Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

GDPR Accuracy and LLM Hallucination: What Regulators Said

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-10-29 · gdpr accuracy principle · gdpr article 5(1)(d) · chatgpt gdpr accuracy · llm hallucination gdpr personal data · edpb chatgpt taskforce accuracy

Original article: https://reliablecontext.com/articles/gdpr-accuracy-principle-and-llm-hallucination



In this article

1. The short answer
2. What Article 5(1)(d) says
3. Accuracy and purpose: when "up to date" matters
4. Input data and output data
5. The noyb complaints: rectification requests a model cannot satisfy
6. The UK position: opinions, inferences and mistakes
7. A worked example: one false statement, two architectures
8. Controls that help
9. Documentation to keep
10. Limits

The short answer

When a large language model (LLM) produces a false statement about an identifiable person, European regulators treat that output as personal data that must satisfy the accuracy principle in Article 5(1)(d) of the General Data Protection Regulation (GDPR). The European Data Protection Board's (EDPB) ChatGPT Taskforce said so plainly in May 2024: the probabilistic nature of the system does not suspend the principle, and transparency notices that warn users about errors are "not sufficient to comply with the data accuracy principle" (EDPB ChatGPT Taskforce report, paras 30-31). For an organization deploying LLMs, the practical consequence is that accuracy has to be designed and documented, and a person must be able to get a wrong statement about them corrected.

This article reads the legal text, the regulator statements and the two best-known complaints, then sets out controls that make the duty workable. It covers EU and UK law separately, and it is educational, not legal advice.

What Article 5(1)(d) says

Article 5(1)(d) GDPR requires personal data to be "accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy')" (Regulation (EU) 2016/679, Article 5).

Four connected provisions turn the principle into obligations:

Provision	Text in brief	Relevance to LLM output
Article 5(2), accountability	The controller is responsible for, and must be able to demonstrate compliance with, the principles	Accuracy must be evidenced, not asserted
Article 16, rectification	The data subject has the right to obtain rectification of inaccurate personal data without undue delay, and to have incomplete data completed, including by a supplementary statement	A person can demand that a false statement stop being produced

Provision	Text in brief	Relevance to LLM output
Article 18(1)(a), restriction	Processing can be restricted while accuracy is contested, for a period enabling the controller to verify it	Contested outputs may need to be suppressed during verification
Article 19, notification	Rectification, erasure or restriction must be communicated to each recipient, unless impossible or disproportionate	Corrections must propagate to where the data went
Article 25(1), data protection by design	Appropriate technical and organisational measures designed to implement the principles effectively	Accuracy belongs in system design, not only in policy

Sources for the table: Articles 5, 16, 18, 19 and 25. Infringements of the basic principles in Article 5 fall in the higher tier of administrative fines, up to EUR 20 million or 4% of total worldwide annual turnover, whichever is higher (Article 83(5)(a)).

Three phrases in Article 5(1)(d) matter most for language models.

- **"Having regard to the purposes."** Accuracy is judged relative to why the data is processed. A dataset of historical news articles is not inaccurate because a person has since changed jobs. An assistant that answers "who is the CFO of X" is processing for a purpose that depends on currency.
- **"Every reasonable step."** The standard is reasonableness, not perfection. That leaves room for statistical systems, provided the steps taken are proportionate and documented.
- **"Erased or rectified without delay."** The remedy assumes the controller can locate and change the inaccurate data. For a fact stored in a database that is trivial. For a statement a model generates from its parameters, it is the hard part.

Accuracy and purpose: when "up to date" matters

The purpose test cuts both ways. If an LLM is used to draft fiction, its outputs are not presented as facts about real people, and the accuracy question is framed differently. If it is used in customer service, recruitment, credit, compliance screening or search, its statements about people are offered as facts and will be acted on.

The EDPB Taskforce made exactly this point about ChatGPT. It observed that "the purpose of the data processing is to train ChatGPT and not necessarily to provide factually accurate information", that the training approach "leads to a model which may also produce biased or made up outputs", and that outputs "are likely to be taken as factually accurate by end users, including information relating to individuals, regardless of their actual accuracy". Its conclusion: "In any case, the principle of data accuracy must be complied with" (EDPB Taskforce report, para 30).

Input data and output data

The Taskforce distinguished two places where accuracy applies. Input data covers data collected, for instance, through web scraping, and the content users provide as prompts. Output data covers what the system returns (para 29). The distinction is useful for any LLM deployment:

- **Input accuracy** is a data-quality question about what was collected and retrieved. It can be managed with ordinary data governance: source selection, deduplication, correction.
- **Output accuracy** is about what the model asserts. A model can produce a false statement about a person even when no false statement about that person existed anywhere in its inputs, by combining true fragments wrongly.

The Taskforce also recorded that, at the time, OpenAI suggested users shift "from rectification to erasure when rectification is not feasible due to the technical complexity of ChatGPT" (para 34), and it reminded the controller of its Article 25(1) duty to build in safeguards (para 35). The report describes its contents as preliminary views of an ongoing coordination between supervisory authorities, not a decision.

The EDPB's later Opinion 28/2024 (adopted 18 December 2024) addressed AI models more generally. It stated that AI models trained with personal data "cannot, in all cases, be considered anonymous", and that the assessment depends on whether personal data, which it describes as possibly "accurate or inaccurate (because statistically inferred)", can be extracted from the model directly or by querying it (EDPB Opinion 28/2024). That matters for accuracy because it means personal data in a model's outputs, including wrong data, is not automatically outside the GDPR.

The noyb complaints: rectification requests a model cannot satisfy

Two complaints filed by the privacy organization noyb show how the principle meets practice.

Austria, April 2024. noyb filed a complaint with the Austrian data protection authority on behalf of a public figure whose date of birth ChatGPT repeatedly gave incorrectly "instead of telling users that it doesn't have the necessary data". According to noyb, OpenAI refused to rectify or erase the data, arguing that correction was not possible, said it could filter or block data on certain prompts only in a way that would filter all information about the complainant, and did not adequately answer the access request about the data, its sources and recipients. The complaint invoked Articles 5, 15 and 16 and asked the authority to investigate, order compliance with the access request and impose a fine (noyb, 29 April 2024).

Norway, March 2025. noyb filed a complaint with the Norwegian Datatilsynet after ChatGPT told a user that a Norwegian man had murdered two of his children and been sentenced to 21 years in prison. The fabricated story mixed in true details: the number and gender of his children and his home town. noyb reported that ChatGPT now searched the internet when asked about specific people, which stopped this output, but argued that the incorrect data "may still remain part of the LLM's dataset". It invoked Article 5(1)(d) and asked the authority to order deletion of the defamatory output, fine-tuning to eliminate inaccurate results, and a fine. noyb's position on disclaimers: "you can't bypass the legal obligation to ensure the accuracy of the personal data you process via a disclaimer" (noyb, 20 March 2025).

Neither complaint had a published decision when this article was checked on 28 October 2026: noyb's pages for both cases report no outcome, and no decision by the Austrian or Norwegian authority was found.

As of 28 October 2026 we found no published final decision in either case. They are complaints, which state one party's view; they show the questions authorities are being asked, not the answers.

The Norwegian case illustrates the most damaging pattern of hallucination about people: a fluent statement that is mostly true, with one false and serious element. A reader cannot tell which part is wrong, and the true details make the false one more credible.

The UK position: opinions, inferences and mistakes

The UK GDPR keeps the same accuracy principle. The Information Commissioner's Office (ICO) explains that the Data Protection Act 2018 defines "inaccurate" as "incorrect or misleading as to any matter of fact" (ICO, Principle (d): Accuracy). Its guidance adds distinctions that are useful for model output:

- **Opinions.** "A record of an opinion is not necessarily inaccurate personal data just because the individual disagrees with it", but the record must make clear that it is an opinion and, where appropriate, whose.
- **Mistakes.** "It is acceptable to keep records of mistakes, provided those records are not misleading about the facts."
- **Currency.** Data used for a purpose that relies on it remaining current should be kept up to date.
- **Challenges.** Controllers must carefully consider challenges to accuracy.

The ICO's AI guidance separates two meanings of the word. Data protection accuracy is the legal duty above. Statistical accuracy "refers to how often an AI system guesses the correct answer, measured against correctly labelled test data". The ICO states that the principle "does not mean that an AI system needs to be 100% statistically accurate", but that controllers must account for incorrect inferences and their impact, and that records should indicate outputs "are statistically informed guesses rather than facts", note the provenance of the data and the system used, and be updated if an inference rests on inaccurate data (ICO, accuracy and statistical accuracy). The ICO recommends measuring precision and recall rather than overall accuracy, and documenting retraining thresholds for concept drift.

Both ICO pages state that the guidance is under review following the Data (Use and Access) Act. UK readers should check current wording.

Point	EU (EDPB Taskforce, May 2024)	UK (ICO guidance)
Does the accuracy principle apply to model output about people?	Yes, "in any case"	Yes; inferences must be recorded as such
Are disclaimers enough?	Transparency measures help but are "not sufficient"	Records must show outputs are guesses, not facts; decisions must reflect that
Must a system be perfect?	Not stated as such	No: not "100% statistically accurate"
Measurement	Not specified	Precision, recall, error trade-offs, drift thresholds
Status	Preliminary views of a taskforce	Guidance under review

A worked example: one false statement, two architectures

The following case is fictional. **Halden Therapeutics AG is a fictional company**, and Dr Lena Brandt is a fictional person.

A compliance assistant at a fictional supplier is asked: "Who leads clinical operations at Halden Therapeutics AG, and has she been subject to any regulatory action?" It answers: "Dr Lena Brandt, Head of Clinical Operations since 2023. In 2021 she was named in an FDA warning letter for data integrity failures at a previous employer." The first sentence is true. The second is false: the 2021 warning letter named a different site, and Dr Brandt was not mentioned.

Dr Brandt submits a rectification request under Article 16. Compare how two architectures can respond.

Step	Model answers from its parameters	Answers drawn from a source-grounded record
Locate the data	The statement is generated, not stored; no row to correct	The assistant cited a record field "regulatory actions: person" with a source and excerpt
Verify (Art. 18(1)(a))	Reproduce the prompt and inspect outputs; results vary by sampling and version	Open the cited source; the excerpt names a site, not a person, so the link was wrong
Rectify (Art. 16)	Suppress outputs by filter, fine-tune, or wait for a retrain; none guarantees the statement will not recur	Remove the link, record why, mark the field "none found" with the sources searched
Notify recipients (Art. 19)	Unknown who received the statement unless outputs are logged	Answers that cited the field can be found and re-issued
Demonstrate (Art. 5(2))	Filter configuration and evaluation results	Before and after record states, with dates and reasons

The point is not that one architecture is lawful and the other unlawful. It is that rectification "without delay" presupposes that the controller can find the inaccurate data and change it. Where a statement about a person is drawn from an identifiable record and cited, that presupposition holds. Where it is produced from the model's parameters, the controller has to rely on filters and retraining and must be able to explain why those steps are reasonable.

Controls that help

None of the following is a statutory list. They are measures that map to the provisions above.

1. **Classify person-related use cases by purpose.** Identify where outputs about identifiable people are presented as fact (screening, due diligence, HR, customer service) and treat those as accuracy-critical.
2. **Ground factual statements about people in retrievable records.** Retrieve from a governed source and cite it at the level of the claim; see the site's note on testing LLM citations claim by claim.
3. **Prefer abstention to invention.** Where no source supports a statement about a person, the answer should say so. The Austrian complaint turned on a system giving a wrong date "instead of telling users that it doesn't have the necessary data". The site's method on declaring unknowns treats "not found" as an explicit value rather than a blank.

4. **Label inferences.** Where an output is an inference or estimate, say so in the record and the interface, as the ICO recommends.
5. **Log outputs about people.** Without a record of what was said to whom, Article 19 notification and Article 15 access requests cannot be answered.
6. **Build a correction path.** Define how a rectification request reaches the record, the retrieval index and any cached answers, and how long each takes.
7. **Measure.** Track error rates on person-related questions with precision and recall, and set thresholds for intervention.
8. **Document.** Keep the purpose analysis, the evaluation results, the correction log and the reasons for the measures chosen, to meet Article 5(2).

Documentation to keep

An accountability file for an LLM system that produces statements about people would typically contain: the purpose and the reasons accuracy matters for it; the sources the system is allowed to draw on; evaluation results on person-related questions, with dates and versions; the rectification and restriction procedure and its log; output logs or a justification for not keeping them; and the record of each challenge and how it was resolved. The hallucination glossary entry gives the terminology.

Limits

This article summarizes EU and UK texts and regulator statements as of 28 October 2026. The EDPB Taskforce report records preliminary views; the noyb documents are complaints; the ICO guidance is under review. National supervisory authorities decide cases individually. Nothing here is legal advice.

A record that keeps a cited, verbatim, dated source behind every value, and declares unknowns instead of leaving blanks, is one way to make statements about people correctable; that is the design Reliable Context describes in its published guarantees.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf
2. https://eur-lex.europa.eu/eli/reg/2016/679/oj
3. https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en
4. https://noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it
5. https://noyb.eu/en/ai-hallucinations-chatgpt-created-fake-child-murderer
6. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/accuracy
7. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/what-do-we-need-to-know-about-accuracy-and-statistical-accuracy

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.