Public consultation on EDPB Guidelines 02/2026 on anonymisation

From: Vasanth Rajendran, individual expert contributor (personal capacity; views are my own)

From: Vasanth Rajendran, individual expert contributor (personal capacity; views are my own)

Dear EDPB,

I welcome these Guidelines and submit the following comments as an individual expert. I am a Software Development Manager at Amazon, where I lead a team shipping production agentic AI systems, and I serve as a Privacy Bar Raiser with sign-off authority on privacy design for product launches. I chair the IEEE P1968 working group developing a recommended practice for the governance of autonomous AI agent systems (IEEE standards project approved September 2026), I developed the SAFE-AGENT-L governance framework presented at the AAAI-26 AI and Law Bridge workshop, and I provided expert input to NIST's TEVV-Athlon work (AI 200-2). My comments focus on the technical bar for anonymisation where AI systems are involved, drawing on production experience with AI training data and model evaluation.

1. Apply the three criteria to AI training corpora and to model outputs: anonymised input does not imply an anonymous model (Section 3.4, pp. 18-24; para 46).

The three criteria (No Record Isolation, No Linkage, No Inference) are formulated for datasets, but the increasingly common claim is that a model trained on personal data is itself anonymous. Model memorisation, membership inference, and extraction attacks mean training data can be recovered from model weights, and generative models can regurgitate training data in outputs, implicating all three criteria at both the corpus and the output level. I recommend the Guidelines require a separate model-output re-identification risk assessment in addition to any assessment of the training corpus, and state that a model trained on personal data is presumed to remain personal data unless the controller demonstrates otherwise through a dedicated assessment: for example, differential-privacy guarantees with stated parameters, or empirical

extraction and membership-inference testing with documented methodology and thresholds. Anonymised input does not imply an anonymous model, and the Guidelines should say so explicitly.

2. Assess "means reasonably likely" against current AI capabilities, with explicit reassessment triggers (Section 2.4.2, p. 11; paras 43-44).

The draft correctly requires the technical analysis to consider the state of the art and future developments in re-identification. That assessment must now be made against current AI capabilities, not historical attack catalogues: ML-assisted linkage and inference using public foundation models are among the means reasonably likely to be used by a motivated intruder today. I further recommend making anonymisation a continuing obligation with explicit reassessment triggers: the release of materially more capable public models, the emergence of new linkable auxiliary datasets, or any material change in how the data or model is shared. Anonymisation cannot be a one-time determination where the adversary's toolkit improves every year.

3. Confront the identifier-removal fallacy and give quantified ways to evidence "insignificant in reality" (para 22; Sections 3.4-3.5, pp. 18-29).

Two sides of the same evidentiary coin. First, the Guidelines should confront directly, with an example, the persistent fallacy that removing direct identifiers equals anonymisation: identifier-stripped high-dimensional data routinely fails both No Linkage and No Inference, and any anonymisation claim for AI training data must rest on the full three-criteria assessment, documented, not on identifier removal alone. Second, to make the "insignificant in reality" test in para 22 operational, the Guidelines should recognize formal privacy methods as ways to evidence it: differential privacy with a stated epsilon and documented composition accounting, or k-anonymity with a stated k and demonstrated resistance to the three criteria. Together, the fallacy correction and the quantified methods give controllers both the floor and the ladder.

4. Clarify when synthetic data counts as anonymous (general; Section 3.4).

Synthetic data is increasingly presented as an anonymisation measure in AI training pipelines. But synthetic data that preserves statistical properties of real personal data can leak information about real individuals through reconstruction and membership-inference attacks on the

generator itself. I recommend stating that synthetic data is anonymous only if the generation process satisfies the three criteria in Section 3.4, and that training a generative model on personal data in order to produce "synthetic" data does not by itself achieve anonymisation. The generator, like any model trained on personal data, needs its own assessment under point 1 above.

5. Require the simplified approach for publicly released data and models (para 45).

The draft offers a contextualised approach assessing each relevant entity's means and a simplified approach disregarding differences between entities. For publicly released datasets or openly distributed model weights, the recipient set is unbounded and entity-by-entity assessment is impossible in principle. I recommend stating explicitly that public release requires the simplified approach, effectively assuming the most capable recipient, and that contextual arguments cannot support an anonymisation claim for openly published data or models. This closes the most dangerous gap in the framework: a contextual assessment valid on release day becoming false the moment the data meets the open internet and its AI-assisted adversaries.

I thank the EDPB for this important work and for the opportunity to comment.

Respectfully, Vasanth Rajendran (personal capacity)