

# Creative Value Monitor

[creativevaluemonitor.ai](https://creativevaluemonitor.ai)

Positioning Paper | v0.4 | June 2026

Creative Value Monitor (CVM) measures the value that creative work contributes to AI. It tracks how much creative work goes into AI systems and how much value those systems draw from it, the inputs and the outputs. The results are published as an index that can be followed over time.

AI models are built from creative work, from books and journalism to music, images, code, and the open web, yet there is no agreed way to say which work contributed to AI, how much, or what it is worth. No attribution means no remuneration for creators, and as remuneration falls away, so does the incentive to keep producing the work that AI most depends on.

Current policy and market conditions create the opening to correct this. Policy in the EU and the UK is moving towards the transparency and disclosure obligations that make independent measurement both necessary and possible for the first time, and the case for it runs deeper than fairness to creators. AI developers need a healthy creative economy to keep building good models, which makes a working market a shared interest rather than a transfer from one industry to another: creators are paid for what their work contributes, and developers keep access to the material that most improves their systems.

The CVM method tracks both the inputs and the outputs: on the input side, the presence and influence of creative works within training data; on the output side, how often and how influentially those works are used when AI systems retrieve, cite, summarise, and answer. It draws on attribution research that has only recently become feasible at the scale of a large model, combined with the statistical approach similar to what collecting societies use, sampling usage and estimating usage in proportion to contribution.

The CVM Index, a public, trackable measure of the value AI systems draw from creative work over time, will be published at [creativevaluemonitor.ai](https://creativevaluemonitor.ai) as a resource for creators and policy makers: creators to evidence how their work contributes to and is used by AI, policy makers to inform and ground enforcement and remuneration policy.

## Creative works are the missing line item for AI developers

In the first half of 2026, the companies building frontier AI are coming to the public markets at a scale without precedent. SpaceX has priced the largest listing in history, with xAI folded in ([Fortune](#)). Anthropic filed confidentially on 1 June at a valuation of \$965 billion ([Fortune](#)). OpenAI followed on 8 June, targeting up to \$1 trillion ([TechCrunch](#)). Together the three carry a combined valuation of roughly \$3.6 trillion, and their listings could draw around \$200 billion from public investors ([TheStreet](#)).

Every one of the models the developers built was built using creative work. Books, journalism, music, images, code, and the open web are the raw material of a large language model, and the quality of that material is what separates a capable model from a weak one. The value being raised this year is, in large part, value extracted from the work of writers, artists, journalists, musicians, and the institutions that support them.

The value and cost of that work appears nowhere on the books. Compute and infrastructure are counted to the dollar. The creative work the products are made from is treated as a low-cost externality, and there is no agreed way to measure what any of it is worth. This paper sets out the problem, the opening that current policy and market conditions create, the research that makes a solution feasible, and a proposed solution to track the value of creative works in AI called the Creative Value Monitor.

## **The problem: no attribution, no payment, on either side of the model**

The economic relationship between AI and creative work is broken in the same way on both sides of the model: at training, where work is used as a training input, and at inference, where work is used to produce answers as an output. In neither place is there currently an implemented method to say which work contributed, how much, or what it should be paid. Without that, there is no basis on which any creators could be remunerated for their contribution to AI.

On the training side, the numbers that exist are revealing. Anthropic agreed to pay \$1.5 billion to settle a class action over pirated books, the largest copyright recovery in United States history. It works out at about \$3,000 for each of roughly 500,000 eligible works, paid once, drawn from a library that the court found to contain more than seven million pirated books. The same court ruled that training a model on books is fair use, and that the violation lay in how the books were acquired, not in the act of training ([NPR](#); [The Guardian](#)). The practical signal to creators is therefore stark: copyright enforcement is not going to deliver them a continuing income from AI. Set against Anthropic's valuation, the settlement is around 0.15%, or roughly twelve days of revenue at the company's current run rate.

The licensing market that has formed alongside the litigation is small and concentrated. One widely cited tracker puts the entire disclosed market for AI content licensing at roughly \$800 million a year, almost all of it flowing to a handful of large rights holders ([Media & the Machine](#)). Google reportedly pays Reddit around \$60 million a year to train on posts that Reddit's users wrote ([CBS News](#)). OpenAI's deal with News Corp is reported at \$250 million over five years ([WSJ](#)). The individual creator, whose work sits inside the same training corpora, receives nothing effectively because there is no mechanism that could identify their contribution and pay them for it.

On the inference, output, side, as AI systems answer questions, summarise sources, and act on a user's behalf, they break the link between a reader and the page that a work lives on. By early 2026, analyses of AI-driven search were reporting steep falls in the traffic that funds publishing ([Ahrefs. via Hindustan Herald](#); [Eyeful Media](#)). One February 2026 study found that Google's AI Overviews were associated with a fall of around 58 % in click-through rates

for the top organic results, almost double the figure a year earlier. Other analyses put referral traffic from AI search engines at a small fraction of what traditional search once sent, and traffic to the largest publishers down by a quarter or more year on year. Here too, creative work is being used to generate AI responses and is generating value for developers, and here too there is no measurement and no payment back to the creator that informed the output from AI.

The common thread is the absence of measurement. There is no shared, independent way to establish what a given creative work contributes to an AI system, either as training input or as material used at inference. Until that exists, attribution is impossible, remuneration has nothing to attach to, and creators negotiate blind with developers who hold all the information.

## The opportunity

Three forces are converging on a short window, and together they make an independent measurement layer both necessary and timely.

### Policy is moving toward disclosure and remuneration

Regulators in the UK and Europe are now requiring exactly the kind of evidence that does not yet exist.

In Europe, the European Parliament adopted a resolution on copyright and generative AI on 10 March 2026, led by the rapporteur Axel Voss and passed by 460 votes to 71 with 88 abstentions ([European Parliament resolution](#); [report](#)). It calls for an itemised list identifying each copyright-protected work used in training, extends that transparency requirement to inference, retrieval-augmented generation, and fine-tuning, and asks that transparency cover the act of crawling itself. It calls for fair remuneration of creators, including for past use, for a European register held at the EU Intellectual Property Office, and for a new licensing market built on voluntary collective agreements accessible to individual creators and small enterprises. It states plainly that a single flat-rate global licence will not be sufficient. The EU AI Act reinforces the direction: obligations on providers of general-purpose AI models, including a sufficiently detailed summary of training content, have applied since August 2025, and the Commission's enforcement powers over those providers, with fines of up to 3 % of global turnover, apply from 2 August 2026 ([EU AI Act Service Desk](#); [European Commission](#)).

In the UK, the Competition and Markets Authority used its powers under the Digital Markets, Competition and Consumers Act 2024 to designate Google with Strategic Market Status and, on 3 June 2026, imposed binding Publisher Conduct Requirements on Google Search. Publishers can now opt out of having their content used in AI Overviews and to fine-tune Google's AI models, Google must provide clear attribution through prominent links, and the measure is explicitly intended to strengthen publishers' ability to negotiate licensing. The CMA called it a world first ([Digital Watch Observatory](#); [VideoWeek](#)). Separately, the UK Government published its Report on Copyright and AI on 18 March 2026 but set out no preferred policy option, stating that it will not legislate until it has a workable solution for transparency, rights-holder control, and legal certainty ([GOV.UK](#)).

Every one of these measures, transparency, attribution, opt-out, fair remuneration, collective licensing, presupposes an economic evidence base that nobody currently provides.

## **A market-based solution to a market-based problem**

The same period has made clear that the destination is a functioning market, not litigation. The legal routes differ by jurisdiction, but they point the same way. In the United States, courts are finding that training on lawfully obtained work is fair use, which removes copyright enforcement as a route to ongoing creator income ([The Guardian](#)). Europe has no fair use doctrine: training is governed by the text and data mining exception and its opt-out under the 2019 copyright directive, the framework within which the Parliament's resolution operates ([EUR-Lex](#)). The United Kingdom has declined to settle the question until a workable solution for transparency, rights-holder control, and legal certainty exists ([GOV.UK](#)). In none of the three does enforcement deliver a continuing income to creators. The European Parliament has rejected a flat global licence in favour of a market with collective licensing ([European Parliament](#)). And the most concrete proposal on the table has come from an AI developer, not a rights-holder lobby.

In a Financial Times opinion piece on 20 March 2026, Mistral's chief executive Arthur Mensch proposed a revenue-based levy on the European revenues of every company commercialising AI in Europe, including American and Chinese firms, with the proceeds flowing into a central European fund for cultural production, in exchange for legal certainty over training on publicly available content. The company subsequently put the rate at 1 to 1.5 % of revenue ([AFP, via losco News](#)). Mistral's illustration: a vendor earning 5 billion euros in European AI revenue would contribute 50 to 75 million euros a year, and one earning 20 billion would contribute 200 to 300 million. The proposal draws on the existing private copying levy and frames creators and AI developers as allies rather than adversaries ([Financial Times](#); [Liora](#)).

*Whether the mechanism is collective licensing, micropayments, or a levy, every one of these market designs needs the same thing to function: a way to price contribution. A licence has to be valued. A levy has to be distributed. A pool has to be split. None of that is possible without measurement.*

## **Continued access to quality creative work is a problem for developers**

The case for paying creators is usually made as a matter of fairness. The stronger case, and the one that makes this a positive-sum proposition, is that AI developers need a healthy creative economy to keep building good models.

The supply of human-written text is finite. Epoch AI estimates the effective stock of quality public human text at around 300 trillion tokens and projects that language models will fully utilise it between 2026 and 2032 ([Epoch AI](#)). Synthetic data does not close the gap: it has been shown to help reliably only in narrow, verifiable domains such as mathematics and code ([Epoch AI](#)). The AI developers themselves now concede the point. In an Atlantic interview in April 2026, OpenAI's Sam Altman accepted that reasoning ability can be developed from synthetic data, but that human values and human culture require human data ([The Atlantic, via Phemex](#)). The cultural and creative domain is therefore the one place

where fresh human work remains essential. Model-collapse research underlines the risk: training models recursively on their own output degrades capability and erases the diversity in the tail of the distribution, which is exactly where original creative work lives ([A Tale of Tails, ICML 2024](#)).

The implication is direct. The marginal value of human creative work is rising at the same moment as the economics that sustain its production are being eroded. If creators withdraw, lock their work away i.e. opt-out, or stop producing, the supply of the material AI most needs degrades with them. *A functioning creator economy is not a charitable concern for the AI industry, it is the industry's supply chain.*

## Creative contributions to AI can now be measured

Measuring the contribution of creative works to AI development is the subject of an active and fast-maturing research field.

The foundational technique is the influence function, introduced to modern machine learning by Koh and Liang in 2017, which estimates how a model's behaviour would change if a given training example were removed, without retraining the model ([Koh and Liang, 2017](#)). The barrier was computational cost at the scale of a large language model. That barrier has fallen. Anthropic's own researchers (Grosse and colleagues, 2023) adapted an efficient approximation to apply influence functions to models with tens of billions of parameters ([Grosse et al., 2023](#)). The TRAK method (Park and colleagues, 2023) attributes model behaviour at scale using projected gradients ([Park et al., 2023](#)). Most directly relevant, In-Run Data Shapley (Wang and colleagues, 2024) made it possible, for the first time, to attribute a model's behaviour to its pretraining data at negligible additional cost during training, and reported a finding that matters greatly for this work: a well-written work can measurably improve a model even when the model never reproduces a line of it ([Wang et al., 2024](#)). A small amount of strong material teaches a model more than a large volume of weak material, which is exactly why the labs pursued books and quality archives. More recent work presented at NeurIPS 2025 has scaled influence-function-based data valuation to current LLMs and frames its purpose explicitly around the fact that data providers remain uncredited ([NeurIPS 2025](#)).

The two sides of the model differ in how they can be measured. Inference contribution is observable from the outside today: retrieval frequency, citation rates, summary inclusion, and traffic displacement can all be measured without access to a model's internals, and regulation is beginning to mandate the underlying data. The CMA's conduct requirement obliges Google to provide publishers with disaggregated engagement metrics, including impressions and click-through rates, and extends its controls to grounding as well as fine-tuning ([PPC Land](#)). Training contribution is partly observable too: the presence and volume of works within documented corpora such as Common Crawl can be estimated from the outside. Influence is the measure that requires access to the model or to the training run itself, and frontier developers protect both closely; an independent monitor should not assume that access. The influence methodology can instead be developed and validated on open-weight models, where internals are fully accessible, and applied to frontier systems as the disclosure obligations described earlier come into force and as developer cooperation develops. The policy regime therefore does two jobs at once: it creates the demand for

independent measurement, and it progressively creates the access that training-side measurement needs.

Two precedents show how a measurement standard becomes the foundation of a market. When carbon accounting began, no one could measure a carbon footprint precisely either, but the organisations that defined the methodology, principally the Greenhouse Gas Protocol, became the infrastructure layer for the entire carbon market; the platforms, reporting tools, and trading systems came afterward ([GHG Protocol](#)). The methodology was the product. And the principle of distributing revenue from statistical evidence of usage is not new: collecting societies such as PRS for Music, GEMA, SACEM, and ALCS have done it at scale for decades, sampling usage and distributing proportionally rather than tracking every play. ALCS alone has distributed around 750 million pounds to roughly 130,000 writers since 1977 ([ALCS](#)).

*Attributing the contribution of creative work to AI systems is technically feasible, at the scale that matters, including at the pretraining stage and including cases where nothing is copied. What is missing is not the science. It is an independent body doing the measurement, consistently and in public.*

## **The proposal: the Creative Value Monitor**

The Creative Value Monitor (CVM) is a measurement layer for the AI content economy. It provides a neutral, public benchmark for the value of creative work to AI.

CVM does not take a position on which legal or commercial mechanism should govern the AI-creator relationship.

CVM monitors and publishes data on both sides of the model.

- Training contribution: the presence, volume, uniqueness, domain authority, temporal value, and influence of creative works and sources within training corpora.
- Inference contribution: how often and how influentially creative works are used when deployed AI systems retrieve, summarise, cite, answer, and act, including retrieval frequency, summary-inclusion rate, citation rate, traffic displacement, and authority within agent-driven systems.

These measures are combined into the Creative Value Monitor Index (the CVM Index), a published, trackable measure of how creative value is extracted across AI systems over time.

CVM measures the economic contribution of creative works to AI.

The first phase of CVM's work is to develop and publish the methodology and the data, as an independent, public evidence base.

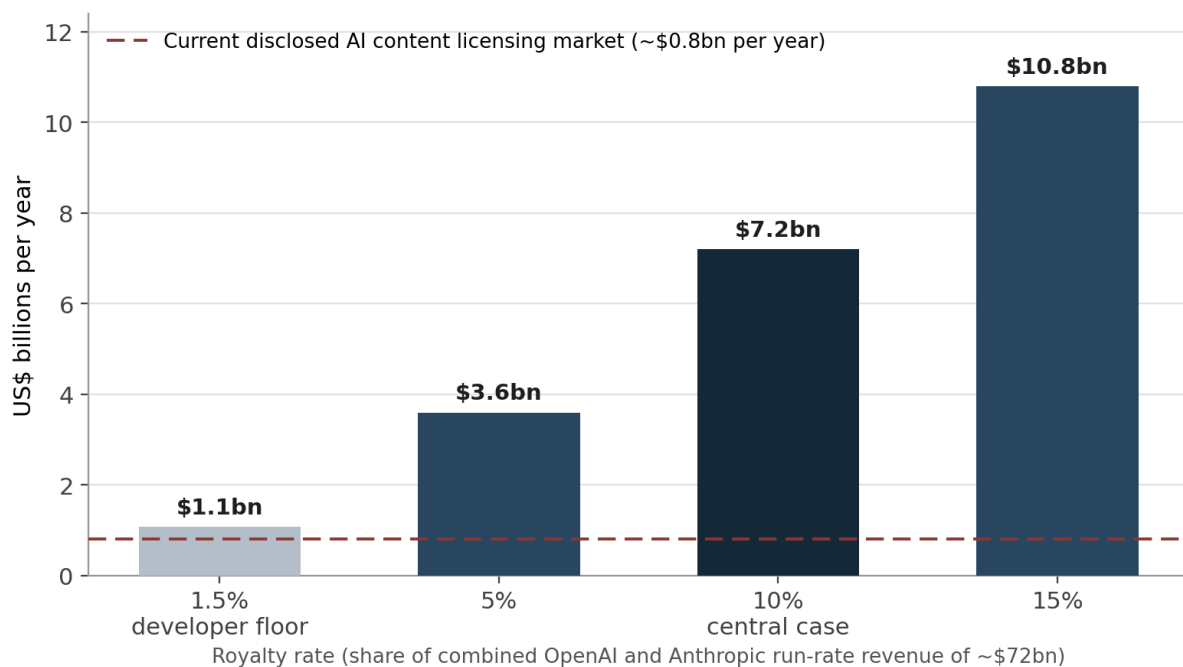
Payment infrastructure, the rails that would turn measurement into money in creators' hands, may be considered as a later phase and is not part of this proposal.

The immediate value is that an independent measurement layer would be useful to policy makers, publishers and AI developers who want to better understand the economic contribution of creative works to AI.

## The size of the market that measurement would serve

What should AI developers pay for the creative work their products are built from? Comparable markets provide the anchors. Music streaming services pass roughly 70 % of revenue to rights holders because the content is the product. Trade publishers pay authors 10 to 15 %. A United States radio blanket licence runs at a low single-digit percentage of revenue. AI sits between these: its output is transformed, but the model is built from the work and competes with it in the same markets. A defensible band is therefore 5 to 15 % of revenue, with 10 % as the central case. The only developer-authored number sits below that band: Mistral's proposed levy of 1 to 1.5 %.

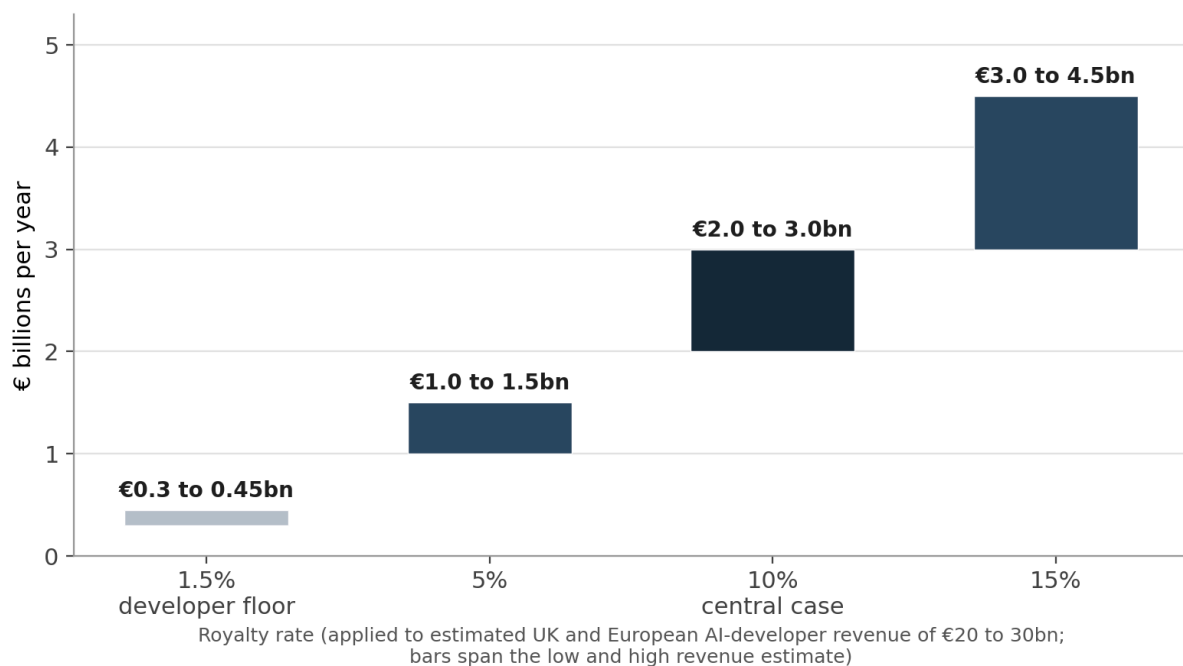
OpenAI and Anthropic alone are turning over roughly \$72 billion a year between them at current run rates. A pool drawn from that revenue at 5 to 15 % would run from around \$3.6 billion to nearly \$11 billion a year globally, against a disclosed licensing market of roughly \$800 million today ([Media & the Machine](#)). The 10 % central case is about \$7.2 billion a year, close to nine times the entire current market. Even the developer-proposed floor of 1 to 1.5 % would produce \$0.7 to \$1.1 billion a year, on its own roughly the size of the entire market that exists today.



*Figure 1. Global annual creator remuneration pool by royalty rate, applied to OpenAI and Anthropic combined run-rate revenue of about \$72 billion, against the current disclosed AI content licensing market of about \$0.8 billion. The 1.5 % bar marks the developer-proposed levy floor.*

CVM is built for the UK and European markets, where the policy momentum is strongest, so the relevant figure is the value generated in those markets. Companies rarely disclose regional AI revenue, so this can only be estimated, but the order of magnitude is clear. Europe's generative AI market was estimated at upward of \$16 billion in 2025, with the UK the largest single national share, and that figure is growing fast ([MarketsandMarkets](#)).

Applying the same band to a UK and European AI-developer revenue base in the region of 20 to 30 billion euros implies an annual pool of roughly 1 to 4.5 billion euros, with the 10 % central case at roughly 2 to 3 billion euros a year and the developer-proposed floor at roughly 200 to 450 million euros.



*Figure 2. UK and Europe annual creator remuneration pool by royalty rate, applied to an estimated UK and European AI-developer revenue base of 20 to 30 billion euros. Bars span the low and high revenue estimate. The 1.5 % bar marks the developer-proposed levy floor.*

These figures describe the cost of a supply chain, not a transfer from one industry to another. The benchmarks above are what comparable industries pay for the inputs their products are built from, and in each case the paying industry remains profitable and the supplying industry stays in production. As set out earlier, original human creative work is the input AI most needs and cannot reliably synthesise; a pool of this order is what keeping that supply in production would cost. Both sides gain: developers secure continued access to the material that most improves their models, and creators are paid in proportion to what their work contributes.

For context, the UK creative industries contributed 145.8 billion pounds in gross value added in 2024 ([DCMS, via Bristol Creative Industries](#)) and are growing at around four times the rate of the wider economy. The point of CVM is to make sure that, as AI is built on this work, the value can be seen, measured, and eventually returned.

# What we are building, and the next step

The first phase of the Creative Value Monitor is to develop the methodology and publish the data, establish the independent evidence base and the credibility for the methodology.

It positions the measurement layer at the centre of the market that forms around it.

We are convening a small group of partners across cultural institutions, academic research, regulatory bodies, and the creative industries to pressure-test the methodology and shape the first published outputs.

CVM is a Creative Thinking Systems initiative as an applied AI research and development company working at the intersection of artificial intelligence and the cultural and creative industries. CVM is one of several initiatives through which Creative Thinking Systems works to enhance, incentivise, and protect human creativity in the age of AI.

Conor Roche, Founder. [conor@creativethinkingsystems.com](mailto:conor@creativethinkingsystems.com)

---

## References

### AI listings, valuations, and revenue

1. SpaceX IPO pricing. Fortune, 3 June 2026.  
<https://fortune.com/2026/06/03/spacex-ipo-share-price-index-funds-valuation-public/>
2. Anthropic confidential IPO filing at a \$965 billion valuation. Fortune, 1 June 2026.  
<https://fortune.com/2026/06/01/anthropic-confidentially-files-ipo-965-billion-valuation/>
3. Anthropic IPO, valuation, and run rate. CNBC, 5 June 2026.  
<https://www.cnbc.com/2026/06/05/tech-download-anthropic-ipo-ai-valuations.html>
4. OpenAI confidential IPO filing, targeting up to \$1 trillion. TechCrunch, 8 June 2026.  
<https://techcrunch.com/2026/06/08/following-anthropic-openai-files-confidentially-for-ipo/>
5. OpenAI revenue run rate. Reuters, via Yahoo Finance.  
<https://finance.yahoo.com/news/openai-cfo-says-annualized-revenue-173519097.html>
6. Combined valuation of the three AI listings. TheStreet, 10 June 2026.  
<https://www.thestreet.com/investing/stocks/openai-makes-ipo-decision-amid-anthropic-spacex-fervor>

### Settlement and the licensing market

7. Anthropic to pay authors \$1.5 billion (largest copyright recovery; about \$3,000 per work for roughly 500,000 works). NPR, 5 September 2025.  
<https://www.npr.org/2025/09/05/g-s1-87367/anthropic-authors-settlement-pirated-chat-bot-training-material>
8. What authors need to know about the Anthropic settlement (eligibility; roughly 500,000 of 7 million). Authors Guild, April 2026.

- <https://authorsguild.org/advocacy/artificial-intelligence/what-authors-need-to-know-about-the-anthropic-settlement/>
9. Anthropic settlement and the fair-use ruling on training. The Guardian, 5 September 2025.  
<https://www.theguardian.com/technology/2025/sep/05/anthropic-settlement-ai-book-lawsuit>
  10. Google and Reddit licensing deal, reported at around \$60 million a year. CBS News.  
<https://www.cbsnews.com/news/google-reddit-60-million-deal-ai-training/>
  11. OpenAI and News Corp deal, reported at \$250 million over five years. Wall Street Journal. <https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba>
  12. Disclosed AI content licensing market, estimated at roughly \$800 million a year. Media & the Machine, June 2026.  
<https://mediaandthemachine.substack.com/p/ai-content-licensing-deals-june-2026>

### **Inference and the traffic economy**

13. Google AI Overviews and the fall in publisher click-through rates (Ahrefs study, February 2026). Reported via Hindustan Herald, 3 June 2026.  
<https://hindustanherald.in/google-ai-publisher-rights-uk-cma/>
14. Referral traffic decline from AI search engines. Eyeful Media analysis.  
<https://www.eyefulmedia.com/blog/pay-per-crawl>

### **EU policy**

15. European Parliament resolution on copyright and generative AI, 10 March 2026 (adopted texts, TA-10-2026-0066).  
[https://www.europarl.europa.eu/doceo/document/TA-10-2026-0066\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-10-2026-0066_EN.html)
16. Report on copyright and generative AI, rapporteur Axel Voss (A10-0019/2026), including itemised disclosure extended to inference and crawling. European Parliament.  
[https://www.europarl.europa.eu/doceo/document/A-10-2026-0019\\_EN.html](https://www.europarl.europa.eu/doceo/document/A-10-2026-0019_EN.html)
17. Summary of the resolution and the vote (460 to 71, 88 abstentions). Council of Europe / Merlin. <https://merlin.obs.coe.int/article/10499>
18. EU AI Act: general-purpose AI obligations and the August 2026 enforcement powers (fines up to 3 % of global turnover). EU AI Act Service Desk.  
<https://ai-act-service-desk.ec.europa.eu/en/faq>
19. AI Act transparency rules and GPAI obligations. European Commission, Shaping Europe's Digital Future.  
<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
20. Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market (text and data mining exceptions, Articles 3 and 4). EUR-Lex.  
<https://eur-lex.europa.eu/eli/dir/2019/790/oj>

### **UK policy**

21. CMA imposes binding Publisher Conduct Requirements on Google (opt-out of AI Overviews and fine-tuning; attribution; world first). Digital Watch Observatory, 3 June 2026.

- <https://dig.watch/updates/uk-cma-targets-ai-search-content-use-in-new-google-conduct-requirements>
22. Publishers gain opt-out from Google's AI features following CMA intervention. VideoWeek, 3 June 2026.  
<https://videoweek.com/2026/06/03/uk-publishers-gain-opt-out-from-googles-ai-features-following-cma-intervention/>
  23. Publisher Conduct Requirement detail: controls across AI Overviews, AI Mode, Gemini, and Vertex AI; engagement metrics obligation; main obligations in force 3 December 2026. PPC Land, June 2026.  
<https://ppc.land/uk-regulator-forces-google-to-give-publishers-ai-opt-out-rights-today/>
  24. UK Government Report on Copyright and Artificial Intelligence, 18 March 2026. GOV.UK.  
<https://www.gov.uk/government/publications/report-and-impact-assessment-on-copyright-and-artificial-intelligence/report-on-copyright-and-artificial-intelligence>

### **Market-based proposals**

25. Mistral's proposed European AI revenue levy for a central cultural fund. Financial Times, 20 March 2026.  
<https://www.ft.com/content/d63d6291-687f-4e05-8b23-4d545d78c64a>
26. Mistral levy, illustrative European figures and the private copying levy precedent. Liora, 24 March 2026. <https://liora.io/en/mistral-ai-proposed-levy-european-creators>
27. Mistral levy rate confirmed at 1.0 to 1.5 % of revenues. AFP, via losco News.  
[https://www.iosconews.com/news/nation/article\\_71f533a6-9578-54fb-8233-5ad5e50f6f7d.html](https://www.iosconews.com/news/nation/article_71f533a6-9578-54fb-8233-5ad5e50f6f7d.html)

### **Data limits and model collapse**

28. Will we run out of data to train large language models? (about 300 trillion tokens; full utilisation 2026 to 2032; synthetic data reliable only in narrow domains). Epoch AI.  
<https://epoch.ai/publications/will-we-run-out-of-data-limits-of-llm-scaling-based-on-human-generated-data>
29. Sam Altman on synthetic data: reasoning yes, human values and culture require human data. The Atlantic interview, reported by Phemex, April 2026.  
<https://phemex.com/news/article/altman-hints-at-ai-models-trained-on-synthetic-data-77495>
30. A Tale of Tails: Model Collapse as a Change of Scaling Laws. ICML 2024.  
<https://arxiv.org/pdf/2402.07043>

### **Attribution research**

31. Koh and Liang, Understanding Black-box Predictions via Influence Functions. ICML 2017. <https://proceedings.mlr.press/v70/koh17a.html>
32. Grosse and colleagues (Anthropic), Studying Large Language Model Generalization with Influence Functions. 2023. <https://arxiv.org/abs/2308.03296>
33. Park and colleagues, TRAK: Attributing Model Behavior at Scale. 2023.  
<https://arxiv.org/abs/2303.14186>

34. Wang and colleagues, In-Run Data Shapley (pretraining-stage attribution; contribution without memorisation; copyright implications). 2024. <https://arxiv.org/abs/2406.11011>
35. What is Your Data Worth to GPT? LLM-Scale Data Valuation with Influence Functions. NeurIPS 2025. <https://openreview.net/forum?id=zPKeJAEo27>

### **Precedents and royalty benchmarks**

36. Greenhouse Gas Protocol, the most widely used greenhouse gas accounting standard (WRI and WBCSD). GHG Protocol. <https://ghgprotocol.org/about-us>
37. Music streaming payouts to rights holders. Spotify Loud & Clear. <https://loudandclear.byspotify.com/>
38. Author royalty rates in trade publishing. The Bindery Agency. <https://www.thebinderyagency.com/articles/royalties-how-do-publishers-pay-authors>
39. US radio music royalty rates (BMI blanket licence). Music Business Worldwide. <https://www.musicbusinessworldwide.com/confirmed-bmi-royalty-rate-from-terrestrial-us-radio-is-rising-by-nearly-24/>
40. ALCS distributions to writers since 1977. ALCS. <https://www.prnewswire.co.uk/news-releases/alcs-annual-awards-winners-announced-at-2026-ceremony-302699546.html>

### **Creative economy and AI market scale**

41. UK creative industries gross value added, £145.8 billion in 2024. DCMS figures, reported by Bristol Creative Industries, February 2026. <https://bristolcreativeindustries.com/new-funding-and-support-for-creative-businesses-creative-industries-growth-four-times-overall-economy/>
42. Europe generative AI market size and UK share, 2025. MarketsandMarkets. <https://www.marketsandmarkets.com/Market-Reports/europe-generative-ai-market-48345665.html>