Practical implementation comments on Guidelines 03/2026

Niuexa | Italy | 10 October 2026

Niuexa is an Italian company working on practical AI adoption, automation, professional training and visibility in AI-mediated search. We welcome the opportunity to comment on the draft guidelines. Our contribution concerns clarity and usability for smaller organisations and their implementation partners. It does not report a model-training experiment or quantitative findings, and does not propose exemptions from data protection obligations.

1. Make the scope usable in mixed technical workflows

Relevant text: paragraphs 2-6, printed page 5.

We recommend adding a short decision tree and worked examples distinguishing collection for training or fine-tuning from retrieval at the time of a user's request, ordinary search indexing, and use of an organisation's own documents. The examples should identify when a workflow moves from one purpose to another, such as when retrieved material is subsequently retained for training.

This would help readers identify which parts of a system the guidelines address without treating a technical label as a legal conclusion. Calling a product a retrieval system, an agent or a search tool should not determine the analysis by itself. The relevant processing steps, purposes, retention and later reuse should be described. The examples should retain the draft's clarification that activities outside its scope may still be subject to data protection rules.

For small organisations, a diagram showing the data flow and the boundary of each example would make this distinction easier to apply than a product-level classification. We suggest including one mixed workflow in which the same supplier provides retrieval and a separate training or fine-tuning service, with the data paths explicitly separated.

2. Extend the responsibility examples to an implementation chain

Relevant text: paragraphs 16-20, printed pages 8-9.

We recommend an additional example involving a small business, an AI implementation partner and a third-party model or dataset provider. It should show which factual decisions require examination: who determines the purpose, selects sources and data categories, sets instructions and retention, decides on subsequent reuse, and can act on a correction or deletion request.

A compact responsibility worksheet could help the parties document these facts and identify missing information. It should distinguish facts they have verified from supplier statements and unresolved questions. Roles should remain dependent on the actual arrangements and conduct; the worksheet should not assign a role merely because a party calls itself a consultant, platform, customer or processor.

The example should also show how an organisation records and escalates a material information gap. An unknown collection history should remain visible as unknown. A contractual label or a supplier's general assurance should not be presented as evidence that every relevant processing step has been examined.

3. Provide a proportionate provenance and checking template

Relevant text: paragraph 11, printed pages 6-7; paragraphs 30-38, pages 10-12; paragraphs 39-42, page 13.

We suggest a reusable example record linking a dataset or collection batch to its stated purpose, source or supplier, collection period, available publication or update dates, selection and exclusion rules, filtering steps, retained versions, checks performed and unresolved limitations. The template should explain what is necessary for each example rather than encourage collection of additional personal data merely to complete fields.

It would be useful to distinguish the date of collection from the age or last update of the underlying information. Similarly, validating a format is different from verifying factual accuracy. An example could demonstrate how a syntactically valid record is still outdated or unsupported, and how that finding is handled before further use.

For filtering and sampling, we recommend recording the method, what it can and cannot detect, and the response when a check fails. A sample that finds no issue should not be described as proof that the whole dataset contains none. Synthetic replacement or pseudonymisation should likewise be accompanied by an explanation of the assessment performed, rather than treated as a universal assurance.

Finally, source reliability, technical accessibility, source preferences and the assessment of permitted processing should remain distinct in the example. An official or well-maintained source may improve an accuracy assessment, but that fact alone does not answer every question about its proposed use.

These additions would help smaller organisations turn the guidelines into documented decisions while preserving their substance. We recommend adaptable examples and clear limits, rather than a checklist presented as automatic compliance or certification.

Reference

EDPB Guidelines 03/2026 on web scraping in the context of generative AI, version 1.0, adopted for public consultation on 7 July 2026.

Official consultation and contribution form.