

12.03.2023

CC Signals – towards a new social contract in the era of AI?

Janina Radny

What is the problem?

AI bots in the web

- Estimated **37% to over 50%** of web and API traffic **comes from automated bots**, e.g.
 - traditional web crawlers
 - malicious bots
 - AI bots (80% crawlers and 20% fetchers)
- Associated problems:
 - Decreasing human user interaction
 - Incoherent attribution practices
 - Very high website traffic (up to 39k requests/minute)
- Mutual benefits (“social contract”)** are no longer provided



Reactions

- Some creators take content down
- Content is hidden behind paywalls
- Cambridge University Press announced to use most restrictive [CC BY-NC-ND licence more often as default](#)
- Websites block crawling for training purposes or entirely
- Interventions with offensive tools, e.g. [data poisoning](#), [AI tarpits](#) and [labyrinths](#)
- ...
- Many of these measures are also impeding human users or machine use for public interest
- Not desirable when favouring openness

Wikipedia becomes fee-based for AI bots

The Wikimedia Foundation wants to secure Wikipedia's survival in the AI era. Those who profit should pay for the effort they cause.



(Image: [Gerd Altmann](#), gemeinfrei (Creative Commons CC0))

Nov 12, 2025 at 8:50 am CET 2 min. read

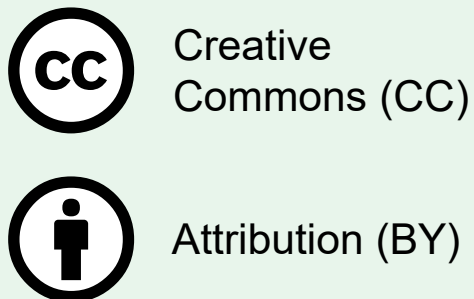
By [Daniel AJ Sokolov](#)

CC Signals as new social contract?

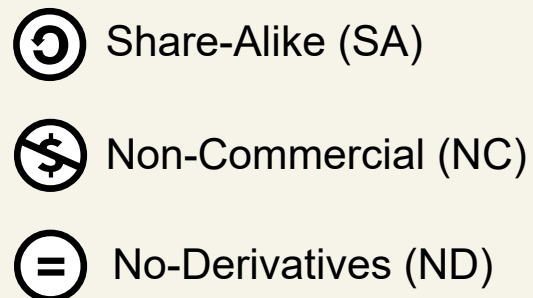
Why not just regular CC licences*?

- CC licences consist of basic and additional elements
- Combinations define how individual works can be reused

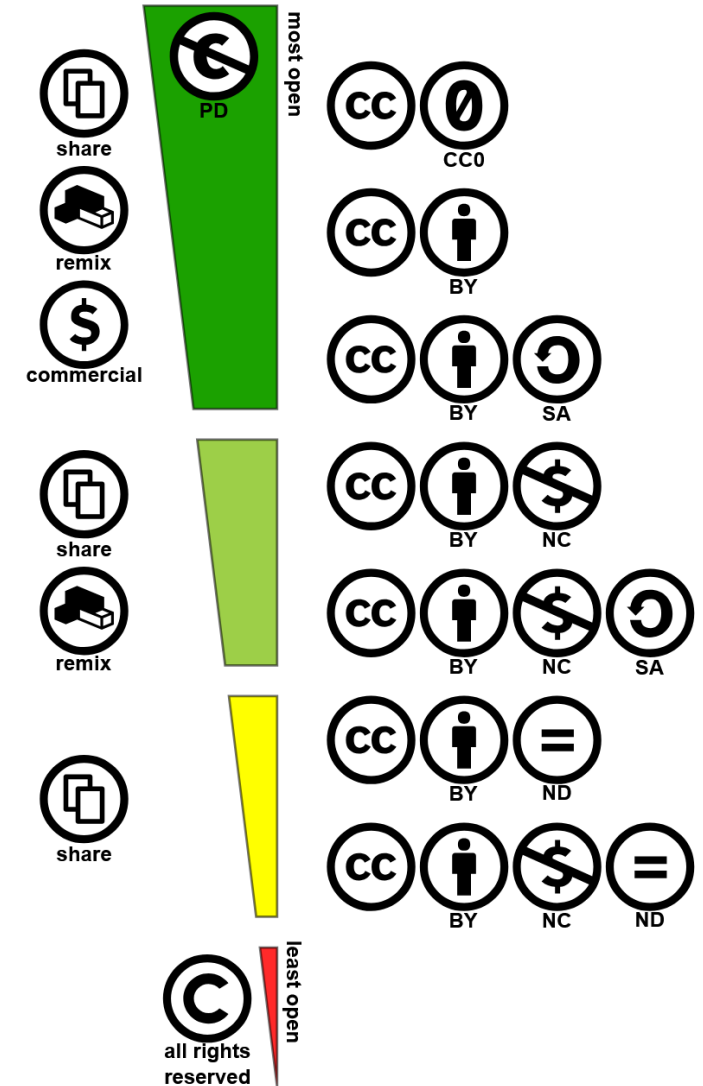
Basic elements



Additional elements



*Learn more about CC licences and why we recommend CC BY [here](#)

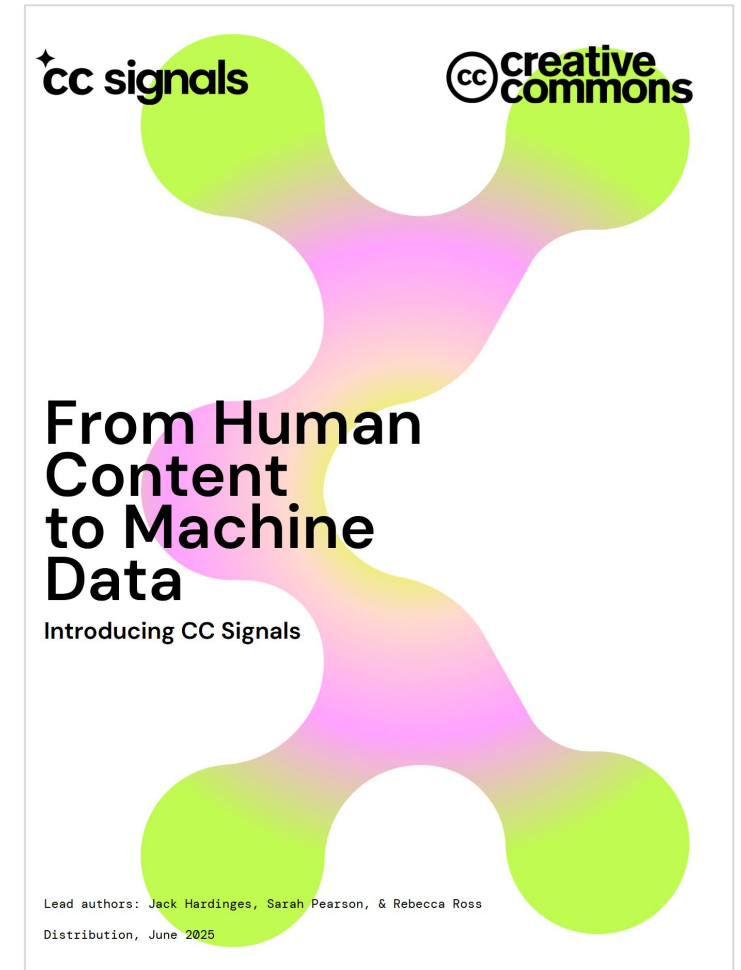


Shaddim; original CC license symbols by Creative Commons, CC BY 4.0, via Wikimedia Commons

Why not just regular CC licences*?

- CC licences consist of basic and additional elements
- Combinations define how individual works can be reused
- “[...] by design, they **apply only when copyright applies**, [...]”
- “Given that **AI training falls outside the scope of copyright** in many scenarios [...]”
- “[...] CC licences are **not well-designed for imposing reciprocal terms** on AI developers.”
- CC Signals as alternative approach, not an extension of existing CC licences

*Learn more about CC licences and why we recommend CC BY [here](#)

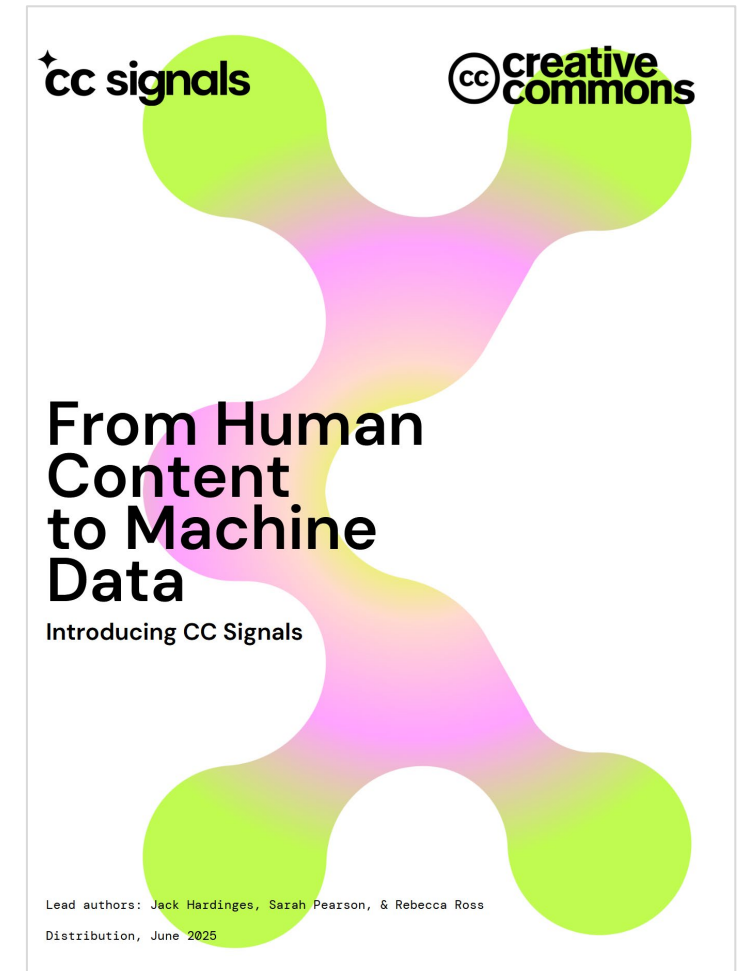


The intention of CC Signals

“[...] to offer a new way for **stewards of large collections of content** to indicate their preferences for how machines (and the humans controlling them) should **contribute back to the commons.**”

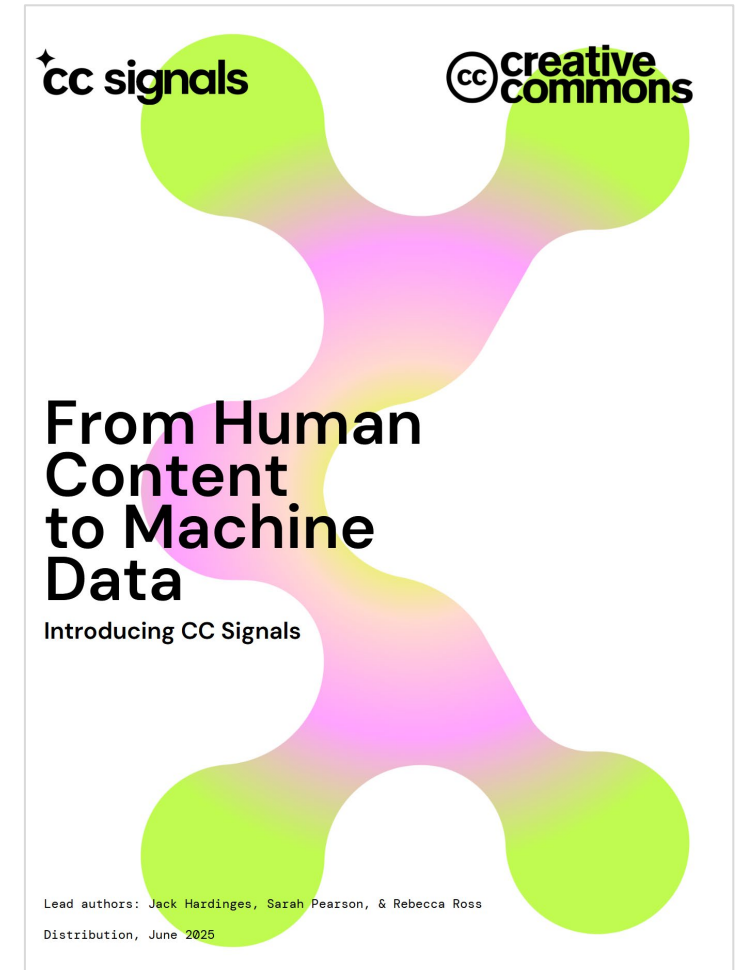
Differences to “regular” CC licences:

- Different declaring party
- Definition of required actions to use content



Data stewards as declaring party

- “We are **focusing on supporting** CC signals to be applied by Declaring Parties that are **stewards of large collections of content** because it is the **most efficient way** to shape new norms across the ecosystem.”
- “An individual work is infinitesimally small in the context of modern machine reuse [...]”
- Stewards of collections might be most immediately affected by crawling bots
- Contested in [community feedback](#) from individual creators
- To be sharpened in future iterations



Modules: first iteration June 2025

- **Credit**
 - Citation of training dataset is required
 - Citation of collection as source with link for retrieval augmented generation (RAG)
- **Direct contribution**
 - Monetary or in-kind support of the Declaring Party
 - Based on type and scale of machine reuse and financial means of the reusing party
- **Ecosystem contribution**
 - Monetary or in-kind support back to the ecosystem
 - Based on use of asset and financial means
- **Open**
 - Must satisfy frameworks of openness, e.g. Open Source AI Definition (OSAID)



Construction of a signal

- Credit must always be given
- Additional modules are possible, but **mutually exclusive**
- Preliminary state, might change in future iterations

CC Signals:



(&)



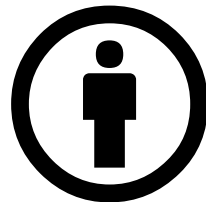
OR



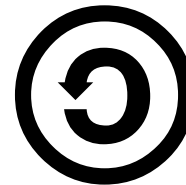
OR



CC licences:



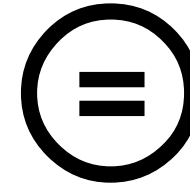
(&)



(&)



(&)



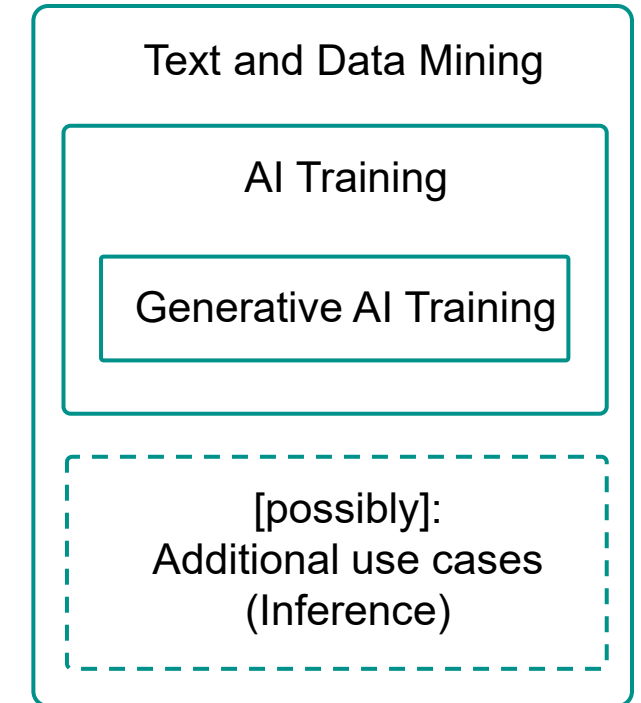
Technical implementation

- Inclusion of signal in robots.txt and HTTP Content-Usage header
- Construction follows: Category + Preference + Signal
- Specification of use based on [vocabulary](#) of the Internet Engineering Task Force (IETF) AI Preferences Working Group

```
User-Agent: *
Content-Usage: ai-use=n;exceptions=cc-cr
Allow: /
```

```
200 OK
Date: Mon, 09 Jun 2025 12:42:03 UTC
Content-Type: text/plain
Content-Usage: train-genai=n;exceptions=cc-cr-ec
```

Example in robots.txt and HTML header. Source: github.com/creativecommons



(Very) high level overview of the vocabulary.
Reproduced from openfuture.eu

Responses to the proposal

From a legal perspective

Creative Commons:

- “[...] the signals are primarily designed to address **the social layer of data governance.**”
- “However, **when applied by a Declaring Party with copyright authority** over the content, CC signals **are likely to have legal implications** under copyright law.”

Fabian Rack, lawyer and author at irights.info:

- Legally binding situations could occur, provided that copyright applies
- No connection between Signals and CC licences:
 - Machines could “choose” to get CC licenced material rather than crawling the database
- Legal enforceability remains to be seen
- Motivation for AI developers: Preserve flow of training content

From the community perspective

- Community sees room for improvement:
 - Concerns that CC is prioritizing AI companies over creators
 - Confusion and disagreement about the CC licenses and AI training
 - Strong calls for opt-outs
 - Asking politely for AI developers to give back in exchange for datasets is not enough
 - Who is this tool meant to serve and the use cases it is meant to address?
 - More clarity about implementation and interoperability

Conclusion

Could CC Signals be a new social contract?

My opinion:

- First iteration of a new concept: flaws are likely and best practice is rare
- Contains a lot of idealism and wishful thinking
- “We recognize that CC signals will rely on willing participation by AI developers to adhere to it.”
- Idealism and wishful thinking started the commons
- CC Signals is one of several comparable approaches, e.g.
 - Verhulst et al. (2025): [Reimagining Data Governance for AI: Operationalizing a Social License for Data Reuse.](#)
- Join the discussion here: <https://github.com/creativecommons/cc-signals/tree/main>



[Wishful thinking](#), Mary Mapes Dodge, 1884

Thank you for your interest!
Let's share thoughts!

References

- AI Bots in Q2 2025: Trends from Fastly's Threat Insights Report | Fastly [WWW Document], 2025. URL <https://www.fastly.com/blog/ai-bots-q2-2025-trends-fastlys-threat-insights-report> (accessed 25.11.25).
- Edwards, B., 2025. Open source devs say AI crawlers dominate traffic, forcing blocks on entire countries [WWW Document]. Ars Technica. URL <https://arstechnica.com/ai/2025/03/devs-say-ai-crawlers-dominate-traffic-forcing-blocks-on-entire-countries/> (accessed 12.11.25).
- Hardinges, J., Pearson, S., Rosa, R., 2025. From Human Content to Machine Data: Using Collective Action to Develop a New Social Contract. URL: https://creativecommons.org/wp-content/uploads/2025/06/Human-Content-to-Machine-Data_Final.pdf
- Mueller, B., Foundation, W., Danis, C., Lavagetto, W.F.G., Foundation, W., 2025. How crawlers impact the operations of the Wikimedia projects. Diff. URL <https://diff.wikimedia.org/2025/04/01/how-crawlers-impact-the-operations-of-the-wikimedia-projects/> (accessed 12.11.25).
- Rack, F., 2025. CC Signals – Anstandsregeln fürs KI-Training? iRights.info. URL <https://irights.info/artikel/cc-signals-ki/32652> (accessed 9.11.25).
- Ross, R., 2025. We Asked, You Answered: How Your Feedback Shapes CC Signals. Creative Commons. URL <https://creativecommons.org/2025/08/27/we-asked-you-answered-how-your-feedback-shapes-cc-signals/> (accessed 11.26.25).
- Vizard, M., 2025. Report Surfaces Significant Spikes in Web Traffic Driven by AI Bots. Techstrong IT. URL <https://techstrong.it/ai/report-surfaces-significant-spikes-in-web-traffic-driven-by-ai-bots/> (accessed 25.11.25).
- Wikipedia becomes fee-based for AI bots | heise online [WWW Document], 2025. URL <https://www.heise.de/en/news/Wikipedia-becomes-fee-based-for-AI-bots-11074755.html> (accessed 11.25.25).