LEGAL, REGULATORY & COMPLIANCE CONSULTANTS

RESPONSE SUBMITTED BY HANDLEY GILL LIMITED TO

THE INFORMATION COMMISSIONER'S OFFICE'S CONSULTATION ON GENERATIVE AI & DATA PROTECTOIN

CHAPTER 1: THE LAWFUL BASIS FOR WEB SCRAPING TO TRAIN GENERATIVE AI MODELS

01 MARCH 2024

ABOUT HANDLEY GILL LIMITED

At Handley Gill, our specialist consultants pride themselves on helping you get the job done… right. We combine a cost-effective, pragmatic and robust approach with the in-depth technical knowledge and expertise necessary to provide quality data protection / data privacy, content regulation, online safety, artificial intelligence (AI), information access, human rights, environmental, social and governance (ESG) and wider legal advice, compliance and assurance services to our clients.

Our consultants have experience across the public and private sectors, working in-house as well as in professional services organisations, advising individuals and organisations in the UK, EEA, USA and across the world spanning a number of industries including:

- regulated industries, such as law firms and other legal professionals, financial institutions and other financial services providers, insurers and insurance intermediaries;
- retail, branding, advertising & marketing;
- technology start ups and scale ups;
- content providers, including publishers, broadcasters, social media platforms and online and editorial content creators;
- political parties and lobbying groups;
- law enforcement entities;
- charitable and non-governmental organisations;
- employment / recruitment agencies;
- the public sector;
- sport and fitness; and,
- healthcare and wellness.

Section 1: Your views on the proposed regulatory approach

1. Do you agree with the analysis presented in [this document](#) ['Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models']?

   ☐ Yes

   ☒ No

   ☐ Unsure / Don't Know

   Please explain your answer:

   While the analysis recites the existing legal obligations and provides some guidance, it makes a number of unsubstantiated assertions/proceeds on the basis of inappropriate suppositions and fails to highlight appropriate challenges or to set out the Information Commissioner's position.

   While the analysis notes that "Training generative AI models on web scraped data can be feasible", given that generative artificial intelligence (AI) models are now ubiquitous, we would have welcomed an indication of the circumstances in which the Information Commissioner accepts that web scraping to harvest training data is permissible and the circumstances or factors that would suggest that it is in fact non-compliant with data protection law.

   We would welcome clarity as to the Information Commissioner's position on the lawfulness of web scraping without prior authorisation under UK law, including in circumstances where, for example, website terms and conditions explicitly prohibit unauthorised web scraping. We note in this regard that the Information Commissioner joined other data protection and privacy authorities in warning data controllers to take measures to protect against unlawful data scraping[1].

   On the question of 'Is legitimate interests a valid lawful basis for training generative AI models on web-scraped data?', in relation to the 'Purpose test: is there a valid interest', the analysis states *"The key question is this: if you don't know what your model is going to be used for, how can you ensure its downstream use will respect data protection and people's rights and freedoms?"*. We disagree that this is the key question in relation to the purpose test; it is in fact relevant to the balancing test.

---

[1] Information Commissioner's Office 'Joint statement on data scraping and data protection', 24 August 2023: https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/08/joint-statement-on-data-scraping-and-data-protection/

In relation to the 'Necessity test: is web scraping necessary given the purpose?', the analysis implies that the Information Commissioner will accept that web scraping is necessary. No evidence is given for the assertion that *"The ICO's understanding is that currently, most generative AI training is only possible using the volume of data obtained though large-scale scraping"* and *"there is little evidence that generative AI could be developed with smaller, proprietary databases"*. AI developers could choose to seek permission to licence content from publishers, social media platforms and others for the purpose of training, which would not only serve to protect copyright and other intellectual property rights in the information but would also assist data controllers in protecting personal data and afford greater transparency and choice to affected data subjects. We acknowledge that this would involve additional cost and would impose a certain administrative burden on web developers, but do not consider that the Information Commissioner has accepted that amounts to a necessity in any other context. Similarly, we would welcome the Information Commissioner's position on the interaction between unauthorised and indiscriminate web scraping, including of children's personal data, and the criminal offences of the Unlawful obtaining etc of personal data under s170 Data Protection Act 2018.

In relation to the 'Balancing test: do individuals' rights override the interest of the generative AI developer?', while we note that the Information Commissioner has acknowledged that unauthorised web scraping for the purpose of harvesting training data will be invisible to data subjects, no recognition is given to the fact that it is unlikely to have been within the reasonable expectations of affected data subjects.

In so far as the guidance suggests that individuals may *"lose control"* over their personal data, we consider that such expositions are unhelpful as individuals neither own nor have an unfettered right to control the use of their personal data.

2. As we explain in the consultation, the legitimate interests test could be met if technical and organisational measures to limit the use of the Gen AI model are in place. Do you agree with the analysis we have presented?

☐ Yes

☐ No

☒ Unsure / Don't Know

Please explain your answer:

While some indication is given of the technical and/or organisational measures that could be relevant in certain circumstances, no indication is given as to the

Information Commissioner's position on the sufficiency of any one or more given measures.

In light of this, we are concerned that the indication that *"There are a number of considerations that may help generative AI developers pass the third part of the legitimate interests test, relevant to both the development and deployment of a model"* may afford a false sense of security to AI developers.

3. This is the first in a series of publications on the ICO's analysis of personal data processing involved in Gen AI. What aspects of this topic would you like us to consider in future publications?

Obligations to protect against inaccurate personal data in accordance with the principle of data protection by design and default.

Section 2: Impacts on your organisation

4. Are you answering on behalf of an organisation?

☒ Yes

☐ No

5. Which of the following describes your organisation?

☐ Developer of Gen AI models
☐ Supplier of Gen AI models
☐ Current user of Gen AI models
☐ Potential user of Gen AI models
☒ Advisor to organisations developing or using Gen AI models
☐ Other

6. Do you think the proposed regulatory approach will result in benefits or costs for your organisation?

☐ Benefit(s) to your organisation

☐ Cost(s) or burden(s) to your organisation

☐ Both

☒ Neither

☐ Unsure / don't know

Please explain your response:

7. What is the name of your organisation?

Handley Gill Limited

8. If you would like your response to remain anonymous when reporting please tick this box

☐ I wish to remain anonymous

9. How many staff does your organisation have globally?

☒ 0 to 9 members of staff

☐ 10 to 249 members of staff

☐ 250 to 499 members of staff

☐ 500 to 999 members of staff

☐ 1,000 to 2,499 members of staff

☐ More than 2,500 members of staff

☐ Unsure / Don't know

10. Approximately what percentage of your staff are based in the UK?

☒ 81% to 100%

☐ 61% to 80%

☐ 41% to 60%

☐ 21% to 40%

☐ 0% to 20%

☐ Unsure / Don't know

11. If you have UK based staff, where are they?

☒ England

☐ Northern Ireland

☐ Scotland

☐ Wales

☐ Unsure / Don't know

12. Before completing this survey, do you have any final comments you have not made elsewhere?

N/A

13. We may wish to contact you for further information on your responses. If you are happy to be contacted please provide an email address below.

info@handleygill.com