March 2026

Attribution & AI Outputs

A Creative Commons Issue Brief: Backgrounders on topics related to AI & the Commons

This issue brief considers how attribution works in the context of AI—where systems generate content based on large amounts of existing data—and why attribution still matters, even when it is difficult to implement.

Introduction

Attribution means giving appropriate credit to the people and sources behind a piece of work. Attribution helps people understand where information comes from, decide whether to trust it, and give credit to those whose work made it possible.

Attribution serves a number of important functions, including:

- Creators of new works may credit prior works and other creators as a show of respect, or cite a third party to demonstrate evidence for a claim. For instance, consider:
    - A student's essay or scholarly research paper that cites sources for its claims.
    - An artist that provides credit to another artist's work to acknowledge and show respect for their inspirations.
- Creativity builds upon existing knowledge and past ideas. Attribution provides essential pathways from a new work to those past building blocks. For instance:
    - Consider someone reading a scholarly research paper and then examining the underlying sources to interrogate whether the paper's claims are justified.
    - To examine and validate scientific research, a researcher may look at the attributed inputs for the research and attempt to reproduce it.
    - A listener for a piece of music might explore credited influences and listen to works from those artists.
- By directing attention to other works, attribution can provide value. For instance:

- Consider the way links on the internet help drive traffic to websites, and how websites then may make money from that traffic through ads or subscriptions.
- Some research indicates that, in some circumstances, attribution can sometimes be more important and valuable to creators than economic payments.

Attribution and the CC Legal Tools

Attribution is a cornerstone of Creative Commons. All six CC licenses include attribution requirements. The licenses state:

> *You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.*

We have worked with our community over time to provide guidance and best practices for different circumstances, including new types of technology and creativity. Attribution is also recommended within CC's public domain tools, even where it is not a legal requirement.

Opportunities and Challenges in AI Data Attribution

Attribution remains essential in AI-based creativity and knowledge-sharing outputs, which is why it is a requirement within the CC signals framework. When AI systems produce outputs without attribution, important context can be lost. Information may appear without any indication of where it came from, making it harder to trust the information, check whether it is accurate, and credit the people whose work made it possible. Over time, this can weaken how knowledge is created, shared, and improved.

For example, if an AI system summarizes facts from multiple news articles but does not name the sources, readers cannot easily verify the claims. Journalists and news organizations also lose credit for their work, which can reduce their ability to sustain high-quality reporting.

Lack of access to a model's training data can also undermine the ability to study and reproduce an AI model, which is especially important in the context of open source approaches to AI.

Data used at different points of development and use of AI present different challenges and opportunities with respect to attribution.

Attribution of datasets used in model training

As discussed in our Issue Brief on Copyright and Generative AI, developers create AI models through machine analysis of data, so that models can capture statistical patterns within training data. These patterns include language structure, basic facts about the world, and how words relate to images.

There are different ways that model developers can provide attribution to the training data. For example, if an AI developer trained on a defined dataset like LAION, they could cite directly to LAION. Models may also be trained on many different works, such as by crawling parts of the website, in which case individual domains or URLs could be cited.

At the same time, there are challenges and limitations with attribution in this context, notably:

- Comprehensibility and practicality: Consider a model built on millions of websites or billions of data points. A long list of links or filenames could be useful for other developers but not necessarily comprehensible for other users.
- Lack of available information: A developer who trains a model on public web data may know the website address where a file came from but not necessarily any further information about any underlying works contained in the site. A single page may contain text from one author, comments from another, and images from a third provider—and the model developer may have no information about those underlying works, besides the fact that they all were found at a given URL.

Attribution of AI outputs

When a model is used to produce an output, attribution can also be provided to some degree. For instance, a generated output could provide attribution to a model and its training datasets to the extent known.

That said, more granular attribution to individual parts of the training data can be challenging.

- Model training derives insights from the training data and then models make statistical predictions to generate outputs. Models themselves are not designed to store the training data and thus do not access or draw directly from it after training.
- Because training sets can be very large, it is difficult to isolate and measure the extent to which a particular data point in the training data impacts a specific output.
- The same or similar information may appear in many different places in the training data, and thus it can be hard to distinguish which one influenced a given output.

AI models may also access and use other data when producing outputs. For example, when you ask a Large Language Model (LLM) to tell you about a topic in the news, it may fetch recent news articles while also using the model to summarize key facts. In this example, outputs can more easily be attributed to those specific inputs. For instance, if an AI model fetches a news article on a particular website and then generates a summary of that news story, it can include a link to that news article specifically. However, the issues of attributing sources within a model remain, and additional information provided via the model (in this case, the summarization of key facts) will persist.

Consideration

More granular forms of attributing particular outputs to parts of training data is the subject of ongoing technical research, which may impact what types of attribution are possible and practical.

Given what is possible and practical, attribution must also be solved via policy and social norms. For example, citing sources within academic research is a social norm driven by professional courtesy, reputation, and institutional guidelines, even if it does not necessarily carry legal repercussions.

Effective solutions will need to be grounded in serving both the interests of creators and follow-on users. While the lack of attribution can harm our knowledge ecosystem, onerous or impractical requirements can also stifle the activities of people who want to reuse materials for legitimate purposes.

Additional Reading and References

- Going beyond open data – increasing transparency and trust in language models with OLMoTrace
- Enhancing Training Data Attribution with Representational Optimization
- The New Art Forgers
- What's a Name Worth?: Experimental Tests of the Value of Attribution in Intellectual Property
- Sufficiently Detailed? A proposal for implementing the AI Act's training data transparency requirement for GPAI, Open Future

Notes on Terminology

Attribution is used here to mean giving appropriate credit to sources for a work. A related question is labeling that work as "AI generated", or, in other words, using attribution to indicate the use of AI, rather than to particular "input" data.

Provenance is also a related concept, referring to information about the origins of a piece of data.