AI Sweden’s comments on EDPB Guidelines 02/2026 on Anonymisation

AI Sweden is the Swedish national center for applied artificial intelligence. Our mission is to accelerate the use of AI for the benefit of our society, our competitiveness, and for everyone living in Sweden. We are broadly funded and not for profit, as a part of Lindholmen Science Park, and collaborate with over 180 partners from the private and public sectors as well as academia. Among our projects, AI Sweden leads LeakPro, an open-source framework for auditing the privacy risk of AI models and synthetic data, and works with health, public-sector and industry partners on the anonymisation and sharing of data for AI development.

AI Sweden welcomes the direction of the draft. The guidelines frame identifiability as a question of likelihood rather than theoretical possibility (para 8), and set the threshold at a likelihood that is insignificant in reality (para 22), which allows attack success, cost and access to be weighed together. Anonymity is assessed from the perspective of defined entities and the means realistically available to them (para 6, 11–12, 26, 29), and the content, purpose and effect test in para 16 gives a workable account of when information relates to a person. Failure of the record isolation or linkage criteria triggers further analysis rather than a verdict (para 52), and AI models and synthetic data are brought explicitly within scope (para 82).

The comments below fall into three parts. The first addresses the technical framework, where we identify points that would benefit from clearer formulation and propose concrete wording. The second examines systemic and logical constraints within the draft that make practical compliance difficult. The third considers the framework as it stands and how it would apply in practice to AI models that are released publicly.

Technical comments

The three criteria differ from Opinion 05/2014 and the difference should be stated

Para 2: The guidelines retain a three-criteria structure and present it as an update of the criteria in Opinion 05/2014. The new criteria, however, are not a renaming of the old ones. “No Record Isolation” is narrower than “singling out”: it asks only whether a record is unique within the given data, while the question of whether a unique record can be matched to a person has been moved to the follow-up analysis in para 98–100. “No Inference” is considerably wider than the 2014 criterion: it now covers membership inference (para 79), linkage that cannot be established on a one-to-one basis (para 80), and the extraction of record-level data from aggregates, models and synthetic data (para 76–77, 82). Much of the weight of the assessment has therefore moved to the third criterion.

This shift is easy to miss. The names invite a one-to-one mapping to the 2014 criteria, and practitioners who have built assessment methods around singling out, linkability and inference as defined in 2014 will apply them unchanged. The result would be assessments that pass the first two criteria and test the third against a definition that is now too narrow.

AI Sweden suggests that para 2 states that the criteria have been redefined, not only renamed, and summarise how each differs from its counterpart in Opinion 05/2014. A short comparison table would serve the same purpose.

The simplified approach needs a defined adversary

Para 22, 45, 48, 49, 67, 71: The simplified approach is presented as the more cautious of the two: by disregarding differences between entities, it is said to provide greater confidence (para 45) and to ensure that the GDPR standard is met in cases of doubt (para 48). Its pass condition is given in para 49: the controller asks “if re-identification is even possible in theory”, and “if this is not possible, the data can be safely considered anonymous”.

That pass condition contradicts the standard stated in para 22. Para 22 says the likelihood of identification need not be zero, only insignificant. Para 49 asks whether re-identification is possible at all, which is a zero-risk test. For any data that still carries information about individuals, some probability of re-identification remains, even after noise addition or differential privacy; it can be made small, not zero. No real data can therefore pass para 49 as written. In practice assessors will read “in theory” loosely, and the outcome will then depend on which attacks and which auxiliary data each assessor happened to consider. That is the dependence the simplified approach was meant to remove, so the confidence claimed in para 45 and 48 does not follow from the definition in para 49.

The problem is sharper for the No Inference criterion. Para 67 defines an inference as meaningful only if it "could not be obtained from general knowledge or from data about the population at large", and para 71 tests this by comparing what an adversary learns with and without the individual's record. Both depend on what a particular adversary already knows. Under the simplified approach there is no particular adversary, so there is no prior to compare against and the test in para 71 cannot be applied. If the gap is filled by assuming an adversary who holds all auxiliary information, then a meaningful inference exists for almost any data and the criterion always fails. For synthetic data and models, where No Inference is in practice the only criterion that can be applied, the simplified approach would then always conclude that the data is personal.

AI Sweden suggests that para 49 define the simplified approach as an assessment against one explicit adversary, assumed to hold all auxiliary information available to any relevant entity, including non-public information, together with knowledge of the anonymisation process and the current state of the art in re-identification techniques, with effectiveness judged under para 87. This replaces an unmeetable condition with one that can be met, keeps the approach conservative, and makes assessments of the same data by different assessors comparable.

Likelihood as a measurable quantity

Para 8, 22, 29, 41: The guidelines correctly frame identifiability as a question of likelihood rather than absolute possibility, and set the threshold at a likelihood that is "insignificant in reality". AI Sweden supports this framing. However, the guidelines do not require that this likelihood be demonstrated. Para 29 lists the factors that influence it and para 41 requires the testing to be documented, but nothing obliges the controller to state what adversary was assumed or to produce a result that another party could reproduce. As a result, a controller can conclude that the likelihood is insignificant without producing any evidence that could be checked by a supervisory authority or a third party.

Such a demonstration is feasible. Re-identification risk can be estimated empirically against a stated adversary, in a way that is reproducible and can be repeated when the state of the art changes, as para 35 and 84 call for. The guidelines do not need to prescribe how this is done, but they should make clear that it is expected.

The guidelines already require the controller to identify relevant entities and their means (para 11–12, 30). What is missing is the corresponding requirement to measure and record the outcome of testing against those entities. AI Sweden therefore suggests that para 41 be extended so that the documentation of testing includes the adversary model assumed, the method used to estimate re-identification risk, the result obtained, and the threshold applied. We do not propose that the EDPB fix a numerical threshold, as the appropriate level depends on context. The proposal is that "insignificant" be shown through a measurement

that can be audited. AI Sweden develops open-source tooling for this type of assessment (LeakPro) and would welcome a reference in the guidelines to the use of scientifically grounded, openly available auditing methods, consistent with para 84.

The No Inference counterfactual and differential privacy

Para 71, fn 43: The test in para 71(b), comparing inferences drawn from A(D) and from A(D') with one record removed, is the definition of differential privacy, used as a test on the output rather than as a property of the process. The guidelines do not name differential privacy, so the test gives neither the controller nor the assessor a way to state how close two inferences must be for the inference to count as not meaningful. Differential privacy expresses exactly that through its parameters. A controller who has applied differential privacy has formally bound what para 71(b) asks about, yet has no stated way to use that to satisfy the criterion.

AI Sweden suggests that the guidelines name differential privacy as the formal counterpart of the test in para 71(b), and state that a documented differential privacy guarantee, with its parameter and justification recorded under para 41, constitutes evidence that the No Inference criterion is met.

Models and synthetic data are not aggregate data

Para 82, 85, 86: Para 82 treats AI models and synthetic data as data "aggregated to represent correlations", and para 85 lists the factors that determine whether a re-identification technique is effective: whether the data is aggregated or record-level, its dimensionality, resolution, diversity, the number of individuals, and the available auxiliary information. These factors describe tables. A trained model has no rows, columns or resolution, and the same is true of the generator behind a synthetic dataset. What determines whether an attack on a model succeeds is different: how much of the training data the model has memorised, whether the attacker has access to its outputs only or to its weights, and how many queries the attacker can make. None of these appear in para 85.

Classifying models as aggregate data also carries over the rule of thumb in para 86 that re-identification is less likely to be effective against aggregates. That assumption does not hold for models. Memorisation is not reducible to overfitting; a model can generalise well on average and still encode rare or outlying records, which is exactly what membership and extraction attacks exploit. However, it is essential to recognise that a model may memorise training data without that information being realistically extractable by an adversary. Such attacks are among the most effective techniques currently documented, and the guidelines refer to them in para 77.

AI Sweden suggests that para 82 state that models and synthetic data constitute a separate category, and explicitly clarify that the anonymisation status of the training dataset and that of the resulting model are separate questions. A model trained on personal data is not necessarily personal data itself. Furthermore, para 85 should be complemented with the factors relevant to that category: the degree of memorisation of individual records, the form of access available to the attacker (outputs only or weights), and the number of queries the attacker can make. For synthetic data, the size of the released sample relative to the original data is a further factor. Access conditions can also change over a model's lifetime. The same model might initially be available through an API and later released with open weights. It would be useful if the guidelines clarified how such changes should affect an existing anonymisation assessment, in line with para 35 and 88. These are also the factors a controller can adjust and document, which is why they belong in para 85. The attack categories relevant to this assessment are already set out in EDPB Opinion 28/2024 on AI models, namely membership inference, attribute inference, model inversion and reconstruction, and the extraction or regurgitation of training data. AI Sweden suggests that para 82 refer to that list, so that the two documents describe the same assessment.

The "record" definition and synthetic data

Para 5, 55, 82: A record is defined in para 5 as all values in the given data that are "known to relate to the same individual", and under para 55 the No Record Isolation criterion is met when no such record is unique. For an ordinary de-identified table this works. For synthetic data it does not, because no synthetic row is known to relate to anyone. Taken literally, synthetic data contains no records, so the criterion is met whether or not the generator has memorised real people. The same holds for any output by a model trained on the original data; a synthetic data generator is one such model.

This is not a theoretical concern. Generators do reproduce training records, in full or in part, when they overfit or memorise. A synthetic row that is a near copy of a real patient relates to that patient in the same way a de-identified row would. The current wording hides such rows from the first criterion only because the controller cannot say which real person they belong to.

Para 82 places synthetic data under No Inference but does not say whether the other two criteria apply, and the definition of record makes them inapplicable in practice. AI Sweden suggests that para 82 state explicitly which criteria apply to synthetic data. If the assessment is meant to run under No Inference only, para 82 should say that reproducing a record from the original data, in full or in part, is a specific and meaningful inference. If all three criteria are meant to apply, the definition in para 5 should be amended so that a record is data which relates to an individual, whether or not that relation is known to the controller. Either route closes the gap; what matters is that memorised records are caught by one of them.

Systemic and logical constraints in the proposed framework

While the technical criteria provide a valuable structure for assessment, certain logical constructs within the draft create systemic dead-ends for practitioners attempting to comply in good faith. We wish to highlight three areas where the legal logic risks diverging from statistical and practical reality.

The all-or-nothing standard for large datasets

Para 36: The guidelines state that a dataset can only be considered anonymous if the likelihood of re-identification is insignificant for all included individuals. While logically sound for small, targeted datasets, this binary approach lacks proportionality when applied to web-scale datasets used in AI training, which often contain billions of records. If a single reliable extraction occurs (as referenced in para 87), the entire dataset is legally contaminated and treated as personal data. This creates an extremely fragile legal foundation for AI developers, where a single outlier negates the anonymisation of millions of others.

The infinite chain of responsibility for public data

Para 20, 21, 89: The draft correctly notes that publicly available data can be accessed by anyone with means reasonably likely to be used (para 89). However, under para 20 and 21, an entity assessing anonymity must account for the means and auxiliary information held by any third party to whom the data is indirectly transferred. For open-release models or public datasets, this effectively demands omniscience from the controller regarding the combined capabilities of unknown future actors globally. We suggest acknowledging that controllers can only reasonably assess risk against a defined, foreseeable adversary rather than a theoretical global collective.

The dichotomy of general knowledge and individual inference

Para 67 and 71 state that an inference is only meaningful (and thus violates anonymity) if it relies on the given data and could not be obtained from "general knowledge or from data about the population at large". This rule relies on an assumption from traditional statistics: that there is a clear, conceptual boundary between knowledge about a broad group and knowledge about a specific person.

However, in the context of high-dimensional data, which is standard for AI and modern datasets, this boundary evaporates. When data contains dozens or hundreds of attributes, a "population" can be defined so narrowly (e.g., by combining age, a specific postal code, profession, and a medical condition) that the statistical group practically consists of only one

person. In such cases, "general knowledge about the population at large" mathematically merges with "an inference about a specific individual". Because the legal distinction the guidelines rely on breaks down in high-dimensional space, it becomes practically impossible for a controller to prove that an inference is not meaningful under this definition.

Practicability for publicly released AI models

Para 35, 36, 84, 87, 89, 92: Taken together, four elements of the draft bear heavily on AI models that are released publicly, including open-weight models. Para 89 states that “if data is publicly available, anybody may be able to access it with means reasonably likely to be used”, which for an open-weight model implies that any attacker holds the weights. Para 92 notes that “developments in AI, and especially agentic AI, will likely further reduce the time and costs necessary to access and deploy these techniques”. Para 36 requires the likelihood of re-identification “to be insignificant for all of these individuals”, and para 87 adds that a technique “can be successful even if it only succeeds against a single individual in a larger dataset”, provided the result is sufficiently reliable to distinguish that individual. Para 35 observes that “the likelihood of re-identification typically increases over time”, and para 84 that the state of the art “is always evolving and should be considered during this assessment”. For a model trained on very large corpora, the combined effect is that anonymity can be lost by a single reliable extraction, by any attacker, at any later time. As the draft stands, anonymity cannot be assessed with reasonable confidence before release, so organisations developing open-weight models face significant legal uncertainty even when they have taken appropriate measures to minimise privacy risks.

While these elements stem from strict legal interpretations, their combined effect makes a clear connection to Opinion 28/2024 essential for compliance. The aforementioned opinion describes practical measures a controller can take before release, such as filtering of training data, differentially private training, and testing against membership and extraction attacks.

AI Sweden suggests that the guidelines state that the anonymity of a model is assessed as set out in Opinion 28/2024, that the measures described there are relevant to the means reasonably likely to be used against a publicly released model, and that documentation of those measures is the evidence expected under para 41.