

Copyright & generative AI: regulating data mining

Dr Hayleigh Boshier

Policy Context: The government is considering how it should regulate AI training models. Originally, it proposed an extension to the copyright exception for data mining. Now, it is developing a code of practice to increase the availability of licenses for data mining. The government aims to reward investment in creativity while enabling AI innovation.

Research: Copyright is not necessarily a barrier to AI innovation. A broad data mining exception would breach international law and could harm both creative industries and AI firms.

Advice: Confirm that under current law AI firms need licenses for data mining. Consider variations or alternatives to an any-purpose copyright exception. Use copyright as a tool to balance interests, understanding that AI firms and creative industries are not separate stakeholders.

Evidence

- > **There is a lack of evidence that AI firms are facing a barrier to innovation as a result of copyright.** Only 13 out of 88 responses to a [UK IPO consultation](#) were in favour of a broad any-purpose text and data mining (TDM) exception, and these largely came from researchers, libraries and archive institutions such as The British Library.
- > **Upholding the copyright system could contribute to AI firms' value and their incentives to innovate.** This is because AI firms may be rights holders themselves – although the application of copyright law to AI-generated work has not yet been clarified. In the UK IPO consultation, tech firms did not argue for an any-purpose copyright exception. They preferred an opt-in/-out system.
- > **An any-purpose copyright exception would likely be contrary to international agreements.** This is because the Berne Convention requires exceptions to be for a specific purpose and to not harm the copyright-holders' interests.
- > **It would also divide and polarise sectors, instead of encouraging collaborative innovation.** The proposed 'any purpose' exception would deprive the creative industries of the remuneration and control over their works that copyright entitles them to. At the same time, it would displace creative workers, as pointed out by representatives giving [evidence](#) that their workers have already been put out of work due to AI.

Policy advice

- > **Confirm that under current copyright law AI firms need licences for data mining.** The government has implied this by suggesting an exception could be made for this purpose, but this has not been affirmed. AI firms need clarity to ensure their businesses are financially sustainable. Some argue that licences aren't needed because AI models copy 'tokenised' aspects of works. But copyright is not intended to be technologically specific and the under the current rules AI firms would be required to obtain licences.
- > **If copyright is shown to be a barrier to innovation, consider a limited copyright exception.** Such as when there isn't a licensing structure available, as currently applies for education institutions. This could incentivise creative sectors to enable AI firms to more easily purchase licenses. A drawback is that it removes control from rightsholders and creators to opt-out of their works being used in AI training. Alternatively, an opt-in exception would respect the right to determine the use of one's works.
- > **Investigate the impact and unintended consequences of other options.** Opt-out exceptions can burden rightsholders to find out which AI models are training on their data, This is impossible without transparency requirements in place for AI firms. Whilst some AI firms claim that data can be removed, other computer scientists say that this doesn't remove what the AI has learned and is therefore meaningless in practice. Alternatively, sector-specific exceptions may be tailored, e.g., for archiving and preservation.
- > **Remember that copyright exists to balance and protect the interests of both the creators and the rightsholders, which do not always align.** Both creators (e.g., musicians, writers) and rightsholders (e.g., record labels and streaming services) may seek revenue from AI uses of their works. Rightsholders tend to have more resources than creators to advocate for themselves, copyright law has mechanisms for recognising this imbalance in bargaining power e.g., through equitable remuneration.
- > **Consider setting out a principle that regulation should support AI and other industries to work well together.** Copyright balances interests to encourage the creation and dissemination of culture and knowledge, which AI firms and creative industries may do best in collaboration. The distinction between AI firms and creative industries is increasingly blurred.

Work with me

Hayleigh is a Reader in Intellectual Property Law at Brunel University London and member of the Brunel Centre for AI. She is well-recognised in the field of intellectual property law, in particular for her research in copyright law and the creative industries. She is the author of *Copyright in the Music Industry* and producer of the *Whose Song is it Anyway?* podcast.

Contact her at hayleigh.bosher@brunel.ac.uk if you would like to work with her on this, or related issues such as whether copyright applies to AI-generated works, and if so, who is the rightsholder and what rights they should have.