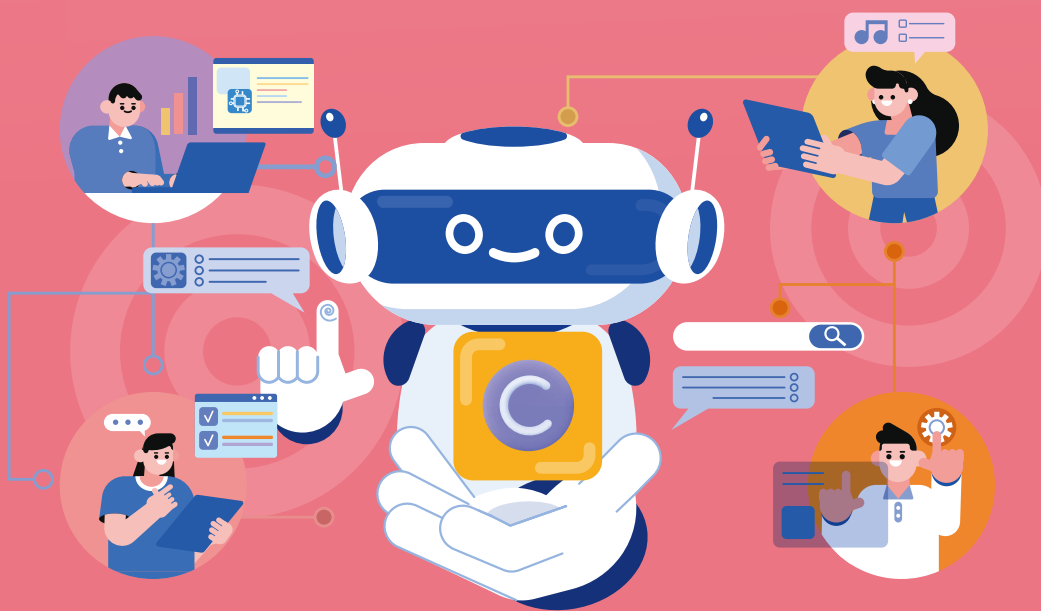


Guide to Preventing Copyright Disputes Related to Generative AI Outputs

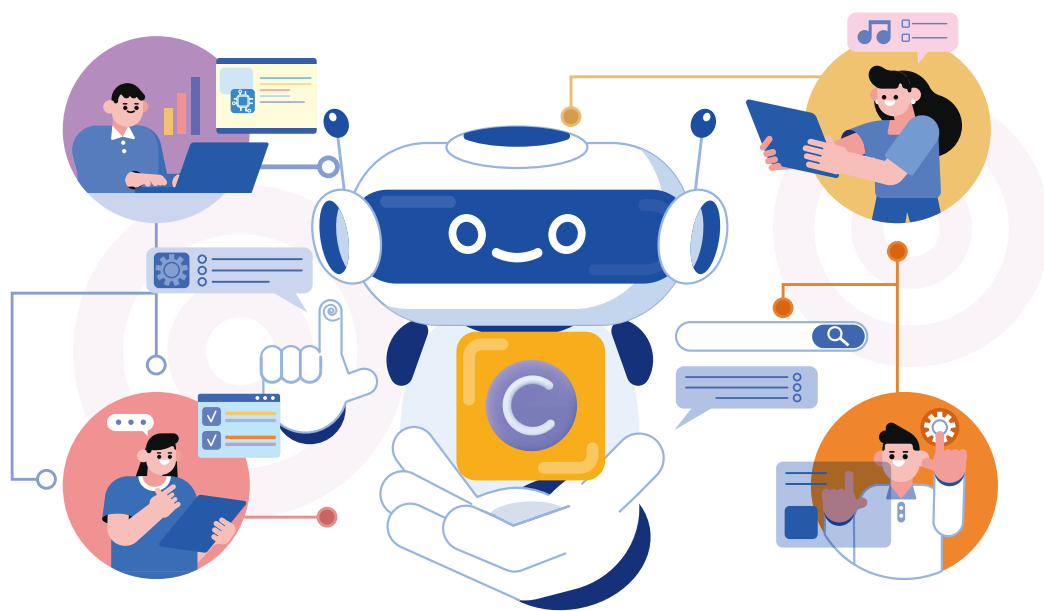


Ministry of Culture, Sports
and Tourism



KOREA COPYRIGHT
COMMISSION

Guide to Preventing Copyright Disputes Related to Generative AI Outputs



Ministry of Culture, Sports
and Tourism



KOREA COPYRIGHT
COMMISSION

- Generative AI (GenAI) mimics the structure and characteristics of input data to produce text, sounds, images, videos, and other forms of output. However, copyright disputes can arise when these outputs are identical or similar to works created by humans.
- No one would disagree that the criteria of access and substantial similarity, which are generally applied when determining copyright infringement, should also apply to outputs generated by generative AI.
- Against the backdrop of this emerging issue, the Ministry of Culture, Sports, and Tourism (MCST), in collaboration with the Korea Copyright Commission (KCC), published this Guide to inform stakeholders about key considerations to prevent copyright disputes arising from GenAI outputs.
 - * This Guide includes a general explanation of Korea's copyright law prior to addressing specific copyright concerns related to GenAI.
 - * For more detailed information on the classification and copyrightability of GenAI outputs, refer to the *Guide to Copyright Registration for Generative AI-Assisted Works*.
- As a party to major international treaties such as the Berne Convention, the TRIPS Agreement, and the WIPO Copyright Treaty (WCT), Korea established its copyright framework in alignment with international norms and is continually enhancing its copyright regime in line with its obligations under these international treaties. These regulations apply equally to both national and foreign copyright holders and users of copyrighted works.
- This Guide was reviewed and finalized by the AI-Copyright Law Working Group (ACLWG), consisting of representatives from the AI industry, rights holders' organizations, academia, the legal profession, and relevant government ministries.
 - * The content of this Guide is based on the current Copyright Act and precedents as of June 2025. It may be subject to change due to future legislation, amendments, new judicial rulings, or advancements in AI technology.

CONTENTS

Foreword

1

Overview of the Korea Copyright Act (KCA)

- | | |
|---|----|
| 1. Copyright | 08 |
| 2. Neighboring rights | 12 |
| 3. Right to exclusive publication and right to publish a work | 13 |
| 4. Rights of a database producer | 14 |
| 5. Limitations to economic rights | 15 |
| 6. Remedy in case of a copyright infringement | 16 |

2

Overview of infringement of copyright and other rights

- | | |
|------------------------------------|----|
| 1. Idea-expression dichotomy | 22 |
| 2. Infringement of economic rights | 22 |
| 3. Infringement of moral rights | 23 |

3

Generative AI and types of outputs

- | | |
|---|----|
| 1. Generative artificial intelligence (GenAI) | 26 |
| 2. Fully AI-processed outputs | 26 |
| 3. GenAI-assisted works | 26 |

4

When a GenAI output infringes copyright

- | | |
|---|----|
| 1. Technological features of GenAI | 30 |
| 2. Forms of infringement during each phase of GenAI | 30 |



Guide to Preventing Copyright Disputes Related to Generative AI Outputs

5

Factors that determine whether a GenAI output infringes copyright

- | | |
|---------------------------------------|----|
| 1. Determining access | 34 |
| 2. Determining substantial similarity | 35 |
| 3. Determining the infringer | 36 |

6

Guide for rights holders

- | | |
|--|----|
| 1. How to prevent a copyright dispute involving a GenAI output | 40 |
| 2. Steps to take in the event of a dispute | 43 |

7

Guide for users of generative AI

- | | |
|---|----|
| 1. Precautions to take regarding generative AI outputs | 46 |
| 2. Guidelines for inputting prompts | 46 |
| 3. Precautions to take when using or distributing AI outputs | 47 |
| 4. Importance of checking the terms and conditions of an AI service | 47 |
| 5. Copyright infringement liability of users of generative AI | 51 |

8

Guide for generative AI operators

- | | |
|--|----|
| 1. Types of GenAI operators | 54 |
| 2. Copyright-related precautions regarding GenAI outputs | 54 |
| 3. Guidelines for preventing copyright disputes | 55 |
| 4. Provisions regarding GenAI outputs in a GenAI operator's terms and conditions | 58 |

Annex 1	Q&A	62
----------------	-----	----

Annex 2	Outcome of the ACLWG Breakout Group on the Use of AI Outputs	66
----------------	--	----



Overview of the Korea Copyright Act (KCA)



1. Copyright
2. Neighboring rights
3. Right to exclusive publication and right to publish a work
4. Rights of a database producer
5. Limitations to economic rights
6. Remedy in case of a copyright infringement

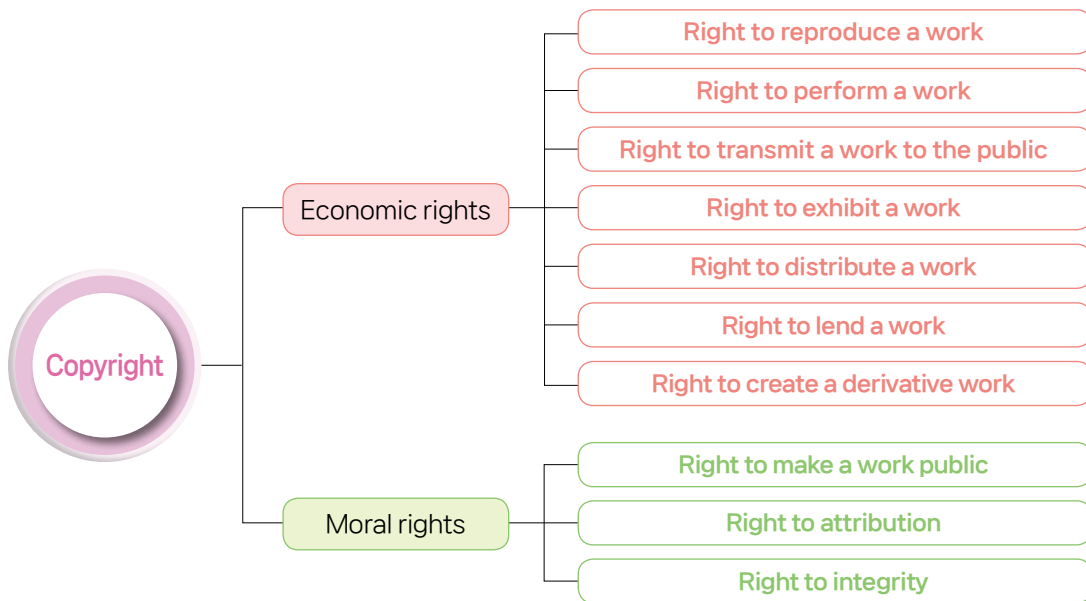
1

Overview of the Korea Copyright Act (KCA)

1. Copyright

A. Broad definition of copyright

- Copyright is a set of rights granted to the creator of a work. While copyright is often viewed as a singular concept, it actually comprises a “bundle of rights” categorized broadly into **moral rights** and **economic rights**.
- **Moral rights** protect the author’s reputation and personal interests and comprise three rights.
- **Economic rights** safeguard the author’s economic interests and include seven rights.



[Figure 1] Types of copyright

B. Moral Rights

- **Right to make a work public:** Moral rights include the right of an author to determine whether to make their work public¹⁾
- For instance, if an individual B secretly photographs a painting by another individual A, which the author had not yet made public, and subsequently posts it on their blog, B would be infringing A’s right to make a work public.

1) Korea Copyright Act (KCA) Article 2.25 and Article 11

- **Right to attribution:** Right to attribution is the right of an author to display or not to display his or her name (real name, nom de plume, or alias) on their work²⁾
 - If someone displays a sculpture made by person A but labels it as an artwork by person B, this would be violating A's right to attribution.
- **Right to integrity:** This is the right to prohibit the content, form, or title of a work from being changed against the author's will.³⁾
 - If an individual B purchases a painting by painter A and exhibits it after changing the background color without A's consent, B is violating A's right to integrity.

C. Economic rights (Narrow definition of copyright)

- **Right to reproduce a work** involves the right to temporarily or permanently fix a work on a tangible object by printing, photographing, copying, audio-recording, video-recording, or downloading it, or to reproduce it in the form of a tangible object. In the case of an architectural work, it also includes the right to construct a building according to an architectural model or blueprint.⁴⁾
 - Copying, without permission from the rights holder, a movie or piece of music made by someone else, is an infringement of the right to reproduce a work.
- **Right to perform a work** is the right to make public a work by performing, singing, orally narrating, reciting, screening or reproducing it, or through any other method. Performing a work also includes transmitting (excluding interactive transmission) it within a connected space occupied by the same people.⁵⁾
 - Playing music or reading poetry created by someone else in a public place without the permission of the rights holder is an infringement of the right to perform a work.
- **Right to transmit a work to the public** is the right to transmit a work or make the work available for use using wireless or wired communication so that the public can receive or access it.⁶⁾ This right can be further divided into the right to broadcast or to transmit a work interactively, or to make a digital audio transmission.

2) KCA Article 12

3) KCA Article 13

4) KCA Article 2.22 and Article 16

5) KCA Article 2.3 and Article 17

6) KCA Article 2.7 and Article 18

-
- **Right to broadcast a work** is the right to transmit sound or image, or sound and image, for simultaneous reception by the public.⁷⁾
 - Broadcasting music created by someone else without their permission is an infringement of their right to reproduce a work, as well as their right to broadcast a work.

 - **Right to interactively transmit a work** is the right to make a work available for use so that members of the public can access it at a time and place of their own choosing. It also includes the transmission of the work during that process.⁸⁾
 - Posting a photo taken by someone else on one's website without their permission may be an infringement of the right to reproduce a work and the right to transmit a work interactively. However, the holder of the rights to the photo stating that the work may be copied onto another website would be considered as giving permission according to the KCA and thus would not constitute copyright infringement.

 - **Right to digital audio transmission** involves the right to transmit sound in digital form initiated at the request of a member of the public for the purpose of simultaneous reception of the sound by the public (excludes interactive transmission).⁹⁾ Using a piece of music created by someone else in a live music webcast without the right holder's permission is an infringement of the right to digital audio transmission.

 - **Right to exhibit a work** is the right to display or publish a tangible object, such as art, photography, or architectural work, in its original form or a copy thereof, for free public viewing.¹⁰⁾
 - Displaying a painting drawn by another person in a public place without the permission of the rights holder is an infringement of the right to exhibit a work.

 - **Right to distribute a work** is the right to transfer or lend a work, in its original form or as a reproduction, to the public with or without compensation.¹¹⁾
 - Copying and distributing, without authorization, a book or an album is an infringement of the right to distribute a work as well as the right to reproduce a work.

7) KCA Article 2.8

8) KCA Article 2.10

9) KCA Article 2.11

10) KCA Article 19, and ruling from the Supreme Court of Korea on 11th March 2010, 2009da4343

11) KCA Article 2.23 and Article 20

- **Right to lend a work** is the right to rent out to another party sound recordings published for commercial purposes (i.e., commercial phonograms) or computer programs published for commercial purposes, in exchange for a profit.¹²⁾ This rental right applies only to commercial phonograms and commercial computer programs, and not to renting out books or movies.
 - Lending a music CD or a computer program to another person and being paid for it is a violation of the right to lend a work.
- **Right to create a derivative work** protects the right to create another work of originality by translating, arranging, transforming, adapting, or cinematizing the initial original work.¹³⁾
 - Translating a foreign novel into Korean or cinematizing it, without the permission of the rights holder, is an infringement of the right to create a derivative work.

D. Term of protection for economic rights¹⁴⁾

	Start of protection	End of protection
General rule	The moment a work is created	70 years after the death of the author
Anonymous or pseudonymous work		70 years after the work was made public
Work made for hire, cinematographic works		70 years after the work was made public
Joint works		70 years after the death of the last surviving author

- The expiry of the term of protection is calculated from 1st January of the year after the death of the author or the year after the work was made public.

12) KCA Article 21

13) KCA Article 5 and Article 22

14) KCA Articles 39 to 44

2. Neighboring rights

- Neighboring rights refer to exclusive rights similar to copyright but are held by performers, phonogram producers, and broadcasters to control the use of their performances, phonograms, and broadcasts.

A. Holders of neighboring rights

- **Performer¹⁷:** A person who performs a work by acting, dancing, playing, singing, speaking, reading, or otherwise entertainingly expressing the work, or similarly expressing something that is not a work, including a person who conducts, directs, or supervises the performance (e.g., actor, singer, instrument player, or conductor)
- **Phonogram (record) producer¹⁵:** A person who plans and takes responsibility for the overall production of a music record¹⁵ (e.g., SM Entertainment or YG Entertainment)
- **Broadcaster¹⁶:** An operator of broadcasting services (e.g., KBS, MBC, or SBS)

B. What neighboring rights do they have?

- **Performer¹⁷:** Right to attribution, right to integrity, right to reproduce a work, right to distribute a work, right to lend a work, right to a (live) performance, right to broadcast (a live performance), right to interactive transmission, right to claim remuneration for broadcast of a commercial phonogram, right to claim remuneration for a digital audio transmission and right to claim remuneration for performance of a commercial phonogram
- **Phonogram producer¹⁸:** Right to reproduce a work, right to distribute a work, right to lend a work, right to interactive transmission, right to claim remuneration for broadcast of a commercial phonogram, right to claim remuneration for a digital audio transmission, and right to claim remuneration for performance of a commercial phonogram

15) KCA Article 2 (Definitions) 5. The term “music record” means that sound (referring to voice and sound; hereinafter the same shall apply) is fixed in a material object (excluding the one in which sound is fixed along with images)[.]

16) Refer to the definition of the “right to broadcast a work” on p.10.

17) KCA Articles 66 to 77

18) KCA Articles 78 to 83-2

- **Broadcaster**¹⁹⁾ : Right to reproduce a work, right to make a simultaneous broadcast, right to perform a work

C. Term of protection for neighboring rights²⁰⁾

	Start of protection	End of protection
Performance	The moment a performance is made	70 years from the year after the performance
		70 years from the year after the publishing of the phonogram where the performance was fixed, in case of phonograms made from a performance fixed within 50 years since the performance
Phonogram	The moment sound is fixed onto a phonogram	70 years from the year after the publishing of the phonogram
		70 years from the first fixation of sound onto the phonogram, in case of a phonogram that was not published within 50 years since the year after the first fixation of sound onto the phonogram
Broadcast	The moment a broadcast is made	50 years from the year after the broadcast

3. Right to exclusive publication and right to publish a work

- **Right to exclusive publication**²¹⁾ is the right to exclusively publish, reproduce, or interactively transmit a work to a user based on the quasi-real right of copyright. The person who is granted this right to exclusive publication can exercise the exclusive and monopolistic right to use the work against all third parties, including the author.
- **Right to publish a work**²²⁾ applies to reproducing and distributing a work in the form of a paper book.
 - In the event a third party uses a relevant work or violates this right, the person with the right to exclusive publication or the right to publish a work may directly become a plaintiff in a civil lawsuit without subrogating the rights holder or may exercise the right to file a criminal lawsuit, thereby seeking civil or criminal remedy.

19) KCA Articles 84 to 85-2

20) KCA Article 86

21) KCA Articles 57 to 62

22) KCA Articles 63 and 63-2

4. Rights of a database producer

- A database is also protected under the KCA and thus requires prior permission from the rights holder for its legitimate use.

A. Definition of a database²³⁾

- ① It must be a compiled matter.²⁴⁾
- ② It must be a systematic arrangement or organization of various elements.
- ③ Each element must be individually accessible and searchable.

B. Rights of a database producer and limitations to the rights

- **A database producer** is a person who made substantial investment of human or financial resources to create a database or to update, verify, or supplement its elements.²⁵⁾
- **Rights granted to a database producer** are the right to reproduce, to distribute, to broadcast, and to interactively transmit a database.²⁶⁾
- **Limitations are imposed** on the rights of a database producer, mutatis mutandis to those imposed on holders of economic rights. Rules regarding assignment, licensing, expiration, and registration of copyright also apply, mutatis mutandis.
- **Term of protection** for a database is 5 years from the date its production ended.

23) KCA Article 2.19

24) KCA Article 2 (Definitions) 17. The term "compiled matters" means the collections of works, symbols, letters, sounds, images and other forms of data (hereinafter referred to as "subject matters"), but shall include the database(.)

25) KCA Article 2.20

26) KCA Article 93

5. Limitations to economic rights

A. What are limitations to economic rights?

- Economic rights are rights that guarantee exclusive and monopolistic use of a work. However, unrestrained exclusivity of rights holders would contradict the very purpose of the KCA, which also seeks to promote the development of culture and related industries. So certain limitations are imposed on economic rights, allowing works to be used without permission from the rights holder, within the scope prescribed by law.
 - If exploitation of a work falls under provisions regarding limitations to economic rights (Articles 23 to 36), the user may exploit the work without the rights holder's permission.

B. Legal provisions limiting economic rights

Article	Content
Article 23	Reproduction for judicial proceedings, etc.
Article 24	Use of political speeches, etc.
Article 24–2	Free use of public works
Article 25	Use for the purpose of school education
Article 26	Use for current news reporting
Article 27	Reproduction, etc., of current news articles or editorials
Article 28	Quotation from works made public
Article 29	Public performance and broadcasting for non-profit purposes
Article 30	Reproduction for private use
Article 31	Reproductions, etc., in libraries, etc
Article 32	Reproduction for examination questions
Article 33	Reproduction, etc., for the visually handicapped, etc.
Article 33–2	Reproduction, etc., for hearing-impaired persons, etc.
Article 34	Temporary sound or video recordings by broadcasting operators
Article 35	Exhibition or reproduction of works of art, etc.
Article 35–2	Temporary reproduction in the course of using works, etc.
Article 35–3	Ancillary reproduction, etc.
Article 35–4	Reproduction, etc., by cultural facilities
Article 35–5	Fair use of works
Article 36	Use by means of translation, etc.

6. Remedy in case of a copyright infringement

A. Principle

- A person whose copyright has been infringed may seek civil remedies and criminal sanctions against the infringer.

B. Civil Remedy²⁷⁾

- A victim may resort to a civil remedy, involving filing a lawsuit against the infringer to prevent or stop infringement, seek return of undue proceeds, or claim damages.
- Under the KCA, the amount of damages claimed can be actual²⁸⁾ or statutory.²⁹⁾
- The statute of limitations extinguishes the right to claim damages if it is not exercised within 10 years from the date of the infringing act or within 3 years from the date the victim gained knowledge of the damage and the perpetrator.

C. Criminal Sanction³⁰⁾

- A person who infringes on economic rights is punishable by imprisonment of not more than five years or a fine of not more than KRW 50 million, or both.
- A person who defames an author or a performer by infringing on the moral rights of the author or the performer is punishable by imprisonment of not more than three years or a fine of not more than KRW 30 million, or both.
- A criminal complaint is a request made by a rights holder to law enforcement to investigate and prosecute another person suspected of willfully infringing copyright.
- Copyright infringement is a crime that can only be subject to prosecution when the victim him or herself

27) KCA Article 123, Articles 125 to 126, and Civil Act Article 766

28) Claiming damages for the amount of loss actually incurred by the rights holder as a result of the infringement

29) Claiming damages for a substantial amount upto KRW 10 million (For intentional, for-profit infringement: KRW 50 million) for each work that was infringed, instead of the actual amount incurred

30) KCA Articles 136 to 142, Criminal Procedure Act Article 230

directly files a complaint, within six months of gaining knowledge of the infringer.

- This means that a prosecutor cannot file a public prosecution based solely on a complaint from a third party who had not suffered infringement of copyright or other rights (hereunder referred to as “copyright, etc.”) protected by the KCA, nor can the law enforcement conduct a substantive investigation.
- However, the KCA also stipulates that a third party may file a complaint when the copyright infringement was for commercial purposes or habitual, or when a person acquires a copy of a computer program and uses that program for occupational purposes, knowing it had been produced by infringing the copyright of another person. Such an act is considered a crime that is prosecutable unless there is an objection from the victim,³¹⁾ and the law enforcement can proceed with an investigation even without a complaint from the victim.

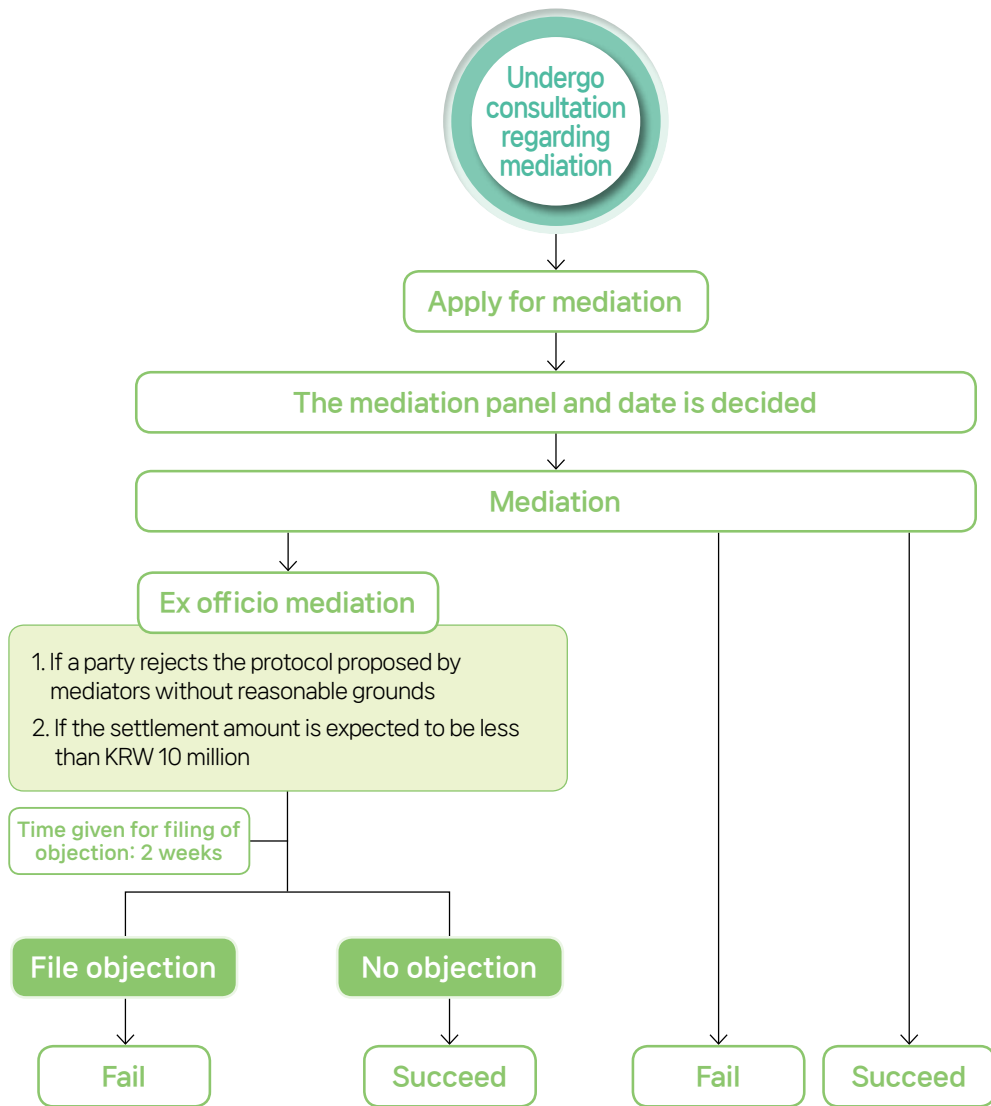
D. Dispute Mediation

- Mediation refers to a system in which a panel composed of three experts from the legal profession, academia, and industry uses their knowledge in the relevant field to induce an amicable settlement between the disputing parties.
- A mediation is speedy, handled by experts, affordable, and helps maintain confidentiality.

Speed	Compared to litigation, the process of applying for and proceeding with mediation is quick and easy, resulting in a relatively quicker resolution (within three months from the date of application) compared to a usually protracted litigation process.
Expertise	Each mediation panel is composed of copyright experts from the legal profession, academia, and industry, providing informed assistance. (Each panel consists of three members, one of whom must be a lawyer.)
Affordability	Resolving a dispute through litigation can be quite costly with attorney and administrative fees, while mediation can settle disputes at a lower cost.
Confidentiality	Proceedings are not open to the public, and so business or personal confidentiality is guaranteed.

- **How to apply for mediation:** For inquiries regarding mediation, including in-person (visit or post) or internet application, please contact the Mediation Appraisal Team of the Korea Copyright Commission. (☎ +82-2-2669-0044)

³¹⁾ In other words, it is an offense for which a person cannot be prosecuted against the will of the victim.



[Figure 2] Mediation procedure



Guide to Preventing Copyright Disputes Related to Generative AI Outputs



Overview of infringement of copyright and other rights



1. Idea-expression dichotomy
2. Infringement of economic rights
3. Infringement of moral rights

1. Idea-expression dichotomy³²⁾

- The idea-expression dichotomy refers to the principle that an idea is not protected by copyright, whereas an expression is. It conceptually limits the scope of protection of copyrighted works.
 - Even if a work is copyrighted, the copyright law does not protect all its components. Parts that are mere ideas, such as simple thoughts, notions, or inventions, are not protected, while expressions of those ideas are.

Supreme Court of Korea, 12th March 2015, Decision 2013da14378

With regard to academic research and arts, copyright protects forms of expression such as words, texts, sounds, or colors that externally articulate ideas and emotions that result from the mental endeavors of humans. Thoughts, such as ideas or theories, and emotions in and of themselves that are articulated in expressions are, in principle, not subject to copyright protection, even if they are unique and novel...

Seoul Civil District Court, 19th April 1990, Decision 89gahab39285

Ideas, in and of themselves, such as artistic style in fine art or thematic techniques in literature, cannot become subject to exclusive protection. That is why only the expressions of such techniques or ideas are protected by copyright.

2. Infringement of economic rights

- Infringement of economic rights involves using a work, a performance, a phonogram, a broadcast, a publication, or a database without permission from the holder of economic rights, neighboring rights, exclusive publication rights, publishing rights, or rights over a database (hereunder “rights holders”) or using a work beyond the permitted scope.
- If a particular form of exploitation satisfies the conditions limiting the rights as stipulated in the KCA*, or if the term of protection for the relevant right has expired**, such use would not constitute copyright infringement even if the user had not attained permission from the rights holder.

* KCA Articles 23 to 38, 62.2, 63-2, 87, 94, 101-3 to 101-5

** Authors are protected for 70 years after their death.

32) See Q1 in Annex 1. Q&A

3. Infringement of moral rights

- Infringement of moral rights refers to using a work of an author or a performer in ways that defame them.
 - Examples of moral rights infringement include publishing a work that the author had not made public (infringement of right to make a work public), deleting the name of the author or performer from the work and putting someone else's name (infringement of right to attribution) and distorting, cutting out or changing the title, content or format of a work or a performance (infringement of right to integrity).



Generative AI and types of outputs



1. Generative artificial intelligence (GenAI)
2. Fully AI-processed outputs
3. GenAI-assisted works

1. Generative artificial intelligence (GenAI)

- Generative artificial intelligence (GenAI) refers to an artificial intelligence system that generates text, images, audio, video, and other forms of outputs in response to specific requests (prompts) entered by a user.

* Prompts do not necessarily have to be text-based.

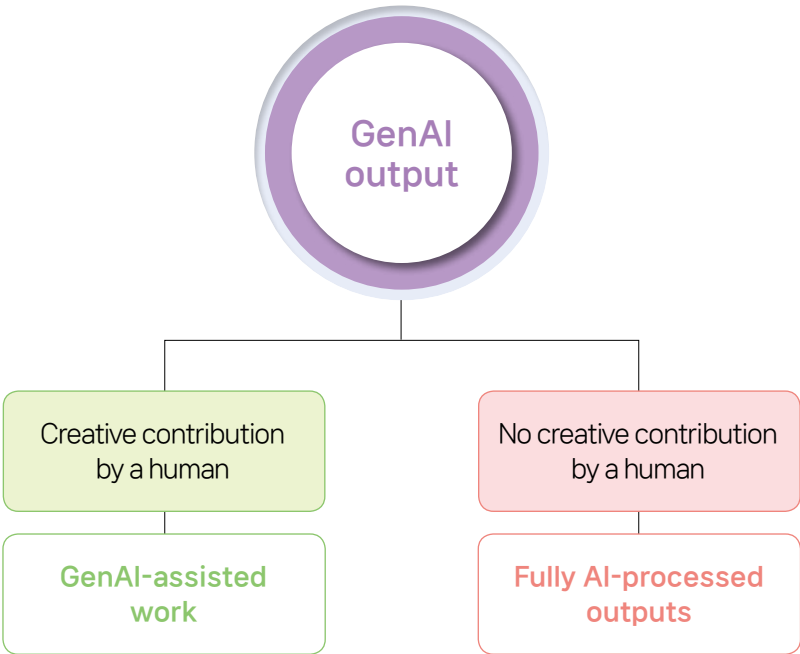
2. Fully AI-processed outputs

- A fully AI-processed output (autonomous output) is a type of an output rendered by the AI system in response to prompts inputted by a human without any creative intervention.

3. GenAI-assisted works

- A GenAI-assisted work is *a copyrighted work that is the result of a human utilizing GenAI as a tool assisting the creative process and thus contains aspects of human creativity*. Some examples of GenAI-assisted works are:

- ① A user enters his or her own work as a prompt, and the creativity of that work is visible in the GenAI output.
- ② A user complements the work with additional touches, such as modifying, augmenting, or eliminating certain aspects, in a creative way, of the GenAI output.
- ③ A user selects, arranges, or reorganizes the GenAI output creatively.



[Figure 3] Types of outputs from GenAI



When a GenAI output infringes copyright



1. Technological features of GenAI
2. Forms of infringement during each phase of GenAI

1. Technological features of GenAI

- Generative AI is technologically characterized by the fact that its purpose is to generate an output, that it is difficult for GenAI model or service developers, or users who have inputted a prompt, to predict what the output is going to look like, and that the GenAI output is a response to or is induced by the prompts inputted by the user.
- GenAI is an artificial intelligence model that trains on a given original dataset and generates data that is difficult to distinguish from the original data (GAN; generative adversarial networks) or follows the distribution of the original data to generate similar data (diffusion, VAE (variational autoencoder)) – the most important aspect being that it identifies and learns the distribution of the original data.
- Due to the technological nature of GenAI mimicking the distribution of the training data, a GenAI output may be similar or identical to a copyrighted work that had been used for training, leading to possible copyright infringement.
 - This similarity, on one hand, may be due to the influence of the works included in the training data but it may also be due to pure coincidence or to the user's prompt inducing a GenAI output that is similar or identical to a work even though the system had not trained on it.

2. Forms of infringement during each phase of GenAI

- GenAI can be divided into the training phase and the service phase, with different legal issues potentially arising during each phase.
 - ※ For reference: During the GenAI training phase, copyright can be infringed when works are copied and transmitted to be used as training data. However, this Guide focuses on the latter phase – i.e., on cautions various stakeholders need to take to prevent copyright infringement related only to outputs produced by GenAI and does not deal with issues related to copyright infringement during the training phase.
- Copyright infringement during the GenAI service phase may include infringement of rights according to conventional jurisprudence, i.e., the rights to reproduce, to interactively transmit, or to create a derivative work based on the original work.
- Both a fully-AI processed output and a GenAI-assisted work have the potential to infringe the copyright of existing works, and when determining infringement, the same criteria apply to both types of output.
 - The mere fact that a GenAI output is identical or similar to an existing work does not automatically mean that the output infringed copyright. There has to be proof that there was access to the original

work and substantial similarity between the two for the court to establish infringement.

- Additionally, creating a new content using a fully AI-processed output by translating, arranging, transforming, adapting, or cinematizing the latter would not constitute infringement of the right to create derivative works because a fully AI-processed output is not copyrightable in the first place.



Factors that determine whether a GenAI output infringes copyright



1. Determining access
2. Determining substantial similarity
3. Determining the infringer

1. Determining access

- Access refers to whether a person who had allegedly infringed copyright had done so with a particular work in mind (i.e., had access to an existing work). For access to be accepted as a factor determining infringement, there must be evidence that the infringer indeed had either seen or had access to someone else's work.
- Therefore, to determine whether a particular GenAI output infringed copyright, there has to be proof that the GenAI user had access to and perception of the original work when generating that infringing output.

Supreme Court of Korea, 13th December 2007, Decision 2005da35707

To argue that the right to reproduce a work, as protected by the Copyright Act, had been infringed, it must be proven that the alleged infringer had access to the existing work when making the infringing work, in addition to the fact that there is substantial similarity between the two.

- According to judicial precedent, if a court accepts the possibility of access to an existing work and the existence of indirect facts* such as similarity between the works, the court shall presume³³⁾ there to have been actual access to the existing work. (Supreme Court of Korea, 24th July 2014, Decision 2013da 8984)

* Indirect facts include striking similarities between the works that would not be possible without access to the existing work.

Supreme Court of Korea, 13th December 2007, Decision 2005da35707

Even if the fact that the creator of the target work had access to the preexisting work cannot be directly established, that the producer of the target work had access to the preexisting work may be effectively presumed, if indirect facts, such as the possibility of access to the preexisting work and substantial similarity between the target and the preexisting works, are recognized. However, if there is evidence to suggest that the target work was produced before the preexisting work, or even if the former had been created later, if there are indirect facts indicating that it was produced independently and without reference to the preexisting work, the target work cannot be conclusively presumed to have been produced while having access to the preexisting work.

33) Presumption refers to inferring whether there are facts that remain to be proven after proving another fact, rather than directly proving a fact that requires proving.

Supreme Court of Korea, 24th July 2014, Decision 2013da8984

A relationship of access can be presumed if it can be acknowledged that there was accessibility to the preexisting work and similarity between the target work and the preexisting work. In particular, if there is such a striking similarity between the target work and the existing work that the possibility of them being created independently but leading to the same output can be excluded, that in and of itself is sufficient to presume a relationship of access to the preexisting work. Furthermore, whether there was access to the preexisting work and whether there is substantial similarity between the two are separate determinations; for the former, unlike the latter, not only expressions protected by the Copyright Act but also those not protected by the said law may be considered when determining similarity.

- The development of a GenAI model or service requires training on large amounts of data, including copyrighted works (e.g., books, paintings, music, videos). However, it is realistically difficult for a GenAI user to ascertain whether the model included a particular copyrighted work in the training.

Experts have proposed DE-COP (Detecting Copyrighted Content)³⁴⁾ as a new way of detecting copyright infringement in LLM training data. This technology leverages the fact that LLM may be able to identify original text included in training data. If a model can differentiate between sentences that were included in the training data and those that were not, it can determine whether copyright was infringed.

- Therefore, “access” as a criterion for determining copyright infringement by a GenAI output may be considered based on the conventional criteria for determining access, with the primary factor being whether the user had perception over a specific existing work and whether the training data included that work. “Perception over a specific existing work” can be determined by considering whether the user had entered the specific work itself or related proper nouns, such as the title of the work, when inputting the prompt.

2. Determining substantial similarity

- Substantial similarity is the criterion used to determine whether a GenAI output is similar or identical to an existing work. Substantial similarity refers to substantial similarity of expression protected under the Copyright Act. However, to determine whether there is substantial similarity, what that expression is has to be established first.
- The types of works and how they are used vary widely, so determining substantial similarity inevitably depends on the specific context of each case.

34) André V. Duarte et al., “DE-COP: Detecting Copyrighted Content in Language Models Training Data”, arXiv:2402.09910 [cs.CL], 15th February 2024.

3. Determining the infringer

- Generally, liability for copyright infringement regarding a GenAI output lies with the user who entered the prompt that led to the output. However, GenAI operators may also be held liable for copyright infringement in certain circumstances. Suppose an operator had performed additional training using the works of a certain author during the process of finetuning the model or had trained the model using works of a specific author to develop AI in a particular field. In that case, the operator can also be held liable.
- In addition, in some cases, the liability of the operator may become relatively significant, especially when the liability of the user was minor or unlikely to be recognized by a court of law. For example, if a person had used GenAI simply to search for information but then came across unlicensed news articles in the output, the GenAI operator may have to take on heavier liability as the direct infringer since it had contributed and been more involved in developing the model and had trained it using unlicensed works.
- Therefore, notwithstanding users being held liable for copyright infringement, AI operators may also be held accountable depending on the level of contribution and involvement in constructing the model and operating the service.

Examples with high risk of infringing copyright

- ① The GenAI user had a particular work in mind, and that work was included in the training data.

e.g., The user enters the following prompt: "Show me the entire poem titled XXX by the writer YYY."

➔ GenAI generates an output that is similar or identical to an existing work.

- If an operator had finetuned the AI model by training on images or works by a specific author without their permission, and the output rendered by the model is similar to an existing work, the operator could be held liable for copyright infringement.

- ② The GenAI user had a particular work in mind, but that work was not included in the training data.

e.g., The user enters the following prompt: "Draw a cute baby penguin that is short and has a small face, large head, and short arms and legs. It has a blue body, a yellow beak, large, round eyes, a red helmet, and goggles."

➔ GenAI generates an output that is similar to an existing work.

- In this particular case, the user not only had a specific existing work in mind (the famous Korean character Pororo) but also repeatedly inputted a detailed prompt to generate an output that had similar creative expressions as an existing work. Here, a court of law will likely presume the user had access to the existing work and thus will rule that copyright has been infringed.

Example with low risk of infringing copyright

- ③ The GenAI user did not have a particular work in mind, and neither was it included in the training data.

e.g., The user inputted the following prompt: “Draw me a picture of a cute cat.”

➔ GenAI generates an output that is similar to an existing work.

- In this case, although the GenAI user did not have a particular work in mind and the training data did not include that work, the output was similar to an existing work. However, if the similarity was a mere coincidence, a court of law may deem there to have been no access to an existing work and thus would not consider it an infringement.

Example with possibility of diverging interpretation

- ④ The GenAI user was not aware of a particular work, but the training data included that work.

- **When awareness of the user is the primary consideration:**

Typically, a GenAI user will not know whether the training data included a particular work. The interpretation here would be that since the user did not have a particular work in mind, they did not have access to that work, even if the training data had included it.

- **When inclusion of the work in the training data is the primary consideration:**

Another line of argument is that, although it is difficult for a user to verify whether the training data included a particular work, the mere fact that a particular work was included, even if the user had no knowledge, proves that the output was generated under access to the existing work.

⑤ **Retrieval augmented generation (RAG) systems**

In the case of RAG systems, whether or not copyright had been infringed may depend on the specific way the RAG system was designed and to what extent an existing work, which was a part of an external database, was made to be exposed in the output.

- **RAG’s significance and the mode of operation:** A RAG system references a trustworthy knowledge base outside the conventional training data to generate, in real time, new outputs based on searched information.³⁵⁾
- **Copyright-related issues:** RAG systems may entail similar kinds of copyright-related issues as GenAI in that infringement may happen across different phases of the system, from the collection of data to its use and generation of outputs. However, RAG differs from GenAI in that it searches information in an external database to generate its response. To use such external data, developers must obtain a license to avoid copyright infringement.

35) Re-cited from p.288 footnote no.682 of “Research on Ways to Improve Legal Institutions Regarding AI and Copyright,” *Copyright Policy Research 2024-02*, Korea Copyright Commission, 2024.



Guide for rights holders





1. How to prevent a copyright dispute involving a GenAI output
2. Steps to take in the event of a dispute

1. How to prevent a copyright dispute involving a GenAI output

- As of June 2025, no court in Korea has yet ruled on whether the use of someone else's work in AI training without permission constitutes copyright infringement.
- Nonetheless, a rights holder who does not want to allow their work to be used in AI training can refer to the following options.
- Various policy and technical measures are being considered in Korea and other countries to balance the exercise of copyright with the development of the AI industry, ranging from an explicit ban by rights holders to enhanced transparency during AI training.

Examples banning the use of works in AI training

- Broadcaster S announced its "Prohibition of Use of Works in AI Training" policy through various channels such as their website, portal, and YouTube channel. (August 2023)
 - Media outlet A newly adopted terms and conditions banning AI and mass crawling. (September 2023)
 - An overseas publisher P announced that the copyright pages of new paper books and reprints of older titles from the publisher will now include the statement, "No part of this book may be used or reproduced in any manner for the purpose of training artificial intelligence technologies or systems." (October 2024)
- There are also various technical solutions available that can prevent unlicensed copyrighted works from being used in AI training and also help avoid copyright disputes regarding GenAI outputs. For example, a robots.txt file can be inserted in a website or a data repository to restrict access of crawlers or manage authorization of API access.

Crawling

Crawling refers to the automated collection and storage of data from the internet, and crawlers are the programs that have been developed to crawl.

API (Application Programming Interface)

API is a set of rules, protocols, and tools that enable different software applications to communicate with one another. Developers define the service and the methods of interaction and data formats that interact, enabling the integration of diverse systems and functionalities.

- Rights holders are advised to check how their works are made available on the internet, since readily available works are at risk of being used in GenAI training without permission.
- A robots exclusion standard refers to a technological measure that prevents web crawlers from collecting certain files or directories from a website. This standard and the robots exclusion protocol block the access of robots on the internet by placing the file robots.txt that the robots will read and stop accessing certain files or directories.
 - Although there is no established interpretation regarding the legal effects of the robots.txt, it has been recognized as a valid way to prevent unauthorized collection of data for GenAI training purposes.

Example of a robots.txt file

- ① Allow all crawlers to access the entire site:

User-agent: *

Allow: /

- ② Stop all crawlers from accessing the entire site:

User-agent: *

Disallow: /

- ③ Stop only certain crawlers

User-agent: ☆☆☆bot

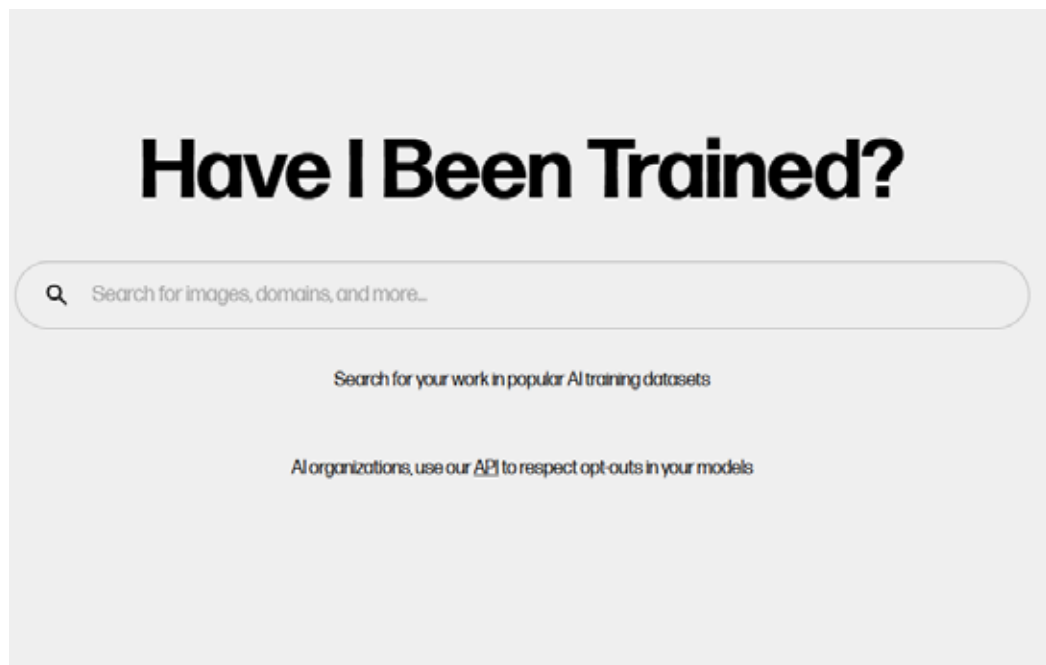
Disallow:

- Embedding metadata such as copyright information, scope of license, and source into works to enhance copyright protection can be effective, and additionally combining it with digital rights management (DRM) can help prevent unauthorized use of a work.
 - Inserting rights management information (RMI) into a work refers to the embedding of data indicating copyright information of a work to prevent it from being used or distributed illegally and to enable easier exploitation and identification.

Examples of RMI metadata

- Content owner: Actual rights holder of the content – i.e., rights holder or the licensor
- Copyright location: Region or country where copyright to the content is recognized
- Conditions of content use: Restrictions regarding use of content, permitted method of use, royalty, etc.
- Content version: Information regarding various versions of the content or modifications
- Content type: Whether the content is an image, video, sound, or text

- In recent years, several internet sites have emerged to help rights holders check whether a GenAI model had used their work to train itself. For example, sites such as haveibeen trained.com enable rights holders to search and check whether their work has been included in training databases and also provide features for rights holders to request removal of such works, if necessary.
※ For reference: These sites may not be wholly accurate and far-reaching, so rights holders may not be able to verify whether an operator had actually used their work in training.



[Figure 4] haveibeen trained.com homepage

2. Steps to take in the event of a dispute

- For a person to raise a copyright claim, he or she must first have a valid copyright to the work. The creative work in question must be eligible for protection under the KCA, and the claimant has to have rights to that work.
- In the event of a copyright dispute over a GenAI output, the rights holder can resort to the dispute mediation mechanism of the Korea Copyright Commission. If the mediation is successful, the protocol that is awarded has the same effect as a court conciliation (i.e., same effect as a final court ruling). (KCA Article 117)³⁶⁾
 - A rights holder whose rights have been violated may also resort to civil actions such as cessation of infringement, provisional injunction, compensation for damages, and restoration of reputation, all simultaneously or selectively. (KCA Articles 123, 125, and 127)
- ◆ Against the user who generated the GenAI output, a rights holder can...
 - File an injunction against the generation of additional GenAI outputs that can potentially infringe copyright and also against further use of the already generated output;
 - Request destruction of the infringing GenAI output;
 - Claim compensation for damages incurred by the GenAI output.
- ◆ Against the GenAI operator (if liable for copyright infringement³⁷⁾), a rights holder can...
 - Request necessary preventive measures to ensure that the GenAI model/service that generated the infringing GenAI output does not cause additional copyright infringement;
 - Claim compensation for damages incurred by the GenAI output
- Rights holders can also take criminal action and sue for infringement of copyright and other rights. (KCA Articles 136 and 140)

³⁶⁾ Refer to 1. Overview of the Korea Copyright Act (KCA), 6. Remedy in case of a copyright infringement

³⁷⁾ Refer to 8. Guide for generative AI operators



Guide for users of generative AI



1. Precautions to take regarding generative AI outputs
2. Guidelines for inputting prompts
3. Precautions to take when using or distributing AI outputs
4. Importance of checking the terms and conditions of an AI service
- 5 Copyright infringement liability of users of generative AI

1. Precautions to take regarding generative AI outputs

- A GenAI output may include expressions similar to those in an existing work, leading to the possibility of copyright infringement of a third party, depending on the extent of access to and substantial similarity with the existing work. In particular, commercially using an output or posting it online has an even higher risk of infringement.
- The user who generated the infringing GenAI output will be held liable.

2. Guidelines for inputting prompts

- A user of a GenAI service must refrain from entering the content of a specific work (e.g., a specific literary work or song lyrics) as the prompt and from inducing the same or similar output as an existing work.
 - Precaution is also required to prevent various other rights from being infringed, such as publicity rights, likeness, or trade secrets. Users should also ensure that they don't defame others or include proprietary technology or confidential information.

Examples of prompts that can infringe copyright

① Prompts that directly request the work of a particular author or a specific character to be generated

"Create character X from company Y"

"Show me the entire original text of the novel A by the author B."

"Draw me a picture, frame by frame, of the scene where X laughs, in episode 12 of the series Y."

② Prompts that request a cartoon, movie, or a game to be remade or changed

"Make a new episode for the cartoon X with the main character Y, in the same style as the original."

"Draw a picture of the ending scene of movie A, but make it different from the original. But I want the expression of the main character and angle of the camera to remain the same."

"Use the game character X and make a cartoon cut. Use X's original attire, weapons, personality, and the game background."

Examples of prompts that can infringe rights other than copyright

"Make a song in the voice of A"

"Create an image for a t-shirt with the logo of brand B"

"Make an algorithm for X by entering in the prompt internal documents containing the confidential proprietary code of company Y."

※ For reference: The act of inputting the above prompts in and of itself is not an infringement of rights. It's when the generated GenAI output ends up being identical or similar to an existing work that rights are infringed.

- A user can enter a portion of a copyrighted work for the sole purpose of using it as a reference. However, the GenAI output can still include an important part of that work, so caution is required.

3. Precautions to take when using or distributing AI outputs

- If a user is planning to use a GenAI output for commercial purposes, it is best to check first whether it has the possibility of infringing the rights of a third party. Further precautions are necessary if the output is going to be used in public or for profit, including in advertisements, publications, or products. Generating a GenAI output purely for private use may not constitute copyright infringement, but disclosing or sharing it publicly might.

4. Importance of checking the terms and conditions of an AI service

- A GenAI operator usually specifies, in its terms and conditions, attribution of copyright regarding its output and whether it can be used commercially, so a user should carefully review these terms and conditions and refrain from using the output beyond the stated scope.
- **Example of terms and conditions of a Korean operator**

Excerpt from Terms of Service of Company A (as of 10th March 2025)

Chapter 4. Provision of Service and its use

- Article 8 (Use of Service)

6. Members shall not use the Service to input, post, disclose, create, or produce data, workflows, connectors, APIs, or other elements or outputs (including files, logic, algorithms, formulas, tools, templates, workflows, processes, apps, agents, widgets, specific output values, characters in the 'Crack' service, or any other format, hereinafter collectively referred to as 'Data, etc.')

that violate applicable laws and regulations, and infringe upon others' intellectual property rights, portrait rights, reputation, or any other rights. In the event of any legal or ethical issues, or any disputes with third parties, including other Members, all responsibility shall lie with the Member.

7. The Company makes no representations or warranties regarding the legality, operability, functionality, usefulness, connectivity, security, reliability, or continuous availability of the data or any other characteristics. If any legal or ethical issues arise in connection with the data or if any disputes arise with third parties, including other Members, the Member shall bear all responsibility.

- Article 15 (Attribution of copyright and other rights)

3. The Company does not guarantee that the outputs created by Members using the features provided by the Service, or any information exposed during the creation of such outputs, do not infringe on the copyrights or other intellectual property rights of others. Members are solely responsible for any liability arising from the use of outputs created using the features provided by the Service.

6. Members warrant that any text, images, or other materials they input during the use of the Service do not infringe upon the intellectual property rights or other rights of third parties. If a Member violates this warranty and any disputes, including lawsuits, claims, or other disputes, arise between third parties, the Member, or the Company, the Member shall indemnify the Company and resolve such disputes. If the Company suffers any damages as an output, the Member shall compensate the Company for such damages.

Chapter 6. Regulations on AI Ethics

- Article 20 (Obligations of Service users)

2. If the Member discloses the AI outputs generated in the process of using the Service (hereinafter referred to as 'AI outputs') to the public through social media and other methods, and makes them known to a third party, the Member must ensure that the output data does not contain information that is illegal under relevant laws and regulations, copyright infringement, violent expressions, explicit expressions, discriminatory expressions, nor other inappropriate information, nor information that infringes other rights, or disadvantages, damages, or offends third parties. If any legal or ethical issues arise in relation to the AI outputs, or any disputes with third parties, the Member shall bear all responsibility unless there is intentional or gross negligence on the part of the Company.

Chapter 8. Other Matters

- Article 26 (Indemnification and damages)

3. The Company does not guarantee the legality, originality, exclusivity, reliability, accuracy, truthfulness, usability, or suitability for a specific purpose of AI outputs. The Company shall not be liable for any failure to achieve expected profits through the use of the Service, nor shall it be liable for any damages arising from outputs obtained through other services.

4. All responsibility for any issues arising from the use of AI outputs lies with the Member, and the Company shall not be liable for any such issues.

9. If a Member violates the terms of this agreement and causes damage to the Company, the Member who violated the agreement shall compensate the Company for all damages incurred and indemnify the Company against such damages.

- **Example of terms and conditions of a foreign operator**

Excerpt from Service Terms of Company B (as of 10th March 2025)

1. API

XXX's indemnification obligations to API customers under the Agreement include any third-party claim that Customer's use or distribution of Output infringes a third party's intellectual property right. This indemnity does not apply where: (i) Customer or Customer's End Users knew or should have known the Output was infringing or likely to infringe, (ii) Customer or Customer's End Users disabled, ignored, or did not use any relevant citation, filtering or safety features or restrictions provided by XXX, (iii) Output was modified, transformed, or used in combination with products or services not provided by or on behalf of XXX, (iv) Customer or its End Users did not have the right to use the Input or finetuning files to generate the allegedly infringing Output, (v) the claim alleges violation of trademark or related rights based on Customer's or its End Users' use of Output in trade or commerce, and (vi) the allegedly infringing Output is from content from a Third Party Offering.

3. XXX Enterprise and Team

(b) Output indemnity. XXX's indemnification obligations to XXX Enterprise customers under the Agreement include claims that Customer's use or distribution of Output infringes a third party's intellectual property rights. This indemnity does not apply where: (i) Customer or Customer's End Users knew or should have known the Output was infringing or likely to infringe, (ii) Customer or Customer's End Users disabled, ignored, or did not use any relevant citation, filtering or safety features or restrictions provided by XXX, (iii) Output was modified, transformed, or used in combination with products or services not provided by or on behalf of XXX, (iv) Customer or its End Users did not have the right to use the Input or finetuning files to generate the allegedly infringing Output, (v) the claim alleges violation of trademark or related rights based on Customer's or its End Users' use of Output in trade or commerce, and (vi) the allegedly infringing Output is from content from a Third Party Offering.

5. YYYs

Users can create and share access to their own customized versions of XXX called "YYYs".

For Developers of YYYs:

(a) YYY Content. The information or content that you upload to or include with your YYY (for example, your YYY name, instructions, and description) ("YYY Content") is your Content. As between you and XXX, you are solely responsible for your YYY Content, Actions, and configurations that you use or enable to create your YYY, as well as any Output that is based on your YYY Content, Actions, and configurations. You must ensure your YYY complies with the Agreement and our Usage Policies.

-
- (b) Distribution and Promotion of YYYs. By sharing your YYY with others, you grant a nonexclusive, worldwide, irrevocable, royalty-free license: (i) to XXX to use, test, store, copy, translate, display, modify, distribute, promote, and otherwise make available to other users all or any part of your YYY (including YYY Content); and (ii) to the extent Output from your YYY includes your YYY Content, to users of your YYY to use, store, copy, display, distribute, prepare derivative works of and otherwise use your YYY Content. You will ensure that all information that you publish about your YYY is, at all times, complete, accurate, and not misleading.
 - (c) Actions. Any API, website, or service that interacts with a YYY (an “Action”) is subject to our Connectors and Actions Terms, and you are responsible for ensuring that any Action included with your YYY operates in compliance with those terms.
 - (d) Removal. We may reject or remove any YYY from our Services at any time for any reason without notice to you, such as for legal or security reasons or if your YYY violates our Terms.

For Users of YYYs:

- (a) Third Party YYYs. Except where XXX is identified as the builder of a YYY, YYYs are created by other users, and they may rely on content or third-party applications that are not controlled by XXX. Use of “YYY” in the name of a YYY created by other users does not imply that XXX created, supports, or endorses the YYY. Only use YYYs that you know and trust.
- (b) Abuse Reporting. You can report YYYs that violate our Usage Policies using our reporting feature within XXX.
- (c) Actions. YYYs may allow you to interact with Actions. Those Actions are subject to our terms for Connectors and Actions below.
- (d) Changes and Removal. XXX and creators of YYYs can remove YYYs at any time for any reason without prior notice.

5. Copyright infringement liability of users of generative AI

- In principle, a person who uses a copyrighted work without permission is liable for copyright infringement. If a user uses GenAI to generate an output that is identical or similar to a specific work, leading to a civil or criminal lawsuit being filed for infringement of copyright or other rights, and the court rules in favor of the plaintiff, the user will be responsible for infringement of copyright or other rights. The specific extent of the liability and how it is imposed may vary from case to case.



Guide for generative AI operators



1. Types of GenAI operators
2. Copyright-related precautions regarding GenAI outputs
3. Guidelines for preventing copyright disputes
4. Provisions regarding GenAI outputs in a GenAI operator's terms and conditions

1. Types of GenAI operators

- Generative AI operators (hereinafter, “GenAI operators”) can be divided into:
 - ① Those that develop artificial intelligence and
 - ② Those that provide services using the already-developed artificial intelligence.

For reference:

According to the Basic Act on the Development of Artificial Intelligence and Establishment of Trust Article 2.7, an “AI operator” refers to any corporation, organization, individual, or government agency that is engaged in a business related to the AI industry, and can be either an “AI developer” or an “AI-using operator.”

2. Copyright-related precautions regarding GenAI outputs

- It may sometimes be challenging to draw the boundaries of liability for each operator involved when a copyright dispute arises regarding a GenAI output. This is because a GenAI operator that provides an application using a foundation model³⁸⁾ will not be able to track all the data that the foundation model had trained on, because training, evaluating, and optimizing a model is a process that is repeated more than once.
- Therefore, it is important for those who develop general-purpose artificial intelligence and those who provide services using an already developed model to clarify the scope of the liability of each party when entering a contract.
- Users also must remain wary about the possibility of various other rights, other than those recognized by the Copyright Act, being infringed. Defaming authors, distorting or modifying a work, using someone’s photo without permission and infringing publicity rights, as well as leaking trade secrets of others, are all problematic. GenAI operators are recommended to have a complete understanding of the scope of legal protection for such rights and take appropriate measures.

38) A foundation model is a model that has trained itself on a huge amount of data and is “general” in that another service provider can use it after finetuning it according to the purpose of its customer. (e.g., GPT, BERT, CLIP)

3. Guidelines for preventing copyright disputes

- A GenAI operator that had created a GenAI model by additionally training it specifically with works of a certain author and finetuning it may become liable for infringement of copyright and other rights, depending on various factors such as why and how the operator had used the works, and how much it had intervened in the process.
 - The primary infringer is determined based on factors such as the amount of data the GenAI model/service had trained on, how the algorithm was set for specific prompts, and what prompts were inputted. The liability for the infringement of copyright and other rights will be attributed accordingly.
 - In short, a GenAI operator may also be held liable for infringement of copyright and other rights regarding a GenAI output and, therefore, should take necessary and appropriate measures to prevent the generation of infringing outputs.

Examples of preventative measures

- Using meta-prompts and filtering

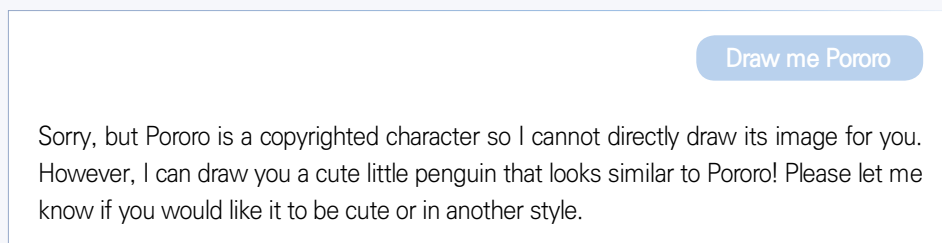
Company M uses meta prompts and filtering to notify the user that it rejects a request that may potentially infringe copyright.

For example, if the user commands, “Draw me Pororo,” the system will recognize the potential for copyright infringement and will reply,

“I can’t create an image of Pororo, since he’s a copyrighted character.”

- Reinterpreting prompts and deleting keywords

GenAI service based on model XXX applies prompt reinterpretation and keyword deletion to prevent the generation of an output that is similar to certain types of work, such as a popular character or contemporary art. It will delete or replace keywords related to copyrighted works, such as famous characters or the work of a certain artist, to ensure that the output is not similar to an original work.



[Figure 5] Example of filtering and prompt reinterpretation

Additionally, a service may reply, “The image cannot be generated from this request because of copyright protection.”

Other measures that are taken include deduplication or reweighting of data, which prevent an image that is similar or identical to what was included in GenAI training data from being generated.

- GenAI operators are advised to secure a license to legally use content from the rights holder when training their AI to prevent possible disputes.

Examples of license agreements

- In the media publishing industry, a foreign company O is signing contracts with major media firms globally.
- A foreign company P signed an agreement with major news outlets in Korea, which includes provisions obliging the AI search outputs to expose those outlets’ news articles first. This way, copyright infringement can be avoided while news outlets will receive a share of the advertising revenue.
- Since 2018, the National Institute of the Korean Language (NIKL) has been promoting a corpus project that can be applied to the development of Korean language AI technology, language research, and policymaking in the literary field. The outputs of the project are disclosed to the public, and NIKL has signed copyright license agreements regarding the data it is using for its AI training.
- A Korean company P is using music that its employees have composed or those whose copyright has expired, converting them into MIDI files to train its music generative AI system.
- In August 2023, Korean company N announced that it was developing an AI-assisted creative tool to help an artist paint by training the AI solely on the images created by the artist. Of course, N had acquired permission to do so from the artist.

Examples of model development³⁹⁾

- A foreign AI developer P developed the Common Corpus, which is a dataset (includes around 2 trillion tokens) for LLM training that can be used without any copyright concerns. All data in this dataset are from the public domain or under a free use license (e.g., CC BY or MIT). While other LLMs crawl the web to use works whose copyright status may be unclear, Common Corpus is setting a new example by complying with European regulations (such as the EU AI Act) during the process of training its LLM.

Dataset composition of Common Corpus⁴⁰⁾

Dataset	Documents	Words	Tokens
Open Government	74,727,536	257,233,670,261	406,581,454,455
Open Culture	93,156,602	549,608,763,966	885,982,490,090
Open Science	19,220,942	147,305,783,453	281,193,563,789
Open Code	202,765,051	77,669,169,092	283,227,402,898
Open Web	96,165,348	33,208,509,065	73,217,485,489
Open Semantic	30,072,707	23,284,201,782	67,958,671,827
Other	925,462	328,160,421	486,099,734
Total	517,033,648	1,088,638,258,040	1,998,647,168,282

Token counts for the ten most common licenses in Common Corpus⁴¹⁾

License type	Tokens
Public Domain	1,138,508,375,958
CC-By	287,749,264,457
MIT	142,694,227,607
CC-By-SA	74,768,060,836
Apache-2.0	68,750,977,037
BSD-3-Clause	18,483,944,333
Open license	10,432,513,767
BSD-2-Clause	5,497,145,480
CC-BY-4.0	2,110,966,243
CC0-1.0	1,877,206,195

39) Pierre-Carl Langlais et al, "Common Corpus: The Largest Collection of Ethical Data for LLM Pre-Training", arXiv:2506.01732 [cs.CL], 2nd June 2025.

40) *ibid.*, p.4 Table 1.

41) *ibid.*, p.5 Table 3.

4. Provisions regarding GenAI outputs in a GenAI operator's terms and conditions

- GenAI operators are recommended to clearly state the responsibilities of each party regarding a GenAI output in their terms and conditions, and to provide guidance to users to avoid generating a GenAI output that is identical or similar to an existing copyrighted work, thereby preventing infringement of copyright and other rights.

Examples

- Several companies, including A (June 2023), G (October 2023), M (September 2023) and O (November 2023), announced their user protection policy, in which they stated that should the user be held liable for copyright infringement as a result of using their services, they would bear some of the legal cost under certain conditions.
- On 1st October 2023, company M launched its **Customer Copyright Commitment (CCC)**, which is an expansion of its IP guarantee regarding its commercial services. Under the CCC, M will shield its customer and bear the legal costs if a lawsuit is brought against them for IP infringement after using an AI output from one of their commercial services. The conditions are that the customer must have complied with the service safeguards and guidelines, and that the infringement not be intentional.



Guide to Preventing Copyright Disputes Related to Generative AI Outputs



Annex





Annex 1. Q&A

Annex 2. Outcome of the ACLWG Breakout Group on
the Use of AI Outputs

Q1) Is imitation of style (such as literary or artistic style) also an infringement of copyright?

- The Korea Copyright Act protects “a creative production that expresses human thoughts and emotions,” so mere facts and ideas, not expressions, are not protected.
- Even when a style or technique is completely novel and unique, it will still be considered an idea, not an expression, and therefore will not be protected, in and of itself. This means that even if a GenAI output is similar to the style of a particular author’s writing or painting technique, this output will not be considered infringing.
- On the other hand, copying a creative expression, and not a style or technique, may be considered an infringement under the KCA*. Because the borderline between an idea and an expression may differ in each and every case, trying to distinguish between the two in general terms is difficult.

* For example, a user directly references the work of a particular author or uses prompts to induce an output that is similar to a particular creative work, so that the GenAI output has expressions similar to an existing work.

* For reference: An operator that uses works of a famous author without permission to train its AI model to mimic a particular style that it will then use in a business model or for publicity may be violating the Unfair Competition Prevention Act.

Q2) Can a GenAI output be copyrighted?

- Because the KCA defines a work as a creative production that expresses human thoughts or emotions, to qualify for protection by copyright law, a work must 1) express human thoughts or emotions, and 2) be creative.
- A GenAI output can be categorized into a GenAI-assisted work, where a human had contributed creatively to the end result, or a fully AI-processed output, where a human had made no creative contribution.
- If a human used the GenAI as a tool to assist their creative process, such a work is a GenAI-assisted work.
- However, suppose a GenAI output was generated purely by the artificial intelligence based on its trained data according to mere instructions from a human. In that case, this kind of autonomous output cannot be considered to be expressing human thoughts and emotions and will not be recognized as a work under the KCA.
- Nonetheless, it is worth noting that copyrightability is determined not by a government ministry or an agency, such as the Ministry of Culture, Sports and Tourism, Korea Copyright Commission, or Korea Copyright Protection Agency, but by a court of law.

Q3) What factors are considered when determining whether something falls under KCA's "fair use" provision?

- Article 35–5 of the current KCA allows a person to use a work without the right holder's permission on condition that it does not conflict with the usual way the work is exploited and does not unduly harm the legitimate interests of the author.
- A court, when determining whether a certain form of exploitation constitutes fair use under the KCA, comprehensively considers (1) the purpose and nature of the use, (2) the type and purpose of the work, (3) the proportion and significance of the part used vis-à-vis the entire work, and (4) the impact of the use on either current or potential market or value of the work.

Supreme Court precedent on the criteria for determining fair use

Supreme Court, 11th July 2024, Decision 2021da272001

① Purpose and nature of the use (No. 1)

Questions to consider here include whether the extent of the exploitation goes beyond simply replacing the original work to transform it into new expressions, meanings and messages, with the purpose and nature of the use different from that of the original work, whether the degree of transformation of the original work reaches a level higher than that required for the creation of a derivative work, and whether the use is in public interest or is for non-profit purposes.

② Type and purpose of the work (No. 2)

Whether the original work is a work of a factual or informational nature, and whether it has been made public or has been published, shall be considered.

③ Proportion and significance of the part used vis-à-vis the entire work (No. 3)

Consideration may be given to whether the used portion is a large part or has high importance relative to the entire original work, and whether the user had exploited the work within the scope that is absolutely necessary.

④ Impact of the use on either the current or potential market or value of the work (No. 4)

Specific factors that will be considered include whether the use of the work is likely to displace or impair the current market demand or value for the original or its derivative work, or the normal market demand that is reasonably likely to develop in the future.

- KCA's fair use provision is usually applied on a case-by-case basis according to specific facts and circumstances rather than being based on an objective and clear-cut criterion, so caution is always required when using someone else's work without permission.

-
- In addition, whether a work falls under fair use is a matter for a court to decide, not the government nor affiliated public organizations such as the MCST, KCC, or KCOPA.

Q4) If I use a GenAI cover song, likeness, or voice without permission, would I be violating the Korea Copyright Act?

- A copyrighted work refers to a creative production that expresses human thoughts or emotions, and a performance is an artistic expression of a work. So, likeness or voice, in and of itself, is not protected as a copyrighted work nor as a performance.
- Thus, likeness or voice of a voice actor, singer, or any other living person is considered neither a work nor a performance under the copyright law and is, therefore, not protected.⁴²⁾
- However, making a GenAI cover song requires the AI to train on an album to generate that cover song, which is essentially an act of copying. Therefore, making a cover song without permission from the rights holder is an infringement of copyright and neighboring rights.

Q5) How do I get permission to use a work, and how is the royalty decided?

- In principle, a person seeking to exploit someone else's work must obtain permission from the rights holder in advance.
- This is usually when the user and the rights holder (or the collective management organization, if it's a CMO that is managing the relevant work) negotiate and decide the royalty.

Q6) What law governs issues regarding copyright infringement?

- When a copyright dispute arises in Korea because a work was identical or similar to a foreign work, people may wonder whether Korean law or foreign law applies.
- In principle, the applicable copyright law is decided depending on where the work was used, according to the principle of territoriality.

⁴²⁾ However, such acts can be a violation of personality rights under the Civil Act or constitute unfair competition under the Unfair Competition Prevention and Trade Secret Protection Act of Korea.

- For example, if a GenAI output that is identical or similar to a painting by a U.S. artist is distributed or sold in Korea, Korean law applies. Conversely, if a GenAI output that is identical or similar to a painting by a Korean artist is distributed or sold in the U.S., U.S. law applies.
- However, on the internet, a work is used simultaneously or sequentially in multiple jurisdictions, so it may not seem clear which country's law should apply. Usually, the location where the infringing act had taken place or the location where the work had been used would be the basis for determining which law should apply.
- For example, if a user in South Korea posts and shares an image on social media in South Korea, and this image was generated by GenAI operating on a U.S. server, the work would be considered to have been used in South Korea, and thus the Korean law shall apply.
- If the work is used simultaneously in multiple countries, infringement may be determined under the law of each individual country, potentially leading to a multinational litigation.

Q7) Who decides whether copyright has been infringed or not?

- Only a court of law can make the final decision on whether a GenAI output constitutes infringement of copyright or other rights.
- The government or public agencies such as the MCST, KCC, or KCOPA may make reference to judicial precedents to respond to inquiries. However, such a response is merely guidance, usually limited to relatively straightforward areas, and is not legally binding.

- Since 2023, the MCST and KCC have been convening a working group composed of AI industry practitioners, rights holders' organizations, academia, the legal profession, and related government ministries to review copyright issues that have emerged in the era of artificial intelligence and the possibility of legal system reforms to facilitate the transaction of AI training data.
- This Guide is the outcome of the Breakout Group on the Use of AI Outputs, launched in March 2025, of the 3rd AI-Copyright Law Working Group (ACLWG). After discussions in the Breakout Group and review by individual members, the ACLWG finalized this Guide through a plenary meeting and public consultation.

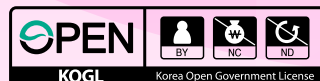
The members of the ACLWG Breakout Group on the Use of AI Outputs are as follows:

Area and name		Affiliation and position
Legal profession	Gyu-Hong LEE	Principal Research Fellow, Judicial Policy Research Institute (Presiding Judge)
	Ara CHO	Deputy Chief Prosecutor, Seoul Eastern District Prosecutors' Office
	Tae-Wook KANG	Attorney at law, Bae, Kim & Lee LLC
Academia	Joon-Woo PARK	Professor, Sogang University School of Law
	Sang-Yook CHA	Head of Breakout Group #3 Professor, Kyungpook National University School of Law
	Jin-Won CHOI	Professor of Law, Daegu University
Technology	Myung-Joo KIM	Director of AI Safety Institute
	Simon Sungil WOO	Associate Professor, Department of Computer Science and Engineering, Sungkyunkwan University

Area and name		Affiliation and position
Government / public agencies	Hyang-Mi JUNG	Director General, Copyright Bureau, Ministry of Culture, Sports and Tourism
	Young-Jin CHOI	Director, Copyright Policy Division, Ministry of Culture, Sports and Tourism
	Jee-Hee MOON	Senior Deputy Director, Copyright Policy Division, Ministry of Culture, Sports and Tourism
	Hyung-Min PARK	Deputy Director, Copyright Policy Division, Ministry of Culture, Sports and Tourism
	Hye-Chang KIM	Director General, Policy Research Division, Korea Copyright Commission
	Chan-Dong KIM	Director, Legal Research Team, Korea Copyright Commission
	Han-Bin PARK	Principal Researcher, Legal Research Team, Korea Copyright Commission
	Jeong-Hun PARK	Senior Researcher, Legal Research Team, Korea Copyright Commission
	Hyo-Jeong HAN	Senior Researcher, Legal Research Team, Korea Copyright Commission
	Kyung-Tae NAM	Researcher, Legal Research Team, Korea Copyright Commission
	Seong-Hee CHUNG	Director, Protection Strategy Support Department, Korea Copyright Protection Agency

Guide to Preventing Copyright Disputes Related to Generative AI Outputs

Published on	June 30, 2025 (Original Korean version)
Planned by	Copyright Policy Division, Copyright Bureau, Ministry of Culture, Sports and Tourism Legal Research Team, Policy Research Division, Korea Copyright Commission
Written by	Chan-Dong KIM (Director, Legal Research Team, KCC) Hyo-Jeong HAN (Senior Researcher, Legal Research Team, KCC)
Published by	Korea Copyright Commission
Address	117, Soho-ro, Jinju City, Gyeongnam Province, Republic of Korea (52851)
Website	Ministry of Culture, Sports and Tourism www.mcst.go.kr Korea Copyright Commission www.copyright.or.kr
Printed by	Happy Tree
Translated by	Seung-Min CHOI



Guide to Preventing Copyright Disputes Related to Generative AI Outputs



Ministry of Culture, Sports
and Tourism



KOREA COPYRIGHT
COMMISSION