

# Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models

February 2026



Ministry of Culture, Sports  
and Tourism



KOREA COPYRIGHT  
COMMISSION

## Preface

- This *Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models* aims to present fundamental criteria and considerations that may serve as reference points when determining the applicability of the fair use doctrine to the training of generative artificial intelligence, taking into account a wide range of situations in which works protected under the Copyright Act — copyrighted works, related rights subject matter, and databases\* — are used during the training process, as well as analyzing domestic and international cases concerning fair use

\* For the purposes of this Guide, the term “works” is used without distinction to refer both to “works” encompassing copyrighted works, related rights subject matter, and databases, and to “works” as manifestations of creative expression.

- Under the Copyright Act, limitations and exceptions to copyright include not only specific and enumerated restrictions on copyright (Articles 23 through 35-4), but also a comprehensive provision on fair use (Article 35-5). The latter is a general provision limiting economic rights and was designed to enable flexible response to rapidly changing social, cultural, and technological environments. It also provides a legal foundation capable of being reasonably applied to new forms of exploitation of copyrighted works.

\* Hereinafter, this Guide will refer to Article 35-5 as “fair use” or the “fair use provision” as a comprehensive limitation on copyright.

- However, because the fair use provision under the Copyright Act must be assessed on a case-by-case basis in light of specific facts, unauthorized use of another person’s work requires careful examination for each individual case. Thus, whether a particular use qualifies as fair use cannot be definitively determined until a final judgment is rendered by a court.
- The Republic of Korea is a party to major international treaties, including the Berne Convention, the WTO Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), and the WIPO Copyright Treaty (WCT). Based on these international norms, Korea

has established its domestic legal framework and has continuously improved its copyright system while complying with international treaty standards. The copyright-related provisions established under this framework apply equally to both domestic and foreign copyright holders and users of works within Korea.

- This *Guide* was finalized following review in 2025 by the Special Sub-Group of the AI-Copyright Law Working Group (ACLWG), composed of experts from academia and the legal profession, as well as after consultations with relevant government ministries and collecting of opinions from stakeholders.

\* This *Guide* is based on the Copyright Act and judicial precedents in force as of February 2026. Its contents may be subject to change in the event of subsequent legislative amendments, new court decisions, or further developments in artificial intelligence technologies.

\*\* It is not an authoritative interpretation of the fair use doctrine by the Ministry of Culture, Sports and Tourism or the Korea Copyright Commission. The cases and examples included here are provided solely for illustrative and explanatory purposes.

\*\*\* Recent overseas court decisions relating to generative AI mentioned in this *Guide* are decisions of lower courts, and their legal conclusions may be altered through review by higher courts in the future.

\*\*\*\* Some footnotes and references have been revised during translation from the original Korean to English to enable better understanding by global readers.

# CONTENTS

---

## Preface

2

## I Generative Artificial Intelligence

6

1. Generative Artificial Intelligence and Copyright 8
2. Concept and Scope of GAI Training 9
3. Training Process of GAI 10

## II Can GAI Infringe Copyright?

12

1. Use of Copyrighted Works 14
2. Use of Works During Training 15
3. Principle of Prior Authorization 17
4. Copyright Licensing Agreements Currently Signed 17
5. Trend in Major Countries 20



### **III What is Fair Use 26**

1. Overview of the Fair Use Provision 28
2. Factor 1 – The Purpose and Character of the Use 31
3. Factor 2 – Type and Purpose of the Work 37
4. Factor 3 – The Amount and Substantiality of the Portion Used 39
5. Factor 4 – The Effect on the Current Market or Value,  
or Potential Market or Value of the Work 41
6. Summary of Fair Use Considerations for Each Factor 46
7. Case Studies Related to Fair Use 50

### **IV Responding to Copyright Disputes Involving Generative AI 64**

1. Probability of a Dispute 66
2. Response 66

### **[Appendix] AI-Copyright Law Working Group (ACLWG) Special Sub-Group 74**

I

---

# Generative Artificial Intelligence

# Guide on Applicability of the Fair Use Doctrine

to Training of Generative AI Models



- 
1. Generative Artificial Intelligence and Copyright
  2. Concept and Scope of GAI Training
  3. Training Process of GAI



# Generative Artificial Intelligence

## 1. Generative Artificial Intelligence and Copyright

- The rapid advancement of generative artificial intelligence (GAI) is accelerating innovation across society and creating new sources of industrial value.
- Countries around the world recognize AI technology as a core future growth engine and are trying to secure a competitive advantage in global AI leadership. Against this backdrop, some are calling for policy support measures that permit, as broadly as possible, the use of training data and encourage AI development.
- However, rightsholders express very real concerns that their works are being used at scale without their permission as AI training data, resulting in the erosion of the value of their works while reasonable compensation is not being paid and their jobs are at risk.
- Constitutionally protected copyright and AI-driven technological innovation are values that must not require sacrifice of one for the other. Rather, adoption of mechanisms that enable their coexistence and mutual reinforcement are needed.
- At this juncture, the most critical policy challenge is to identify a balanced approach that respects the legitimate rights of rightsholders and incentivizes creative endeavors, at the same time avoiding undue constraints on AI innovation and promoting mutual growth and development.
- Ultimately, innovation in AI development and the protection of copyright are not inherently conflicting values. When harmonized as complementary principles, they can together drive broader cultural and economic progress for society as a whole.

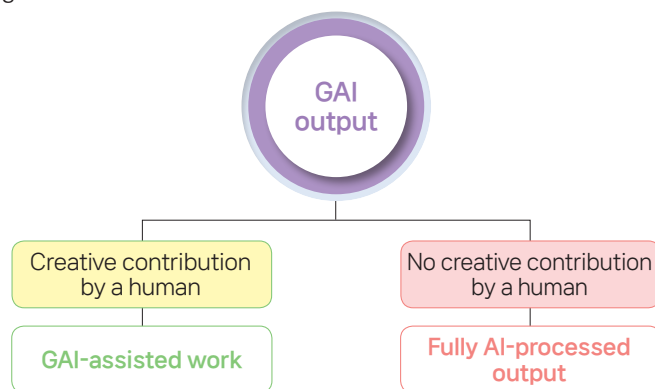
## 2. Concept and Scope of GAI Training

- **GAI training** refers to a series of processes through which data, including works, are collected and pre-processed, which are used by the model to learn statistical rules and patterns and then fixed as internal parameters (variables that show the characteristics of a particular system or a function). This process includes various stages such as data acquisition, data cleaning and transformation, model training, and performance evaluation and optimization, and may involve the reproduction or transformation of works at each stage.

- GAI produces text, audio, images, video, and other outputs by imitating the structure and characteristics of input data. When such outputs are identical or substantially similar to human-created works, copyright disputes may arise.

\* The generation of outputs by GAI is a separate process from the training. Issues regarding copyright infringement of AI-generated outputs are addressed separately in the *Guide for the Preventing Copyright Disputes Related to Generative AI Outputs*.

- A **GAI output** refers to content — such as text, audio, or video — generated in response to a user’s specific input or prompt. A **fully AI-processed output** refers to an AI-generated result that contains no human creative input. On the other hand, a **GAI-assisted work** refers to a result created using generative AI as a tool in the creative process, where a human’s creative contribution is evident and is legally recognizable.



[Diagram] Types of GAI output



# Generative Artificial Intelligence

## 3. Training Process of GAI

- GAI training generally goes through the following stages: Data collection → Data preprocessing → Model training → Model evaluation and optimization
- **Data collection** refers to the process of gathering original data (such as text, images, music, and video) for use in GAI training.

### Various means of obtaining training data

- Use of open or public datasets
- Acquisition of private datasets from third parties
  - Licensing agreements with rightsholders or their agents
  - Acquisition of private datasets from third parties (e.g., database operators or data brokers)
- Web crawling (collecting data using automated programs)
- Web scraping (selectively extracting only the necessary data)
- Direct creation of data (e.g., direct collection, surveys)
- Use of synthetic data (virtual data mimicking real data used for training AI)
- Collection of data from or digitization of offline sources (e.g., digitizing analogue texts and images)

- **Data preprocessing** refers to technical procedures —such as data deletion (cleaning), splitting (tokenization), and enhancing consistency (normalization) — performed to improve data quality and convert data into a form more suitable for training.
- **Model training** refers to the process where preprocessed data is used to train an AI model.
- **Model evaluation and optimization** involve assessing model performance using validation datasets. Model structure, learning rates, and parameters are adjusted to optimize the process.

# Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models

to Training of Generative AI Models

# II

---

**Can GAI Infringe  
Copyright?**

# Guide on Applicability of the Fair Use Doctrine

to Training of Generative AI Models



1. Use of Copyrighted Works
2. Use of Works During Training
3. Principle of Prior Authorization
4. Copyright Licensing Agreements Currently Signed
5. Trend in Major Countries



# Can GAI Infringe Copyright?

## 1. Use of Copyrighted Works

- An author, under the Copyright Act, has the right to reproduce their work\*. (Article 16)  
\* A work refers to a creative work that expresses human thoughts and emotions. The Copyright Act protects only works. Works whose term of protection has expired or works not eligible to be protected will not be included in the scope of protection, meaning they may be used without the risk of infringing rights, regardless of whether fair use apply.

### Copyright Act

#### Article 7 (Works Not Protected)

No work which falls under any of the following subparagraphs shall be protected under this Act:

1. Constitution, Acts, treaties, orders, municipal ordinances and municipal rules;
2. Bulletins, public notifications, directives and others similar thereto which are issued by the State or local governments;
3. Judgments, decisions, orders, or adjudications of courts, as well as resolutions and decisions made by the administrative appeals procedures, or other similar procedures;
4. Compilations or translations of those as referred to in subparagraphs 1 through 3 which are produced by the State or local governments;
5. Current news reporting which is the mere delivery of facts.

#### Article 24-2 (Free Use of Public Works)

(1) A work produced on business and already made public by the State or a local government, or a work of which the author's property right is owned in its entirety by the State or a local government under a contract, may be used without permission: Provided, That the same shall not apply when the work falls under any of the following cases:

1. Where it includes any information pertaining to national security;
2. Where it corresponds to an individual's privacy or confidential business information;
3. Where it includes any information of which disclosure is restricted under other Acts;
4. Where it is registered with the Korea Copyright Commission under Article 112 and is managed as State-owned property under the State Property Act or as public property under the Public Property and Commodity Management Act.

## 2. Use of Works During Training

- **Reproduction** refers to fixing a work on a tangible object temporarily or permanently, or making a copy by printing, photographing, copying, sound or visual recording, or other means (Article 2.22). It includes not just copying that is visible, such as printing or photographing, but also copying into forms that can be played back, such as audio-recording or video-recording. There is essentially no limit to the method of copying. Storing digital works on electronic media such as cloud storage or computer hard drives also constitutes reproduction.
- Works are stored during each stage of GAI training and such storage — regardless of the form of the storage medium — constitutes reproduction under the Copyright Act. Unless there are specific circumstances that fall under copyright limitations or exceptions, using copyrighted works without the permission of the rightsholder may lead to liability for copyright infringement.
- GAI training refers to the entire process of collecting and preprocessing data, and training, evaluating and optimizing a model. Reproduction may not occur at every individual stage but if it does during any one of the stages, the copyright holder can claim their right of reproduction.
- During the data collecting stage, AI developers typically gather data or works from publicly accessible websites or repositories, and these data and works are then stored, transmitted, or transformed. Copyright-protected works may also be included. Furthermore, automated collection of data (e.g., crawling or scraping) may also entail creating copies, and storing and using of the data (in part or in whole).
- Data preprocessing consists of technical processes such as removing unnecessary information (cleaning), splitting (tokenization), and enhancing consistency (normalization) to convert and store data in forms suitable for machine-reading. It is commonly performed by loading works onto cloud-based or other types of storage environments for processing, during which the whole or part of a work may be fixed and utilized within the storage system.



## Can GAI Infringe Copyright?

- During text tokenization or format conversion, a work may be split into multiple parts or converted into another format. At this point, data that is substantially identical or similar to the original work may be generated, stored, and used.
- Model training involves repeatedly inputting and analyzing preprocessed data to identify statistical rules and patterns, which are then converted into internal parameters. During training, large volumes of copyrighted works may be processed in bulk and repeatedly, and copies may be stored and used during the process.
- Model evaluation refers to the process of inputting evaluation datasets and confirming the results to verify model performance while model optimization refers to improving performance by adjusting the parameters, additionally training the model using supplementary data, analyzing errors, and refining the model. When evaluation or additional data are inputted during evaluation and optimization, copyrighted works may again be stored and processed just as in data collection, preprocessing, and training.

### 3. Principle of Prior Authorization

- Generally, using another person’s copyrighted work requires prior authorization from the rightsholder. No particular form or special procedure is required to obtain authorization, as long as permission comes from (1) the copyright holder (2) regarding the scope, method, and conditions of use.
- When the purpose of use is GAI training, authorization from the rightsholder is required because copyrighted works are copied in the process, unless the use is justified under statutory limitations or exceptions to copyright, either specific or comprehensive as in the fair use provision.

### 4. Copyright Licensing Agreements Currently Signed

- In recent years, rightsholders and AI developers have entered into licensing agreements for the use of copyrighted works in AI training in various jurisdictions, including Korea, the United States, and the European Union (as of November 25, 2025).

#### 〈Major Licensing Agreements in Korea〉

Date	AI Developer	Rights Holder / Publisher	Content of license
Feb. 2025	Perplexity	Edaily	(Literary sector) Edaily Economist, Edaily M Content <ul style="list-style-type: none"> <li>• Technical support (AI search, data analytics tools) provided in exchange for content without monetary payment</li> <li>• Advertising revenue generated through source-attributed search to be shared with the media outlets</li> </ul>
Feb. 2025	Perplexity	Maeil Business Newspaper	(Literary sector) Maeil Business Newspaper and MBN Content <ul style="list-style-type: none"> <li>• Technical support (Desktop beta-version AI search tool to be inserted into Maeil Business Newspaper and MBN websites) provided in exchange for content without monetary payment</li> <li>• Advertising revenue generated through source-attributed search shared with the media outlets</li> </ul>



# Can GAI Infringe Copyright?

## ◀ Major Licensing Agreements in the United States ▶ ※ Approximately 40 cases<sup>1)</sup>

Date	AI Developer	Rights Holder / Publisher	Content of license
May 2025	Amazon	The New York Times	(Literary sector) NYT and The Athletic content Annual payments of approximately USD 20–25 million (about 1% of NYT’s 2024 total revenue)
Aug. 2024	Musical AI	Symphonic Distribution	(Music) Works of artists under Symphonic Revenue-sharing arrangements with rightsholders though terms vary by contract
Mar. 2023	NVIDIA	Getty Images	(Images/Audiovisuals) Getty Images Creative Library and iStock stock photo library Detailed contractual terms not disclosed

## ◀ Major Licensing Agreements in the European Union ▶

Contract Date	AI Developer	Rights Holder / Publisher	Content of license
Jan. 2025	Mistral AI (France)	AFP (French wire service)	(Literary sector) Multi-year content license; Approximately 2,000 articles per day integrated into the “Le Chat” chatbot Detailed contractual terms not disclosed
Dec. 2023	OpenAI (Germany)	Axel Springer (German media outlet)	(Literary sector) Three-year agreement; News content provided for use in services and model training Detailed contractual terms not disclosed

## ◀ Other Notable Examples ▶

Case 1	Universal Music Group (UMG)’s AI related partnerships and future plans <sup>2)</sup>
	<ul style="list-style-type: none"> <li>• Universal Music Group (UMG) entered into a wide range of AI-related agreements with major platforms such as YouTube, TikTok, and Meta, as well as emerging companies including BandLab and SoundLabs.</li> <li>• In Japan, UMG Japan is collaborating with a telecommunications company KDDI to develop AI-based music experiences connecting artists and fans.</li> <li>• UMG has also selected as partners and signed agreements with AI developers such as ProRata and KLAY that offer products designed to “both protect music and enhance its monetization.”</li> </ul>

1) Overall, reported agreements include approximately 26 cases involving literary works, 6 involving music, 9 involving images, 8 involving audiovisual works, 2 involving code, and 4 undisclosed cases.

2) Music Business Worldwide, Sir Lucian Grainge on UMG’s AI policy: “We will NOT license AI models that use an artist’s voice without their consent.” 13<sup>th</sup> Oct. 2025. <https://www.musicbusinessworldwide.com/sir-lucian-grainge-we-will-not-license-ai-models-that-use-artists-voices-without-consent/> (Last visited on 16th Oct, 2025)

<b>Case 2</b>	Licensing Agreement between Swedish Performing Rights Society (STIM) and an AI Developer	
<ul style="list-style-type: none"> <li>• This is the world's first licensing agreement by a collective management organization (CMO) for musical works.</li> <li>• STIM, representing approximately 100,000 artists, entered into a broad license agreement with an AI developer. However, the works subject to licensing are strictly limited to those for which the right holders had provided explicit prior consent (opt-in).</li> <li>• The license encompasses fair compensation for both AI training (input) and use of output. AI-generated music may be used in subscription/streaming services. However, separate agreements are required for some types of exploitation such as large-scale advertisements and films. The AI developer has the obligation to provide rights notices, limit scope of use, perform real-time monitoring, and pay royalties based on contribution.</li> </ul>		
<b>〈 Role of Each Party in the STIM AI Music Licensing Scheme 〉</b>		
Party	Main role	Basis for compensation
STIM (Swedish Performing Rights Society)	Manages rights and is the licensor	<ul style="list-style-type: none"> <li>• As a CMO, entered into a collective AI license agreement with AI developers</li> <li>• Established compensation standards and fair-trade norms</li> </ul>
Songfox (AI Music Developer)	Is the licensee and responsible for technology	<ul style="list-style-type: none"> <li>• Entered into a contract with STIM to legally use STIM-affiliated works for AI model training and generation</li> <li>• Provides AI-based music creation services to users.</li> </ul>
Sureel.ai (Third-party Tech Provider)	Provides transparency and attribution tracking technology	<ul style="list-style-type: none"> <li>• Provides technical infrastructure to track and quantify influence (attribution) of Songfox's AI outputs on STIM works in real-time</li> </ul>
<b>〈 Stage-by-Stage Structure of STIM AI Music Licensing 〉<sup>3)</sup></b>		
Stage	Details	Basis for compensation
(1) Training	Only songs for which rightsholders had opted-in used in training	Designed to generate compensation upon utilization for training
(2) Provision of service	Output generated by the trained model and consumed by user	Royalty calculated based on amount of use and output ratio
(3) Use of generated output	When output streamed, sold or otherwise used, the system remunerates	Royalty distributed based on downstream consumption

3) Korea Copyright Commission, "A Study on the Basic Structure and Technical System of STIM's AI License for Music," *Copyright Issue Trend* Vol. 69, 2025. 11-1, p. 10, [Figure] Recited.



# Can GAI Infringe Copyright?

## 5. Trend in Major Countries

- Although prior licensing needs to be attained for use of works, there are also limitations and exceptions to copyright that allow the use of works without prior permission under certain conditions.
- In the context of AI training, discussions primarily center on fair use and the text and data mining (TDM) exceptions.

### A. Fair Use

- The fair use provision is a flexible and broad limitation on economic rights that applies broadly to various situations — such as changes in the technological or usage environment — rather than being limited to specific cases like reproduction for educational purposes or non-profit performances.
- Adopted by countries such as the United States (1976) and South Korea (2011), it is recognized for its ability to respond flexibly to new types of copyright exploitation.

### B. TDM Exception

- The TDM exception is a specific limitation on economic rights that is applied to text and data mining (TDM)\*.

\* TDM refers to a technology that analyzes big data to identify information with value, such as statistical rules and trends.

- Adopted by the EU (2019) and Japan (2018), the provision generally permits training if explicit statutory requirements\* are met.

\* These requirements vary by country.

### 〈 Summary of Fair Use and TDM Exception in Major Countries 〉

Category	Country	Legal Basis	Commercial Use	Opt-out*	Characteristics
Fair Use	U.S.	Copyright Act Article 107	Possible, on case-by-case basis	×	Interpretation centered on case-law
	S. Korea	Copyright Act Article 35-5	Possible, on case-by-case basis	×	Interpretation centered on case-law

Category	Country	Legal Basis	Commercial Use	Opt-out*	Characteristics
TDM exception	EU	DSM <sup>4)</sup> Directive Article 3	×	×	Dual structure based on user/purpose
		DSM Directive Article 4	○	○	
	UK	CDPA <sup>5)</sup> Article 29A	×	×	Limited to non-commercial research
	Japan	Copyright Act Article 30-4	○	×	Permitted regardless of purpose (commercial & non-commercial) or type of user
Hybrid	Singapore	Copyright Act Articles 190, 191, 244	○	△ <sup>6)</sup>	Permitted regardless of purpose (commercial & non-commercial) or type of user

\* Opt-out, as in EU's DSM Directive Article 4, refers to a rightsholder explicitly reserving their rights against TDM exception for works made available to the public online, using machine-readable means or other appropriate methods.

### C. Constraints and Ways Forward

- A key constraint to using either the fair use provision or the TDM exception for GAI training is that it must be contested in court. In fact, there are several copyright-related litigations between rightsholders and AI developers regarding the use of training data globally, in jurisdictions with either the fair use or the TDM exception framework.
- In one lawsuit involving GAI training in the U.S., a court recently ruled\* that training on legally purchased books could constitute fair use whereas a German court claimed\*\* that the TDM exception did not apply to AI training.

\* For further reference, see *Andrea Bartz, Charles Graeber and Kirk Wallace Johnson v. Anthropic PBC*, No. 3:24-cv-05417, *Richard Kadrey, et al v. Meta Platforms, Inc*, No. 3:23-cv-03417-VC, *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc*. No. 1:20-cv-613-SB (D. Del. Feb. 11, 2025)

4) Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

5) Copyright, Designs and Patents Act 1988

6) Singapore doesn't have an explicit legal provision on opt-out. However, a rightsholder maintains the right to control access to works through technical protection measures, which is considered to have a similar effect as an opt-out mechanism.



## Can GAI Infringe Copyright?

\*\* For further reference, see *GEMA v. OpenAI* (LG München I, Urteil vom 11.11.2025, 42 O 14139/24)

- In the U.S., a jurisdiction with the fair use defense, the Copyright Office issued a report *Copyright and Artificial Intelligence Part 3: Generative AI Training* (May 2025) to review whether the fair use<sup>7)</sup> defense could be applied to GAI training.

**Reference** Summary of "Conclusion," p.107, *Copyright and Artificial Intelligence, Part 3: Generative AI Training*

- The fair use doctrine has been flexible enough to accommodate past technological revolutions and can continue to do so regarding artificial intelligence.
- Several stages in the development of generative AI involve using copyrighted works in ways that encroach on the owners' exclusive rights. Whether the fair use defense can also be applied, considering relevant case-laws, has been reviewed.
- Copyrighted works can be used in AI training in various ways and are likely to be transformative. Whether such exploitation constitute fair use, however, will depend on what works were used, from what source, for what purpose, with what controls and how all of these factors affect the market.
- When a model is used for analytical or research purposes, the outputs are unlikely to substitute expressive works used in training. However, using huge amounts of works to produce expressive content that competes with original works in existing markets goes beyond what can be recognized as fair use.

- The EU, which adopted the TDM exception, published the *General-Purpose AI Code of Practice* in July 2025 where it discussed the

7) U.S. Copyright Act §107 · Limitations on exclusive rights: Fair use

Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—

- (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- (4) the effect of the use upon the potential market for or value of the copyrighted work.

The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.

criteria for the TDM exception while Japan reviewed whether its TDM exception could be applied to AI in the *Views on AI and Copyright* (March 2024).

Reference	<i>General Purpose AI Code of Practice</i>
	<ul style="list-style-type: none"><li>• The EU's <i>General-Purpose AI Code of Practice</i> provides criteria for fulfilling copyright compliance obligations regarding Article 53 of the EU AI Act.<sup>8)</sup></li><li>• <b>Measure 1: Establish and maintain a copyright policy.</b> Signatories commit to develop and regularly update a robust copyright policy, which must be in the form of a single document.</li><li>• <b>Measure 2: Guarantee lawful access via web crawling.</b> Signatories must ensure that data collected via web crawling for TDM and AI training is lawfully accessed by:<ol style="list-style-type: none"><li>(1) Prohibiting circumvention of technological protection measures (including technological or contractual limitation to control access, such as in subscription models or paywalls);</li><li>(2) Excluding from web-crawling websites known to repetitively infringe copyright.</li></ol></li><li>• <b>Measure 3: Ensure web-crawling can recognize any indication of reserved rights.</b> Machine-readable rights signals like robots.txt and other opt-out indications must be recognizable and be complied to. Rights holders may resort to various other methods other than the robots exclusion protocol to express reserved rights, based on DSM Article 4.3.<ul style="list-style-type: none"><li>* i.e., Signatories must consider explicit intentions of the rightsholder, not just the adoption of machine-readable rights indication.</li></ul></li><li>• <b>Measure 4: Alleviate risk of infringing output.</b> Appropriate and proportionate technical safeguards must be adopted to prevent generation of infringing output. Also, user policy, terms and conditions, or any other similar document must clearly prohibit use of AI in ways that infringe copyright.</li></ul>

8) EU AI Act Article 53: Obligations for Providers of General-Purpose AI Models

1. Providers of general-purpose AI models shall:

(...)

(c) put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;

(d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.



## Can GAI Infringe Copyright?

**Reference** *Views on AI and Copyright*

- Japan published its *Views on AI and Copyright* explaining how its current Copyright Act should be interpreted and provides concrete guidance, centered on Article 30-4<sup>9)</sup>, on which acts constitute copyright infringement in the use of GAI.
- **Article 30-4 of the Copyright Act:** For “case in which it is not a person’s purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work” – i.e., for non-enjoyment purposes – the work may be used without the copyright holder’s consent to the extent necessary, provided that such use does not unreasonably prejudice\* the legitimate interests of the copyright holder.
  - \* (1) The rightsholder or website operator has implemented technical measures to prevent reproduction for AI training; (2) It is reasonably foreseeable that a database work, organized in a form suitable for information analysis, including data on the relevant website, is likely to be sold in the future; (3) Technical measures in (1) are circumvented to carry out reproduction or other acts for AI training.
- **Training AI on specific works or artistic styles:** Training the AI specifically on works of a particular author so that the main purpose seems to be to generate a part or all of a work’s creative expressions is indeed for enjoyment. This line of reasoning assumes the adoption of technologies such as LoRA (Low-Rank Adaptation)\*. The final determination must be made case by case through judicial review.
  - \* LoRA (Low-Rank Adaptation): Technology involving additionally training an AI model on multiple images to make it generate outputs that resemble a specific artwork or photographic style.
- **Pirated works:** Training data must be collected in ways that do not encourage infringement of copyright and specific obligations need to be imposed on AI developers so that they refrain from collecting data from pirated sources.
- **News article databases:** An additional requirement is imposed that there must be both a future plan to sell the database and a demonstrable history of past database sales.

9) Japan’s Copyright Act Article 30-4 (Exploitation without the Purpose of Enjoying the Thoughts or Sentiments Expressed in a Work)  
It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person’s purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:  
(i) if it is done for use in testing to develop or put into practical use technology that is connected with the recording of sounds or visuals of a work or other such exploitation;  
(ii) if it is done for use 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; the same applies in Article 47-5, paragraph (1), item (ii));  
(iii) if it is exploited in the course of computer data processing or otherwise exploited in a way that does not involve what is expressed in the work being perceived by the human senses (for works of computer programming, such exploitation excludes the execution of the work on a computer), beyond as set forth in the preceding two items.

- 
- AI developers argue that it is necessary either to introduce a TDM exception or to provide clear standards for fair use, while rightsholders maintain that the use of copyrighted works must be accompanied by reasonable remuneration. Accordingly, there is a need to clearly articulate interpretive guidance for determining fair use in the context of GAI training, and to establish standards that enable the fair use provision to function effectively within the evolving technological environment.



---

# What is Fair Use

# Guide on Applicability of the Fair Use Doctrine

to Training of Generative AI Models



1. Overview of the Fair Use Provision
2. Factor 1 – The purpose and character of the use
3. Factor 2 – The Type and purpose of the work
4. Factor 3 – The Amount and Substantiality of the Portion Used
5. Factor 4 – The Effect on the Current Market or Value,  
or Potential Market or Value of the Work
6. Summary of Fair Use Considerations for Each Factor
7. Case Studies Related to Fair Use



# What is Fair Use

## 1. Overview of the Fair Use Provision

### A. Purpose of the Fair Use Provision

#### Copyright Act Article 35-5 (Fair Use of Works, etc.)

- (1) Except as provided in Articles 23 through 35-2 and 101-3 through 101-5, where a person does not unreasonably prejudice an author's legitimate interest without conflicting with the normal exploitation of works, he or she may use such works. (Amended by Act No. 14083, Mar. 22, 2016)
- (2) In determining whether an act of using works, etc. falls under paragraph (1), the following shall be considered: (Amended by Act No. 14083, Mar. 22, 2016)
  1. Purposes and characters of use including whether such use is for or not-for-profit;
  2. Types and natures of works, etc.;
  3. Amount and substantiality of portion used in relation to the whole works, etc.;
  4. Effect of the use of works, etc. on the current or potential market for or value of such work etc.

- Article 35-5 of the Copyright Act is a general clause on fair use, allowing courts to flexibly determine the legality of specific uses even when they do not fall under the individual statutory limitations and exceptions (e.g., use for educational purposes, use by libraries or non-profit performances) to economic rights. It was incorporated into the Copyright Act in December 2011 to respond to various circumstances arising from digitalization of works and changes in the distribution environment that may not be readily captured only by the existing exceptions and limitations that were enumerated in the previous version of the Act. Furthermore, it was introduced to not only adequately protect copyright but also to promote use of works in a fair manner so that use of works could be vitalized and a more sound and robust copyright ecosystem could evolve. (KCC, MCST, *Explanation of the Revised Copyright Act* (2012), p 32~33)

### B. Content of the Fair Use Provision

- Paragraph 1 of Article 35-5 stipulates that “where a person (1) does not unreasonably prejudice an author's legitimate interest (2) without conflicting with the normal exploitation of works, he or she may use such works.”

- “Legitimate” in (1) refers to the fair, reasonable, and socially acceptable claim that the author has over his or her work while “interest” refers to the economic gains an author can earn from the work. In other words, an “author’s legitimate interest” refers to the socially acceptable level of financial and material gains an author can earn from their work.  
\* The same logic applies to related rights (Article 87) and rights of database producers (Article 94).
- “Normal exploitation of works” in (2) refers to the method the work is expected to be used in the market, meaning that any form of exploitation that competes against the exploitation of the work in its normal market is not allowed.
- Paragraph 1 of Article 35–5 is also the legislative form of the three-step test established in international treaties. For a fair use defense to be accepted according to Paragraph 2, the use has to fulfill the criteria stipulated in Paragraph 1.

**Reference** The Three-Step Test in international treaties

- The Three-Step Test articulated in copyright-related treaties establishes limits on the discretionary power of parties when setting exceptions or limitations to authors’ rights. It was first articulated in Article 9(2) of the Berne Convention regarding reproduction rights and subsequently expanded to cover all economic rights under Article 13 of the TRIPS Agreement.
- The enactment and interpretation of copyright limitation provisions in the domestic law must comply with the Three-Step Test as mandated by the international treaties.
- According to the Three-Step Test, exceptions and limitations must (1) apply only to certain special cases, (2) not interfere with normal exploitation of the work, and (3) not unfairly infringe on the rightsholder’s legitimate interests.
- Requirements of (1): Exceptions and limitations to copyright must not be vague or broad; they must apply only to specific and restricted circumstances.
- Requirements of (2): They must not interfere with the actual or potential economic benefits the rightsholder derives from the work.
- Requirements of (3): They must not excessively infringe upon the rightsholder’s legal economic interests.

- Paragraph 2 of Article 35–5 of the Copyright Act stipulates that when determining whether a specific use of a work (1) unreasonably prejudices



# What is Fair Use

the author's legitimate interest, and (2) conflicts with normal exploitation, the following four key factors must be comprehensively considered.

- ▲ Factor 1: The purpose and character of the use
- ▲ Factor 2: The type and purpose of the copyrighted work
- ▲ Factor 3: The amount and substantiality of the portion used in relation to whole work
- ▲ Factor 4: The effect of the use on the current market or value, or potential market or value of the work

## C. Determining fair use

- Determining whether a particular use of a copyrighted work constitutes fair use requires a holistic assessment of all four factors. Even if one or two factors appear unfavorable, conclusion must be reached by weighing all factors, including the purpose and character of the use, the type and purpose of the work, the amount and substantiality of the portion used, and the effect on the work's market, when determining whether a work used in GAI training was fair use.
- While the importance of the four factors\* of fair use may vary, no single factor is dispositive. The weight assigned to each factor depends on the specific facts and circumstances of the individual case.
  - \* This applies also to the specific sub-factors under each of the four factors.
- The Supreme Court of Korea has ruled that courts may consider additional factors beyond the four, such as the specific context or manner of the use. (Supreme Court Decision 2021Da272001, July 11, 2024)

**Case law** Supreme Court Decision 2021Da272001, July 11, 2024

- "(...) (T)he factors listed illustratively in Paragraph 2 must be considered comprehensively. Even factors not explicitly listed, such as the context or manner of use, may be taken into account during the determination."

\* Disclaimer: The following commentary on AI training is based on the Copyright Act and legal precedents current as of January 2026. It is not an official legal interpretation by either the Ministry of Culture, Sports and Tourism (MCST) or the Korea Copyright Commission (KCC). Contents are subject to change based on future legislative amendments, new precedents, or technological advancements.

## 2. (Factor 1) The purpose and character of the use

- The first factor evaluates the purpose and character of the use of the copyrighted work. Specific elements that are considered include whether the use is transformative, whether it is for commercial or non-profit purposes, and the specific context or method\* of the use.  
\* The Guide has interpreted the context or method of use as part of factor 1.
- Regarding factor 1, the Supreme Court had considered whether the use had gone beyond simply replacing the original to being transformative enough by including distinct new expressions, meanings or messages, and whether the use had been for public interest or non-profit purposes.

**Case law** Supreme Court Decision 2021Da272001, July 11, 2024

“With regard to “purpose and character of the use (Sub-paragraph 1),” whether the use goes beyond merely substituting for the original work to provide new expression, meaning, or message, whether the use serves a purpose and character that are distinct from the original, whether the degree of transformation exceeds what is typically required to create a derivative work, and whether the use serves the public interest or is non-profit in nature may all be considered.”

\* Under Korea’s copyright law, “public interest” is a relative concept, which is not determined solely by industrial development or technological innovation. Rather, it is assessed by balancing the personal interests of the rightsholder with the social benefits achieved through fair use.

### A. Transformative Use

#### (1) Meaning

- Transformative use refers to whether the use merely substitutes the original work or adds new value by altering it with new expressions, meanings, or messages. If the purpose and character of the use are identical to those of the original work, it may weigh unfavorably against fair use.
- While transformative use is a critical factor in establishing fair use, it is not automatically equated with the public interest. In fact, the Supreme Court (Case 2021Da272001) treated transformative use and the public interest involved in the use as distinct considerations. This



# What is Fair Use

means that just because a court accepts the public interest defense, that does not mean it also considers the output to be transformative, and vice versa.

## (2) Implications for AI training

- If a work was used to train an AI model that generates outputs unrelated to the purpose or character of the original work, such use may be viewed in favor of factor 1. This is particularly true if the source expressions, either in part or in whole, are not reflected in the GAI output and the output serves a different purpose.

\* For further reference, see *Andrea Bartz, Charles Graeber and Kirk Wallace Johnson v. Anthropic PBC*, No. 3:24-cv-05417, *Richard Kadrey, et al v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC, *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.* No. 1:20-cv-613-SB (D. Del. Feb. 11, 2025).

- Implementing technical measures to reject prompts that generate the reproduction or imitation of specific copyrighted works will also be considered favorably. An AI system that fundamentally blocks requests to reproduce or imitate the expressions contained in a certain work significantly reduces the likelihood of substituting the market of the original work.
- A court may also view that using certain works to train a GAI model so that it creates new value or new utility, thereby significantly enhancing public interest, is in favor of factor 1, provided there is little to no infringement on the existing market or the rightsholder's economic interests.
- On the other hand, excessively training on a specific work increases the likelihood that the GAI output will share a similar purpose or character with the original, which weighs against factor 1. If the expressions and creative elements of the source work are reflected in the GAI output, it is more likely to conflict with the “normal exploitation” of the work and harm the rightsholder's economic interests.

- If the use does not significantly contribute to the public interest and also fails to create new value or utility while simultaneously infringing on existing markets or economic interests, it will be viewed unfavorably regarding factor 1. In addition, if the GAI output simply mimics the source work or merely functions as an economic substitute by reconstructing the work, it is unlikely to satisfy factor 1.

## B. Commercial Purpose

### (1) Meaning

- Using a work for a commercial purpose does not necessarily refer only to obtaining direct monetary compensation; it also includes cases where the use indirectly makes economic contribution or gives financial benefit to a company.
- Using a copyrighted work for a commercial purpose increases the likelihood of encroaching upon the rightsholder’s economic interests, which may be disadvantageous compared to non-commercial use when a court is determining factor 1 in a fair use defense.
- Korean courts have, in some cases, held that use for commercial purposes is permitted within a significantly narrower scope than use for non-commercial purposes. In other cases, however, courts — considering the legislative history related to the purpose and character of use — have ruled that the importance of “commerciality” is relatively limited.

**Case law** Seoul Central District Court Decision 2021Gahap512773, January 14, 2022

“Under the Copyright Act, a work may be used where such use does not conflict with the normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author. In determining whether these conditions are met, the following factors must be considered: the purpose and character of the use; the type and purpose of the copyrighted work; the proportion and significance of the portion used in relation to the whole; and the effect of the use on the current market or value, or potential market or value of the work (Article 35-5 of the Copyright Act). Although not all forms of exploitation have to be non-commercial for fair use to be accepted, use for commercial purposes is generally permitted within a considerably narrower range than use for non-commercial purposes.”



# What is Fair Use

**Case law** Patent Court Decision 2024Na10249, January 23, 2025

“Paragraph 2(1) of Article 35–3 of the former Copyright Act (prior to its amendment by Act No. 14083 on March 22, 2016) originally stipulated “the purpose and character of the use, such as whether it is commercial or non-commercial” as a factor to be considered in determining fair use. However, following the amendment to the current Copyright Act, the phrase “such as whether it is commercial or non-commercial” was deleted from the corresponding provision, while maintaining the same overall intent. The reason for this deletion was, “The fair use provision should serve the important purpose of promoting cultural development and related industries by facilitating the use of works across diverse fields. However, the stated purpose of use and factors that have to be considered were overly restrictive, making it difficult to achieve that purpose, and therefore required revision.” This indicates that the amendment reflects the legislature’s intent to accord reduced weight to the commercial nature of use when determining whether the use of a copyrighted work constitutes fair use.”

## (2) Implications for AI training

- The fact that an output serving a public purpose — such as benefiting social safety, welfare, or the environment — is generated and utilized for commercial purposes does not, by itself, weigh against a favorable assessment under the first factor. For example, in the case of GAI training conducted by a university, a public research institute, or a non-profit organization, the use is more likely to be recognized as non-commercial. However, if a non-profit research organization conducts training with support from a for-profit business, and the training is closely related to that business’ commercial activities, that may be deemed commercial.
- As such, non-commercial and commercial purposes may be intertwined, in which case, the legal status or financial structure of the user may not be considered as the sole determinant in factor 1. For instance, an industry-academia collaborative research conducted jointly by a university and a business may be non-commercial during the initial research stage. However, if the GAI model is developed and then used to provide commercial services at a later stage, a more comprehensive assessment of various sub-factors will have to be considered – for example, the purpose during the initial research stage, scope of copyrighted works used, possibility of

commercialization, transparency while using works, and the structure of the finances. Conversely, even a for-profit enterprise may be recognized as engaging in non-commercial training if it trains an AI model that generates outputs unrelated to its business activities.

- In short, just because a GAI system was developed or trained for commercial purposes does not automatically preclude a finding of fair use. Even where a commercial purpose exists, if the technology creates new value and serves public interest, the use may be assessed favorably under the first factor.

\* For further reference, see *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2nd Cir. 2015)

- On the other hand, if there is direct compensation from the users for the services rendered by the developer or the latter gains indirect economic benefits through advertising or promotional activities, these circumstances may be considered unfavorable under the first factor. In particular, if a developer trains a GAI model to provide commercial services, and the use of copyrighted works directly conflicts with the rightsholder's economic interests, the use is highly likely to be regarded as commercial.

### C. Context or Method of Use

#### (1) Meaning

- The Supreme Court has held that, in determining fair use, courts may consider not only the four statutory factors in Paragraph 2 of Article 35-5 of the Copyright Act, but also additional sub-factors such as the context and method of the use. In fact, in a case where a user had posted online freely downloadable high school and university entrance exam questions that included excerpts from copyrighted works, without any copy-prevention measures to prevent infringement of copyright, the court ruled that such exploitation did not constitute using a work in legitimate ways.



# What is Fair Use

**Case law** Supreme Court Decision 2021Da272001, July 11, 2024

“In determining whether a particular use of a copyrighted work falls within a case that does not unreasonably prejudice an author’s legitimate interest without conflicting with the normal exploitation of works, as provided under Article 35-3(1) of the former Copyright Act, an assessment must comprehensively consider the factors illustratively enumerated in Paragraph 2, which are purposes and characters of use (sub-paragraph 1), type and nature of works (sub-paragraph 2), amount and substantiality of portion used in relation to the whole work (sub-paragraph 3), and the effect of the use of works on the current or potential market for or value of such work (sub-paragraph 4). Even factors not explicitly listed, such as the context or method of use, may be taken into account during the determination.”

(…)

“However, the defendant did not take any necessary measures such as those to prevent copying and infringement of copyright and other rights, as well as making available to the public online materials relevant to this case for a long period, without permission from the rightsholder.”

“The defendant’s method of using the works cannot be deemed legitimate.”

- The Supreme Court’s stance is that in addition to whether measures were taken to prevent illegal copying, whether the work itself was a pirated version or whether the route of accessing the work was illegal can all be considered when determining factor 1 of fair use.

\* If a copyright claim is raised, whether the user had made effort to attain permission from the rightsholder may also be a sub-factor during an overall analysis of the circumstances.

## (2) Implications for AI training

- Training data for GAI may be obtained through various means, such as using automated web crawling programs, selectively collecting only necessary materials by web scraping, purchasing datasets, or using open or public datasets.

\* When assessing factor 1, it may be relevant whether rightsholders have taken measures to restrict web crawling or scraping.

- Even where a work is technically accessible, using it for GAI training without permission from the rightsholder may be viewed as exceeding the scope of authorization, which can weigh unfavorably under factor 1.

- Furthermore, collecting access-restricted works (such as works behind log-in requirements or paywalls) without permission may also weigh unfavorably under factor 1, as well as circumventing technical protection measures, which is a violation of another provision in the Copyright Act (Article 104-2) and is thus unfavorable. This applies especially where works subject to explicit access restrictions or robot exclusion standards (robots.txt) are collected without permission.
- Also, collecting works from a website in ways that infringe the applicable Terms and Conditions, or accessing a work in ways that are not allowed could potentially infringe copyright and thus also weigh against factor 1.
- If the datasets were purchased from or provided by a third party, the user needs to confirm whether the scope of use clearly covered AI training, and whether the third party lawfully held rights to the dataset.
- Using works published by government bodies, public institutions, or research organizations does not automatically weigh in favor of the first factor merely because the works are publicly available or serve public interests.
- Finally, even when a work seems to have been used transformatively, if the use involved unlawful reproduction or improper access, such circumstances may still weigh against a favorable finding under factor 1.

### 3. (Factor 2) The Type and purpose of the work

#### A. Meaning

- Assessing the type and purpose of a copyrighted work refers to considerations such as whether the original work is factual or functional, or literary or artistic, and whether it has been made public.
- Factual or functional works are those intended primarily to convey facts or information, or to perform a functional role, such as computer code. By contrast, literary or artistic works are works created primarily for the enjoyment of the expressive content itself. News articles may



## What is Fair Use

fall within the category of factual or informational works. However, an article that contains a critique or interpretation of social phenomena based on a journalist's thoughts or emotions is not regarded as a work whose purpose is purely factual or informational.

- The reason why whether the work has been made public is assessed is to take into consideration the balance between protecting the rights and interests of rightsholders on one hand and the right for the public to access information on the other. Using a work that has not been made public is generally weighed unfavorably under factor 2. However, this does not mean that using a work that has already been made public and is accessible by the public automatically weighs in favor of fair use. In fact, even if a work has been made public, if it is distributed through a restricted access structure — such as paid content or commercial services — the work remains within a sphere where the rightsholder can meaningfully control its use. In such cases, the rightsholder's control over use should also be prioritized.
- The Supreme Court, in applying factor 2, has indeed considered whether the original work was factual or informational in nature, and whether it was published or been made public.

**Case law** Supreme Court Decision 2021Da272001, July 11, 2024

“Regarding the type and nature of a work (sub-paragraph 2), elements such as whether the original work is factual or informational in nature, and whether it has been published or been made public are considered...”

- Use of factual, informational, or functional works like computer codes that serve a functional role tends to weigh in favor of factor 2 whereas use of works with a high degree of creativity may weigh against.
- In applying factor 2, works created primarily for enjoyment of the work itself — such as novels, films, visual art, and music — are generally treated less favorably in determining fair use than works created mainly for fact-finding or information delivery. In addition, use of works that have not been made public may be viewed less favorably than works that have already been published or released.

---

## B. Implications for AI training

---

- When factual or informational works, including materials intended to convey information, are used for GAI training, a court may consider it to be relatively favorable under factor 2. Because factual works contain only a limited amount of creative expression, their use is generally viewed more favorably under factor 2 than that of literary or artistic works, which are authored primarily for enjoyment of their expressions. This is why, conversely, literary or artistic works may be assessed unfavorably when used in AI training. Since their creative expressions are their core value, a GAI model that had trained on such expressions is likely to generate identical or similar expressive characteristics – hence the unfavorable assessment of factor 2.
- In the context of GAI training, freely accessible online works and works provided under restricted or paid access may be treated differently, as the nature of their use and the surrounding social expectations are not the same. However, the mere fact that a work is publicly available does not, by itself, support a finding of fair use. Aspects such as the existence of robots.txt restrictions, as well as the scope and scale of the intended use, may be weighed differently.
- Works that have not been made public or content made available only under restricted access — such as subscription services or paid databases — may be viewed less favorably, as their use carries a greater risk of directly affecting the rightsholder’s economic interests.

## 4. (Factor 3) The Amount and Substantiality of the Portion Used

### A. Meaning

---

- The “amount and substantiality of the portion used” refers to an assessment of whether the AI developer used the copyrighted work only to the extent necessary for the purpose and character of the use under factor 1. The assessment is made from both a quantitative perspective (the amount used) and a qualitative perspective (importance of the part used). Because factor 3 is closely connected to the purpose and character of the use (factor 1) and the likelihood



# What is Fair Use

of the use substituting for the original work (factor 4), both the quantitative proportion and the qualitative importance of the portion used are examined together. In fact, the Supreme Court had said whether the part that was used was quantitatively or qualitatively important vis-à-vis the entire work and whether the user had ensured the use was limited to what was reasonably necessary to achieve the intended purpose were to be considered when determining fair use.

**Case law** Supreme Court Decision 2021Da272001, July 11, 2024

“When considering the amount and substantiality of the portion used, the assessment looks at whether, in relation to the original work as a whole, the portion used represents a small quantitative share or has limited qualitative importance, and whether the user relied on the work only to the extent necessary for the intended purpose.”

- From a quantitative perspective, an AI developer reproducing an entire copyrighted work may be viewed unfavorably under factor 3. Nonetheless, if reproducing the entire work was unavoidable or reasonably necessary in light of the purpose of the use, there is still possibility of a favorable assessment.

\* For example, in a case where Google cooperated with libraries to scan entire books, the court recognized fair use because only limited portions of the books were made publicly available, and full scanning was an unavoidable technical step to achieve the intended purpose. (For further details, refer to *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2nd Cir. 2015))

## B. Implications for AI training

- If an entire copyrighted work was used in GAI training, that, on its own, weighs unfavorably under factor 3. However, if the entire work had to be used for technical or other necessary reasons, the conclusion regarding fair use could change when all sub-factors are considered comprehensively. In fact, given the structure of AI models, a certain minimum amount of data may have to be inputted to enable effective training and performance improvement. In such a case, the fact that the entire work was used may not, by itself, be a sufficient basis for treating the use as unfavorable under factor 3.

- However, in the course of using a work, if its core expressive elements were directly reproduced, or the output was substantially identical to or substitutable for the original, and the use exceeded, quantitatively or qualitatively, what was reasonably necessary in light of the purpose and character of the use recognized as fair under factor 1, the use is unlikely to be assessed favorably.

## 5. (Factor 4) The Effect on the Current Market or Value, or Potential Market or Value of the Work

### A. Meaning

- Factor 4 is a criterium for examining whether the use of a copyrighted work affects the current market or value, or the potential market or value, of the original work.
- The **current market or value** refers to the situation where a market already exists in which users typically obtain permission from the rightsholder and pay royalty. The existence of such market indicates that the rightsholder has a normal and established means of monetization through licensing his or her work.
- The **potential market or value** refers to situations where a market, while not yet fully developed, has a reasonable likelihood of emerging in the future — where users would normally be expected to obtain permission from the rightsholder and pay royalty to use the work. In this context, “potential” refers to the probability of a particular type of use becoming widespread enough for such use to replace or reduce demand for the original work in the rightsholder’s existing market.
- When assessing the effect on the current market or value, or potential market or value, whether the use of a work directly competes with the original work or its derivative work, or whether the use is likely to substantially substitute for demand for the original work, either now or in the future, are important questions that need to be asked. In other words, the assessment of factor four should not be limited only to the markets the rightsholder has already entered. Markets that could be reasonably expected to develop also have to be taken into account. Even if a particular form of use does not currently constitute a clearly



## What is Fair Use

defined market, if there is a realistic possibility that the rightsholder could generate revenue through licensing in the future, or if such a market is reasonably foreseeable due to technological or industrial developments, that kind of use is also viewed as having the potential to be substituted under factor 4. Conversely, if the work is used in an area where the likelihood of a market forming is low, or where a strong public-interest is recognized, the risk of market substitution may be considered relatively low.

- Depending on the case, if the purpose and character of the use (factor 1) are assessed favorably, the analysis under factor 4 may focus on whether the use went beyond what was reasonably necessary to achieve that purpose and thus substitute for demand for the original work.
- In fact, in applying the fourth factor, the Supreme Court considered whether the use of a work had the risk of replacing or diminishing the demand for the current market of the original work or its derivative works, or of undermining the demand in a normal market that had a reasonable likelihood of developing in the future, and whether the use was likely to harm the market value of the original work to a significant degree.

**Case law** Supreme Court Decision 2021Da272001, July 11, 2024

“With respect to “the effect of the use of a copyrighted work on the current market or value, or potential market or value of the work” (sub-paragraph 4), the assessment shall focus on whether the use is unlikely, or only minimally likely, to substitute for demand in the current market for the original work or its derivative works, or substitute for demand in a normal market that has a reasonable likelihood of developing in the future, or otherwise undermine the market value of the original work.”

- In short, whether to recognize fair use based on factor 4 depends on specific sub-factors such as potential loss of sales for the original work, economic harm to the rightsholder, or reduction or elimination of opportunities for licensing the work.

---

## B. Implications for AI training

---

### (1) Potential Loss of Sales or Economic Harm

- The use of a copyrighted work that does not result in loss of sales or any other economic harm may be considered favorable under factor 4. In particular, if the outputs generated through GAI training are unlikely, or only minimally likely, to substitute for the economic value or market demand of the original work, or to undermine its market value, those outputs may weigh favorably. In addition, generating outputs that do not directly reproduce the expressive elements of the original work but instead have their own independent utility is less likely to form a competitive relationship with the original work and thus be regarded as favorable. Also, if the training does not directly render or recreate the original work, even commercial use may weigh in favor under factor 4 as long as the use does not directly affect demand in the market of the original work.
- Any loss of sales or economic harm to the work will likely be weighed against the user under factor 4. If the AI output directly or substantially reproduces the original's core expressions, that will likely be weighed against the user. In addition, an output substituting the original work or eroding its potential market demand will likely be weighed against the user.

### (2) Licensing Opportunities

- A licensing market for using works in GAI training already exists in some areas or may reasonably be expected to emerge or develop in the future, depending on the type of work, the training method, and the model's characteristics.
- If a certain exploitation clearly does not undermine licensing opportunities and thus has little or no potential to replace the market, this may weigh in favor under factor 4. If the training is limited to non-commercial, public-interest purposes (such as model research, validation or performance evaluation), and the use of works does not meaningfully interfere with the



## What is Fair Use

rightsholder's reasonable ability to enter the market or earn a revenue, that use may be seen positive under factor 4. Furthermore, using a work in ways that do not affect licensing opportunities and do not substitute for the market or intrude on the work's normal methods of use may also weigh in favor.

- On the other hand, using a work in ways that clearly undermine licensing opportunities, replace or weaken the market may weigh against the user.

**Case law** Supreme Court Decision 2021Da272001, July 11, 2024

"A market already existed in which anyone seeking to use the works managed in trust by the plaintiff as training materials obtained a license from the plaintiff and paid royalties. In addition, a further market — one in which parties obtain a license from the plaintiff and pay royalties to transmit past exam questions and other study materials with excerpt from those works — was reasonably likely to develop in the future as a normal market for the rights holder.

However, the defendant's posting activities in this case differed from those normal methods of use in that the defendant transmitted the works that exam questions were based on, without obtaining a license from the plaintiff. Because this created a substantial risk that demand in that market for the works would be substituted or that their market value would be harmed, the posting constituted an act that affected the current market or value, or the potential market or value, of the works in question."

- Repetitive and large-scale unauthorized training can lead to unlicensed use, which, when left to become widespread, can disrupt normal licensing practices and market norms, weaken a copyrighted work's market value, and effectively wipe out a rightsholder's ability to earn revenue. This can weigh against fair use under factor 4. An AI developer, by collecting and using works for training without permission and then exploiting the trained model in commercial services, may interfere with or damage the rightsholder's reasonably foreseeable opportunity to earn money through future licensing for similar kind of use. In addition to replacing the current market, this kind of exploitation may be viewed as encroaching on the work's normal area of use or undermining the formation of a legitimate licensing market and may therefore weigh against fair use under factor 4.
- That said, whether using a particular copyrighted work in training data

---

substitutes for demand for the existing work under factor 4 can vary depending on market conditions, including the nature of the training data.



# What is Fair Use

## 6. Summary of Fair Use Considerations for Each Factor

\* Applicability of fair use is determined by considering each factor below. However, no single factor decides the final outcome although the weight of each factor may vary.

### Factor 1: Purpose and character of the use

Sub-factor	Favorable for fair use
Transformative use	<ul style="list-style-type: none"><li>* <i>Transformative use is the criteria for determining whether a use of a work was merely substituting for the original or whether it had transformed the original to have new expressions, meanings and messages, thus creating new value.</i></li><li>• Training a model to generate output whose purpose or characteristic is unrelated to the copyrighted work used in training</li><li>• Installing technical measures or otherwise making effort so that AI model rejects prompts requesting it to generate output identical or similar to the work used in training</li><li>• Clearly contributing to public interest by training the AI model to create new value or new purpose, imposing very little to no harm to the existing market for the copyrighted work or economic interests of the rightsholder</li></ul>
Commercial purpose	<ul style="list-style-type: none"><li>• Training a model to generate output for benefit of the public, such as for social safety, welfare and environment</li><li>• A for-profit enterprise training a model for non-commercial, for-public purposes, unrelated to its business activities</li></ul>
Context or method of use	<ul style="list-style-type: none"><li>• N/A</li></ul>

### Factor 2: Type and purpose of the work

Sub-factor	Favorable for fair use
Type	<ul style="list-style-type: none"><li>• Using factual or functional works intended merely to deliver facts or information</li></ul>
Made public or not	<ul style="list-style-type: none"><li>• N/A</li></ul>

Unfavorable for fair use	Supreme Court's basis (Case2021Da272001)
<ul style="list-style-type: none"> <li>• Training the model excessively on a particular work so that the model generates output whose purpose or characteristic is identical or similar to the original</li> </ul>	<ul style="list-style-type: none"> <li>• "With regard to "purpose and character of the use (Sub-paragraph 1)" whether the use goes beyond merely substituting for the original work to provide new expression, meaning, or message, whether the use serves a distinct purpose and character from the original, whether the degree of transformation exceeds what is typically required to create a derivative work, and whether the use serves the public interest or is non-profit in nature may all be considered."</li> </ul>
<ul style="list-style-type: none"> <li>• Creating no new value nor purpose, not making any significant contribution to public interest, nor harming the existing market for the copyrighted work or economic interests of the rightsholder</li> </ul>	
<ul style="list-style-type: none"> <li>• Seeking to receive direct compensation from users for the service rendered or indirect compensation through advertising or publicity</li> <li>• A non-profit research organization training a model to be used by a for-profit enterprise for business purposes under sponsorship from the enterprise</li> </ul>	
<ul style="list-style-type: none"> <li>• Accessing the work possible but GAI training forbidden</li> <li>• Using works only provided under certain control, such subscription or fee-based content or database</li> </ul>	

Unfavorable for fair use	Supreme Court's basis (Case2021Da272001)
<ul style="list-style-type: none"> <li>• Using literary or artistic works intended for enjoyment of the work itself</li> </ul>	<ul style="list-style-type: none"> <li>• "(E)lements such as whether the original work is factual or informational in nature, and whether it has been published or been made public are considered..."</li> </ul>
<ul style="list-style-type: none"> <li>• Exploiting works that have not been made public</li> </ul>	



# What is Fair Use

## Factor 3: Amount and substantiality of the portion used

Sub-factor	Favorable for fair use	Unfavorable for fair use
Amount of work used (quantitative portion)	<ul style="list-style-type: none"><li>Using an entire work to train a GAI model or to generate services may be unfavorable for fair use solely under factor 3, however other considerations may be taken into account if the entire use was technically unavoidable and necessary.</li></ul>	
Substantial value (qualitative importance)		

## Factor 4: Effect on market\* or value of the work

Sub-factor	Favorable for fair use	Unfavorable for fair use
Harm to sales and other economic loss	<ul style="list-style-type: none"><li>Using a work in ways that pose no harm on sales and other economic interests of the work</li></ul>	<ul style="list-style-type: none"><li>Using a work in ways that harm sales and other economic interests of the work</li></ul>
Loss of licensing opportunities	<ul style="list-style-type: none"><li>Using the work in ways that don't clearly damage licensing opportunities and thus cannot be substitute for or weaken existing market</li></ul>	<ul style="list-style-type: none"><li>Clearly encroaching upon licensing opportunities and thus likely to substitute for or weaken existing market</li></ul>

\* The concept of the "market" encompasses not only the rightsholder's existing market, but also potential or future markets that the rightsholder could reasonably be expected to enter. Courts may further consider broader market impacts, including a decline in the overall market for works of the same category and the loss of employment opportunities for rightsholders.

Supreme Court's basis (Case2021Da272001)

- "(T)he assessment looks at whether, in relation to the original work as a whole, the portion used represents a small quantitative share or has limited qualitative importance, and whether the user relied on the work only to the extent necessary for the intended purpose."

Supreme Court's basis (Case2021Da272001)

- "(T)he assessment shall focus on whether the use is unlikely, or only minimally likely, to substitute for demand in the current market for the original work or its derivative works, or substitute for demand in a normal market that has a reasonable likelihood of developing in the future, or otherwise undermine the market value of the original work."



# What is Fair Use

## 7. Case Studies Related to Fair Use

\* The following cases are examples illustrating each of the four fair use factors discussed in this *Guide*. They are intended to demonstrate that the determination of fair use is not based solely on whether a particular factor weighs for or against fair use, but rather on a comprehensive consideration of all relevant circumstances.

- Whether fair use is accepted is determined by the court’s holistic assessment of the specific facts of each individual case. Accordingly, actual judicial decisions may differ from the cases illustrated here. Even within similar fields, whether the fair use defense is accepted or not can vary depending on the particular factual circumstances of each case.

\* “Case Studies Related to Fair Use” illustrated below are references based on the Copyright Act and judicial precedents as of February 2026. They are not official interpretations on the case by the Ministry of Culture, Sports and Tourism nor by the Korea Copyright Commission, and their contents are subject to change in response to future amendments to laws, developments in case law, or technological advancements.

### A. General-Purpose AI Foundation Models

- The cases presented below are intended to illustrate that, even within the same or similar categories of cases, whether fair use applies may vary depending on the specific factual circumstances and the context in which the copyrighted works were used.

\* In the following cases, certain facts have been simplified for ease of understanding. The explanations are solely for reference purposes to explain the general criteria for determining fair use and do not present a determination or conclusion regarding any specific case.

- A for-profit enterprise developed a general-purpose AI foundation model and trained the model using publicly available online posts\* and lawfully purchased books\*\*.

\* Excluding content subject to access restrictions (e.g., login or paid access), posts for which the uploader has explicitly forbid AI training through means such as robots.txt, and unlawfully uploaded or pirated content

\*\* The assumption here is that the books were lawfully purchased through ordinary commercial channels and that no separate license was obtained specifically for the purpose of training AI.

\* Note: A general-purpose AI foundation model is a base model designed for broad applicability. Its training dataset typically consists of a mixture of copyrighted and non-copyrighted materials. For the purposes of this *Guide*, this analysis of fair use is limited to the use of copyrighted works only.

## 1. Purpose and Character of the Use

### (1) Likely to weigh in favor of fair use

The developer had used publicly available online posts and lawfully purchased books to train the AI, and the use of works was not intended to reproduce or substitute the expressions of specific posts or books, but rather to enable general-purpose conversational and generative capabilities. Also, the developer had made efforts to implement technical measures to reject prompts that generate outputs identical or substantially similar to the works used in the training.

\* An overseas case where the judge accepted that training on lawfully purchased books without obtaining a separate license may constitute fair use can be found in *Andrea Bartz, Charles Graeber and Kirk Wallace Johnson v. Anthropic PBC*, No. 3:24-cv-05417.

\* However, it should be noted that the purchase of a book confers ownership only of the physical copy and does not include automatically authorize reproduction of the work. The U.S. Copyright Act Section 202 explicitly distinguishes ownership of a material object from ownership of copyright. Accordingly, future judicial decisions may reach different conclusions.

### (2) Likely to weigh against fair use

The developer had trained an AI model or provided services that generate outputs that were identical or substantially similar to the sentences, the structure or expressions of a specific post or a book, or the model had absorbed the essence of linguistic expression of the original, i.e., how words were selected and arranged at the level of sentences, paragraphs, or the entire document.



# What is Fair Use

## 2. Type of Copyrighted Work

\* Training a general-purpose AI foundation model may involve circumstances that weigh both in favor of and against factor 2 (type and purpose of the copyrighted work).

(1) Likely to weigh in favor of fair use

The model had been trained using short phrases or general, non-professional expressions primarily to convey facts and information, such as social media posts or reviews, and the outputs showed relatively low degree of creativity.

\* Works produced and made public by the State or local governments in the course of their official duties (Article 24-2(1) of the Copyright Act), as well as works not protected under the Copyright Act (Article 7), may be used regardless of fair use. For further details, see p.14.

(2) Likely to weigh against fair use

Works that were used in training have a high degree of literary or artistic creativity and strong individuality of expression.

## 3. Amount and Substantiality of the Portion Used

(1) Likely to weigh in favor of fair use

The developer had no choice but to input hundreds of millions of data, and the amount and extent of the copyrighted works used may be considered reasonable considering the objective of enabling general-purpose language comprehension and generation capabilities.

\* In other words, the extent of the use is within the scope reasonably necessary to achieve the purpose recognized under factor 1.

(2) Likely to weigh against fair use

The extent of the use of works was not structurally unavoidable given the architecture of the AI model, nor was it reasonably necessary to achieve the objective of enabling general-purpose language comprehension and generation capabilities.

\* In other words, the extent of the use exceeds, qualitatively and quantitatively, the scope reasonably necessary to achieve the purpose recognized under the first factor.

---

#### 4. Effect on the market or value of the work

(1) Likely to weigh in favor of fair use

Outputs from the trained model or the services based on the outputs did not substantially substitute for the viewing, use, or sale of the original works. Also, works with a relatively low level of creativity — such as posts, comments, reviews, photographs, or videos shared on social media or on personal blogs — or mundane works created for non-commercial purposes, particularly where it was difficult to identify or locate the copyright holder, were also used to train the AI.

(2) Likely to weigh against fair use

AI training or the provision of related services resulted in a decline in sales of the copyrighted work, caused economic harm to the rightsholder, and deprived the rightsholder of licensing opportunities, enough for there to be a significant risk of market substitution or market dilution. Also, a licensing framework for AI training under reasonable terms — such as a separate licensing system specifically for AI training purposes — as well as a collective management system were already in existence, making it relatively easy to obtain authorization. In other words, a current or potential market was highly likely to be formed. This means that the exploitation of works is likely to be denied the fair use defense.

#### B. Applied AI services and models: Cases where fair use may be recognized

---

- For fair use to be accepted, the training of GAI must not unreasonably prejudice the legitimate interests of the rightsholder, and the copyrighted works used for training must serve a purpose or character different from that of the original. In other words, if a model is trained to generate outputs whose purpose or character is different from the works used in training, there is higher likelihood that fair use may be accepted.



# What is Fair Use

- A use is considered transformative if the model generates outputs that make substantial contribution to the public interest — namely, if the GAI model creates new value or new uses, and the outputs cause little or no harm to the existing market for the original works or to the rightsholder’s economic interests. In particular, GAI training conducted for research or educational purposes has a relatively higher possibility to being recognized as fair use, if it is non-commercial.
- In addition, lawfully obtaining the works used for training and installing measures to prevent the generation of outputs identical or substantially similar to the original works will weigh in favor of a finding of fair use. Furthermore, if the outputs do not result in loss of sales, economic harm, or a clear deprivation of licensing opportunities for the original works, such that there is no risk of market substitution, such exploitation may also be deemed fair use.
- Based on these criteria, the sections below present major examples of AI training cases that may be recognized as fair use.

## 1. AI Training Based on Public Data for Non-Profit Purposes

A government-funded research institute used public data, as defined under the Act on Promotion of the Provision and Use of Public Data, to train its natural language processing (NLP) model as a part of its research on social inequality.

- (1) Purpose and character of the use: Non-profit research purpose → Weighs in favor of fair use
- (2) Type of copyrighted work: Public data as defined under Article 3(4) of the Act on Promotion of the Provision and Use of Public Data may be used for commercial purposes → Weighs in favor of fair use

### Act on Promotion of the Provision and Use of Public Data

**Article 3 (Basic Principles)** (4) No public institution shall prohibit or restrict the use of public data for gain, except as otherwise expressly provided for in any other Act or except in cases falling under Article 28 (1).

**Article 17 (Scope of Public Data to be Released)** (1) The head of each public institution shall provide citizens with the public data held and managed by the public institution: Provided, That the same shall not apply to data; including any of the following information:

1. Information subject to non-disclosure under Article 9 of the Official Information Disclosure Act;
  2. Information which involves any third person's right protected under the Copyright Act or any other statute and the use of which is not duly authorized under the relevant law.
- (2) Notwithstanding paragraph (1), if the information referred to in the subparagraphs of paragraph (1), is technically separable, the head of the relevant public institution shall exclude such information when providing its public data.
- (3) The Minister of the Interior may propose a plan for obtaining authorization to use the public data excluded from provision, classified as involving any third person's right under

- (3) Amount and substantiality of the portion used: Although the entire public dataset did not have to be used for the analysis of social inequality, the whole dataset was used → Weighs against fair use
  - (4) Effect on the market: There was no impact on the market for the original works, nor risk of market substitution → Weighs in favor of fair use
- Because the public data was used for non-profit purposes such as research and there was no likelihood of loss of sales of the original works or decrease of licensing opportunities, the use is highly likely to be recognized as fair use.



# What is Fair Use

## 2. Developing AI that Summarizes Science-and-Technology-Related Aspects of Open-Access Papers

A university research team collected papers published as open access and developed an AI science-and-technology summarization model. It then released the model also as open source (or open weights).

- (1) Purpose and nature of use: The use was transformative in that the team extracted only the science-and-technology parts that correspond to ideas in the published papers. → Weighs favorably
  - (2) Type of work: Academic papers are works with scholarly value, intended to communicate and spread expert knowledge → Weighs unfavorably
  - (3) Amount used for training: Even though the full paper was used, using the entire content was necessary to achieve the transformative goal of accurately extracting the science-and-technology-related parts. (i.e., The use was within the scope of transformative purpose under factor 1.) → Weighs favorably
  - (4) Market impact: The exploitation of works did not replace the original work. Rather, it may even increase opportunities for the original paper to be used. → Weighs favorably
- The works that were exploited were freely accessible online and were used for transformative and non-profit purposes such as research. Because the likelihood of replacing or compromising licensing opportunities was low, this kind of case is highly likely to be recognized as fair use.

## 3. Developing AI for Data Analysis of Science and Engineering Papers

An AI developer used the full text of science and engineering related papers that were freely available to build a model that could automatically analyze data and support research.

- (1) Purpose and character of use: The works were used in training not for the content itself but to develop methods for automated data analysis. So, the purpose and character of use differ from those of the original works. → Weighs favorably
  - (2) Type of work: Academic papers have scholarly value and are intended to communicate and spread expert knowledge. → Weighs unfavorably
  - (3) Amount used for training: Although whole works were used for GAI training, using the full paper was necessary to achieve the transformative goal of automated data analysis and research support. → Weighs favorably
  - (4) Market impact: There was no impact on the existing market, and no possibility of substituting for the original work. → Weighs favorably
- The papers were not used to reproduce their expressive elements but were used transformatively to develop automated data-analysis methods. Only the data (not the full body text) were used, and the impact on existing demand for the works was minimal. Because there was no real harm to the original work's market, it is highly likely to be recognized as fair use.

#### 4. Developing an AI Model to Analyze a Criminal's Movement Patterns Using Audiovisual Works

An AI developer trained a model using audiovisual works such as drama series or films that the company had lawfully downloaded or purchased on CDs to develop crime-prevention applications such as those analyzing criminal movement patterns.

- (1) Purpose and character of use: The training aimed to generate outputs for crime analysis, which was unrelated to the purpose of the original dramas series or films. In other words, the purpose and character differ from the original work. → weighs favorably
- (2) Type of work: The exploited works were primarily meant to be enjoyed for their content. → Weighs unfavorably



# What is Fair Use

- (3) Amount used for training: The entire work was used for GAI training because differentiating and separately copying only the “necessary” parts while leaving out the “unnecessary” parts were realistically unfeasible. In other words, using the full video was unavoidable to achieve the transformative goal of movement-pattern analysis. → Weighs favorably
- (4) Market impact: No likelihood of substituting for the original work in its current or potential market. → Weighs favorably
  - Although the works were being used by a for-profit company, the audiovisual works were attained legally and were not used to copy the expressive elements of the work but used transformatively for movement-pattern analysis. Because they were used only to analyze human motion patterns and did not cause real market harm or pose a substitution risk, it is highly likely to be recognized as fair use.

## C. Applied AI Services and Models Unlikely to be Accepted as Fair Use

- The fair use defense is less likely to be accepted in cases where GAI training harmed the rightsholder’s legitimate interests, lacked a transformative purpose, couldn’t justify as serving a social or public interest or was primarily for commercial gain. If the use of works had little to do with real public benefit, had failed to create new value or new use case, or interfered with the existing market for the work or harmed the rightsholder’s economic interests, fair use is unlikely to be accepted. Also, fair use may not be accepted if, during the training process, original’s expressions were copied as-is, or if the model’s outputs reproduced the original work in ways that were essentially the same or substantially similar.
- Additionally, using works as training data after being obtained illegally, or when the license status was unclear, directly harms the rightsholder’s own opportunity to license the work and would thus weigh against fair use. Furthermore, if a model is trained excessively on specific works, or if the outputs substitute for the original work’s market or reduce demand for sales or use, fair use is unlikely to be accepted.

- Based on these criteria overall, the following section presents some AI training examples where fair use is unlikely to be recognized.

### 1. AI Summarization Service Trained on Full-Text News Articles

An AI developer, without permission from the relevant news publisher, crawled the web and trained its model on entire news articles, and then ran a commercial service that automatically provided summaries of articles.

(1) Purpose and character of use: The use was commercial. It could not be seen as transformative because it served the same or a very similar purpose as the publisher's original news articles. In particular, the works were collected without authorization, making the use even more weighed against fair use. → Weighs unfavorably

\* For details on how to determine whether a use is "transformative," see the "Summary Table of Fair Use Factors," p. 43.

(2) Type of work: News articles often include a journalist's interpretation and commentary on social events, not just raw facts. → Weighs unfavorably

(3) Amount used for training: The developer used the entire work for GAI training. → Weighs unfavorably

(4) Market impact: If the AI summary provides the key content of the news articles enough so that users can attain information without visiting the original article, that may decrease use of the original, encroaching on the publisher's subscription and advertising revenue. → Weighs unfavorably

▸ The fact that the works were collected without permission, that the purpose doesn't seem to be transformative, and that outputs could directly substitute for the original market and cause economic harm all lead to the unlikelihood that fair use might to be accepted.



# What is Fair Use

## 2. Training an AI model using lawfully purchased textbooks and selling teaching materials

An edutech company bought hundreds of digital textbooks from multiple publishers and used them to train an AI model that could create textbooks or workbooks.

(1) Purpose and character of use: The works were being used commercially. The use was not transformative because the purpose was the same as or similar to that of the original works.  
→ Weighs unfavorably

\* For details on how to determine whether a use is “transformative,” see the “Summary of Fair Use Considerations for Each Factor” on p.46.

(2) Type of work: The works that were used had scholarly and educational value and were creatively compiled or newly authored by selecting, arranging, and restructuring many materials for teaching. → Weighs unfavorably

(3) Amount used for training: Entire works were used for GAI training.  
→ Weighs unfavorably

(4) Market impact: The AI output may replace or weaken the publishers’ textbook or workbook market. → Weighs unfavorably

▸ Even though the original textbooks were purchased legally, the purpose of the use was not transformative, and the outputs may replace or weaken licensing opportunities of the original works. Thus, fair use is unlikely to be accepted.

## 3. Training of AI to Generate Commercial Stock Images Using Works Without Permission

A commercial AI model was developed to generate images after unauthorized crawling and training on millions of high-resolution paid images from paid stock-image websites. The company subsequently sold these generated outputs.

(1) Purpose and character of the use: The use was commercial and for the same or a similar purpose as the original works. It was not transformative. → Weighs unfavorably

\* For details on how to determine whether a use is “transformative,” see the “Summary of Fair Use Considerations for Each Factor” on p.46.

(2) Type of work: Photographs and works of art are artistic works intended to be enjoyed as expressive works in themselves. → Weighs unfavorably

(3) Amount used for training: Entire works were used to train the GAI. → Weighs unfavorably

(4) Market impact: The use was likely to substitute for or weaken the market for stock-image or licensing. → Weighs unfavorably

▸ The works were collected without permission, and the purpose of the use wasn’t transformative. Also, the outputs were likely to directly replace the market of the original works, imposing a risk of economic harm. In particular, paid image platforms are usually designed to restrict access, using watermarking, controlling API access, embedding robots.txt and using other technical measures to prevent unauthorized collecting of its images. Therefore, training an AI model using such unauthorized use of works is unlikely to be accepted as fair use.

#### 4. Producing AI Cover Songs After Training on Musical Works

A developer purchased thousands of songs from an album of a popular singer from a music platform, used the songs to train an AI model and then ran a paid AI cover-song generation business.

(1) Purpose and character of the use: The use was commercial and served the same or a similar purpose as the original works. So, it was not transformative. → Weighs unfavorably

\* For details on how to determine whether a use is “transformative,” see the “Summary of Fair Use Considerations for Each Factor” on p.46.



## What is Fair Use

- (2) Type of work: A musical work is an artistic work that creatively expresses a particular feeling or emotion by harmonizing pitch, duration, and intensity of various sounds. → Weighs unfavorably
- (3) Amount used for training: Entire works were used to train the GAI. → Weighs unfavorably
- (4) Market impact: The use was likely to substitute for or weaken the market for the original music. → Weighs unfavorably
  - The purpose of the use was not transformative, and the outputs could directly replace the market of the original works and cause economic harm. In particular, if using the works to train the AI model was beyond the scope of the signed license, that fact can also act against fair use.
  - ※ Merely imitating a singer's voice or singing style is not considered as a direct act of copying, and may therefore be less likely to be recognized as infringement of related rights. However, if a singer's actual recordings (sound sources) are copied and used for training, that constitutes infringement of the reproduction right.

# Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models



# IV

---

## Responding to Copyright Disputes Involving Generative AI

# Guide on Applicability of the Fair Use Doctrine

to Training of Generative AI Models



- 
1. Probability of a Dispute
  2. Response

# IV

## Responding to Copyright Disputes Involving Generative AI

### 1. Probability of a Dispute

- The licensing counterpart cannot be pinpointed  
AI developer

AI developers complain that they cannot realistically identify each and every right holder to sign a license agreement with for every work that they use in GAI training.
- Parties cannot agree on remuneration  
Rights holder vs. AI developer

Rights holders and AI developers have differing views on remuneration for use of works in GAI training because their perspectives on use of copyrighted works differ, increasing the probability of a dispute.
- Output infringes copyright  
GAI user

Outputs generated by AI that are identical or similar to the original works that the model trained on may infringe copyright.\*

\* For more details on infringement of copyright by GAI output, refer to *Guide to Preventing Copyright Disputes Related to Generative AI Outputs*.

### 2. Response

- A system needs to be set up to prevent copyright disputes arising between rightsholders, AI developers and GAI users and to speedily deal with them when they do.

#### A. Prevention of Dispute

##### (1) Supporting Facilitation of Transactions

- The Ministry of Culture, Sports and Tourism (MCST) will provide support to identify AI developers' demand for training data (including preferred datasets) and to facilitate one-on-one matching with rightsholders so that relevant parties are able to reach an agreement on reasonable transaction terms (e.g., criteria for calculating remuneration, method of payment).

[Table: Current status of collective management organizations (CMOs)]

Field	Name of CMO	Website
Music (lyrics and compositions)	Korea Music Copyright Association	komca.or.kr
	Korea Society of Composers, Authors and Publishers	koscap.or.kr
Music performances	Federation of Korean Music Performers	fkmp.kr
Music productions	Recording Industry Association of Korea	riak.or.kr
Literary, photographic, theatrical, audiovisual & artistic works (excl. TV or movie script writers)	Korea Literary Arts Copyright Association	kolaa.kr
Literary works (TV script writers)	Korea Television and Radio Writers Association	ktrwa.or.kr
Literary works (movie script writers)	Korea Scenario Writers' Association	scenario.or.kr
Broadcasted performances	Korea Broadcasting Performers Association	kobpra.or.kr
Audiovisual works	Korean Film Producers Association	kfpa.net
News works	Korea Press Foundation	kpf.or.kr
Public sector works	Korea Cultural Information Service Agency	kcisa.kr

- If rightsholders and AI developers diverge greatly regarding the use of works, the MCST, acting as a neutral mediator, will provide the disputing parties with an opportunity for negotiation and help reach an agreement. (If negotiations fail, the MCST can assist the parties to apply for dispute mediation.)

## (2) Providing Rights Information and Building a Distribution Platform

- The MCST is also planning to build and provide an integrated database of rights management information\* to help AI developers cut their cost of tracking down rightsholders. Such database will be useful when AI developers want to use copyrighted works as AI training data and need to sign a license with the relevant rightsholder.

# IV

## Responding to Copyright Disputes Involving Generative AI

\* Rights management information includes:

- Details that help identify a work
- Information that identifies the rightsholder of copyright or other rights protected by the copyright law
- Details on how the work can be used, under what conditions

### (3) Expanding Access to Public-Sector Works for AI Training

- To make it easier for AI developers to use public-sector content as their training data, the MCST will add new license types to its Gonggong Nuri\* system (introduced on Jan 28, 2026) – “Type 0” and “Type AI” – to promote more access to quality public domain works created or held by the public sector for AI training.

- **Type 0:** Free to use for any purpose, including AI training, without any conditions

- **Type AI:** Maintain the existing conditions under Gonggong Nuri Types 1 to 4 but also may be freely used for AI training (to be shown together with the relevant Type 1–4 label).

\* Gonggong Nuri is the Korean name for the Korea Open Government License (KOGL), a license labeling system for government/public-sector works. It shows the reuse conditions — such as attribution, commercial use or modification — by classifying works into Types 0–4 and Type AI.

- As new types of licenses are rolled out, continuous support will be given, such as visiting and training of relevant persons-in-charge in public sector institutions and helping transform existing licenses into the new ones.

### B. Resolution of Dispute

- The Korea Copyright Commission (KCC) will strengthen its copyright counselling services for rightsholders, AI developers and GAI users who need legal information and advice.

Establishment and operation of new counselling channel specializing in AI

Telephone: Main line of KCC (+82-1800-5455) → Press “0” → AI related counselling

- When a copyright dispute arises regarding use of copyrighted works in AI training or related to GAI outputs, the KCC will support relevant parties to come to a speedy and fair resolution through its copyright mediation services.

\* The KCC has specialized panels dedicated to professional and efficient resolution of disputes related to AI and copyright.

**Reference** Copyright mediation system

- **Overview:** Copyright mediation is a system designed to resolve copyright disputes quickly by helping the parties reach a mutually acceptable agreement through a panel of experts.
- **Legal basis:** Copyright Act Articles 113 and 114 through 118-2.
- **Effect:** A written mediation protocol has the same effect as a court conciliation. If the agreed terms are not carried out, enforcement is possible without filing a separate lawsuit.
- **Application fee:** The fee is typically from minimum KRW 10,000 up to several hundred thousand won, depending on the case.
- **Advantages of mediation over litigation:** Mediation saves time and cost, helps keep sensitive information confidential, and allows the parties to maintain ownership of the process.
- **How to apply:** Applications can be submitted online (adr.copyright.or.kr) or offline (in person / by post). For inquiries, contact the Korea Copyright Commission's Mediation & Appraisal Team (+82-2-2669-0044).
- **Process**

```

graph TD
    A[Parties receive guidance on mediation] --> B[Parties submit application]
    B --> C[KCC forms panel and sets date for mediation]
    C --> D[Mediation]
    D --> E[Mediation successful]
    D --> F[Mediation failure]
    F --> G[Ex-officio decision, when:  
1. A party rejects, without reasonable grounds, draft protocol proposed by the panel;  
2. The value of the settlement is expected to be less than KRW 10 million.]
    G --> H[No objection]
    G --> I[Objection]
    H --> J[Mediation successful]
    I --> K[Mediation failure]
  
```

[Diagram] Copyright mediation procedure

# IV

## Responding to Copyright Disputes Involving Generative AI

**Reference** Data-related dispute mediation system

- **Overview:** When a dispute arises around producing, trading, or using data, a mediation panel made up of experts can help the parties reach an agreement to resolve the dispute fairly and quickly.
- **Legal basis:** Framework Act on the Promotion of Data Industry and Utilization (Articles 34 to 41).
- **Effect:** A mediation protocol has the same legal effect as a court conciliation under the Civil Procedure Act.
- **Cost:** Free of charge, except for cases where the Data Dispute Resolution Committee (DDRC) decides that mediation costs must be borne by a party to the dispute
- **Advantages of mediation over litigation:** Mediation saves time and cost, and helps keep sensitive information confidential.
- **How to apply:** Applications can be made online ([www.ddrc.go.kr](http://www.ddrc.go.kr)). For inquiries, contact the Data Dispute Resolution Committee (+82-53-230-1957).
- **Process**

```

    graph LR
      A[Parties submit application] --> B[DDRC forms panel  
(of no more than 3 persons)]
      B --> C[Verification of facts and testimonies  
by each party]
      C --> D[Proposal of draft protocol]
      D --> E[Mediation successful]
      D --- F(If parties accept the protocol)
  
```

[Diagram] Data dispute mediation procedure

# Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models



---

# Appendix

# Guide on Applicability of the Fair Use Doctrine

to Training of Generative AI Models



---

AI-Copyright Law Working Group (ACLWG)  
Special Sub-Group

- Since 2023, the Ministry of Culture, Sports and Tourism (MCST) and the Korea Copyright Commission (KCC) have been operating a working group made up of representatives from the AI industry, rightsholder organizations, academia, the legal community, and relevant government ministries, with the aim of reviewing key copyright issues that have become more prominent with the rise of AI. The working group also reviewed possible legal and policy improvements as well as looking into ways to better promote transactions in AI training data.
- In 2025, the Special Sub-Group of the 3<sup>rd</sup> Phase AI-Copyright Law Working Group reviewed the draft of this *Guide*, followed by discussions in the full working-group meetings and a round of consultations with the public and relevant government ministries.
- The members of the ACLWG Special Sub-Group are as follows.

### 〈Members of ACLWG Special Sub-Group〉

Area	Name	Affiliation & Title
Legal Profession	Gyu-Hong LEE	Principal Research Fellow, Judicial Policy Research Institute (Presiding Judge)
	Min-Jeong KIM	Chief Prosecutor, Trial and Litigation Division 2, Incheon District Prosecutors' Office
	Sang-Hyeok LIM	Partner Attorney, Shin & Kim LLC
Academia	Jun-Woo PARK	Professor, Sogang University Law School
	Seung-Jae CHOI	Professor, Department of Law, Sejong University
	Cheol-Nam LEE	Professor, Chungnam National University Law School
	In-Cheol KIM	Professor, IPR Major, Sangmyung University
	Jin-Won CHOI	Professor, Department of Law, Daegu University
Government / Public Institutions	Hyang-Mi JEONG	Director General, Culture and Arts Policy Bureau, Ministry of Culture, Sports and Tourism (MCST)
	Young-Jin CHOI	Director, Copyright Policy Division, MCST
	Yong-Su KIM	Senior Deputy Director, Copyright Policy Division, MCST
	Ji-Hyeon LEE	Deputy Director, Copyright Policy Division, MCST
	Su-Jeong SHIN	Assistant Director, Copyright Policy Division, MCST
	Hyung-Min PARK	Assistant Director, Copyright Policy Division, MCST
	Hye-Chang KIM	Director General, Policy Research Bureau, Korea Copyright Commission (KCC)
Chan-Dong KIM	Team Leader, Legal Research Team, KCC	

## Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models

---

Area	Name	Affiliation & Title
Government / Public Institutions	Hyo-Jeong HAN	Senior Researcher, Legal Research Team, KCC
	Hee-Young OH	Researcher, Legal Research Team, KCC

# Guide on Applicability of the Fair Use Doctrine

## to Training of Generative AI Models

---

**Published on** 26<sup>th</sup> February 2026

**Planned by** Copyright Policy Division, Copyright Policy Bureau,  
Office of Culture and Media Industry, Ministry of Culture,  
Sports and Tourism (MCST)  
Legal Research Team, Policy Research Bureau, Korea Copyright  
Commission (KCC)

**Written by** Chan-Dong Kim, Team Leader  
(Legal Research Team, Korea Copyright Commission)  
Hyo-Jeong Han, Senior Researcher  
(Legal Research Team, Korea Copyright Commission)

**Published by** Korea Copyright Commission (KCC)

**Address** 117, Soho-ro, Jinju-si, Gyeongsangnam-do, Republic of Korea

**Website** Ministry of Culture, Sports and Tourism [www.mcst.go.kr](http://www.mcst.go.kr)  
Korea Copyright Commission [www.copyright.or.kr](http://www.copyright.or.kr)

**Printed by** KyungSung

**Translated by** Seungmin CHOI

---