Data Is The New Oil And Electricity

As artificial intelligence becomes increasingly pervasive in our daily lives, policies and laws need to be examined for the protection of and access to the data driving it, experts said.

We live in a world where almost every aspect of our daily lives can be captured and analysed in real time, thanks to artificial intelligence (AI). AI has thus become an essential tool for shaping businesses and technological advancements. But the ever more sophisticated AI that will be needed by tomorrow's businesses is built on data, leading to concerns over how this data is generated and managed.

"The development of AI has evolved from being technology-driven to becoming data-driven. The protection of the access to relevant data will thus have a tremendous impact on the future of AI," said Professor Liu Kung-Chung at the Artificial Intelligence and Intellectual Property Conference, organised by the Singapore Management University (SMU) from 28 to 29 November 2019, in collaboration with the Max Planck Institute for Innovation and Competition, Germany; and the Chinese University of Hong Kong, Faculty of Law.

Speakers at the conference addressed the global need for a more comprehensive picture not just of the very nature of AI itself, but also its commercial applications; its interaction with different kinds of intellectual property (IP); IP administration; software and data; its social and economic impact on overall innovation policy; and, ultimately, even AI's alleged eligibility as a legal entity in its own right.

These are issues close to Professor Liu's heart: a Lee Kong Chian Professor of Law (Practice) at SMU and the founding Director of SMU's Applied Research Centre for Intellectual Assets and the Law in Asia (ARCIALA), he stands at the forefront of IP law, with a particular emphasis on Asia.

Be aware of the types of data involved

According to Professor Liu, whenever we talk about data in the context of AI, we have to be clear about which type of data we are referring to, so as to avoid miscommunication. In general, the types or typologies of data used by AI fall into three groups: custom-made training data, user- and machine-generated big data, and copyright-protected data. These three types are not precise concepts, and some flexibility and overlap is to be expected.

Early in its development, AI platforms were trained to perform specific tasks almost entirely on custom-made training data. This human-generated dataset naturally involves much creativity, requiring extensive intellectual and professional input—and, therefore, the need for IP protection. Thankfully, the purposefully constructive nature of the training data lends itself readily to copyright protection.

"As a whole, I think custom-made training data should qualify quite easily as works worthy of copyright protection," Professor Liu said.

On the other hand, the emergence of more sophisticated AI in recent years has enabled the AI itself to generate its own training data. This has led to the inevitable question of how IP laws should be applied, specifically with regard to copyright.

"It is a legal question we need to look at, but basically, we think [AI-generated data] doesn't seem to be able to qualify as protected work because it is just the result of the exercise of a software algorithm," Professor Liu explained.

Protecting big data from a regulatory perspective

Meanwhile, user- and machine-generated big data presents its own separate set of issues—less in terms of copyright, and more towards enabling socially beneficial access to information.

"We must protect big data not from a proprietary perspective, but from a regulatory perspective. How can we protect big data from being prevented from forming and accumulating, and allow it to be free-flowing?" Professor Liu said.

With regard to the access of copyright-protected data for the development of AI, Professor Liu suggested that necessary copying, digitisation and processing during text and data mining (TDM) should not constitute copyright infringement for two reasons. First, there is no 'use' for the works in the traditional sense, namely for personal enjoyment, appreciation, comprehension or inspiration (non-expressive use), he said.

Second, TDM does not affect the economic results of those works (non-consumptive use or non-market-encroaching use), Professor Liu said. He added that the Japanese approach is preferred over the EU approach in efforts to construct a specific fair use clause that is more conducive to AI, because it does not have an opt-out feature.

The leaps and bounds by which AI has progressed over recent years has ushered in an unprecedented era of data-driven products and services. From chatbots to healthcare facilities, AI is increasingly becoming more ubiquitous and essential to people's daily lives. "Data has become the new 'oil and electricity'," Professor Liu said. "It is a new source of economic and social value."

Asian Scientist Magazine is a media partner of the Singapore Management University Office of Research & Tech Transfer.