COPYRIGHT IN THE AGE OF AI: A CASE COMMENT ON ANI VS. OPENAI, CS(COMM) 1028/2024

by Rajlatha Kotni (Partner), Swagita Pandey (Associate) and Anushka Tripathi (Intern)

INTRODUCTION

The rapid growth of artificial intelligence (AI), especially large language models like OpenAI's ChatGPT, has raised profound legal and ethical questions around data usage, intellectual property, and accountability. In November 2024, leading Indian news agency Asian News International (ANI) filed a lawsuit against OpenAI in the Delhi High Court, alleging that OpenAI's ChatGPT used ANI's copyrighted news content without permission to train its AI models. ANI also claims that ChatGPT generated fabricated news stories and falsely attributed them to the agency, potentially harming its reputation. Given the landmark implications of the case, the Court appointed two amicus curiae, Mr. Adarsh Ramanujan, an expert in intellectual property law, and Dr. Arul George Scaria, an academician in copyright law to assist the Court in scrutinizing the case. It is noteworthy that this case is part of a broader global debate over the use of copyrighted material in training AI models, with similar lawsuits filed by media organizations in the U.S., Canada, and Europe.

ISSUES

The Court has framed four key legal issues-

1. Whether training an AI model on copyrighted content without a license constitutes copyright infringement under Indian law?
2. Whether generating user responses using copyrighted data constitutes infringement?
3. Whether this use falls under 'fair use' as per Section 52 of the Indian Copyright Act and will the use of publicly available data, including news content, is fair use especially when the content is used in a transformative way (i.e., training an AI).
4. Whether Indian courts have jurisdiction over this matter as OpenAI's servers are based abroad but given the fact that the ChatGPT is accessible in India and allegedly impacts ANI's business and reputation in India?

CONTENTIONS MADE BY ANI

- ANI contends that OpenAI has unlawfully infringed its copyright by storing, utilizing, and reproducing copies of ANI's copyrighted material for the purpose of training its AI model. ANI emphasizes that the public availability of its content does not eliminate the requirement for obtaining permission prior to its use, and OpenAI's actions constitute an unauthorized appropriation of its intellectual property.

- ANI asserts that the ChatGPT model generates responses that are either verbatim or bear substantial similarity to its original content. This practice, ANI argues, amounts to the unauthorised reproduction of its works, thereby infringing upon its exclusive copyright and undermining the legal protection afforded to its creative and journalistic output.

- ANI further alleges that ChatGPT has disseminated outputs that falsely attribute interviews or reports to ANI. ANI maintains that such misleading attributions not only misrepresent its factual reporting but also damage its reputation and the reliability of its news service, thereby causing additional harm beyond mere copyright infringement.

www.lexport.in

NEW DELHI: K-1/114, First Floor, Chittaranjan (C.R.) Park, New Delhi - 110019, India | +91-11-2627 0506 | delhi@lexport.in
BANGALORE: 516, 10th A Cross, 29th Main, Sector 1, HSR Layout, Bangalore - 580034, India | +91-80-2268 0308 | bangalore@lexport.in

CONTENTIONS MADE BY OPENAI

- OpenAI contends that its practices are legally permissible because the content in question is publicly accessible.

- OpenAI points to the fact that similar legal claims in jurisdictions such as the United States, Canada, and Germany have not resulted in injunctions or finding of copyright infringement against it.

- OpenAI maintains that it operates with full transparency, disclosing its data usage practices on its website. It asserts that its AI model does not simply copy ANI's content; rather, it generates responses by analyzing and learning from a wide range of publicly available data. The company further denies that its system reproduces ANI's content verbatim, emphasizing that the training process does not store or recall identical copies during user interactions.

- OpenAI refutes the claim that it has falsely attributed interviews or reports to ANI. It argues that no formal grievances have been raised by ANI regarding inaccuracies in attribution, and that any mistakes that occur are rare, promptly corrected, and do not amount to a systematic pattern of misrepresentation.

- OpenAI also asserts that ANI had previously opted to restrict its content by blocking its web crawlers via a blocklist mechanism, and that OpenAI has honored this configuration.

- A significant pillar of OpenAI's defence is its challenge to the court's jurisdiction. The company argues that it does not maintain a physical presence in India, with its servers and training operations located overseas. Consequently, OpenAI contends that the Indian courts lack the authority to adjudicate the matter, especially given that the training activities occur outside Indian territory.

ROLE OF AMICUS CURIAE

The Delhi High Court appointed two amicus curiae Mr. Adarsh Ramanujan and Dr. Arul George Scaria to provide independent expert analysis on the copyright and technology issues involved. Their role is to assist the court in navigating the complex intersection of AI, copyright law, and digital rights. Their involvement is significant especially since India's copyright and intellectual property laws when originally drafted, the emergence of artificial intelligence was not foreseen. As a result, these laws do not contain specific provisions addressing the legal complexities introduced by AI technologies and Indian jurisprudence on AI is still in its infancy.

LEXPORT'S ANALYSIS

Jurisdiction and Applicability of the Indian Copyright Act:

- Section 62 of the Indian Copyright Act, 1957 permits a copyright owner to initiate legal proceedings in any forum where the owner resides or carries on business. Since ANI is Delhi-based, this provision strengthens its standing even if the contested data processing occurs abroad. However, OpenAI contended that Section 20 of the Code of Civil Procedure requires lawsuits to be filed where the defendant resides or where the cause of action arises. Since OpenAI has no physical presence in India as its servers and the primary storage used for training are located outside India, the Indian courts lack jurisdiction. The court deferred final jurisdictional rulings but leaned toward accepting jurisdiction under Section 62, given ANI's Delhi-based operations and ChatGPT's accessibility in India. Amicus

www.lexport.in

Curiae Dr. Arul George Scaria and Adarsh Ramanujan supported this view, noting that digital services operating in India fall under its legal purview.

- Another significant provision in this context is Section 75 of the Information Technology Act, 2000 (IT Act), which stipulates that the provisions of the IT Act shall be applicable to offences or contraventions committed outside India by any person, regardless of nationality. The section clarifies that the IT Act applies to any offence or violation involving a computer, computer system, or computer network situated in India.

Liability of AI-Generated Outputs:

- At the heart of the dispute is whether the output generated by ChatGPT which may, in certain instances, closely resemble ANI's original content constitute a copyright infringement. While ANI contends that any reproduction of its content, even if transformed into AI output, infringes its exclusive rights, OpenAI takes a contrasting view.

- OpenAI emphasizes that, after tokenization and subsequent processing (where texts are converted into numerical tokens), the model does not "store" or retrieve the original articles in a recognizable or retrievable form. As such, even if outputs occasionally reflect similarities, these are an emergent property of the model's learning process rather than instances of direct copying.

- OpenAI defended its practices by highlighting its "opt-out" mechanism, which allows content owners to block its web crawlers (e.g., GPTBot) via robots.txt directives or domain blocklisting, as seen in its response to ANI Media's copyright claims. However, ANI's legal team countered that blocklisting its domain (aninews.in) in October 2024 did not fully resolve the issue, as its news content could still be indirectly accessed through third-party platforms or mirror sites that replicate its articles, perpetuating unauthorized use.

- *The New York Times* initiated a high-profile lawsuit in December 2023 against OpenAI and Microsoft, seeking billions in damages for allegedly training their AI models on the newspaper's copyrighted articles without permission. The suit also raised concerns about AI-generated "hallucinations" falsely attributing misinformation to the *Times*, which the plaintiffs argued harmed its reputation and bypassed its paywall.

- The Delhi High Court held in *Akuate Internet Services Pvt. Ltd. v. Star India Pvt. Ltd.*, *2013 SCC OnLine Del 3344*, that while creative works enjoy copyright protection, the underlying facts and information constituting public knowledge cannot be exclusively controlled. This principle highlights the delicate balance between protecting intellectual property and preserving public rights, a balance that is now at the heart of the ANI versus OpenAI controversy.

Data Storage and Copyright Protection:

- A central technical aspect in the case is the method by which AI models like ChatGPT are trained. OpenAI scrapes enormous amounts of text, then tokenizes that content (i.e., converts it into a numerical format suitable for machine learning) rather than storing long-form original articles.

- Courts have grappled with whether temporary or non-expressive reproductions—those that occur during the transformation process are to be considered "copies" under copyright law. ANI argues that

www.lexport.in

NEW DELHI: K-1/114, First Floor, Chittaranjan (C.R.) Park, New Delhi - 110019, India | +91-11-2627 0506 | delhi@lexport.in

BANGALORE: 516, 10th A Cross, 29th Main, Sector 1, HSR Layout, Bangalore - 580001, India | +91-80-2268 0308 | bangalore@lexport.in

even temporary storage infringes its exclusive rights of reproduction, whereas OpenAI maintains that these processes are akin to human learning rather than publication.

- Dr. Scaria argued that tokenization does not retain original content and aligns with permissible "learning-based use". Ramanujan countered that even temporary storage for training infringes copyright unless exempted under fair use.

Transformative Use Under Copyright Law in the Context of AI:

- The concept of transformative use plays a pivotal role in modern copyright discourse, especially in the evolving landscape of artificial intelligence (AI). Under copyright law, a use is considered transformative when the new work adds significant new expression, meaning, or message, effectively transforming the original content into something with a different purpose or character. This principle has gained prominence in legal debates surrounding AI-generated content, such as in the emerging contention between ANI and OpenAI.

- The Delhi High Court recently dealt with the contours of transformative use in *Digital Collectibles Pte Ltd and Ors. v. Galactus Funware Technology Pvt Ltd and Anr*. CS(COMM) 108/2023. The Court observed that mere replication or commercialization of copyrighted material, even in a digital medium, does not automatically qualify as transformative. The judgment highlighted that for a use to be truly transformative, it must alter the original work with new expression, purpose, or character that fundamentally recontextualizes it.

ANTICIPATED RATIO DECIDENDI

While a final decision has not yet been rendered, the ratio decidendi in this case may revolve around several key legal findings:

1. Whether training an AI model on copyrighted news content constitutes reproduction under Indian copyright law.
2. Whether fair dealing can be extended to AI training in the absence of explicit legislative guidance.
3. Whether reproduction by generative AI is merely derivative, or if it is transformative—adding new value or functionality that reinterprets the original journalistic work.
4. Whether Indian courts have jurisdiction when the alleged infringement affects an Indian entity and its operations.
5. The liability framework for AI-generated content that causes reputational harm.

******

www.lexport.in

NEW DELHI: K-1/114, First Floor, Chittaranjan (C.R.) Park, New Delhi - 110019, India | +91-11-2627 0506 | delhi@lexport.in
BANGALORE: 516, 10th A Cross, 29th Main, Sector 1, HSR Layout, Bangalore - 580034, India | +91-80-2268 0308 | bangalore@lexport.in