

WIPO Conversations on AI - 11th Session (2025)
Dr. Ana Ramalho, Senior Copyright Counsel (Google)

Good afternoon everyone, and thank you for the opportunity to be here. I am Ana Ramalho and I am a Senior Copyright Counsel at Google.

The first point I'd like to make is: we - all the actors in this ecosystem - need to invest in AI infrastructure and support pro-innovation frameworks to ensure researchers, creators, technologists, and businesses have the tools they need to convert ideas into new scientific discoveries or artistic works. At the same time, we need to make sure that we respect the balance between AI model providers and creative industries, broadly speaking. Only with these considerations in mind can we have an ecosystem that works for everyone.

While fair use and text and data mining exceptions allow for model training, we should all work towards good practices in the AI sphere. In my intervention, I would like to focus on a key good practice in relation to infrastructure, which is the question of opt-out mechanisms - and specifically, robots.txt.

This is an industry standard that has existed for 30 years. Search engines in particular have a long history of respecting these technical rights reservations made by web page owners. It's simple, it's scalable, and all the parties involved (chiefly, web publishers) are very familiar with it. Robots.txt are also both human- and machine-readable, providing a level of transparency in that anyone can check whether or not a page is allowed to be accessed in an automated manner by a specific crawler.

With this in mind, we launched Google-Extended, a standalone product token that builds on robots.txt and that can be used by web publishers to manage whether their content help improve Gemini Apps and Vertex AI generative APIs, including future generations of models that power those products. Web publishers can disallow their entire website, a single or several directories, or a single or several webpage(s), or any combination thereof. In short, we believe that making simple and scalable controls, like Google-Extended, available through robots.txt is an important step in providing transparency and control that should apply with respect to content from the open web.

Of course, as AI applications expand, managing different uses at scale can become increasingly complex. AI developers and creative and web communities should engage to explore the limitations of robots.txt, and improvements of the standard - since, because of its format, robots.txt is flexible and there's room for growth. We are committed to continued engagement with the different communities to explore this, and we will continue to facilitate a public discussion where all voices can be heard. Thank you very much for your attention.