standards for a healthy information ecosystem and democracy. Given that the AI sector is led by some of the largest companies in the world, they are often in a position to dictate the terms of their partnerships, and this relative market power can be influential when it is used to ensure that AI is used in the public interest (Cihon et al., 2021).

Journalists and news managers should also educate themselves on not only the use of AI in news production, but also remain vigilant and acknowledge what the economic and market incentives as a result of AI technologies could mean for both media pluralism and sustainability: '...the owners of news organizations have a particular responsibility in preserving and promoting ownership cultures that promote media freedom and pluralism, as they search for sustainable avenues to operate within the AI economy' (Sjøvaag, 2024).

6. Conclusions: new visions and a holistic approach

This review of academic and grey literature illustrates the complexities and multi-layered issues related to the relationship between Generative AI, the media ecosystem, and the integrity of information and democracy. From the technicalities of information extraction by AI crawlers and scrapers, the displacement of audiences to AI-summaries or AI multimodal tabs and the loss of public interest media principles in the process, to the inadequacy of legal frameworks, the differences in leveraging power between large and small, local or investigative outlets and, at the more macro level, the implications this has for the health of democracies and cultures, the challenges are multiple and interconnected, and so are the potential solutions.

This review has been organised to cover potential pathways forward identified in the literature, some operating at the micro level of newsrooms, but most relating to structural approaches to address power asymmetries. Specifically, this review has covered:

- technical solutions for gatekeeping content extraction or content summarisation;
- licensing arrangements – from individual, to collective to statutory;
- law and policy approaches, including copyright reform, competition policy, platform regulation, and levy proposals;
- approaches to new business models, namely focused on audience loyalty and diversification;
- public and philanthropic investment;
- commons-based approaches to pooling and governing media data, such as data collaboratives; and coalitions and collective action approaches;
- the use of and governance of AI tools in newsrooms;

- the need for media, data and AI literacy and to reinforce and establish institutional and social norms.

While the report frames the findings from this literature review through a diagnosis-solutions binary framework, these potential solutions are entangled between what they propose and the problems they try to address. Some of the potential ways forward, such as bilateral licensing deals, can widen the gap between large publishers and smaller independent media outlets or can amplify the legitimacy of large-scale scraping by large Generative AI players, rather than addressing the need to foster public interest media principles and epistemic integrity in AI and digital platforms.

Within this complexity, however, some key intersecting and overarching themes become clear:

- The current changes driven by Generative AI technologies both reflect and result in not only technical or economic puzzles, but also deeper and structural power asymmetries that are rewiring how information is produced, interpreted, distributed and used, bypassing democratic mechanisms for shared knowledge, contestation and accountability.

- These structural changes have taken the sector to a critical juncture that technical fixes or media literacy alone cannot address. While technical solutions, literacy and governance frameworks at the newsroom level are part of the toolkit, the literature points to the need for holistic solutions that strengthen the collective and bargaining power of media. This involves a multiplicity of solutions, including decisive regulatory approaches and the creation of and investment in alternative public digital infrastructure.

- Holistic solutions require collective action and collaborative ecosystems so that the media ecosystem can regain leveraging power vis-a-vis the industry and market behind generative AI, without leaving behind local, niche and investigative outlets from different languages and contexts.

- The literature points towards the need for new business models for media, particularly through diversification and a modular approach to business strategies, as well as by strengthening audience loyalty through community-focused spaces and services. However, it is

limited in examples of business development strategies that have been tested or scaled for different types of media organisations. There is a need to identify and foster approaches that work and can scale for different types of media and contexts.

- The literature highlights deeper engagement with audiences both as part of new business model strategies and as part of a more democratic governance of media, aligning tensions between public interest and the trends towards personalisation and reduced agency that are part of AI system architectures.

- Although under threat, public interest journalism and media are arguably more important than ever to uphold information integrity, the visibility of a diversity of realities and experiences, and as a mechanism of accountability for healthy democracies.

- Investment – both public and philanthropic – and regulation need to adapt to this state of complexity with multilayered ways forward. Particular attention should be paid to solutions that address the structural issues at the meso and macro levels, enabling spaces for horizontal collaboration, and, at the same time, protecting and nurturing smaller and independent media organisations from diverse languages and contexts.

Ultimately, the role of Generative AI in the media and information sector is marked by large power asymmetries and threats to the principles of journalism, public interest media and their role in democracy. In this critical juncture, for the media sector to gain bargaining power and ensure its sustainability and the integrity of the knowledge ecosystem, a collaborative approach and a multiplicity of actions able to tackle structural challenges, from the micro to the macro level, will be essential.

Appendix: additional details on the methodology

The systematic strand of this literature review was conducted on Google Scholar using seven different boolean string searches. The search strings and the final number of results from each search are summarised in the table below. As a validation step, a keyword sanity check was conducted by selecting ten articles identified as highly relevant through non-systematic literature searches and analysing the ten most frequently occurring words or phrases in the title and abstract of those texts, excluding stop words, and corroborating that these were adequately represented in the boolean searches.

Search String	Final number of results included
(news OR "news outlets" OR outlets OR media) AND (ownership OR aggregators) AND (AI OR GenAI OR GPAI)	23
"public interest" AND (news OR "news outlets" OR outlets OR media) AND (ownership OR aggregators) AND (AI OR GenAI OR GPAI)	32
"media independence" AND (GPAI OR AI OR GenAI OR "Big Tech")	17
media AND (IP OR copyright) AND (GPAI OR AI OR GenAI OR "Big Tech")	28
"media business models" AND (GPAI OR AI OR GenAI OR "Big Tech")	16
"media sustainability" AND (GPAI OR AI OR GenAI OR "Big Tech")	12
"media integrity" AND (GPAI OR AI OR GenAI OR "Big Tech")	12

The search strategy, exclusion and inclusion criteria, and review protocol were established before the systematic search took place. The primary search was conducted between 29 October and 5 November 2025, with a supplementary search conducted on 15 January 2026. For each search term, results were filtered using custom date ranges to include publications from January 2023 onwards. This time constraint was chosen due to the rapidly evolving and changing AI media landscape,

in which technologies have developed substantially in recent years, to avoid outdated or irrelevant information.

The first 100 results for each search were downloaded to a Zotero folder shared by the principal researchers. Publications that were not accessible due to broken links, or content that had been removed, for example, were excluded at this stage. This resulted in 693 articles being downloaded for screening.

Screening and Selection Criteria

The articles from the systematic search strand were screened in two phases. In the initial screening, duplicate articles were removed, and the remaining articles were included or excluded based only on the criteria of whether their content was relevant to the focus of this paper. This was achieved by reading the titles, abstracts and relevant sections of the text that contained key search terms to assess their relevance to the pre-determined research questions. Due to the wide-ranging nature of publications captured in Google Scholar searches, a large number of articles met the boolean search terms, but did not directly address the focus of this study. Following the first screening, 284 articles were identified as relevant in addressing the research questions in this study.

In the second screening, each of the 284 articles was assessed according to the following inclusion and exclusion criteria described in the table below.

Criteria	Include	Exclude
Publication Type	Journal articles, reports, white papers, 'grey literature' (e.g., NGO or think tank publications), newspaper articles, blog posts, book chapters, pre-prints, PhD dissertations.	Book reviews, Masters theses dissertations, promotional content, non-substantive web pages (such as homepages without relevant content), articles or published materials which cannot be accessed (either via Open Access or Institutional Access), publications with inadequate scholarly quality and analytical rigor, as determined through close reading of the text.
Geographic Scope	Worldwide, with a particular focus on Europe.	Studies exclusively focused on local issues outside of Europe and without clear generalizable insights.

Criteria	Include	Exclude
Language	Publications in English (relevant outputs in Spanish and Catalan were included as additional languages spoken by the authors).	Publications in other languages not spoken by the authors.
Topical Relevance	Publications that address at least one of the three research questions for this process.	Publications that do not substantively engage with the research questions, despite containing the search terms.
Time Period	Publications from January 2023.	Publications prior to January 2023.

The second screening process led to 91 articles being identified and selected for inclusion in the final review and coding process. The first and second screening phases were conducted by one of the principal researchers, with close collaboration and input from the other researchers where clarification or discussion regarding article inclusion was required.

References

Abiri, G. (2024). Generative AI as Digital Media. *Harv. J. Sports & Ent. L.*, *15*, 279. **https://arxiv.org/abs/2503.06523**

Adebiyi, O. I., & Adeusi, O. C. (2025). *Developing standardized approaches to copyright attribution and ownership for AI-assisted and AI-generated works in digital ecosystems*. **https://gsconlinepress.com/journals/gscarr/sites/default/files/GSCARR-2025-0164.pdf**

Alonso, E., & Lucchi, N. (2025). AI And Copyright. *European Intellectual Property Review*, *47*(9), 515–526. **https://openaccess.city.ac.uk/id/eprint/35488/**

ALT Advisory. (2025). *Tech and Journalism in the Global South*. International Fund for Public Interest Media; Gordon Institute of Business Science. **https://content.gibs.co.za/cmscontent/media/pm1huamm/251119-ctrl_j_africa-tech_and_journalism_in_the_global_south_v3.pdf**

Amon, A., Yin, Z., Wang, Z., Palikhe, A., & Zhang, W. (2025). *Uncertain Boundaries: Multidisciplinary Approaches to Copyright Issues in Generative AI* (arXiv:2404.08221). arXiv. **https://doi.org/10.48550/arXiv.2404.08221**

Atkin, C. (2026). *Ads for News, News for Ads. The role of government in reforming the digital advertising market* [Policy Brief]. Forum on Information and Democracy. **https://checkmyads.org/policy-brief/check-my-ads-and-fid-partner-on-a-policy-brief/**

Balasingham, B., Kozak, M., & Ruiz Palacios, T. A. (2025). Fair P(l)ay in the Digital Arena – In Search of a Balanced Relationship Between Press Publishers and Digital News Aggregators. *IIC - International Review of Intellectual Property and Competition Law*, *56*(6), 1092–1116. **https://doi.org/10.1007/s40319-025-01594-5**

Banaji, S. (2025, January 17). *Totalitarian tech? Billionaires, hate and the undermining of social media integrity* [Online resource]. London School of Economics and Political Science. Media@LSE. **https://blogs.lse.ac.uk/medialse/**

Barnett, J. (2025). The Free Content Illusion. *SSRN Electronic Journal*. **https://doi.org/10.2139/ssrn.5414819**

Barr, Alistair. (2025, September 9). Inside the battle for the future of the web. *Business Insider*. **https://www.businessinsider.com/google-microsoft-openai-fight-standards-limit-ai-access-websites-2025-9?op=1**

BBC, & EBU. (2025). *News Integrity in AI Assistants. An international PSM study*. **https://www.ebu.ch/Report/MIS-BBC/NI_AI_2025.pdf**

Bearne, S. (2025, September 9). Publishers fear AI summaries are hitting online traffic. *BBC News*. **https://www.bbc.com/news/articles/c0mlvryx0exo**

Beckett, C., & Yaseen, M. (2023). *Generating change: A global survey of what news organisations are doing with AI*. **https://eprints.lse.ac.uk/129643/**

Benson, R. (2018). Can foundations solve the journalism crisis? *Journalism*, *19*(8), 1059–1077. **https://doi.org/10.1177/1464884917724612**

Bleyer-Simon, K., Da Costa Leite Borges, D., Brogi, E., Carlini, R., Kermer, J. E., Palmer, M., Parcu, P. L., Reviglio Della Venaria, U., Trevisan, M., & Verza, S. (2024). *Monitoring media pluralism in the digital era: Application of the media pluralism monitor in the European member states and in candidate countries in 2023*. European University Institute. **https://cadmus.eui.eu/bitstream/handle/1814/77028/Media%20Pluralism%20Monitor%20Report_2024.pdf?sequence=3**

Boyles, J. L. (2024). A New(s) Copyright Balancing Act: How American Journalism Institutions Approached the Early Era of Artificial Intelligence and Fair Use. *Journalism Studies*, *25*(15), 1893–1910. **https://doi.org/10.1080/1461670X.2024.2403522**

Böyük, M. (2025). Techno-Polarization on the East-West Axis in AI-Based Algorithmic Communication. *Insight Turkey*, *27*(2), 365–388. **https://www.jstor.org/stable/48829619**

Brantner, C., Karlsson, M., & Kuai, J. (2025). Sourcing behavior and the role of news media in AI-powered search engines in the digital media ecosystem: Comparing political news retrieval across five languages. *Telecommunications Policy*, *49*(5) **https://doi.org/10.1016/j.telpol.2025.102952**

Brevini, B. (2023). Global Digital Lords and Privatisation of Media Policy: The Australian Media Bargaining Code. *Javnost - The Public*, *30*(2), 268–283. **https://doi.org/10.1080/13183222.2023.2207427**

Brogi, E., & Sjøvaag, H. (2024). *Good practices for sustainable news media financing*. Council of Europe. **https://edoc.coe.int/en/media/11879-good-practices-for-sustainable-news-media-financing.html**

Brown, Peter D. & Jaźwińska, Klaudia. (2025). *Journalism Zero: How Platforms and Publishers are Navigating AI* [Tow Center for Digital Journalism]. Columbia University. **https://doi.org/10.7916/JKV1-F205**

Burger, J., Künzler, M., & Autenrieth, U. (2024). Innovations in local media (funding) in Switzerland. In M. Young, A. Hermida, & C. Castenada (Eds), *Novel Directions in Media Innovation and Funding*. Global Journalism Innovation Lab. **https://open.library.ubc.ca/soa/cIRcle/collections/facultyresearchandpublications/52383/items/1.0440952**

Campos-Rueda, M., & Goyanes, M. (2023). Public service media for better democracies: Testing the role of perceptual and structural variables in shaping citizens' evaluations of public television. *Journalism*, *24*(11), 2493–2513. **https://doi.org/10.1177/14648849221114948**

Carlini, R. (2025). Market power and opinion power. Assessing the risks for media pluralism and editorial independence deriving from ownership's concentration. *Rivista Italiana Di Informatica e Diritto*, *7*(1), 12–12. **https://cadmus.eui.eu/entities/publication/71cb2671-8a02-4e0c-98ef-2800a70f71dd**

Carlini, R., Menéndez, N., & Schiffrin, A. (2026). *IPD Working Paper: How to Update EU and US Copyright Regimes in the Age of AI*. Initiative for Policy Dialogue (IPD). **https://cadmus.eui.eu/server/api/core/bitstreams/dc36decc-acd5-433a-9d27-d933404a9661/content**

Chandrasekhar, R. (2025). *Legal frictions for data openness: Reflections from a case-study on re-use of the open web for AI training* (p. 75). Centre Internet et Société - CNRS; Inno3; Open Knowledge Foundation. **https://hal.science/hal-05009616**

Chapekis, A., & Lieb, A. (2025, July 22). Google users are less likely to click on links when an AI summary appears in the results. *Pew Research Center*. **https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/**

Cihon, P., Schuett, J., & Baum, S. D. (2021). Corporate Governance of Artificial Intelligence in the Public Interest. *Information*, *12*(7), 275. **https://doi.org/10.3390/info12070275**

Civitates. (2025). *Beyond The Headlines 2025*. Civitates. **https://civitates-eu.org/wp-content/uploads/2025/09/Beyond-The-Headlines-2025.pdf**

Cloudflare Radar (2026, May). AI Insights; **https://radar.cloudflare.com/ai-insights?dateRange=52w**

Colangelo, G. (2026). News Publishers, Digital Platforms, and Bargaining Codes: Debunking the Free-Riding Myth. *GRUR International*, *75*(1), 19–29. **https://doi.org/10.1093/grurint/ikaf130**

Colussi, J., de Souza Paes, P., & Gomes-Franco e Silva, F. (2025). Journalism and Economic Sustainability: An Analysis of Google News Showcase in Argentina, Brazil and Colombia. *Journalism Practice*, *0*(0), 1–17. **https://doi.org/10.1080/17512786.2025.2463472**

Cook, C. E. (2023). Assessing conditions for inter-firm collaboration as a revenue strategy for politically pressured news media. *Journal of Media Business Studies*, *20*(1), 52–71. **https://doi.org/10.1080/16522354.2021.2002106**

Cools, H., & Diakopoulos, N. (2024). Uses of Generative AI in the Newsroom: Mapping Journalists' Perceptions of Perils and Possibilities. *Journalism Practice*, 1–19. **https://doi.org/10.1080/17512786.2024.2394558**

Council of Europe (2026, April). Parliamentary Assembly Resolution on Copyright enforcement in the artificial intelligence environment. **https://pace.coe.int/en/files/35917/html**

Darling, C. (2024). News startups and business model innovation. In M. Young, A. Hermida, & C. Castenada (Eds), *Novel Directions in Media Innovation and Funding*. Global Journalism Innovation Lab. **https://open.library.ubc.ca/soa/cIRcle/collections/facultyresearchandpublications/52383/items/1.0440952**

D'Arma, A., Michalis, M., Lowe, G. F., & Zita, M.-B. (2025). Public Service Journalism in Digital Markets. In *Challenges and Developments in Public Service Journalism*. University of Westminster Press. **https://www.jstor.org/stable/10.16997/14610450.4**

de-Lima-Santos, M.-F., Munoriyarwa, A., Elega, A. A., & Papaevangelou, C. (2023). Google news initiative's influence on technological media innovation in Africa and the Middle East. *Media and Communication*, *11*(2), 330–343. **https://www.cogitatiopress.com/mediaandcommunication/article/view/6400**

de-Lima-Santos, M.-F., Yeung, W. N., & Dodds, T. (2025). Guiding the way: A comprehensive examination of AI guidelines in global media. *AI & SOCIETY*, *40*(4), 2585–2603. **https://doi.org/10.1007/s00146-024-01973-5**

Diab, R. (2025, October 9). *Why Lawsuits Over AI Summaries Will Fail: There is No Right to Traffic | TechPolicy.Press*. Tech Policy Press. **https://techpolicy.press/why-lawsuits-over-ai-summaries-will-fail**

Díaz-Noci, J., Peña-Fernández, S., Meso-Ayerdi, K., & Larrondo-Ureta, A. (2024). The influence of AI in the media work force: How companies use an array of legal remedies. *Tripodos*, (55), 33–54. **https://tripodos.com/index.php/Facultat_Comunicacio_Blanquerna/article/view/1113**

Ding, Z., Smith-Renner, A., Zhang, W., Tetreault, J. R., & Jaimes, A. (2023). *Harnessing the Power of LLMs: Evaluating Human-AI Text Co-Creation through the Lens of News Headline Generation* (arXiv:2310.10706). arXiv. **https://doi.org/10.48550/arXiv.2310.10706**

Directa (2026, March 12). Neix CoopMèdia, la primera xarxa de mitjans cooperatius catalans. *directa.cat*. **https://directa.cat/neix-coopmedia-la-primera-xarxa-de-mitjans-cooperatius-catalans**

Dobrescu, P., & Durach, F. (2025). The Reconfiguration of the Public Sphere. In P. Dobrescu & F. Durach, *Competing Visions* (pp. 117–162). Springer Nature Switzerland. **https://doi.org/10.1007/978-3-031-74685-7_4**

Dornis, T. W., & Lucchi, N. (2025). *Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of C-250/25*. **https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5391439**

Doyle, G., Barr, K., & Boyle, R. (2025). Public service media as critical media infrastructure for the digital era. *Media, Culture & Society*, *47*(6), 1132–1149. **https://doi.org/10.1177/01634437251330119**

Dragomir, M., & Nemeth, R. (2026). Buying time, not markets: The limits of donor funding in Hungary's captured media system. *Journalism*, 14648849261416556. **https://doi.org/10.1177/14648849261416556**

Dragomir, M., Rodríguez Castro, M., & Aslama Horowitz, M. (2024). Public Service Media and Platformization: What Role Does EU Regulation Play? *Journalism and Media*, *5*(3), 1378–1394. **https://doi.org/10.3390/journalmedia5030087**

Dutkiewicz, L., Krack, N., Kuczerawy, A., Valcke, P., & Smuha, N. (2025). *Artificial Intelligence and Media*. **https://lirias.kuleuven.be/retrieve/807625**

Eeman, E. (2026, January). *WayFinder. Review Preview—Trends Report 25-26*. **https://www.linkedin.com/posts/ezra-eeman-8a5ba64_wayfinder-25-26-reviewpreview-trends-report-activity-7413868907855400960-UGEI/?rcm=ACoAAAAoIXQBPaa1bG37g3aRLIALer8AXchuKm4**

European Commission. (2019). *Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market*. Official Journal of the European Union. **https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng**

European Parliament. (2025). *Draft report on Copyright and generative artificial intelligence – opportunities and challenges* (2025/2058(INI)). European Parliament. **https://www.europarl.europa.eu/doceo/document/A-10-2026-0019_EN.html**

European Parliament. (2026). *Report on copyright and generative artificial intelligence – opportunities and challenges* (Nos A10-0019/2026). European Parliament. **https://www.europarl.europa.eu/doceo/document/A-10-2026-0019_EN.pdf**

European Publishers Council. (2026). *Our AI Complaint*. **https://www.epceurope.eu/our-ai-competition-complaint**

European University Institute. (2025). *Monitoring Media Pluralism in the European Union: Results of the MPM2025.* Publications Office. **https://doi.org/10.2870/2415644**

Exner, M. (2025). *How can journalism survive AI?* **https://www.publix.de/en/news/how-can-journalism-survive-ai**

Flew, T. (2026). News, Consumers and Citizens. In T. Flew, A. Stepnik, & T. Koskie (Eds), *Valuing News* (pp. 23–46). Springer Nature Singapore. **https://doi.org/10.1007/978-981-95-2341-2_2**

Flew, T., Iosifidis, P., Meese, J., & Stepnik, A. (2024). Digital platforms and the future of news: Regulating publisher-platform relations in Australia and Canada. *Information, Communication & Society*, *27*(16), 2727–2743. **https://doi.org/10.1080/1369118X.2023.2291462**

Georgescu, A., & Hausner, R. (2026, February). AgoraEU 2028–2034 Funding for culture: media and EU values in the new multiannual financial framework. Briefing EPRS | European Parliamentary Research Service. **https://www.europarl.europa.eu/RegData/etudes/BRIE/2026/782660/EPRS_BRI(2026)782660_EN.pdf#:~:text=The%20Commission%20uses%20current%20prices%20in%20its,civil%20society%20in%20current%20prices%20has%20doubled**

Goudarzi, S. (2024). To protect democratic values, journalism must save itself. *Bulletin of the Atomic Scientists*, *80*(5), 314–320. **https://doi.org/10.1080/00963402.2024.2388468**

Gutierrez Lopez, M., Porlezza, C., Cooper, G., Makri, S., MacFarlane, A., & Missaoui, S. (2023). A Question of Design: Strategies for Embedding AI-Driven Tools into Journalistic Work Routines. *Digital Journalism*, *11*(3), 484–503. **https://doi.org/10.1080/21670811.2022.2043759**

Hardinges, J. (2025, November). *Pay-To-Crawl*. Creative Commons. **https://creativecommons.org/wp-content/uploads/2025/11/Pay-to-Crawl-Issue-Brief-Nov-2025.pdf**

Hardinges, Jack, Pearson, Sarah, & Ross, Rebeca. (2025, June). *From Human Content to Machine Data Introducing CC Signals*. **https://creativecommons.org/wp-content/uploads/2025/06/Human-Content-to-Machine-Data_Final.pdf**

Hinchliffe, L. J. (2025, June 23). Are AI Bots Knocking Digital Collections Offline? An Interview with Michael Weinberg. *The Scholarly Kitchen*. **https://scholarlykitchen.sspnet.org/2025/06/23/are-ai-bots-knocking-digital-collections-offline/**

Hingle, A., & Knodel, M. (2025, April 3). Robots.txt Is Having a Moment: Here's Why We Should Care. *Tech Policy Press*. **https://techpolicy.press/robotstxt-is-having-a-moment-heres-why-we-should-care**

Hoffmann, J. (2025). *Technological Determination of AI-Relevant Press and Copyright Law and Generative Content's Relevance for EU Competition Law- The referral in Case C-250/25, Like Company v. Google Ireland Ltd.* **https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5411443**

Holder, P., Mateen, H., Anya, S., & Tabakovic, H. (2024). *Paying for News: What Google and Meta Owe Us Publishers*. **https://www.ssrn.com/abstract=4704237**

Hyun, D., Park, S., & Kim, E. (2025). Towards a New Copyright Paradigm for Generative AI: Bridging Human-Machine Creativity Through Legal and Policy Reform. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4487–4494. **https://openaccess.thecvf.com/content/ICCV2025W/WCCA/html/Hyun_Towards_a_New_Copyright_Paradigm_for_Generative_AI_Bridging_Human-Machine_ICCVW_2025_paper.html**

I, T., & Cath, C. (2025, October 31). Creepy AI Crawlers Are Turning the Internet into a Haunted House | TechPolicy.Press. *Tech Policy Press*. **https://techpolicy.press/creepy-ai-crawlers-are-turning-the-internet-into-a-haunted-house**

Islam, G., & Greenwood, M. (2024). Generative Artificial Intelligence as Hypercommons: Ethics of Authorship and Ownership. *Journal of Business Ethics*, *192*(4), 659–663. **https://doi.org/10.1007/s10551-024-05741-9**

Jarvis, J. (2025). APIs for news. *NiemanLab*. **https://www.niemanlab.org/2025/12/apis-for-news/**

Jay, H. (2025). *Public Service Media in an Age of Personalisation A Critical Research Agenda*. Responsible Innovation Centre. **https://downloads.bbc.co.uk/rd/responsible-innovation/responsible-innovation-centre-personalisation-public-interest.pdf**

Jones, B., Strait, A., Parnell, B.-Á., Horzyk, A., & Perez, J. (2025). *Journalism and generative AI: Data, deals and disruption in the news media*. Zenodo. **https://doi.org/10.5281/zenodo.14968194**

Jones, K. (2023). *AI governance and human rights: Resetting the relationship*. Royal Institute of International Affairs. **https://doi.org/10.55317/9781784135492**

Kanza, Y., Krishnamurthy, B., & Srivastava, D. (2024). A Geospatial Perspective on Data Ownership, the Right to be Forgotten, Copyrights, and Plagiarism in Generative AI. *Proceedings of the 32nd ACM International Conference*

on Advances in Geographic Information Systems, 477–480. **https://doi.org/10.1145/3678717.3691269**

Keller, P. (2025). *Beyond AI & Copyright*. **https://openfuture.eu/publication/beyond-ai-and-copyright/**

Keller, P. (2026). Copyright, AI and the limits of voluntary licensing. *Open Future Blog*. **https://openfuture.eu/blog/copyright-ai-and-the-limits-of-voluntary-licensing/**

Khanal, S., Zhang, H., & Taeihagh, A. (2025). Why and how is the power of Big Tech increasing in the policy process? The case of generative AI. *Policy and Society*, *44*(1), 52–69. **https://ideas.repec.org/a/oup/polsoc/v44y2025i1p52-69..html**

Krishna, S., & Nottingham, M. (2025, February 27). IETF setting standards for AI preferences. *IETF News*. **https://www.ietf.org/blog/aipref-wg/**

Kuai, J. (2024). Unravelling Copyright Dilemma of AI-Generated News and Its Implications for the Institution of Journalism: The Cases of US, EU, and China. *New Media & Society*, *26*(9), 5150–5168. **https://doi.org/10.1177/14614448241251798**

Kumar, A. (2024, February 15). *Written Evidence to the House of Lords Communications and Digital Committee Inquiry on The Future of News: Impartiality, Trust, and Technology* [Report]. UK Parliament. **https://committees.parliament.uk/writtenevidence/128179/pdf/**

Lefèvre, B., Errando, A., Afilipoaie, A., Ranaivoson, H., & Wiart, L. (2025). Exploring ethical and regulatory challenges of AI integration in European newsrooms. *Medijske Studije*, *16*(31), 56–76. **https://researchportal.vub.be/en/publications/exploring-ethical-and-regulatory-challenges-of-ai-integration-in-/**

Lim, J. S., Shin, D., Zhang, J., Masiclat, S., Luttrell, R., & Kinsey, D. (2023). News Audiences in the Age of Artificial Intelligence: Perceptions and Behaviors of Optimizers, Mainstreamers, and Skeptics. *Journal of Broadcasting & Electronic Media*, *67*(3), 353–375. **https://doi.org/10.1080/08838151.2022.2162901**

Limelight Foundation. (2026). Be bold, back the brave. *Limelight Foundation*. **https://limelight.foundation/**

Longo, E. (2024). Grounding media freedom in the EU: The legal basis of the EMFA. *Rivista Italiana Di Informatica e Diritto*, *7*, 111–124. **https://flore.unifi.it/handle/2158/1436697**

LSE - POLIS. (2024). *AI and the newsroom next door. Experiments and best practices for small publishers* (2024 Innovation Challenge). **https://static1.**

squarespace.com/static/64d60527c01ae7106f2646e9/t/691217ceac084a12e45e84c7/1762793423014/2024+JournalismAI+Innovation+Challenge+Report.pdf

Lucchi, N. (2025). *Generative AI and Copyright: Training, Creation, Regulation* [Policy Department for Justice, Civil Liberties and Institutional Affairs Directorate-General for Citizens' Rights, Justice and Institutional Affairs]. European Parliament. https://www.europarl.europa.eu/thinktank/en/document/IUST_STU(2025)774095

Macpherson, L. (2025a, September 22). Is There a Middle Ground in the Tug of War Between News Publishers and AI Firms? Part 1: Framing the Problem. *Public Knowledge*. https://publicknowledge.org/is-there-a-middle-ground-in-the-tug-of-war-between-news-publishers-and-ai-firms-part-1-framing-the-problem/

Macpherson, L. (2025b, September 22). Is There a Middle Ground in the Tug of War Between News Publishers and AI Firms? Part 2: Framing Solutions. *Public Knowledge*. https://publicknowledge.org/is-there-a-middle-ground-in-the-tug-of-war-between-news-publishers-and-ai-firms-part-2-framing-solutions/

Mansell, R., Durach, F., Kettemann, M. C., Lenoir, T., Procter, R., Tripathi, G. P., & Tucker, E. (2025). *Information Ecosystems and Troubled Democracy: The State of Knowledge on News Media, AI, and Data Governance*. Nordicom, University of Gothenburg. https://doi.org/10.48335/9789189864238

Mensch, A. (2026, March 20). Mistral CEO: AI companies should pay a content levy in Europe. *Finacial Times*. https://www.ft.com/content/d63d6291-687f-4e05-8b23-4d545d78c64a#:~:text=Mistral%20CEO%3A%20AI%20companies%20should,holders%20and%20bring%20legal%20certainty

Meyer, T., Kerkhof, A., Cennamo, C., & Kretschmer, T. (2024). Competing for attention on digital platforms: The case of news outlets. *Strategic Management Journal*, *45*(9), 1731–1790. https://doi.org/10.1002/smj.3600

Molitorisz, S., & Attard, M. (2024). Is Australia's News Media Bargaining Code a Model for Saving Journalism? *Global Perspectives*, *5*(1), 126792. https://online.ucpress.edu/gp/article-abstract/5/1/126792/204717

Money-Kyrle, O. (2026, February). *Meeting the Challenges of a Changing Journalism Landscape through Funding, EU Regulation and Dialogue* [Policy Brief]. European Federation of Journalists. https://resilientmedia.eu/?p=2148

Nel, F., & Rymajdo, K. (2024). *News futures 2035 final report*. https://clok.uclan.ac.uk/id/eprint/52306/1/NF2035-Final-Report_Print-Out-Copy_WP.pdf

Newman, N. (2026). *Journalism and technology trends and predictions 2026*. Reuters Institute for the Study of Journalism. **https://doi.org/10.60625/RISJ-PS1D-NP11**

Newman, N., Ross Arguedas, A., Robertson, C. T., Nielsen, R. K., & Fletcher, R. (2025). *Digital news report 2025*. Reuters Institute for the study of Journalism. **https://ora.ox.ac.uk/objects/uuid:24de5b16-d5bb-40da-a55c-4e7c28ab6dff**

Nogueroles, E. (2025, October 24). *The Quiet Takeover: How AI Is Rewriting Newsrooms and Academics • Media Power Monitor*. **https://mpmonitor.org/the-quiet-takeover-how-ai-is-rewriting-newsrooms-and-academics/**

Oleart, A., & Rone, J. (2025). Reversing the Privatisation of the Public Sphere: Democratic Alternatives to the EU's Regulation of Disinformation. *Media and Communication*, *13*(0). **https://doi.org/10.17645/mac.9496**

Park, S., Fisher, C., & Picard, R. (2026). Funding the News: Current Business Models. In T. Flew, A. Stepnik, & T. Koskie (Eds), *Valuing News* (pp. 47–71). Springer Nature Singapore. **https://doi.org/10.1007/978-981-95-2341-2_3**

Park, S., Fulton, J., Fujita, M., McGuinnes, K., Nolan, D., Lee, J., Howden, S., Atkinson, S., Yao, P., & Kim, W. (2026). *Synthesis and Priorities. News Futures: Media Policy Roundtable 2026.* University of Canberra. **https://apo.org.au/sites/default/files/resource-files/2026-04/apo-nid333994.pdf**

Pattison, C., Ricks, V., & Wihbey, J. (2025, August 11). How AI-Driven Search May Reshape Democracy, Economics, and Human Agency. *Tech Policy Press*. **https://www.techpolicy.press/how-ai-driven-search-may-reshape-democracy-economics-and-human-agency/**

Pérez-Seijo, S., & Nuno Vicente, P. (2024). AI Governance in European Public Service Media: Policies and Guidelines for the Use of Artificial Intelligence Systems. In Martín Vaz Álvarez, José Miguel Túñez López, & Ana Gabriela C. Frazão Nogueira (Eds), *The Public Value Blueprint* (p. pp 169-185). Springer Cham. **https://link.springer.com/chapter/10.1007/978-3-031-60824-7_13**

Peukert, C. (2025). *The economics of copyright and AI: Empirical evidence and optimal policy* (PE 778.859). Policy Department for Justice, Civil Liberties and Institutional Affairs Directorate-General for Citizens' Rights, Justice and Institutional Affairs. **https://www.europarl.europa.eu/RegData/etudes/STUD/2025/778859/IUST_STU(2025)778859_EN.pdf**

Philea. (2026). *Journalism and Media Funding in Europe*. Journalism Funders Forum. **https://philea.eu/philea-publishes-journalism-and-media-funding-in-europe/**

Pion, C. (2025, October 16). Towards Digital Sovereignty: German Media Pilot Shared Data Systems. *Public Media Alliance*. **https://www.publicmediaalliance.org/german-media-test-the-foundations-of-a-shared-data-infrastructure/**

Quintais, J. P. (2025a). *Copyright, the AI Act and extraterritoriality* [Policy Brief]. The Lisbon Council. **https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5316132**

Quintais, J. P. (2025b). Generative AI, copyright and the AI Act. *Computer Law & Security Review*, *56*, 106107. **https://doi.org/10.1016/j.clsr.2025.106107**

Radsch, C. (2022). Making Big Tech Pay for the News They Use. *SSRN Electronic Journal*. **https://doi.org/10.2139/ssrn.4192026**

Radsch, C. (2024). *What is the Value of Journalism to AI?* [Expert Brief]. Center for Journalism & Liberty. **https://www.journalismliberty.org/publications/value-of-journalism-to-ai**

Radsch, C. (2025, January 17). *The Case for Requiring Explicit Consent from Rights Holders for AI Training* [TechPolicy.Press]. **https://www.techpolicy.press/the-case-for-requiring-explicit-consent-from-rights-holders-for-ai-training/**

Radsch, C. C., Montoya, K., & Pomeroy, E. (2025). *The Emerging AI Content Licensing Market*. **https://www.journalismliberty.org/publications/expert-brief-ai-and-market-concentration-hej73**

Rahman, A. (2025). Opportunities and Challenges of Artificial Intelligence in Public Media Journalism. In A. D'Arma, M. Michalis, G. Ferrell Lowe, & M.-B. Zita (Eds), *Challenges and Developments in Public Service Journalism*. University of Westminster Press. **https://www.jstor.org/stable/10.16997/14610450.9**

Reisner, A. (2025, November 4). *The Company Quietly Funneling Paywalled Articles to AI Developers*. The Atlantic. **https://www.theatlantic.com/technology/2025/11/common-crawl-ai-training-data/684567/**

Ross Arguedas, A., Mont'Alverne, C., Toff, B., Fletcher, R., & Nielsen, R. K. (2024). Ritual Reinforcement: Habit, Emotion, and Identity as Attributes of Trust in News. *Journalism Studies*, *25*(15), 1875–1892. **https://doi.org/10.1080/146167 0X.2024.2401403**

Sandrini, L., & Somogyi, R. (2023). Generative AI and deceptive news consumption. *Economics Letters*, *232*, 111317. **https://www.sciencedirect.com/science/article/pii/S0165176523003427**

Šarčević, T., Karlowicz, A., Mayer, R., Baeza-Yates, R., & Rauber, A. (2024). *U Can't Gen This? A Survey of Intellectual Property Protection Methods for Data in Generative AI* (arXiv:2406.15386). arXiv. **https://doi.org/10.48550/arXiv.2406.15386**

Schiffrin, A., Kostic, B., Nenadic, I., Haas, J., Helberger, N., & Van Drunen, M. Z. (2025). *Safeguarding Media Freedom in the Age of Big Tech Platforms and AI* (Healthy Online Information Spaces – SAIFE Renewed). Office of the Representative on Freedom of the Media Organization for Security and Co-operation in Europe (OSCE). **https://informationdemocracy.org/publications/osce-policy-manual%EF%BD%9Csafeguarding-media-freedom-in-the-age-of-big-tech-platforms-and-ai/**

Schirru, L., & Margoni, T. (2023, August 22). Arts 3 and 4 of the CDSM Directive as regulatory interfaces: Shaping contractual practices in the Commercial Scientific Publishing and Stock Images sectors. *Kluwer Copyright Blog*. **https://legalblogs.wolterskluwer.com/copyright-blog/arts-3-and-4-of-the-cdsm-directive-as-regulatory-interfaces-shaping-contractual-practices-in-the-commercial-scientific-publishing-and-stock-images-sectors/**

Schneider, P. (2025, March 25). The Third Way of Funding Independent Media [Blog]. *Medium*. **https://medium.com/@patrice.schneider/beyond-grants-and-market-returns-fe7fa131f058**

Schneiders, P., & Stark, B. (2025). Ensuring News Quality in Platformized News Ecosystems: Shortcomings and Recommendations for an Epistemic Governance. *Media and Communication*, *13*(0). **https://www.cogitatiopress.com/mediaandcommunication/article/view/10042**

Scott, M. (2025). *M20 Policy Brief 6: Fuelling democracy and growth through viable media – an urgent G20 call to action*. **https://media20.org/2025/07/16/policy-brief-6-fuelling-democracy-and-growth-through-viable-media-an-urgent-g20-call-to-action/**

Sevignani, S., Theine, H., & Tröger, M. (2025). Toward Media Environment Capture: A Theoretical Contribution on the Influence of Big Tech on News Media. *International Journal of Communication*, *19*, 21. **https://ijoc.org/index.php/ijoc/article/view/21987**

Simon, F. M., Nielsen, R. K., & Fletcher, R. (2025). *Generative AI and news report 2025: How people think about AI's role in journalism and society*. Reuters Institute for the Study of Journalism. **https://doi.org/10.60625/RISJ-5BJV-YT69**

Simon, J. (2023). *Press Freedom and Public Interest*. Tow Center for Digital Journalism, Columbia University. **https://doi.org/10.7916/Y3CG-8J77**

Sjøvaag, H. (2024). The business of news in the AI economy. *AI Magazine*, *45*(2), 246–255. **https://doi.org/10.1002/aaai.12172**

Sonni, A. F. (2025). Digital transformation in journalism: Mini review on the impact of AI on journalistic practices. *Frontiers in Communication*, *10*,

1535156. https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1535156/full

Stănescu, G. C. (2026). Artificial Intelligence and its Implications for Media Sustainability in Romania. In M. Todorovic, S. O. Idowu, & S. Puiu (Eds), *Sustainability and Social Responsibility of the Media and in the Media* (pp. 89–110). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-00086-6_5

Stiglitz, J., & Ventura-Bolet, M. (2025). *The Impact of AI and Digital Platforms on the Information Ecosystem* (No. W34318; p. w34318). National Bureau of Economic Research. https://doi.org/10.3386/w34318

Teilmann-Lock, S., & Savin, A. (2025). Beyond the AI-copyright wars: Towards European dataset law? *Computer Law & Security Review*, *58*, 106190. https://www.sciencedirect.com/science/article/pii/S2212473X25000628

TEMS. (2025). *Survey Results: Understanding Media's Data Needs Across Europe* [Survey]. https://tems-dataspace.eu/survey-results-understanding-medias-data-needs-across-europe/

TEMS. (2026). *Trusted European Media Data Space*. https://tems-dataspace.eu/

Thomass, B. (2024). Public service media and public service internet as a counterweight to media concentration. *European Journal of Communication*, *39*(5), 457–471. https://doi.org/10.1177/02673231241270972

Thurman, N., Thäsler-Kordonouri, S., & Fletcher, R. (2025). *AI adoption by UK journalists and their newsrooms: Surveying applications, approaches, and attitudes*. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/RISJ-EA11-Q402

Tobitt, C. (2026, March 4). Who's suing AI and who's signing: News Corp signs $50m-a-year deal with Meta, Reach and Amazon agree usage deal, Danish publishers sue OpenAI. *Press Gazette*. https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/

Tome, Joao. (2025, August). *The crawl-to-click gap: Cloudflare data on AI bots, training, and referrals* [The Cloudflare blog]. https://blog.cloudflare.com/crawlers-click-ai-bots-training/#google-referrals-fall-as-ai-overviews-expand

Tremayne-Pengelly, A. (2025, March 28). Wikipedia Built the Internet's Brain. Now Its Leaders Want Credit. *Observer*. https://observer.com/2025/03/wikimedia-foundation-execs-speak-on-ai-scraping-attribution-and-wikipedias-future/

Triantafyllou, S., Panagopoulos, A. M., & Kapos, P. (2025). AI Pioneers and Stragglers in Greece: Challenges, Gaps, and Opportunities for Journalists and Media. *Societies*, *15*(8), 209. **https://doi.org/10.3390/soc15080209**

Urban, A. (2026). Community Building as a Tool for Sustainability in Hungarian Digital Media. *Journalism and Media*, *7*(1), 48. **https://doi.org/10.3390/journalmedia7010048**

Van den Bulck, H., Kulig, M., Hyzen, A., Puppis, M., & Paulussen, S. (2025). Epistemic welfare and public service media's algorithmic recommender systems: A theoretical framework, operationalization and relevance for governance. *European Journal of Communication*, 1–21. **https://doi.org/10.1177/02673231251385761**

Verdegem, P. (2024). Dismantling AI capitalism: The commons as an alternative to the power concentration of Big Tech. *AI & SOCIETY*, *39*(2), 727–737. **https://doi.org/10.1007/s00146-022-01437-8**

Vetter, M. A., Jiang, J., & McDowell, Z. J. (2025). An endangered species: How LLMs threaten Wikipedia's sustainability. *AI & SOCIETY*, *40*(6), 4309–4322. **https://doi.org/10.1007/s00146-025-02199-9**

WAN-IFRA. (2026). *World Press Trends Outlook 2025-2026: The annual global survey, analysis and report on publishers' business and predictions*. World Association of News Publishers. **https://media.licdn.com/dms/document/media/v2/D561FAQH_23icGqyieA/feedshare-document[…]73878400&v=beta&t=nmIakjaDNPuI0dtY4xyF_4_e5fXrogB16U9acOk7nYQ**

Wasdahl, A., & Srinivasan, R. (2026). Algorithms and authors: How generative AI is transforming news production. *First Monday*. **https://doi.org/10.5210/fm.v31i1.14442**

Wright, K., & Porter, K. (2025). *PSM and AI. Mapping how public service media use AI in journalism*. Public Media Alliance. **https://www.publicmediaalliance.org/wp-content/uploads/2025/08/PSM-and-AI_PMA-Industry-Report.pdf**

Zhao, H., & Berman, R. (2026). *Strategic Response of News Publishers to Generative AI* (arXiv:2512.24968). arXiv. **https://doi.org/10.48550/arXiv.2512.24968**

The Data Tank

May 2026

Belgium

email: **comms@datatank.org**

web: **www.datatank.org**