IAPP **AI Governance Global Europe 2026**

Training 1-2 June
Workshop 2 June
Conference 3-4 June
DUBLIN

#IAPPAIGG26

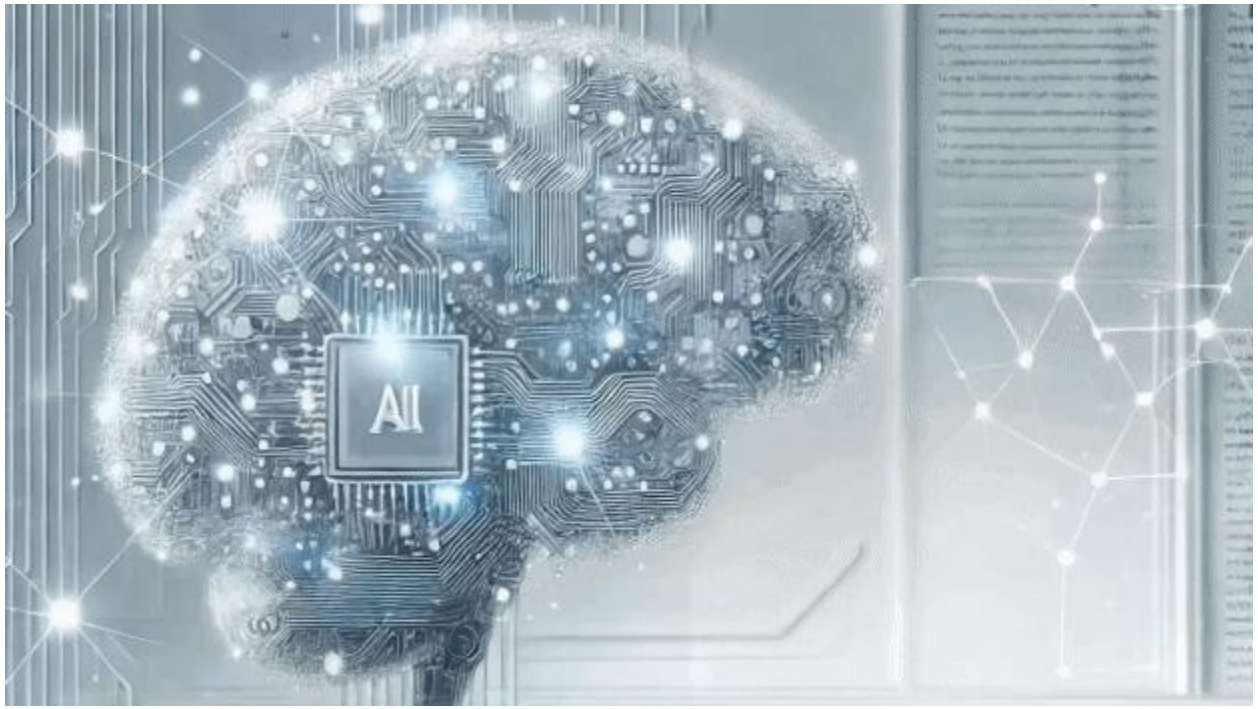
The Impact of AI on Copyright Law

Dr. Barry Scannell – Partner, William Fry

Barry.Scannell@WilliamFry.com

T + 353 1 639 5393

Setting the Scene: AI and Copyright in Flux



Setting the Scene: AI and Copyright in Flux

- Generative AI systems rely on massive datasets - often scraped without licences.
- Copyright law was not designed with machine learning in mind.
- The central tension: AI innovation vs. the rights of creators.
- This presentation unpacks the EU's response – and why it matters globally.


WILLIAM FRY

AI Authorship: Who is the Creator?

- EU and Irish law: only humans can be authors.
- Most jurisdictions have a human authorship requirement.
- US stance: *Zarya of the Dawn* – AI-generated images not protected.
- *Thaler v. Perlmutter*: courts reject AI authorship – reaffirm human creativity.
- China: 2023 Beijing ruling upholds copyright where human input is “creative” – hybrid model.

AI Authorship: Who is the Creator?

- In September 2022, artist and AI expert Kris Kashtanova (KK) was granted a copyright was informed by the USCO that the regiregistration for the graphic novel, Zarya of the Dawn, by the US Copyright Office (USCO). But then, at the end of October, KK stration may be cancelled. The reason? The images in the graphic novel were created using the text-to-image AI system Midjourney.

- The USCO to provide details of her process “to show that there was substantial human involvement in the process of creation of this graphic novel”.

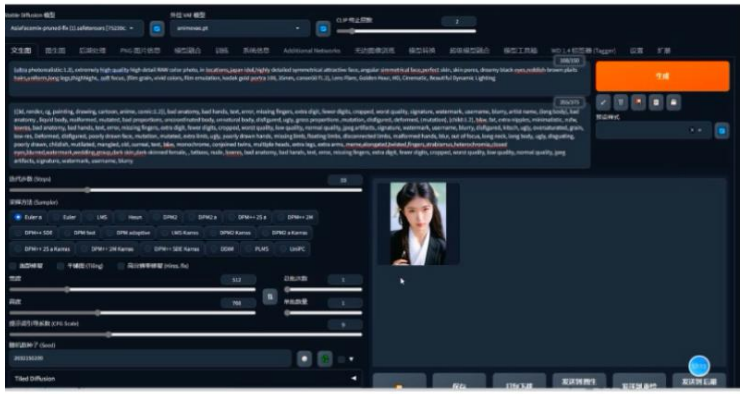
- Under US law, only a human can be an author. The USCO is currently defending itself against a federal lawsuit filed earlier this year by Stephen Thaler. Last year Thaler failed to have an AI system named as author for an AI generated image for which he sought a registration, under the doctrine of work for hire.

AI Authorship: Who is the Creator?

- **Case Background**: Stephen Thaler sought copyright registration for a visual art piece generated by his AI system, the "Creativity Machine."
- **Court Decision**: The Court reaffirmed that human authorship is required for copyright protection, denying Thaler's application.
- **Legal Arguments**: Thaler's common law property and work-for-hire doctrines were dismissed, as these presume the existence of a valid copyrightable work.
- **Impact**: The ruling underscores that copyright serves as an incentive for human creativity, setting a precedent as AI becomes more involved in creative processes.

AI Authorship: Who is the Creator?

- **Beijing Internet Court Ruling**: Affirmation of copyright protection for AI-generated images, setting a precedent in recognising AI-assisted creative works as original intellectual property.
- **Mr. Li's Case**: The court validated the creative decisions made by the AI operator, from conceptualisation to final selection, as a basis for copyright.
- **Criteria for Originality:** Emphasis on the role of human intellectual investment in designing character presentations, selecting prompts, and fine-tuning parameters in AI-generated works.
- **Definition of Artistic Works:** The court classified AI-generated images as artistic works with aesthetic significance, eligible for copyright under Chinese law.
- **Divergence:** This contrasts with the U.S. approach, which requires substantial human involvement and does not recognise AI as an author.



（图 2）

（图 3）

6. 在上述参数不变的情况下，将“Additional-Networks”中的模型 lord-hanfugirl-v1-5.safetensors”的权重修改为 0.75。生成结果如图 4 所示。

The Real Issue: Text and Data Mining

- Generative models ingest billions of works – often without consent or licence.
- This implicates copyright in the training phase.
- Key questions:
  - Is unlicensed scraping lawful?
  - Are opt-outs enforceable?
  - Who bears the burden of compliance?

The Real Issue: Text and Data Mining

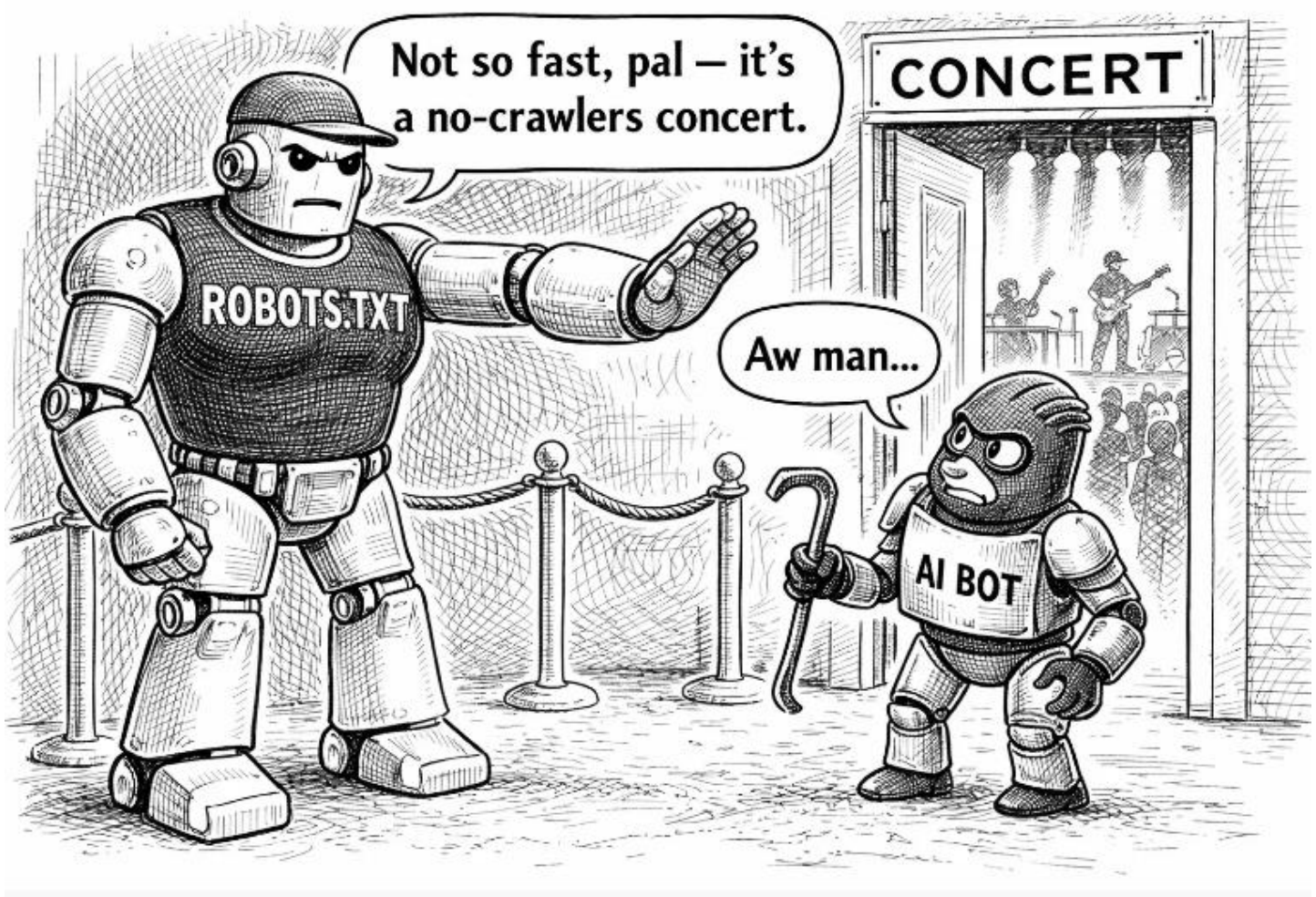
- LAION-5B is one of the largest text/image datasets available today. It has been used by myriad companies to create deep learning models. One such deep learning model is called Stable Diffusion - on which new AI apps such as Lensa AI rely.
- LAION-5B is a dataset of 5.85 billion image-text pairs, which is 14 times larger than LAION-400M, the previous biggest openly accessible image-text dataset in the world.
- LAION provides the dataset under a Creative Commons licence which they say poses no particular restrictions. However, LAION recommends that the images should only be used for research purposes - and they also point out that the images themselves are under copyright.

Text and Data Mining Copyright Exceptions

EU's Answer: Article 4 of the Copyright Directive

- Directive (EU) 2019/790 – CDSM Directive.
- Article 4(1): commercial text and data mining (TDM) is permitted with **lawful access**.
- Article 4(3): rightsholders may opt out by "**appropriate means**" (robots.txt, metadata).
- Creates an enforceable mechanism to exclude works from AI training.

What “Appropriate Means” Looks Like



- Machine-readable: robots.txt, HTTP headers, embedded metadata
- Explicit clauses in licensing terms
- Consideration of the Laion Case

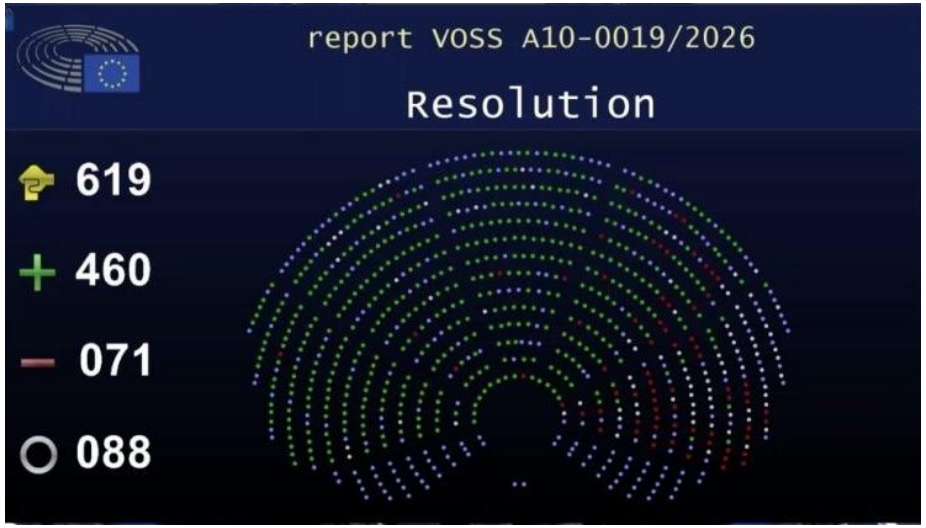
Control begins with infrastructure.

Why the EU's TDM Provisions are a Disaster

- No central opt-out registry.
- No remedy for past infringement.
- Enforcement is jurisdictionally fragmented and technically difficult.
- Still: Article 4(3) is the EU's core rightsholder protection against AI ingestion.

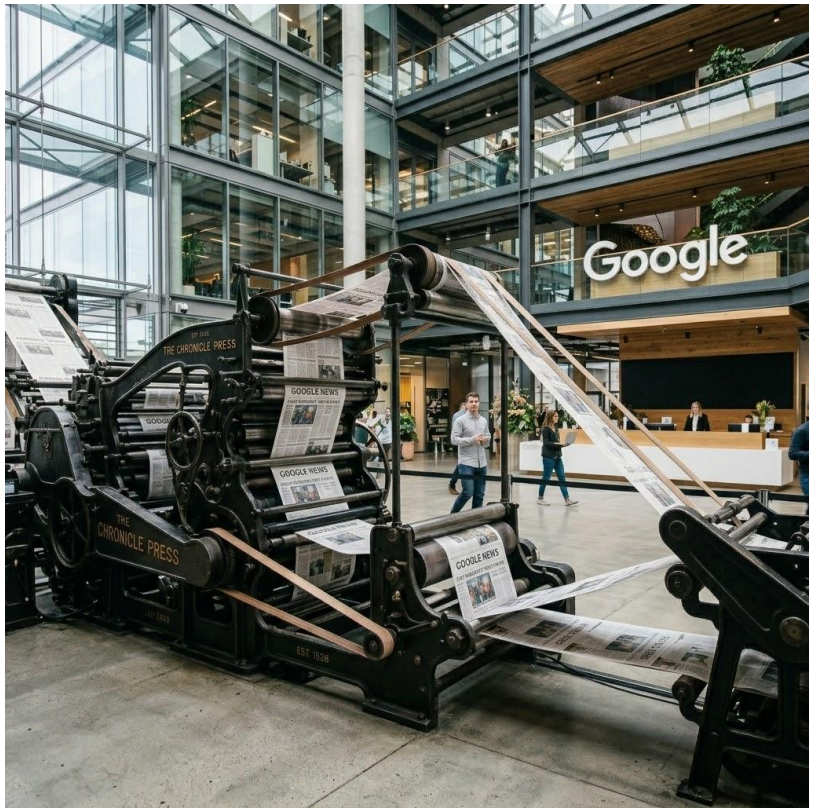
Voss Report: Recalibrating EU copyright for AI

- **Licensing-first recalibration of the CDSM regime.** Article 4(3) opt-out retained as the substantive baseline, supplemented by voluntary sectoral licensing. No mandatory collective licensing.
- **EUIPO as trusted intermediary.** Centralised opt-out standards, training-data documentation templates, and a proposed register of works used in training and rightholder opt-outs.
- **Transparency with teeth.** Rebuttable presumption that a model placed on the EU market was trained on protected works if transparency obligations are not met. Costs recovery mechanism modelled on Article 21(5) DSA.
- **Press and news media carve-out.** Stronger and more exclusive control for press publishers over RAG and inferencing uses, with express consent and an extension of related rights contemplated.
- **Extraterritorial scope.** Compliance triggered by placing on the EU market, irrespective of where training occurred. Mirrors the controversial reading of Article 53(1)(c) AI Act.

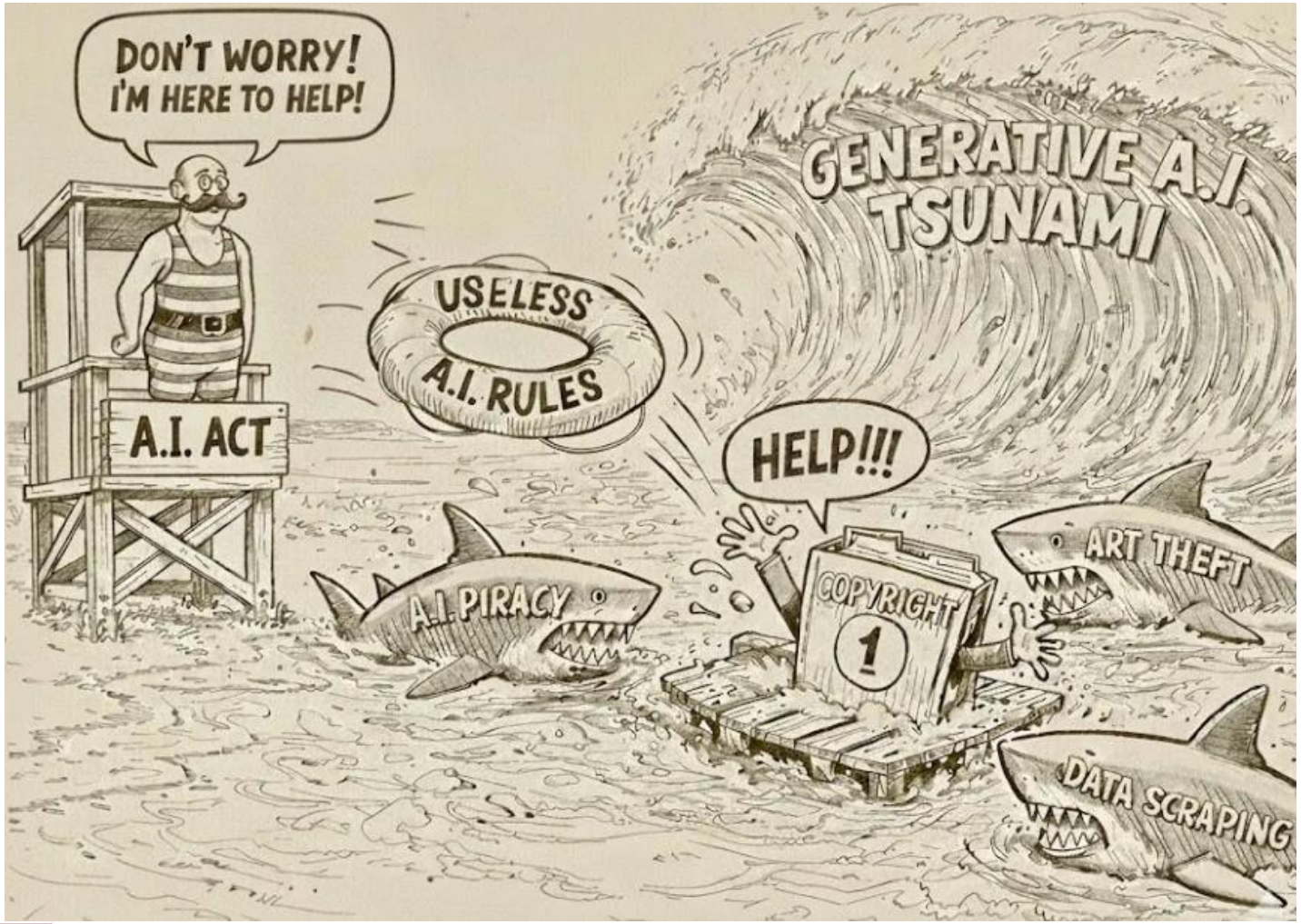


The press publishers' moment

- Article 15 CDSM created an ancillary right for press publishers over digital uses of their content.
- The Like Company referral now tests whether this right applies to chatbot summarisation that goes beyond very short extracts.
- If the CJEU holds that it does, RAG-based products and search-style AI assistants face a structural licensing requirement for press content in the EU.
- The Voss Report contemplates a stronger and more exclusive control regime for press publishers, including an extension of the related right to expressly cover AI uses.
- The press publishers' position is now the leading edge of the European licensing debate.



The AI Act Enters the Scene



Regulation (EU) 2024/1689 – AI Act (in force from February 2025).

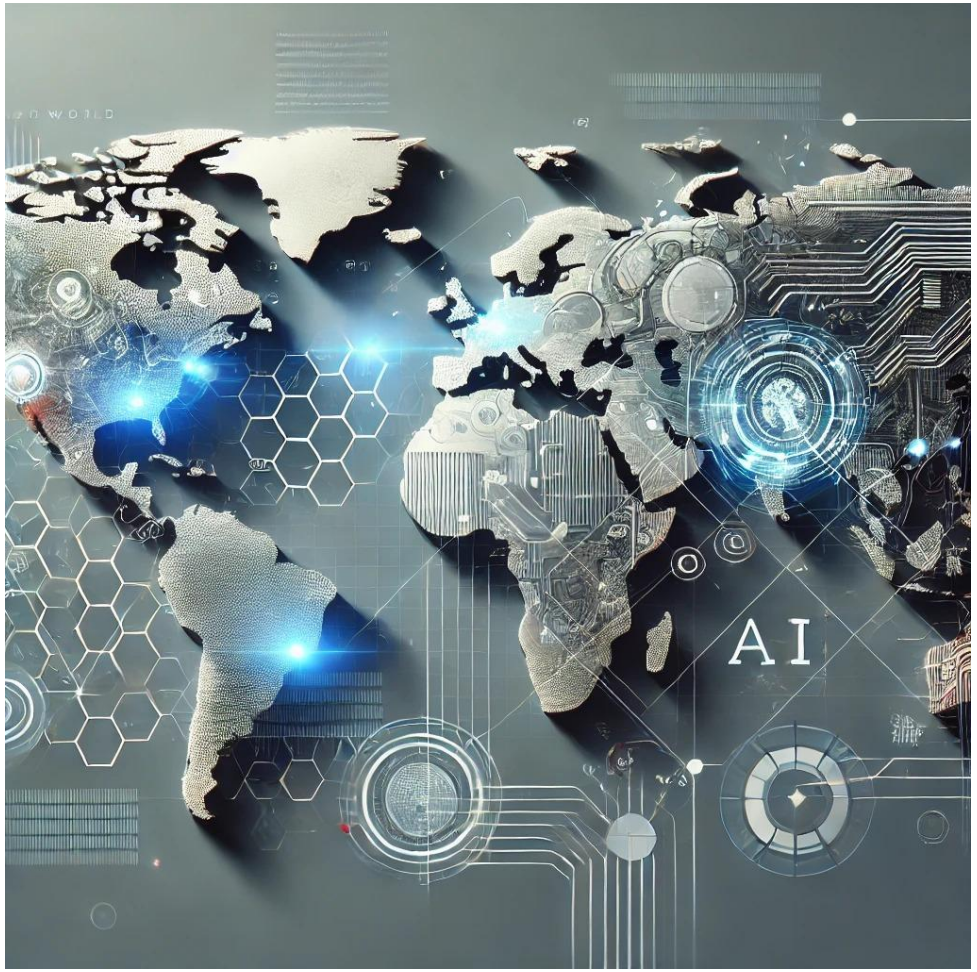
Applies to General-Purpose AI Models (GPAI) from August 2025.

Article 53 requires GPAI providers to:

- Create a copyright compliance policy.
- Detect and honour Article 4(3) opt-outs.
- Publish a training data summary.
- Cooperate with regulators.

Applies extraterritorially to non-EU providers operating in the EU.

Recital 106 – Territorial Neutrality



- Any provider placing a GPAI model on the Union market must comply with EU copyright rules, regardless of where training took place.
- Purpose: eliminate regulatory arbitrage and ensure a level playing field.
- Effect:
  - No safe harbour by training in non-EU jurisdictions.
  - Triggers full Article 53 obligations for any GPAI model placed on the EU market.
  - Applies to both developers and deployers of GPAI.

The GPAI Code of Practice

- I.2.1 – Maintain a formal copyright policy, with assigned responsibilities
- I.2.2 – Do not crawl paywalled or piracy-linked domains
- I.2.3 – Detect and comply with opt-outs under Article 4(3) DSM
- I.2.4 – Check copyright status of third-party datasets
- I.2.5 – Prevent memorisation that leads to infringing outputs
- I.2.6 – Provide contact point and complaints mechanism for rightsholders

The Code of Practice signatories

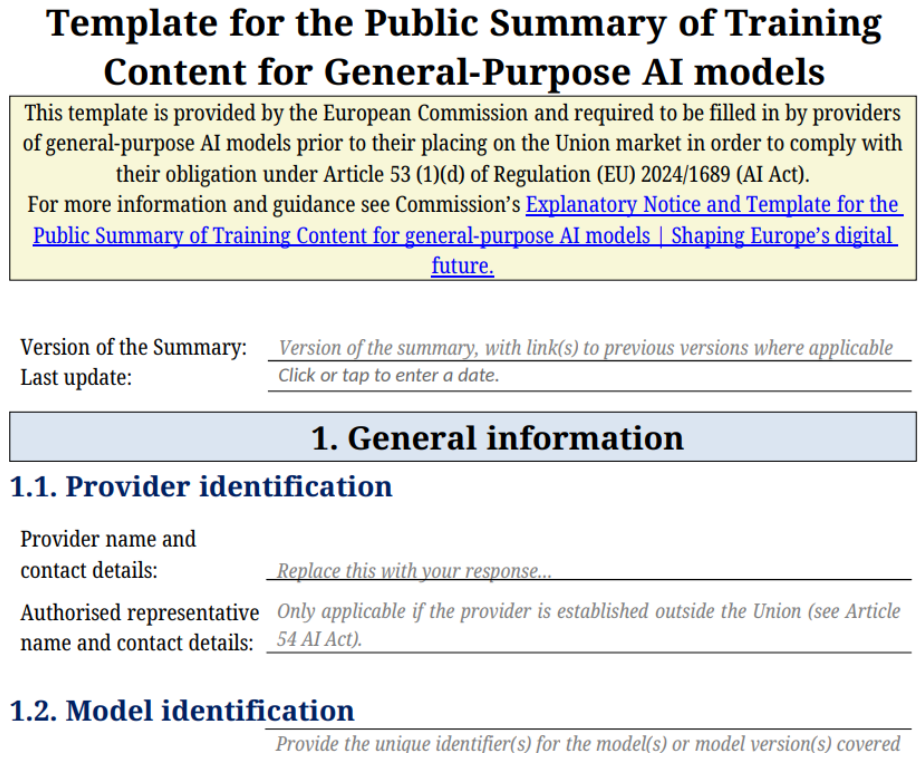
- The General-Purpose AI Code of Practice was finalised in July 2025.
- Signatories include OpenAI, Anthropic, Google, Microsoft, Mistral, and Amazon.
- Meta refused to sign and challenged the framework publicly. xAI signed only the safety chapter, not the copyright chapter.
- Signatories receive a presumption of compliance with the underlying Article 53 obligations. Non-signatories must demonstrate equivalent measures on their own evidence.
- Enforcement leverage now runs through this distinction. The AI Office can apply heavier scrutiny to non-signatories without amending the law.

GPAI Guidelines

- Clarifies how Article 53 obligations will be interpreted by the Commission.
- Opt-outs must be respected unless technically unfeasible.
- Training data summaries must document how opt-outs were handled.
- Disclosure of past non-compliance must be included and justified.
- Downstream modifiers may be treated as “providers” under Article 53.
- Transitional rules give existing models extra time only if non-compliance is transparently explained.
- Reinforces a phased but enforceable compliance model built on transparency and accountability.

The training data summary template

- The Commission published the mandatory template in July 2025.
- Providers must disclose principal data sources, modalities used, the order of magnitude of works, the treatment of opt-outs, and any past non-compliance
- The template forces structured transparency without requiring full dataset disclosure.
- Early summaries have varied widely in granularity. The AI Office is expected to publish supplementary guidance during 2026 to address inconsistency.
- For deployers, the training data summary is the primary public artefact for vendor due diligence.



Template for the Public Summary of Training Content for General-Purpose AI models

This template is provided by the European Commission and required to be filled in by providers of general-purpose AI models prior to their placing on the Union market in order to comply with their obligation under Article 53 (1)(d) of Regulation (EU) 2024/1689 (AI Act).
For more information and guidance see Commission's Explanatory Notice and Template for the Public Summary of Training Content for general-purpose AI models | Shaping Europe's digital future.

Version of the Summary: *Version of the summary, with link(s) to previous versions where applicable*
Last update: *Click or tap to enter a date.*

1. General information

1.1. Provider identification

Provider name and contact details: *Replace this with your response...*

Authorised representative name and contact details: *Only applicable if the provider is established outside the Union (see Article 54 AI Act).*

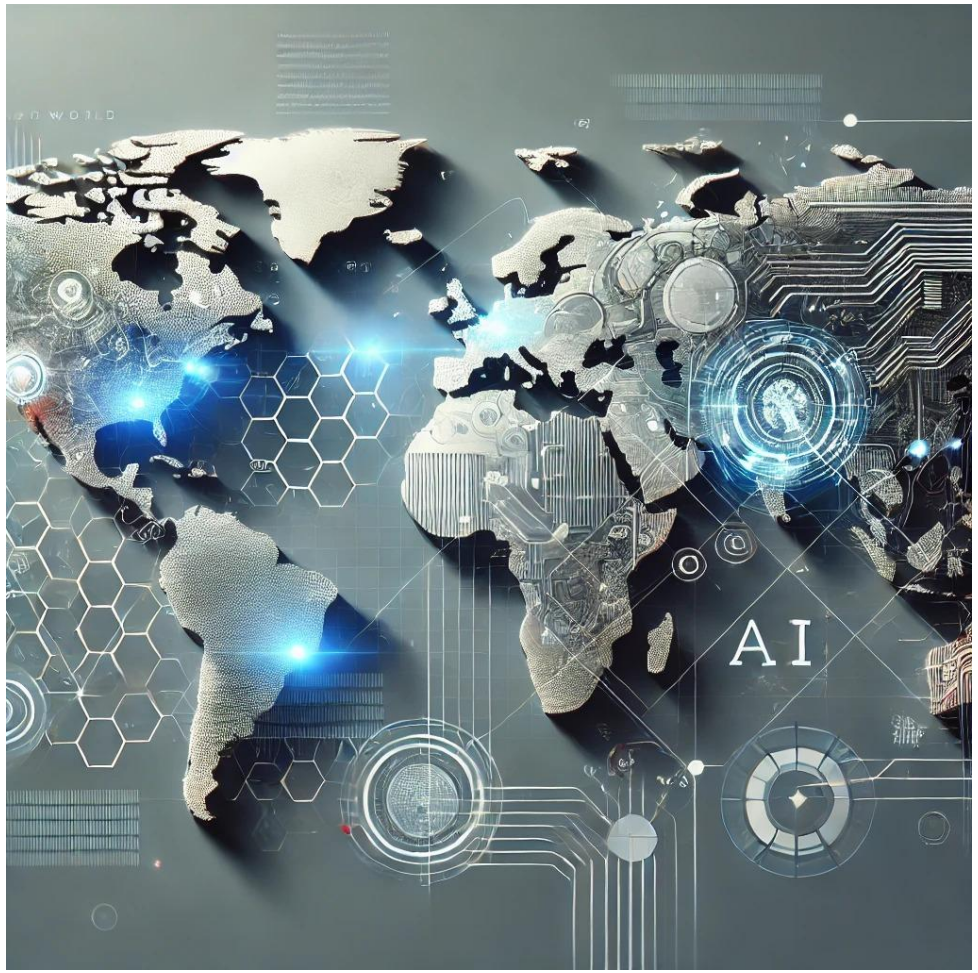
1.2. Model identification

Provide the unique identifier(s) for the model(s) or model version(s) covered

Transparency and the trade secrets problem

- The training data summary creates a governance tension. Providers must disclose enough to satisfy Article 53 and the public interest in transparency.
- Providers must also protect competitively sensitive information about pipeline design, dataset construction, and filtering methodology.
- The Commission's template attempts to thread the needle through high-level categories rather than itemised disclosures.
- For governance teams the practical question is ownership. Who inside the organisation owns the trade-off between disclosure and protection?
- This is the same accountability gap GDPR exposed in 2018, surfacing in a different domain.

Enforcement Realities - Jurisdiction



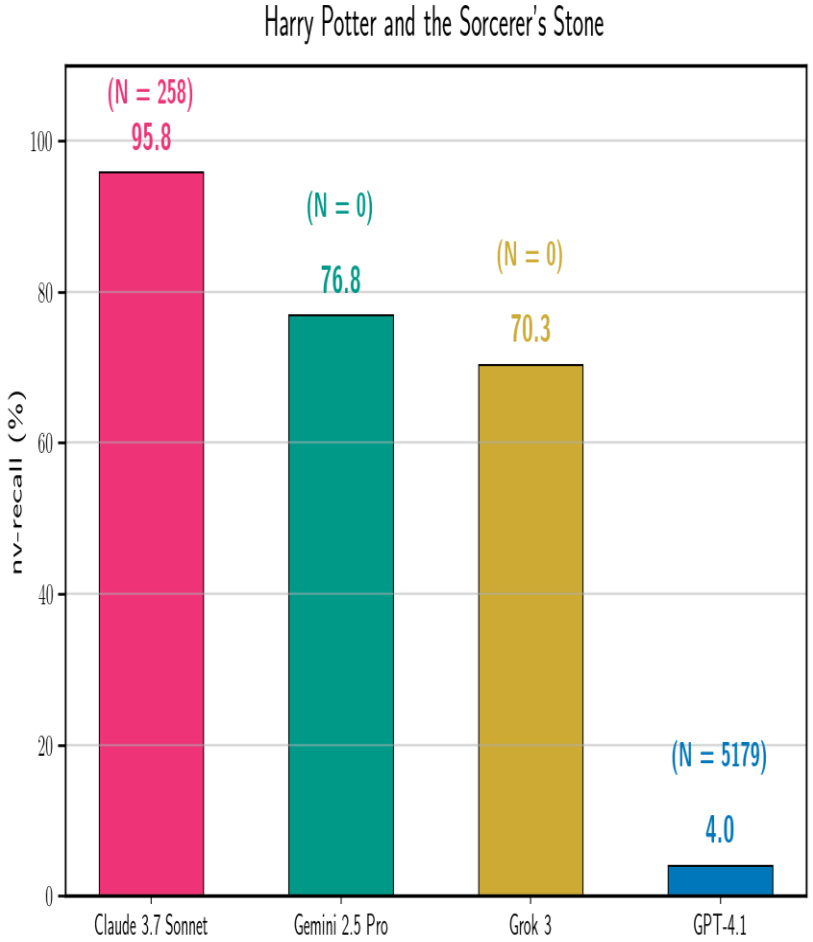
- Where does infringement occur?
  - Where the data is scraped?
  - Where the model is trained?
  - Where it is deployed or generates outputs?
- Infringement may span multiple territories
- Enforcement can only proceed country by country
- AI firms often based outside jurisdiction

Enforcement Realities– Proof

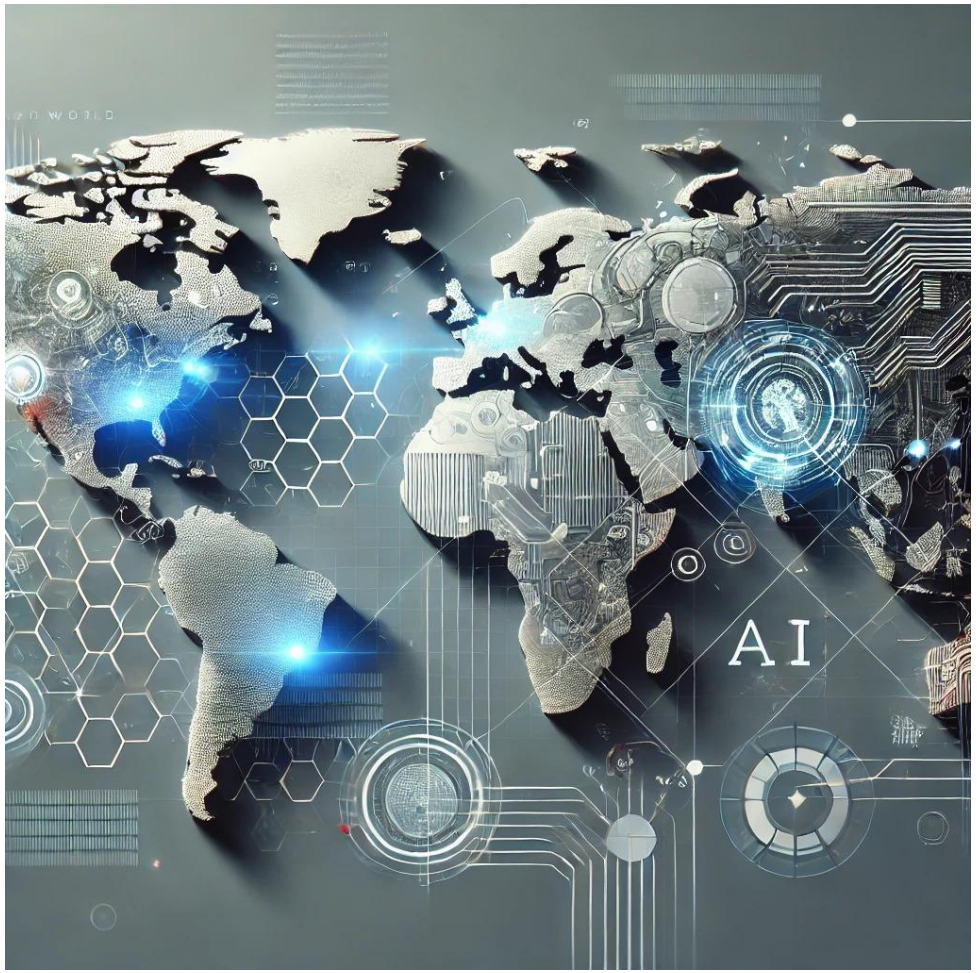
- How do you prove a work was ingested?
  - Training datasets not publicly disclosed
  - Outputs rarely include exact matches
- Embedding, style replication, and hallucinated attribution are common
- Regurgitation
- Very difficult to show causal copying at scale
- Discovery rights are weak in many jurisdictions

Memorisation

- Memorisation creates a practical copyright risk because a model can reconstruct distinctive training text when a prompt aligns with learned “grooves”.
- Extraction turns that latent risk into an operational one: it can be tested, repeated, and costed, which makes it easier to evidence in disputes and harder to dismiss as theory.
- The “book extraction” results sharpen the stakes because large-scale recovery moves beyond snippets into substantial reproduction, which is where infringement arguments become far more direct.
- Guardrails do not remove memorised content; they sit as a safety layer over weights that may still contain recoverable material, and that matters especially for open-weight release.
- Data permanence raises a governance problem: if memorised content persists as a recoverable imprint, “deletion” and the right to erasure become technically and legally contested in practice.



Global Copyright Litigation



- USA – Cases Against OpenAI
- United Kingdom – *Getty Images v Stability AI*
- European Union – *Laion* Decision, *GEMA* cases

The Fear of Going First

Global Copyright Litigation

Germany*: Kneschke v. LAION*

- **Hamburg Higher Regional Court (10 Dec 2025)** upheld dismissal of Kneschke's claim against LAION over use of a photograph in dataset creation for AI-related purposes.

- **TDM exception confirmed for the download:** the court held the image download and analysis was covered by Section 44b UrhG / Article 4 DSM (unlike the first instance, which left this open).

- **Opt-out standard tightened:** a natural-language "no bots/scraping" notice on the stock agency site was held not effective because online opt-outs must be machine-readable under Section 44b(3) UrhG.

- **Research exception interpreted broadly:** use also justified under Section 60d UrhG / Article 3 DSM, treating dataset creation for future knowledge gain as "scientific research", even if others may use it commercially.

- **Scope still unresolved and appeal possible:** the court limited itself to pre-training measures (image–text comparison) and allowed a further appeal to the Federal Court of Justice, leaving wider questions on generative model training open.

Global Copyright Litigation

United Kingdom*: Getty Images –v- Stability AI*

- **Core holding:** Model weights are not “copies” of training images and do not store or reproduce the training data; they reflect learned patterns from training.
- **Where infringement sits:** The legally relevant copying, if any, happens during training when images are downloaded and stored (the “materialisation” step), not in the existence of the trained model.
- **Secondary infringement shut down:** Even if training involved infringement elsewhere, importing or distributing the model in the UK did not make the model an “infringing copy” for secondary infringement purposes.
- **Trade mark outcome was narrow:** Limited findings tied to specific Stable Diffusion versions and watermark evidence; broad dilution and tarnishment arguments failed for lack of real-world proof.
    - **Responsibility point matters:** The court rejected the “passive tool” framing and treated the developer’s choices on datasets, filters, and the wider generation “pipeline” as central to liability analysis.

Global Copyright Litigation

Germany*: GEMA –v- OpenAI Ireland*

- **First European ruling** holding an AI developer directly liable for unlicensed use of copyrighted works in training and outputs (GEMA v. OpenAI, Munich Regional Court).
- **Training and verbatim output treated as "reproduction"** under EU copyright rules, with the model said to contain reproducible determinations of lyrics.
- **Text and data mining exceptions narrowed:** permitted analysis does not extend to permanent "embodiment" or memorisation of protected works in model parameters.
- **Compliance implications for GPAI providers:** points towards blanket licensing, stronger dataset documentation, and traceability expectations under AI Act Article 53.
- **Strategic consequence:** heightened EU litigation risk pending appeal, widening divergence from UK and US approaches and strengthening rights-holder leverage (including collective licensing).

Global Copyright Litigation

CJEU*: Like Company –v- Google Ireland*

- **First CJEU referral on generative AI and EU copyright:** *Like Company v Google Ireland* (Gemini) tests how EU law applies to both LLM training and chatbot outputs.
- **Press publisher rights in focus:** whether chatbot summaries that go beyond "very short extracts" trigger Article 15 CDSM and communication to the public.
- **Training as "reproduction":** whether tokenisation and pattern-learning from protected works is reproduction under Article 2 InfoSoc.
- **Limits of the TDM exception**: if training is reproduction, whether Article 4 CDSM (text and data mining) shields commercial LLM training.
- **Practical impact:** outcome may drive licensing and dataset due diligence requirements for AI providers, and strengthen press publishers' enforcement leverage across the EU.

Global Copyright Litigation

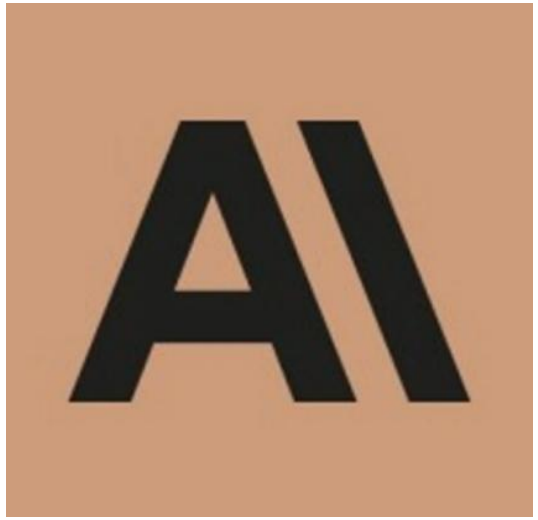
California*: Kadre –v- Meta*

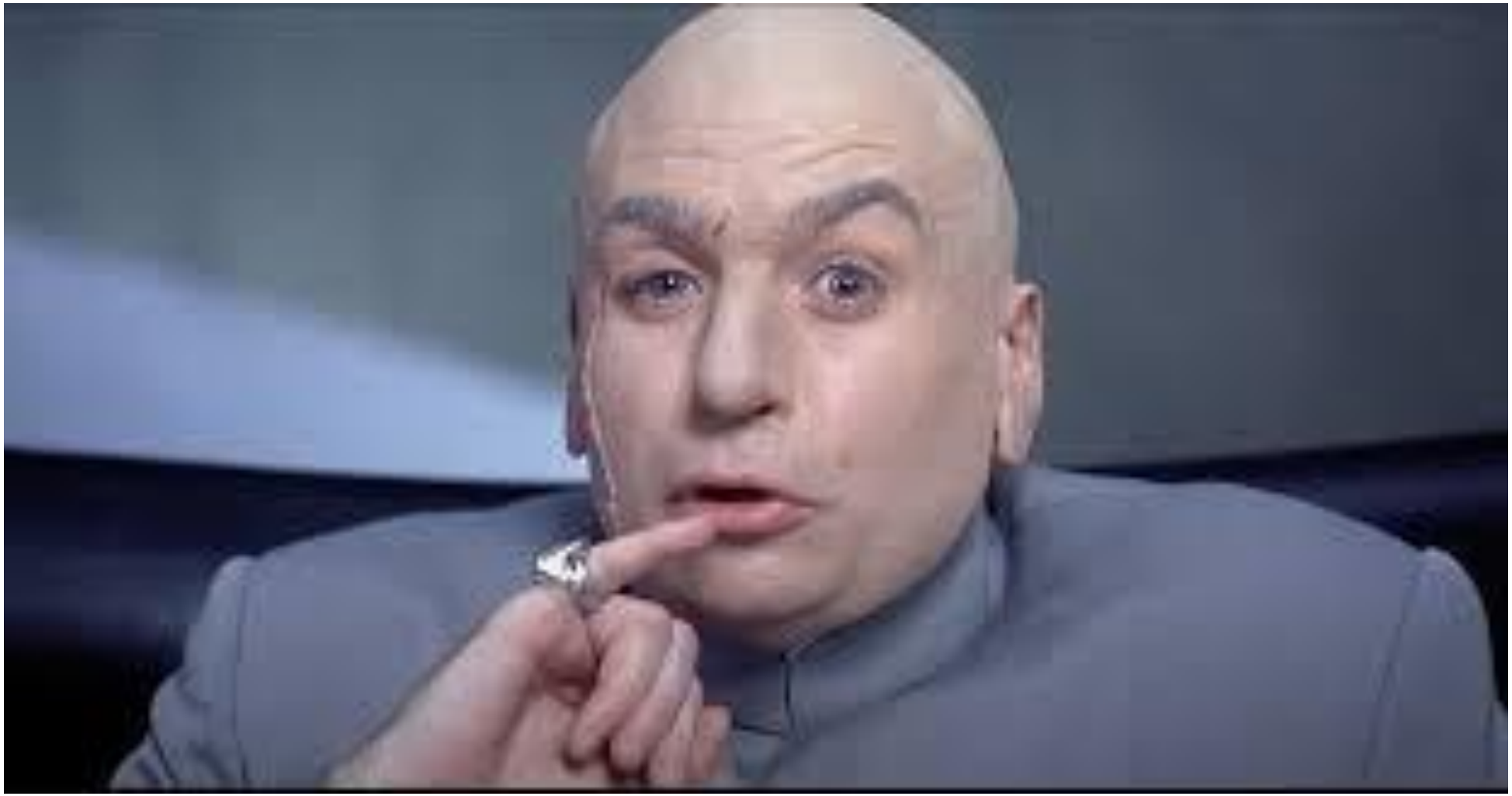
- **Claim:** Authors allege Meta copied books without permission to train LLaMA, constituting copyright infringement and related claims.
- **Court:** Northern District of California, Judge Vince Chhabria, consolidated class-style actions.
- **Fair use outcome (25 June 2025):** Plaintiffs lost on summary judgment; training use held highly transformative (factor one).
- **Market harm (factor four):** Court rejected plaintiffs' market theories and found no meaningful evidence that LLaMA's training harmed the market for their books; licensing-for-training theory not recognised on this record.
- **DMCA outcome (27 June 2025):** Meta won partial summary judgment on DMCA §1202 because, with copying deemed fair use, removal of copyright management information could not further infringement; ruling framed as record-specific, not a blanket precedent.

Global Copyright Litigation

California*: Bartz –v- Anthropic*

- **Split ruling on fair use:** training on lawfully obtained books held "spectacularly transformative"; unlawful acquisition of pirated books could not be cured by fair use.
- **Data provenance becomes central:** the legality of the pipeline turns on how inputs were sourced, not only what the model does with them.
- **Scanning bought books permitted (narrowly):** format-shifting for internal search and practical use accepted where it does not create a market substitute.
- **Output risk needs proof:** claims of memorisation or infringing outputs require evidence of infringing generation, not theoretical possibility.
- **EU tension flagged:** judgment's provenance logic may sharpen pressure under EU TDM opt-outs and AI Act Article 53, especially for cross-border training and data collection.
- **LED TO THE BIGGEST COPYRIGHT SETTLEMENT IN HISTORY…**

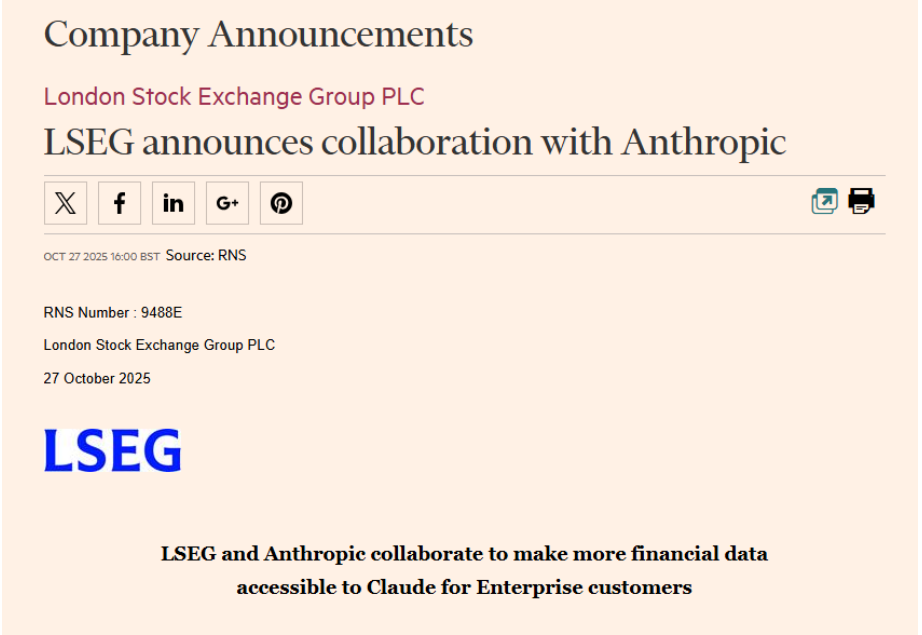




$1.5 BILLION DOLLARS

The Licensing Wave

- New York Times v OpenAI continues in litigation. The same period has produced a series of voluntary licensing deals.
- OpenAI has signed agreements with News Corp, Axel Springer, Associated Press, Le Monde, Schibsted, Vox Media, The Atlantic, and Time.
- Anthropic has signed with the Financial Times.
- Reddit has licensed to both OpenAI and Google.
- Microsoft has signed sectoral arrangements across European publishers.
- The litigation-to-licensing arc is not a side effect. It is the equilibrium the Voss Report assumes.



Company Announcements

London Stock Exchange Group PLC

LSEG announces collaboration with Anthropic

OCT 27 2025 16:00 BST Source: RNS

RNS Number : 9488E

London Stock Exchange Group PLC

27 October 2025

LSEG

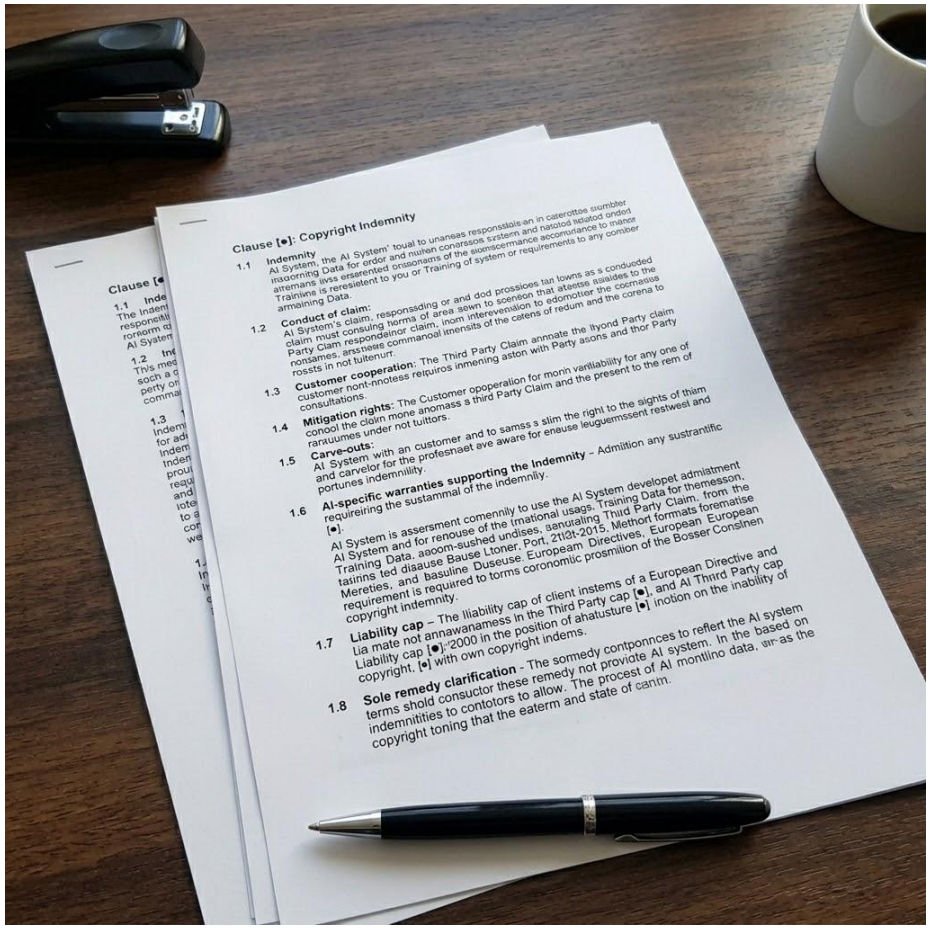
LSEG and Anthropic collaborate to make more financial data accessible to Claude for Enterprise customers

Music, Voice, and Performers

- Suno and Udio faced litigation from Universal, Sony, and Warner.
- The claims allege wholesale ingestion of copyrighted recordings and outputs that reproduce protected elements of the catalogue.
- Voice cloning sits at the boundary of copyright and personality rights. EU law engages through Article 3(2) of the Rental and Lending Directive on performers' rights, and through Member State personality and image rights regimes.
- The US NO FAKES Act proposes a federal right of publicity addressing AI-generated likenesses and voices.
- For governance teams in entertainment, advertising, and any voice-based product, this is now a board-level question.

The Contractual Cascade

- Article 53 obligations sit on providers. They flow downstream through procurement contracts.
- Deployer protections should include representations on training data provenance and lawful access, warranties on opt-out compliance, and commitments on output filtering and memorisation testing.
- Indemnities for copyright infringement should be scoped to both inputs and outputs, with carve-outs and caps modelled on the major hyperscaler positions (Microsoft, Google, Anthropic).
- Restrictions on model substitution without consent.
- Audit rights, including rights to receive the training data summary and supporting compliance evidence on demand.



Regulatory Divergence: EU

- Legally enforceable opt-outs (CDSM Art. 4).
- Mandatory copyright policies (AI Act, Art. 53).
- Transparency through Code of Practice and training data summaries.

Regulatory Divergence: United States

Biden Executive Order on Semiconductor Export Controls (Jan 2025):

- Imposed new restrictions on advanced AI chip exports to 50+ jurisdictions due to national security concerns. Ireland was one of only 18 countries excluded from these restrictions, reflecting its status as a trusted US partner and critical hub in the global semiconductor supply chain.

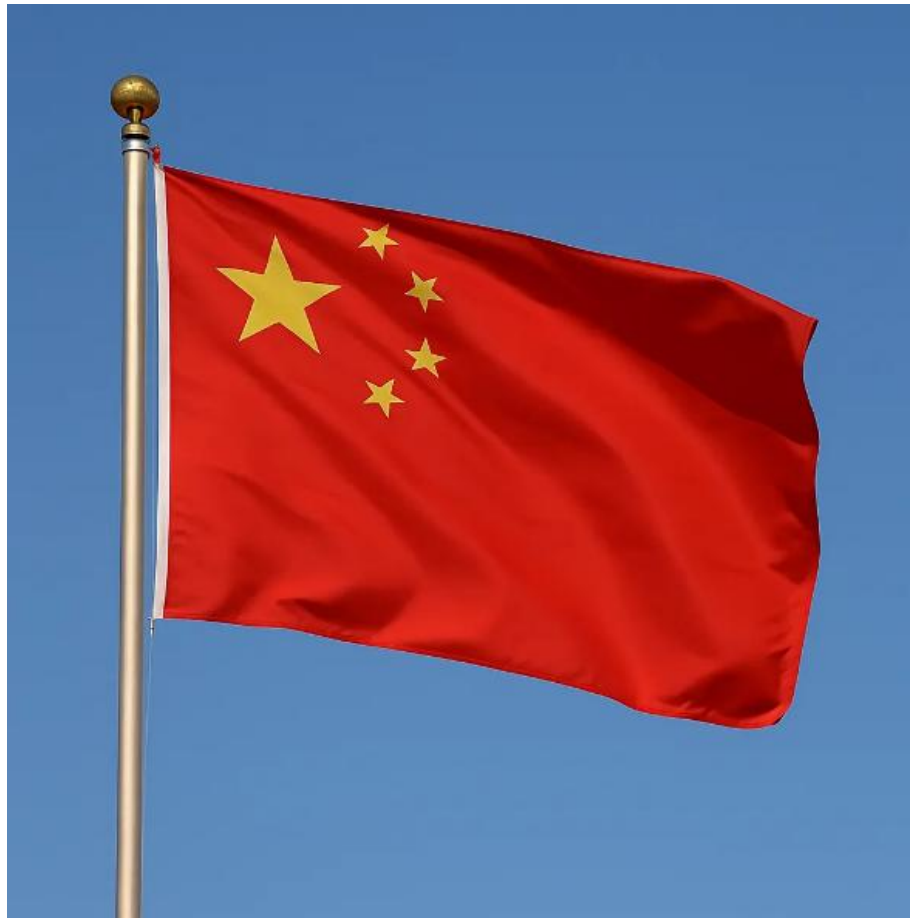
Trump Executive Order (January 2025)

- Rescinded Biden's AI Safety Executive order

California

- In September Gov. Gavin Newsom vetoed California's AI Safety Legislation

Regulatory Divergence: China



- **Generative AI Regulations :** China's *Interim Measures for the Management of Generative Artificial Intelligence Services* mandate that AI service providers use training data from lawful sources and ensure that AI-generated content does not infringe upon others' intellectual property rights.
- **Consent and Attribution Requirements:** The regulations require providers to obtain consent when personal information is involved and to label AI-generated content appropriately to prevent deception.
- **No Fair Use Doctrine:** Unlike the U.S., China does not have a broad fair use exception. Therefore, unlicensed use of copyrighted material for AI training poses significant legal risks.
- **Judicial Recognition of AI-Generated Works:** Chinese courts have begun to recognise AI-generated content as eligible for copyright protection.
- **Mandatory Labeling of AI-Generated Content:** Starting September 1, 2025, China will enforce regulations requiring clear labeling of AI-generated content, including visible marks and metadata identifiers, to enhance transparency and combat misinformation.

EU Copyright Law: What Does the Future Hold?

- The European Commission opened a call for evidence on copyright modernisation on 13 May 2026. It closes on 25 June.
- The exercise is a reconsideration of the CDSM Directive, drafted in 2019 before generative AI existed. The TDM provisions in Article 4 were almost afterthoughts at the time.
- Three structural problems sit underneath any licensing-led reform: identification of training inputs at scale, valuation of individual fragments, and enforcement against models whose training data cannot be reverse-engineered.
- The harmonisation gap is the deeper issue. There is no EU-wide doctrine on authorship of AI-generated outputs. What is protectable in Ireland is not protectable across much of the Union.
- Digital replica rights, addressed in Voss and in the Commission's evidence call.

How Did Things Go?
(We Really Want To Know)

Did you enjoy this session? Is there any way we could make it better? Let us know by filling out a speaker evaluation.

1. Open the IAPP Events app.
2. Select **IAPP AIGG Europe 2026**.
3. Tap "Schedule" on the bottom navigation bar.
4. Find this session. Click "Rate this Session" within the description.
5. Once you've answered all three questions, tap "Done".

Thank you!