



Public consultation reference: 10/2024
Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR
20 November 2024

European Data Protection Board
Rue Wiertz 60
B-1047 Brussels
Belgium

For the European Data Protection Board (EDPB)

On behalf of Chamber of Progress – a tech industry association supporting public policies to build a more inclusive society in which all people benefit from technological advance – I write in response to the public consultation on “Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR”.

You can find our detailed comments below in Annex I.

Sincerely,

A handwritten signature in black ink that reads "Kayvan Hazemi-Jebelli".

Kayvan Hazemi-Jebelli
Senior Director, Europe

ANNEX 1.

Chamber of Progress Submission to EDPB on Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR

Chamber of Progress welcomes this opportunity to respond to the public consultation on “Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR”¹ (Guidelines). While we appreciate the European Data Protection Board’s (EDPB) efforts to provide clarity around legitimate interests processing, we have significant concerns that the proposed Guidelines take an overly restrictive approach that could severely hamper technological innovation, particularly in artificial intelligence, without providing proportionate privacy benefits. Our response focuses on three key areas where we believe the Guidelines require improvement.

First, following the Guidelines would disregard the technical reality of how modern AI systems actually work. These systems do not store or process personal data in traditional ways, but rather create statistical representations and probabilistic relationships between tokens. This apparent technical misunderstanding leads to inappropriate application of data protection concepts that could needlessly restrict beneficial AI development.

Second, the Guidelines take an excessively cautious approach to data processing across multiple dimensions. The requirement that processing for purposes of generative AI be “strictly necessary” sets an unnecessarily high bar. The Guidelines’ apparent adoption of an “absolute” rather than “relative” concept of personal data, combined with an overly broad interpretation of special categories data and problematic treatment of reasonable expectations, creates an impractical framework that could make many beneficial data processing activities impossible.

Third, while the Guidelines acknowledge certain economic interests, they fail to adequately consider the broader societal benefits that data-driven innovation enables. A more balanced approach would better serve both privacy protection and technological progress.

Technical Reality of Modern AI Systems

The Guidelines’ approach to legitimate interests processing fails to account for how modern AI systems, particularly large language models (LLMs), actually function. These systems do not store or process personal data in traditional ways that the GDPR was

¹ https://www.edpb.europa.eu/system/files/2024-10/edpb_guidelines_202401_legitimateinterest_en.pdf

primarily designed to regulate. Instead, they create statistical representations and probabilistic relationships between tokens - the basic units of text that might be syllables, words, or parts of words.

When an AI model processes training data, it does not memorise or store personal information. Rather, as the Hamburg Commissioner for Data Protection and Freedom of Information has concluded in its Discussion Paper on Large Language Models and Personal Data, "The mere storage of an LLM does not constitute processing within the meaning of article 4 (2) GDPR. This is because no personal data is stored in LLMs."² The model maintains aggregated statistics based on aggregated learnings that represent, for example, how often certain words appear together across millions of tokens in the training dataset.

This technical reality has important implications for privacy risk assessment. The training process fundamentally transforms input data into statistical representations, where individual data points are not preserved or retrievable. Accordingly, and as the European Data Protection Supervisor noted in its June 2024 Orientations on Generative AI, "service providers of generative AI models may use legitimate interests under the GDPR as a legal basis for data processing ... taking into account the specific conditions of processing carried out by these systems."³

The Guidelines' apparent assumption that AI training data processing carries the same privacy implications as traditional data processing reflects a fundamental misunderstanding of the specific conditions of processing carried out by those systems. This could needlessly restrict beneficial AI development. A more technically informed approach would recognize that:

- 1. The purpose of AI training is not to process personal data but to understand language patterns and relationships;**
- 2. The statistical nature of model training means individual data points are not preserved or retrievable;**
- 3. Appropriate technical safeguards like de-identification and filtering can further minimise any theoretical privacy risks;**
- 4. Output-level controls can effectively address any residual privacy concerns without restricting beneficial training.**

²

https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/240715_Discussion_Paper_Hamburg_DPA_KI_Models.pdf

³ https://www.edps.europa.eu/system/files/2024-06/24-06-03_genai_orientations_en.pdf

Overly Cautious Approach to Data Processing

The Guidelines adopt an excessively restrictive approach to data processing across multiple dimensions that goes beyond what the GDPR requires and what practical innovation demands.

First, the Guidelines' interpretation that processing must be "strictly necessary" and "carried out only in so far as is strictly necessary" (paras. 28-29) sets an unnecessarily high bar in the context of generative AI. While necessity is an important principle, the Guidelines' interpretation effectively requires controllers to prove a negative - that no less intrusive means could possibly achieve their purpose. This is particularly problematic for AI development where broader training datasets often lead to more accurate and less biased models. The GDPR's actual text requires processing to be "necessary" for legitimate interests, and case law does not require processing to be "strictly necessary" in the context of generative AI. This distinction matters for practical implementation, and the EDPB should leave that question open.

Second, the Guidelines appear to adopt an 'absolute' rather than 'relative' concept of personal data. This approach would mean that information must be treated as personal data if any entity anywhere could theoretically link it to an individual, regardless of whether the processing organisation has any realistic means to do so. This interpretation appears to be in tension with CJEU precedent in Breyer (C-582/14)⁴, which supports a more contextual approach focused on whether the controller has legal means to identify the individual.⁵ As the Court held in that case, information would not constitute personal data where identification "was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant."⁶

The Guidelines' treatment of special categories data is similarly overbroad, focusing on whether it is "objectively possible to infer sensitive information"⁷ rather than actual risks and context. This approach could make AI training practically impossible since any sufficiently large dataset might theoretically allow certain inferences. A more reasonable interpretation would follow CJEU guidance requiring "a certain degree of probability"⁸ rather than mere theoretical possibility.

Additionally, the Guidelines' statement that "reasonable expectations do not necessarily depend on the information provided to data subjects"⁹ creates significant uncertainty about how controllers can establish reasonable expectations. This appears to diminish

⁴ C-582/14 - Breyer, <https://curia.europa.eu/juris/liste.jsf?num=C-582/14>

⁵ Ibid, para. 49

⁶ Ibid., para. 46

⁷ Guidelines, para. 40

⁸ Case C-21/23 - ND v DR, <https://curia.europa.eu/juris/documents.jsf?num=C-21/23>, para. 90

⁹ Guidelines, para. 53.

the value of transparency measures and privacy notices that the GDPR itself emphasises as key safeguards.

More practically workable approaches to these issues would include:

- 1. Recognizing that "necessary" does not mean "strictly necessary" but rather "reasonably necessary" in the context of generative AI;**
- 2. Adopting a relative approach to personal data that considers realistic identification capabilities;**
- 3. Taking a risk-based approach to special categories data focused on actual rather than theoretical privacy impacts;**
- 4. Acknowledging that proper transparency measures can help establish reasonable expectations.**

These interpretations would better align with both GDPR's text and practical innovation needs while maintaining strong privacy protections.

Broader Societal Benefits

While the Guidelines acknowledge certain economic interests, they fail to adequately consider the broader societal benefits that data-driven innovation enables. AI systems trained on diverse datasets can boost economic growth, enhance scientific research, improve healthcare outcomes, and foster cultural exchange. The EU's own Digital Strategy and AI Strategy recognize these benefits, yet the Guidelines' restrictive approach could severely limit their realisation.

As noted in recent economic analysis, AI is expected to generate trillions in value across sectors like healthcare, energy, and scientific research.¹⁰ Moreover, these technologies play a crucial role in preserving European cultural diversity and linguistic representation. Without appropriate access to training data, European AI systems cannot effectively reflect local languages, customs, and cultural contexts, potentially leaving European citizens reliant on systems trained primarily on non-European data.

Conclusion

The EDPB's Guidelines, while well-intentioned, require revision to create a workable framework that balances privacy protection with technological progress. **We recommend:**

- 1. Updating the Guidelines to reflect the technical reality of modern AI systems, particularly recognizing how their statistical nature fundamentally differs from traditional data processing;**

¹⁰ <https://implementconsultinggroup.com/article/the-economic-opportunity-of-generative-ai-in-eu27>

- 2. Adopting more practical interpretations of necessity, personal data, and special categories;**
- 3. Considering the broader societal benefits that balanced regulation can unlock.**

Chamber of Progress stands ready to engage with the EDPB to improve guidelines that protect privacy while fostering technological advancement that benefits all Europeans. We believe that with appropriate revisions, “Guidelines 1/2024 on processing of personal data based on Article 6(1)(f)” GDPR can better serve both objectives.