Creative Commons Position Paper on Preference Signals

Prepared for IAB Workshop on AI-CONTROL (aicontrolws), submitted 2024-08-02

When Creative Commons (CC) was founded over 20 years ago, sharing on the internet was broken. With the introduction of the CC licenses, the commons flourished. Licenses that enabled open sharing were perfectly aligned with the ideals of giving creators a choice over how their works were used. Those who embrace openly sharing their work have a myriad of motivations for doing so. Most could not have anticipated how their works might one day be used by machines.

CC has supported people participating in the commons by enabling more permissive use and reuse than default copyright would otherwise allow. Building atop the framework of copyright rules that grant a bundle of exclusive rights to the creator, the aim of CC's approach is to reject the all-or-nothing framework and create options for sharing that reflect a more generous and collaborative spirit than default copyright. The CC licenses exist on a spectrum of permissiveness, all underpinned by the goal of enabling access to and sharing of knowledge and creativity as part of a global commons, built on mutual cooperation and shared values.

Generative AI is raising new questions about the use of existing content. Many of the largest generative AI models have crawled significant portions of the publicly accessible web to collect training data, including CC-licenced works. While some of the resulting uses of AI could be positive for society, and some creators are using and embracing generative AI tools, there is a general lack of transparency about what content has been used, how the training was conducted, and what the models will ultimately be used for. Many creators, including those that have shared their works under CC licenses and tools, are understandably feeling disempowered as they question the use of their content for purposes of AI training.

There is evidence that CC licensed content is often preferred as training data, given their attributes of openness and the idea of "free" use often associated with them. This may not necessarily be consistent with creators' intention in openly sharing, especially when that sharing took place before the public launch and proliferation of generative AI models.

On one hand, If someone uses a CC-licensed work with any new or developing technology, and if copyright permission is required, then the CC license allows that use without the need to seek permission from the copyright owner so long as the license conditions are respected. This is one of the enduring qualities of our licenses — they have been carefully designed to work with all new

technologies where copyright comes into play. On the other hand, it has always been possible to imagine use cases for a CC licensed work that could make its creator feel uncomfortable or want to object. Bad actors and unintended consequences are not new and are generally offset or at least mitigated by the overwhelming positive impact of open sharing. What is new with generative AI is that the unanticipated uses are happening at scale. With concentrated power there is a risk of concentrated benefits and creators are questioning anew whether the bargain is worth it. The creative works are being used outside of their original context in a way that does not distribute any of the usual rewards back to the creator, either financial or reputational.

Arguably, copyright is not the right framework for defining the rules of this newly formed ecosystem. As the CC licenses exist within the framework of copyright, they are also not the correct tools to prevent or limit uses of content to train generative AI. Nor do we think blunt instruments such as broad opt-outs (creators choosing to prevent their content from being used for AI training) are the answer. Based on our community consultations to date, we believe a new suite of tools that complement the CC licenses are required to communicate sharing preferences for AI training.

Preference Signals

In our consultations to date, we have uncovered many of the limitations of using instruments such as robots.txt as an indicator of opt-in or opt-out for generative AI training. In many cases, robots.txt and a website's terms of service are inconsistent, and robots.txt is a limiting protocol when it comes to creator content in the commons as a public good (including, but not limited to art, culture, science, journalism, scientific data). Further, approaches that propagate the limiting binary of blunt instruments of opt-out do not take into consideration the values and social norms embedded in sharing content on the web. CC's approach is to develop and advocate for tools that empower creators and contribute to a healthy and ethical commons for the public good.

Preference signals for AI are the idea that an agent (creator, rightsholder, entity of some kind) is able to signal their preference with regards to how their work is used to train AI models. Preference signals would represent a range of creator preferences, all rooted in the shared values that inspired the CC licenses. At the moment, preference signals are not meant to be legally enforceable. Instead, they aim to define a new vocabulary and establish new norms for sharing and reuse in the world of generative AI.

For preference signals, or any other approach to communicating or controlling how content is used in generative AI training models, to be successful, our community consultations to date posit that the system must:

- Address the lack of transparency within AI models with regards to when and how creator content is being used;

- Center human labor and creativity and consider if there is a specific set of activities that need to be protected in the process of human creating and sharing;
- Prioritize AI for the public good and address the risk of a small minority of players capturing the benefits. Increased public investment and participatory governance of AI should be explored as a meaningful part of the solution. Safeguarding public AI infrastructure is also an act of future-proofing and preserving the commons to enable pro-social organizations building the capacity to train AI models to use the commons as training material.
- Support the development of policy frameworks that foster participatory governance. Reliance on commercial players to set forth industry norms that influence the future of the open commons is imprudent and ill-advised

Use Cases and Other Considerations

Blunt-opt out instruments that are largely enacted opaquely from creators ignore a creators' right to communicate how they wish to share their content. Like the CC licenses, a more nuanced approach that lives on a spectrum of choice also becomes a tool to communicate individual values and informed consent. Use cases identified to date show the need for individual, sector, and regional and cultural considerations.

Content creators have a range of uses cases for preference signals, including but not limited to:

- As a creator, I don't want to be part of a corpus
- As an artist, I don't want AI to be able to copy my exact style or deliberately make an exact copy
- I want to profit from this use, I want my work included so I get a cut of the future profits
- As a creator, I want to act with informed consent, especially as conditions of use change, e.g. "Now your content is being used for this new thing - do you agree?"
- As a user, I want to grant a specific trusted organization the ability to choose preferences on my behalf
- If my data is going to save a life, I would like it to be used for AI training. And the opposite: if it will be used to take away lives, I refuse that it be used for AI training.
- I don't want a commercial entity to benefit from using my content.
- I want attribution.
- I want to be anonymous.

General users also have a range of use cases, including: wanting to ensure they are fairly using the content of others (e.g. not wanting to create harm or job loss), the assurance that any content or datasets that they are using are reliable, and more. More specifically, user types such as researchers may need comprehensive access to datasets and AI training models as well as being able to confidently attribute the data they are using and to understand its provenance. Use cases for sectors

including scientific research, cultural institutions, education, and more, all around the world point to the undeniable need for choice beyond opt-out.

Next Steps in Exploring Preference Signals (and other mechanism of AI control)

We are still in the research phase of exploring preference signals and recognize the importance that we carry forth these explorations in tandem with the many others who are exploring and building responsible approaches to generative AI. The below steps and considerations are equally important for AI controls more generally and in the consideration of opt-out approaches.

- We need clarity on who we are aiming to serve with AI controls generally, whether opt-outs or a spectrum of choice;
- We must ensure that our plans identify and build for global needs, communities, and within varying copyright jurisdictions;
- We need to better understand who most benefits from blunt opt-out approaches. Is it large, commercial rights-holders and large tech companies who can afford to license? Is there a risk that commercial rights holders will use opt-opt as a negotiation strategy?
- Will opt-ins (or preference signals) be irrevocable? Is it reasonable for creators on the internet to opt-out completely?

As generative AI models evolve and the long-term implications of including content in AI models remain unclear, there is a risk that creators will become less inclined to share in the commons. If they do share, they may hesitate to do so under open licenses. Regardless of the exact approach or approaches of AI control, collectively, we must begin to define a new vocabulary and encourage a new set of norms for sharing and reuse in the world of generative AI. This is not solely about individual creator control. It is about collectively defining a new social contract for the digital commons in an era shaped by AI technology.