

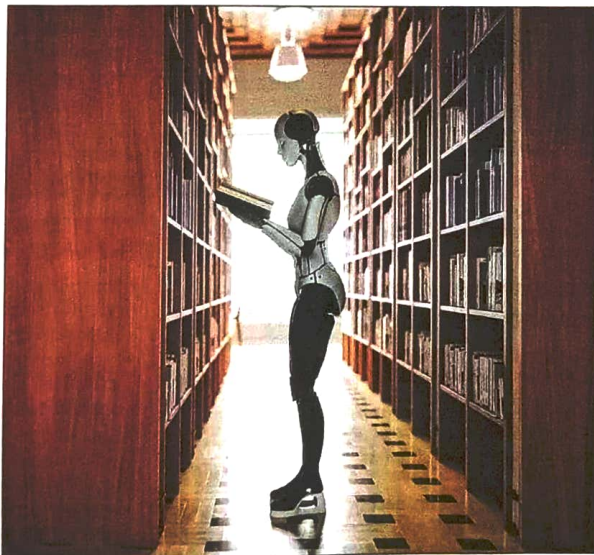
The future of information

The ANI Media Pvt. Ltd. versus OpenAI OpCo LLC case was much more than a battle between a news agency and the company that owns a generative Artificial Intelligence model. The judgment of the Delhi High Court in this case has thus raised an important issue: how should copyright law respond when machines *learn* from information rather than merely *reproduce* it?

Nearly a decade ago, the historian, Yuval Noah Harari, had predicted in his book, *Homo Deus: A Brief History of Tomorrow*, that the 21st century would be increasingly defined by data. 'Dataism', according to Harari, is the world view that interprets the universe as a flow of data and signifies a world in which information becomes the chief economic and social asset. In this world, people who can process information most efficiently have the greatest influence. This is the shift embodied in today's AI revolution. We are not just an information society anymore but a society in which information is continuously processed, restructured and created by computational systems.

Data mining — the analysis of large datasets to uncover patterns, relationships and predictions — is at the core of this transformation. Generative AI is different from traditional search engines that retrieve information; it is a system that learns from massive amounts of publicly available information

How should copyright law respond to Artificial Intelligence? Ashok Ramappa Patil and Sumit Kumar Singh examine a recent Delhi High Court judgment



and finds statistical relationships between words and ideas. Large Language Models are developed by collecting data, web scraping, tokenisation, embeddings, model training and, then, generating responses during inference, which is occasionally backed by Retrieval-Augmented Generation. These systems are more than just information storage and retrieval apparatuses. They exhibit computational knowledge that can create completely new

outputs from the information that is stored.

It is this technological framework that the Delhi High Court was forced to consider. The court had to grasp how LLMs work before determining whether OpenAI violated ANI's copyright. It is a significant change in copyright law. As models are used to create new works of art, questions of infringement cannot be answered without questions about whether the use of copyright-

ed works in the training of these models constitutes reproduction, whether the transformation of expressive works into mathematical representations constitutes copying, and whether an AI-generated answer is a reproduction or simply the result of statistical learning. The litigation thus transcends the question of 'who owns a work?' and examines questions like 'what is learning?' and 'when does learning become infringement?'

Paul Keller's 2025 White Paper, "Beyond AI and Copyright: funding a sustainable information ecosystem", situates the current discussion in the context of restructuring of the information ecosystem itself. His observations were premised on the underestimation of the effect of AI on the access to information. While copyright has been a topic of public discussion, the more pressing question is how people are increasingly interacting with knowledge via AI. When AI-generated summaries are used instead of reading newspapers, journals or archives, it is not just a question of copyright but a question of control of information in a democratic society. Keller said that societies require public AI infrastructure. Focusing on the foundational AI model, he proposed systems that are governed by the public and would therefore be transparent, accountable and equitable in providing access to knowledge. He claimed that the information ecosystem is not limited to individual

authors. The knowledge commons that fed into the immense value that AI systems gleaned has been created and maintained by researchers, universities, libraries, archives and public institutions. Copyright protection alone would have ignored the collaborative effort that made the creation of knowledge possible. He suggested redistributive mechanisms which would return part of that value to cultural institutions, to public service media and to the development of public AI systems to secure long-term sustainability of the knowledge ecosystem as a whole.

In this context, the ANI case takes on meaning and significance much greater than the outcome or the parties involved. It marks the onset of a legal debate on the future of information. The purpose of copyright law is to address reproduction by human actors and generative AI challenges that as it learns based on computational abstraction, rather than on copying. Courts will thus have to differentiate between information and expression, learning and reproduction, innovation and infringement, none of which copyright law has ever considered.

The Delhi High Court has started this conversation by affirming the importance of grasping the inner workings of AI in today's copyright debates. However, the bigger question remains — will technological advancements hinder progress in the development, verification and preservation of knowledge systems or institutions? As Harari foresaw, the 21st century is one of information and, as Keller reminds us, maintaining a democratic and sustainable information ecosystem is as crucial as protecting copyright itself.

Ashok Ramappa Patil is Vice-Chancellor and Sumit Kumar Singh a research assistant and student at the National University of Study and Research in Law, Ranchi