GDPR & AI

Open Questions and Recommendations

Personal Data: A Critical Component for Effective AI Model Development

1. Data Volume and Sources in AI Training

The training of AI models, particularly large language models (LLMs), requires vast datasets, often involving millions or billions of data points. These data points are sourced from a wide range of materials, including web documents, books, code repositories, image, audio, and video data, as well as proprietary data and data licensed from third parties.

2. The Inclusion and Importance of Personal Data

While personal data typically constitutes a small fraction of overall training data, it plays a crucial role in model development. Examples of personal data included in training datasets include names and biographical information published in sources like Wikipedia, professional data such as job titles and educational histories posted on professional websites, and commentary on living persons, found in newspaper articles and other publications. The inclusion of some personal data contributes to reducing bias and improving model accuracy and performance. Information about people is useful in understanding language without losing meaning and context. LLMs use such data to learn how language incorporates concepts about relationships between people and the world. For instance, a model generating text about a historical or current event needs to accurately identify and use the proper names of people, places, and organizations involved. Thus, excluding, masking, or filtering out personal data from training data can hinder an LLM's ability to understand language and harm the quality of the model's outputs. Similarly, as generative AI (GAI) applications like personalized medicine, individualized tutors, and personal agents evolve, models will need to continue learning from personal data to be effective and meet consumer expectations.

3. Challenges in Filtering Personal Data

Identifying and filtering personal data within large-scale training datasets presents significant challenges. Difficulties include distinguishing between factual information and fictional content, determining whether a person is living or deceased, and identifying whether a word is a name and what data may be reasonably linked to it. Attempting to filter out personal data can result in quality issues and unpredictable outcomes. All filters have false positive and false negative rates, which vary depending on the content and filter configuration. High false positive rates are particularly detrimental, as they lead to the removal of valuable training data that are not actually personal data. For example, it was found that a filter designed to remove a certain identifier also removed numerical coordinates that were critical to the geometric understanding of the AI. In some cases, false positives are unavoidable and can create significant problems for model training.

4. Challenges in Removing Data Influence Post-Training

Identifying the specific training data points responsible for a particular model output is extremely difficult. Even if these data points could be identified, removing their influence from a trained model remains an unsolved computer science research problem. While retraining the model without certain personal data is theoretically possible, it is extremely resource-intensive and may have quality implications.

1. GDPR's Principle of Data Minimization *vs* the large-scale data needs of AI

- **Problem –** The GDPR's data minimization principle (Article 5(1)(c)) mandates limiting data collection to what is strictly "necessary" for a specified purpose. However, modern AI, especially General Purpose AI (GPAI) and Large Language Models (LLMs) require massive datasets for effective training, accuracy, and to mitigate biases. With the rapid progress of AI, it is crucial to avoid making assumptions about what data is needed to train a model, how long data needs to be retained, and the impact of deleting, pseudonymizing, or anonymizing training data.

  At the training stage, organizations can implement data minimization safeguards such as developing a responsible data collection framework and using technologies like data scrubbing and synthetic data. However, data minimization doesn't necessarily mean using small data volumes for AI training. There are important trade-offs to take into account. Dataset size is essential for model quality, for mitigating bias, and for ensuring model safety.

- **Risk –** A rigid application of data minimization risks impeding the development and utility of AI applications, thereby weakening the EU's global competitiveness and hindering its ability to capitalize on AI's economic potential. Limiting dataset size not only hinders innovation but can also directly compromise AI performance, safety, quality and fairness.

Recommendation 1 – Advocate for a flexible, contextual interpretation of data minimization that is proportionate to the risks and benefits of AI.

- Proportionality is key when applying data minimization to AI. We should move beyond a purely quantitative, restrictive approach and recognize that "necessary" data in the AI context includes the large data volumes essential for robust model training, high-quality outputs, and bias mitigation. The French DPA acknowledged that using large datasets for AI training do not necessarily equate to a violation of the data minimisation principle (CNIL).

2. GDPR's purpose limitation principle *vs* General-Purpose AI and the AI Act

- **Problem:** The GDPR's purpose limitation principle (Art. 5(1)(b)) mandates that personal data be collected for "specified, explicit and legitimate purposes." This principle, while sound for traditional data processing, clashes with the inherent nature of training General-Purpose AI (GPAI). That is because GPAIs are designed for a wide and evolving range of applications, many of which are unforeseen at the training stage, but are determined by the users of these systems in the deployment phase. The general purpose of developing and building a GPAI model should, in itself, be considered a sufficient and legitimate purpose under the GDPR at the development phase.

- **Risk:** A narrow interpretation of "specified purpose" that fails to recognize general purpose AI model development as a sufficiently specific legitimate purpose in itself could act as a significant barrier to GPAI development, preventing the emergence in Europe of novel and beneficial AI applications that cannot be anticipated *ex-ante*.

Recommendation 2 – Recognize "developing and building a general purpose AI model" itself as a legitimate purpose under GDPR.

- Training general-purpose AI models should be recognized as a legitimate and permissible purpose in itself, so long as appropriate accountability measures and safeguards are reasonably and sufficiently implemented (CIPL, 2024). Given the wide potential societal benefits of such models - and the demand from academia, businesses, civil society and citizens - there is a clear legitimate interest in the research and development of general purpose AI models.

3. GDPR's Accuracy Principle *vs* AI Hallucinations

- **Problem:** The GDPR's accuracy principle (Article 5(1)(d)) requires personal data to be accurate and up-to-date, posing a direct challenge to the probabilistic nature of LLMs, which are known to generate "hallucinations" – outputs that are factually incorrect or nonsensical despite sounding plausible. Applying the GDPR's accuracy requirement in the traditional sense may be neither feasible nor appropriate for these complex systems (Christakis, 2024).

- **Risk:** Holding LLM providers to an absolute accuracy standard under the GDPR could create disproportionate enforcement risk and stifle innovation without meaningfully improving data subject protection in this unique context.

Recommendation 3 – Advocate for a risk-based, contextual interpretation of "accuracy" under the GDPR

- When it comes to AI, "accuracy" should be understood as *reliability appropriate to the intended purpose* and focus on *mitigating harmful inaccuracies* through transparency and user empowerment mechanisms (e.g., disclaimers, fact-checking features). The EDPB should clarify that "statistical accuracy" for AI-generated outputs is distinct from the "GDPR's personal data accuracy obligations" (UK ICO). “Statistical accuracy” should be assessed contextually, focusing on mitigating real-world harms.

4. GDPR's user controls *vs* the complex technical nature of AI

- **Problem:** Implementing GDPR rights to rectification and erasure face significant *practical limitations* due to the nature of AI model training on vast, unstructured datasets. It is often technically impossible or unrealistic as it would require a complete re-training of the model, as acknowledged by the French DPA (CNIL).
- **Risk:** Attempting to directly implement rectification and erasure *within the training data of a model* is often technically impossible without compromising model functionality or requiring complete retraining, rendering such efforts disproportionate. This creates significant legal uncertainty and hinders responsible AI development.

Recommendation 4 – Focus the implementation of safeguards and user controls at the application layer of AI systems

- While data protection considerations are relevant throughout the AI lifecycle, the application layer is where the most effective safeguards can be implemented. These safeguards can include fine-tuning of models to reduce bias and improve accuracy, classifiers to filter out harmful or inappropriate content, and user controls over their data and prompts . By implementing these safeguards at the output stage, we can mitigate the risks associated with AI while maximizing its benefits.

5. GDPR's restrictions on Special Category Data *vs* made-for-Europe AI models

- **Problem –** GDPR Article 9's strict prohibition on the processing of *special category data* (race, ethnicity, health data, etc.) could create obstacles to the development of beneficial AI applications, especially in vital sectors like healthcare. Many such applications *require* processing sensitive data (Novelli et al, 2024). Moreover, AI models need to reflect wider cultural and social contexts, including sensitive topics, to be effective, accurate, and unbiased. The European Parliament, in a recent study, echoes these concerns, noting that "*the GDPR, which imposes limits on the processing of special categories of personal data, might prove restrictive in a context dominated by the use of AI in many sectors of the economy, and faced with the mass processing of personal and non-personal data*".

  The AI Act, on the other hand, recognises the need for special category data (SCPD) for bias mitigation efforts. AIA Article 10(5) explicitly allows for processing SCPD for "bias detection and correction" in high-risk AI systems. However, this allowance is limited to high-risk systems, creating a gap for other AI systems and, crucially, General Purpose AI (GPAI) models.

- **Risk –** Overly strict interpretation of Art. 9 GDPR could prevent the development of life-saving AI applications in healthcare, accessibility, and other areas reliant on sensitive data. Crucially, it would also hinder the training of AI models on European data, which must include sensitive topics to accurately reflect European languages, cultures, and societal specificities. Removing all information with potential references to sensitive information would lead to AI models that are less performant and relevant for European users, do not accurately represent minority groups or cultures and would fail to uphold cultural, religious and linguistic diversity. This directly harms the EU's cultural and linguistic diversity goals and undermines the development of AI that truly serves European citizens.

  Finally, restricting European AI developers from processing SCPD, even for legitimate and beneficial purposes, puts them at a significant disadvantage compared to developers in jurisdictions with more flexible data protection frameworks.

Recommendation 5.1 – Advocate for a pragmatic interpretation of GDPR Art. 9

- Call for the EDPB to develop guidelines that inform a more pragmatic interpretation of GDPR Art. 9. This could be done by broadening the scope of applicable legal bases under GDPR Article 9(2), for instance exploring the relevance of "scientific research" (Art. 9(2)(j)) or “substantial public interest” (Art. 9(2)(g)) for AI development. This would provide a more comprehensive framework for lawful SCPD processing in AI contexts, and ensure the development of AI models adapted to the European continent.

Recommendation 5.2 – Encourage the use of Privacy-Enhancing Technologies

- Strongly encourage and incentivize the use of Privacy-Enhancing Technologies (PETs) to minimize privacy risks when processing SCPD for AI.

Recommendation 5.3 - Extend the AIA's SCPD Allowance Beyond High-Risk Systems

- Advocate for extending the AIA's Article 10(5) allowance for SCPD processing for bias mitigation to all AI systems and GPAI models, not just those classified as "high-risk." Limiting this crucial provision to high-risk systems creates a counterproductive restriction. Bias detection and correction, and the development of representative, culturally relevant AI, are essential for all AI systems, regardless of their risk classification.

6. GDPR enforcement *vs* the need for sector-specific expertise for AI

- **Problem –** GDPR enforcement can exhibit "privacy myopia," prioritizing maximum privacy protection without sufficient consideration of innovation and other societal interests (Barczentewicz, 2025). That is contradictory to the spirit of GDPR Recital 4, which provides that

"the processing of personal data should be designed to serve mankind", and that data protection should be balanced with other EU fundamental rights, including the right to freedom of thought, expression and information, the right to education or the freedom to conduct a business.

The novel challenges and opportunities of AI therefore call for a broader perspective from regulators. However, public consultations are not always conducted by authorities in charge of regulating AI, including for industry-critical guidance from the EDPB (e.g. Art. 64(2) EDPB opinions). Effective AI regulation also calls for *sector-specific AI expertise* within regulatory bodies. Some Member States still haven't decided who will enforce the AI Act, which creates significant uncertainty for companies.

- **Risk –** A rigid GDPR enforcement approach, without a nuanced understanding of AI and its societal benefits, creates a chilling effect (as already demonstrated by several major AI actors delaying the launch of AI features in the EU). The implementation of the AI Act and the application of GDPR to AI will both require cross-regulators collaboration and public consultations if it is to be effective, pertinent and grounded in market reality.

Recommendation 6.1 – Call for mandatory interregulator collaboration and public consultations

- Implement the EDPS' idea of a Digital Clearinghouse for the collaboration of privacy, competition, and consumer protection authorities on digital policies enforcement. Mandate public consultations on all guidance documents from the EDPB.

Recommendation 6.2 – Broader Accountability Framework for DPAs

- Explore mechanisms to ensure DPAs demonstrably consider the broader societal and economic impacts of their enforcement decisions, ensuring a balanced approach that goes beyond a singular focus on maximizing privacy protection and incorporates the AI Act's integrated risk-benefit perspective.

ADDITIONAL INFORMATION - RECENT EUROPEAN DEVELOPMENTS

EDPB Opinion on AI models (December 2024)

On 18 December 2024, the EDPB published its opinion on certain data protection aspects related to the processing of personal data in the context of AI models. The opinion came at the request of the Irish DPC under Art.64(2) GDPR, following injunctive proceedings taken by the Irish DPC against X for its use of personal data to train its generative AI model.

The EDPB's opinion presents several **constructive elements**. It acknowledges the societal opportunities and benefits offered by AI. Furthermore, it helpfully recalls that the GDPR framework is designed to encourage responsible innovation, and it reaffirms that GDPR is meant to promote data protection in a way that is also balanced with other fundamental rights, such as the freedom to conduct a business and freedom of expression and information. Crucially, the EDPB acknowledges legitimate interest as a valid legal basis for the development and deployment of AI models, representing a welcome confirmation that avoids what would have been an outright prohibition of AI training in Europe.

However, the opinion also introduces some additional burdens or uncertainties:

- **Uncertainty due to "Case-by-Case Analysis" Overload:** The Opinion's excessive reliance on "case-by-case analysis" provides insufficient legal certainty for businesses. This fragmented approach leaves substantial discretion to individual DPAs, creating an unpredictable regulatory landscape.
- **Onerous Compliance Burden and High Anonymization Standard:** The EDPB sets a high bar for demonstrating AI model anonymity. This creates an overly burdensome compliance regime, particularly for SMEs and smaller AI actors, potentially requiring significant expenditure on legal and technical expertise without clear guarantees of compliance.
- **Key Issues Left Unaddressed:** The Opinion fails to address crucial questions such as the permissibility of processing Article 9 data for AI training and offers limited guidance on key aspects of GDPR application in the AI context. This lack of clarity increases uncertainty for AI developers.