

# Mapping Europe's AI Media Landscape: From Adoption to Governed Value Capture

May 2026

---

Dr August Bourgeus

Prof Dr Marlen Komorowski

imec-SMIT, Vrije Universiteit Brussel



©2026, imec-SMIT, Vrije Universiteit Brussel (and other partners)

All rights reserved. No part of this report may be reproduced, stored or made public by any means whatsoever, whether electronic or mechanical, without prior permission in writing from the authors.

# Table of Contents

<b>Table of Contents</b> .....	<b>3</b>
<i>Summary</i> .....	4
<b>Introduction</b> .....	<b>5</b>
<i>Approach</i> .....	5
<i>Background</i> .....	6
<b>Layer 1: Compute</b> .....	<b>7</b>
<i>What is the compute layer?</i> .....	7
<i>What is Europe’s position in the compute layer?</i> .....	7
<i>Case study: EuroHPC</i> .....	8
<i>Key Learnings</i> .....	10
<b>Layer 2: Models</b> .....	<b>11</b>
<i>What is the model layer?</i> .....	11
<i>What is Europe’s media sector’s position in the model layer?</i> .....	11
<i>Case study: GPT-NL (Netherlands)</i> .....	14
<i>Case study: NorLLM (Norway)</i> .....	15
<i>Case study: OpenGPT-X / Teuken &amp; Soofi (Germany)</i> .....	15
<i>Key Learnings</i> .....	16
<b>Layer 3: Applications</b> .....	<b>17</b>
<i>What is the application layer?</i> .....	17
<i>What is Europe’s media sector’s position in the application layer?</i> .....	17
<i>Case studies: Archive search and retrieval</i> .....	18
<i>Case studies: Newsroom productivity</i> .....	20
<i>Case studies: Audience-facing products</i> .....	21
<i>Key Learnings</i> .....	22
<b>Conclusion: From AI Adoption to Governed Value Capture</b> .....	<b>24</b>
<i>Routes for governance and value creation</i> .....	25

## Summary

This report maps Europe's media sector's position in the AI landscape across three layers: compute, models, and applications. It asks how media organisations are affected by AI, how they contribute to AI systems, and how they can capture value from them. The central finding is that media organisations are no longer passive users: their archives, content, editorial expertise, trusted brands, and user relationships are increasingly valuable for AI training, retrieval, and product development. Yet this does not automatically translate into control, revenue, or strategic autonomy. The key challenge is therefore to move from AI adoption to governed value capture.

At the **compute layer**, Europe has developed stronger public infrastructure through EuroHPC and AI Factory Antennas. These infrastructures support sovereign experimentation, dataset processing, model adaptation, and fine-tuning. However, public compute does not remove dependence on commercial cloud. Many media applications still require continuous deployment, integration, uptime, and low-latency inference, making hybrid compute environments likely.

At the **model layer**, Europe has produced several sovereign and language-oriented initiatives, but media organisations are only selectively involved. Most initiatives are led by research institutions or public AI programmes, with media actors treated as data providers or future users rather than co-owners. GPT-NL, NorLLM/Kopinor, and OpenGPT-X/Teuken are important exceptions. They show that media archives can become strategic and monetisable assets when governance, IP, licensing, and revenue-sharing are addressed from the start.

At the **application layer**, Europe's media sector is most active. The report identifies three main areas: archive search and retrieval, newsroom productivity, and audience-facing products. Here, media organisations can create value without training foundation models themselves, by combining existing models with trusted archives, metadata, editorial workflows, and human oversight. Value can take the form of productivity gains, better archive access, stronger user engagement, or commercial B2B services. However, many applications still depend on US cloud infrastructure or proprietary models.

Two cross-cutting themes emerge.

- First, local language and cultural specificity matter: generic models may underrepresent smaller European languages, regional references, and national media cultures.
- Second, AI adoption alone does not guarantee value capture. Value is captured by those who control infrastructure, data, models, interfaces, rights, and business arrangements.

The report identifies six routes for media organisations to govern and create value from AI: bilateral licensing with frontier model providers; collective or sectoral licensing; collaborative sovereign model initiatives; internal application-based value creation; commercialisation of internal tools as B2B products; and governed access, attribution, and monetisation layers.

The main lesson is that successful AI initiatives align technology with governance, IP, data ownership, and value-sharing from the start. Future European media AI initiatives should move beyond siloed experimentation toward collective approaches, especially to ensure that smaller and regional media players are not left behind.

## Introduction

This report is a deliverable in the VLAM program<sup>1</sup>, a partnership between imec, media organisations, and academic institutions to build multimodal AI models that reflect Flemish linguistic and cultural nuances.

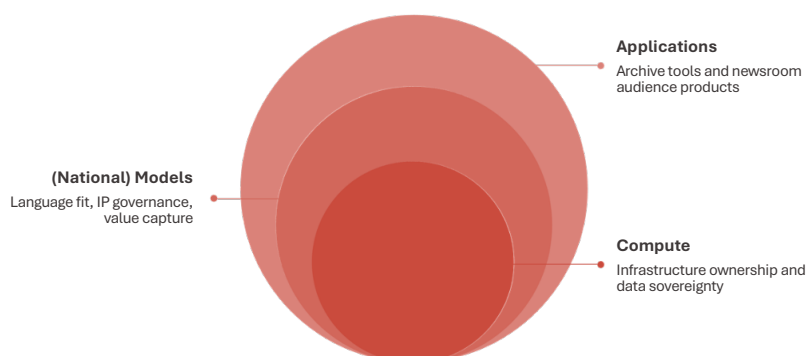
It maps Europe's media sector's position in the AI landscape, with a focus on how media organisations are affected by, contribute to, and capture value from AI initiatives across Europe. Such a mapping is needed because the landscape is shifting rapidly and lacks a consolidated overview: new model initiatives, infrastructure programmes, application-layer tools, licensing arrangements, and governance approaches are emerging in parallel across Europe.

For media organisations, this creates both uncertainty and strategic opportunity. They are increasingly affected by AI systems that reshape content production, search, distribution, audience interaction, rights management, and business models. At the same time, media organisations are not only passive users of AI. They contribute valuable content, archives, editorial expertise, trusted brands, and user relationships to AI systems.

**The strategic question is therefore not only how media organisations adopt AI, but how they can retain control and capture a fair share of the value created.**

## Approach

This report maps the landscape across three AI stack layers:



Across the three layers, the report asks two recurring questions: who controls the relevant infrastructure, data, models, and workflows; and who captures the economic or strategic value created from them? This report is based on structured desk research conducted between February and April 2026.<sup>2</sup> The research draws on three main source types: academic and policy literature on AI infrastructure, model development, and media sector AI adoption; publicly available documentation from European initiatives, national language model programmes, and media AI consortia; and industry reporting covering licensing agreements, partnership announcements, and application-layer tool deployments.

Sources were selected for recency and relevance to the European media context, prioritising initiatives active since 2023. Facts and figures were checked against a mix of official, academic, industry, and news sources; where public verification was not possible, claims are presented as indicative rather than definitive.

---

<sup>1</sup> <https://www.imec.be/nl/vlaamse-innovatiemotor/impactdomeinen/media-entertainment/vlam-programma>

<sup>2</sup> AI tools, such as Claude and Mistral's LeChat, were used to support literature research, brainstorming, and writing improvement throughout the process.

## Background

Two structural forces frame the entire analysis.

- First, **the (US-based) AI stack is becoming less transparent.** According to Stanford HAI, more than 90% of notable AI models in 2025 came from industry; The strongest models disclose fewer details about training data and model construction.<sup>3</sup> Media companies are increasingly being asked to build on systems they cannot properly inspect.<sup>4</sup> At the same time, open-source AI models are being released and are mostly used for specific tasks, but they often underperform proprietary models. Using open-source models could, in the long run, be a much cheaper option than proprietary models<sup>5</sup>.
- Second, **the information environment is deteriorating while media archives become more valuable.** The amount of AI-generated or AI-assisted web content is growing rapidly<sup>6</sup>. Recent estimates suggest a substantial rise, but the exact share remains methodologically uncertain.<sup>7</sup> This increases the value of provenance-marked journalism. Media archives are no longer just editorial output. They are training material, leverage and a strategic asset. Several European media organisations have entered licensing deals<sup>8</sup> with AI companies. However, such arrangements may prove less sustainable at renewal as market conditions shift and publishers' bargaining position evolves.<sup>9</sup>

Because the AI stack is becoming opaquer and media archives are now highly valuable training and retrieval assets, European media organisations face a critical choice. They can remain users of external AI systems, in which case control and value may shift to infrastructure and model providers, or they can actively shape, govern, and valorise the AI components that depend on their content, data, and editorial expertise.

---

<sup>3</sup> <https://hai.stanford.edu/ai-index/2026-ai-index-report> <https://arxiv.org/abs/2512.10169>

<sup>4</sup> Rydenfelt, H., Kiskola, J., & Olsson, T. (2026). AI in Media Consumption: Charting the Futures of Journalism. Journalism Studies, 1–24. <https://doi.org/10.1080/1461670X.2026.2627469>

<sup>5</sup> Irugalbandara, C., Mahendra, A., Daynauth, R., Arachchige, T. K., Dantanarayana, J., Flautner, K., ... & Mars, J. (2024, May). Scaling down to scale up: A cost-benefit analysis of replacing OpenAI's LLM with open source SLMs in production

<sup>6</sup> Dolezal, J., Alam, S., Graham, M., & Bohacek, M. (2026). The Impact of AI-Generated Text on the Internet. arXiv preprint arXiv:2604.26965. <https://ai-on-the-internet.github.io/>

<sup>7</sup> <https://hai.stanford.edu/ai-index/2026-ai-index-report>

<sup>8</sup> <https://nananwachukwu.github.io/media-capture-watch/>

<sup>9</sup> <https://tollbit.com/state-of-the-bots/q3-q4-2025/>

## Layer 1: Compute

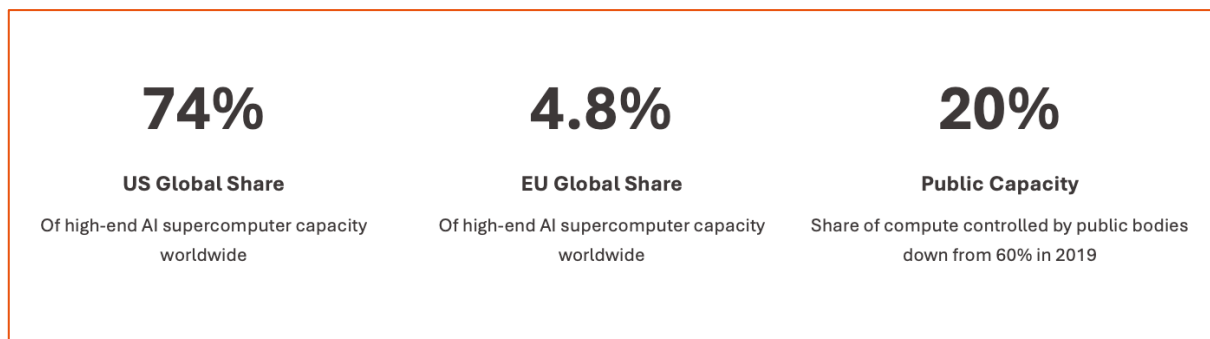
### What is the compute layer?

Training and running AI systems require enormous processing power. This processing happens on specialised hardware (primarily graphics processing units, or GPUs) housed in large data centres. The compute layer refers to this underlying infrastructure: who owns it, who can access it, and at what cost. For anyone building an AI system, from a frontier model or an AI tool, access to sufficient compute is a prerequisite.

For media organisations, the compute layer is mostly invisible but structurally significant: it determines where sensitive content is processed, which providers gain operational leverage, and how much of the value created by AI applications flows to external infrastructure providers. Day-to-day AI tools (transcription, recommendation, summarisation) run on commercial cloud infrastructure operated by, among others, Amazon, Microsoft, or Google. Training or fine-tuning a model on proprietary media content often requires a more deliberate computing arrangement, and the choice of infrastructure carries governance implications: who processes your data, under what legal terms, and in which jurisdiction.

### What is Europe's position in the compute layer?

Compute is heavily concentrated. The United States controls an estimated 74% of global high-end AI supercomputer capacity, while China holds 14% and the EU approximately 4.8%.<sup>10</sup> This infrastructure is overwhelmingly privately owned. In 2019, public bodies such as universities controlled 60% of computing capacity, but that share has now fallen below 20% as privately controlled infrastructure has proliferated.<sup>11</sup> The dominant providers (AWS, Microsoft Azure, Google Cloud) are all US-headquartered, meaning that European organisations that rely on them for AI workloads face questions about data sovereignty, legal jurisdiction, and pricing dependence.



Europe's response is primarily public and collective. Through EuroHPC, the EU has built a network of AI Factories that offer shared supercomputing access to research consortia, SMEs, and public institutions, explicitly designed as a sovereign alternative to commercial hyperscalers.

<sup>10</sup> <https://arxiv.org/pdf/2504.16026>; <https://www.federalreserve.gov/econres/notes/feds-notes/the-state-of-ai-competition-in-advanced-economies-20251006.html>

<sup>11</sup> <https://sciencebusiness.net/news/ai/state-ai-2025-five-key-charts-europeans>

Provider	Ownership & Jurisdiction	Data & IP Control	Cost / Access model <sup>12</sup>	Suited for	Value-capture implication for media
<b>LUMI (Finland)</b>  <b>EuroHPC</b>	Public EuroHPC infrastructure ; EU jurisdiction	Stronger EU-based control; public research infrastructure	EuroHPC access calls; public/research-oriented access	Training, fine-tuning, dataset processing, episodic large-scale workloads	Used by Nordic media AI efforts (YLE, NRK). Belgian organisations eligible as LUMI consortium members (Belgian Antenna)
<b>JUPITER (Germany, Jülich)</b>  <b>EuroHPC</b>	Public EuroHPC infrastructure ; EU jurisdiction	Stronger EU-based control; public research infrastructure	EuroHPC access calls; competitive allocation	Large-scale training, fine-tuning, AI-optimised workloads	Used for TildeLM, Scandinavian model training; Belgian organisations eligible as LUMI consortium members (Belgian Antenna)
<b>VSC (Vlaams Supercomputer) (Belgium)</b>  <b>Regional cluster</b>	Public Flemish/university infrastructure ; Belgian/EU jurisdiction	Data remains under Belgian/EU legal framework; institutional oversight	Allocation-based for academic users; industrial access via university partnerships	Research, training, fine-tuning, archive processing, proof-of-concept development	No media sector case studies yet; imec and Flemish universities who connect to media sector already have VSC access
<b>American Cloud</b>  <b>Commercial hyperscalers</b>	Private US-headquartered providers: Microsoft, Amazon, Google	Strong contractual control needed; US jurisdiction and Cloud Act-related concerns	Pay-per-use; commercial contracts	General-purpose deployment, scaling, inference, model APIs, newsroom tools	Most media tools currently hosted on commercial cloud (ARD, BBC, Axel Springer, Le Monde assistant)
<b>European Cloud (OVH, STACKIT,..)</b>	Private European providers, e.g. OVHcloud, STACKIT	EU jurisdiction; potentially stronger sovereignty positioning than US hyperscalers	Pay-per-use; commercial contracts	Deployment, hosting, scaling, inference, potentially sovereign AI services	Possible middle route between public HPC and US cloud, but media-specific examples unclear

Table 1: AI Compute Examples.

## Case study: EuroHPC

Europe has historically suffered from a structural and strategic dependency on commercial hyperscalers headquartered in the United States. However, by 2026, the European High Performance Computing Joint Undertaking (EuroHPC JU) aims to substantially alter this infrastructural landscape through the deployment of AI Factories and AI Factory Antennas.<sup>13</sup>

The European Union has established an interconnected network of 19 AI Factories and 13 AI Factory Antennas, explicitly designed to provide customised, AI-optimised supercomputing capacity to startups, small and medium-sized enterprises (SMEs), and research consortia without requiring them to make

<sup>12</sup> Published GPU pricing from commercial providers (roughly EUR 2–4/hour per GPU for inference-grade hardware) understates true total cost once storage, data transfer, and engineering overhead are included. However, large-scale model adaptation may require thousands to tens of thousands of GPU-hours, before storage, data transfer, orchestration, failed runs, evaluation, and engineering overhead are included (<https://lambda.ai/pricing>). EuroHPC public calls offer this at no direct cost for eligible projects.

<sup>13</sup> [https://www.eurohpc-ju.europa.eu/eurohpc-ju-selects-ai-factory-antennas-broaden-ai-factories-initiative-2025-10-13\\_en](https://www.eurohpc-ju.europa.eu/eurohpc-ju-selects-ai-factory-antennas-broaden-ai-factories-initiative-2025-10-13_en)

prohibitive capital investments in private hardware.<sup>14</sup> This infrastructure includes prominent facilities such as the LUMI AI Factory in Finland, the JUPITER system, and the newly contracted HammerHAI<sup>15</sup> AI Factory in Germany.

Belgium has been selected as an AI Factory Antenna site<sup>16</sup>, with EUR 4,3 million in federal investment (AI Antenna programme) and confirmed links to LUMI (Finland) and JUPITER (Germany). Table 2 provides European consortia with viable, sovereign alternatives for the computationally intensive phases of model development.<sup>17</sup>

Access Mode	Target Audience	Capacity Allocation Profile	Turnaround and Approval Timeline	Potential Use Cases for the Media Sector
<b>Playground Access</b>	Entry-level SMEs and Startups	Limited introductory computing resources	~2 working days	Proof-of-concept testing, initial model exploration
<b>Fast Lane Access</b>	HPC-familiar organizations	Up to 50,000 GPU hours	~4 working days	Mid-scale model fine-tuning, adaptation of base models, and dataset vectorization
<b>Large Scale Access</b>	Advanced consortia and enterprises	Exceeding 50,000 GPU hours	Subject to cut-off dates	Large-scale foundation model pre-training from scratch
<b>Collaborative Projects</b>	Public and private research entities	Resource intensive, subsidized cost	Horizon Europe / Digital Europe cycles	Scientific research where AI models represent a core workflow component
<b>Commercial Access</b>	Industrial applications	Pay-per-use allocation based on market rates	Negotiated SLAs	Enterprise deployment, proprietary model hosting, and continuous inference operations

Table 2: Access Models of EuroHPC Clusters (based on [https://www.eurohpc-ju.europa.eu/ai-factories/ai-factories-access-modes\\_en](https://www.eurohpc-ju.europa.eu/ai-factories/ai-factories-access-modes_en)).

While new public infrastructure is optimised for research, development, and episodic training runs, continuous, low-latency deployment (inference) for audience-facing media applications often requires hybrid cloud architectures.<sup>18</sup> This leads to an inherent tension: although AI factories are ideally suited to supporting research into training medium-sized AI models for the media sector, they are currently insufficient to drive mass-scale commercial AI innovation across the EU. Furthermore, partnership consortia are primarily composed of research institutions rather than commercial actors.<sup>19</sup>

For media companies, therefore, deploying real-time retrieval agents, generative summarisation tools, or automated breaking-news pipelines will likely involve using EuroHPC facilities for periodic model updates

<sup>14</sup> [https://www.eurohpc-ju.europa.eu/ai-factories\\_en](https://www.eurohpc-ju.europa.eu/ai-factories_en)

<sup>15</sup> Focused on manufacturing use cases.

<sup>16</sup> <https://matz.belgium.be/nl/nieuws/ai-antenna-de-federale-regering-investeert-43-miljoen-digitale-sovereiniteit>

<sup>17</sup> [https://www.eurohpc-ju.europa.eu/ai-factories\\_en](https://www.eurohpc-ju.europa.eu/ai-factories_en)

<sup>18</sup> <https://digiqt.com/blog/ai-agents-in-news-media/>

<sup>19</sup> <https://www.interface-eu.org/publications/ai-factories>

and fine-tuning. Conversely, it is unclear if they must rely on commercial hyperscalers or sovereign private cloud deployments to meet the strict uptime, security, and latency requirements demanded by the application layer and enterprise B2B clients.<sup>20</sup> We delve deeper into the application layer below.

Summarising, this programme is important for sovereign experimentation, but it does not eliminate dependence on commercial cloud in deployment and tooling. In practice, many media workflows will not focus on compute but will still rely partly on commercial cloud for deployment, tooling, and day-to-day operations.

## Key Learnings

- **Compute is a strategic issue for the media sector, even when it remains invisible.**

AI applications all depend on underlying compute infrastructure. For media organisations working with proprietary archives or sensitive content, compute choices affect cost, control, data governance, and strategic autonomy.

- **Compute choices shape data governance and IP control.**

Processing media archives on European public infrastructure offers stronger jurisdictional clarity and institutional oversight. Commercial hyperscalers offer mature tooling and scalability, but introduce additional questions around sovereignty, contractual control, and legal exposure.

- **Europe now offers a stronger public route for AI experimentation and fine-tuning.**

EuroHPC (including LUMI, JUPITER), VSC, and the linked Belgian AI Factory Antenna provide media-related research consortia and SMEs with access to sovereign compute for proof-of-concept testing, dataset processing, model adaptation, and fine-tuning.

- **Public compute is useful for development, but not sufficient for daily media operations.**

European public infrastructure is well-suited for episodic workloads such as training runs, fine-tuning, and R&D projects. Continuous media applications (such as real-time retrieval, recommendation, newsroom assistants, or audience-facing tools) will often still require cloud-based deployment from commercial (European or non-European) providers.

- **Media organisations will likely rely on hybrid compute environments, making workload governance essential.**

Public European infrastructure can support research and model development, while commercial cloud providers remain important for scaling, integration, uptime, and low-latency services. The key question is not simply which infrastructure is technically available, but which workloads should run where, under which data terms, and with which implications for cost and control.

- **European media organisations are still marginal in the compute layer.**

The compute landscape is mainly driven by research institutions, public bodies, supercomputing centres, and industrial AI actors. Media-specific needs (secure archive processing, rights-cleared training, provenance, and trusted retrieval) must therefore be actively articulated in future infrastructure initiatives.

---

<sup>20</sup> <https://digiqt.com/blog/ai-agents-in-news-media/>

## Layer 2: Models

### What is the model layer?

If compute is the engine, models are what the engine produces. A foundation model is a large AI system trained on vast amounts of text, audio, or other data, which learns to recognise patterns, generate language, answer questions, and perform a wide range of tasks. The most well-known examples (GPT-4, Gemini, Claude) are trained by large technology companies on datasets comprising much of the publicly available internet, at a cost of hundreds of millions of dollars per training run. The result is a general-purpose system that can be used directly or adapted for specific purposes. That adaptation process is called fine-tuning. Rather than training a model from scratch, an organisation takes an existing foundation model and continues training it on a smaller, more targeted dataset.

### What is Europe's media sector's position in the model layer?

At the frontier, the model layer is dominated by a small number of US companies and, increasingly, Chinese open-source efforts. Building and maintaining frontier models requires a combination of capital, compute access, and proprietary data that very few organisations globally can sustain. Europe has innovative commercial providers such as Mistral and Aleph Alpha that compete with OpenAI, Google, or Anthropic, particularly on multilingualism and compliance, but often lag behind hyperscalers in model size and performance. For media organisations, this creates a structural dependency: the models underpinning their AI tools are built by companies whose incentives, training data choices, and governance structures they cannot inspect or influence.

Europe's response has not been to compete directly on frontier scale. Instead, European (public) foundation model initiatives tend to follow a different logic. Rather than entering a financially exhaustive race to maximise parameter size, they focus on:

#### Digital Sovereignty & Cultural Context

Emphasising linguistic and cultural contextualisation to ensure AI reflects European values.

#### EU AI Act Adherence

Strictly aligning with the comprehensive regulatory framework for ethical and responsible AI development.

#### Openness & Public Access

Prioritising public access and open science principles over purely commercial advantage.

#### Equitable Copyright Governance

In a limited number of cases, establishing frameworks for fair compensation and rights retention for content creators.

Across these initiatives, three common objectives recur: cultural and linguistic contextualisation, data sovereignty<sup>21</sup> and security, and socioeconomic contribution through local capability development.<sup>22</sup> The result is a growing but fragmented landscape of national model initiatives across Europe.

For media organisations, the **key questions are therefore not only which model performs best, but who controls it, which data it was trained on, under which legal conditions, and whether media content contributors retain rights, oversight, and a route to value capture.**

Table 3 gives an overview of several European and US AI models relevant for the media sector.

<sup>21</sup> [https://www.linuxfoundation.org/hubfs/Research%20Reports/lfr\\_gosim25\\_072325.pdf?hsLang=en](https://www.linuxfoundation.org/hubfs/Research%20Reports/lfr_gosim25_072325.pdf?hsLang=en)

<sup>22</sup> <https://www.fticonsulting.com/insights/articles/building-national-large-language-model-llm>

Model	Base Model / Training Approach	Openness & licensing	Data governance & IP	Media Sector Involvement
<b>GPT-NL</b> TNO + SURF + NFI (Netherlands, public)	Trained from scratch on a curated corpus; RAG-optimised design	Mixed	Data processed in Netherlands; EU jurisdiction; permissively licensed public corpus; no IP claims on model	Dutch-language focus; NDP Nieuwsmedia (Dutch publishers) as data contributors with formal IP and revenue-sharing agreement
<b>NorLLM</b> National Library of Norway + partners (Norway, public)	Fine-tuned on Norwegian national corpus including Kopinor-licensed publisher content	Open: model weights released under Norwegian licence	Data processed in Norway; publisher content used under collective licence; remuneration model agreed before training	Norwegian-language; Kopinor (Norwegian collective rights body) licensed publisher archives as training data with remuneration model
<b>Teuken-7B</b> Fraunhofer + German research consortium (Germany, public)	Trained from scratch on 24 EU official languages; EuroHPC / JUPITER compute	Open: Apache 2.0	EU jurisdiction; EuroHPC infrastructure; training data sources documented	Pan-European multilingual; no specific media sector involvement
<b>OpenEuroLLM</b> Pan-European consortium (EU, public/mixed)	Open family of multilingual large language foundation models, with shared governance and open-source orientation.	Open	EU jurisdiction by design; governance shared across member institutions	All EU official languages; coordination layer across national efforts; no specific media sector involvement
<b>GPT-5</b> OpenAI (US, private)	Trained from scratch on proprietary data at frontier scale; architecture undisclosed	Closed: API access only; weights not released; some open variants released	Data sent to OpenAI servers; US jurisdiction; training data practices not fully disclosed; EU data residency options available but limited	Multilingual but English-dominant; major media licensing deals (Axel Springer, Le Monde, Schibsted, AP) as training data inputs
<b>Gemini</b> Google (US, private)	Trained from scratch; natively multimodal (text, image, audio, video); architecture proprietary	Closed: API via Google Cloud; some open variants (Gemma) released	Google Cloud data processing terms; US jurisdiction risk; GDPR-compliant configuration available	Multilingual; Google News Initiative partnerships with European publishers; used for transcription and archive search by Mediahuis
<b>Mistral Large 3 / Mistral Small 4</b> Mistral AI (France, private)	Trained from scratch; MoE architecture; European-headquartered; building own EU datacenter infrastructure	Open-weight (Apache 2.0) for most models; some proprietary variants; enterprise API available	Can be self-hosted in EU; Apache 2.0 licence permits fine-tuning and commercial deployment; EU data sovereignty by design	Multilingual including strong European language coverage; partnerships with European enterprises (ASML, CMA CGM, ESA); Le Chat media-facing product

Table 3: Examples of European AI Models.

Most European initiatives in our analysis have open-sourced their models, although concrete implementations are still to be developed in many cases. At the same time, there is growing friction between open-source and proprietary approaches. For example, when the Chinese open-source model DeepSeek was released to the public, American lawmakers quickly recognised the need to support American open-source providers in competing on the global stage. Since then, US companies have followed a mixed strategy, with some open-sourcing selected models and others restricting their open-source strategies.<sup>23</sup>

This debate is also becoming more relevant for media organisations because dependence on proprietary models can quickly become a cost issue. Recently, Copilot announced a steep price increase through usage-based and premium-request billing changes.<sup>24</sup> Other providers are following similar paths. This increases pressure on European actors to rethink dependency and to explore open-source models such as Mistral, Qwen, Kimi, and European national models as possible foundations for media-sector applications.

At the same time, European media organisations have not waited for national model programmes to resolve the IP question. Alongside, and largely independently of public model initiatives, a growing number of publishers have begun negotiating directly with US frontier model developers to license their archives as training and retrieval data. Axel Springer concluded a multi-year agreement with OpenAI in 2023<sup>25</sup>, covering training data access and product integration across its European titles, including Bild and Welt. Le Monde and Prisa followed with similar bilateral deals with OpenAI in 2024<sup>26</sup>. The Financial Times concluded a licensing agreement with OpenAI in 2024<sup>27</sup>, and entered a program with Google in 2026<sup>28</sup>. Schibsted, the Norwegian-Swedish publishing group, entered a partnership with OpenAI covering both content licensing and development of AI-powered editorial tools<sup>29</sup>. In the Netherlands, NDP Nieuwsmedia negotiated a collective arrangement linking publisher data contributions to the GPT-NL initiative (see below) rather than to a US provider. Potentially, the only known example of a multi-publisher licensing model that keeps both the model and the revenue within a national ecosystem.

These deals establish a commercial precedent: European media archives have measurable value as AI training data, and publishers can extract revenue from that value. However, the terms of most bilateral agreements are commercially confidential, making it difficult to benchmark what fair remuneration looks like or whether current deals adequately reflect the long-term value of content contributions<sup>30</sup>. Several media law scholars and publisher associations have also raised concerns about the sustainability of these arrangements at renewal, as the bargaining position of individual publishers relative to frontier model developers may weaken once initial deals are signed and content has already been used for training<sup>31,32,33</sup>.

For European media organisations, this context is directly relevant. This is the core model-layer question for European media: not whether media archives have value as AI training data but whether that value is captured individually through bilateral deals or collectively through a governed framework that retains IP, control, and ongoing revenue within the media ecosystem. The EU Copyright Directive's text-and-data

---

<sup>23</sup> <https://www.economist.com/business/2025/08/21/china-is-quietly-upstaging-america-with-its-open-models>

<sup>24</sup> <https://docs.github.com/en/copilot/reference/copilot-billing/models-and-pricing#model-multipliers-for-annual-copilot-pro-and-copilot-pro-subscribers>

<sup>25</sup> <https://www.reuters.com/business/media-telecom/global-news-publisher-axel-springer-partners-with-openai-landmark-deal-2023-12-13/>

<sup>26</sup> <https://openai.com/index/global-news-partnerships-le-monde-and-prisa-media>

<sup>27</sup> [https://aboutus.ft.com/press\\_release/openai](https://aboutus.ft.com/press_release/openai)

<sup>28</sup> <https://pressgazette.co.uk/platforms/google-search-features-deals-ai-financial-times/>

<sup>29</sup> <https://schibsted.com/news/schibsted-media-partners-with-openai/>

<sup>30</sup> <https://committees.parliament.uk/committee/170/communications-and-digital-committee/news/212361/uk-creative-industries-face-a-clear-and-present-danger-from-generative-ai/>

<sup>31</sup> Oderinwale, H., & Kazlauskas, A. (2025). The Economics of AI Training Data: A Research Agenda. ArXiv, abs/2510.24990

<sup>32</sup> Adam Buick, Copyright and AI training data—transparency to the rescue?, Journal of Intellectual Property Law & Practice, Volume 20, Issue 3, March 2025, Pages 182–192, <https://doi.org/10.1093/jiplp/jpae102>

<sup>33</sup> <https://mediacopilot.ai/ai-pay-for-news-statutory-licensing/>

mining provisions and the Press Publishers Rights Directive provide a legal foundation for remuneration claims, but enforcement and valuation methodologies remain unresolved at a European level<sup>34</sup>.

In summary, the model layer matters for media organisations for two main reasons.

- **First, it determines how well an AI system understands the content they produce:** models trained predominantly on English-language internet data are likely to underperform on European languages, local cultural references, and journalistic source material.
- **Second, it is increasingly where intellectual property and value-capture questions become acute.** Media archives are valuable training data, and whether a media organisation contributes that data to an external model, a shared national model, or a governed sectoral initiative is as much a legal and strategic question as a technical one.

## Case study: GPT-NL (Netherlands)



The development of GPT-NL (launched in 2023) serves as the preeminent European case study for how media organisations can actively collaborate with technology developers to train base models without forfeiting the economic value of their intellectual property. Launched by the Netherlands Organisation for Applied Scientific Research (TNO), the Netherlands Forensic Institute (NFI), and SURF, GPT-NL is a highly structured initiative backed by EUR 13.5 million in funding from the Ministry of Economic Affairs and Climate Policy.<sup>35</sup>

Unlike early international models that relied on indiscriminate, unauthorised web scraping of copyrighted material, GPT-NL explicitly partnered with NDP Nieuwsmedia (the trade association representing over 30 national and regional news titles, including DPG Media and Mediahuis) and the ANP news agency.<sup>36</sup> This landmark collaboration secured access to over 20 billion tokens of highly curated, rights-cleared journalistic content, resulting in the GPT-NL Public Corpus, which contains 36 billion<sup>37</sup> pre-processed Dutch tokens.

This ethical data acquisition was achieved through the establishment of a "Content Board." This institutional mechanism provides data contributors and rights holders with active oversight and a permanent voice regarding how their intellectual property is utilised in the model's ongoing development.

GPT-NL's official revenue-sharing mechanism allows rights holders to opt into a 50% proportional share of net revenues from professional licences.<sup>38</sup> The remaining 50% is strictly ring-fenced and reinvested into the continuous operation and refinement of the GPT-NL infrastructure.<sup>39</sup> Furthermore, strict technical agreements and architectural constraints are embedded into the model to prevent the verbatim extraction of copyrighted articles by end-users, thereby protecting the publishers' primary digital assets from cannibalisation.<sup>40</sup> The GPT-NL model demonstrates that it is possible to build linguistically nuanced foundational AI while treating media archives as governed, monetisable assets rather than as an exploitable public commons.

The deployment trajectory also shows that GPT-NL is moving beyond research and model development: after training started in June 2025 and first use was prepared for Q4 2025, the model was already being

---

<sup>34</sup> <https://data.consilium.europa.eu/doc/document/ST-15147-2025-REV-1/en/pdf>

<sup>35</sup> [www.gpt-nl.nl](http://www.gpt-nl.nl)

<sup>36</sup> <https://auteursbond.nl/nieuwsmedia-leveren-archieven-aan-gpt-nl-vergoeding-voor-freelance-journalisten-nog-onduidelijk/>

<sup>37</sup> [https://huggingface.co/datasets/GPT-NL/GPT-NL\\_Public\\_Corpus](https://huggingface.co/datasets/GPT-NL/GPT-NL_Public_Corpus)

<sup>38</sup> [https://gpt-nl.nl/publish/pages/8907/gptnl-pres-25-040-dnb-for\\_public\\_publication.pdf](https://gpt-nl.nl/publish/pages/8907/gptnl-pres-25-040-dnb-for_public_publication.pdf)

<sup>39</sup> [https://gpt-nl.nl/publish/pages/7064/annex\\_6\\_revenue\\_sharing\\_mechanism.pdf](https://gpt-nl.nl/publish/pages/7064/annex_6_revenue_sharing_mechanism.pdf)

<sup>40</sup> <https://www.tno.nl/en/newsroom/2025/07/large-dataset-news-organizations-dutch/>

tested by April 2026 in municipal virtual assistant services through ICTU's Gem project in cities such as Utrecht, Rotterdam, and Eindhoven<sup>41</sup>.

### Case study: NorLLM (Norway)



Norway's approach introduces a complementary model to GPT-NL. NorLLM is developed by NorwAI (the Norwegian Research Center for AI Innovation), a consortium involving Schibsted, VG, and NRK as industry data partners alongside universities and research institutes. The model family covers Norwegian, Sami, and Scandinavian languages and is openly available via Hugging Face.

Additionally, on 19 December 2025, Kopinor (the Norwegian rights management organisation) signed an agreement with the National Library of Norway establishing the first government-funded national licensing arrangement for newspaper content used in AI training.<sup>42</sup> The Norwegian state pays NOK 45 million (approximately EUR 3.8 million) annually to license journalistic content for AI training purposes. Publishers can opt out of specific content, and clear data governance requirements, documentation standards, and license terms are stipulated for any models trained on the corpus. Norway is the first country in the world to establish such an arrangement.<sup>43</sup>

The NorLLM/Kopinor combination demonstrates a two-track approach: direct industry participation in model development (Schibsted/NRK) alongside a state-mediated licensing mechanism (Kopinor) that compensates publishers whose content is used even when they are not consortium members. Both tracks are potentially relevant for future media sector initiatives.

### Case study: OpenGPT-X / Teuken & Soofi (Germany)



The German research consortium, involving FZ Jülich, Fraunhofer, and DFKI, produced Teuken-7B (a 7B-parameter, open model covering 24 EU languages), released in 2024. WDR (a German public service broadcaster) participated as a media partner, demonstrating that public broadcasters can be core members of the consortium rather than peripheral stakeholders. Deutsche Telekom operationalised the output commercially via 'AI Foundation Services', establishing a tiered commercialisation structure through Fraunhofer. Currently, Fraunhofer and consortium members are working on Teuken 7B's successor, Soofi.

The follow-on initiative, Soofi, is developing a 100-billion-parameter model targeting EU languages and industrial applications. Funded by approximately EUR 20 million from the German Ministry of Economic Affairs under the IPCEI-CIS framework, and training from March 2026 on European AI Factory infrastructure, it is targeted for Q3 2026 release. The three-tier architecture (base model, reasoning model, and agent layer) makes Soofi one of the most ambitious European sovereign model initiatives currently in development.<sup>44</sup>

---

<sup>41</sup> <https://www.ictu.nl/artikelen/tno-en-ictu-starten-met-praktijkproef-gpt-nl-voor-virtuele-assistent-gem/>

<sup>42</sup> <https://ifrr.org/page/article-detail/ifrr-member-kopinor-signs-historic-agreement-on-newspaper-content-for-ai-in-norway/>

<sup>43</sup> <https://ifrr.org/page/article-detail/ifrr-member-kopinor-signs-historic-agreement-on-newspaper-content-for-ai-in-norway/>

<sup>44</sup> <https://www.soofi.info/aktuelles/>

## Key Learnings

- **European media organisations remain marginal in the model layer, despite contributing some of its most valuable inputs.**

Most European model initiatives are led by public research institutions, supercomputing centres, or national AI programmes. Media organisations are usually treated as future users or data sources, not as co-owners or co-governors. The main exceptions identified are GPT-NL, NorLLM, and OpenGPT-X/Teuken, where media actors participated more directly.

- **The model layer matters because it shapes both language fit and strategic dependency.**

Models trained mainly on English-language internet data are unlikely to fully capture European languages, local cultural references, journalistic formats, and national media contexts. European initiatives, therefore, tend to prioritise linguistic and cultural contextualisation, sovereignty, and legal compliance rather than frontier-scale performance.

- **Media archives have become valuable model-layer assets.**

Licensing deals between European publishers and US model providers show that media archives have measurable value as training and retrieval data. However, bilateral deals often remain confidential and may transfer long-term control and value capture to external model providers.

- **Governance, IP, and value-sharing need to be solved before training begins.**

The strongest cases are those where data rights, contribution models, and revenue-sharing arrangements were agreed upon up front. GPT-NL and NorLLM/Kopinor show that rights-cleared training data, publisher oversight, and compensation mechanisms can be built into the model initiative from the start.

- **Open models create opportunity, but not automatic valorisation.**

Many European initiatives release models openly or as public infrastructure. This supports sovereignty, experimentation, and local capability-building, but does not, in itself, create a sustainable business model. Several initiatives still lack clear routes to adoption, maintenance, commercial use, or long-term financing after the research/project phase.

- **Some European initiatives show that media-sector involvement at the model layer is achievable — and point to distinct governance routes.**

GPT-NL addresses three problems simultaneously: Its Content Board gives rights holders ongoing oversight, while its revenue-sharing mechanism returns part of commercial licensing income to data contributors and reinvests the remainder in the model infrastructure. NorLLM complements this by combining direct media participation with a state-mediated licensing model through Kopinor. OpenGPT-X/Teuken adds public broadcasters as active consortium members, with publicly funded model outputs subsequently commercialised through service layers and industrial partners such as Deutsche Telekom.

- **For future media-sector initiatives, the key challenge is not only building a model, but governing and sustaining value over time.**

Many European model projects are time-bounded, publicly funded, and technically impressive, but their long-term continuation remains uncertain. For media organisations, the relevant questions are: who maintains the model, who pays for updates, who controls access, who captures revenue, and how content contributors are protected over time.

## Layer 3: Applications

### What is the application layer?

Media companies generally do not build foundation models themselves. They create value at the application layer by integrating AI into archives, editorial workflows, verification, recommendations, and user-facing services, either directly or with industry partners. This is where AI adoption becomes most visible, but also where questions of dependency and value capture become operational: which external models are used, which data flows back to providers, and whether the resulting value remains internal, public, or commercialised. This makes the application layer the most mature and realistic entry point for many media actors.

The application layer sits on top of the compute and model layers. It is where AI becomes most visible and useful to actual users. For example, journalists searching an archive, editors checking a draft, audiences receiving a personalised recommendation, or a radio producer generating a script summary. Unlike the model layer, the application layer does not require owning or training a foundation model. A media organisation can build highly effective AI tools by connecting existing open-source or commercial models to its own data, metadata, and editorial workflows via APIs and retrieval systems.<sup>45</sup>

### What is Europe's media sector's position in the application layer?

The application layer has historically been a focus for Europe's media. Publishers focus on the application layer, as it is often the most visible to audiences. News content is becoming more valuable as training input. The latest AI Index report by Stanford HAI points to a shift toward proprietary data as high-quality public web data becomes harder to rely on. Publisher's archives are, besides editorial content, also training material, leverage, and potentially a strategic asset.<sup>46</sup> Publishers have licensed their archives to frontier AI providers (see above), receiving upfront revenue but transferring a degree of governance and IP control to an external party. On the other hand, they can invest in building or co-building application-layer tools that keep data, governance, and value creation within the sector. The tension is sharpened by two market dynamics.

First, Reuters Institute's 2026 media trends report found that Google search traffic to more than 2,500 news sites declined by 33% globally between November 2024 and November 2025, while publishers expected search traffic to fall further over the next three years. This strengthens the case for media organisations to invest in direct audience relationships, archive-based products and provenance-rich content strategies. Second, Answer Engine Optimisation (AEO) is emerging as a strategic priority.<sup>47</sup> This strengthens the case for media organisations to invest in direct audience relationships, archive-based products and provenance-rich content strategies.

Additionally, the information environment is getting noisier. The report cites research suggesting that by mid-2025, 35% of newly published websites contained AI-generated or AI-assisted text. For media and journalism, this means polluted content, increased pressure on verification, and greater value in reporting that still shows where it came from.<sup>48</sup> The DPP European Media Trends 2026 report<sup>49</sup> confirms that AI has become the dominant driver of technological change across European broadcasters, with a quarter of companies shifting toward greater in-house development, mostly enabled by AI-assisted coding tools. Here as well, digital sovereignty is a growing priority in business resilience planning across the sector.

---

<sup>45</sup> <https://www.akamai.com/blog/partners/hidden-value-ai-transforms-media-archives-revenue-engines>

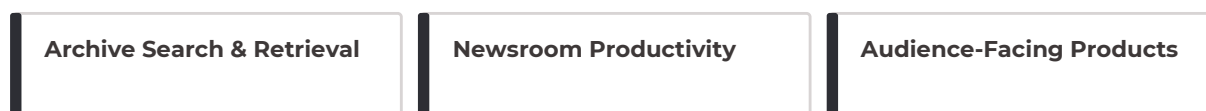
<sup>46</sup> <https://hai.stanford.edu/ai-index/2026-ai-index-report>

<sup>47</sup> <https://www.forbes.com/sites/lutzfinger/2025/06/19/answer-engine-optimization-aeo--what-brands-need-to-know/>

<sup>48</sup> <https://hai.stanford.edu/ai-index/2026-ai-index-report>

<sup>49</sup> <https://www.thedpp.com/news/ai-digital-sovereignty-impacting-tech-strategy-european-media-trends-2026-report>

At the application layer, Europe has more actionable media-sector cases than at the compute or model layer. Our analysis identifies three main AI application areas:



These are the areas where media organisations are already translating AI into operational, editorial, and commercial value. In most of these applications, media organisations do not train foundation models from scratch. Instead, they build on general-purpose models, often developed outside the media sector and not specifically trained on local languages, cultural references, or proprietary media content.

However, this does not mean that media organisations have no control or distinctive advantage. Their own data can still be used in the development of application-layer tools: for example, through retrieval-augmented generation, archive indexing, fine-tuning, metadata enrichment, workflow integration, or human editorial validation. In practice, this means that value is created less by owning the base model itself and more by combining existing models with trusted content, domain-specific data, editorial standards, and user-facing services. The key strategic issue is whether this value remains limited to internal efficiency gains or can be transformed into durable, governed assets or services.

Public broadcasters have built production-scale AI tools without owning a base model, using their archives, trusted brands, and editorial governance as differentiators. Other media organisations have gone further by commercialising internally developed tools as standalone B2B products. These cases show that application-layer innovation can create value in different ways: internally, through productivity gains, better archive access, improved workflows, and stronger user engagement; or externally, through licensing, service provision, or the creation of commercial products for other media organisations.

The strategic advantage of European media organisations at this layer is therefore not technical scale, but the combination of high-quality local-language content, editorial credibility, trusted brands, and regulatory alignment with the EU AI Act and GDPR. These are assets that generic AI tools cannot easily replicate, and they form the basis for the three application areas discussed below.

## Case studies: Archive search and retrieval

Archive retrieval is one of the most mature and best-documented AI application areas in European media. The examples in the table below illustrate that such applications are already operational across public broadcasters and commercial publishers at scale, often built without ownership of a base model or infrastructure.

The cases suggest two broad patterns. The first is public or internal infrastructure, where archive AI improves access and productivity but does not itself become a commercial product. VPRO/NPO, BBC, meemoo and Beeld & Geluid fit this pattern. The second is commercial archive-enabled infrastructure, where archive access, metadata and AI-assisted discoverability are packaged as a paid service. Reuters Connect is the clearest example, because Reuters explicitly markets AI-powered discoverability features for its archive (video, images, text) within a commercial content platform. This service is offered as a monetizable B2B service.

These cases are therefore assessed not only as technical applications, but as different ways of turning archives into accessible, reusable, and potentially monetisable assets.

Case	Type of Application	Data Governance	Value Capture / Market Status
<b>Archief van de Toekomst</b> <sup>50</sup> VPRO / NPO (Netherlands)	Uses AI/machine learning to make 500+ VPRO Tegenlicht broadcasts searchable through image/speech recognition, and text analysis; Built with Medialab, SudoX and Beeld & Geluid	The project is described as based on open standards and open-source technology; The public source does not describe external AI-provider data transfer	Internal/public-good model: free public access, no direct commercial revenue model. Operational since 2022; strong proof that archive enrichment is technically feasible and publicly valuable.
<b>Prompto</b> <sup>51</sup> RTBF/NRB (Belgium)	RAG system; Claude Sonnet 3.5 on AWS	Runs on AWS; Further details on content access, retention, and rights governance are not publicly specified	Internal value creation: developed by NRB for RTBF use. No external commercialisation reported. Useful example of RAG-based archive access in a Belgian media context.
<b>BBC Archive Search</b> <sup>52</sup> BBC (UK)	Internal archive search system using BBC R&D capabilities and AI-assisted search across BBC archive repositories	Operates within the BBC's internal archive and rights environment; Exact model architecture and use of external providers are not fully public	Internal productivity and public-service value: supports programme makers and journalists. No direct B2B commercialisation of archive search to date; strong example of archive governance and internal AI capability-building.
<b>Mediahuis NewsSearch</b> <sup>53</sup> Mediahuis (Belgium / Netherlands)	Uses Google Cloud AI, Elasticsearch, Vertex AI and Gemini to search archives, answer questions, summarise and suggest editorial angles	Runs through Google Cloud infrastructure; data governance depends on contractual and technical controls with Google Cloud	Internal productivity model: operational internal tool, no external commercialisation reported.
<b>Beeld en Geluid Archival tools</b> (Netherlands)	Open-source speech recognition and image tagging; partner in NPO Archief van de Toekomst	Public infrastructure; rights remain with contributing organisations	Public/shared infrastructure model: supports access, discoverability, and reuse of audiovisual heritage. Operational; no direct commercial AI product model.
<b>Meemoo tools</b> (Belgium)	Open-source speech recognition + semantic tagging	Public infrastructure; rights remain with contributing organisations	Public/shared infrastructure model: creates institutional and cultural value through better access and preservation. Operational; no direct commercial AI product model.
<b>Reuters Connect</b> <sup>54</sup> (UK)	Commercial content platform offering access to Reuters content, with AI-assisted search, metadata, and discoverability features	Proprietary Reuters platform; Reuters controls the archive, metadata, licensing terms, and customer access	Commercialised archive infrastructure model: archive access and AI-assisted discoverability are packaged as a paid B2B service for publishers, broadcasters, platforms, and brands. Strong example of archive and metadata capabilities becoming a market-facing product.
<b>EBU<sup>55</sup> NEO</b> <sup>56</sup> (Switzerland)	RAG-based multilingual news chatbot grounded strictly in EBU member content	Strict RAG architecture: AI is confined exclusively to retrieved member articles	Value captured through strengthening the EBU member ecosystem and demonstrating a trusted alternative to commercial AI news tools

Table 4: Examples of AI Applications in Archival Search and Retrieval.

<sup>50</sup> <https://avdt.vpro.nl/>

<sup>51</sup> <https://tech.ebu.ch/files/live/sites/tech/files/shared/techreports/tr093.pdf>

<sup>52</sup> [https://www.upf.edu/web/emerald/news/-/asset\\_publisher/sdagEwNgX9gZ/content/improving-bbc-archive-search-capabilities-using-ai/maximized](https://www.upf.edu/web/emerald/news/-/asset_publisher/sdagEwNgX9gZ/content/improving-bbc-archive-search-capabilities-using-ai/maximized)

<sup>53</sup> <https://annual-report.mediahuis.com/2024/2024/naar-een-centraal-technologieplatform>

<sup>54</sup> <https://reutersagency.com/content-delivery-platforms/reuters-connect/>

<sup>55</sup> EBU Tech Report TR 093 (2025). [tech.ebu.ch/files/live/sites/tech/files/shared/techreports/tr093.pdf](https://tech.ebu.ch/files/live/sites/tech/files/shared/techreports/tr093.pdf)

<sup>56</sup> <https://tech.ebu.ch/neo>

## Case studies: Newsroom productivity

Beyond access to archives, AI is embedded in editorial and production workflows across European media organisations. The cases in the table below show two distinct patterns: internal tools that remain in-house (Mediahuis, VRT, Schibsted), and internal tools that were subsequently commercialised into standalone products sold to other publishers (Arc XP, United Robots).

A risk to newsroom productivity is that AI could soon determine which topics to cover and how to tell stories, with minimal human oversight. This prompts the question: in automated news writing, where does gatekeeping happen, and who or what serves as the gatekeeper?<sup>57</sup>

Case	Type of Application	Data Governance	Valorisation / Market Status
<b>VRT Smart Content Assistant</b> <sup>58</sup>	Internal prompt library/content repackaging tool for titles, summaries, descriptions, social posts and platform-specific formats	VRT states that the tool helps staff repackage stories in the right tone of voice in a safe, controlled environment. Future Media Hubs reports 113,000+ interactions and 200+ unique users since launch.	Internal productivity tool; No evidence of external commercialisation; Application-layer AI is already operational inside Flemish public media
<b>Schibsted AI Tools / VG</b>	AI-supported investigative journalism and newsroom tools	Schibsted reports that VG used an adapted GPT to analyse 6,000 pages in an investigation into illegal adoptions <sup>59</sup> ; Schibsted also says its newsrooms have AI policies and AI-responsible teams <sup>60</sup>	Mostly internal capability building; strategic value lies in governance and skills; Open sourced of one AI tool that turns news articles in video <sup>61</sup>
<b>United Robots</b> <sup>62</sup>	Automated routine reporting using structured data	United Robots describes automated reporting for sports, real estate, traffic incidents and weather warnings; the company traces its origin to Mittmedia and registered as a company in 2016	Clear B2B Content-as-a-Service model; AI component was fully valorized as a standalone company that now licenses its "robot journalism" software globally; <sup>63</sup> It is structured-data automation, not archive search or generative foundation-model work
<b>Arc XP</b> <sup>64</sup>	Publisher technology platform originally associated with The Washington Post	The Washington Post states Arc Publishing became Arc XP and provides digital experience technology for media and enterprise clients;	B2B software spin-out model, but financially difficult
<b>Chat DPG</b> <sup>65</sup>	An innovative AI-driven tool designed to enhance newsroom workflows built on OpenAI GPT-4 (2024)	Internal developed tool	Mostly internal capability building

Table 5: Examples of AI Applications in Newsroom Productivity.

These cases show that newsroom AI is not a single category. VRT and Schibsted are primarily examples of internal capability-building: AI helps staff repackage content, search documents, analyse material and

<sup>57</sup> Voinea DV. Reconceptualizing Gatekeeping in the Age of Artificial Intelligence: A Theoretical Exploration of Artificial Intelligence-Driven News Curation and Automated Journalism. *Journalism and Media*. 2025; 6(2):68.

<sup>58</sup> <https://www.vrt.be/nl/over-ons/nieuws-over-vrt/nieuwe-collega-aan-boord-de-smart-content-assistant-verpakt-vrt-verhalen-naar-ek-platform>

<sup>59</sup> <https://schibsted.com/news/how-vgs-journalists-used-ai-to-find-signs-of-illegal-adoptions/>

<sup>60</sup> <https://schibsted.com/tech-ai/>

<sup>61</sup> <https://schibsted.com/news/schibsted-open-sources-ai-tool-that-turns-news-articles-into-videos/>

<sup>62</sup> <https://www.unitedrobots.ai/>

<sup>63</sup> <https://pressgazette.co.uk/news/real-estate-robot-journalism-automated-property/>

<sup>64</sup> <https://www.arcxp.com/>

<sup>65</sup> <https://www.futuremediahubs.com/future-media-hubs/cases/transforming-newsrooms-how-dpg-media-leverages-ai-chatdpg>

work more efficiently. United Robots is a B2B automation provider built around structured data and natural-language generation. Arc XP (US-based) shows that media-originated technology can become a commercial platform, but also that B2B media software is capital-intensive and not automatically profitable. The key distinction is whether AI remains an internal productivity layer or becomes a reusable product, service, or capability that can generate value beyond the original newsroom.

## Case studies: Audience-facing products

Audience-facing AI products are closest to direct revenue and distribution impact, but they are also the least settled. The Reuters Institute Digital News Report 2025 notes that while 80% of media leaders describe AI as important for news distribution and recommendation, fewer than 30% of audiences express a strong interest in any single personalisation option. Article summaries and translations rank highest.<sup>66</sup> Reuters Institute research also finds that interest in AI-driven news personalisation remains below 30% across all tested options, although article summaries and translations perform better than more ambitious forms of personalisation.

Current media experiments therefore cluster around relatively low-risk formats: summaries, translations, audio briefings, personalised recommendations and conversational interfaces. The Independent’s Bulletin<sup>67</sup>, for example, uses Google Gemini to produce bullet-point briefings that are reviewed by journalists before publication. BBC News has also created a growth, innovation and AI department<sup>68</sup> focused on more personalised content, especially for younger audiences.

Case	Type of Application	Data Governance	Valorisation / Market Status
<b>NRK (Norway)<sup>69</sup></b>	AI-generated article summaries; Live transcription; Automated translation into Sámi	OpenAI API; Every summary journalist-reviewed before publication	Internal public good; Not commercialised. 1 in 5 users expand summaries; Engaged readers spend 60s vs. 27s
<b>The Independent / Bulletin (UK)<sup>70</sup></b>	AI-rewritten summaries published as a separate product; Bylined to original journalist	Google Gemini; Dedicated editorial team of reviews every piece before publication	Commercially viable; Separate domain and app in development
<b>Axel Springer / Hey_ on BILD<sup>71</sup> <sup>72</sup>(Germany)</b>	Conversational AI assistant; Personalised summaries and follow-up responses grounded in BILD journalism	Azure OpenAI (Microsoft partnership); Axel Springer also holds a direct OpenAI content licensing deal	Audience-facing commercial product; Part of announced third-pillar AI strategy; No external licensing
<b>Taboola / Deeper Dive AI<sup>73</sup></b>	Conversational AI for content discovery within Taboola’s recommendation network	Third-party platform; Publisher content accessed via existing Taboola distribution agreements	Operational on multiple publishers’ infrastructure

<sup>66</sup> Reuters Institute Digital News Report 2025

<sup>67</sup> <https://www.the-independent.com/tech/intelligence-the-independent-financial-times-apple-geordie-greig-b2717465.html>

<sup>68</sup> <https://www.theguardian.com/media/2025/mar/06/bbc-news-ai-artificial-intelligence-department-personalised-content>

<sup>69</sup> <https://reutersinstitute.politics.ox.ac.uk/news/how-norways-public-broadcaster-uses-ai-generated-summaries-reach-younger-audiences>

<sup>70</sup> <https://pressgazette.co.uk/publishers/digital-journalism/independent-bulletin-ai-written-audiences-prefer-works-commercially/>

<sup>71</sup> <https://news.microsoft.com/source/2024/04/29/axel-springer-and-microsoft-expand-partnership-across-advertising-ai-content-and-azure-services/>

<sup>72</sup> <https://hey.bild.de/>

<sup>73</sup> <https://www.taboola.com/deeperdive/>

<b>Schibsted<sup>74</sup> / Curate + personalised audio (Norway, Sweden)</b>	Personalised homepage ranking; AI-generated audio articles; Interest-prediction model for subscription conversion	Proprietary in-house; Top editorial slots manually curated; Editorial signals weighted in algorithm	Internal only; Campaigns delivered improvement in subscription sales conversion
<b>Ask INMA<sup>75</sup> (Global)</b>	RAG-based AI answer engine trained on INMA's proprietary archive of case studies, reports, and presentations; Members query it like a knowledge assistant	Fully proprietary; ~90% of content walled off from public AI and search; Member-only access; Built by Techlabee.ai (EU AI company)	B2B audience-facing product for media professionals; No commercialisation; Value captured through membership retention and differentiation

Table 6: Examples of AI Applications for Audience-Facing Tools.

Audience-facing deployment represents the frontier of media AI and the layer closest to direct revenue impact, but also the layer with the least settled evidence on what audiences actually want. The examples span synthetic audio, personalisation, and conversational interfaces. Most European public broadcasters are in the early stages of experimentation. Still, industry enthusiasm outpaces documented audience demand. NRK's summary tool and The Independent's Bulletin are the two European cases with actual outcome data; both show positive engagement signals, but both are also conservative in scope (summaries, not generative content). Schibsted's Curate data is the strongest evidence for personalisation ROI, but it operates on editorial ranking logic rather than generative AI. These cases show how media organisations are testing whether AI can strengthen direct audience relationships, but the evidence for durable revenue capture remains uneven.

## Key Learnings

- **The application layer is Europe's strongest position in the AI stack and the most realistic near-term entry point for media organisations.**

It does not require owning or training a foundation model. Value is created by combining existing models with trusted content, domain-specific data, editorial standards, and user-facing services.

- **Two value-capture patterns define the AI application field.**

The internal or public-good model delivers productivity, access, and engagement gains but generates no external revenue and depends on sustained institutional funding. The commercialised-infrastructure model (where internal tools become B2B products or services) can create durable revenue streams, but is capital-intensive and not automatically profitable.

- **Many media companies are already active at this layer, which is both an opportunity and a constraint.**

Future media initiatives must demonstrate clear added value over what already exists in the market.

- **Industry enthusiasm for audience-facing AI significantly outpaces documented audience demand.**

Summaries, translations, and audio briefings have the strongest evidence base. Conversational interfaces and full personalisation have high editorial trust risk and weak audience uptake data.

- **AI applications by European media are predominantly siloed and organisation-specific.**

<sup>74</sup> <https://wan-ifra.org/2025/08/how-schibsteds-ai-model-helped-boost-subscription-sales/>

<sup>75</sup> <https://www.inma.org/blogs/earl/post.cfm/100-days-of-ask-inma-how-ai-unlocked-the-world-s-smartest-news-media-knowledge-base>

No European case reviewed has successfully built a shared application layer across competing media organisations within the same market. Bringing multiple media organisations together is the most ambitious and least-precedented design challenge and the one most dependent on resolving governance before building.

- **A shared application layer requires explicit governance design around contribution, free-riding, and value distribution.**

Partners who contribute less data, infrastructure, or expertise must not be able to extract equal or greater value without agreed compensation mechanisms. None of the European cases reviewed have fully resolved this within a competitive media market, making this one of the most important open questions for future sector-wide initiatives.

## Conclusion: From AI Adoption to Governed Value Capture

This report shows that Europe's media sector is no longer merely affected by AI from the outside. Media organisations are increasingly part of the AI value chain: they provide high-quality content and archives, experiment with AI-enabled workflows, license data to model providers, build application-layer tools, and participate in emerging governance debates around attribution, rights, and compensation.

Yet across the compute, model, and application layers, the same strategic question returns: **can media organisations govern the assets they contribute and capture a fair share of the value created, or will value continue to shift toward external infrastructure and model providers?**

This is important because AI dependency is not a neutral technical choice for media companies. As AI tools become embedded in newsroom workflows, archive systems, distribution, and audience products, the costs and conditions of access may change quickly: GitHub Copilot<sup>76</sup>, for example, is moving to usage-based billing from June 2026, illustrating how AI services that start as predictable subscription tools can shift toward consumption-based pricing. At the same time, larger publishers are better positioned to negotiate licensing deals or pursue legal action against AI companies, while smaller and regional media organisations often lack the legal, technical, and financial capacity to do so individually; even among journalism nonprofits, litigation against OpenAI remains exceptional. The mapping across the three layers shows additional uneven positioning of media companies in Europe.

- At the **compute layer**, Europe has developed a stronger public infrastructure route through EuroHPC (LUMI, JUPITER,...) VSC, and national AI Factory Antennas. This creates opportunities for sovereign experimentation, dataset processing, model adaptation, and fine-tuning. However, public compute does not remove dependency on commercial cloud. Many media applications require continuous deployment, uptime, integration, and low-latency inference. As a result, media organisations will likely operate in hybrid compute environments. The key governance question is therefore which workloads should run where, under which data terms, and with what implications for cost, sovereignty, IP control, and long-term dependency.
- At the **model layer**, Europe has produced a growing number of (public) sovereign and language-oriented initiatives, but media organisations remain only selectively involved. Most national model initiatives are led by research institutions, supercomputing centres, or public AI programmes, with media actors treated as future users or data providers rather than as co-owners or co-governors. The strongest exceptions are GPT-NL, NorLLM/Kopinor, and OpenGPT-X/Teuken. These cases show that media archives can be treated as strategic and monetisable assets, but only if governance, IP, licensing, and value-sharing are designed before or alongside model development. GPT-NL is the clearest example: publishers contributed rights-cleared Dutch-language content through a governed framework, received oversight through a Content Board, and participate in a revenue-sharing model. NorLLM/Kopinor adds a second route through state-mediated collective licensing, compensating publishers even when they are not full consortium members.
- At the **application layer**, European media organisations are most active and closest to value creation. Archive search and retrieval, newsroom productivity, and audience-facing products are already being tested or deployed across public broadcasters, commercial publishers, and cultural heritage institutions. This is the most realistic near-term entry point for media-sector AI because organisations do not need to train foundation models from scratch. They can create value by combining existing models with trusted archives, metadata, editorial workflows, user relationships, and human oversight. However, this strength is also partly dependent on external

---

<sup>76</sup> <https://github.blog/news-insights/company-news/github-copilot-is-moving-to-usage-based-billing/>

infrastructure. Many application-layer tools rely on US cloud providers or proprietary models. This means that the editorial layer may remain European, while compute, model access, data processing, and part of the economic value flow to external providers.

Across all three layers, one finding is consistent: **AI adoption alone does not guarantee strategic value for media organisations**. Value is captured by those who control the relevant assets, interfaces, infrastructure, rights, and business models. For media organisations, the challenge is therefore not simply to use AI, but to move toward governed value capture: retaining control over content, data, IP, workflows, and audience relationships while turning AI experimentation into sustainable operational or commercial value.

A second cross-cutting finding is that **local language and cultural specificity matter**. Generic models developed primarily by US-based companies may be powerful, but they are not optimised for smaller European languages, regional references, local political cultures, public-service traditions, or the specific formats and archives of national media systems. This creates a cultural and editorial dependency as well as a technical one. If European media organisations rely only on external models, they risk building AI tools that underrepresent local languages, flatten cultural differences, and weaken the distinctiveness of European media ecosystems. Sovereign and language-specific initiatives are therefore not only about infrastructure independence; they are also about preserving cultural relevance, editorial quality, and democratic visibility in AI-mediated information environments.

## Routes for governance and value creation

### 1. Bilateral licensing with frontier model providers

Publishers such as Axel Springer, Le Monde, Schibsted, and the Financial Times have licensed content to OpenAI or Google. This route creates direct revenue and confirms that media archives have measurable value as AI training and retrieval data. However, data governance and ownership remain difficult to assess because contract terms are often confidential. Key questions include whether content is used for training, retrieval, product integration, or all three; whether data can be reused in future models; whether publishers retain audit rights; and whether derived model value remains outside their control. This route tends to favour large media organisations with strong brands, large archives, legal capacity, and bargaining power. Smaller media players are less likely to negotiate favourable bilateral deals on their own, which risks widening the gap between large and small publishers.

### 2. Collective or sectoral licensing models

The Norwegian Kopinor arrangement shows how publishers can be compensated through a collective rights mechanism, even when they are not direct members of a model consortium. This route is particularly relevant for smaller and regional media organisations that lack the scale to negotiate directly with major AI providers. A collective model can pool rights, standardise licensing terms, clarify opt-in or opt-out conditions, and create a shared remuneration mechanism. From a data-governance perspective, this requires clear rules on which content is included, how consent or withdrawal is managed, how usage is documented, and how revenues are distributed. Collective licensing is, therefore, a way to reduce asymmetry between large platforms and fragmented media markets.

### 3. Collaborative sovereign model initiatives

GPT-NL demonstrates that publishers can contribute data to a national model while retaining oversight and participating in future revenue. This route requires strong upfront governance: clear contribution rules, IP arrangements, revenue-sharing, safeguards against content extraction, and long-term maintenance commitments. Data governance is central here. Contributors need to know what data is used, where it is processed, how it is documented, how it can be removed or updated, and what rights they retain over model outputs or derivative products. Ownership does not necessarily need to mean that each media

organisation owns the model itself, but it does require shared oversight, transparent decision-making, and a credible mechanism for distributing value. Such initiatives are also important for local-language and culturally specific AI, because they allow national or regional media ecosystems to shape the training data and evaluation criteria.

#### **4. Internal value creation through applications**

Many media organisations create value without external commercialisation. Archive search, newsroom assistants, transcription, summarisation, metadata enrichment, and workflow tools can generate productivity gains, improve archive access, strengthen editorial quality, and support better user engagement. Here, data governance concerns focus on operational control: which internal data is used, whether it is sent to external providers, how prompts and outputs are stored, whether confidential editorial material is retained by vendors, and how human oversight is organised. This route is attractive because it is practical and close to daily workflows. However, when each organisation develops tools in isolation, the result is a fragmented and siloed application landscape. This may limit learning, duplicate costs, and make smaller media organisations dependent on off-the-shelf tools they cannot meaningfully shape.

#### **5. Commercialisation of internal tools as B2B products or services**

Cases such as Reuters Connect, United Robots, Arc XP, and Sophi.io show that media-originated tools can sometimes become commercial products. This route can create durable revenue streams, but it is capital-intensive and not automatically successful. It requires product management, sales capacity, maintenance, customer support, and a clear market beyond the originating newsroom. From a data-governance and ownership perspective, this route requires careful separation between the original publisher's content, customer data, model infrastructure, and product IP. If a tool is trained or improved using multiple clients' data, rules are needed on data reuse, benchmarking, confidentiality, and derivative value. This route is more accessible to large organisations or specialised spin-outs than to smaller media players, unless developed through shared infrastructure or sector-wide partnerships.

#### **6. Governed access, attribution, and monetisation layers**

Tools and arrangements around attribution, monitoring, licensing, and compensation are becoming more important as AI shifts from one-off training to continuous retrieval. TollBit<sup>77</sup>, ProRata, Perplexity publisher partnerships, and emerging provenance systems point toward a future in which media organisations need to govern not only training access, but also repeated AI access to live and archived content. Data governance here means defining who can access content, how often, under what licensing terms, with what attribution, and with which reporting or audit obligations. This route is especially relevant for smaller media organisations, because shared attribution and licensing infrastructure could help them participate in AI markets without individually negotiating with every AI provider.

This last point becomes **especially important in the context of agentic AI**. As AI systems become capable of searching, retrieving, summarising, and acting across digital services, the economic issue shifts from one-time use of content for training to continuous operational access. RAG-based and agentic systems may repeatedly retrieve journalistic content without sending traffic, attribution, or compensation back to publishers. This changes the governance problem: media organisations must define who can access their content, under what conditions, how often, with what attribution, and against which compensation model. Without such mechanisms, the value of journalism may increasingly be captured at the level of AI interfaces rather than by the organisations producing the content.

The central lesson is therefore clear: **the most successful initiatives are not simply those with the strongest technology, but those that align technology with governance, IP, data ownership, and value-**

---

<sup>77</sup> <https://digiday.com/media/the-case-for-and-against-publisher-content-marketplaces/>

**sharing from the start.** Initiatives that resolve these questions early are better positioned to reach deployment, maintain trust, and create sustainable value. Initiatives that treat governance as a later problem risk stalling, fragmenting, or leaking value to external providers.

**Future media-sector AI initiatives should therefore move beyond siloed experimentation.** Many European media organisations are already building useful tools, but mostly within their own institutional boundaries. This benefits larger players with the capacity to experiment, negotiate, and integrate AI systems, while smaller players risk being left behind. A more sustainable path requires collective approaches: shared licensing mechanisms, common data-governance standards, interoperable archive infrastructures, joint procurement or compute access, and collaborative application layers where appropriate. The strategic opportunity for Europe’s media sector is not only to adopt AI faster, but to organise collectively around the assets that make media valuable in the first place: trusted content, local languages, cultural knowledge, editorial standards, and long-term public legitimacy.