
LEGISLATIVE PROBE ON THE BLACK BOX: WHY AI AUDITING IN ARTIFICIAL INTELLIGENCE REGULATION IS KEY TO PROTECTING INDIA'S INTELLECTUAL PROPERTY

Jiya Elizabeth Jikku, BA LLB (Hons.), Jindal Global Law School

ABSTRACT

Artificial Intelligence has perplexed the court of law. Especially in the widespread copyright violations done in the name of training large language models. For bare minimum in regulatory necessity, we need to understand the importance of the intersections at play. The copyright violation, AI Governance, India's interesting position in the AI race and Algorithmic Forensics. Looking at the compelling proposed Copyright Remuneration Collective for AI Training proposed by the Department of Promotion of Trade and Internal Industry and its shortcoming in expecting corporations to be transparent. We dissect current regulations worldwide. Including AI Safety Institutes and their non-enforcement technical expertise that if given legal capacity could revolutionize AI innovation as well as intellectual property protection.

This essay dissects current copyright litigation, corporate realities, technical solutions and the liability of the Indian judiciary to foster legal innovation. Demanding accountability through AI auditing to prosecute those that choose to violate Indian Intellectual Property law in the name of technicality. While also balancing the Indian mission for national AI innovation. By the conclusion we will see investing in AI Data Lineage Engineering in regulation is a double-edged sword that both facilitates the vision for AI Innovation while protecting Indian intellectual property from exploitation.

Introduction

In the current litigation of *Asian News International (ANI) vs OpenAI*¹ the judiciary becomes a battleground in the name of never-ending arguments of statutory interpretation and jurisdiction precedence. Section 52(1)(a)(i) of the *Indian Copyright Act, 1957*² is OpenAI's defense which negates copyright infringement (as defined under Section 51 of the Act) in cases where the work is used for research or learning purposes. In those cases, the use of the work is classified as "fair dealing". OpenAI equated web scrapping Asian News International's articles to human learning. A transformative "dissolving into weights". Where the copyrighted work joins a neural network of mathematical probabilities and parameters.

But the Indian test for "fair dealing" in this case is highly exhaustive. The purpose of the use was clearly commercial with market detriment to Asian News International. Causing them brand damage from the hallucination of the models crediting them. Along with the effect of viewership moving to get the same information from these Generative AI chatbots. Failing two major tests to classify this as fair dealing under Indian law. Commercial purpose and Market Detriment to the intellectual property holders.

But then OpenAI calls into question territorial jurisdiction. Holding up the transformative use doctrine under "fair use" in *United States Copyright Act, 1976*³. Stating the data training centers are located in United States jurisdiction, so it is U.S. law that is applicable. Not Indian frameworks. But if this argument is held up we will see a systematic exploitation of Indian intellectual property for enriching these foreign AI giants. It seems to be heading in that direction because in July 2026 Justice Amit Bansal denied Asian News Network's request for an interim injunction to block OpenAI from storing their news content⁴. Adopting the United States's transformative doctrine to call the use of the copyrighted work "fair dealing".

This is not an unconstitutional judgment. It is one made by carefully considering India's position in the AI landscape. Indian judges cannot give judgements that could potentially hinder our own AI innovation. But the lack of legislation on artificial intelligence is doing two things. Leading to a violation of India's rich intellectual property base. Exploited by foreign

¹ ANI Media Pvt Ltd v. Open AI OpCo LLC, 2026 DHC 5900.

² The Copyright Act, 1957, ss 51, 52(1)(a)(i).

³ The Copyright Act, 1976, ss 107 (U.S.A.).

⁴ Udayan Kishore, 'OpenAI can use copyrighted content to train ChatGPT: Delhi HC', THE NEW INDIAN EXPRESS, July 25, 2026 (Chennai).

corporations. But it is also the reason our own AI innovation culture is stunted. Especially with the monumental verdict that came from the case of *Bartz et al. v. Anthropic*⁵. Which is being called the largest copyright settlement in recorded history. Let us explore this question dynamically through this paper.

Interim Necessity

India is falling victim in a war of jurisdiction in an ever-evolving field. It will take years for a consensus to be reached on the balance between the protection of Indian intellectual property and Indian AI innovation. As it should, we are dealing with a delicate balance. But our current legislation is not supportive of the interim atmosphere that will persist till those infrastructures are set up. Article 53 of the *European Union Artificial Intelligence Act, 2024*⁶ allows creatives, publishers and other copyright holders to explicitly “opt out” of commercial AI companies scraping their intellectual property; by embedding a piece of machine-readable code on the web page intellectual property is on. Which if detected is a direct violation of European Union law. Jurisdiction is not an excuse. As long as the AI model is in use on E.U. soil, it cannot violate the intellectual property protected under this section. Irrespective of where the model was trained or where the corporation operates from.

Creatives in the Indian subcontinent would no longer have economic incentive to keep producing valuable intellectual property if the State fails to protect their right to do so. So while striking this balance, protecting India intellectual property should be the priority. Especially with the geopolitical nature of this dynamic. It is shameful for foreign corporations to be able to steal Indian intellectual property. But we can't pass a judgement that will shoot us in the back later on, hindering AI innovation.

“Proprietary Trade Secrets”

AI companies refusing to disclose the exact sets their LLMs are trained on is a reasonable argument in the face of it when we look at competitive realities⁷. The exact data set and years they spent training the model with their specific methodology are what sets one artificial

⁵ *Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (United States District Court for the Northern District of California, 2025).

⁶ E.U. Regulation 2024/1689, *Artificial Intelligence Act*, June 13, 2024, O.J.E.U. L-2024/1689, Art. 53.

⁷ A Shaji George, T Baskar and D Pandey, ‘Establishing Global AI Accountability: Training Data Transparency, Copyright, and Misinformation’ (2024) 2(3) *Partners Universal Innovative Research Publication (PUIRP)*.

intelligence model apart from another. Giving public access to this is out of the question.

This must be the most commonly brought up argument for why AI corporations do not disclose the datasets their use. *Section 8(1)(d)* of the *Right to Information Act 2005*⁸ negate this. Information including commercial secrets are to be released to regulators and public authorities when public interest is concerned. Almost all jurisdictions around the world agree with this. China's Personal Information Protection Law⁹, the E.U. Sustainability disclosure rules¹⁰, the South Africa's Promotion of Access to Information Act¹¹, the U.S. Freedom of Information Act¹² and for a more definitive example the European Union Artificial Intelligence Act¹³. Article 53 of the EU AI Act explicitly demands providers of "general purpose" AI to maintain detailed technical documentation on how the model was trained and tested. Again, excellent theory but slow implementation. That is the reality, but it is a strong start.

So why do these companies even try this angle? Most don't even keep a record of the database they used to train their AI models because they know having that record is a recipe for disaster. It would become easier to hold them accountable for the vast amounts of copyright violation they partake in. So it is not even that they can't record this data, they choose not to. This documentation deficit is the result of this gap in regulation allowing them to act above the law. When the regulation is not definite, we see this arduous headache of a practice. But when we have more legislations in place, it not only upholds the right of these corporations to innovate it also protecting intellectual property holders.

Now how would AI copyright regulation on these corporation even foster AI innovation. We only need to look at the recent verdict against Anthropic in *Bartz, et al. v. Anthropic PBC*.¹⁴ Recently in a lawsuit pioneered by Bloomsbury; Anthropic agreed to pay a whopping \$1.5 billion in settlement. The largest copyright settlement in recorded history. Roughly 500,000 distinct books were pirated off of shadow sites and for each book pirated and then data scrapped. The publishers and authors will receive a 1,500 dollar compensation each from

⁸ The Right to Information Act, 2005, ss8(1)(d).

⁹ The Personal Information Protection Law, 2021 (China).

¹⁰ E.U. Regulation 2019/2088, *Sustainability-related Disclosures in the Financial Services Sector*, November 27, 2019, O.J.E.U. L-317/1.

¹¹ The Promotion of Access to Information Act, 2000 (South Africa)..

¹² The Freedom of Information Act, 5 U.S.C. ss552 (U.S.A.)..

¹³ E.U. Regulation 2024/1689, *Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, June 13, 2024, O. J. E. U. L2024/1689, Art. 53.

¹⁴ *Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (United States District Court for the Northern District of California, 2025).

Anthropic. 3,000 dollars per book. Along with a 101.5 million dollar payout to the plaintiff's lawyer.

Over 91% of the eligible authors and publishers have successfully submitted the necessary paperwork to claim their share. This judicial precedent was a monumental step to uphold the respect for man-made copyrighted which was getting exploited through loopholes so openly. Sure the case only went through cause the copyrighter work was taken off illegal pirate websites. But it is a clear testament to how lack of regulation is just of detriment to everyone including AI innovation. This might be the one thing stunting India's growth in this field.

India's Interesting Position

If we are to look at the most respected tool used to rank countries based on their performance in AI innovation through multiple complex vertices like economic competitiveness, research, talent, infrastructure, etc. we get Stanford University's Global AI Vibrancy Tool¹⁵. The top country is the United States with frontier models like OpenAI's GPT models, Google's Gemini, and Meta's Llama. Globally commanding 50% of the industry. Closely following that we have a China with a bulk of academic research on the frontier and over 230 distinct AI industrial clusters developing models like Qwen3.7 Max, DeepSeek V4 Pro, etc. The most shocking 3rd place however goes to India herself. Overtaking countries like the United Kingdom and South Korea with a promising talent pool of AI developers and researchers.

Maybe it is because of the fact India has the largest technology workforce in the world according to the World Economic Forum¹⁶. Our digital public infrastructure, according to an article put out by The Economic Times on this milestone, is the reason we have been accelerate our commercial rollout when it comes to AI development¹⁷. This includes the open access to intellectual property.

You might then say why even legislate on this if our current inaction is leading us to this

¹⁵ Stanford Institute for Human-Centered Artificial Intelligence, *Global AI Vibrancy Tool* (2024), available at <https://aiindex.stanford.edu/vibrancy/> (Last visited on July 18, 2026).

¹⁶ Arundhati Bhattacharya, *Why India has the perfect conditions to make the most of digital technology*, WORLD ECONOMIC FORUM, May 16, 2023, available at <https://www.weforum.org/stories/2023/05/india-technology-opportunity/> (Last visited on July 15, 2026).

¹⁷ PTI, *India ranks 4th globally on AI performance; 5th in digital economy rankings: Report*, THE ECONOMIC TIMES, May 29, 2026, available at <https://economictimes.indiatimes.com/tech/artificial-intelligence/india-ranks-4th-globally-on-ai-performance-5th-in-digital-economy-rankings-report/articleshow/131389483.cms> (Last visited on July 15, 2026).

position. The answer is one of the biggest hurdles right now is our inaction. Without legal certainty, Indian corporations do not want to invest heavily in training large artificial intelligence models¹⁸. The current scenario is that most current AI models in India build “wrappers” on top of foreign models¹⁹. So currently it is not even Indian artificial intelligence innovation that is benefitting off our large intellectual property base, but the foreign companies out of our reach. A modern colonial loot.

We can see this point driven even further when we look at *Forbes*’s 2026 “50 AI²⁰” report that lists the top artificial intelligence corporations worldwide. The current trends and shifts. No major Indian AI corporation or startup is on that list. Despite India being 3rd on *Cambridge’s AI Vibrancy Tool*²¹. So AI innovation in India right now is scattered and unorganised. And the only way forward is for India as a country to legislate on the issues hindering major investment. Especially when it comes to the question of intellectual property used to train LLM models. Cause that seems to be the biggest financial risk for investors to invest the large sums needed to train a prominent model.

One Nation, One License, One Payment

The Department Of Promotion Of Industry And Internal Trade in their working paper on generative AI and copyright “*One Nation, One License, One Payment: Balancing AI Innovation and Copyright*²²” brings up another interesting point. The amount of data needed for generative AI models to train on is massive. Each source dissolved into a pixel in a massive neural network. The raw data obtained from internet scraping gets dissolved into weights in the training datasets turning into raw mathematical expressions for machine learning through pattern recognition. Determining exact liability is tedious²³.

¹⁸Prithvika Survepalli & Kanishk Goyal, *Balancing Innovation and Copyright Laws: A Comparative Analysis of AI Training Regulations and Roadmap for Legislative Future of India*, Vol. 3(1), JOURNAL OF LAW AND INTELLECTUAL PROPERTY RIGHTS, [Page] (2024).

¹⁹Nabodita Ganguly, *Indian AI startups confront new risks after Anthropic pulls Fable 5*, LIVEMINT, June 15, 2026, available at <https://www.livemint.com/ai/artificial-intelligence/indian-ai-startups-confront-new-risks-after-anthropic-pulls-fable-5-11781414145265.html> (Last visited on July 17, 2026).

²⁰Rashi Shrivastava ed., *Forbes 2026 AI 50 List: Top Artificial Intelligence Companies*, FORBES, April 16, 2026, available at <https://www.forbes.com/lists/ai50/> (Last visited on July 17, 2026).

²¹Rashi Shrivastava ed., *Forbes 2026 AI 50 List: Top Artificial Intelligence Companies*, FORBES, April 16, 2026, available at <https://www.forbes.com/lists/ai50/> (Last visited on July 17, 2026).

²²DPIIT, *One Nation, One License, One Payment: Balancing AI Innovation and Copyright (Working Paper, Part I)* (Department for Promotion of Industry and Internal Trade, December 2025).

²³Marc Hoag, *Why AI Training Should Fall Outside Copyright’s Domain: A Legal Analysis*, SSRN, (2025) available at <https://ssrn.com/abstract=5291815> (Last visited on July 20, 2026).

This is the way the *idea/expression dichotomy*²⁴ of copyright law is being debated worldwide in relation to artificial intelligence. Take the AI out of the picture. When it comes to copyrighted work the expression itself is what is protected. The specific words, musical notes or code used to propagate the work. The visual uniqueness of an artwork. That's the protected expression. But ideas are free. The concepts, themes, the revelations. That is public domain. This is extended to the debate on whether the copyrighted intellectual property is truly copyrighted. Stating that training models with intellectual property does not amount to copyright infringement because it is a non-expressive interpretation of the data anchored in statistical relationships²⁵.

Along with the fact demanding compensation on each individual intellectual property on a piece-by-piece basis from AI corporations, severely stunts the growth of the field domestically. Making AI innovation a classist field with a high economic barrier to entry based on the hundreds of thousands of sources needed to build efficient artificial intelligence models that hallucinates less. That's the one thing the *Department for Promotion of Industry and Internal Trade*¹ in their working paper on *Generative AI and Copyright*²⁶ gets the most right with the *Collective Royalties Collective for AI Training (CRCAT)*.

An Achilles' Heel of The Collective Royalties Collective For AI Training

In this proposal the collective is made up of organizations not individuals. One representative *Copyright Society* or *Collective Management Organization* for each per class of intellectual property like film, literature, etc. AI developers would have to disclose a detailed summary of all the sources they used to train their models through an *AI Training Data Disclosure Form*.

A *Rate Setting Committee* fixes royalty rates based on expert review of market trends, prices and all other facets. A flat percentage of the global gross revenue of the Developer's revenue. Which the organization then distributes in a preferably pro-rata basis to its copyright holders.

The problem is this entire system relies on the fact that Generative artificial intelligence

²⁴Mark A. Lemley, *How Generative AI Turns Copyright Upside Down*, [Page Number] (Stanford Law School Public Law & Legal Theory Research Paper Series, Working Paper No. 4517702, 2023) available at <https://ssrn.com/abstract=4517702> (Last visited on July 20, 2026).

²⁵ *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417 (ND Cal, June 25, 2025) (United States District Court for the Northern District of California).

²⁶DPIIT, *One Nation, One License, One Payment: Balancing AI Innovation and Copyright*, [Page Number] (Working Paper, Part 1, Department for Promotion of Industry and Internal Trade, 2025).

companies will be transparent when it comes to disclosure which is highly unlikely. The entire unique value proposition (UVP) of that company's model is the particular data set it used and the training technique used to train the model with that data set. Even if *Section 8(1)(d)* of the *Right To Information Act*²⁷ negates any claim by these companies wanting to protect their confidential information in the face of regulatory bodies. Arguments in the many policy scenario forecast reports like "*AI 2027*" by *Daniel Kokotajlo* and *The AI Futures Project*²⁸ team emphasis the geopolitical and quasi-sovereign consequences of this information not being protected. With the frontiers these companies are actively breaching and the state-backed commercial proxy warfare that we see among nations through their artificial intelligence developments. The countries with the strongest artificial intelligence systems are the ones that hold the power tomorrow. An AI race is the modern-day space race.

Regulations are kept lax internationally cause every country wants to foster innovation in this field. Reading the *NITI (National Institute For Transforming India) Aayog's* strategy blueprint "*Responsible AI #AIforAll: Approach Document for India*²⁹" we see the same aims. The principles for responsible AI like equality, transparency, accountability, privacy, etc. are balanced with an understanding of the necessity for AI innovation. Innovation and regulation are two sides of the same coin. In fact, innovation becomes more possible with regulation that protects its' citizens. When the intellectual backbone of our country is protected, we have the room for advancement. India's geopolitical situation in this whole AI architecture is interesting as we saw earlier. A repeat of this same caveat in a call for globalisation in a world that has seen colonisation. Data is the one truly globalised infrastructure.

The Current Nature Of AI Auditing

Currently when we look at how AI Safety engineers and L.L.M. Operation Engineers are utilised they are mainly a corporate entity working with artificial intelligence corporations to stress-test models before release. Building guardrails to prevent AI hallucination and negative thought patterns formed within the model's Blackbox by the L.L.M.s. They observe the model's algorithmic behaviours³⁰ and to some legal capacity ensure compliance to major

²⁷ The Right to Information Act, 2005, §8(1)(d).

²⁸ Daniel Kokotajlo et al., *AI 2027*, AI Futures Project, November 2025, available at <https://ai-2027.com/> (Last visited on July 21, 2026).

²⁹ NITI Aayog, *Responsible AI #AIforAll: Approach Document for India (Part 1: Principles)*, (February 2021), a available at <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf> (Last visited on July 21, 2026).

³⁰ OpenAI, *Software Engineer, AI Safety*, available at <https://openai.com/careers/software-engineer-ai-safety-san->

legislation that could hold the corporations accountable later like copyright laws and the E.U. AI Act. But this is within the profit-motive of the corporation. Compliance is secondary, the profit agendas of the company are priority. Even when evaluators like *METR* (Model Evaluation and Threat Research) were commissioned by corporations like Open AI and Anthropic to stress-test the danger capabilities of frontier models, we saw very troubling results. METR warning these companies to proceed with caution³¹.

When OpenAI was running its own internal safety test the advanced Open AI model *GPT-5.6 Sol* was able to break out of its digital box to breach another network *Hugging Face* to get the answer key to the test they were asked of. Open AI found software flaws to break into the production infrastructure of Hugging Face to steal secret information. Doing something that concerningly uncontrollable to fulfil what it was asked on. Open AI did not notice for weeks, and the model acted entirely autonomously. Because we haven't seen widespread damage yet, no legal action beyond a joint security statement came from this situation from Hugging Face and OpenAI³². But before finding out that OpenAI was behind this breach, Hugging Face had reported this incident to local law enforcement as a hostile intrusion³³. As a violation under the United States federal law of the *Computer Fraud and Abuse Act (CFAA)*³⁴.

Beyond this incident. Multiple internal engineers went on to become whistle blowers. Many researchers alleging rampant concealment of safety concerns from regulators. Like in the tragic suicide behind the case of *The New York Times Company v. Microsoft Corporation, OpenAI, Inc. et al.*³⁵

Suchir Balaji was an Indian American data scientist whose main occupation was to organise the massive amount of data scraped from the internet to train *GPT-4*. The end of 2024, he quit OpenAI and gave an exposing interview to The New York Times. His emails and internal corporate files were essential evidence for the case that preceded him. To attest that OpenAI

francisco/ (Last visited on July 10, 2026)

³¹METR, *Frontier Risk Report (February to March 2026)*, May 19, 2026, available at <https://metr.org/blog/2026-05-19-frontier-risk-report/> (Last visited on July 10, 2026).

³²OpenAI, *OpenAI and Hugging Face partner to address security incident during model evaluation*, July 21, 2026, available at <https://openai.com/index/hugging-face-model-evaluation-security-incident/> (Last visited on July 22, 2026).

³³ Hugging Face, 'Security incident disclosure — July 2026' (*Hugging Face Blog*, 16 July 2026) <https://huggingface.co/blog/security-incident-july-2026> accessed 22 July 2026.)

³⁴The Computer Fraud and Abuse Act, 18 U.S.C. §1030 (U.S.A.).

³⁵ The New York Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (SDNY 27 December 2023) (United States District Court for the Southern District of New York).

scrapped copyrighted books, news articles and more to train their models. Eight days after The New York Times legal team asked the court to designate Balaji as a key data custodian to allow legal collection of his internal OpenAI files for the active lawsuit; Suchir Balaji was found dead in his San Francisco apartment³⁶.

Banks are expected to produce financial statements that are thoroughly crosschecked. Food supply chains are observed. Almost every corporation is regulated to some extent by the nation they function in so why is it that a frontier this fragile is left largely unregulated. In a situation like this, engineer whistle-blowers like Suchir Balaji become a lifeline. AI models are intentionally secretive. It is hard for a claimant to decipher if their work was stolen one it has been fed into a system and transformed. Engineer whistle-blowers are often the ones that are able to produce the irrefutable paper trail. Not only providing the ingestion blueprint. But can play a heavy role in understanding mechanisms the company uses to systematically hide information away in complex neural networks³⁷. We see the role whistleblowers play intertwined with our laws. Balaji's death drove considerable momentum towards the proposition of the *The US AI Whistleblower Protection Act (AIWPA)*³⁸. The New York times had to hire their own set of highly specialised AI forensic experts to comb through the data eventually produced by OpenAI³⁹. Not an affordable course of action that anyone less than The New York Times could afford.

AI Safety Institutes

But we see examples of technical engineering expertise being slowly intertwined with AI regulation. An outstanding example to this is the United Kingdom's "AI Security Institute"⁴⁰ under the United Kingdom's Department for Science, Innovation and Technology (DSIT). The United Kingdom Prime Minister Rishi Sunak reached an agreement with OpenAI's Sam Altman, DeepMind's Demis Hassabis and Anthropic's Dario Amodei to run safety evaluations over their pre-release models. The U.K.'s Safety Institute sees considerable funding but we see

³⁶Alys Davies, *OpenAI whistleblower found dead in San Francisco apartment*, BBC News, December 14, 2024 (London), available at <https://www.bbc.com/news/articles/cd0el3r2nlko> (Last visited on July 22, 2026).

³⁷Apoorv Agarwal et al., *AI Whistleblowers - Regulators of Last Resort?*, Coventry University Research Blog, December 2, 2025.

³⁸The AI Whistleblower Protection Act, 2025, S. 1792, 119th Cong. (U.S.A.).

³⁹Rebecca Bellan, *New York Times says OpenAI hid evidence in ChatGPT copyright trial*, TechCrunch, July 9, 2026, available at <https://techcrunch.com/2026/07/09/new-york-times-says-openai-hid-evidence-in-chatgpt-copyright-trial/> (Last visited on July 22, 2026).

⁴⁰UK AI Security Institute, *UK AI Security Institute*, available at <https://www.aisi.gov.uk/> (Last visited on July 27, 2026).

an International Network Of AI Safety Institutes across countries like the United States, Japan, Singapore, Canada, the European Union and South Korea⁴¹. India's AI Safety Institute⁴² has been established recently in 2025 under the Ministry Of Electronics and Information Technology (MeitY) but it's priority is more so domestic development.

These government bodies test for possibility of criminal misuse of the models for cyberattacks, warfare, fraud, etc. But with no capacity for enforcement, all they can do is report back to the corporation with suggestions who still have full ability to release their models exactly how it is⁴³. These bodies are not involved with intellectual property breaches at all though which is usually assigned to intellectual property offices of each of these countries that lack the technical backing these non-enforcement bodies have. Right now, in India even the Intellectual Property Office of India does not even have jurisdiction in cases of intellectual property breach by artificial intelligence companies. Those cases mainly only being handled by the Indian court system once large-scale damage has happened. By the few parties that can even afford to bring the case to court.

A Road To Opening The Black Box

But the framework adopted by these AI Safety Institutes could be the solution for the Intellectual Property Office of India. An inter-mixing of technical AI Safety Specialists with adequate skill in studying the data lineages of generative artificial intelligence models. Along with those well-versed in AI legal Governance laws and its active evolving interpretation. How it can be read dynamically with Indian intellectual copyright laws and precedent. In an ever-changing world that is the only way forward.

Their jurisdiction can only be expanded by the *Parliament Of India*. The *Ministry Of Commerce and Industry*, specifically the *Department for Promotion of Industry and Internal Trade (DPIIT)*, would have to draft a bill that gets approval from both the *Lok Sabha* and *Rajya*

⁴¹Juhi Kore & Ebani Dhawan, *The Global Landscape of AI Safety Institutes*, All Tech Is Human Blog, March 14, 2026, available at <https://alltechishuman.org/all-tech-is-human-blog/the-global-landscape-of-ai-safety-institutes> (Last visited on July 27, 2026).

⁴²IndiaAI Mission, *India Takes the Lead: Establishing the IndiaAI Safety Institute for Responsible AI Innovation*, January 31, 2025, available at <https://indiaai.gov.in/article/india-takes-the-lead-establishing-the-indiaai-safety-institute-for-responsible-ai-innovation> (Last visited on July 27, 2026).

⁴³Renan Araujo, *Understanding the First Wave of AI Safety Institutes: Characteristics, Functions, and Challenges*, IAPS, October 7, 2024, available at <https://www.iaps.ai/research/understanding-aisis> (Last visited on July 27, 2026).

Sabha. Along with Presidential assent⁴⁴.

Only Part one of the *Working Paper on Generative AI and Copyright*⁴⁵ has been released by the *Department for Promotion of Industry and Internal Trade (DPIIT)*. It's an enthusiastic model. An excellent starting point. We discussed earlier the frameworks they suggested like the *Copyright Royalties Collective for AI Training* and more that solves preliminary problems when it comes to compensation for copyrighted work used to train large language models (LLMs).

The Copyright Royalties Collective for AI Training is a necessary framework that with benefit everyone. But the overly optimistic part to me was that there is a presumption that AI corporations would be transparent about the datasets they used to train their models in the mandated disclosure form. The only liability that arises if they are not transparent is the burden of proof being shifted to the AI Developer if copyright infringement litigation were to struck up against them. In the Anthropic case, we only got a verdict against them because they used the shadow libraries *Library Genesis (LibGen)* and *Pirate Library Mirror (PiLiMi)* to download the millions of unauthorised copyrighted work. You would think that between choosing between the considerably lower flat percentage they would have to pay before releasing their models and the statutory damage per individual work if found liable; AI Developers would choose the flat percentage by being honest in the disclosure and not hiding the work used to train their sources.

But in a world where litigation against large AI corporations is a privilege led by other equally large corporations like news networks and publishing houses; how can an individual copyright holder fairly wage a war to protect their work if they have an inkling their work has been used to train a model. How can an individual copyright holder even evaluate whether their work their work has been used to train a model. In the analytical article "Who's suing whom in AI?" we see two clear classes major corporations like publishing house, entertainment corporation getting the limelight in their AI copyright litigation⁴⁶. Then if an individual wants to protect

⁴⁴MINISTRY OF COMMERCE AND INDUSTRY, Department for Promotion of Industry and Internal Trade, *Annual Report 2025-26*, 45 (2026).

⁴⁵DEPARTMENT FOR PROMOTION OF INDUSTRY AND INTERNAL TRADE, *One Nation, One License, One Payment: Balancing AI Innovation and Copyright*, [Page] (Working Paper, Part 1, 2025).

⁴⁶David McCandless, *Who's Suing Whom in AI?*, INFORMATION IS BEAUTIFUL, June 2026, available at <https://informationisbeautiful.net/2026/whos-suing-whom-in-ai/> (Last visited on July 27, 2026)..

their intellectual property we see large scale class action law suits⁴⁷.

Doesn't that mean it is the fundamental duty on the government body to evaluate these models themselves to protect the rights of the Indian citizen to add to the vibrant intellectual property base of the nation.

Only *Part 1* of the *Working Paper on Generative AI and Copyright*⁴⁸ has been released. Currently framed as a two-part interrogation of two major areas of debate. With Part 1 focusing on input. So the ingestion of copyrighted work to train the models. And Part 2 which is to be released soon dealing with output. So questions of ownership of AI generated work. But the official document structure of the paper maps out multiple other phases to be explored in the future. Part 3 focuses on an operational action plan for the Copyright Royalties Collective for AI Training.

And the only way for this Collective to function to its necessary capacity is to adopt the frameworks we see across the Safety Institutes all around the world. These Institutes are not just staffed by policy makers and advocates. We see leading technical staff from various industries like senior alumni who worked with artificial intelligence and technology. To evaluate, observe and trace the data lineages of these models to cross check whether an individual intellectual property has been exploited as an individual right.

Conclusion

India is in a special position as we discussed. We are third when it comes to AI innovation worldwide, but our AI landscape is still scattered. This is the best situation we could find ourselves in. We still have the power to set a precedent in AI regulation. As complex as the landscape is, without mega giant corporations to mitigate we have the breathing room to be a powerhouse in AI regulation. Especially with our respect for our considerable intellectual property base. Giving the frameworks we see in these AI Safety Institutes the enforcement power that seems constitutionally sane. Not only protecting the right of our citizens to keep producing intellectual property with dignity. But also fostering AI innovation that thrives

⁴⁷Theresa M. Weisenberger, *Case Tracker: Artificial Intelligence, Copyrights and Class Actions*, BAKERHOSTETLER, available at <https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/> (Last visited on July 27, 2026).

⁴⁸DEPARTMENT FOR PROMOTION OF INDUSTRY AND INTERNAL TRADE, *One Nation, One License, One Payment: Balancing AI Innovation and Copyright*, [Page] (Working Paper, Part 1, 2025).

without legal uncertainty.

India has the potential to be more than 3rd in AI innovation internationally. But the only way forward is if our policy keeps up with innovation. We owe it to the intellectual engineering youth of our nation. We owe it to the artists, writers and creators in our nation. The legal system and our government bodies now stand at an important juncture in India's economic story. It is no longer an economic question. But whether India respects their fundamental right to freedom of profession under *Article 19(1)(g) of The Indian Constitution*⁴⁹.

⁴⁹ The Constitution of India, 1950, Art. 19(1)(g).