

ARTIFICIAL INTELLIGENCE TRAINING AND COPYRIGHT INFRINGEMENT IN SINGAPORE

HOW DOES SINGAPORE LAW TREAT THE USE OF COPYRIGHT WORKS FOR AI TRAINING?

AI development uses extensive datasets, which often include copyright works, to train AI systems. Such use may infringe copyright unless the copyright owner has granted a licence or a legal exception applies. This explainer looks at how copyright exceptions and licensing complement each other in Singapore's copyright framework to support the use of copyright works for AI innovation while respecting the interests of creators and copyright owners.

WHERE COPYRIGHT INFRINGEMENT ISSUES ARISE IN THE AI LIFECYCLE

AI TRAINING

Input phase of generative AI



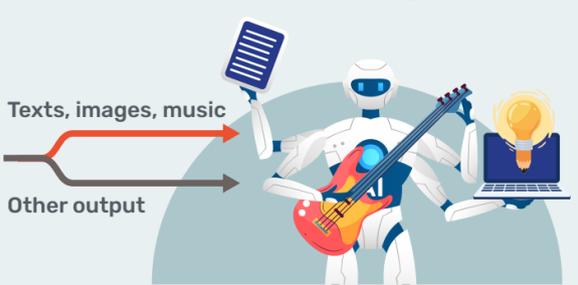
Copyright works may be copied when they are downloaded, stored, or ingested by an AI system for training.

Copyright owners have the exclusive right to control copying of their works. Copying for AI training infringes copyright unless:

- A copyright exception applies, e.g., computational data analysis or fair use
- You obtain a licence to use the works

AI DEPLOYMENT

Output Phase of generative AI



For AI systems that generate content, copyright works used during training may appear in the generated output in response to a user's prompt.

This may raise separate copyright issues at the output phase, even if the training itself does not infringe copyright.

See our previous explainer, [Artificial Intelligence Output And Copyright Infringement In Singapore](#).



COMPUTATIONAL DATA ANALYSIS (CDA) EXCEPTION

(Section 244 of the Copyright Act 2021)

This exception allows copyright works to be used for CDA, subject to safeguards protecting copyright owners. It applies only to the AI training phase.



Since 2021, the CDA exception has supported Singapore's National AI Strategy, push towards data-driven innovation, and efforts to grow our AI and technology sectors. It has provided legal clarity and certainty for research institutions, technology companies, and financial institutions that engage in AI training and data analysis.

HOW THE EXCEPTION ALLOWS CDA ACTIVITIES



Technology-neutral definition of CDA

Includes training for developing all types of AI technologies, general text and data mining, and sentiment analysis.



Applies to commercial and non-commercial uses

E.g., for university research which subsequently takes on commercial application.



Applies to all types of copyright works

E.g., literary, musical, or artistic works, films, and sound recordings.



Cannot be excluded or restricted by contract

Contract terms that try to do so are unenforceable.

CONDITIONS FOR USING CDA EXCEPTION



No using of works beyond the act of copying*

Includes copies made when preparing for CDA, e.g., storage.



No supplying of works to others*

Includes not distributing physical copies or transmitting electronic copies of the works.



Works must be lawfully accessed

Subscribe to paywalled content and database terms of use.



No knowing use of pirated copies

Except for research or study relating to rights infringement.

* Copies may be supplied to a third party only if required to verify CDA results or for collaborative research or study relating to the CDA.

FAIR USE EXCEPTION

(Section 190 of the Copyright Act 2021)

This general exception allows any "fair" use of copyright works. Courts assess fairness based on the specific facts of each case (see page 18 of our [Copyright Infopack](#)). Fair use may apply to both the AI training and deployment phases.

Courts in the United States have found fair use in some cases involving AI training, but outcomes depend on circumstances involved.

(Note: US cases are instructive because the statutory fair use factors in Singapore and the US are similar. Further cases in the US are still pending and current positions may be overturned by subsequent decisions.)



LICENSING WORKS FOR AI TRAINING

Licensing is integral to Singapore's copyright regime and complements the legal exceptions for AI training (e.g., the CDA exception's lawful access requirement supports licensing behind paywalls). This is aligned with global trends, where copyright owners are increasingly partnering with technology companies to structure licensing deals for AI training. Even where an exception could apply, obtaining a licence for AI training could offer additional benefits such as:



Higher Quality & More Reliable Training Data

Compared to raw scraped data, licensed datasets are often of higher quality, curated and consistent. This helps to reduce bias and improve model performance.

Access to Private Datasets

Includes restricted databases or specialised datasets not publicly available.

Collaboration Opportunities

Partnerships with content owners beyond AI training, e.g., co-branding or joint innovation and product development.

PRACTICAL TIPS FOR AI DEVELOPERS



Ensure compliance with CDA exception from the outset

The CDA exception gives legal certainty from the outset if you ensure that you meet all conditions of the exception before using the works (compared to fair use, where the courts determine "fairness" only after the works are used). Developers should regularly review datasets and training practices to remain compliant. Keep proper records on the materials used, e.g., type, source, and how it was obtained, selected, and processed. This helps prove lawful access to a work if there is a dispute.

Consider whether licences should cover AI output

Licences for AI development need not be limited to the training phase. Developers of generative AI systems should consider whether licences should also cover inadvertent reproduction of works in AI-generated output.



Make reasonable checks that datasets do not contain pirated works

The CDA exception requires that you do not knowingly use pirated copies.

Developers should take reasonable steps to verify the sources of training data and avoid flagrantly infringing websites.

