

AI TRAINING AND COPYRIGHT INFRINGEMENT: SCIENCE & TECHNOLOGY

NEWS: Are generative AI models built on stolen creative work? Unpacking two court verdicts

WHAT'S IN THE NEWS?

The legality of using copyrighted content to train generative AI is under global scrutiny, with lawsuits in the US and India challenging companies like OpenAI and Meta. India faces regulatory gaps in AI and copyright law, raising ethical, legal, and economic concerns for its creative industry.

1. Core Issue: AI Training and Copyright Infringement

- The key debate revolves around whether **training generative AI models** (like ChatGPT, Claude AI, Meta's LLaMA) on **copyrighted content without consent** violates copyright law.
- The concern is that these models may **memorise and reproduce** content from copyrighted books, music, or art—raising serious legal and ethical questions.
- **Key cases in the US and India** are testing the limits of existing copyright laws in the context of AI.

2. Major International Lawsuits and Court Responses

A. Writers vs. Anthropic (USA, Aug 2023)

- Plaintiffs: Authors including **George R.R. Martin, Michael Chabon**.
- Allegation: Anthropic used **7 million+ pirated books** to train Claude AI.
- Court's observation: Claude's outputs did not “transform” the texts sufficiently to qualify for **fair use** under US law. No final ruling on infringement.

B. Authors vs. Meta

- Allegation: Meta's LLaMA model trained on copyrighted books.
- Court dismissed parts of the case for **procedural reasons**, not legality.
- Notably, court held Meta **could still be liable** if the AI **memorises and outputs verbatim** text from copyrighted sources.

3. Key Legal Concepts in AI & Copyright

- **Fair Use Doctrine (US):** Limited use of copyrighted work without permission if it is **transformative** — such as for **research, criticism, or parody**.
- **Transformative Use:** Must add **new expression, meaning, or utility** to qualify. AI's use of exact text without meaningful change may not meet this standard.

4. The Indian Context: Legal Ambiguity & Emerging Disputes

Ongoing Legal Challenge

- In 2024, **ANI and Indian Music Industry (IMI)** accused **OpenAI** of using **copyrighted Indian content** to train its models.
- Alleged violation of **Section 65A** of the Indian Copyright Act (circumvention of technical protection measures).
- No final judgment yet — **jurisdiction under question** due to OpenAI's lack of physical/data presence in India.

Challenges in Indian Legal Framework

- No **AI-specific provisions** in Indian copyright law (Copyright Act, 1957).
- No legal clarity on:
 - Whether using data for AI training constitutes **infringement**.
 - Whether AI outputs can be considered **derivative works**.
 - Who owns AI-generated content.

5. Why This Issue Matters for India

- **Economic Impact:** India's creative industries (film, music, books) are vulnerable to **unauthorised AI replication**, undermining creator income.
- **Ethical Concern:** AI models may **use creative work without consent**, attribution, or payment.
- **Innovation vs Regulation:** While AI offers growth opportunities, there is need to **balance IP protection** with technological progress.

6. Global Legal Responses and Comparisons

Region	AI & Copyright Approach
United States	Uses " fair use " doctrine; still evolving through case law.
European Union	Allows AI training only under Text and Data Mining (TDM) exceptions ; creators can opt-out explicitly.
Japan	Allows broad exemptions for AI training; more tech-friendly approach.
India	No explicit law or guidance yet; courts and policy pending.

7. Key Dimensions and Gaps in India

Legal Gaps

- No AI-specific law defining:
 - **Authorship of AI-generated content**
 - **Liability for infringement**
 - **Fair use in context of machine learning**

Ethical and Ownership Issues

- **AI-generated content** may be **indistinguishable from original work**.
- Lack of attribution or royalty mechanisms for creators.
- Raises **moral rights** issues.

Economic Vulnerability

- AI monetizes **free or pirated public data**, while original artists/writers lose potential income.
- No mechanism for **revenue sharing** or licensing.

8. Challenges in Enforcement & Regulation

- **No AI Guidelines** under IT Act or **Digital Personal Data Protection (DPDP) Act**.
- **Digital India Act** is still pending — expected to address AI regulation.
- **Consent & Privacy**: AI may be trained on **personal data** without explicit consent, violating data protection norms.

9. Creator Rights and Licensing Mechanisms

- India lacks **Collective Management Organizations (CMOs)** to negotiate AI licensing, as seen in Europe and the US.
- No structured system for **royalties or attribution** from AI outputs.

10. Questions of AI Accountability

- Unclear **who is liable**:
 - The **developer** of the AI?
 - The **user** deploying the tool?
 - The **platform** hosting the output?

11. Public Good vs Private Profit

- Debate over whether **public domain data** should be used by **private companies** for AI profit.
- Raises questions around **open-source mandates**, **algorithm transparency**, and **ethical governance**.

12. India's Role in Global Tech Governance

- India–Germany **Joint Declaration on Triangular Cooperation (2022)** includes digital ethics and governance.
- India can shape **global AI rules**, particularly for the **Global South**.

13. Way Forward for India

- Draft AI-specific amendments to **Copyright Act, 1957**.
- Clarify:
 - Ownership of AI-generated content
 - Liability for AI misuse
 - Conditions for training data usage
- Launch a **Digital India Act** with provisions on:
 - **Ethical AI governance**
 - **Data consent**
 - **Creative rights**
- Establish a **licensing framework** or **royalty-sharing mechanism** for content used in AI training.
- Promote **public awareness** on AI-generated content and creator rights.
- Encourage **AI transparency standards** and **algorithm audits**.

Source: <https://indianexpress.com/article/explained/explained-ai-copyright-law-10102888/>