



Q&A



Librarians Navigating Copyright in the Age of GenAI

Victoria F. CAPLAN [_lbcaplan@ust.hk](mailto:lbcaplan@ust.hk)

Head (Research & Learning Support)

The Hong Kong University of Science & Technology Library

January 9, 2026



Disclaimer!

- I am NOT a lawyer
- This is NOT legal advice



Q&A

Outline

1. Copyright, Public Domain, & Creative Commons Basics
2. Navigating: Questions to Steer by
3. Using different materials as sources to “harvest” to “train” computer models
4. TDM Licensing
5. Copyright & GenAI text, images, etc. in Hong Kong



1.1. Copyright



Q&A



Requires written or “concrete expression” of an idea

Examples:

- Poem, play or drama
- Music (the composer's rights)
- Paintings, photos, sculptures
- Computer programs & software
- Sound recordings
- Films & TV broadcasts



Caplan, V. Chinese New Year in Hong Kong, 2011.

Example: I took this photo. It is copyrighted to me.

- I can't copyright the concept or reality of “Chinese New Year in Hong Kong”
- I can copyright the photo



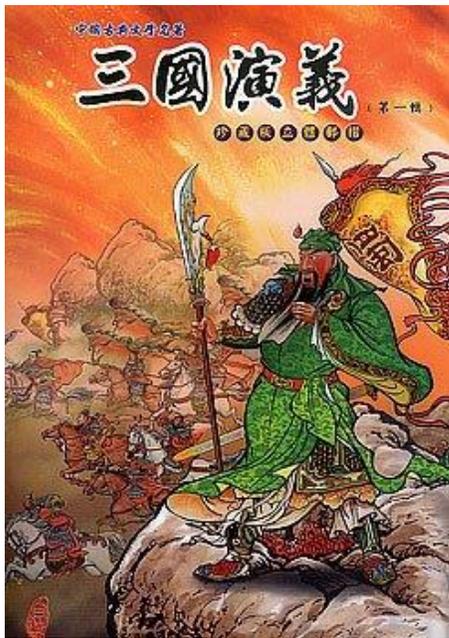
1.3. Public Domain



After a copyright term expires:

Anyone can re-publish it or translate it or make a new version

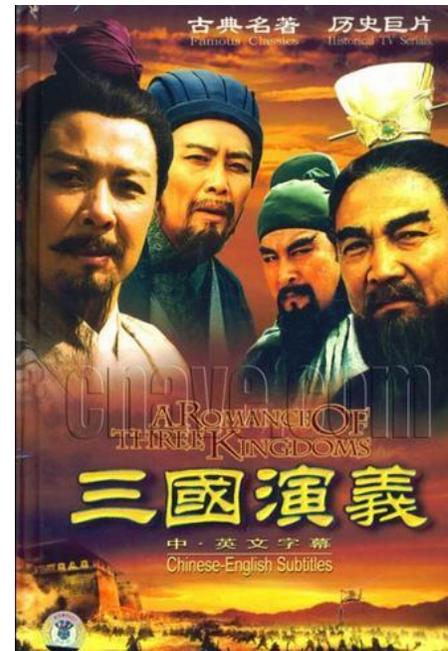
But...Translation, typographic arrangement, or new version may be copyrighted



http://bp1.blogger.com/_A8kK3GAyva0/R-ELBRO1r8I/AAAAAAAAABrs/jV3d0NYf_eU/s400/%E4%B8%89%E5%9C%8B%E6%BC%94%E7%BE%A9%E9%83%B5%E6%91%BA.jpg



<http://www.kilikulu.com/wp-content/uploads/2009/01/redcliff2.jpg>



http://sgpower.tgbus.com/UploadFiles_9949/2007/20090706161126348.jpg



http://www.morningstar.com.tw/pic_M/1099102.jpg

People may use 羅貫中's original work without paying \$\$\$ to his descendants



1.4. Copyright Guides



香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | 李兆基圖書館 LEE SHAU KEE LIBRARY

Guides @ HKUST LIBRARY

HKUST Library / LibGuides / Copyright / Start

Copyright: Start

Provides information on copyright for HKUST students & staff

Start | Fair Dealing for Instructional Purposes | Canvas (LMS) | Playing videos in class | Creative Commons | E-Reserve & Course Packs

Licensed audio or video | GenAI | MOOCs | Help

What is Copyright?

In general, copyright is the right given to the owner of an original work, it is an automatic right for the creator (author). Often, the author then sells it on; for example, a composer sells the copyright to a recording company like Sony.

Who is the Copyright Owner?	The Author of the work
Copyright protects:	<ul style="list-style-type: none"> Literary work (Books, articles) Musical works (Musical compositions) Dramatic works (Plays) Artistic works (Drawings, paintings, sculptures, etc.) Computer programs Sound recordings, films, broadcasts or cable programmes Typographic arrangement...
Copyright last for?	Usually 50 years after the creator of the work dies

Source: Intellectual Property Department (<https://www.ipd.gov.hk/en/copyright/what-is-copyright/index.html>)

Copyright law in HK - Copyright Ordinance (Cap. 528)

The Copyright Classroom - The Basics

Teaching & Research in Hong Kong

Making & Using Photocopies for Study & Research

HKUST Library Statement on Copyright

- Personal
- Private Study, Research, & Instruction
- Reasonable extent
- "Should not conflict with a normal exploitation of a work by the copyright owner nor unreasonably prejudice the legitimate interests of the copyright owner"

Guidelines on the Use of Photocopies in Classroom Instruction

Check out [Guidelines for Photocopying of Printed Works by Not-for-profit Educational Establishments](#), available from the Intellectual Property Department of Hong Kong and their [FAQ](#)

- 3-day rule
- No more than 27 instances in a year per course
- No more than 15 articles from the same newspaper title for one course in any one acad

Also see [Copyright and Education](#) on the HK IPD website.

Learning

The Copyright Classroom : Lesson 1 Learning

The University of Hong Kong Libraries

The University of Hong Kong Libraries

Library / LibGuides / Copyright / Copyright

Copyright

A Guide to Copyright in Teaching, Learning and Research

Copyright

Copyright in Teaching & Learning

Course Packs

Course Materials for MOOC

Copyright in Research

Author's Rights

Creative Commons Licenses

Open Access Publishing

Copyright and Gen-AI

Free Legal Advice Scheme on HKU Campus

Disclaimer: This guide is for informational purposes only and does not constitute legal advice. For specific legal concerns or detailed guidance, please consult a qualified legal professional or contact the Free Legal Advice Scheme on HKU Campus.

Introduction to Copyright

Copyright protects the rights of creators, encouraging creativity and innovation by ensuring that their hard work is properly recognized and rewarded. At the same time, it balances the interests of the public by allowing fair use of copyrighted works for purposes such as education and research.

Infographic | Copyright Basics | Types of Works Protected by Copyright | Duration of Copyright

What Is Copyright?

Copyright is...
A type of intellectual property right that belongs to the creator of an original work.

Expression of Ideas
Copyright protects the expression of ideas, not the ideas themselves.

Exclusive Rights
The law grants copyright owners certain exclusive rights, also known as restricted acts.

Types of Works Protected
Copyright protects nine categories of works.

Duration of Copyright
Generally speaking, copyright lasts for 50 years after the creator's death.

Worldwide Coverage
Works created or first published anywhere in the world are protected by copyright.

Transferrable
Like other intangible personal property, copyright can be licensed or transferred.

The Copyright Classroom - The Basics

The Copyright Classroom : Lesson 1 Learning

The Copyright Classroom : Lesson 1 Learning



<https://libguides.hkust.edu.hk/copyright/>



https://libguides.lib.hku.hk/copyright_hku/home



1.5. Copyright in Hong Kong – Education Exemptions



Copyright Ordinance

- Section 38 – Research & Private Study
- Section 41A - For Purposes of Instruction
- Section 41 – Things done for purposes of instruction or examination
- Section 43 – Performing, playing or showing work in course of activities of educational establishments
- Sections 47-53 – Copying by Librarians

License agreements can create rights & responsibilities beyond the exemptions



1.6. Creative Commons Menu of Choices

License Type	
Attribution - = CC BY	This license lets others distribute , remix, tweak, and build upon your work, even commercially , as long as they credit you for the original creation
Attribution Share Alike = CC BY-SA	This license lets others remix, tweak, and build upon your work even for commercial purposes , as long as they credit you and license their new creations under the identical terms
Attribution-No Derivatives = CC BY-ND	This license allows for redistribution, commercial and non-commercial , as long as it is passed along unchanged and in whole, with credit to you .
Attribution-NonCommercial = CC BY-NC	This license lets others remix, tweak, and build upon your work non-commercially , and although their new works must also acknowledge you and be non-commercial , they don't have to license their derivative works on the same terms.
Attribution-Non-Commercial-ShareAlike = CC BY-NC-SA	This license lets others remix, tweak, and build upon your work non-commercially , as long as they credit you and license their new creations under the identical terms .
Attribution-NonCommercial-NoDerivs = CC BY-NC-ND	This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you , but they can't change them in any way or use them commercially .



Q&A



2. Navigating - Questions to Steer by



W.G.N. (Wicher Gosen Nicolaas) van der Sleen, 1929, cc-by [Wereld Museum, Rotterdam](#).

https://commons.wikimedia.org/wiki/File:COLLECTIE_TROPENMUSEUM_%27Kapitein_achter_het_stuurwiel_van_het_schip_%27Doortje%27_TMnr_10027606.jpg



Q&A



2.1. Questions before Using **Others'** material

What is the status of the item(s)?

- Currently under copyright?
- Creative Commons?
 - What does the license allow me to do?
- Public Domain?

If it is under copyright

- Is the desired activity covered by some of the HK Copyright Exemptions?
- Is the action permitted by a license?
- Who owns or licenses it?
- Who to contact to request permission or to pay \$ for the action?



2.3. Questions for Sharing & Protecting **Own** Material

What does the creator think about people or companies using it for....

- Non-profit educational use?
- For profit & give me some of the \$\$\$?
- Make money from it without sharing the profits with me?
- Do anything (for profit or non-profit) if they acknowledge?
- Do anything, if I can stay anonymous?
- Training LLM models for profit?
- Training in-house LLM for non-profit educational or research purposes?



3.1. LLMs & GenAI Need “Content” to “Train”



Bots and robots “hoover” up humanity’s data, words, images, and sounds



Public Domain Image Archive from PUBLIC DOMAIN REVIEW

Providers

Institution	Images	Digital Rights	Right
BSB Bayerische Staatsbibliothek On the Web / On FOIA		See all 36 images...	Public Domain Mark No
SBB Berlin State Library On the Web / On FOIA			Public Domain Mark No
Berlin State Museums On the Web / On FOIA			Public Domain Mark No
BIBLIOTHÈQUE SAINTE-GENEVIÈVE LIVRES On the Web / On FOIA		See all 35 images...	Public Domain Mark No
BRITISH LIBRARY On the Web / On FOIA		See all 424 images...	Public Domain Mark and others No
Bryn Mawr College Library On the Web / On FOIA		See all 18 images...	Public Domain Mark No

OpenAlex

Search: "whale communication"

1,647 results

Type	Count
article	1,250
book-chapter	104
preprint	87
dissertation	47
dataset	39

Institution	Count
University of St Andrews	64
Woods Hole Oceanographic Institution	55
National Oceanic and Atmospheric Administration	48
United States Department of Commerce	48
Cornell University	45

Topic	Count
Marine animal studies overview	721
Animal Vocal Communication and Behavior	42
Energy Efficient Wireless Sensor Networks	38
Underwater Vehicles and Communication Systems	34
IoT and Edge/Fog Computing	29

Open Access: 47.6% (784)

Modern-tools. cc-by Parvathisri, 2011. https://commons.wikimedia.org/wiki/File:A_vacuum_cleaner.jpg

Public Domain Image Archive Sources - <https://pdimagearchive.org/sources/>

[Open Alex – Whale Communication Query] https://openalex.org/works?page=1&filter=title_and_abstract_search:whale+communication

Open Music Archive <https://www.openmusicarchive.org/>

Open Music Archive

Open Music Archive is a collaborative project, initiated by artists [Eileen Simpson](#) and [Ben White](#), to source, digitize and distribute out-of-copyright sound recordings. The archive distributes these recordings freely, forms a site of exchange, and is a vehicle for future collaborations and distributed projects.

Search and download recordings from the archive

Track Title: Search

Use tags to browse the archive and find recordings.

1920s 1926 1927 1928 1929 ATL 2017 Atlanta Tampa Beat Blues british music hall Camden, NJ charlie poodle Chicago Clara Smith cornet country dallas dance lessons death Discarded One Don't go 'long nobody Eddie Lang idiana female vocal fiddle folk Free-to-air group guitar happy instrumental jazz Johnny Indle Kroo male vocal Memphis money murder music New York Check-paraffol anthology piano Pinetop Smith pump rouse solo violin Virginia Linton vocal Vocalion washboard word When The Deal Goes Down work

Projects Selected Work

Everett/Step 1 Peter De Young
61444 MATTHEW 228, Hershfield: 1 January 2021 - 3 February 2021
Whitshapel Gallery: 27 August 2019 - 5 January 2020
Gulfport Museum and Gallery: 4 July 2019 - 3 November 2019



3.2. LLMs & GenAI Need “Content” to “Train” Deals, Concerns....



An academic publisher has struck an AI data deal with Microsoft – without their authors’ knowledge

Published: July 23, 2024 7:04am BST

Chuttersnap / Unsplash

- Copy link
- Email
- X (Twitter)
- Bluesky
- Facebook
- LinkedIn
- WhatsApp
- Print

In May, a multibillion-dollar UK-based multinational called Informa announced in a [trading update](#) that it had signed a deal with Microsoft involving “access to advanced learning content and data, and a partnership to explore AI expert applications”. Informa is the parent company of [Taylor & Francis](#), which publishes a wide range of academic and technical books and journals, so the data in question may include the content of these books and journals.

According to [reports published last week](#), the authors of the content do not appear to have been asked or even informed about the deal. What’s more, they say they had no opportunity to opt out of the deal, and will not see any money from it.

Academics are only the latest of several groups of what we might call content creators to take umbrage at having their work ingested by the generative AI models currently racing to Hoover up the products of human culture. [Newspapers](#), [visual artists](#) and [record labels](#) are already taking AI companies to court.

While it’s unclear how Informa will react to the rumblings of discontent, the deal

Author



Disclosure statement

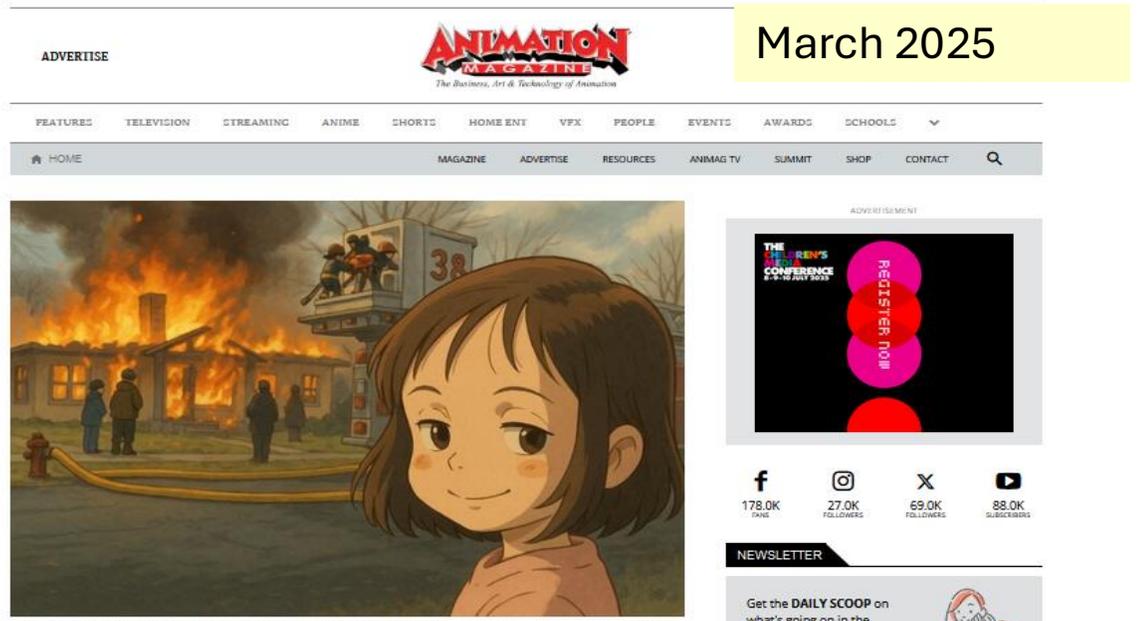
Willett Potter does not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and has disclosed no relevant affiliations beyond their academic appointment.

Partners



University of New England provides funding as a member of The Conversation AI.

[View all partners](#)



OpenAI’s 4o image generation is being used to create “Ghibli style” images based on classic films, world events and internet memes like “Disaster Girl”

“An Insult to Life Itself”: OpenAI’s New Tech Fuels Ghibli Imitation Trend, Raises Questions

By Animation Magazine March 27, 2025

- Facebook
- Twitter
- Pinterest
- WhatsApp

This week, ChatGPT creators OpenAI unexpectedly released the latest version of its image generator tech, built into GPT-4o, which features a “natively multimodal model capable of precise, accurate, photorealistic outputs.” GenAI enthusiasts quickly realized they were able to create images with a “Studio Ghibli style,” flooding the net with Ghibli-style images that will be distributed more widely than ever before.

ADVERTISEMENT

178.0K FANS

27.0K FOLLOWERS

69.0K FOLLOWERS

88.0K SUBSCRIBERS

NEWSLETTER

Get the **DAILY SCOOP** on what’s going on in the **ANIMATION INDUSTRY**

DAILY NEWSLETTER

ADVERTISEMENT

FOR YOUR EMMY® CONSIDERATION IN ALL Ghibli® CATEGORIES

SOLO LEVELING

“...a breakout hit...” - FORBES



Q&A

3.3. LLMs & GenAI Need “Content” to “Train” Lawsuits & Settlements



Next > Tech News

Danish publishers sue OpenAI after failing to strike rights deal

July 2025



Copyright Michael Dwyer/Copyright 2023 The AP. All rights reserved.

By [Cynthia Kroet](#)
Published on 01/07/2025 - 13:28 GMT+2 • Updated 13:29

[Share this article](#) [Comments](#)

The collective action comes as Denmark starts chairing meetings of the EU member states.

An industry group representing almost all of Denmark’s media outlets including broadcasters and newspapers has said it’s suing ChatGPT’s parent company OpenAI for using its content, in a [statement](#) issued on Tuesday.

“Big tech must respect national laws, and as such, DPCMO has a responsibility to act. Pursuing legal action is not a choice; it’s a necessity,” the Danish Press Publications’ Collective Management Organization (DPCMO) said, claiming that: “OpenAI has used DPCMO repertoire in its training, [and] OpenAI does not respect reservations in terms & conditions of a website.”

DPCMO said it reached out to OpenAI “long ago”, in a bid to learn about the product, data, technology and the company as they use information from DPCMO’s members “to train its app ChatGPT”.

But an attempt to negotiate with the publishers as a collective failed, according to the

<https://www.euronews.com/next/2025/07/01/danish-publishers-sue-openai-after-failing-to-strike-rights-deal>

Anthropic to pay authors \$1.5 billion to settle lawsuit over pirated books used to train AI chatbots

September 2025



Thriller novelist Andrea Bartz is photographed in her home, in the Brooklyn borough of New York, Thursday, Sept. 4, 2025 (AP Photo/Richard Drew)

BY [MATT O'BRIEN](#)
Updated 5:15 AM GMT+8, September 7, 2025

[Add AP News to Google](#) [Share](#)

NEW YORK (AP) — Artificial intelligence company Anthropic has agreed to pay \$1.5 billion to settle a class-action lawsuit by book authors who say the company took pirated copies of their works to train its chatbot.

The landmark settlement, if approved by a judge as soon as Monday, could mark a

<https://apnews.com/article/anthropic-copyright-authors-settlement-training-f294266bc79a16ec90d2ddccdf435164>



Q&A

4.1. Text & Data Mining (TDM) Licensing



<https://libguides.hkust.edu.hk/tdm>



Guides @ HKUST LIBRARY

HKUST Library / LibGuides / Text and Data Mining (TDM) from HKUST Licensed Material / Publishers' Policies

Text and Data Mining (TDM) from HKUST Licensed Material

This guide is developed to help HKUST users learn what publishers permit text and data mining via their regular subscriptions.

Publishers' Policies

- Text and Data Mining - HKUST Subscriptions
- Cambridge University Press
- Elsevier
- EIU Viewpoint
- Factiva
- Gale - Cengage
- IEEE
- JSTOR
- Nexis Uni
- Project Muse
- ProQuest
- SAGE
- Springer Nature
- Taylor & Francis
- Web of Science

Text and Data Mining - HKUST Subscriptions

Text and Data Mining (TDM) refers to the process of using automated tools and techniques to extract, analyze, and derive insights from large sets of text and data.

The majority of publishers that support TDM offer the service free of charge. However, there are often rules and requirements that must be followed, and the methods for data mining and delivery can also be quite different.

Commonly Seen Terms and Conditions:

- Use is restricted to **non-commercial** research purposes.
- Only **subscribed and open access content** can be text-mined.
- Follow the **download limit**, e.g. 3 requests per second.
- **Sharing the data** with third parties is prohibited.
- **Delete the data** once the project is completed.
- Use **APIs** to extract data rather than crawling databases with web robots, spiders, etc.

In autumn 2020, HKUST Library's Research Support Services did a small study on text and data mining (TDM) of Library subscribed resources. The findings appeared in a Research Bridge article, [Text and Data Mining: Full-text Databases](#).

Cambridge University Press

Machine Analysis (Text and Data Mining)

Cambridge allows users with **lawful access to its content to perform text and data mining (TDM) for non-commercial purposes**. Users can download, extract, store, and analyze content, provided a link to the original content on Cambridge's site is included. Any health-related content must be deleted upon the research project ends. While TDM results can be shared with library...





Q&A

4.2. Text & Data Mining (TDM) Examples of different licensing

EIU Viewpoint

Text and Data Mining

EIU makes its data available for download and analysis with its “EIU Viewpoint add-in for Excel”. EIU provides a user guide to the [Excel Add-in](#).

The [terms of use](#) allow you to do text and data mining, but any use in either university assignments or for publishing in academic articles must cite EIU as the data source. You must only use the [Excel Add-in](#) or API to access data for use in text and data mining.

Using AI Tools with EIU data

Generative AI and other AI models or products can be developed using EIU content for non-public research and teaching only and when taking specific content protection measures.

- If you are using a third-party platform as a base for an AI model (e.g. OpenAI GPT models), your models must be trained in a secure and “ring-fenced manner”. “Ringfenced” means information that is self-contained to each organization/AI program and isn’t commingled with the rest of the world/internet.
- HKUST’s enterprise version of ChatGPT (via Azure), which is fenced off from training from the underlying OpenAI model, is OK to use.
- Public/non-paying versions of ChatGPT and other generative AI tools may **NOT** be used by HKUST users with the data from EIU.

EIU also offers additional and bespoke licences for AI use. If you would like to discuss these, please email them at licensing@eiu.com.

Factiva

Due to a recent change in their company policy, they are no longer able to offer text mining solutions to customers in the academic sector. (Updated Jun 2025)

Each publisher has its own:

- Rules
- License agreements
- Subject to change
- Subject to negotiation

Librarians must

- Negotiate vigorously
- Record & share agreements
- Inform users
- Update info promptly



Q&A

4.3. Scholarly Publishing GenAI Deal Tracker



Librarians need to keep updated on publishers' arrangements & help inform researchers



Our work Publications Blog People About Contact us

Generative AI Licensing Agreement Tracker

In recent months, several publishers have announced that they are licensing their scholarly content for use as training data for LLMs. These deals illuminate how major publishers are grappling with their strategy amid uncertainty. To understand the dynamics around this fast-developing market, Ithaka S+R has launched a tracker of these licensing deals. In it, we catalog (when the information is available), the publisher, the purchaser, the deal type and size, and the impact of and strategy behind the deal. Roger Schonfeld provides more analysis about these deals in his Scholarly Kitchen piece, "Tracking the Licensing of Scholarly Content to LLMs."

So far, several major publishers have announced deals. For them, there is a substantial near-term revenue upside. The basic idea behind these deals is to generate revenue for the publishing house in exchange for easy, reliable, and legal access to the content for the LLM. A number of companies are in the hunt for this content, including not only OpenAI and Google but also Apple and more specialized providers.

Thus far a standard set of terms or overall model from which to build these deals has yet to emerge. Pricing of course is at the top of mind for everyone, but there are many other considerations as well. There are technical and reputational questions about how corrections or retractions will propagate through an LLM and whether an author can opt out, and there are business model issues such as whether provenance will be tracked through the output from an LLM such that a citation or link can be provided back into the scholarly record, just to take several straightforward examples.

We will update the tracker periodically. You may also access the tracker as a [Google Sheet](#). If you are aware of other deals that we have not yet documented in this tracker, please contact us using the form below.

GenAI Licensing Agreement Tracker : Publisher Deals

last updated on August 1, 2025

	Purchasee (content provider)	Purchaser	Deal type	Date	Financial arrangement	Content type	
	Wiley	Large tech company	Content licensing	Q4, 2024	One time payment, nonexclusive	Academic and professional books	\$
	Wiley	Potato	Content licensing	Announced Oct 2024	Not disclosed, but SVP and GM for AI Growth mentioned that for a revenue share agreement is a possibility they are open to for AI licensing deals	Relevant content to the assistant, which focuses on biology and is built on peer-reviewed literature	?
For more information about this tracker, please visit https://sr.ithaka.org/our-work/generative-ai-licensing-agreement-tracker .	Wiley	Perplexity	Content licensing	Announced May 2025	Not disclosed	Educational collections and resources in areas such as nursing, business, and engineering, including e-textbooks	N l i d r
To share comments, questions, and suggestions for additional entries, please fill out this form	Wiley	Amazon Web Services	Content licensing	Announced May 2025	Not disclosed	Full text scientific journals	N e e t c A I



<https://sr.ithaka.org/our-work/generative-ai-licensing-agreement-tracker/>



Q&A



4.4. Scholarly Publishing GenAI Deal Tracker (b)

https://docs.google.com/spreadsheets/d/e/2PACX-1vRJ8ZPFRAY6GKX_026vvtQk6ExJSbil7X_xSrGDEdqdX1BPzAMZccRFe5mhdxcK5YRwOIXg4pFY7LWm/pubhtml

GenAI Licensing Agreement Tracker : Publisher Deals										
last updated on August 1, 2025										
	Purchaser (content provider)	Purchaser	Deal type	Date	Financial arrangement	Content type	Deal size	Impact	Strategy	Link
	Wiley	Large tech company	Content licensing	Q4, 2024	One time payment, nonexclusive	Academic and professional books	\$23m	Select content used to train LLM	Wiley's former SVP for Strategy is now SVP and General Manager for AI Growth. The job description is "Lead new AI growth opportunities in content licensing, AI application development, and new business models," showing a big commitment to content licensing. They have "have developed licensing frameworks that offer flexible, fair terms tailored to different use cases and development needs." Released a statement against the illegal scraping of their content. "We are working to uncover similar content opportunities with other AI players and remain convinced that the future development of LLMs is best served by the high-quality structured content that Wiley delivers"	https://www.expectst-out/ https://www.illegal-scr
	Wiley	Potato	Content licensing	Announced Oct 2024	Not disclosed, but SVP and GM for AI Growth mentioned that for a revenue share agreement is a possibility they are open to for AI licensing deals	Relevant content to the assistant, which focuses on biology and is built on peer-reviewed literature	?	Used to help build Potato's tools. No details provided, but tools include automated paper review and a lab protocol generator.	This is the first deal announced as part of Wiley AI Partnerships, a "co-innovation program" which "aims to develop new AI applications, assistants, and agents in partnership with innovative companies, to empower researchers and practitioners and help drive the pace, efficiency and accuracy of scientific discovery." They are specifically hoping to work with start-ups and scale-ups.	https://www.deals/arti
For more information about this tracker, please visit https://sr.ithaka.org/our-work/generative-ai-licensing-agreement-tracker .	Wiley	Perplexity	Content licensing	Announced May 2025	Not disclosed	Educational collections and resources in areas such as nursing, business, and engineering, including eTextbooks	Not disclosed, but AI licensing segment is driving a lot of their revenue growth	"Students can access assigned Wiley curriculum materials through their institution's Enterprise Pro subscription, eliminating the need to switch between platforms."	Wiley is using this limited use of AI to increase the value of purchasing a Wiley educational collection. By requiring the purchase of these materials, it maintains that revenue stream while increasing the value proposition (as well as the Perplexity Enterprise Pro value proposition to academic institutions)	https://www.expectst-out/
To share comments, questions, and suggestions for additional entries, please fill out this form .	Wiley	Amazon Web Services	Content licensing	Announced May 2025	Not disclosed	Full text scientific journals	Not disclosed, but Alvest estimated that "Subscription-based access to the AI toolkit could add \$100-\$150 million annually to Wiley's top line by 2027"	AWS has built an open source toolkit for healthcare and life sciences, which "offers a catalog of starter agents and an orchestration framework for organizations to build and customize their agentic systems, supporting use cases from biomarker discovery to clinical trial protocol generation." Until this point, the toolkit relied on open source content.	It appears that in order to use the Wiley agent within the broader AWS toolkit, users will pay an additional subscription fee. This is an interesting model similar to Wiley's Perplexity deal, as it is packaged as an optional enhancement to an existing system where the end user pays for the additional content. The press release also stresses the full text search component of this agentic model heavily (in contrast to more common abstract-only search models). In Wiley's 2025 Quarter 1 Earnings Call, Wiley's AI licensing strategy was described as a two-part model: "We see the content licensing opportunity in two stages. The first as discussed is participating in the near-term development of foundational models. The second is in recurring licensing arrangements over the medium to long term, as these models and applications come online and as information-centric corporates bring our content into their AI environments."	https://news.wiley.com/insights/2025/08/01/earnings-tran
	Wiley	Anthropic	Content licensing	Announced July 2025	Undisclosed, requires dual-subscription to Wiley and Claude to access Wiley content and Claude's MCP integration chatbot	Scholarly research, which highlighted including for life sciences, education, and earth science applications, among others	Not disclosed	Wiley's Model Context Protocol (MCP) integration is part of a Claude for Education pilot program. MCP is "an open standard that will enable integration between peer-reviewed content and AI platforms", with a beta version for Wiley MCP available in fall 2025. Anthropic describes the partnership as "a blueprint for how AI should integrate with scholarly content." Pilot participation will allow Wiley to establish best practices for citations and copyright. Being a part of the Claude for Education initiative promotes Wiley's products to Claude's higher education institutional partners.	The MCP integration requires a premium subscription to Claude. Though there is not an online source on how MCP deals with subscription access to Wiley, other MCP integrations such as Claude and Canva require users to have a paid Canva subscription, indicating that users would need to have a pre-existing Wiley subscription to use the integration. "The future of research depends on keeping high-quality, peer-reviewed content central to AI-powered discovery. This partnership sets the standard for integrating trusted scientific content with AI platforms while creating a scalable solution for other institutions. By adopting the Model Context Protocol, we're ensuring authoritative research remains discoverable in an AI-driven landscape." - Josh Jarrett, Senior Vice President of AI Growth at Wiley in Anthropic's Press Release	https://www.ithaka.org/our-work/generative-ai-licensing-agreement-tracker
	Taylor & Francis / Informa	Microsoft	Content licensing		Informa will be paid \$10m+ for "an initial data access" of the works it has the rights to, with a recurring payment of an undisclosed sum to be made over the subsequent three years. Nonexclusive.	"advanced learning content and data" - seems to include books but not journals	\$10m	Used to "help improve relevance and performance of AI systems"	Facing backlash from authors, although they claim to be "protecting the integrity of our authors' work," limiting verbatim text reproduction as well and paying out royalties according to contracts. There was reported communications to authors from Routledge (a T&F imprint) which stressed the importance of meeting publishing targets for 2024, with some speculation that this is due to the targets being a part of the licensing agreements. Reported first half 2025 revenue of "£15 million incremental, non-recurring data licensing revenue" at 2025 Half-Year Results meeting , though did not disclose the revenue's company or contract source.	https://theccc.deal-with-mix
	Taylor & Francis / Informa	Unnamed big tech firm	Content licensing			"advanced learning content and data" - seems to include books but not journals	\$75m	Used to "help improve relevance and performance of AI systems"	Facing backlash from authors, although they claim to be "protecting the integrity of our authors' work," limiting verbatim text reproduction as well and paying out royalties according to contracts.	https://www.ithaka.org/our-work/generative-ai-licensing-agreement-tracker
	Taylor & Francis / Informa	Unnamed LLM data licensing partner	Data licensing	June / July 2025	Undisclosed	Content related to "Academic Markets" per June 2025 Annual General Meeting Trading Update	Not disclosed	Not disclosed	A third LLM deal was reported in the Q&A session of Informa's Half-Year Results meeting. The LLM partner was not disclosed, though announced as a "different customer" from other AI licensing contracts. Informa wants to respect the confidentiality of the deal to support the "competitive advantage" of their AI partner	https://www.ithaka.org/our-work/generative-ai-licensing-agreement-tracker
	Johns Hopkins University	Unnamed large language model	Content licensing	Announced July 2025	Undisclosed, authors	Books	Not disclosed	Content used to train LLMs	- In an email to authors, JHUP executive director stated that an AI licensing	https://www.ithaka.org/our-work/generative-ai-licensing-agreement-tracker





Q&A

5.1. GenAI & Copyright in Hong Kong



<https://www.ipd.gov.hk/en/copyright/current-topics/public-consultation-on-copyright-and-artificial/index.html>

Intellectual Property Department
The Government of the Hong Kong Special Administrative Region
of the People's Republic of China

Home • About Us • Contact Us • Other Languages

香港 · 區域知識產權貿易中心
HONG KONG · Regional IP Trading Centre

IP Overview Trade Marks Patents Designs **Copyright** Online Services Promotional Activities Mainland and International Cooperation Tools and Resources

COPYRIGHT

Home > [Copyright](#) > [Current topics](#) > Public Consultation on Copyright and Artificial Intelligence (2024) - Completed

Public Consultation on Copyright and Artificial Intelligence (2024) - Completed

Consultation – 2024 Summer

- [Consultation Paper](#)
- [Press Release: Government launches public consultation on enhancing Copyright Ordinance regarding protection for artificial intelligence technology development](#)
- [Legislative Council Panel on Commerce, Industry, Innovation and Technology – Administration’s paper on “Enhancement of the Copyright Ordinance regarding Protection for Artificial Intelligence Technology Development”](#) and Power-point presentation materials (Chinese version only) (July 2024)
- [Public Forum](#)
- [Written Submissions](#)
- [Legislative Council Panel on Commerce, Industry, Innovation and Technology – Administration’s paper on “Enhancement of the Copyright Ordinance regarding Protection for Artificial Intelligence Technology Development – Outcomes of Public Consultation and Proposed Way Forward”](#) (February 2025)

What is Copyright

- Copyright Laws
- Legislative Proposals and Amendments >
- Copyright Licensing Bodies Registry
- Copyright Tribunal
- FAQs and Guidance Notes >
- Current topics**
 - **Public Consultation on Copyright and Artificial Intelligence (2024) - Completed**



5.2 GenAI & Copyright in Hong Kong



<https://www.legco.gov.hk/yr2025/english/panels/ci/papers/ci20250218cb2-240-4-e.pdf>

LC Paper No. CB(2)240/2025(04)

For discussion on
18 February 2025

Legislative Council
Panel on Commerce, Industry, Innovation and Technology

Enhancement of the Copyright Ordinance regarding
Protection for Artificial Intelligence Technology Development –
Outcomes of Public Consultation and Proposed Way Forward

PURPOSE

The Government conducted a two-month public consultation on copyright and artificial intelligence (“AI”) from 8 July to 8 September 2024 and briefed Members on the contents of the consultation at the meeting of the Panel on Commerce, Industry, Innovation and Technology on 16 July 2024. This paper briefs Members on the outcomes of the public consultation and sets out the Government’s proposed way forward for further enhancement of the Copyright Ordinance (Cap. 528) (“CO”) regarding protection for AI technology development as pledged in the Chief Executive’s 2024 Policy Address.

PUBLIC CONSULTATION AND PROPOSALS OF THE CONSULTATION PAPER

2. As mentioned during the briefing to the Panel on the public consultation in July last year, while actively building a comprehensive AI ecosystem, the Government is also mindful of the copyright issues arising from the technological development of AI (especially generative AI). To ensure that the local copyright regime stays robust and keeps pace with the times to align with technological and economic developments and maintain Hong Kong’s competitive advantages, we see the need to focus on exploring further enhancement of the CO regarding protection for AI technology development to support high-quality development of the local innovation and technology (“I&T”), cultural and creative industries, and to

dovetail with Hong Kong’s development into a regional intellectual property (“IP”) trading centre, an international I&T centre and an East-meets-West centre for international cultural exchange as outlined in the 14th Five-Year Plan. Therefore, in accordance with the Chief Executive’s 2023 Policy Address, we conducted a two-month public consultation on this subject from July to September last year.

3. For issues relating to generative AI and copyright, we have reviewed the relevant legislation in Hong Kong and other major jurisdictions as well as the prevailing market situations. The consultation paper looks into the issues in a holistic manner and our views are set out for public consultation as follows:

- (a) **Copyright Protection of Works Generated by Generative AI (“AI-generated works”):** The existing CO already contains applicable provisions to protect the copyright of AI-generated works. Issues relating to the authorship and originality of such works should be determined based on all the circumstances of each individual case;
- (b) **Copyright Infringement Liability for AI-generated Works:** The relevant provisions in the existing CO are broad and technology-neutral, which are sufficient to address copyright infringement involving AI-generated works.
- (c) **Possible Introduction of Specific Copyright Exception:** The possible introduction of copyright exception will allow reasonable use of copyright works for computational data analysis and processing (commonly known as “text and data mining exception” (“TDM exception”)), which covers conventional text and data mining and the training of AI models, and will be applicable to both non-commercial and commercial uses to foster the growth of the AI industry. The proposed TDM exception will ensure that the legitimate interests of copyright owners are safeguarded so as to strike a proper balance of interests between copyright owners and the public; and



Q&A

5.3. GenAI & Copyright in Hong Kong



<https://www.legco.gov.hk/yr2025/english/panels/ci/papers/ci20250218cb2-240-4-e.pdf>

(c) **Possible Introduction of Specific Copyright Exception:** The possible introduction of copyright exception will allow reasonable use of copyright works for computational data analysis and processing (commonly known as “text and data mining exception” (“TDM exception”)), which covers conventional text and data mining and the training of AI models, and will be applicable to both non-commercial and commercial uses to foster the growth of the AI industry. The proposed TDM exception will ensure that the legitimate interests of copyright owners are safeguarded so as to strike a proper balance of interests between copyright owners and the public; and

- No legislation yet – "Possible" - Maybe, in the future
- License agreements can allow greater or less rights than in ordinance



5.4. GenAI & Copyright in Hong Kong



11. Authorship of work

- (1) In this Part *author* (作者), in relation to a work, means the person who creates it.
- (2) That person is taken to be—
 - (a) in the case of a sound recording, the producer;
 - (b) in the case of a film, the producer and the principal director;
 - (c) in the case of a broadcast, the person making the broadcast (see section 8(3)) or, in the case of a broadcast which relays another broadcast by reception and immediate re-transmission, the person making that other broadcast;
 - (d) in the case of a cable programme, the person providing the cable programme service in which the programme is included;
 - (e) in the case of the typographical arrangement of a published edition, the publisher.
- (3) In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author is taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.
- (4) For the purposes of this Part a work is of *unknown authorship* (作者不為人知) if the identity of the author is unknown or, in the case of a work of joint authorship, if the identity of none of the authors is known. (*Amended E.R. 2 of 2019*)
- (5) For the purposes of this Part the identity of an author is regarded as unknown if it is not possible for a person to ascertain his identity by reasonable inquiry; but if his identity is once known it shall not subsequently be regarded as unknown.

[cf. 1988 c. 48 s. 9 U.K.]

Are computer-generated works copyrightable?

In Hong Kong, computer generated works are copyrightable.

- [Copyright Ordinance, section 11](#)

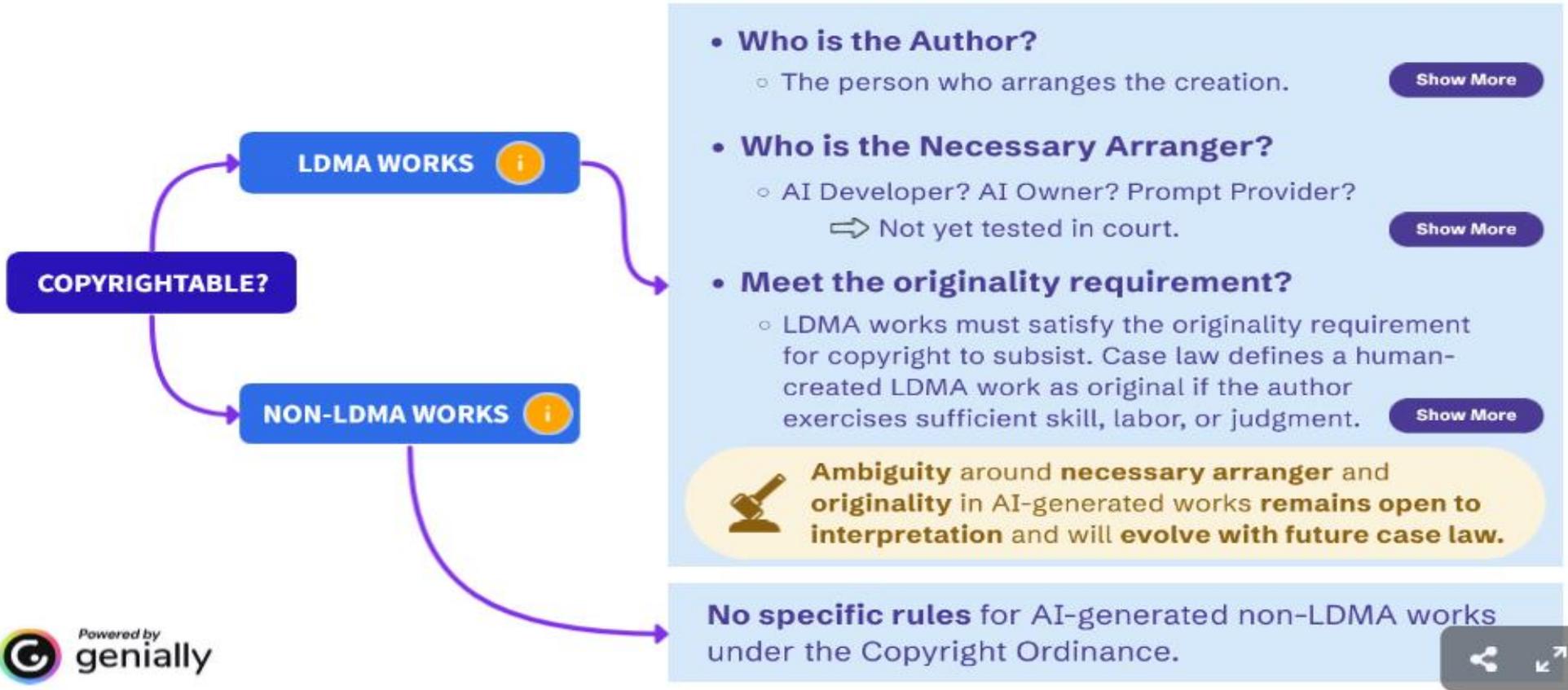


5.5. GenAI & Copyright in Hong Kong



LDMA ???
Literary
Dramatic
Musical
Artistic

Are AI-Generated Works Copyrightable?





Q&A

5.6. GenAI & Copyright in Hong Kong



Protecting Works from AI Training

Protecting Works from AI Training

Copyright owners can take additional steps to **prevent AI tools from using their work**, such as:

Artwork & Images

- Using **Glaze** (from the University of Chicago), which includes "WebGlaze" to help artists keep their work from being scraped and used by tools like MidJourney or StableDiffusion without their consent. A defensive tool.
- Using **Nightshade** (also from the University of Chicago) changes images into 'poison' samples, "so that models training on them without consent will see their models learn unpredictable behaviors..." A proactive tool.

Websites & text

- Implementing **robots.txt** to block web crawlers.
 - This [GitHub site](#) maintains a list of AI agents and robots to block.
 - [No Robots\(.txt\): How to Ask ChatGPT and Google Bard to Not Use Your Website for Training](#), a Dec. 2023 article from the Electronic Frontier Foundation (EFF) also provides information on how to use it.

Publications - Legal clauses & addendums

- USA-based Author's Guild has a "[model clause to prohibit AI Training](#)" (2025) that authors can use when signing publication agreements with publishers

Useful References

- Check policies on AI training via the [Generative AI Licensing Agreement Tracker](#).
- [How Copyright Relates to GenAI](#) from University College of London (UCL): while HKSAR law and British law are different, the HKSAR takes reference from UK law and other common law jurisdictions, and the language of the ordinance for computer generated work is quite similar.

Protecting work from GenAI Harvesting ? Possible? ? Desired?

Creative Commons - CC Signals

Creative Commons is seeking to create a way between trying to block or poison web scrapers and text and data miners, and allowing a free-for-all that may impoverish original authors, artists, musicians, and other creators, while enriching billionaires.

"CC signals are designed to offer a new way for stewards of large collections of content to indicate their preferences as to how machines (and the humans controlling them) should contribute back to the commons when they reuse and benefit from using the content."

Go to [CC Signals: A New Social Contract for the Age of AI](#) to learn more. It includes,

- **CC Signals Implementation**
 - [A report - From Human Content to Machine Data: Introducing CC Signals](#)
 - [Technical Implementation of CC Signals on Github](#)



Q&A

5.7. GenAI & Copyright in Hong Kong



<https://libguides.hkust.edu.hk/ai-literacy/faqs>



Guides @ HKUST LIBRARY

HKUST Library / LibGuides / AI Literacy for End-Users / Responsible Use of AI - FAQs

AI Literacy for End-Users: Responsible Use of AI - FAQs

This guide is created to help HKUST students and staff develop their "AI Literacy" in the realm of higher education learning, teaching, and research. It is a work in progress

Search this Guide Search

- Use AI Wisely
- AI Tools & Directories
- Comparison of GenAI Tools
- Citing & Acknowledging AI
- Responsible Use of AI - FAQs**
- AI and Copyright

This page addresses some common questions for students and staff about using AI tools while ensuring compliance with intellectual property rights and academic standards.



FAQs - Using AI-generated content & Academic Integrity



Using AI-generated content & Academic Integrity

- Can I use AI-generated text or images in my course assignment?
- Can I use AI-generated text or images in journal publications?
- Do I need to disclose AI use for literature discovery?
- Do I need to disclose AI use for grammar checks and rephrasing?
- How should I properly acknowledge the use of AI-generated content in my course assignments and journal publications?

Q: Can I use AI-generated text or images in my course assignment?

HKUST gives faculty members the flexibility to set their own policies regarding the use of AI in their courses. So, consult your instructor for guidance.

In general, CEI suggests that, whether or not AI tools are used, students should be prepared to explain their work, justify their sources, provide drafts, and discuss alternative perspectives. For more details, check [HKUST's Policy for GenAI integration in teaching and learning](#).

Q: Can I use AI-generated text or images in journal publications?

Many publishers have established guidelines on the use of AI. Key principles include:

FAQs - Copyright, Intellectual Property

Copyright & Intellectual Property



- Can I upload journal article full-text PDFs to AI tools to create summaries for my research project?
- Can I upload open access materials to AI tools to create summaries for my research project?
- Can I use library databases to train AI or LLM models?
- Can I use open access materials to train AI or LLM models?
- How to deal with personal data when training AI or LLM models?

Q: Can I upload journal article full-text PDFs to AI tools to create summaries for my research project?

No. Uploading entire copyrighted materials, such as journal articles or book chapters, to AI tools is generally **prohibited by most publishers' licensing agreements**.

These agreements typically restrict sharing full-text content with third-party platforms (including AI tools) to protect intellectual property rights. While some AI tools claim that uploaded content remains private, always review the specific terms of both the content provider and the AI platform before proceeding.

Instead, consider using AI to summarize **your own notes**.



Thank you - 多謝

