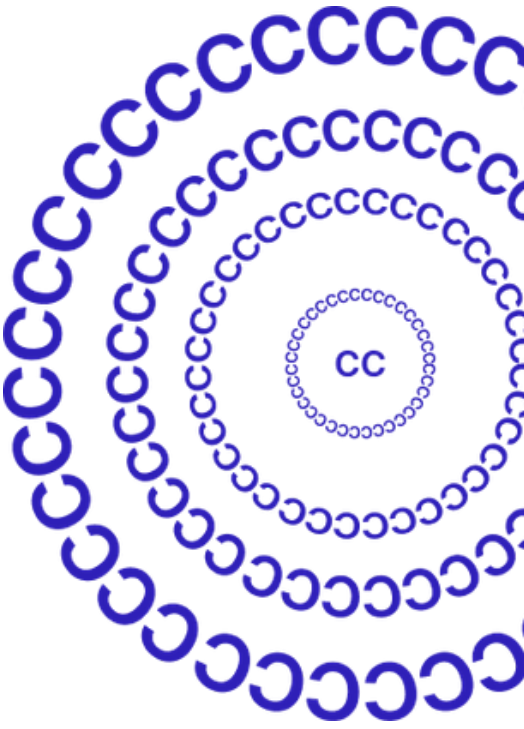
Guidance on CC Licensing in the Age of AI

DOI: 10.5281/zenodo.22261032

Guidance on CC Licensing in the Age of AI

Table of Contents



Introduction

Since 2002, Creative Commons (CC) licenses have served as legally enforceable tools for digital sharing. Beyond their functionality, they also signal a commitment to open sharing in the public interest.

CC's general guidance has been to acknowledge and support all valid uses for each of our current six licenses and two public domain tools.

When making decisions about which CC tool to use, we've long encouraged licensors to share their work under the most flexible CC license or tool applicable, while recognizing that some contexts call for added restrictions. This practice fosters robust knowledge and cultural exchange, which in turn strengthens our collective commons. Access to educational, scientific, and cultural content remains essential to society.

However, that guidance was developed for a world of reuse by people, a premise that doesn't fit as neatly in a world of widespread machine use of content through AI. As Sarah Hinchliff Pearson noted in this blog post, "In the age of AI, we all necessarily feel small. A lifetime's worth of creative output becomes a microscopic portion of data in a system functioning at machine scale.... In this unsatisfactory binary, many are choosing to close off, even those who have historically believed in the power of sharing knowledge." When more people choose to close off knowledge and culture, we all stand to lose.

This guidance addresses how the current suite of CC legal tools apply—or not—in a world where AI systems both rely on and produce knowledge while increasingly acting as an intermediary to original sources. We start with an executive summary of our general guidance and include a note on terminology before examining the broader content and underlying assumptions of this work. We then provide additional context for and links to our existing licensing guidance across three sectors that are among the heaviest users of CC legal tools: open culture, open education, and open science.

Executive Summary

Our general guidance is to:

- **Use the most permissive and flexible CC license that meets your goals for sharing the work.**
- **Understand that more restrictive CC licenses may limit the usability of your work, including (and especially) for human reusers, while doing little to prevent AI usage.**
- **Mark your work clearly with a full and complete licensing statement that clarifies the title, author, source, and license for the work.**

There are several important considerations that inform this guidance. We list them briefly here and provide additional context for them in later sections of this report.

- CC licenses are copyright tools, and they are only invoked where copyright applies.
- There is a perception among AI developers that use of CC-licensed content is generally permitted for AI development, including at times without complying with the applicable CC license conditions. This perception may be based, in part, on the premise that CC licensors are more likely to be comfortable with use of their content in this way. However, the current CC legal tools predate mainstream AI, and most openly-licensed content was released in a pre-AI era. Most licensors could not have anticipated potential AI uses when they chose their license and shared their work.
- Reactions to AI vary widely, but licensors across the commons are increasingly turning to more restrictive licenses in their attempts to deter AI use. Sharing content under more restrictive terms results in a more limited pool of public resources for all of us.

- CC continues to value open sharing in support of a robust commons, and we will continue to balance licensors' choices in sharing their works with the public's rights and interests to access and reuse.

In brief: We encourage people to share openly using CC licenses and tools to help achieve their goals and values, rather than viewing the use of particular CC licenses as a way to impede AI use.

This guidance does not substantially differ from what we've given in previous resources; however, we think it is important to reiterate and further explain how we got here and to contextualize that guidance against the rise of AI. This document is reflective of where things stand today, and we recognize that the broader legal, cultural, and technological context will continue to change.

We hope this guide serves to clarify and support the valuable educational, cultural, and scientific work that is produced across the globe, every day. Our goal is to continue to balance public interest and licensor choice, providing options for a vast and complex world of sharing.

Notes on Terminology

We use "AI," "generative AI," and "AI training" rather broadly here, as a general shorthand for what we know is a complex field of technologies and practices. In some cases, both our guidance and licensor concerns apply not just to AI training but to the use of CC-licensed content throughout the AI value chain more generally. Defining and categorizing AI and machine use of the web is difficult, as there are a wide variety of goals and purposes, technical methods, economic stakes, and cultural implications involved. We thus generally refer to "AI use" to capture the types of uses that intersect with CC licenses. We gloss over those details here, though we have outlined some of them previously in our Categorizing Machine Use of the Web Issue Brief.

This document considers only the licensing and sharing of human-generated (not wholly AI-generated) content and its use in AI. We understand that there are many other considerations on how copyright and/or CC licenses apply to: content that has been wholly or partially generated by an AI tool; AI generated outputs that infringe copyright; and AI models themselves, including model weights, data, and metadata. These questions are beyond the scope of this guidance, though you can see additional thoughts on this topic in our Copyright and Generative AI Issue Brief.

Additionally, we discuss data and datasets throughout this guide, and it is important to contextualize what we mean by those terms. A dataset that accompanies a research article, for example, is inherently different from a dataset that contains a collection of cultural heritage materials, but both may be treated similarly for AI use. In a world where all content is now being treated as "data," it is important to understand that there is significant nuance in the licensing recommendations we would make, depending on what type of content the dataset actually contains.

A discussion about the copyrightability of data and datasets is also beyond the scope of this guide, but you can read some preliminary

thinking about this subject on our FAQ page. Widespread collection and usage of data—including everything from empirical measurements (temperature, land area, mass, etc.) to entire creative outputs like poems, novels, and artworks—involves many considerations that extend beyond copyright, such as ethical concerns regarding privacy and confidentiality, plagiarism and academic integrity, and bias both within the individual components and overall composition of a dataset, and the way that it is used.

Finally, it's worth noting that our conclusions here are subject to change based on future legislation or forthcoming court rulings about AI and copyright in various jurisdictions, which may then offer a clearer global picture. However, that will take several years, if it happens at all. This guidance reflects our best understanding of the copyright and AI context as of the date of publication.

Broader Context

A few key considerations have informed our thinking.

First: **CC licenses are copyright tools, and they only apply where copyright applies. If copyright does not apply to a particular AI use, neither do CC license conditions, so using a more restrictive CC license in an effort to prevent it is not effective.**

CC licenses grant permissions under copyright law and only have force where copyright applies. In many jurisdictions, such as the EU, Japan, and the US, many forms of AI use currently fall under existing exceptions and limitations to copyright, meaning the uses are allowed and not infringing. In these situations, the CC licenses (and their conditions) do not apply to the use.

Second: Because of their long history of being associated with permission for reuse, there is a perception among AI developers that reuse of CC-licensed content is generally permitted even without fulfilling license conditions like attribution and non-commercial (with many commercial AI systems using NC-licensed content). While perceptions alone do not necessarily dictate legal interpretation, we recognize that their impact shifts common understanding and subsequent behavior, both among reusers and licensors.

We recognize that releasing content under a CC legal tool does not automatically imply that the content was *intended* to allow AI use, regardless of legal arguments and interpretations. While the current suite of CC legal tools was legally designed to be technologically neutral, these legal tools predate mainstream AI, and most openly licensed content was released in a pre-AI era. As such, most licensors could not have anticipated potential AI uses–or the novel ways those uses would interact with copyright–when they shared their work under their chosen CC license.

Reactions to AI continue to vary widely. Some licensors are questioning if they should publish or share content at all, as they may feel choosing not to

share is the only way to ensure no interaction with AI. Other licensors are embracing AI as a tool in their content development process, and/or they are actively and intentionally sharing their content for AI use. Some do this in an effort to provide AI systems with more accurate and representative data, or they may believe that the benefit of sharing is more important than the risks. These licensors may also not feel strongly about differentiating between "human" and "machine" use of their content.

Regardless, licensors across the commons are increasingly turning to more restrictive licenses in their attempts to deter AI use. Some educators, researchers, authors, and publishers are applying more restrictive CC licenses, or even moving back to "all rights reserved" copyright restrictions, for works they previously would have released under a less restrictive license. We've seen this happen across multiple sectors, not just those we examine here. This appears to be a direct reaction to AI ingestion of published works that is often done without appropriate attribution or reciprocity.

The retreat toward restrictions has real costs. **When people share content under more restrictive terms, the result is a more limited pool of publicly accessible resources (or *commons*) for all of us.** This means less material is available for researchers, educators, archivists, and the public interest uses that depend on open access.

With that in mind, decisions around open sharing must not be taken lightly. Given that copyright exceptions often mean CC license restrictions don't apply to AI use in the first place, using a more restrictive license isn't effective in preventing it. In fact, a more restrictive license may prevent the exact types of human sharing and reuse licensors are hoping to allow, while at the same time putting no restrictions against AI use. As a result, we continue to guide communities to share as openly as they can, using CC licenses and tools to the extent that it is consistent with their values and goals.

CC has always tried to balance licensor choice with the public's rights and interests to access and reuse. We are in a moment where that balance is

being tested. We've faced profound shifts in the copyright ecosystem before. Each time, we've evolved our tools, policy positions, guidelines, and educational resources in order to keep meeting the commons' needs. That evolution takes time, and in the interim, we aim to provide licensors with clear, practical guidance that they can apply right away.

Guidance in Specific Sectors

In alignment with our ongoing work across Open Culture, Open Education, and Open Science, we continue to provide sector-specific guidance to these communities on how to share openly.

Open Culture

In the Open Culture sector, CC licensors and content stewards include cultural heritage institutions such as libraries, archives, and museums, as well as community and informal collection holders.

> **The CC licenses and legal tools that are most consistent with the goals of optimal engagement and reuse for cultural heritage content include: CC0, the Public Domain Mark (for public domain works worldwide), CC BY, and CC BY-SA.**

For more details about the specific licenses we recommend in specific contexts, please see Recommended Licenses and Tools for Cultural Heritage Content. The Open Heritage Statement also provides general principles to guide equitable access to heritage, including through standardized tooling such as CC licenses and tools.

Open Education

There are a range of stakeholders in the Open Education ecosystem, including learners, teachers, content creators, content stewards, government and philanthropic funders, and corporate entities such as publishers and learning management platforms.

> **The CC licenses and tools that are traditionally and most commonly associated with open education include: CC0, the Public Domain Mark, CC BY, CC BY-SA, CC BY-NC, and CC BY-NC-SA.**

For more details about our recommendations in Open Education, please see previous resources, such as a clarification of the 5Rs of OER and

which licenses support the permissions to retain, reuse, revise, remix, and redistribute educational materials, and an overview of how OER can meet global needs and make lasting impact. We have also developed resources in partnership with other organizations working in open education, including: the 2019 UNESCO Recommendation on Open Educational Resources (OER) (and accompanying video), and the *7 Things You Should Know About Open Education* series published in collaboration with the EDUCAUSE Learning Initiative, where we outline key aspects of open educational content, practices, and policies.

Open Science

In the Open Science sector, CC license holders and users include scientists, researchers, academic institutions, publishers, government agencies and ministries, intergovernmental organizations, and funders of scientific research.

> **The CC licenses and tools that we recommend for open science include: CC0 (with attribution requested) and CC BY.**

For more details about our recommendations in Open Science, please see previous resources such as Licensing Best Practices for Sharing Scientific Data.

Reflections and Next Steps

The future of the commons depends on how we redefine open, public access. The challenge is not simply to preserve science, culture, and knowledge, but to provide access that is meaningful, equitable, and capable of sparking curiosity and creativity to express our shared humanity. Some forms of AI may create new opportunities for discovery, accessibility, and research, while others (particularly mainstream large language models) raise concerns about the environment, the information ecosystem, and labor and human rights. But changes in licensing practices that restrict open, public access to content have considerable negative repercussions on the commons and human progress, while doing little to address the harms of AI.

In today's digital landscape, it can feel like we have only two options: share in a system that is essentially a free-for-all, where "open" means anything goes, or close off access and reuse across the board. This binary is unsatisfactory and unsustainable.

Creative Commons is working to create alternatives that sit between these two extremes. We think there should be standardized tools for sharing knowledge in ways that reflect the evolving spirit and values of the open movement. We're building this through CC Signals, a framework for communicating expectations around AI use of content or data. Our approach to AI is intended to respond to the valid concerns that many have about unfettered machine reuse where sources aren't recognized and where benefit isn't shared broadly with the commons. Our existing legal tools are one important part of the infrastructure of open sharing, and they have been for 25 years. We invite you to observe and work alongside us as we develop new tools to help licensors more clearly communicate not only the copyright permissions that accompany their shared content but also their intentions and expectations behind creating and sharing it with both humans and machines.