AZB & PARTNERS
ADVOCATES & SOLICITORS

Regulating Generative AI – Series I: Balancing Innovation and Legal Protections for Human Content

1. Generative AI ("**GenAI**") has risen to the forefront of technological innovation, enabling automated content creation and sophisticated machine-generated outputs across countless domains. Yet, this rapid progress has generated a complex legal conflict for multiple stakeholders, from content creators and media organizations to technology startups and AI giants. Central to this struggle is the question of who holds the rights to the human-generated material that is used to train AI systems. This tension is vividly illustrated by the open conflict between OpenAI, the developer of ChatGPT, and ANI, a major Indian news agency. The dispute, currently before the Delhi High Court, brings into focus core copyright issues: whether using and storing news content for AI training amounts to infringement, the applicability of "fair use," and the power of Indian courts to intervene when AI servers reside outside the country. At stake is how emerging regulations will balance robust legal protections for rights-holders with the freedom to innovate—a tension playing out not only in India but across jurisdictions worldwide.

2. *Global and indigenous models of governance*

2.1. Around the world, governments are racing to regulate AI in a way that protects society without hampering innovation. The European Union has led with Regulation 2024/1689 ("**EU AI Act**"), which adopts a risk-based classification of AI systems[i]. It bans uses deemed "unacceptable risk" (like social scoring), heavily regulates "high-risk" AI (such as in healthcare or finance), and imposes transparency for limited-risk systems (e.g. requiring users be told when they interact with AI). Notably, the EU AI Act will also require providers of general-purpose AI models to disclose information about their training data and comply with copyright rules.

2.2. In the United States, while there is no omnibus federal AI law yet, regulation has emerged through executive action and sectoral guidelines. President Biden's 2023 Executive Order on AI directed federal agencies to ensure AI safety and called for standards to identify AI-generated content (to combat deepfakes) while urging the U.S. Copyright Office and Patent Office to address AI's intellectual property challenges. The U.S. Copyright Office and Patent Office has publicly published some of its conclusions basis its reviews, noting that the United States requires enhanced federal laws to protect citizens against the risk of digital replicas (alluding to deepfake content)[ii], and that existing copyright questions raised by AI can be resolved under current law without new legislation, emphasizing that AI-assisted works require case-by-case inquiries into human authorship[iii].

2.3. Meanwhile, China has taken a prescriptive approach: the Interim Measures for Generative AI (effective 2023)[iv] jointly issued by the Cyberspace Administration of China along with six other regulators, require AI providers to use only lawfully sourced data, avoid infringing others' IP, obtain consent for personal data, and label AI-generated

content. These rules align with China's broader goals of ensuring AI follows "social morality and ethics" and safeguarding national security and public interests.

2.4. Other jurisdictions vary – for example, the UK has opted not to pass a single AI law yet, instead issuing guiding principles to be applied by existing regulators in each sector. In short, global regimes range from the EU's comprehensive, binding regulations to the US and UK's softer, adaptive frameworks, and to China's state-centric, stricter controls. Each reflects a balance between fostering AI growth and mitigating risks.

2.5. In India, there is no specific codified law dedicated to AI yet. Instead, a patchwork of policies and general laws guide AI governance. The government released a National Strategy for AI (2018)[v] and Principles for Responsible AI (2021)[vi] to outline ethical AI development. These are policy frameworks, not binding law, but signal India's intent to encourage AI for social good with safeguards. Existing statutes in technology and data protection fill some gaps. The Information Technology Act, 2000 ("**IT Act**"), to be replaced by a proposed Digital India Act, and the new Digital Personal Data Protection Act, 2023 (not yet in force) ("**DPDP Act**") are technology-agnostic laws that apply to AI activities involving data handling. For example, an AI system's collection and processing of personal data must comply with the privacy protections of the extant Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011 ("**SPDI Rules**"), and subsequently the DPDP Act. Similarly, if an AI system causes harm through a computer resource in India, provisions like Section 75 of the IT Act could extend jurisdiction for offenses committed abroad.

2.6. The evolving landscape of AI governance in India took a pivotal turn in March 2024, with the Ministry of Electronics and Information Technology ("**MeitY**") issuing advisories that highlighted the balance between innovation and responsibility. Initially, the advisories mandated prior government approval for deploying under-tested AI models, sparking industry concern over stifling innovation[vii]. However, a revised approach emphasized self-regulation and accountability, marking a shift toward collaborative governance.[viii]

2.7. An Advisory Group, chaired by the Principal Scientific Advisor, was constituted to develop an 'AI for India-Specific Regulatory Framework.' A Subcommittee on 'AI Governance and Guidelines Development' was tasked with providing actionable recommendations for AI governance in India. Its report, published for public consultation on January 6, 2025[ix], aims to guide the development of a trustworthy and accountable AI ecosystem. The AI Governance principles align with globally recognized frameworks, including OECD, NITI Aayog, and NASSCOM. They emphasize transparency, accountability, and safety in AI systems, advocate for privacy, security, and fairness, and promote sustainable innovation. These principles encourage a "digital by design" governance approach leveraging technology to enhance compliance and regulation while ensuring inclusivity and human-centred oversight. The report also addresses intellectual property and AI biases. It examines challenges such as liability for training models on copyrighted data and the copyrightability of AI-generated works, alongside the need for appropriate guidance. On AI biases, the report highlights risk of discrimination at scale, stressing transparency and risk mitigation. These issues underscore the need for governance frameworks that ensure ethical AI deployment. By aligning with global standards, addressing key issues, and formulating a "digital by design" approach, the

recommendations aim to shape an AI ecosystem that is ethical, transparent, and future-ready.

2.8. Importantly, India's intellectual property laws – especially the Copyright Act, 1957 ("**Copyright Act**") – apply fully to AI. There is no AI-specific IP legislation, meaning that use of copyrighted material in AI training or outputs is assessed under the same rules developed for human conduct. The Copyright Act grants authors exclusive rights (reproduction, adaptation, communication to public, etc.)[x] and lists limited exceptions under Section 52 ("fair dealing" and other permitted uses). As discussed below, the interpretation of these general IP provisions in the AI context is now being tested in Indian courts. Overall, India's regulatory landscape for AI is in flux: policymakers have hinted that upcoming legislation (like the Digital India Act) may address AI more directly, but for now legal professionals must navigate AI issues through existing IT, data protection, and IP laws, informed by guiding principles and comparative global developments.

3. *Challenges by media companies for usage of data by AI systems*

3.1. Across jurisdictions, media and content companies are pushing back against GenAI models' use of their copyrighted material. In the United States, for example, the New York Times Company filed a high-profile lawsuit in 2023 against OpenAI (creator of ChatGPT) and its investor Microsoft[xi]. The Times alleges that OpenAI scraped millions of Times articles (and other content from the web) without permission to train its large language model. The complaint states that ChatGPT can, when prompted, sometimes produce verbatim excerpts from Times articles or detailed summaries of their investigative findings, effectively reproducing copyrighted content and undermining the Times' subscription and licensing revenues. OpenAI has called the lawsuit "without merit" and moved to dismiss some claims[xii].

3.2. Another landmark case is Getty Images v. Stability AI, raised in the United Kingdom[xiii]. Getty Images, a global stock photo provider, accuses the makers of Stable Diffusion (an image-generating AI) of copying over 12 million Getty photos (along with captions and watermarks) to train the AI model without a license. Getty argues this massive unauthorized use of its curated, professional photographs violates copyright and even claims trademark infringement, especially as some AI-generated images mimicked Getty's watermark logo, implying the mechanisms creating the output trained on Getty's library, and identified the watermark as a required element of finished works. Stability AI, in its defence, does not deny using Getty images but contends that the training process does not store or reproduce the photos in a way that infringes individual works. It argues that the AI model does not "memorize" images, rather, it abstracts features and stores only numeric weights, so that no substantial part of any single photograph can be retrieved from the model. In the UK proceedings, Stability AI has even asserted a novel fair dealing argument, claiming that any similarity of its outputs to Getty's images is incidental and that the outputs are essentially "pastiche" works.

3.3. These global disputes highlight a common legal question: can AI developers use copyrighted content without consent to develop new AI tools, or does this practice trespass on content owners' rights? In the U.S., the issue is framed by the flexible fair use doctrine – OpenAI and Stability AI effectively argue that using copyrighted data to teach an AI model is a transformative, non-expressive use, akin to how Google's scanning of

books to enable search was deemed fair use by courts. By contrast, content owners argue that AI training is an unauthorized act of reproduction and that the resulting outputs may substitute for the originals, thus harming the market for the creative works. A recent study using data archaeology reveals that OpenAI models, including ChatGPT and GPT-4, have memorized a vast range of copyrighted books, with memorization correlating to their online availability—further complicating fair use debates and reinforcing calls for transparency in AI training data[xiv].

3.4. Indian law does not (yet) have direct precedents for these AI copyright issues, but the principles are generally less permissive than U.S. law. India's Copyright Act grants a bundle of exclusive rights to creators, and its exceptions (termed "fair dealing") are narrower and enumerated (for example, use for private research, criticism, or news reporting) rather than a broad, open-ended fair use clause. Using entire news articles or millions of photos to train an AI may prima facie amount to a reproduction of such works. Unlike in the U.S., Indian law has no statutory concept of transformative use, and any defence must fit one of the Section 52 exceptions or similar doctrines. For instance, using an article for the purpose of "reporting current events" is allowed, but an AI scraping news to generate answers is not clearly the kind of news reporting the statute envisions. Additionally, India has no sui generis database rights, but copyright can subsist in original databases or compilations, so mass-extraction of a media website's content could implicate those rights as well. In sum, the legal challenges raised by the New York Times and Getty would likely find sympathy under Indian copyright principles – absent a specific exception, such wholesale unlicensed use of content would be seen as infringing. Indeed, these very issues are now playing out in India with domestic media companies taking OpenAI to court, as discussed next.

4. *Preliminary examination of questions raised in the OpenAI v ANI case before the Delhi High Court*

4.1. India's first major confrontation over GenAI and copyright is the case of ANI Media Pvt. Ltd. v. OpenAI, filed in 2023 and currently pending before the Delhi High Court. ANI, a prominent Indian news agency, alleges that OpenAI misused its news articles (and those of other publishers) to train ChatGPT without permission. Several other Indian publishers, including the owners of major outlets like NDTV, Hindustan Times, and Indian Express, have since joined or expressed support for the suit. The Delhi High Court has framed four key preliminary legal issues for considerations in this case, which resonate with the questions being debated globally. We examine each in turn:

4.2. <u>Does storing copyrighted news data for AI training amount to infringement?</u>

(a) OpenAI allegedly stored and used ANI's news content in its training dataset for ChatGPT. The court will decide if this act of copying content into an AI training corpus violates the Copyright Act.

(b) Under Section 14 of the Indian Copyright Act, the right "to reproduce the work in any material form" belongs exclusively to the copyright holder. Downloading or scraping articles and storing them (even digitally) for AI training creates copies of those articles. Unless a defence applies, this is an unauthorized reproduction.

(c) OpenAI's potential defences draw on both facts and law. Factually, OpenAI might argue that it did not store entire articles in a human-readable way but processed them into abstract data (similar to Stability AI's argument in the Getty case). However, from a legal standpoint, even intermediate copying for analysis can infringe if not permitted. Indian law does recognize an exception for transient or incidental storage as part of a technical process. Section 52(1)(b) and (c) exempt certain temporary copies made in the course of using a work or for facilitating electronic transmission. If OpenAI's use of ANI's content was truly transient caching, it might invoke this exception. But training an AI typically involves storing data for a non-trivial duration and iterative processing, which goes beyond the fleeting network cache contemplated by Section 52. Moreover, the scale (entire articles) and purpose (commercial AI training) would likely lead a court to view this as substantive copying, not merely incidental processing.

(d) Indian jurisprudence suggests that copyright subsists in the expression of facts (e.g. the writing of a news report), even if the underlying news itself is not protected (the facts of the recent India v Pakistan conflict may not be copyrightable, but a creative commentary on the same would be eligible for protection). Simply because the data is used for technology (i.e. machine learning for GenAI applications) would not remove it from the ambit of copyright – the act of making a copy is what matters. In *Eastern Book Company v. D.B. Modak*[xv], the Supreme Court found infringement where a competitor copied a compilation of case law headnotes into an electronic database, underscoring that digital reproduction without authorization violates Section 14 (even though the content was being used in a new medium). Similarly, OpenAI's intermediate copying of news text for model training can infringe copyright, as Indian law does not recognize a general exemption for "intermediate" or "technical" copies outside specific contexts (like certain computer program uses). Indian precedent on digital copying reinforces that the medium is irrelevant – unauthorized reproduction, whether via photocopy (*Sony Pvt. Ltd. v. HMD Pvt. Ltd., 2018*) or digital means (*Super Cassettes Industries v. MySpace, 2011*), is actionable. In MySpace, the Delhi High Court held the platform liable for users' unlicensed uploads of videos, indicating that even temporary data storage on a server can infringe unless excused by law. Here, OpenAI's direct role in scraping and storing content makes the case even stronger against it. The absence of a clear exception for AI training in the Indian Copyright Act suggests that, on current law, storing copyrighted news data to "train" an AI is unauthorized and infringes copyright, subject to the fair dealing analysis discussed later.

(e) It appears likely that storing copyrighted news content to train an AI does constitute infringement of the reproduction right under Indian law, unless OpenAI can show that the copying was minimal or purely transient. The case will force the court to consider if training data copies are more like infringing photocopies or more like temporary RAM copies. Given the lack of a clear exemption for data mining in the Copyright Act, the safer view is that ANI's rights were prima facie infringed when its articles were ingested into the training dataset.

4.3. <u>Do AI-generated responses using copyrighted material infringe copyright?</u>

(a) The Copyright Act gives owners' rights over adaptations and reproductions of their works. Section 14 covers the rights "to reproduce the work..." and "to make any

translation or adaptation of the work". An AI-generated response that closely paraphrases or excerpts from a protected article could be seen as an unauthorized reproduction or adaptation. The nuance here is that the output is generated by the AI in response to a prompt, not directly by a human acting intentionally to copy the text. Nevertheless, if the AI's response contains a substantial part of the original expression from ANI's article, it infringes just as a human plagiarizing the article would.

(b) Indian courts determine infringement by assessing substantial similarity and the qualitative value of what was taken. The Supreme Court's seminal test in *R.G. Anand v. Deluxe Films*[xvi] asks whether an ordinary observer would "*unmistakably infer that the subsequent work is a copy of the original*". If ChatGPT produces text that would lead a reader to recognize ANI's article in it (whether via identical phrasing or a very close paraphrase), that output is likely infringing. Notably, ANI alleges instances where ChatGPT's news answers showed "90 percent similarity" to ANI's original wording (even if not a word-for-word copy). Such near-verbatim outputs plainly implicate the reproduction right. Even without verbatim copying, paraphrased summaries can infringe if they capture the protectable elements of the original. For example, an unauthorized translation of a literary work is an infringement in India – the Madras High Court in *Blackwood & Sons v. A.N. Parasuraman*[xvii] held that translating an English book into Tamil without permission infringed the original work. By analogy, AI-generated re-writings or condensations of ANI's articles could infringe the adaptation right, especially if such re-writings / condensations build on the creative expressions used in the underlying copyrighted content. The Bombay High Court's injunction against an unlicensed translated version of the Bhagavad Gita in the case of *Bhaktivedanta Book Trust v. Thomson Press (India) Limited*[xviii] is a reminder that the adaptation right is robust – one cannot simply rephrase or re-express someone else's text without permission.

(c) The Court may also consider whether the AI's output "*usurps*" the market of the original work. The reproduction right is infringed not only by literal copying but also by creating a work that "*substitutes for the original*". If a user can get the substance of an ANI article from ChatGPT, then the output is effectively playing the role of a derivative news product. Given the current allegations (verbatim quotes and close paraphrasing), the outputs likely do infringe unless excused by fair dealing.

4.4. <u>Does the use of copyrighted content by OpenAI qualify as "fair use" (fair dealing) under Section 52 of the Copyright Act?</u>

(a) OpenAI's key defence is likely to be that its activities fall under "fair dealing" exceptions in Section 52 of the Act, meaning they are permitted uses of copyrighted material without license. The crux is whether OpenAI's copying, and use of ANI content can fit into any of these categories or be deemed analogous to them. Section 52 provides a list of scenarios where use of copyrighted material is not an infringement. Relevant clauses include use for research or private study, criticism or review, reporting current events, transient or incidental storage (as noted above), and certain library or educational uses. India's fair dealing is narrower than the open-ended U.S. fair use. For OpenAI to succeed, it must bring its use within one of these permitted purposes.

(b) OpenAI might argue that ingesting ANI's content to train the model is a form of research or private study. Indeed, Section 52(1)(a)(i) allows fair dealing with any work for private use, including research. However, Indian courts have construed "*research*" in this clause to mean research by an individual or non-commercial scholarly use, not the wholesale copying of works for a commercial product. In the case of *The Chancellor, Masters & Scholars of the University of Oxford & Ors. v. Rameshwari Photocopy Services & Ors.*[xix], the court did extend fair dealing to a university's course-pack preparations, treating the students' use as private study and the photocopy shop as an agent of that use. But crucially, that was in an educational, non-profit context benefiting students. OpenAI's training of ChatGPT, by contrast, is part of a for-profit enterprise and not "*private*" in the ordinary sense – it is akin to an R&D project for a commercial service. There is no Indian precedent squarely holding that a company's data mining of copyrighted works for product development counts as "research" under Section 52. On the contrary, courts are wary of broadening fair dealing to cover general commercial use. Justice Endlaw in the aforesaid case observed that copyright law's exceptions are meant to further the public interest (education, research) and that copyright "is not an inevitable, divine, or natural right" but a statutory grant to encourage creativity balanced against public welfare. Here, OpenAI would need to convince the court that training an AI on news data has a comparably strong public interest element (e.g. advancing knowledge or AI innovation) that justifies reading it into "*research*" or another exception. Given that Section 52 is exhaustive in listing exceptions, this is a challenge – Indian courts generally do not create new exceptions judicially beyond those listed.

(c) Another angle is whether ChatGPT's act of providing answers containing ANI-sourced information could fall under "reporting current events" (Section 52(1)(a)(iii)) or some form of commentary. This exception permits fair dealing for "the reporting of current events… including the reporting of a lecture in public," typically understood to allow news media to quote or summarize others' content when reporting news. OpenAI is not a news reporter, but when ChatGPT tells a user about a news event, one might characterize that as a form of news reporting (albeit automated). However, courts will likely require that such use be genuine reporting and accompanied by "*fair practice*" such as attribution. In *Super Cassettes Industries Ltd. v. Hamar TV*, a TV channel invoked "reporting current events" to justify using T-Series' music in its programming, but the court rejected this defence because the use of entire songs was unrelated to actual news reporting and lacked transformative context. Justice Shakdher in that case laid down that courts must adopt a "liberal approach" to interpreting criticism or reportage exceptions but also examine whether the use was truly for those purposes or a pretext. For ChatGPT, simply regurgitating news content on request may not qualify as "reporting" in the intended sense, since ChatGPT is not creating its own journalistic work – it is arguably merely redistributing ANI's work. Additionally, the lack of source attribution (users may not be told that the content originates from ANI) weighs against fair dealing, as fair dealing for news typically assumes the source is acknowledged.

(d) In *Civic Chandran v. Ammini Amma*, the defendant had dramatized portions of a published play in his own stage play to critique the original. The court treated this as fair dealing for criticism or review, laying out a three-factor test: (1) the quantity and value of the matter taken, (2) the purpose of taking it, and (3) the effect on the potential market of the original work. Notably, it held that even substantial copying

can be fair if done to critique or parody the original and not to compete with it. This "transformative" use (turning the original into a critique) was allowed because it furthered public discourse and did not usurp the original's market (people would not watch the new play as a substitute for the old drama). If we apply similar reasoning to OpenAI, one could argue that training ChatGPT is transformative – the AI is not providing a verbatim archive of ANI stories but using them to generate new text and engage in conversations, potentially providing a social benefit (AI services). However, the purpose behind ChatGPT's use of ANI content is not to comment on ANI's journalism nor to critique it, it is purely to utilize the information for AI responses. That is a commercially driven purpose rather than a protected purpose like commentary, criticism, or news reporting in the public interest. Also, unlike a parody or critique, ChatGPT's use is not creating a new work about the original – it is simply drawing from the original to answer queries. This makes it harder to fit under existing fair dealing heads.

(e) The fair dealing defence in this case is novel and its success is uncertain. On one hand, one could argue that training an AI is a non-expressive, transformative use – the AI does not republish articles, it "learns" from them, arguably similar to a human learning fact and writing an original piece later. If the court is forward-looking, it might carve out some flexibility here, especially considering India's ambition in AI development. On the other hand, Indian law's enumerated exceptions do not plainly cover AI training, and courts may defer to the legislature for any new exemption. It is possible the court will rule that OpenAI's use is not protected by fair dealing as currently written, effectively putting the onus on AI developers to seek licenses or on lawmakers to update the law. The outcome on this issue will be crucial in setting the balance between IP rights and AI innovation in India.

5. *Regulatory Balancing Act: Protecting Rights vs. Fostering Innovation*

5.1. The tension between innovation and regulation is at the heart of the AI debate. On one side, strong regulation is advocated to protect intellectual property, privacy, and data security:

(a) Without legal protections, GenAI could disincentivize content creation. Media companies, publishers, artists, and software developers invest heavily in content, and if AI firms can appropriate this content freely to train models, it undermines the value of copyrights and licenses. Robust IP enforcement or new licensing regimes would ensure creators are compensated and encourage a sustainable creative ecosystem. For example, the Indian news outlets in the OpenAI case argue that unregulated AI scraping poses "a clear and present danger" to the value of their copyrights. Strong regulation (or court rulings) in favour of such rights could push AI developers to negotiate licenses or use public domain/Government-licensed data, thereby respecting the incentive structure copyright law is meant to preserve.

(b) GenAI models are often trained on vast troves of data, which may include personal information scraped from websites or social media. In the absence of regulation, individuals' privacy could be compromised – AI might learn and regurgitate personal data or sensitive information. Regulations can enforce principles of data minimization and consent. The EU's GDPR already limits how personal data can be processed, and India's DPDP Act will require compliance with consent and purpose

limitations. Privacy regulators have started taking action – notably, Italy's Data Protection Authority temporarily banned ChatGPT in 2023 over privacy concerns, prompting OpenAI to implement age checks and allow users to opt-out of data use. Proponents of strong AI oversight argue that clear privacy rules (like requiring companies to anonymize training data or honour "do not train" flags) are critical to prevent AI from becoming a surveillance or data leak tool.

(c) There are also security concerns – AI systems could be manipulated to reveal confidential training data or produce harmful outputs. Without regulation, an AI might inadvertently output proprietary code or classified information if such data were in its training set. Strong regulation could mandate testing and safeguards to prevent data leakage and to harden AI against malicious use. Additionally, ensuring that AI models are transparent and auditable can help detect biases or vulnerabilities, thereby protecting users and society from downstream harms. Regulators might require AI providers to conduct risk assessments (as the EU AI Act will) and share these with authorities, ensuring models do not compromise financial systems, critical infrastructure, or individual security.

5.2. On the other side, there are warnings that excessive regulation may stifle AI innovation and economic growth:

(a) Heavy-handed rules could raise the cost of developing AI dramatically. For instance, if every piece of training data required individual licensing or compliance checks, only tech giants could afford to build advanced models, squeezing out startups and researchers. Innovation often thrives in sandboxes – too many legal hurdles might deter entrepreneurs from experimenting with new AI applications. Countries like the UK explicitly chose a lighter, principles-based approach to avoid hampering their AI sector. Over-regulation could slow down beneficial AI advancements in medicine, education, and business by entangling developers in red tape or legal uncertainty.

(b) AI is a frontier technology race. If India (or any country) imposes onerous regulations while others foster AI more freely, companies might relocate R&D to more permissive environments. An excessive regulatory burden could cause India to lag in AI adoption and innovation, ceding ground to jurisdictions with balanced or innovation-friendly rules. Policymakers fear a scenario akin to the biotech or cryptocurrency domains, where too strict a regime at home drives talent and capital elsewhere. Thus, industry advocates urge a calibrated approach that protects core rights without stifling the AI ecosystem's growth.

(c) A law that is too rigid might become quickly outdated in the fast-evolving AI field or inadvertently ban harmless uses. For example, a rule broadly prohibiting use of copyrighted data could outlaw even benign text-and-data mining for research or for training AI to detect plagiarism. The challenge is that AI's technical inner workings are complex – poorly tailored laws might capture socially beneficial AI uses in an attempt to target the harmful ones. Therefore, some argue for industry self-regulation and technical solutions as a first resort (such as content owners using meta-tags to signal "do not scrape" and AI companies voluntarily complying), combined with narrow laws targeting truly abusive practices. This lighter touch, coupled with government investment in AI education and infrastructure, is seen as key to maintaining the momentum of innovation.

5.3. In practice, the regulatory trend is trying to find a middle path. Many jurisdictions aim to "encourage AI innovation while protecting fundamental rights". The EU, for instance, coupled its strict rules with support for sandboxes and research exceptions to help innovators comply. India, too, will have to strike this balance. It may consider developing a framework that, for example, recognizes a text/data mining exception with conditions – allowing AI training on data as long as certain transparency and opt-out mechanisms are honoured (an approach seen in the EU and Japan) – thus giving AI developers access to data but respecting rights of those who opt out. On the flip side, clear guidelines or perhaps a statutory license model could be introduced, so that using copyrighted Indian content for AI requires either authorisation or payment of a fee to a collective, ensuring creators benefit from AI uses of their work. Finding this equilibrium is crucial to both protect stakeholders like media companies and to ensure that regulation does not become a proverbial wrench in the gears of AI progress.

6. *Conclusion*

6.1. The emergence of GenAI has set up a complex intersection between law and technology. As seen from the global and Indian developments, regulators and courts are grappling with how best to protect legal rights without unduly restraining technological innovation. A one-sided approach will either chill the promise of AI or leave creators and citizens vulnerable to exploitation. The ideal outcome – for industry and society – is a balanced regime that provides legal certainty and safeguards: AI innovators should have clarity on what is permitted (perhaps through well-defined exceptions or licensing frameworks), and content owners and users should have their intellectual property, privacy, and security rights respected.

6.2. In the Indian context, the OpenAI v. ANI case will likely be a bellwether. If the court can craft a nuanced doctrine (for example, by recognizing certain AI uses as fair dealing under conditions, or by affirming infringement but urging the legislature to act), it could set the stage for an equilibrium. In the future, as AI-generated content proliferates and models train on each other's outputs, the very foundation of intellectual property rights will face an existential challenge—one where the original human-authored work may become increasingly untraceable in a recursive loop of AI-to-AI learning. This evolution will likely fuel a shift from traditional copyright frameworks toward more dynamic, model-specific licensing regimes, where AI developers must track provenance through watermarking, data transparency obligations, or blockchain-based attribution systems. Additionally, as competition intensifies, regulatory focus may expand beyond fair use debates to address issues of monopolization, proprietary model extraction, and cross-border enforcement. The next major lawsuits may not just be media companies versus AI firms but rather AI firms litigating against each other over unauthorized "knowledge distillation"—a battle where corporate espionage, data exfiltration, and model cloning become the new frontier of intellectual property disputes.

6.3. Similarly, forthcoming Indian legislation, potentially the Digital India Act or amendments to the Copyright Act, may explicitly address AI. The hope is for regulations that foster collaboration – encouraging tech companies to work with content creators (via licensing deals or partnerships) and to build compliance (privacy filters, attribution systems) into their AI. This way, innovation can continue in a manner that also creates new revenue models and respect for rights, rather than a zero-sum game. In conclusion,

regulating GenAI requires a scalpel, not a sledgehammer. Strong protections are necessary to prevent abuse and protect what society values – creative works, personal data, security – but those protections should be implemented in an innovative-friendly manner. By learning from global examples and carefully weighing the risks and rewards, India and other nations can craft a legal environment where AI innovation and legal protections co-exist in a constructive balance, ultimately benefiting consumers, creators, and the economy.

---

[i] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[ii] https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf

[iii] https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf

[iv] https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

[v] https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf

[vi] https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf

[vii] https://regmedia.co.uk/2024/03/04/meity_ai_advisory_1_march.pdf

[viii] https://www.azbpartners.com/bank/meity-liberalizes-ai-advisory-dated-march-1-2024-following-industry-concerns-and-issues-revised-advisory-on-march-15-2024/

[ix] https://indiaai.gov.in/article/report-on-ai-governance-guidelines-development

[x] For example, Section 14 of the Copyright Act, 1957 grants authors several exclusive rights, including: (i) the right to reproduce the work in any material form (including electronic storage); (ii) the right to issue copies to the public; (iii) the right to perform the work in public or communicate it to the public; (iv) the right to make cinematograph films or sound recordings in respect of the work; (v) the right to make translations or adaptations; and (vi) the right to do any of the aforementioned acts concerning translations or adaptations of the work

[xi] https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf

[xii] nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf

[xiii] https://www.judiciary.uk/wp-content/uploads/2025/01/Getty-Images-and-others-v-Stability-AI-14.01.25.pdf

[xiv] https://arxiv.org/abs/2305.00118

[xv] https://indiankanoon.org/doc/1062099/

[xvi] https://indiankanoon.org/doc/1734007/

[xvii] https://indiankanoon.org/doc/1685540/

[xviii] https://www.casemine.com/judgement/in/5a65cbaa4a93263320777425

[xix] https://indiankanoon.org/doc/114459608/