View on our website.

GIBSON DUNN

Artificial Intelligence and Privacy, Cybersecurity & Data Innovation Update

December 27, 2024

Unboxing the EDPB's Opinion on AI Models

Few European Data Protection Board (EDPB) opinions have been awaited as eagerly as the EDPB's opinion on AI models (Opinion)[1]. The build-up to publication of the Opinion raised levels of expectation that were almost impossible for the EDPB to meet.

The EDPB finally delivered the Opinion just in time for Christmas, and it is almost as notable for what it does not cover as for what it does. A number of important issues concerning AI models are not addressed at all, and much of what the Opinion does cover is drafted in heavily qualified language that leaves substantial room for interpretation and will not be straightforward to apply in practice.

However, two points in particular stand out and should in our view be welcomed by those developing and deploying AI models. The first is that the EDPB has **avoided taking a hard line that training AI models with personal data means that those models can never be considered anonymous**. Instead it stresses the need for a case-by-case assessment based on the likelihood of personal data being extracted from the model and the likelihood of obtaining personal data from queries. However, the threshold set by the EDPB for a model to be considered anonymous is a high one, and controllers are likely to have substantial difficulties in practice with giving effect to data subjects' rights in relation to models that are not considered anonymous.

The second is that the EDPB **has not ruled out the possibility of controllers relying on legitimate interests for developing and deploying AI models**, including training AI models on personal data scraped from publicly-accessible websites. Again, the EDPB stresses the

requirement for a case-by-case assessment, and identifies factors that should be taken into account by controllers, including in relation to web-scraping. As with the issue of anonymity, the EDPB sets a high bar, and the Opinion is light on detail as to how the EDPB's recommendations can be applied in practice.

It may be tempting to criticise the EDPB for taking such a cautious approach – after all, it leaves some of the most pressing questions unanswered, and creates the potential for significant fragmentation in approach at member state level. However, some of the limitations in the Opinion result from the way in which the issues were brought before the EDPB; it was always going to be difficult for the EDPB to give concrete answers to some of the questions put to it, and, given the rapid pace of technological development in the AI field, the EDPB would have been unwise to try.

Key Takeaways

- The EDPB's view is that training AI models with personal data does not necessarily prevent those models being anonymous. Whether they are actually anonymous depends on the likelihood of extraction of personal data, either through direct extraction from the model or from the model's outputs.
- The EDPB has set a high bar for anonymity, and developers will need to be able to demonstrate the design and functioning of their models, including by maintaining comprehensive documentation. The EDPB's position that AI models may not be anonymous is likely to give rise to serious issues, particularly in relation to the exercise of data subjects' rights.
- The EDPB has not ruled out controllers relying on legitimate interests for developing or deploying AI models, including in relation to training models using personal data scraped from public websites. Again, the EDPB has set a high bar, and its position on necessity is likely to create significant practical difficulties for those training LLMs and similar foundation models.
- Supervisory authorities may be able to impose corrective measures in relation to the deployment of AI models that are not anonymous, where those models have been developed through unlawful processing of personal data. This applies even where one party develops the model and another deploys it. Those acquiring AI models will need to carry out careful due diligence on the development phase, and will need to consider appropriate contractual protection.

Background to the Opinion

The Opinion arose from a request from the Irish Data Protection Commission (IDPC) for an opinion in relation to AI models and the processing of personal data. That background is important, because the EDPB can be criticised only so far for the limited scope of the Opinion; **an opinion under Article 64(2) GDPR should not be confused with guidelines or recommendations issued by the EDPB on its own initiative under Article 70(1) GDPR**. An opinion under Article 64(2) is directed to the questions put to the EDPB, so its scope is, to a degree, dictated by the scope of those questions. Nevertheless, given the keen interest in the Opinion and the broader significance of the issues discussed, this raises important questions about when the EDPB should be issuing guidelines or recommendations on its own initiative

rather than relying on individual supervisory authorities to frame the questions it considers, as well as about the transparency of the Article 64(2) process. In its 2024-2025 Work Programme, the EDPB has planned to issue guidelines on anonymisation, pseudonymisation and data scraping in the context of generative AI.

The Opinion addresses three main issues. First, when can an AI model trained on personal data be considered anonymous? Secondly, can controllers rely on legitimate interests as a lawful basis under GDPR for processing personal data in the development and deployment of an AI model? Thirdly, what are the consequences of unlawful processing of personal data during the development of an AI model?

Scope of the Opinion

The Opinion is concerned only with AI models that are trained with personal data.[2] That reflects the definition of AI models used by the IDPC in its request[3], but it does mean that the Opinion does not address AI models that may process personal data but that were not themselves trained with personal data.

The **Opinion also does not cover certain issues that may arise under the GDPR when using AI models**, including the processing of special category data, automated decision-making, purpose limitation, data protection impact assessments and data protection by design and by default.[4] These are important considerations that are already being addressed in other jurisdictions (such as in California, with the draft automated decisionmaking technology (ADMT) regulations recently advanced to formal rulemaking by the California Privacy Protection Agency), and may need to be addressed by the EDPB in order to avoid supervisory authorities taking diverging approaches. Therefore, on these issues supervisory authorities should proceed cautiously and be open to considered dialogue with controllers on developing best practices.

When can AI models be considered anonymous?

The first question addressed by the EDPB is when an AI model that is trained with personal data can be considered anonymous.

Here, the EDPB considers **three categories of AI models**. The first category is AI models that are specifically designed to provide personal data about individuals whose personal data was used to train the model. The EDPB dispenses with these quickly – these models inherently involve the processing of personal data, and cannot be considered anonymous.[5] Examples given by the EDPB are AI models fine-tuned on an individual's voice in order to mimic that individual's voice, and models designed to reply with personal data from the training data set when prompted for information about a specific individual. It remains to be seen how broadly supervisory authorities interpret this category of AI models; certainly, many of the current generation of generative AI models (such as some large language models (LLMs)) are *capable* of outputting personal data from the data used to train them when prompted to do so (e.g. "tell me all about <celebrity>"), even if they are not designed uniquely for that purpose.

As to AI models that are *not* designed to provide personal data about individuals whose personal data was used to train the model, the critical question posed by the EDPB is whether information

relating to those individuals can be obtained from the model *with means reasonably likely to be used*.[6] If so, the model cannot be considered anonymous. If not, the model can be considered anonymous and is outside the scope of the GDPR.

Here, the EDPB notes that the exploitation of vulnerabilities in AI models may result in leakage of personal data, and also identifies the possibility of accidental leakage of personal data through interaction with the model. Whilst the EDPB does not say so expressly, the EDPB evidently considers that *means reasonably likely to be used* may include means that would be unlawful under the GDPR and other EU and member state law. This is an interesting expansion of the approach taken by the CJEU in *Breyer*[7], which focused on whether a provider had *legal means* which enable it to identify the data subject.

On the basis that personal data may in certain cases be obtained from AI models trained with personal data, the EDPB **concludes that AI models trained on personal data cannot be considered anonymous in all circumstances, and that a case-by-case assessment is required** (one of many case-by-case assessments that the EDPB encourages in the Opinion).[8]

As to what that assessment should involve, the EDPB encourages supervisory authorities to focus on two areas: **whether personal data relating to the training data can be extracted** from the model itself and **whether output produced when querying the model relates to data subjects whose personal data was included in the training data set**.[9] In each case, the question is whether the personal data can be obtained with reasonable means, and in order for the model to be considered anonymous the likelihood of obtaining the data through those means must be 'insignificant'.[10] The EDPB stresses that a "thorough evaluation" of the risks of identification is likely to be required.

Helpfully, the EDPB identifies measures that might reduce the risk of identification, as well as factors that supervisory authorities should take into account in evaluating the residual risk of identification, including the design of the AI model itself, the selection of data sources used for training the model, the design of the training process itself and measures designed to limit personal data included in model outputs (e.g. output filters).

One point that stands out in particular is the need for **comprehensive documentation**. Providers who wish to make claims that their models are anonymous should be prepared to produce documentation to support that position, including documentation on the specific measures used at each stage of the model lifecycle to reduce the risk of identification.

It is worth noting that the EDPB appears to diverge from the approach taken by a number of supervisory authorities, notably the Hamburg DPA in its discussion paper on LLMs[11], which have expressed the view that LLMs themselves do not contain personal data, although their outputs may do so. This may be because the Opinion is not limited to LLMs specifically and therefore does not assume that data is necessarily stored within the model in tokenised form. However, **the EDPB's reference to whether personal data can be extracted from the *output* of a model as a factor in determining whether the model is anonymous** suggests that the EDPB's view is at odds with that of the Hamburg DPA and likeminded supervisory authorities. This is likely to give rise to serious issues in practice, and in particular whether and

how controllers can give effect to data subjects' rights under Chapter III of the GDPR in relation to AI models that are not considered anonymous on the EDPB's view.

When can legitimate interests be relied on in developing and deploying AI models?

The second question addressed by the EDPB is whether, and in which circumstances, controllers can rely on the legitimate interests basis[12] for developing or deploying AI models.

Perhaps the most important point to take away is that the EDPB does not rule out controllers relying on legitimate interests, either in general or in any specific case. **In particular, the EDPB does not rule out the possibility of relying on legitimate interests for training AI models with data derived from web-scraping**. However, as with the question on anonymity of AI models, the Opinion does not give concrete examples of cases where controllers can rely on the legitimate interests basis. Instead, the EDPB stresses the requirement for a case-by-case assessment, adopting the three-step test in Article 6(1)(f) GDPR (i.e. identifying a legitimate interest pursued by the controller or a third party; establishing necessity of the processing for pursuit of that interest; and balancing the legitimate interest against the interests, rights and freedoms of the data subjects). Much of the EDPB's analysis here draws on its prior work on legitimate interests, including its guidelines from earlier this year[13].

One interesting point to note in the context of lawfulness is that the EDPB gives **violation of intellectual property rights as an example of a factor that may be relevant when evaluating whether the controller can rely on legitimate interests**. This echoes a similar point made by the ICO in its first call for evidence on generative AI[14] and in its outcomes report[15], in the context of the lawfulness principle. **This is questionable**. It is true that (as the EDPB notes) the CJEU has clarified that the *interest* pursued by the controller must not be contrary to law[16], but that is not to say that any violation of intellectual property rights *in pursuing* that interest renders the processing unlawful within the framework of GDPR. It should be noted here that the owners of the intellectual property rights may well not (and often will not) be the data subjects. Does training an AI model with personal data in breach (even inadvertent breach) of a licence for that data render the processing unlawful? What about the use of third party software to train an AI model in breach (even inadvertent breach) of a licence for that software? Such an approach would represent a remarkable expansion of EU data protection law into areas that have nothing to do with the protection of personal data, and in which data protection law does not belong.

In relation to the necessity limb, the EDPB's assessment sets a high bar, although this is broadly consistent with the EDPB's prior guidelines on legitimate interests. One potential difficulty for those developing AI models is the EDPB's position that, *"if the pursuit of the purpose is also possible through an AI model that does not entail processing of personal data, then processing personal data should be considered as not necessary"*.[17] A number of AI models, including LLMs, require an extremely large training corpus, and for practical purposes this necessitates training those models using data scraped from publicly available websites. This will, in many cases, necessarily include personal data. If those training LLMs and similar foundation models are required to demonstrate to supervisory authorities, every single time and on a case-by-case basis, that it was not feasible to train the model without processing personal data, this will act as a significant impediment to current model training activities. It would have been helpful if the EDPB had done more to recognise the practical reality facing those training foundation

models, when considering the necessity limb. How supervisory authorities now apply the necessity limb in practice will be of critical importance.

Much of this section of the Opinion is given over to the balancing test, and two points in particular: **data subjects' reasonable expectations and mitigating measures that may be employed by controllers**. In relation to reasonable expectations, the EDPB repeats a point made in its own prior guidance that the fulfilment of transparency requirements under GDPR is not sufficient in itself to consider that data subjects reasonably expect the processing in question. This continual downplaying of the significance of data protection notices is unhelpful; after all, what is the point of the transparency requirements if not to inform data subjects' expectations as to how their personal data will be processed? The EDPB also repeats a point made in its prior guidelines on legitimate interests, that mitigating measures should not be confused with measures that the controller is legally required to adopt anyway, an unhelpful and unnecessary distinction that is difficult to apply in practice.

In **relation to web-scraping specifically**, those looking for a categorical statement from the EDPB as to whether and when this is in line with data subjects' reasonable expectations may be disappointed: the EDPB does not express a firm view either way, but does explain that the steps taken to inform data subjects should be considered. The EDPB does not elaborate on this, which is a pity given that in many cases of web-scraping informing data subjects about the use of their data to train AI models (beyond making a notice generally available to the public) is practically impossible.

In relation to mitigating measures, the EDPB gives examples of measures **that facilitate the exercise of individuals' rights (including rights of objection and erasure)** and enhanced transparency measures. The former in particular are likely to be extremely challenging to implement in practice, especially in relation to personal data derived from web-scraping, where the controller has no prior relationship with the data subject. The EDPB's recommendations in relation to web-scraping specifically may be more helpful: in the development phase, the EDPB recommends that controllers consider, for example, excluding content from websites that are likely to present particularly high risk or from websites that have objected to scraping by using mechanisms such as robots.txt or ai.txt. Similarly, in the deployment phase, the EDPB recommends that controllers consider technical measures to prevent the output of personal data (such as through regurgitation of training data) and also measures to facilitate the exercise by individuals of their rights, in particular in relation to erasure of personal data (controllers may see a glimmer of light in the EDPB's reference to the erasure of personal data from model *output data*, rather than from the model itself).

What are the implications of unlawful processing of personal data in the development of an AI model?

The final question addressed by the EDPB concerns the impact of unlawful processing of personal data, during the development of an AI model, on the lawfulness of use of the model in the deployment phase.

Here, **the EDPB considers three scenarios**. In the first scenario, a controller unlawfully processes personal data to develop an AI model, the personal data is retained in the model and it

is subsequently processed by the same controller. In this scenario, the EDPB's position is that the power of the supervisory authority to impose corrective measures on the initial processing would, *in principle*, affect the subsequent processing. However, whether the development and deployment phases of an AI model are separate processing activities, and the impact of unlawfulness in the development phase on processing in the deployment phase, is to be assessed on a case-by-case basis (that phrase again). In other words, the EDPB stops short of saying that a supervisory authority can require a controller to delete or stop using an AI model that has been unlawfully trained on personal data, but appears not to rule that out.

The second scenario is the same as the first, except that the controller using the model in the deployment phase is different from the controller who developed the model. The EDPB's view here is the least conclusive of the three scenarios – it stresses the need for (you guessed it) a case-by-case assessment, and in particular the degree of due diligence carried out by the deployer on the original processing carried out by the developer. The EDPB appears here to allow more flexibility than in the first scenario, but does not rule out the possibility of corrective measures relating to the initial processing also affecting the subsequent processing. One point is clear, however: those acquiring AI models will need to carry out careful due diligence on developers of AI models, and will need to document their findings and should be prepared to share them with supervisory authorities. Acquirers of AI models will also need to consider any contractual protection that may be required in the event that a corrective measure relating to the developer's processing has an impact on the acquirer's subsequent use of the model.

The third scenario involves unlawful processing in the development phase of an AI model, in circumstances where the model itself is anonymised and personal data is subsequently processed in the deployment phase. Here, the EDPB's position is that the GDPR does not apply to the operation of the model, and that the unlawfulness in the training stage does not affect the subsequent processing of personal data. It does not matter whether the subsequent processing is carried out by the developer of the AI model or by a third party controller. There is, in other words, no general doctrine of 'fruit of the poisonous tree' that would enable a supervisory authority to require a controller to delete or stop using an anonymised AI model, merely because that model has been trained unlawfully with personal data. However – and here we come full circle – the EDPB emphasises the need for supervisory authorities to examine thoroughly a controller's claim that its model is in fact anonymous.

[1] Opinion 28/204 on certain data protection aspects related to the processing of personal data in the context of AI models, available here.

[2] Opinion, paragraph 26.

[3] *Ibid*, paragraph 21.

[4] *Ibid*, paragraph 17.

[5] *Ibid*, paragraph 29.

[6] *Ibid*, paragraph 31.

[7] Case C-582/14, *Breyer*.

[8] *Ibid*, paragraph 34.

[9] *Ibid*, paragraph 38.

[10] *Ibid*, paragraph 43.

[11] Available here.

[12] Article 6(1)(f) GDPR.

[13] Guidelines 1/2024 on processing personal data based on Article 6(1)(f) GDPR, available at https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2024/guidelines-12024-processing-personal-data-based_en .

[14] https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/generative-ai-first-call-for-evidence/

[15] https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/response-to-the-consultation-series-on-generative-ai/

[16] Case C-621/22, *Koninklijke Nederlandse Lawn Tennisbond*, paragraph 49; Opinion, footnote 54.

[17] Opinion, paragraph 73.

Gibson Dunn lawyers are available to assist in addressing any questions you may have regarding these issues.

Please contact the Gibson Dunn lawyer with whom you usually work, or any leader or member of the firm's Artificial Intelligence or Privacy, Cybersecurity & Data Innovation practice groups:

Artificial Intelligence:
Keith Enright – Palo Alto (+1 650.849.5386, kenright@gibsondunn.com)
Cassandra L. Gaedt-Sheckter – Palo Alto (+1 650.849.5203, cgaedt-sheckter@gibsondunn.com)
Vivek Mohan – Palo Alto (+1 650.849.5345, vmohan@gibsondunn.com)
Robert Spano – London/Paris (+33 1 56 43 13 00, rspano@gibsondunn.com)
Eric D. Vandevelde – Los Angeles (+1 213.229.7186, evandevelde@gibsondunn.com)
Frances A. Waldmann – Los Angeles (+1 213.229.7914, fwaldmann@gibsondunn.com)

Privacy, Cybersecurity & Data Innovation:
Ahmed Baladi – Paris (+33 1 56 43 13 00, abaladi@gibsondunn.com)
Ashlie Beringer – Palo Alto (+1 650.849.5327, aberinger@gibsondunn.com)
Joel Harrison – London (+44 20 7071 4289, jharrison@gibsondunn.com)
Jane C. Horvath – Washington, D.C. (+1 202.955.8505, jhorvath@gibsondunn.com)
Lore Leitner – London (+44 20 7071 4987, lleitner@gibsondunn.com)
Vera Lukic – Paris (+33 1 56 43 13 00, vlukic@gibsondunn.com)
Rosemarie T. Ring – San Francisco (+1 415.393.8247, rring@gibsondunn.com)