Kluwer Copyright Blog

Report on a roundtable on academic publishing and genAI deals – GenAI and copyright series at the Institute of Brand and Innovation Law

Alina Trapova (University College London) · Wednesday, December 18th, 2024

Being an academic is a vocation. We are not in it for the money (hopefully), but mostly (hopefully) for the impact that we can make on our students' and colleagues' lives, as well as to contribute to the process of healthy law and policy-making. It is a job with lots of responsibility, joys, surprises and disappointments, but one thing is for sure – publishing is a big part of our workload. Career development is a factor, but all of us are, in one way or another, driven by the idea of knowledge dissemination. To that end, we work very closely with academic publishers. In the complicated landscape of genAI and copyright law, several different themes have emerged as particularly thorny and triggering the interests of different stakeholders. One such topic is academic publishing and genAI deals between publishers and tech companies. On 19 November 2024, the Institute of Brand and Innovation Law (IBIL) at UCL Laws hosted a closed doors roundtable on academic publishing. The roundtable is part of IBIL's series on genAI and copyright.

The genAI and copyright series

Led by Prof Sir Robin Jacob, IBIL is one of only a small number of UK-based university research centres which focus solely upon IP law. The Institute was established in 2007 with a distinctive objective – it seeks not only to undertake academic research, but also to pay attention to the practical application of IP law in a rounded and inclusive manner.

The Institute is actively exploring the impact of genAI on copyright law via a dedicated series of events, roundtables, lectures and publications. The series organizes these into four streams:

- **Inputs**– exploring copyright implications of content used in AI training, focusing on issues including infringement, text and data-mining exceptions, lawful access and licensing;
- **Outputs**– examining copyright issues around AI-generated outputs, from authorship and ownership to originality, and their integration into existing copyright frameworks;

- **Policy**– addressing the legal challenges for AI and copyright, aiming to inform legislative reforms, policy-makers and to contribute to the development of best practices; and
- **Roundtables**

Roundtable on academic publishing

GenAI thrives on good data. In copyright law terms, very often that data could correspond to 'individual human expression'. Academic publications are a source of such reliable good data. In order to engage in text and data-mining (TDM) some tech companies are reported to have approached different publishers, hoping to secure access to their catalogues on reasonable terms. Others have publicly available TDM policies (see here, here and here). Different publishers have behaved differently – some have signed those deals without consulting authors, others have provided authors with an opt-in and a large group are currently said to be negotiating.

On 19 November 2024, IBIL hosted a closed doors roundtable on this topic bringing together different stakeholders from various backgrounds for an afternoon of discussions. To foster an open and inclusive dialogue, the meeting operated under the Chatham House Rule, meaning that participants are free to use the information received at the meeting, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed, outside of this meeting. One afternoon is certainly insufficient time to solve the issues around academic publishing and genAI, but the goal of the roundtables is to bring awareness by hopefully opening the avenue for a conversation from new angles and common grounds. This roundtable succeeded in this as it started an honest conversation where protecting authorial integrity and attribution appeared to be one of the central values when any such deals are negotiated.

Overall, there was a concern among the attendees that authors' works have already been used without permission in large quantities. There is a shared interest in ensuring that AI companies have legal routes to license content in the quantity required by this revolutionary technology. In addition, there was also a feeling that for the public good it would be better for models to be trained on authoritative scholarly materials rather than scraping whatever information is found on platforms such as X, Reddit or the open web.

The topics discussed at the meeting included:

- **The nature of academic publishing** – what makes it different to other publishing? It was mentioned that there is a strong public access ethos in academia which is perhaps less visible in other types of publishing. The emphasis on developing and disseminating knowledge is very prominent. That said, the price of access is much higher with academic publishing which from an economic perspective makes it less accessible.

- **'Old' contracts** – many of the contracts that were signed by authors and academics pre-date the LLM/genAI era. They vary in wording and sometimes include broad terms. Contractual interpretation dictates that the provisions must be given the meaning a reasonable person would have understood the intention of the parties to have been at the time the contract was concluded. This is an objective test, so actual intention of the parties is irrelevant. Therefore, where those 'old' contracts do not cover genAI and LLMs (as a matter of contractual interpretation), permission should be sought afresh (this is precisely the position of the Authors' Guild in the US).

- **Essential values to be respected** – several values are at stake here. This post does not seek to reflect a verbatim of the roundtable, but picks several themes to report on. One such theme is attribution. The output of genAI nowadays can often be treated as being in direct competition with the academic works it has been fed with; thus, ensuing respect for rules on attribution and, to that end, false attribution is crucial. An interesting discussion here revolved around the nuance about whether the outputs of AI models are a direct substitution for the inputs and the impracticality of citing source materials in large language model (LLM) training vs retrieval-augmented (RAG) models (where they certainly would be). Therefore, the EU AI Act's commitment to transparency (Articles 53(1)(c) and (d); see more here and here) is commendable. Nonetheless, it was discussed that some authors would not want attribution in these genAI outputs as they do not feel they have any control over what that output could be and would not want their name associated with any distortion/misleading statement. That said, calls were made to emphasise that 'transparency' needs to work in practice, ie transparency needs to be tied to trust. How does one ensure that AI model providers have told the truth when they declared the content on which they have trained their systems? A related problem that emerged in the discussion was the question of trade secrets. Copyright is only one IP right in the way.

- **Jurisdiction** – According to the EU AI Act and its Recital 106 "*providers of general-purpose AI models should put in place a policy to comply with Union law on copyright and related rights*". *It then goes on to highlight that "any provider placing a general-purpose AI model on the Union market should comply with this obligation, **regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place**.*" While this was more of a theme for the tech industry as the obligation is targeted at providers, interesting thoughts were expressed by the participants. It was felt that despite the AI Act not having effect in the UK, should there be a legislative initiative in the UK (which is a topic on which the UK government has just launched a public consultation on), one may wonder whether there is any reason to depart from such extraterritorial narrative (on the problems of extraterritoriality see here).

Future roundtables

Roundtables and stakeholder dialogues are difficult to organize (and chair). Solutions are tricky to champion in these settings, but what they mostly contribute to is to raise awareness of the many perspectives, emotions and rights involved. They would hopefully shape an open-minded and balanced approach in licensing, legislating, policy-making, publishing and researching. The next IBIL roundtable will take place on 10 February 2025 at UCL's Faculty of Laws and focuses on visual outputs. If you are a stakeholder and would be interested in participating in a future roundtable, then please register your interest by completing the form here.

_____

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe here.

This entry was posted on Wednesday, December 18th, 2024 at 7:32 am and is filed under Artificial Intelligence (AI), Britain' and 'exit' and refers to the UK leaving the European Union (EU). A referendum – a vote in which everyone (or nearly everyone) of voting age can take part – was held on 23 June 2016, to decide whether the UK should leave or remain in the EU. Leave won by 51.9% to 48.1%. Britain's departure from the EU is scheduled to take place at 11pm UK time on 29 March 2019.">Brexit, Text and Data Mining (TDM)

You can follow any responses to this entry through the Comments (RSS) feed. You can leave a response, or trackback from your own site.