UK Intellectual Property Office
Copyright and AI: Consultation 2025

Text and Data Mining Exceptions

Artificial Intelligence (AI) services are important socially and economically. The quality of the services depends in part on the quality of the data used to train and evaluate the services. The United Kingdom (UK), like other governments, is considering a range of options for text and data mining (TDM). Typically the policy choices are to treat commercial and noncommercial services the same or different, and to make the TDM exception apply to all copyrighted works or to offer an opt-out option.

Knowledge Ecology International (KEI) has encouraged governments to extend TDM exceptions to both commercial and noncommercial works. With regard to any opt-out options, KEI has asked governments to limit such opt-out options to a narrow set of cases where they are appropriate.

The most vocal opponents of TDM exceptions to train AI are those from the cultural industries, and news reporting. A one-size-fits-all policy on the opt-out or opt-in nature of using copyrighted works or otherwise protected data to train AI ignores the significant differences that exist between different types of data and copyrighted works.

Policy makers are encouraged to distinguish between works from cultural industries, such as recorded music, visual arts or screenwriting, from other works and data.

Medical Products and Services

KEI has often highlighted the importance of inclusion of training data when AI services are used for the development and use of medical treatments, for example, and this is just one of the areas where inclusiveness is critical and beneficial for society.

In written testimony to the United States Trade Representative regarding its 2025 Special 301 List (docket number USTR-2024-0023), KEI offered these comments:

> The most emotive and politically important voices to make it an infringement of copyright to train AI services are cultural industries and journalism. KEI has been concerned that restrictive policies on the uses of copyrighted works to train AI will be extended to much broader classes of works. We have singled out science, drug development and legal issues, to illustrate cases where society is best off if AI services have access to everything possible in terms of data, and where omissions may have considerable downsides. (2023. James Love. We Need Smart Intellectual Property Laws for Artificial Intelligence: "One-size-fits-all" regulation will sideline medical and research benefits promised by the advent of artificial intelligence, Scientific American, August 7, 2023; https://www.scientificamerican.com/article/we-need-smart-intellectual-property-laws-for-artificial-intelligence/)

> In regard to scientific and medical information, note that authors are rarely paid for their works, and a handful of companies control a large number of journals, many of them foreign-owned. It would be a very bad outcome if the Holtzbrinck Publishing Group, the private equity firm BC Partners, Wiley and the Relx Group, are able to significantly limit which companies can use the leading medical journals to train AI services, or for that matter, if any publisher can opt-out of the science being used to train programs that are used for drug discovery or to treat and protect patients.

Below are a few additional examples to illustrate areas where AI-related policies should differ from cultural industries.

Legal Services

For legal services, AI tools should be able to extend deeper and to more widely accessible analysis of the obligations and rights one has under the law. This could potentially benefit underserved populations, and provide for more accurate legal assistance, at lower costs. To be reliable, this has a completeness requirement. The AI needs to have access to comprehensive data. One does not want to see blind spots in reasoning and missing precedents and relevant rules.

Weather Forecasting

For weather forecasting, the more comprehensive the data, the more accurate the forecasts. Complex pattern recognition benefits from diverse data sources, and fragmented data access can create less accurate or even dangerous prediction gaps.

Engineering and Manufacturing

For engineering and manufacturing, comprehensive data will improve the efficacy and safety of the decisions. Given the cumulative nature of acquired knowledge, it may be challenging to distinguish between truly original and protected works from older public domain knowledge. One challenge for the training of AI services is to understand highly technical terminology and context, and to avoid misinterpretation of data. The risks of misinterpretations are magnified when using data that is incomplete.

Financial Services

In an ideal world, AI in the financial services sector can provide better market stability, risk assessments, fraud detection and fairer lending practices, with more comprehensive data.

Education and Academic Research

The role of AI in education is accelerating, and faces a number of challenges that should not be underestimated by anyone. As far as TDM exceptions, they are important for enabling more comprehensive literature reviews, and creating opportunities for cross-disciplinary insights to emerge, benefiting researchers with narrower formal training.

Patent Examinations

AI services used to assist in patent examinations can require large amounts of data to train effectively, including patent documents of course, but also scientific literature, and other relevant publications. Prior art searching is one obvious area, but also in evaluating other standards, such as non-obviousness. As in other areas, if certain works or data related to an invention and its industrial application are not available for training the AI services, the outcomes are less useful, and create more legal uncertainty.

Trade Policy

As noted in KEI's recent comments to USTR, "Any country that provides robust exceptions for using copyrighted material for AI will have a significant advantage in terms of training such services. But a lack of harmonization may create a situation where services developed in one country . . . will not be legal in another, because of non-authorized use of copyrighted works to train the service. This makes it a significant trade issue."

The UK needs to engage with norm setting fora around the world. One important forum is the World Intellectual Property Organization (WIPO). WIPO's work on AI has often focused on the concerns expressed by musicians, screen writers, actors, and visual artists, and less so in areas such a biomedical service or other areas where there are compelling reasons to have robust exceptions. The UK should encourage WIPO to expand the opportunities to showcase the importance of TDM exceptions in a wide range of areas where comprehensive or inclusiveness is important.

Competition Concerns

Economies of scale may be significant in general, and for specific types of services, excessive concentration of control of data will be concerning. Creation of opt-in or opt-out rules for text and data mining will create entry barriers, some extremely high, for some types of services. We have focused frequently on biomedical research, where an opt-in rule creates the risk that a handful of highly concentrated biomedical journal publishers that hold the copyrights on published articles could engage in restrictive licensing to a handful of large pharma companies.

Beyond the risks of concentration raised by creating and opt-in or opt-out options, are those where economies of scale and the costs of assembling data lead to unwanted concentration. One can look at the experiences with Internet Search or Mapping services to appreciate how highly concentrated services can become. It may be important to treat some types of data or databases as essential facilities, and mandate some types of sharing of access.

Data Spaces

Data spaces is a term used to describe a framework for secure and trusted data sharing. The objective is to create systems where organizations can exchange or provide controlled access to data, under agreed-upon rules, standards and infrastructure.

In decentralized systems, data remains held by various entities, but can be accessed by authorized parties for specific uses. For example, medical records data around the world is held by different government and private entities, and is subject to rules to protect the privacy of patients. Queries can be run on decentralized datasets, with the results benefiting researchers, while avoiding excessive concentration of the control over records, and allowing each organization holding the records to protect patient privacy according to relevant ethical and national legal obligations.

Metadata and Formalities

For a variety of reasons, it will be useful to have better metadata attached on published articles, photographics, designs and data of all sorts.

Some WIPO administered treaties on copyright and related rights place practical restrictions on requirements for formalities, which make it more difficult to identify works used to train AI or to provide remuneration, if that is a policy objective.

Metadata adds crucial context, explaining the meaning, origin, and characteristics of the data, and allows machine learning models to interpret the information more accurately.

Standards for metadata are important, challenging to develop and typically are a cross border concern. KEI has been asking the WIPO Standing Committee on Copyright and Related Rights to have as an agenda item discussions on metadata, beginning with recorded music and photographs.

Sui generis IP regimes

When software was becoming more important, there was an extensive debate over the appropriate type of intellectual property protection, including proposals for *sui generis* regimes of protection. Court decisions in the United States and elsewhere would eventually determine that software would benefit from copyright, trademark, patent, trade secret and contract protections.

While governments sort out the appropriate types of protection for copyrighted works and data uses as inputs as well as the protection of the generative AI outputs, it is useful to review the earlier discussions over the protection of software, and this could be the basis for a future request for comments.

James Love
Knowledge Ecology International (KEI)
110 Maryland Avenue NE, Suite 308
Washington, DC 20002
USA
https://keionline.org
james.love@keionline.org

Annex: Selected KEI Comments and Publications on AI Topics

2025, February 3. KEI Comments to US Trade Representative on 2025 Special 301 Review, docket USTR-2024-0023). https://www.keionline.org/40497

2024. November 12. KEI Briefing Note 2024:8. The Basic Proposal for the Design Law Treaty (DLT) and its inappropriate restrictions on transparency. (AI mentioned). https://www.keionline.org/wp-content/uploads/kei-bn-2024-8.pdf

2024, September 10. KEI Comments to USPTO on Experimental Use Patent Exception, Patent and Trademark Office, Docket No.: PTO-C-2024-0023, Document Citation: 89 FR 53963. (AI mentioned) https://www.keionline.org/40192

2024, March 14. KEI Comments to USPTO Request for Comments: Unlocking the Full Potential of Intellectual Property by Translating More Innovation to the Marketplace, Notice by the Patent and Trademark Office on 03/15/2024, Document Citation: 89 FR 18907, Agency/Docket Number: Docket No.: PTO-C-2024-0004. (AI mentioned) https://www.keionline.org/39965

2023, August 7. We Need Smart Intellectual Property Laws for Artificial Intelligence: "One-size-fits-all" regulation will sideline medical and research benefits promised by the advent of artificial intelligence. Article by James Love in Scientific American. https://www.scientificamerican.com/article/we-need-smart-intellectual-property-laws-for-artificial-intelligence/

2023, July 7. KEI Comments in response to the Office of Science and Technology Policy (OSTP) Request for Information (RFI), regarding National Priorities for Artificial Intelligence, as noticed on May 23, 2023 in the Federal Register, 88 FR 34194. https://www.keionline.org/38868

2023, July 7. KEI comments at WIPO conversation on intellectual property (IP) and artificial intelligence (AI). https://www.keionline.org/33491

2023, June 14. KEI Blog by Koenraad Rang. "New EU AI legislation." https://www.keionline.org/38819

2023, June 12. KEI response to NTIA request for comments on AI Accountability Policy, expressing concerns over trade agreement provisions which limit the transparency of source code and algorithms. https://www.keionline.org/38810

2023, January 31. KEI Comments to USTR 2023 Special 301 Review, Docket USTR-2022-0016. https://www.keionline.org/38283

2020, February 14. Submission of Knowledge Ecology International comments on WIPO's draft issues paper on Intellectual Property and Artificial Intelligence. https://www.keionline.org/32253

2020, January 10. KEI Comments on Intellectual Property Protection for Artificial Intelligence Innovation, for USPTO Request for Comments, Re: 84 FR 66176, Docket No. PTO-C-2019-0038. https://www.keionline.org/32253