On behalf of Shoosmiths LLP I am pleased to respond to the second part of your generative AI consultation: Purpose limitation in the generative AI lifecycle.

If you would like more details of Shoosmiths' Privacy and Data team including our advice experience and other recent consultation responses please see https://www.shoosmiths.com/expertise/services/privacy-and-data

Response to call for evidence

We support the ICO's analysis of the principles of purpose limitation and in particular the careful application of the principles to each separate stage of the AI lifecycle. We agree that it provides a means of exerting a degree of control over indiscriminate re-use of personal data.

The following points would benefit in our view from further consideration:

The role of transparency

Alongside establishing a lawful basis, our experience suggests that this is one of the most difficult elements of generative AI development to reconcile with current data protection law, and the most difficult for users to navigate in practice. We welcome the reminder of the role that the transparency principle has to play in considering purpose limitation (and in particular the compatibility test); however, this is only one of a number of transparency obligations in the UK GDPR which will be relevant to the development of generative AI. We would therefore welcome comprehensive guidance on whether, and how, transparency principles can be achieved when using personal data for AI training without express consent. It would be helpful to have separate guidance on transparency with a link from this piece. Otherwise, the paragraph in the draft guidance starting "When the developer has no direct relationship" (fourth paragraph under "Our analysis") has in our view the potential to mislead readers by suggesting that strategies such as public messaging campaigns, prominent privacy information and the use of PETs are sufficient to address *all* the transparency challenges users face under current laws. In addition it would be helpful to understand whether these approaches in relation to publicity (which presumably stem from the "disproportionate effort" exemption in Article 14(5)(b)) will apply in *any* given situation or whether there are controllers or processing activities where they will not satisfy legal requirements.

Interplay with Article 89 (research purposes regime)

Further consideration of Articles 14(4) and (5) would be helpful, together with analysis of the role that Article 89 plays when navigating re-use of data collected for a different purpose. Existing ICO guidance on Article 89 refers to a possible link with generative AI but this is not explored in the guidance issued to date.

Further, linking to earlier observations, it would be helpful to have the ICO's view on whether *only* an organisation which is complying with the requirements of Article 89 and s.19 of the Data Protection Act 2018 can rely on the "disproportionate effort" exemption in Article 14(5)(b) in the context of generative AI.

Legitimate interests as a lawful basis for processing

In the ICO's draft guidance on web scraping (part one of the consultation) we especially welcome the acknowledgement that issues such as breach of intellectual property law can undermine efforts to comply with data protection law. We would go further and point out that the same difficulty plainly extends into the concept of "legitimate interests" itself. As the ICO's own guidance on legitimate interests says, "Anything illegitimate, unethical or unlawful is not a legitimate interest[…]If the interest is not legitimate then you do not meet the first part of the test and you are not able to use legitimate interests as your lawful basis. There is no need to consider the rest of the test as the other parts are not able to legitimise processing that is illegitimate from the outset". It would seem appropriate to include this within the discussion of legitimate interests as a possible lawful basis for processing.

Web scraping will often rely on use of data collected in violation of website terms and conditions. It will very often infringe copyright, at least in the commercial context. To disregard this central difficulty when considering “legitimate interests” as a lawful basis for web scraping activity risks undermining the facts underpinning the key legitimate interests tests. It also demonstrates the difficulty in relying on a cross-sectoral approach to the regulation of AI within existing frameworks. However, given that this is the current position, we would urge the ICO to resist any call to restrict discussion of the lawfulness of data scraping to “data protection” matters only. It is important that at least one regulator is able to take a holistic view.

It would also be useful to understand the ICO’s view on the requirement under Art. 6(1)(f) UK GDPR to consider the weight to be given within the balancing test “in particular where the data subject is a child”.”