

EIFL Webinar

Libraries, Copyright and AI for science and research

AI in copyright laws and litigation: an update

18 September 2024

Jonathan Band
policybandwidth



Attribution 4.0 International

Does Data Ingestion Infringe?

- Training AI requires ingestion of millions of works.
- Software analyzes these works to discover patterns, relationships, and trends, which are reflected in the AI “model.”
- Most of the countries that have examined this issue have determined that ingestion does **not** infringe.

Does Data Ingestion Infringe?

EU Copyright in the Digital Single Market Directive (applies in 27 EU member states)

- Art. 3: permits text and data mining for non-commercial scientific research of lawfully accessed works.
- Is TDM the same thing as training AI? TDM defined as “any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.”
- Art. 4: permits text and data mining for other purposes, subject to opt-out.
- Most companies that scrape works from the Internet respect robot exclusion protocols.
- Article 3 (but not Art 4) cannot be overridden by contract.

Does Data Ingestion Infringe?

- **UK** TDM exception permits “computational analysis” for “the sole purpose of research for a non-commercial purpose.”
- **Ukraine** has an even narrower TDM exception, permitting the making of copies “from a legitimate source for the purpose of searching for text and data included in or related to scientific publications for research purposes.”
 - Appears to permit only the assembly of a corpus for the purpose of searching for information, not the sort of computational analysis necessary for training AI. Further, the corpus could contain only scientific publications.
 - Like Article 4 EU DSM Directive, “this provision shall apply if the use of works has not been expressly prohibited by copyright holders in an appropriate manner, in particular, by computer-readable means from digital content available on the Internet”.

Does Data Ingestion Infringe?

- **Singapore** permits the copying of works for the purpose of “computational data analysis.”
- Copyright Act defines computational data analysis of a work as:
“(a) using a computer program to identify, extract and analyse information or data from the work or recording; and (b) using the work or recording as an example of a type of information or data to improve the functioning of a computer program in relation to that type of information or data.”
- “an example of computational analysis under paragraph (b) is the use of images to train a computer program to recognize images.”

Does Data Ingestion Infringe?

- **Japan** also has an exception for data ingestion.
- It permits uses of works “in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data...).”
- Like Article 4 EU DSM Directive and Singapore’s TDM exception, this provision contains no limitations such as a scientific research purpose.

Does Data Ingestion Infringe?

- **US:** Several courts have found that the assembly of large databases for AI functions, such as search and plagiarism detection, are non-infringing under a **fair use** theory.
- But these AI functions were more basic (and less likely to have competitive impact) than generative AI
- Over 30 pending cases, including several class actions.
- **Israel** Ministry of Justice opinion letter finds that copying involved with machine learning is a fair use.

Thank you!

Contact

jband@policybandwidth.com