Comments from Astroware on the EDPB Guidelines 02/2026 on Anonymisation

Response to the public consultation on Guidelines 02/2026, which closes on 30 October 2026

Astroware builds HaloGuard EU, a small guard model that sits in front of the output of AI assistants and agents. Its privacy filter looks for personal data in conversations and masks what it finds. Companies that use it receive the masked conversation, a map back to the original values, and the kinds of data found. Because of that map, we do not describe the filter's output as anonymous. At most it is pseudonymised, and it should be treated as personal data.

We welcome the Guidelines, in particular the three criteria and the clear statement in paragraph 50 that easily reversed pseudonyms are not anonymous. We offer three comments from the point of view of a company whose tool masks personal data in text sent to and from AI systems.

1. Add guidance and an example for free text, including text sent to AI systems.

The framework and its examples are built around records and attributes in structured datasets, such as the patient table in Example 10 and the purchase records in Example 11. The draft does not discuss free text: chats with AI assistants, support tickets, emails, or the documents an AI agent reads.

Free text has no fixed list of attributes. Identifying details sit inside sentences: "my manager at the Lisbon office", "the only night-shift nurse on the ward". Paragraph 24 already makes this point with "the one with the northern accent". We suggest a worked example showing how No Record Isolation, No Linkage and No Inference apply to a set of conversations where names, email addresses and phone numbers have been masked but descriptive details remain.

2. State that masking by automated detection is not presumed to produce anonymous data.

Most text masking relies on automated detection of personal data. Our published results show how far such detection goes and where it stops. Across six public benchmarks, our filter in strict mode finds 96.3% to 100% of personal values, in loose mode 88.9% to 99.1%, and the open model openai/privacy-filter at default settings 79.1% to 93.7%. Our filter also masks more text around personal data than the open model: about 2 to 3.5 times the personal data itself, against about 1 to 2 times. Each benchmark and its notes are listed at astroware.ai/haloguard-eu/privacy-benchmarks.

Three points follow from these results:

- **Detection is not complete.** Even in strict mode, five of the six benchmarks show some personal values not found.
- **Results vary widely by tool, setting and data.** On the same benchmark of records in 12 EU languages (REDACT), the share found ranges from 89.0% to 99.3%.
- **What a detector looks for matters as much as how well it finds it.** In loose mode our filter looks for direct identifiers such as names, email addresses, phone numbers and ID numbers. Strict mode adds any date, codes, places, organisations and job titles. These are the attributes that, combined, allow an individual to be singled out (paragraph 24, Example 5).

So the fact that masking was applied says little on its own about what is left in the text. We suggest the Guidelines say expressly that the output of detection-based masking is not presumed

anonymous. It should be tested against the three criteria like any other given data. The measured results of the detector on comparable data are relevant evidence for that test.

3. Confirm that reversible masking in AI pipelines is pseudonymisation.

A common pattern in AI products works like this. A company masks personal data in a user's message, sends the masked text to a hosted AI model, then puts the original values back into the reply using a map it keeps. Paragraph 50 treats easily reversed pseudonyms as a simple case of personal data. Paragraph 15 says that data processed on behalf of a controller is personal for the processor too.

We suggest a short example that applies both paragraphs to this pattern. It would confirm that the masked text remains personal data for the company and for a model provider acting on its behalf. It could point to the EDPB Guidelines 01/2025 on Pseudonymisation. This would not make masking less useful as a safeguard. It would make clear that it is not anonymisation.

Paragraph 40 asks controllers not to describe data as "anonymous", "de-identified" or "de-personalised" if individuals are still identifiable. Controllers often take that wording from their tool providers. The Guidelines could encourage providers of masking tools to describe their output in the same way. We describe ours as looking for personal data and masking what it finds.

A note on agentic AI. We agree with paragraphs 92 and 93 that AI, and especially agentic AI, will likely lower the time and cost of gathering and combining information from public sources. Free text is where this matters most. A detail that looks harmless on its own, such as a job title and a city, can be matched against public profiles by an agent. This supports applying the criteria to masked free text with the same care as to structured data.

Astroware
astroware.ai