Comments on Guidelines 02/2026 on Anonymisation

Submitted by: SentinelNexus AI (VionNexus SAS, in formation), Le Bourget, France — sentinelnexusai.com
Date: October 2026

We are a small French company building an EU-hosted security layer that sits between applications, AI agents and large language models. It detects prompt injection, masks personal data in prompts before they reach a model, controls agent tool calls, and keeps an audit log. The prompts we process routinely contain personal data, so the draft directly affects how we design and describe our product. We welcome the three-criterion framework and the explicit treatment of AI models. Our comments focus on four practical gaps.

1. Hashing needs explicit guidance

The draft discusses encryption (para. 91) and fingerprint-based pseudonyms (Example 6, para. 25), but we found no specific guidance on cryptographic hashing in the main text. Hashing is the most common technique in logging, analytics and AI security tooling, and it is widely described on the market as "anonymisation". A deterministic, unsalted hash of an input fails the No Linkage criterion (para. 60): anyone who holds the same input can confirm a match. For low-entropy inputs such as email addresses, phone numbers or IBANs, it can be reversed by enumeration. We suggest a short section or worked example distinguishing (a) unsalted hashes, (b) salted or keyed hashes where the controller keeps the key, and (c) keyed hashes where the key is destroyed. It would also help to state that an integrity hash computed over records containing personal data (for example, a hash-chained audit log) is a security measure, not anonymisation.

In the interest of transparency: our own product stores a SHA3-512 fingerprint of each prompt instead of the raw text. Reading para. 40 and the No Linkage criterion, we will treat and describe these fingerprints as pseudonymous personal data, not as anonymous data. An explicit example in the guidelines would help other vendors avoid the "hashed, therefore anonymous" claim.

2. Masking direct identifiers in free text before an LLM

A growing number of AI gateways replace direct identifiers in prompts with placeholders (for example "[EMAIL_1]") before forwarding them to a third-party model provider, and keep the mapping locally. Free text, however, often contains indirect identifiers (job title, employer, town, a rare event) that can fail No Record Isolation (para. 55) or No Inference (para. 67). We ask the EDPB to state, ideally with an example, that format-based masking of direct identifiers in free text should be presumed to be pseudonymisation unless the inference criterion is demonstrably met. We also ask for clarity on how para. 37 and the controller-perspective rule for processors apply when the mapping table is held by a processor (the gateway) rather than by the controller.

3. Evidence for AI models

We support the recognition that prompting is a way of querying data (para. 82), that extraction attacks exist (para. 77), that membership can be inferred (para. 79), and that AI agents can be used for re-identification (para. 83). Controllers now need to know what evidence is sufficient. We suggest the guidelines list the minimum tests expected before an AI model or its outputs are described as anonymous (membership inference, training-data extraction, extraction of hidden instructions or context), and state that documentation under para. 41 should record the model version, the test set, the results and the date. Automated, repeatable test suites run at each model update would be a proportionate way for SMEs to evidence the periodic reassessment described in para. 35.

4. Proportionate guidance for SMEs

The simplified approach (paras. 45–49) is helpful for small organisations that cannot model every possible recipient. However, "insignificant" likelihood of identification is not illustrated with any order of magnitude. A worked example showing how a small controller documents a simplified assessment, possibly reusing the structure of the EDPB DPIA template consulted on in 2026, would make the framework usable without specialist advice. We also suggest listing typical reassessment triggers: a data breach (para. 42), the release of new models or agents that lower the cost of linkage (para. 92), and changes in recipients or access.

Thank you for the opportunity to comment.

Abdur Rahim, co-founder, SentinelNexus AI — sentinelnexusai.com