

# Copyright's tryst with generative AI

**C**opyright law has always been a product of technology. It was created in 1710 to deal with the outcome of the invention of the printing press, to protect publishers against any unauthorised publication while encouraging learning, and to further their economic interests.

Since inception, copyright law has adapted itself to various technologies from the time of the printing press to the photocopying machine, to the recording device, and to the Internet. In each stage, the law has worked its way around technology. However, today there is a belief that generative AI has the potential to upset the copyright law. Such a debate is not new: it surfaces roughly every 20 years with each technological advent. So far, copyright law has been successful in forbidding commercial reproduction of works protected by copyright; currently, the law faces the task of prohibiting AI platforms from training on the works of the creators. There is a shift in the approach of using copyright law. In the past, the law dealt with copies of the original works; now, it has to deal with training of copyrighted material by AI platforms and not with the reproduction of copies itself.

## At a crossroads

Generative AI companies, specifically Open AI, have found themselves at a crossroads with copyright law across countries. AI platforms employ a technology called Internet scraping by which Large Language Models (LLM) train the platform on all available knowledge. For training purposes, the platform accesses both copyrighted and non-copyrighted content. The copyright infringement cases are fought on subject matters such as literature, music, and photographs.

Recently, the Federation of Indian Publishers as well as the Asian News International initiated copyright infringement claims against Open AI before the Delhi



**Sundar Athreya**  
H.

Legal Academic associated with Kiit Law School



**N.S. Amogh Simha**

Advocate practising at the Madras High Court

In the past, the law dealt with copies of the original works; now, it has to deal with training of copyrighted material by AI platforms and not with the reproduction of copies itself

High Court for training the AI platform with the works of the publishers without their prior consent. Similar cases are pending before the American courts, where the respondents have taken the claim of 'fair learning' and 'fair use in education' as an exception provided by the U.S. Copyright Act. In these cases, Open AI has developed an opt-out mechanism which allows the publishers to opt-out from the data set training. But this strategy applies only to future and not past training.

In the ongoing case in India, Professor Dr. Arul George Scaria, amicus curiae, has suggested that the court should address the issue of whether unlearning the information from the content used during training is technically and practically feasible. Further, he has also underscored the need for keeping in mind the effect of the future of AI development in India; access to legitimate information including copyrighted materials; and a direction from the court to Open AI to address falsely attributable sources.

Among other things related to Open AI, it has been argued that the Indian courts lack competence to hear the case. Leaving that aside, the LLM platforms may find themselves in an uncharted territory in India, as the Indian Copyright Act adopts a different exception test and not the 'fair use' test established in the U.S. It adopts the enumerated approach, where the exact exceptions are already stated, the scope to manoeuvre is limited, and education exceptions are confined within classrooms and not beyond. In India, this could be effectively used by the right-holders in their favour. However, the law could potentially be used to prohibit access to books, much against the original purpose for which it was created.

The opt-out mechanism developed by Open AI may also have a huge impact on the future of generative AI, as the efficiency of the AI depends on the material that it is trained upon. If in future,

the technology is not trained on quality material, that could obfuscate the budding AI platforms, which will not have the benefit that Open AI has. The court should ensure a level playing field between generative AI with deep pockets and generative AI without deep pockets so as to strike the right balance.

## Solutions to the problem

The claims by parties have the potential to impact the core of creation, art, and copyright law, since any creation stands on the shoulders of its predecessors. Generative AI/human creativity functions on the basis of learning from existing creativity, which acts as a nourishment to churn further creativity. Copyright law should not be turned on its head to prohibit future creators from having access to this benefit.

Further, the arguments of the publishers in the case at hand has a potential of viewing human creation and machine creation differently in future and setting different consequences for both. It is pertinent to remember that a human being is not expected to create further without learning; at the same time, the law as it stands does not make any differentiation between human creation and machine creation.

The foundational norms of copyright law offers solutions to the existing problem. Copyright in a work does not apply to the idea/information; rather, it is applicable only to the expression of the information. As long as the AI platform only uses the existing information for learning purposes, and does not thieve on the expression of the idea, it does not amount to infringement as per the law. When AI robs the copyright protected content, the existing norms of copyright law has its net in place to catch the infringements. The founding doctrine should not be compromised for the best interests of creativity as it acts as a medium between generative AI and creativity.