Publishers Association

Publishers Association response to the IPO Artificial intelligence call for views: Copyright and related rights

Introduction

The Publishers Association is the member organisation for UK publishing, representing companies of all sizes and specialisms. Our members produce digital and print books, research journals and educational resources across genres and subjects.

We are delighted to respond to the IPO's Artificial Intelligence call for views. In our submission we focus on the 'copyright and related rights' chapter of the consultation. Our response argues that the existing copyright framework is sufficient to ensure the development of AI and in fact, is already doing so.

We argue that intellectual property laws are not an impediment to the development of emerging technologies, rather that they underpin the value of the information that is being used. Instead we assert that a significant change in copyright exceptions to allow unfettered free access to UK content for AI purposes would merely result in a wholesale transfer of value from the UK content sector to predominantly large technology firms and, more importantly, would not support greater investment or higher-quality outputs.

We would like to present the possibility of a collaborative approach between the content and technology sectors by encouraging commercial negotiation and licensing solutions for paid-for content which can bring about both access and secure investment. We believe that this approach can and will achieve the UK government's aim to make the UK a global centre for AI and data-driven innovation.

Publishers' Investment in AI

In October 2020, the Publishers Association released a report with Frontier Economics exploring AI investment in the UK publishing industry. The report, which is based on sector interviews, case studies and an industry-wide survey, found that overall, the majority of publishers, irrespective of size and sector, consider AI will have a transformative impact over the next five years.

The report included case studies of publishers' investment in AI in order to show the breadth of investment and innovation. A key example was from education publisher, McGraw Hill's ALEKS platform ("Assessment and Learning Knowledge Spaces"). This is a platform that uses AI to deliver tailored content to maths, science and business school students, and Higher Education students. AI is applied to understand precisely which part of the knowledge space a student is in. This information then determines which topic to teach the student next and in doing so delivers truly smart student-centric learning. McGraw Hill cites student benefits as increased engagement, improved study time effectiveness and higher quality learning (learning gradually rather 'cramming') resulting in deeper understanding and improved recall rates. McGraw Hill's other AI platform, SmartBook is another example of AI being used in learning resources to create the learning techniques of the future. These two AI products

The Publishers Association Limited is a company limited by guarantee incorporated in England and Wales. Registration number: 3282879. Registered Office: 50 Southwark Street, London SE1 1UN.

were the result of collaboration between the publisher, academia and EdTech. The current legal framework protects and supports this type of collaboration and joint venture.

Other examples in the report from academic publishing were the Elsevier and Taylor & Francis content recommendation platforms (Mendeley Suggest and a partnership with UNSILO's Recommend product), which support research outputs by using AI and ML to recommend relevant research to academic researchers. These examples show the investment already being made by UK content companies in bringing new products using these technologies to market.

The report also makes reference to the fact that AI content-generation is also being undertaken by publishers. The example used is the Springer Nature's Lithium-Ion batteries research publication, written entirely by AI and in collaboration with the University of Frankfurt and Digital Science (a science service partner).

The report found that 60% of large industry members viewed government policy and regulation as a potential barrier to AI investment and confirmed that issues of text and data mining of copyrighted materials, patenting of AI-generated content and copyright protection and legal responsibility for AI-generated works are all relevant to publishers' investment decisions. The report makes the case for legal certainty regarding UK IP law, so that the investment cases both for the development of AI products and the UK content that underpins quality outputs is preserved.

1. **Do you agree with the above description of how AI may use copyright works and databases, when infringement takes place and which exceptions apply? Are there other technical and legal aspects that need to be considered?**

We agree with the descriptions of how AI may use copyright works. We assert that the current provision of database rights and copyright exceptions, such as the TDM exception (s29a CDPA), are suitable and provide balanced access to content. Publishers also permit more access under the TDM exception that the description implies, allowing use of copyrighted AI outputs under the TDM non-commercial exception. Where an AI provider is not able to access content, because it is not publicly available or covered by the non-commercial TDM exception, then licensing models are available, and commercial negotiations for access are open and transparent.

2. **Is there a need for greater clarity about who is liable when an AI infringes copyright?**

There are certainly unanswered questions in relation to how infringement should be managed. If liability for infringement is considered in terms of input, for example, then we would argue that there should be clarification that the beneficiary of that activity should be liable i.e. 'the person by whom the arrangements necessary for the creation of the work are undertaken', whether a natural person or a legal entity. In other words, the responsibility should lie with the individual or entity who has control over the infringement, the ability to stop future infringement and can compensate the copyright owner.

There are further distinctions to be made in this area. For example, an AI system which is "fed" copyright works as part of its training is very different to a system which is trained to find copyright works to incorporate into its own training. Such systems already exist and have a degree of freedom to choose which works they incorporate into their training. This poses particularly difficult questions for infringement as the system's (human) operators may not be able to predict in advance if the system will copy or otherwise use a work unlawfully or whether those operators had requisite knowledge of any actual infringement. This problem is compounded in the case of neural network systems which contain decision-making algorithms potentially immune to human scrutiny. This "black box" obstacle presents challenges for many areas of regulation and ethics, especially a question about the difference between 'active' and 'passive' AI activity. Although we believe the human operator would be liable for the infringement, regardless of this requisite knowledge, this is an issue that will need to be addressed more substantially in other legal areas, such as data protection (where a data controller is required to provide transparency of how, where and why personal data is processed).

We would encourage the IPO to collaborate with other governmental bodies on developments in this area as we expect they will have direct influence on issues of liability and regulation of AI in IP. We very much support the IPO in its efforts to clarify these issues and we urge collaboration as well as additional guidance on the appropriate use of AI technology and the importance of licensing solutions from the copyright owner or where applicable, the collective management organisation. For example, the IPO could explore the possibility of a requirement that algorithms (using metadata) must prevent use of copyright works without a licence. However, the use of licensing must remain voluntary – not compulsory.

We also support the proposal from the British Copyright Council that suggests a statutory requirement for the retention of auditable records that show what input the AI has used. This record would give users more insight into how the AI tools have been created and therefore more confidence in the quality of the output.

3. Is there a need to clarify existing exceptions, to create new ones, or to promote licensing, in order to support the use of copyright works by AI systems? Please provide any evidence to justify this.

The current copyright framework, including existing exceptions and licensing models, sufficiently supports and encourages the use and development of AI. We believe it to be a "red herring" to say that intellectual property stands in the way of AI development and argue that the creation of new exceptions to stimulate it would be severely misplaced. Instead, we assert that commercial negotiation and licensing solutions are abundantly suitable to ensure that rightsholders can collaborate with technology companies to bring about quality AI outputs. We believe that it is through this approach that the UK government can realise its aim of making the UK a global centre for AI and data-driven innovation.

As Science, Research and Innovation Minister Amanda Solloway referenced in a recent speech, strong AI requires a strong IP framework. High-quality AI outputs are based on high-quality data and information, which are generated by a strong IP framework. UK publishers

need to incentivise investment in order to produce said high-quality data and broad exceptions would undermine the skill and investment involved in producing, managing and curating it. In turn this means UK publishers would not be able to sustain this important way of working, consequently weakening the quality of their outputs. Broad commercial or "catch all" exceptions for AI would defeat the purpose of producing high-quality outputs by undermining the underpinning content. The rate of AI production would be high, but the resulting outputs would not be of a sufficiently high standard. This is especially concerning when dealing in very high stakes areas of research, such as clinical decision-making for example.

We assert that the existing exceptions framework is fit for purpose. The UK's TDM exception allows researchers to make copies of copyright material for the purpose of computational analysis if they already have lawful access to the work, and provided the making of copies is for research of a non-commercial purpose.

Where the exception does not cover access, then the other route is licensing. And the promotion of licensing is hugely welcome. Licensing models are evolving and collaborations between publishers and technology firms are developing all the time. We are not aware that there is any evidence to suggest that these licensing agreements are not suitable for AI or that a lack of a new exception or broadening the current exceptions is a barrier to investment.

We are aware that the question specifically requests evidence to support positions and we include an example below of how licensing can be used effectively by data acquisition companies to collect copyrighted material. Elsevier's *Scopus* is a world-leading citation database of peer-reviewed literature and hosts over 1.7 billion cited references from nearly 12,000 publishers. This data was obtained through commercial negotiations with publishers and obtaining licenses for the data. This model is also used by similar platforms like *Web of Science* owned by Clarivate. This highlights that commercial negotiations regarding access to products are possible on a grand scale and that IP protected data can be licensed effectively.

Further to this point, the claim that AI needs data to develop is an argument that has long been used when new technology emerges. When big data analytics and data science emerged, the same arguments for support to access copyrighted material emerged – AI does not require anything different to these technologies. The fundamental value of data and content has not changed. In fact, the market for data acquisition has only grown in importance and value.

We believe that a collaborative approach between businesses is the best way to achieve high-quality AI outputs and secure investment. These collaborative approaches based on commercial licensing include: licensing agreements with large consultancies developing information analytics tools or conducting specific research projects; major technology companies developing AI service layers or industry-specific tools; start-ups building AI data analytics tools for the pharma industry; and companies using AI to conduct primary research in pharma/life sciences. These licensing agreements have served to enable the development of AI without infringing copyright or the need to use exceptions.

In summary, we believe that the existing copyright exceptions are fair and balanced and, more importantly, that licensing and commercial collaboration between sectors is the best route to successful AI development. This is not to say that we do not believe that progress can be made, but that this progress should be in the incentivisation and championing of licensing, including in some cases licensing by collective management organisations, rather than change in IP legislation. We maintain that a significant change to copyright law to allow unfettered free access to UK content for AI purposes would merely result in a wholesale transfer of value from the UK content sector to predominantly large technology firms and would not support greater investment or higher-quality outputs.

4. Is there a need to provide additional protection for copyright or database owners whose works are used by AI systems? Please provide any evidence to justify this.

We believe that increasing support for licensing models for AI systems would incentivise the use of lawful content and facilitate access in fast and practical manner. The key is ensuring that these processes both work and are enforced effectively to ensure that AI systems can function.

5. Should content generated by AI be eligible for protection by copyright or related rights?

One view we heard from publishers was responsibility should be jointly attributed to the individuals and organisations responsible for providing the underlying data and algorithms, with an example earlier in 2020 of a Chinese court ruling that an AI-generated work was the work of a legal entity. The question of legal responsibility is particularly important for publishers who use AI algorithms to summarise content or make editorial recommendations where there is scope for AI algorithms to introduce bias or inaccurate representation.