CCIA Briefing

AI and Copyright: what is an "opt out" and can it work?

This briefing builds on an earlier CCIA explainer that covered AI and copyright and the need for a TDM exemption more generally.

What is an opt out?

An "opt out" or "rights reservation" model gives rights holders greater control over whether and how their content is used by allowing them to express a preference over whether it can be used for text and data mining. This mechanism exists in, for example, the EU Copyright Directive and is proposed as an option to accompany a TDM (text and data mining) exception in the UK.

What are the costs of an opt out?

The costs for rights holders to opt out using existing technical tools and solutions is negligible. However, an opt out is not without cost, as it means:

- Additional systems are required for AI developers to implement an opt out, these are to some extent fixed costs and will affect novel applications and newer developers most.
- Some content that rights holders have opted out could have contributed to improving AI models, weakening their ability to make good predictions and support informed decision-making. This could mean AI models that are less useful for UK businesses and undermine progress in tackling risks such as algorithmic bias.

The measure should therefore be understood as a compromise between an imperative to support AI innovation and economic growth, and reassuring rights holders that they can control whether and how content is used.

Is an opt out practical?

An opt out, as a regulatory measure, is relatively novel. This does not mean it is technically impractical and cannot be done on a machine-readable basis reflecting emerging commercial practice.

Technical tools and opt-out solutions already exist and are available to rights holders. For example, rights holders can already rely on the universally accessible robots.txt protocol to opt-out from TDM and prevent the use of their content to train AI models. Recent research by the Reuters Institute found that, as of late 2023, 48% of the most widely used news websites across 10 countries were blocking OpenAI's web crawler by using the robots.txt protocol.

This protocol has limitations in terms of granularity, but represents a robust and universal basis for improvement. Many companies, including Google and Apple, have developed reliable tools and mechanisms to enable rights holders to better control and manage their rights with extensions to the robots.txt protocol that enable crawling for search indexing but not AI training, for example. In July 2023, Google went further by "inviting members of the web and AI communities to weigh in on approaches to complementary protocols" with a video to "explore additional machine-readable means for web publisher choice and control for emerging AI and research use cases.

Companies that enable websites to host content are also addressing this issue. Cloudflare, for example, created tools for customers to easily block well-behaved AI bots and allows all customers (including those on its free tier) to block all AI bots.

While there is not yet a unified global standard, it is clear that an opt-out mechanism is technically practical and achievable as part of a compromise to enable AI development and rights holder control.

What can UK policy do to support opt-out mechanisms?

There is ongoing technical work to establish a common "machine readable" format based on robots.txt.

Rights holders across different creative sectors may have different preferences for different types of machine-readable opt-outs, while AI developers, in turn, require clarity, market adoption, and technical feasibility to make sure this can all work in practice. While there is not yet a unified global standard, it is clear that an opt-out mechanism is technically practical and achievable as part of a compromise to enable AI development and rights holder control.

Policymakers should resist the urge to require an overly prescriptive common solution. Given the diversity and dynamism of the AI sector, it would be better to allow for the evolution of standards and best practices. Any policy requirement should be for an opt out defined at a sufficiently high level that it allows for standardisation to develop in the market globally over time.

How can UK rights holders be confident that an opt out is being respected?

There are broadly two options:

- To the extent a TDM is in place with an opt out, rights holders would have the same legal rights that they do now. With a practical, machine-readable standard, there would be little incentive for developers to take the legal risks associated with using

content subject to an opt out instead of using other data that is not. This should be the preferred option.

- The Government could implement conventional oversight for the systems and processes used to implement an opt out. This would be akin to the process seen in other digital regulations and require AI developers to document how they have implemented the opt out (many are likely to do so anyway) and satisfy an appropriate regulator that this is effective.

There are trade-offs. The more onerous the compliance process, the less likely developers are to train leading models in the UK. Any costs are likely to affect emerging models and smaller developers most and diminish the quality of AI tools available to UK consumers and businesses. However again there is no reason to assume that compliance cannot be achieved by conventional means familiar to UK policymakers and regulators.