

The Information Commissioner's response to the UK Government's consultation on copyright and artificial intelligence

About the Information Commissioner

1. The Information Commissioner has responsibility for promoting and enforcing data protection and information rights. This includes responsibilities under the UK General Data Protection Regulation (UK GDPR), the Data Protection Act 2018 (DPA 2018), the Freedom of Information Act 2000 (FOIA), the Network and Information Systems Regulations 2018 (NIS), and the Privacy and Electronic Communications Regulations 2003 (PECR).
2. The Information Commissioner is independent from Government and upholds information rights in the public interest, promoting openness by public bodies and data privacy for individuals. The Commissioner provides guidance and support to individuals and organisations, aimed at helping organisations to comply, and takes appropriate action where the law is broken.

Introduction

3. Artificial intelligence (AI) has the capacity to fundamentally change how people communicate, create, access services and experience the world. But in order to deliver on that promise it needs to be developed and deployed responsibly, and in accordance with all applicable law.
4. The Information Commissioner's Office (ICO) welcomes the opportunity to respond to the UK Government's consultation on

copyright and AI,¹ supporting its ambition to deliver a framework that supports human creativity, incentivises innovation and promotes legal certainty. This objective aligns with the Government's support for targeted growth-driving sectors² as set out in the Industrial Strategy, but also the AI Opportunities Action Plan³ published earlier this year.

5. The ICO is committed to supporting Government in delivering on that plan, building on our extensive work on AI⁴ and guidance we provide on AI regulation.⁵ We help create an environment where responsible AI innovators are supported while people and their rights are protected through our own innovation services⁶ as well as those we provide jointly with other Digital Regulation Cooperation Forum⁷ regulators.
6. As the consultation on copyright and AI explains, AI models are trained on large amounts of data. Often, a substantial part of this data is personal data, bringing this processing under the ICO's purview. The ICO regulates the processing of personal data across the AI value chain, from the point of the initial collection of training data all the way to the deployment, monitoring or fine-tuning of a model.
7. As a pragmatic regulator, our decision-making involves considerations about securing an appropriate level of protection for personal data while promoting public trust and confidence in its processing. We fulfill our duties as a regulator while supporting innovation and competition. We recently set out how the ICO does this as well as how our work promotes sustainable economic growth in our response to the Prime Minister.⁸

Synergies between ICO's approach and the consultation

8. The ICO does not regulate UK copyright law. However, because some copyrighted material may also meet the definition of personal data,⁹ we believe it is important for Government to take into account the synergies and potential overlaps between the two regimes to ensure regulatory clarity and coherence for people and business alike.

¹ [Copyright and Artificial Intelligence - GOV.UK](#)

² The Creative Industries, and Digital and Technologies, are two of the sectors in the Industrial Strategy: [Invest 2035: the UK's modern industrial strategy - GOV.UK](#)

³ [AI Opportunities Action Plan - GOV.UK](#)

⁴ [Our work on Artificial Intelligence | ICO](#)

⁵ [Artificial intelligence | ICO](#)

⁶ [ICO Innovation Services | ICO](#)

⁷ [AI and Digital Hub | DRCF](#)

⁸ [Letter from the Information Commissioner](#)

⁹ [What is personal data? | ICO](#)

9. The key objectives of the consultation align well with ICO's priorities. More specifically, the consultation's objectives are:
 - **Supporting rights holders' control** of their content and ability to be remunerated for its use.
 - **Supporting the development of world-leading AI models in the UK** by ensuring wide and lawful access to high-quality data.
 - **Promoting greater trust and transparency** in the sector.
10. Relatedly, the ICO and the data protection framework we oversee seek, among other things, to:
 - Maintain people's control over their personal data, including instances where copyrighted material contains or comprises such data.
 - Support the development of AI models that use the personal data of UK data subjects responsibly by ensuring this processing is lawful, transparent and fair.
 - Promote greater transparency in the sector as the foundation of public trust in the technology.

More regulatory certainty may be needed

11. AI developers need clarity on how to lawfully collect and process data to train AI models, including generative AI. Our work on generative AI seeks to build on our existing positions to provide such certainty. We are also working closely with the Competition and Markets Authority on issues AI brings to both competition and data protection law, and our upcoming Joint Statement on Foundation Models will provide more clarity on our respective regulatory approaches.
12. Creatives and publishers also need clarity over the levers they have at their disposal in terms of retaining control of their published and creative work. These issues have been highlighted during the passage of the Data (Use and Access) (DUA) Bill. Additionally, the ICO's recent consultation on generative AI and data protection¹⁰ attracted significant input from creatives about the challenges they face from the increased use of AI.¹¹

Transparency is part of the solution

13. We concur with the Government's view that greater transparency from AI developers is necessary to address issues such as the ability of people to exercise control over their data while also ensuring AI

¹⁰ See the outcomes of the consultation here: [Information Commissioner's Office response to the consultation series on generative AI | ICO](#). See the original consultation here: [ICO consultation series on generative AI and data protection | ICO](#)

¹¹ A substantial portion of the respondents to ICO's consultation on generative AI and data protection came from the creative industries. The ICO and the British Screen Forum also co-hosted a roundtable with the representatives from the sector.

training is happening in compliance with the law. This includes transparency about the material – which may include personal data - used to train models, how that material is acquired, and the content that is generated by these models. This aligns with our recommendations¹² following the conclusion of our consultation series¹³ on generative AI published in December 2024.¹⁴

14. Meaningful transparency from AI developers in terms of training data and outputs is urgently needed. As we set out below, technical and organisational tools and measures to increase transparency such as the public disclosure of operational web crawlers can be part of the solution.

Consider possible unintended consequences

15. The consultation proposes widening existing exemptions for text and data mining (“TDM”) to the development of AI models, known as an “opt out”. Should the Government proceed with this approach, it will be important to provide clarity that this will not in and of itself constitute a determination of the lawful basis for any personal data processing that may be involved under data protection law. This would aim to avoid any potential misinterpretation by AI developers or others. A substantial amount of the material involved in TDM may be personal data and the lawfulness of processing would need to be evaluated on a case-by-case basis. The ICO is ready to work closely with Government and other stakeholders to identify and mitigate potential unintended consequences of any chosen approach.¹⁵
16. Apart from the obligations of entities undertaking the TDM, entities choosing to “opt in” their IP data for that processing may also have data protection obligations. There is a risk that IP rights holders may seek to “opt in” content containing personal data for TDM, without first considering possible data protection compliance issues. For example, a wedding photographer may be a controller for the photos of people they have photographed so they will need to ensure that sharing that data for TDM – by not “opting out” - is fair, transparent and lawful. The interplay between data protection and the future copyright regime will need to be clearly articulated to avoid the risk

¹² [Generative AI developers, it's time to tell people how you're using their information | ICO](#)

¹³ [ICO consultation series on generative AI and data protection | ICO](#)

¹⁴ [Information Commissioner's Office response to the consultation series on generative AI | ICO](#)

¹⁵ It may be useful to share that the ICO is also providing feedback to the EU AI Office's Working Groups developing the voluntary Code of Practice on General Purpose AI, which highlights IP and transparency considerations. Even though that code reflects the requirements of the EU AI Act and not data protection, our contributions seek to support consistency between the EU and UK approach to processing personal data to train AI.

of any ambiguity inadvertently leading to processing practices that may be lawful under copyright but unlawful under data protection.

Opportunities and limitations of technical and organisational measures

17. We welcome the consultation's commitment that any reforms relying on technical measures would come into force when effective, proportionate, and accessible technological solutions are in place. As the consultation points out, there are currently considerable limitations in terms of technical measures to control what data is mined, from whom and for what purposes.

Web crawling

18. As noted above, there is currently a lot of opacity around the identity of entities undertaking web crawling, the purposes of that processing or the justification for its frequency. Our understanding is that blocking web crawlers may be a blunt instrument with negative consequences such as websites not being properly indexed for search engines or for web archiving purposes. As another example, in online advertising, contextual advertising products rely on crawling web pages to categorise content and target relevant audiences. Similarly, brand safety products use web crawling to ensure that ads are not displayed next to harmful or inappropriate content. Technical standards such as robot.txt can also be ignored by crawlers.
19. These limitations should be acknowledged and considered in developing any future regime and the efficacy of standardisation efforts examined. For example, we are interested in the suggestion that more details around web crawlers could be publicly disclosed, including ownership and the purposes for which content is being crawled. More efforts can be made to allow digital publishers to refine the level of access they provide to crawlers depending on the purposes of the processing (eg whether this will be used for training AI or for search engine indexing).¹⁶

The role of organisations hosting personal data

20. We believe it is important to hear from organisations hosting personal data that is 'mined' or scraped (eg websites, social media, etc) and in particular how they monitor the entities scraping their website to collect data, as well as how and if they disclose that information in their privacy policy. In the long term, this could help

¹⁶ For example, Apple stated that its web crawler, Applebot, will adhere to digital publishers' directions not to crawl content for training its foundation models. See: [Applebot model training and individual privacy rights – Apple Support \(UK\)](#)

Government and regulators get a better picture of the scale, nature and purposes of current data mining activities.

21. The ICO and counterparts at the Global Privacy Assembly issued a joint statement in 2024, asking social media companies to play their part in this context and in particular, in relation to unauthorised web scraping.¹⁷

Licensing

22. The consultation on copyright and AI seeks to explore licensing opportunities for training data, which is also an area of interest for the ICO. As we explained in the outcomes report of our consultation on generative AI,¹⁸ we believe that data protection compliance requirements should be part of licensing agreements when personal data is involved, and those who rely on or put in place these licences should ensure that those requirements are met.

Conclusion

23. There are various touchpoints between data protection and copyright law, and we are keen to work closely with Government to ensure consistency between the two regimes to avoid any regulatory gaps or misinterpretations.
24. Some issues may be matters for legal reform, others for technical measures, yet others for public awareness campaigns. Understanding the different contours of the problem is necessary for providing a nuanced and efficient solution that does not place unnecessary burdens on AI developers or lead to legal ambiguity. For example, on the issue of creating unauthorised 'digital replicas' and people's control over their digital likeness, a combination of different approaches may need to be considered.
25. More broadly, the creation of sustainable, lawful and mutually beneficial business models allowing IP holders to be remunerated while also allowing innovators to build AI models with their data could have a positive effect on data protection. That could happen if access to non-personal training data would mean less personal data would be necessary to train AI models or more standardised and legally robust methods of collecting them could become the norm. This in turn could supercharge AI development in the UK, leading to sustainable growth that relies on all sectors of the economy.

¹⁷ [Global privacy authorities issue follow-up joint statement on data scraping after industry engagement | ICO](#)

¹⁸ [The lawful basis for web scraping to train generative AI models | ICO](#)