

RSIS Commentary is a platform to provide timely and, where appropriate, policy-relevant commentary and analysis of topical and contemporary issues. The authors' views are their own and do not represent the official position of the S. Rajaratnam School of International Studies (RSIS), NTU. These commentaries may be reproduced with prior permission from RSIS and due credit to the author(s) and RSIS. Please email to Editor RSIS Commentary at RSISPublications@ntu.edu.sg.

The Copyright Dilemma Shaping the Future of Generative AI

By Jose Miguelito Enriquez

SYNOPSIS

Do Generative Artificial Intelligence models violate copyright law? Tensions between AI developers and copyright holders have risen as courts and government agencies race to answer this question. Whatever the answer may be, it will surely define the future of AI development.

COMMENTARY

The development of Generative Artificial Intelligence (Gen AI) models depends heavily on the training data developers have access to. As Gen AI models have become more common, developers have found themselves subjected to significant scrutiny over issues surrounding the use, training, and development of their AI systems.

One question that lies at the heart of AI development has recently become increasingly contentious and will surely shape the future trajectory of AI development: *Do Generative AI models violate copyright law?*

The Brewing Legal Battle

This copyright dilemma is two-fold. The first concerns the output of a Gen AI model, specifically, whether AI-generated creations can be copyrighted. Consistent with current copyright laws, there must be human involvement in the creation of the output for copyright protections to be awarded, though it is unclear how substantial this involvement should be. The second issue focuses on the input side, or whether the use of copyrighted material in training Gen AI models without prior authorisation constitutes infringement.

Thus far, most high-profile court cases have focused on the second, more complicated

issue. A lawsuit filed by the [New York Times](#) against AI developers OpenAI and Microsoft alleges that the use of the Times' news articles to train AI chatbots causes them to unfairly "compete with (the Times) as a source of reliable information".

A separate lawsuit against OpenAI by [several authors](#) made a broader claim that due to the unauthorised use of their books as AI training data, every response generated by ChatGPT constitutes infringement. This claim has since been dismissed by the court.

AI developers have been steadfast in their defence of the use of copyrighted material in training Gen AI models. The crux of their argument lies in the doctrine of "[fair use](#)", a concept in US copyright law that allows the use of protected material under limited circumstances, including when the material is used for educational purposes or in one's exercise of freedom of speech.

Representatives of some AI companies have also drawn a parallel between an AI being trained on books and artwork and [a human reading those same books and perceiving the same art](#). In response, copyright advocates have argued that the scale by which Gen AI models produce content compared to human creations makes such a parallelism a faulty analogy.

Domestic and International Responses

Government organs, such as legislatures and intellectual property offices, in some countries, are racing to clarify the obligations of AI developers under copyright and IP law. Some regulators, including in [China](#) and the [European Union](#), have already issued Gen AI-specific regulations that impose a responsibility on developers to conform with existing copyright and IP law.

Still, many governments are trying to navigate this complicated problem as they seek to balance copyright issues with the AI developers' ability to pursue innovation. Indeed, achieving the equilibrium between these two concerns has proven to be difficult. Despite this, some mechanisms have been proposed.

In the United Kingdom, where [AI-related legislation](#) is still under consideration, a committee in the [House of Lords](#) published a report on potentially regulating generative AI and large language models (LLMs), where they concluded that copyright and IP laws must "ensure creators are fully empowered to exercise their rights" and that developers need to be transparent about their training data to "help rightsholders make informed decisions over the use of their data".

The House of Lords report also raised the possibility of implementing either an *opt-in* or *opt-out* mechanism that would allow rightsholders to grant permission to AI developers to use their work as training data. It has likewise encouraged the use of government-held datasets under the public domain for AI training as an alternative to using copyrighted material.

In the United States, where many AI-related copyright lawsuits have been filed, Federal congressional actions have likewise been contemplated. [Several senators](#) have voiced support for a licensing arrangement between rightsholders and AI

companies for the use of copyrighted material for AI training. While developers and some policy experts have argued that licensing deals are impractical due to the scale of data required to train AI models, such arrangements [are not unheard of](#).

On the other hand, a bill recently filed in the [US House of Representatives](#) would require developers to submit a notice to the Register of Copyrights whenever copyrighted material is used to train a Gen AI model. Either of these proposals would be a major advancement in AI copyright enforcement in the US as several AI-related laws at the state level have focused on [data privacy and the use of personal data](#) in AI training.

To prevent protracted and costly legal battles, some countries have sought to open a dialogue between rightsholders and developers to seek an amicable resolution to the copyright question. Singapore made such a recommendation in its draft [Model AI Governance Framework](#) published earlier this year. However, due to the contentious nature of the issue, it will be difficult for interested sectors to arrive at a consensus. For example, an initiative by the UK Intellectual Property Office to craft a voluntary AI copyright code [failed to materialise](#) after talks between rightsholders and developers stalled.

International organisations have also taken notice of these copyright issues. In its [Guide for AI Governance and Ethics](#), the Association of Southeast Asian Nations (ASEAN) cited intellectual property infringement as a potential risk in adopting Gen AI. More recently, the [United Nations General Assembly](#) passed a resolution regarding the safe and secure development of AI which included a provision calling for “appropriate safeguards on intellectual property and copyright while promoting innovation”.

Finding the Right Solution

Resolving the copyright dilemma will not be easy. While governments deliberate over their preferred enforcement mechanisms, rightsholders will be left with no choice but to sue AI companies. However, frequently bringing in the judiciary to formulate a concrete interpretation of the law is not the most appropriate solution.

When major personalities and organisations file lawsuits against AI developers, their interests may not necessarily align with freelance creative individuals who have fewer resources and are unable to litigate against potential infringement of their work by developers. Moreover, the judgment rendered by the courts may be limited and not apply to all rightsholders.

In any case, policymakers should ensure that their preferred mechanism satisfies two conditions. On the one hand, arrangements should provide protections for rightsholders against indiscriminate use of their copyrighted work, especially for commercial purposes. At the minimum, rightsholders should have the agency to decide whether their work can be used as AI training data. On the other hand, developers should have enough space to properly train their models and pursue innovations. Indeed, some rightsholders might choose not to authorise the use of their work in training, but this does not mean that all paths to innovation will close. On the

contrary, this only ensures that AI training is conducted in a manner compliant with copyright law.

The longer Gen AI copyright concerns are left at an impasse, the higher the stakes will get for all interested parties. [Gen AI models will only get more sophisticated](#), which means that the demand for training data will rise. Rightsholders will inevitably see this higher demand as a higher risk for copyright infringement, and tensions between the two camps will only intensify. Hence, it is imperative for governments to swiftly seek dialogue with developers, rightsholders and other creative individuals to resolve the Gen AI copyright dilemma.

Jose Miguelito Enriquez is an Associate Research Fellow in the Centre for Multilateralism Studies at S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore. His research interests include digital economy governance in ASEAN, populist foreign policy, and Philippine politics and foreign policy.

S. Rajaratnam School of International Studies, NTU Singapore
Block S4, Level B3, 50 Nanyang Avenue, Singapore 639798