

## Disclaimer:

**This primer is intended solely for informational purposes. The content reflects the current discussion on copyright issues as of [Month/Year] and does not constitute legal advice or a legally binding opinion. Due to its general nature, the information provided cannot address the specifics of individual cases. For advice tailored to a particular situation, readers should consult qualified legal experts.**

## PRIMER 1 — AI INPUT & COPYRIGHT

### 1. Introduction

The rapid development of generative artificial intelligence (GenAI) has placed unprecedented pressure on existing copyright frameworks. At the centre of legal scrutiny is the use of copyrighted content as input for training large-scale AI models, including large language models (LLMs), diffusion models, and multimodal systems.

Training such models requires at least one of the following activities: the collection, reproduction, extraction, storage, transformation, and analysis of vast amounts of data—processes which constitute copyright-relevant acts under EU law. As emphasised in the European Parliament Study on Generative AI and Copyright (Lucchi, 2025), no existing copyright exception fully covers the end-to-end workflow of AI model training.

Developers and deployers operating in or supplying the EU market must therefore navigate a multi-layered legal framework consisting of:

1. Copyright in the Digital Single Market Directive (CDSM Directive, 2019/790)
2. The Information Society Directive (2001/29) and Database Directive (96/9)
3. The AI Act (Regulation 2024/1689), especially Art. 53
4. The GPAI Code of Practice (2025)
5. Emerging enforcement trends and litigation.

### 2. Copyright Law and the Reproduction of Works in AI Training

#### 2.1 Copyright-Relevant Acts of Use

The CDSM Directive operates against the background of the InfoSoc Directive (2001/29), which grants rightsholders exclusive control over:

- reproduction (Art. 2),
- communication to the public (Art. 3),
- distribution (Art. 4).

Every technical step in AI model development engages these rights. The following steps constitute acts of reproduction or extraction under EU law:

- web scraping and crawling;
- downloading datasets (raw or pre-processed);

- storing training corpora;
- (intermediate) caching and pre-processing;
- fine-tuning and retraining cycles;
- distribution of trained model weights.

Copies made “even as intermediate steps during training” are reproductions in the sense of the InfoSoc Directive (Rosati, 2025).

This legal position aligns with academic commentary (Peukert, 2024); (Lucchi, 2025); (Guadamuz, 2024), all of which emphasise that AI training is a technically copy-intensive process that cannot be characterised as mere 'extraction of ideas' or 'inspiration'.

### 3. The TDM Exceptions in the CDSM Directive

The CDSM Directive contains two distinct exceptions authorising limited acts of reproduction/extraction for Text and Data Mining (TDM):

- Article 3 — TDM for scientific research
- Article 4 — General TDM exception applicable to commercial actors

These are essential for AI developers but remain narrow in scope.

#### 3.1 Article 3 CDSM — TDM for Scientific Research

Article 3 provides a mandatory exception for reproductions and extractions made for TDM by research organisations or cultural heritage institutions, provided:

- the purpose is scientific research;
- the entity has lawful access;
- copies are stored securely;
- they may be retained for research-verification purposes.

This exception is strictly limited to organisations operating under a public-interest mission, such as:

- universities,
- non-profit research institutes,
- public archives or libraries.

Private-sector developers cannot rely on Article 3 unless participating in a recognised scientific research collaboration. Even then, Article 3 does not authorise training generative AI models for commercial deployment, because the purpose of commercial model training exceeds scientific research (Quintais, 2025); (Lucchi, 2025).

#### 3.2 Article 4 CDSM — General TDM Exception

Article 4 allows any user, including commercial actors, to perform reproductions/extractions for TDM unless the rightholder has expressly reserved their rights (“opt-out”).

Requirements:

- lawful access (undefined but restrictive);

- no rights reservation by rightsholders.

#### Scope Limitations

1. Article 4 only authorises reproductions and extractions for the purpose of TDM, not the downstream model training steps that involve transformation, persistent storage, re-use, distribution, etc.
2. TDM exceptions do not authorise the creation of a generative model capable of emitting outputs similar to protected works.
3. Opt-out mechanisms significantly narrow the practical availability of Article 4.

Scholars (Peukert, 2024); (Lucchi, 2025) argue that Article 4 was never intended to support industrial AI model training and does not withstand AI “stress tests”. However, the opposite position is also represented in the specialist literature, namely that the TDM exception also applies to the training of AI systems (Beurskens, 2025).

#### 3.3 The Rights Reservation (Opt-Out) Problem

Article 4(3) provides that rightsholders may opt out “in an appropriate manner, including by machine-readable means”. However, no harmonised standard exists.

Currently, there are multiple opt-out systems in use especially:

- robots.txt (REP) – widely used but technically unreliable (Li et al., 2025), not legally robust
- TDMRep – early machine-readable proposal
- C2PA metadata – for content provenance and usage restrictions
- ISCC – perceptual hashing for content recognition
- contractual restrictions (Terms of Use prohibiting TDM).

However, the lack of standardisation has created extreme legal uncertainty. Hence, on 1 December 2025, the European Commission launched a formal consultation<sup>1</sup> to develop EU-wide protocols for expressing TDM opt-outs under:

- AI Act Art. 53(1)(c)
- GPAI Code of Practice (Copyright Chapter)

The consultation aims to:

- define a harmonised technical standard,
- improve machine-readable enforcement,
- consider registries of rights reservations,
- ensure compliance by developers of GPAI models.

---

<sup>1</sup> See: Commission launches consultation on protocols for reserving rights from text and data mining under the AI Act and the GPAI Code of Practice at <https://digital-strategy.ec.europa.eu/en/consultations/commission-launches-consultation-protocols-reserving-rights-text-and-data-mining-under-ai-act-and>.

This initiative is highly significant: once adopted, the new protocol will likely become the *de facto compliance requirement* for all AI developers operating in the EU. There are also indications that while several GenAI developers' source and use content available online without prior authorisation of copyright holders, a market for direct licensing is slowly emerging (European Union Intellectual Property Office., 2025).

#### 4. The AI Act: A Parallel and Complementary Compliance Regime

##### 4.1 Article 53 AI Act — Copyright Obligations

Providers of general-purpose AI models must:

1. Adopt, maintain, and implement a copyright compliance policy (Code of Practice, Measure 1.1).
2. Ensure lawful access when reproducing/extracting content for crawling (Measure 1.2).
3. Identify and comply with rights reservations expressed by rightsholders (Measure 1.3).
4. Publish a sufficiently detailed summary of training data (Art. 53(1)(d)).
5. Mitigate output-related copyright risks (Measure 1.4).
6. Provide a copyright complaint mechanism (Measure 1.5).

These requirements go beyond copyright law and constitute regulatory obligations enforceable by the European AI Office and national authorities.

##### 4.2 GPAI Code of Practice (2025)

The General-Purpose AI Code of Practice<sup>2</sup> is a voluntary but Commission-endorsed mechanism for demonstrating compliance with Articles 53 and 55. Key commitments for signatories (Microsoft, Google, OpenAI, Mistral, etc.) include:

- documentation of training datasets,
- machine-readable rights reservation detection,
- lawful-access verification,
- filtering of infringing outputs,
- designation of a copyright point of contact.

#### 5. Lawful Access: A Critical Unresolved Concept

The notion of lawful access is one of the most contentious elements in Article 4 CDSM and AI Act compliance.

Current Interpretations

- Content that is freely accessible online (without technical or contractual restrictions) is likely lawfully accessible.<sup>3</sup>
- Content behind paywalls, logins, or technical barriers is not.

---

<sup>2</sup> See at: <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>.

<sup>3</sup> But there are still discussions going on, whether this is actually the case (Rosati, 2024).

- Content protected by terms prohibiting crawling, even if technically accessible, is arguably not lawfully accessible (pending CJEU clarification – see below).
- Automated scraping which bypasses access controls is unlawful.

The Court of Justice of the European Union (CJEU) has been called upon to interpret precisely these issues. The preliminary ruling request in *Like Company v. Google Ireland* (Case C-250/25) presents the first opportunity for the CJEU to clarify whether, and under what conditions, AI training and AI-generated outputs implicate copyright and related rights under EU law.<sup>4</sup>

## 6. Current Litigation Trends

There are currently several legal disputes on the key issues about AI & copyright, which will have crucial impact on the interpretation of applicable copyright rules. It is important to observe the following ongoing and recent cases:

### 6.1 EU Case — GEMA v. OpenAI (Germany 2025)<sup>5</sup>

In November 2025, the Munich Regional Court held that:

- copying song lyrics for training requires authorisation;
- TDM exceptions do not automatically apply;
- training copies are reproductions regardless of form.

This is the first major EU case affirming that training on copyrighted works without authorisation is infringement.

### 6.2 US Case — Tremblay & Awad v. OpenAI<sup>6</sup>

Allegations include:

- large-scale ingestion of copyrighted books;
- ChatGPT generating summaries demonstrating internal copies.

While US law differs, the case highlights global expectations for licensing.

### 6.3 US and UK Litigations against Midjourney<sup>7</sup>, Stability<sup>8</sup>, Runway<sup>9</sup>

Claims focus on:

- reproduction of copyrighted characters;
- training datasets allegedly containing protected art;

<sup>4</sup> See CJEU, Case C-250/25, *Like Company v. Google Ireland*, preliminary reference lodged on 3 April 2025. Referral from Fővárosi Törvényszék (Budapest Metropolitan Court), Hungary. Available at <https://curia.europa.eu/juris/liste.jsf?num=C-250/25&language=en>

<sup>5</sup> See the Urteil des LG München I – 42 O 14139/24 at <https://www.gesetze-bayern.de/Content/Document/Y-300-Z-GRURRS-B-2025-N-30204>.

<sup>6</sup> See *Tremblay P. and Awad M. v. OpenAI INC. et al*, No. 3:23-cv-03223.

<sup>7</sup> See at [https://deadline.com/wp-content/uploads/2025/09/1-Complaint-1\\_Redacted.pdf](https://deadline.com/wp-content/uploads/2025/09/1-Complaint-1_Redacted.pdf).

<sup>8</sup> See at <https://www.judiciary.uk/wp-content/uploads/2025/11/Getty-Images-v-Stability-AI.pdf>.

<sup>9</sup> See at <https://www.meshiplaw.com/litigation-tracker/andersen-v-stability-ai>.

- lack of licensing arrangements.

These cases, filed by groups of artists and large corporations, generally allege that the companies' generative models were trained without permission on vast amounts of copyrighted material scraped from the internet.

#### 6.4 Sora (OpenAI) Controversies

Sora-generated videos include:

- depictions of copyrighted characters (SpongeBob, South Park, etc.),
- violent and racist imagery violating provider terms.

OpenAI announced forthcoming rightsholder granular opt-in controls, an industry sign of compliance pressure.<sup>10</sup>

### **7. Practical Implications for AI Developers and Deployers**

#### 7.1 Data Collection Phase

- Perform a copyright audit of all intended data sources.
- Classify sources into: licensed / public domain / permissible TDM / opted-out / prohibited.
- Avoid scraping content with clear opt-outs (robots.txt, metadata, ToS).
- Prefer licensed high-quality datasets.

#### 7.2 Dataset Construction Phase

- Maintain detailed provenance logs (required under Code of Practice).
- Use automated tools to detect opt-outs (C2PA, ISCC, TDMRep).
- Store copies securely and limit retention.

#### 7.3 Model Training Phase

- Ensure training operations comply with both TDM rules and AI Act obligations.
- Implement deduplication and memorisation-reduction techniques.
- Retain records needed for regulatory inspection.

#### 7.4 Model Deployment Phase

- Publish a “sufficiently detailed” training summary.
- Provide user documentation on lawful usage.
- Implement copyright-risk mitigation filters.
- Operate a complaint system for rightsholders.

### **8. Conclusion**

---

<sup>10</sup> See: [Sora update #1 - Sam Altman](#) on 4 October 2025: “We are going to make two changes soon (and many more to come).

The EU copyright regime does *not* provide a broad exception allowing developers to train generative AI on copyrighted content. The scope of TDM exceptions (Articles 3 and 4 CDSM) is limited, subject to rights reservations, and inadequate to justify full training workflows. The AI Act introduces additional regulatory obligations, particularly regarding transparency and compliance with opt-outs.

The legal direction is clear: AI model training on copyrighted content requires either a licensing basis or a demonstrable compliance pathway built around lawful access, rights reservation detection, and transparency.

## 9. References

Anderl, A., & Appl. (2025). KI-Einsatz im lautereren Leistungswettbewerb. *ailex* 2025/13.

Beurskens. (2025). Training generativer KI nur auf Lizenzgrundlage? *RDi* 2025, 1.

European Union Intellectual Property Office. (2025). *The development of generative artificial intelligence from a copyright perspective*. Publications Office.

<https://data.europa.eu/doi/10.2814/3893780>

Guadamuz, A. (2024). The EU's Artificial Intelligence Act and copyright. *The Journal of World Intellectual Property*, n/a(n/a), 1-7. <https://doi.org/10.1111/jwip.12330>

Li, Y., Song, W., Zhu, B., Gong, D., Liu, Y., Deng, G., Chen, C., Ma, L., Sun, J., Walsh, T., & Xue, J. (2025). *ai.txt: A Domain-Specific Language for Guiding AI Interactions with the Internet* (arXiv:2505.07834). arXiv. <https://doi.org/10.48550/arXiv.2505.07834>

Lucchi, N. (2025). *Generative AI and Copyright* (p. 175). European Parliament.

[https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST\\_STU\(2025\)7740\\_95\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)7740_95_EN.pdf)

Peukert, A. (2024). *Copyright in the Artificial Intelligence Act – A Primer* (SSRN Scholarly Paper No. 4771976). Social Science Research Network. <https://doi.org/10.2139/ssrn.4771976>

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

Rosati, E. (2024). Is text and data mining synonymous with AI training? *Journal of Intellectual Property Law & Practice*, 19(12), 851–852. <https://doi.org/10.1093/jiplp/jpae092>

Rosati, E. (2025). TDM exceptions (not just the three-step test) don't allow all unlicensed AI development. *The IPKat*. <https://ipkitten.blogspot.com/2025/10/tdm-exceptions-not-just-three-step-test.html>

## PRIMER 1 — AI INPUT & COPYRIGHT (practical version)

---

### 1. Why This Primer Matters for AI Developers and Deployers

Generative AI (GenAI) models — including LLMs, diffusion models, and multimodal systems — require the ingestion of huge quantities of text, code, images, audio and video. Every step in this process involves **copyright-relevant acts**, especially **reproduction**, **extraction**, and **storage**.

---

### 2. The Legal Landscape at a Glance

#### 2.1 Copyright Basics

AI training triggers the **exclusive right of reproduction** under:

- Article 2 InfoSoc Directive (2001/29/EC)
- Articles 3 & 4 CDSM Directive (2019/790)

Copies may be:

- temporary
- fragmented
- obfuscated
- transformed
- embedded inside datasets
- stored in model weights through statistical representations

All still count as acts of reproduction.

#### What this means for you

→ If you make a copy of a protected work at any point in the training pipeline, copyright law applies.

---

#### 2.2 Text and Data Mining (TDM) — Two Different Exceptions

##### Article 3 CDSM — TDM for Scientific Research

Permits reproductions/extractions when:

- performed by **research organisations** or **cultural heritage institutions**,
- for **scientific research**,
- on content they have **lawful access** to.

Copies may be stored securely for verification.

### Important:

This exception *does not* apply to private companies unless they formally participate in public-interest research or work under a non-profit mission.

### Article 4 CDSM — General TDM Exception

Allows anyone (including commercial actors) to reproduce/extract content for TDM, **unless rightsholders have opted out**.

Common rights-reservations include:

- robots.txt<sup>11</sup>
- C2PA metadata
- ISCC identifiers
- TDMRep headers
- explicit terms of use prohibiting scraping

### Limitations of both exceptions

Current scholarship and EU policy documents converge on one point:

**TDM exceptions do not cover full AI training pipelines.** They allow *analysis* but not *model training that generates new content*.

## 3. Rights Reservations: Opt-Outs and the Coming EU Standard

### 3.1 The current messy ecosystem

Rightholders may reserve rights using:

- robots.txt (“REP”) — widely used, but unreliable (Li et al., 2025)
- metadata (C2PA, IPTC, EXIF) — detectable, but often inconsistent
- perceptual identifiers (ISCC) — emerging technical standard
- TDMRep — early machine-readable convention
- contractual ToS prohibitions — strong legal effect when visible

### 3.2 December 2025 EU Consultation

On **1 December 2025**, the Commission launched a consultation aimed at producing a **harmonised EU protocol for expressing TDM opt-outs** under:

- Article 4(3) CDSM
- Article 53(1)(c) AI Act
- GPAI Code of Practice (Copyright Chapter)

This protocol may include:

<sup>11</sup> For information about shortcomings of robots.txt see for example: (Li et al., 2025)

- a machine-readable rights vocabulary;
- a canonical EU metadata standard;
- or a central/federated opt-out registry.

#### Why this matters:

→ AI developers must **detect and comply** especially with these new opt-outs to meet AI Act obligations.

## 4. The AI Act — Your Regulatory Obligations

Under **Article 53 AI Act** (respectively the corresponding Code of Practice), providers of GPAI models must:

### ✓ Maintain an up-to-date copyright policy

(Code of Practice Commitment 1)

### ✓ Crawl only lawfully accessible content

(Copyright Chapter — Measure 1.2)

### ✓ Detect and comply with rights reservations

(Copyright Chapter — Measure 1.3)

### ✓ Mitigate the risk of copyright-infringing outputs

(Copyright Chapter — Measure 1.4)

### ✓ Publish a sufficiently detailed training data summary

(Art. 53(1)(d) AI Act)

### ✓ Provide a copyright complaint mechanism

(Copyright Chapter — Measure 1.5)

## 5. "Lawful Access": A Key Unsettled Issue

The concept of **lawful access** has not yet been fully clarified. But current best understanding:

### Lawful access: likely yes when

- content is publicly available online *and*<sup>12</sup>
- there is no contractual or technical restriction against scraping.

### Lawful access: likely no when

- content is paywalled, login-protected, region-restricted;
- terms of service prohibit crawling or commercial reuse;

<sup>12</sup> But there are still discussions going on, whether this is actually the case (Rosati, 2024).

- scraping circumvents technical measures;
- content is in private repositories.

The pending **CJEU Case C-250/25 (Like Company v. Google)** may clarify this issue.

### What this means for you

→ You must treat “lawful access” as an affirmative obligation: check **technical access, terms of use,** and **rights reservations** before ingestion.

---

## 6. Copyright-Relevant Steps in AI Training

The following steps are **copyright-relevant**:

1. **Web scraping / crawling** → extraction + reproduction
2. **Dataset construction** → reproduction
3. **Dataset publication / sharing** → communication to the public
4. **Downloading datasets** → reproduction
5. **Training processes** → reproduction + extraction
6. **Fine-tuning / RAG indexing** → new reproductions
7. **Distribution of model weights** → possible derivative work issues

### What this means for you

→ Every stage must be either authorised by licence, covered by a valid exception, or excluded from copyright because the data is public domain or non-protected.

---

## 7. Litigation Trends: What Courts Are Signalling

References to some key cases and litigations:

### 7.1 GEMA v. OpenAI (Munich, 2025) – not final

Court held that AI training on song lyrics without permission = **copyright infringement**.

### 7.2 Tremblay & Awad v. OpenAI (US)

Authors allege ChatGPT was trained on their copyrighted books without consent.

### 7.3 Disney, Universal, Warner Bros. lawsuits vs. Midjourney

Claims:

- unlicensed use of protected character art;
- output reproduces copyrighted characters.

### 7.4 OpenAI Sora controversies

The model generated videos of copyrighted characters → strong indicator of training on unlicensed audiovisual content.

## What this means for you

- Courts are increasingly sceptical of training on “publicly available” content without licensing.
  - Memorisation and identifiable reproduction are high-risk areas.
- 

## 8. Practical Workflows for AI Developers

Below are structured workflows aligned with the AI Act, the CDSM Directive, and your internal guidance.

---

### 8.1 BEFORE TRAINING — Data Intake Workflow

#### Step 1 — Classify Data Sources

- Public domain
- Freely licensed (e.g., CC-BY, CC-0)
- Commercially licensed
- Article 4 TDM-eligible (no opt-out)
- Article 4 opt-out present
- Contractually prohibited
- Illegal/uncertain sources (avoid)

#### Step 2 — Check for Rights Reservations

- robots.txt
- C2PA metadata
- ISCC identifiers
- TDMRep headers
- “No scraping” in Terms of Use
- “No AI training” clauses (**increasingly common**)

#### Step 3 — Assess Lawful Access

Ask:

- Am I permitted to access this content?
- Am I permitted to copy this content?
- Does the provider prohibit automated access?

#### Step 4 — Decide: Licence or Exclude

If rights reserved → **exclude or license**.

If lawful access unclear → **exclude or seek legal review**.

---

## 8.2 DURING TRAINING — Compliance Workflow

### Technical safeguards

- dataset deduplication;
- avoidance of high-risk sources;
- filtering of copyrighted images/audio;
- reduced memorisation via RLHF, LoRA or parameter regularisation.

### Documentation

- provenance logs;
- access-verification logs;
- opt-out detection logs;
- model versioning;
- training summaries for future publication.

### What this means for you

→ **If you cannot demonstrate compliance later, regulators may assume non-compliance.**

---

## 8.3 AFTER TRAINING — Deployment Workflow

### Key obligations

- **Publish training data summary** (AI Act 53(1)(d))
- Provide **AI-generated content disclosure** where relevant (Art. 50 AI Act)
- Implement **copyright-risk mitigation** (Code of Practice)
- Provide a **complaints mechanism**
- Maintain legal and technical documentation for inspection

### Risk mitigation tools

- output filtering for verbatim recall;
- similarity checking;
- blocking prompts seeking to reproduce specific copyrighted works;
- provenance tracking (C2PA).

### What this means for you

→ Transparency and provenance will become standard competitive expectations.<sup>13</sup>

---

<sup>13</sup> For more details see: (Anderl & Appl, 2025).

## 9. Developer Checklists

### ✓ TDM / Input Compliance Checklist

- Lawful access verified
- Article 3 applicability assessed
- Article 4 opt-outs checked
- Contractual restrictions checked
- Machine-readable rights detected
- High-risk datasets excluded or licensed
- Data provenance logged
- Data retention policies applied

### ✓ AI Act Article 53 Compliance Checklist

- Copyright policy implemented
- Crawler respects opt-out protocols
- Training summary prepared
- Copyright complaint mechanism active
- Output filtering applied

---

## 10. Conclusion

The legal environment is converging toward a clear position: **AI training on copyrighted content requires either a licence or strict compliance with TDM rules and AI Act obligations**. The opt-out regime is becoming more enforceable through forthcoming harmonised EU standards.

Developers able to demonstrate **lawful access, respect for rights reservations, transparent documentation**, and **robust output safeguards** will be best positioned for regulatory scrutiny and commercial reliability.

---

## 11. References

Anderl, A., & Appl. (2025). KI-Einsatz im lautereren Leistungswettbewerb. *ailex 2025/13*.

Beurskens. (2025). Training generativer KI nur auf Lizenzgrundlage? *RDi 2025, 1*.

European Union Intellectual Property Office. (2025). *The development of generative artificial intelligence from a copyright perspective*. Publications Office.

<https://data.europa.eu/doi/10.2814/3893780>

- Guadamuz, A. (2024). The EU's Artificial Intelligence Act and copyright. *The Journal of World Intellectual Property*, n/a(n/a), 1-7. <https://doi.org/10.1111/jwip.12330>
- Li, Y., Song, W., Zhu, B., Gong, D., Liu, Y., Deng, G., Chen, C., Ma, L., Sun, J., Walsh, T., & Xue, J. (2025). *ai.txt: A Domain-Specific Language for Guiding AI Interactions with the Internet* (arXiv:2505.07834). arXiv. <https://doi.org/10.48550/arXiv.2505.07834>
- Lucchi, N. (2025). *Generative AI and Copyright* (p. 175). European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST\\_STU\(2025\)7740\\_95\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)7740_95_EN.pdf)
- Peukert, A. (2024). *Copyright in the Artificial Intelligence Act – A Primer* (SSRN Scholarly Paper No. 4771976). Social Science Research Network. <https://doi.org/10.2139/ssrn.4771976>
- Quintais, J. P. (2025). Generative AI, copyright and the AI Act. *Computer Law & Security Review*, 56, 106107. <https://doi.org/10.1016/j.clsr.2025.106107>
- Rosati, E. (2024). Is text and data mining synonymous with AI training? *Journal of Intellectual Property Law & Practice*, 19(12), 851–852. <https://doi.org/10.1093/jiplp/jpae092>
- Rosati, E. (2025). TDM exceptions (not just the three-step test) don't allow all unlicensed AI development. *The IPKat*. <https://ipkitten.blogspot.com/2025/10/tdm-exceptions-not-just-three-step-test.html>