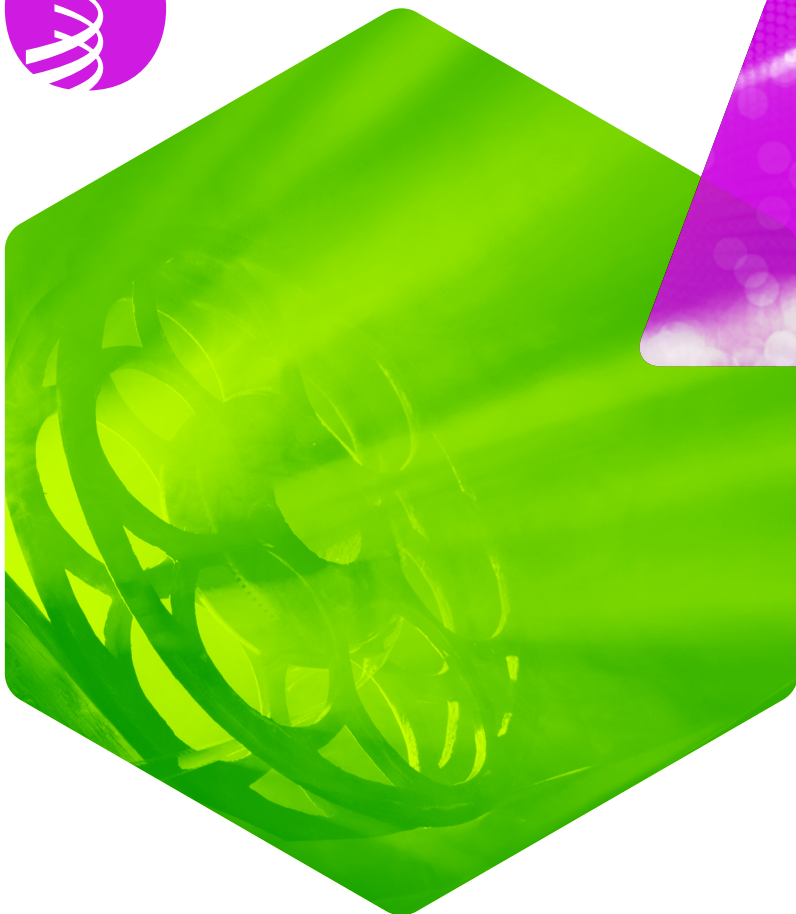


WIPO Conversation
IP and Frontier
Technologies

AI and IP Infrastructure



WIPO

What is IP infrastructure?

IP infrastructure refers to the systems, tools, and processes that power the modern IP ecosystem. For example, for copyright-protected works, such mechanisms enable digital content to circulate globally while helping to ensure that creators and rightsholders are properly identified and, where applicable, compensated. Infrastructure may enable innovative businesses that rely on creative works and help rightsholders protect their works against unauthorized uses.

Existing infrastructure includes:

- copyright registers administered by copyright offices to identify copyright-protected works and their owners;
- collective management organizations (CMOs) and their technical infrastructure for collecting data on the use of copyright-protected works and distributing compensation;
- international standards for generating unique identifiers for copyright-protected works;
- standards, such as the International Standard Content Code (ISCC) for identifying text, image, audio and video; and
- technical tools to discourage and identify infringements, such as watermarking and digital fingerprinting.

IP infrastructure is often developed to address the commercial requirements of specific industries, such as the Entertainment Identifier Registry (EIDR) for movies and television.



Effective IP infrastructure will be critical to human-centric AI policies that ensure a sustainable relationship between human creativity and innovation, and the development and deployment of AI.



An example of existing IP infrastructure: DDEX

The Digital Data Exchange (DDEX) is a standards setting organization in the music industry founded in 2006 with over 100 members across the supply chain of the industry, such as artist associations, record labels, digital retailers and music streaming services. DDEX was formed by copyright owners and distributors to develop the operating standards for the digital distribution of music. With its members, DDEX develops standards to enable information exchange across the value chain using shared message formats. This allows standardized communications about ownership, permissions, sales, etc., enabling services such as music streaming to track rights and payments efficiently.



Current IP infrastructure is specific to the differing business models and needs of various industry sectors. See the section “Further reading”, below, for more information about existing copyright infrastructure.


IP infrastructure and AI

The scale, capabilities and widespread adoption of current AI tools present new issues for IP law and policy.

Generative AI tools may be trained on billions of copyright-protected works and can output text, images, video and music that would previously have required a human author. In response to the development and deployment of AI tools, policymakers are considering how best to respond, including the options of creating new laws and related regulation. Any such new laws and related regulations will depend on existing and new IP infrastructure. Some existing IP infrastructure may be useful in the context of AI, perhaps with modification; but new systems, tools, and processes may also be needed.

While many of the current AI and IP debates center around copyright and highlight the need for copyright infrastructure to evolve in the age of AI, AI raises questions across the entire IP system and all IP rights.






Many stakeholders around the world believe that humans should be made aware when something has been generated by AI. This applies not only to creative works but also to designs, brands and innovations. Technologies such as watermarks or identifiers will have applications across different industries. As AI technologies develop, they are increasingly being used as a tool in the innovation process. Technical solutions that will allow innovators to track both human and AI contributions will become increasingly important in the future in the innovation industries.


The 11th session of the WIPO Conversation illustrated the role that infrastructure plays in the age of AI by focusing on the creative industries and the impact on copyright infrastructure from the development and deployment of AI tools.

This factsheet provides a high-level summary of some of the topics discussed.

 **The copyright infrastructure for AI can be divided into input-side and output-side, although some solutions may apply to both. These are discussed below.**

Collaboration and interoperability

While many copyright infrastructure tools and mechanisms currently exist and others have been proposed, much of the existing infrastructure has been developed for specific copyright-reliant industries, use cases, or geographic markets. Thus, current solutions may fail to address the broader challenges of the global scope of Generative AI and AI developers' interest in broad access to all types of copyright-protected works. Interoperability considerations are highly relevant to promoting consistent protection of creators' and copyright owners' rights and minimizing transaction costs while maximizing revenues.

 **To help address the challenges at the intersection of AI and IP infrastructure, WIPO will launch its AI Infrastructure Interchange (AIII, pronounced "A-triple-I"). AIII is a WIPO initiative to facilitate global dialogue on the technical and operational aspects of the intellectual property system in the context of artificial intelligence. For more information on WIPO's AI Infrastructure Interchange (AIII), see the "Next steps" section below.**


Copyright infrastructure for AI training and inputs

The material used by AI developers to train models, to fine-tune models for specific tasks and, when a model is run, to improve the accuracy and relevance of their outputs may include copyright-protected works. Infrastructure may help rightsholders control and monetize the use of their copyright-protected works and help AI developers access relevant and high-quality training materials.

Unilateral technical approaches to prevent scraping and use

When seeking to control scraping and training, rights holders can unilaterally seek to discourage or prevent their works being scraped from the internet or being used for training or as AI inputs. Such unilateral approaches include:

- watermarking publicly accessible versions of works;
- modifying works in ways not visible to humans but that disrupt the performance of AI trained upon them; and
- detecting and blocking web crawlers used by AI developers.

 IP infrastructure such as watermarking, “poisoning” of AI models, and blocking web crawlers and other approaches currently exist or are under development to allow rights holders to unilaterally take action against scraping and use of their works for AI.

Relevant to the application and utility of these types of unilateral approaches are broadly adopted international legal norms against the circumvention of technological measures used in connection with copyright-protected works and against the removal or alteration of so-called “rights management information” that includes data about the works. These norms are set out in the WIPO Internet Treaties and the domestic laws of many countries. The existence of these protections – created in response to the prior disruption of the creative industries by the Internet – and how they are seen to apply in the age of AI may directly impact the effectiveness of rightsholders unilaterally seeking to control how their works are used for AI training.

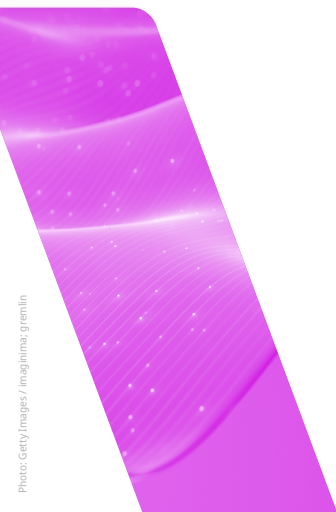


Rightsholder preferences

In addition to technical measures that seek to unilaterally prevent scraping or training, rightsholders often use expressions of rightsholder preferences to signal to AI developers whether their works can be used for training. In the absence of legal requirements to give effect to such expressions, AI developers voluntarily choose whether to comply with preferences.

When rightsholders express their preferences, approaches include opting “in” or “out” of use of their works from copyright exceptions or specific AI-related activities, such as use for training or as inputs to AI models. In general, copyright grants exclusive rights that require authorization from the rightsholder before the work can be used in specific ways (such as copying). But there are exceptions and limitations under national copyright laws, such as for text and data mining, that allow for certain uses of copyright-protected works unless the rightsholder “opts out” in a machine-readable format. Moreover, there are scenarios, such as the use of Creative Commons licenses or voluntary collective licensing, where a rightsholder chooses to “opt-in” to uses, even without a direct authorization to the user of the work. Whether and how such opt-ins apply specifically to the issues raised by AI training and inputs is the subject of active discussion.


Given that AI models, especially Generative AI, may be trained on billions of copyright-protected works, the magnitude of the rightsholder preferences that may need to be collected, stored and checked is a technical challenge to be considered.





Considerations for technical solutions that have been raised include:

- the ease of use for rightsholders and those checking preferences, such as AI developers, including whether preferences are machine-readable (and what that requires in practice);
- the cost of collecting, storing and checking preferences, such as the cost of processing works to generate a unique identifier;
- whether preferences are expressed per website, webpage or copyright-protected work, also referred to as location-based or asset- or work-based;
- how works are identified, such as via an algorithm producing a unique identifier;
- the extent to which such unique identifiers can be consistent across copies and extracts of the same work;
- how to deal with conflicting expressions of preference attached to different copies of a work;
- the granularity of preferences, such as a general opt-in/out, preferences according to AI model type and use case (such as medical research) and/or licensing options and contact details;
- whether the technical approach works for all forms of copyright-protected works, such as text;
- whether preferences are stored with the website, webpage or work and/or in a separate register (e.g., in case information stored with the website, webpage or work is lost when a work is reproduced); and
- the technical feasibility and likely adoption of national or international registries of preferences and/or general or industry-specific registries (such as a registry for the music industry)

 Giving full effect to rightsholder preferences at the scale needed for Generative AI – flexibly dealing with billions of preferences in easy and cost-effective ways – is a significant IP infrastructure challenge.

Transparency of content used in training

Rightsholders have raised concerns that without accurate information about the content that is used to train AI models, it is impossible for them to adequately protect their rights. Since models may have been trained on billions of works, providing granular information on works used may require infrastructure and technological solutions. Such solutions may also seek to address AI developers' concerns over trade secrets. Information may be provided in the form of automated responses to queries about whether specific works were included in a training data set based on using a unique identifier of the works, for example, such as the International Standard Content Code.

Rights management, licensing and delivery


Rights management, licensing and delivery infrastructure may help AI developers use:

- copyright-protected works that cannot be scraped (e.g., because they are only in private archives);
- copyright-protected works that could be scraped, but with better quality (such as image resolution), in different file formats (e.g., music with each voice and instrument recorded separately), with additional metadata, with greater speed and convenience (e.g., via a dedicated delivery system, such as an API), and/or with greater certainty over the right to use the works; and
- copyright-protected works that have been opted-in to particular licensing schemes.

Infrastructure solutions may have a role in supporting licensing itself (e.g., negotiating and tracking the right to use the licensed copyright-protected works) and/or delivering licensed works (e.g., electronic delivery of an archive of work or enabling on-demand access to works via an interface, such as an API).

Solutions are being developed by a variety of stakeholders, including:

- individual rightsholders (such as newspapers and music rightsholders);
- collective management organizations representing multiple rightsholders who are developing/offering collective licenses for AI training, inputs and/or outputs;
- for-profit technology companies offering solutions for bilateral licensing or marketplaces of aggregated content; and
- national government initiatives seeking, for example, to stimulate national AI investment and development or to encourage the development of AI reflecting their national culture.

 **Infrastructure approaches may overlap. For example, a particular method of watermarking a work may or may not be suitable for other infrastructure aims, such as discouraging unlicensed training or labelling AI outputs.**



Copyright infrastructure for AI outputs

Generative AI can output text, images, video and music that would previously have required a human author. These outputs may compete with human works in terms of speed, volume, cost and personalization. There are also debates over whether some, or all, AI outputs infringe copyright-protected works used to build or operate generative AI tools. Infrastructure may help identify AI outputs, compensate rightsholders for the contribution their works made to specific outputs, and mitigate the risks of outputs that infringe copyright-protected works.

Identifying and labeling of AI outputs and works modified by AI

Identifying AI outputs may be desirable for multiple reasons, including:

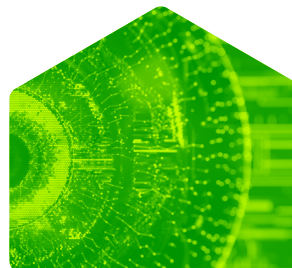
- enabling user choice over whether to consume human works or AI outputs;
- identifying which works are entitled to copyright protection;
- identifying potential deepfakes; and
- filtering AI outputs from the training materials for AI models to avoid degrading their performance (e.g., so-called “model collapse”).

Infrastructure solutions may include methods to:

- label AI outputs;
- register AI outputs;
- record the provenance of works; and
- record changes made to works, including by AI.



Labeling of AI outputs may be useful to copyright registries, collective management organizations and businesses, such as music streaming services, that seek to identify copyright-eligible works (excluding AI outputs) and compensate rightsholders. Labeling for one purpose might require a different solution than for another purpose.



Attribution

Attribution technology seeks to identify works that have contributed to an AI output. This may apply, for example, to works used in training or fine-tuning the model or to works added to prompts at when the model is run (such as in a “Retrieval-Augmented Generation” or “RAG” approach). Such technology may be the basis for calculating compensation for specific rightsholders. Efforts to develop reliable and accurate attribution technology are currently under development, but there are substantial challenges, for example, in precisely attributing which of the billions of pieces of training data a particular output is based on or the weight that individual works have had in creating a particular output.

Unlearning and guardrails

Training AI models may be extremely costly, especially in the case of “foundation” models, which are typically trained on billions of works and can be adapted for many uses, such as tasks involving text or image generation. There is therefore interest in technical solutions to minimize harms where a model has been trained on something undesirable, including copyright-protected works if these should not have been used. Areas of research include methods to enable a model to “unlearn” training examples and “guardrails” that may prevent infringing outputs, by filtering prompts likely to trigger such outputs, for example, or by detecting and suppressing the unwanted outputs.



Further reading

The WIPO Conversation is the leading global forum to facilitate discussion and share knowledge among all stakeholders on the impact of AI on IP.

More information about the 11th session of the WIPO Conversation on AI and copyright infrastructure, including the program, presentations, summary and webcast, can be found on the meeting page. More information about IP and frontier technologies is available on the [WIPO website](#).

WIPO recorded a series of 30 webinars in 2021–2023 on topics relevant to copyright infrastructure, such as metadata, identifiers, technology solutions and WIPO services. These can be found at www.wipo.int/meetings/en/topic.jsp?group_id=346.

Next steps

You can read more about WIPO's AI Infrastructure Interchange (AIII), watch a video podcast series about it and find more details at www.wipo.int/meetings/en/2025/ai-infrastructure-interchange.html.

To keep informed about WIPO's AI work, including news about AIII and the next session of the WIPO Conversation, [sign up for the IP and Frontier Technologies Division's newsletter](#).



© WIPO, 2026 /  Attribution 4.0 International (CC BY 4.0) / The CC license does not apply to non-WIPO content in this publication.
Cover: Getty Images / Marinkazumovskaya, AleksSecrey / WIPO Reference RVZ026-ZEN; DOI: 10.34697/rind49599

wipo.int