# LEGAL OPINION

Lawfulness of Using Client Procurement Data to Fine-Tune an LLM for AI-Powered Spend Analytics Under the GDPR

**Date:** 26 March 2026

**Prepared by:** Kim De Bruyne (김덕배), GDPR-expert

**For:** [Client] — B2B SaaS Procurement Platform

**Classification:** Confidential

## I. EXECUTIVE SUMMARY

**Your company cannot rely on legitimate interest (Art. 6(1)(f) GDPR) to fine-tune an LLM using client procurement data in your current capacity as a data processor under Art. 28 GDPR.** Using processor data for your own AI training purposes would constitute a fundamental breach of the processor relationship, effectively making you a controller for that processing — with severe legal consequences including fines of up to EUR 10 million or 2% of annual worldwide turnover (Art. 83(4)(a) GDPR).

Before any AI model training can occur, a restructuring of the legal relationship with your clients is required. Even with restructured roles, the legitimate interest analysis under current law presents significant challenges that require careful mitigation. However, the proposed Digital Omnibus Package (COM(2025) 837) would, if adopted, explicitly recognise AI development as a legitimate interest — potentially easing (but not eliminating) this analysis.

**Risk assessment: HIGH — without restructuring, the proposed processing is unlawful.**

## II. FACTUAL BACKGROUND

Your company operates a B2B SaaS platform processing procurement data (invoices, purchase orders, vendor contracts) on behalf of enterprise clients. You act as a data processor under Art. 28 GDPR. The data includes business contact details of employees at client companies (names, email addresses, job titles, phone numbers).

Your CTO proposes using this data to fine-tune a Large Language Model (LLM) for an "AI-powered spend analytics" feature — predicting spending patterns and suggesting cost optimisations.

## III. ANALYSIS

### A. The Threshold Issue: Controller vs. Processor Status (Art. 28 GDPR)

#### 1. The Legal Framework

This is the determinative issue that must be addressed before any legitimate interest analysis can proceed.

**[VERIFIED]** **[Grade A]** **GDPR Article 28(3)(a)** provides that a processor:

> *"processes the personal data only on documented instructions from the controller, including with regard to transfers of personal data to a third country or an international organisation"*

**[VERIFIED]** **[Grade A]** **GDPR Article 29** reinforces this:

> *"The processor and any person acting under the authority of the controller or of the processor, who has access to personal data, shall not process those data except on instructions from the controller, unless required to do so by Union or Member State law."*

[VERIFIED] [Grade A] **GDPR Article 28(10)** states the consequence of breach:

> *"if a processor infringes this Regulation by determining the purposes and means of processing, the processor shall be considered to be a controller in respect of that processing."*

### 2. Recital Context

[VERIFIED] [Grade A] **Recital 81** provides the legislative intent behind Art. 28:

> *"The carrying-out of processing by a processor should be governed by a contract or other legal act under Union or Member State law, binding the processor to the controller, setting out the subject-matter and duration of the processing, the nature and purposes of the processing, the type of personal data and categories of data subjects, taking into account the specific tasks and responsibilities of the processor in the context of the processing to be carried out and the risk to the rights and freedoms of the data subject."*

This confirms that the processor's role is **bounded by the contract** with the controller — the processor processes data for the controller's purposes as defined in that contract, not for its own.

### 3. EDPB Guidance on Controller/Processor Distinction

[VERIFIED] [Grade A] **EDPB Guidelines 07/2020 on the concepts of controller and processor** provide the authoritative framework. The Guidelines establish that:

- The concept of controller is a **functional concept**: it allocates responsibilities according to actual roles, not formal designations.

- A **controller** determines the **purposes and means** — the "why" and "how" of processing.

- A **processor** processes personal data **on behalf of** the controller and must not process data beyond the controller's instructions.

- If a processor goes beyond instructions and starts to determine its own purposes and means, **the processor shall be considered a controller** in respect of that processing.

The distinction between "essential means" and "non-essential means" is critical:

- **Essential means** (determined by the controller): *type of personal data processed, duration, categories of recipients, categories of data subjects*

- **Non-essential means** (may be left to processor): *choice of hardware/software, detailed security measures*

### 4. Application to Your Scenario

Using client data to fine-tune your own LLM constitutes **determining a new purpose** for the processing. The original purpose (providing the procurement platform service) was defined by your clients as controllers. AI model training for your own product development is an entirely separate purpose that **you** are determining.

Under Art. 28(10) GDPR, **you would be deemed a controller** for the AI training processing. This is not merely a theoretical risk — it is the direct legal consequence prescribed by the GDPR.

**[VERIFIED] [Grade A]** The **Garante v OpenAI** enforcement decision (EUR 15 million, 20 December 2024) demonstrates that DPAs are actively enforcing against AI companies that process personal data for model training without proper legal basis, finding violations of Art. 5(1)(a), Art. 6, Art. 13, and Art. 25 GDPR.

### 5. CJEU Case Law on Controller Determination

**[VERIFIED] [Grade A]** In **C-683/21 — Nacionalinis visuomenės sveikatos centras** (7 December 2023), the CJEU confirmed that joint controllership arises from **participation in determining purposes and means**, even without a formal agreement.

**[VERIFIED] [Grade A]** In **C-40/17 — Fashion ID** (29 July 2019), the CJEU held that a party can be a joint controller **even without direct access to the personal data** — what matters is the deliberate decision that triggers data processing.

**[VERIFIED] [Grade A]** In **C-210/16 — Wirtschaftsakademie** (5 June 2018), the CJEU found joint controllership where a party had influence on processing and received derived benefit (statistical data), even with unequal involvement.

### 6. Consequences of Processor-to-Controller Reclassification

If you use client data for AI training without restructuring:

**(a) Breach of processing agreement** — Violation of Art. 28(3)(a) (processing beyond instructions), potentially triggering contractual liability and termination clauses.

**(b) Administrative fines** — Art. 83(4)(a): up to EUR 10 million or 2% of annual worldwide turnover for violations of processor obligations under Art. 28.

**(c) Full controller obligations** — As the deemed controller for AI training, you would bear all GDPR controller obligations (lawful basis, transparency, data subject rights, DPIA, etc.) — obligations you are unlikely to have prepared for.

**(d) No valid legal basis** — You collected the data as processor under your clients' legal basis (likely Art. 6(1)(b) — contractual necessity). Using it for your own purpose requires **your own, independent legal basis**, which you currently lack.

**(e) Civil liability** — Art. 82 GDPR: data subjects can claim compensation for material and non-material damage.

## B. Purpose Limitation (Art. 5(1)(b) GDPR)

### 1. The Principle

[VERIFIED] [Grade A] **GDPR Article 5(1)(b)** provides:

> *"[Personal data shall be] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes"*

[VERIFIED] [Grade A] **Recital 50** elaborates:

> *"The processing of personal data for purposes other than those for which the personal data were initially collected should be allowed only where the processing is compatible with the purposes for which the personal data were initially collected."*

### 2. Recital Context — Recital 39

[VERIFIED] [Grade A] **Recital 39** provides the core interpretive context for purpose limitation and data minimisation:

> *"In particular, the specific purposes for which personal data are processed should be explicit and legitimate and determined at the time of the collection of the personal data. The personal data should be adequate, relevant and limited to what is necessary for the purposes for which they are processed. [...] Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means."*

This last sentence — *"processed only if the purpose of the processing could not reasonably be fulfilled by other means"* — is particularly significant for AI model training: it creates a direct link between purpose limitation and the necessity assessment, requiring controllers to consider whether the same objective can be achieved without personal data (e.g., through anonymised or synthetic data).

### 3. Compatibility Assessment Under Art. 6(4) GDPR

[VERIFIED] [Grade A] **Article 6(4) GDPR** provides the compatibility test. Where processing is for a purpose other than the original collection purpose, the controller must consider:

> *"(a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing;* *(b) the context in which the personal data

*have been collected, in particular regarding the relationship between data subjects and the controller;* *(c) the nature of the personal data, in particular whether special categories of personal data are processed;* *(d) the possible consequences of the intended further processing for data subjects;* *(e) the existence of appropriate safeguards, which may include encryption or pseudonymisation."*

### 4. Application to AI Model Training

Applying these criteria to your scenario:

**(a) Link between purposes:** WEAK. The original purpose is providing a procurement SaaS platform (operational). Fine-tuning an LLM is a product development activity. While the subject matter (procurement) overlaps, the nature of processing is fundamentally different — operational service delivery vs. statistical model training.

**(b) Context of collection:** UNFAVOURABLE. Data was collected in a B2B context where data subjects (employees at client companies) reasonably expect their data to be used for the procurement platform they interact with — not to train AI models. Per Recital 47, *"the interests and fundamental rights of the data subject could in particular override the interest of the data controller where personal data are processed in circumstances where data subjects do not reasonably expect further processing."*

**(c) Nature of data:** MODERATE RISK. Business contact details (names, emails, job titles) are not special category data under Art. 9, but procurement data may reveal commercially sensitive information about purchasing patterns, vendor relationships, and spending behaviour.

**(d) Consequences:** MODERATE. Data may become embedded in model weights, creating risks of data extraction or regurgitation. Per EDPB Opinion 28/2024, personal data may remain "absorbed" in model parameters.

**(e) Safeguards:** FAVOURABLE (if implemented). Pseudonymisation, differential privacy, data minimisation techniques, and access controls could mitigate risks.

**Conclusion on purpose limitation:** AI model training is **not obviously compatible** with the original purpose of providing a procurement platform. It is a distinct purpose requiring its own legal basis.

[VERIFIED] [Grade A] **EDPB Opinion 28/2024** (para. 64) confirms: *"When assessing whether the purpose pursued is legitimate, specific and explicit, and whether the processing complies with the data minimisation principle, one should first identify the processing activity at stake."* The development and deployment phases are treated as potentially separate processing activities requiring independent assessment.

### 5. The Scientific Research Exception

Art. 5(1)(b) provides a presumption of compatibility for *"scientific or historical research purposes or statistical purposes"* in accordance with Art. 89(1). However:

- Commercial AI model fine-tuning for a product feature does not qualify as "scientific research" under current law.

- **Note:** The Digital Omnibus Package (COM(2025) 837) proposes to define "scientific research" as including research that *"can also support innovation, such as technological development and demonstration"* and *"may also aim to further a commercial interest."* If adopted, this broader definition could potentially bring some AI development activities within the research exemption, subject to Art. 89(1) safeguards.

## C. Legitimate Interest Balancing Test (Art. 6(1)(f) GDPR)

**Preliminary note:** This analysis assumes the controller/processor issue has been properly restructured (see Section III.A above) such that your company acts as controller (or joint controller) for the AI training purpose.

### 1. The Legal Framework

[VERIFIED] [Grade A] **GDPR Article 6(1)(f)** provides:

> *"processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child."*

[VERIFIED] [Grade A] **EDPB Guidelines 1/2024 on legitimate interest** (8 October 2024) set out the three-step test:

1. The pursuit of a **legitimate interest** by the controller or third party;

2. The **necessity** of processing for the purposes of that legitimate interest; and

3. The interests or fundamental rights and freedoms of data subjects **do not take precedence**.

[VERIFIED] [Grade A] **EDPB Opinion 28/2024** (17 December 2024) applies this framework specifically to AI model development and deployment.

### 2. Step 1: Identifying the Legitimate Interest

The interest must be: (a) **lawful**, (b) **clearly and precisely articulated**, and (c) **real and present** (not speculative).

[VERIFIED] [Grade A] EDPB Opinion 28/2024 (para. 69) provides examples of potentially legitimate interests in AI contexts:

> *(i) developing the service of a conversational agent to assist users; (ii) developing an AI system to detect fraudulent content or behaviour; and (iii) improving threat detection in an information system.*

**Your interest — "developing AI-powered spend analytics to predict spending patterns and suggest cost optimisations for clients"** — is:

- **Lawful:** Product improvement and innovation are recognised commercial interests. No law prohibits AI-based spend analytics.

- **Clearly articulated:** Yes — predicting spending patterns and suggesting cost optimisations for enterprise clients.

- **Real and present:** Yes — assuming the CTO has concrete development plans, not speculative future use.

[VERIFIED] [Grade A] The CJEU in **C-621/22 — KNLTB** (4 October 2024) confirmed: *"A legitimate interest does not have to follow from a legal norm but it must be lawful."* Commercial interests, including direct marketing, are recognised as legitimate.

**Assessment: The interest likely qualifies as legitimate.** However, it must be articulated with precision — "improving our AI model generally" would be too vague; "developing spend analytics for our existing client base" is sufficiently specific.

### 3. Step 2: Necessity of the Processing

[VERIFIED] [Grade A] EDPB Opinion 28/2024 (paras. 70-75) requires:

- Whether the processing **will allow pursuit** of the legitimate interest; AND

- Whether there is **no less intrusive way** of achieving the purpose.

**Critical questions for your scenario:**

**(a) Is personal data necessary?** The spend analytics feature could potentially be developed using:

- Fully anonymised data

- Synthetic data

- Aggregated statistical data without personal identifiers

If business contact details (names, emails, job titles) can be stripped before training without materially degrading model performance, processing personal data is **not necessary** and cannot satisfy this step.

[VERIFIED] [Grade A] EDPB Opinion 28/2024 (para. 73): *"If the pursuit of the purpose is also possible through an AI model that does not entail processing of personal data, then processing personal data should be considered as not necessary."*

**(b) Data minimisation:** Even if some personal data is needed, is the full dataset necessary? Can you train on a representative subset? Can contact details be pseudonymised while retaining transactional patterns?

**(c) Proportionality of data volume:** SAs will examine *"whether [the amount of personal data] is proportionate to pursue the legitimate interest at stake, also in light of the data minimisation principle"* (Opinion 28/2024, para. 73).

**Assessment: Necessity is the weakest link.** For spend analytics (predicting spending patterns, suggesting cost optimisations), the core value lies in transactional data (amounts, categories, timing, vendor types) — not in personal identifiers. It is likely that the model can be trained effectively on pseudonymised or anonymised procurement data. If so, legitimate interest under Art. 6(1)(f) fails at this step.

**Recommendation:** Conduct a technical assessment to determine whether personal data is genuinely necessary for model performance, or whether anonymised/pseudonymised data achieves comparable results.

### 4. Step 3: Balancing Test

Assuming necessity is established, the balancing test considers:

#### *(a) Data Subjects' Interests, Rights and Freedoms*

[VERIFIED] [Grade A] EDPB Opinion 28/2024 (paras. 77-80) identifies relevant data subject interests including:

- **Self-determination and control** over personal data

- **Financial interests** where AI models affect professional activities

- **Fundamental rights** under Art. 7 (private life) and Art. 8 (data protection) EU Charter

- **Risk of extraction or regurgitation** of personal data from the trained model

The data subjects here are **employees at your client companies** — individuals in a professional (B2B) context, not consumers or vulnerable individuals. This moderately favours the controller's interest.

#### *(b) Impact Assessment*

[VERIFIED] [Grade A] Per EDPB Opinion 28/2024 (paras. 82-86), SAs should consider:

**Nature of data:** Business contact details and procurement data. Not special category data, but commercially sensitive.

**Context:** B2B professional context. Data subjects are acting in their professional capacity.

**Consequences:** Risk of data being "absorbed" in model parameters (Opinion 28/2024, para. 31). Potential for membership inference attacks, data extraction, or regurgitation of training data. Commercially sensitive procurement patterns could be inadvertently revealed.

### *(c) Reasonable Expectations*

[VERIFIED] [Grade A] **Recital 47** GDPR:

> *"At any rate the existence of a legitimate interest would need careful assessment including whether a data subject can reasonably expect at the time and in the context of the collection of the personal data that processing for that purpose may take place."*

[VERIFIED] [Grade A] EDPB Opinion 28/2024 (paras. 87-92) emphasises reasonable expectations:

- Whether data was **publicly available** — No, this is private business data provided under contract.

- The **nature of the relationship** — Indirect; data subjects are employees of your clients, not your direct customers.

- Whether data subjects are **actually aware** their data could be used for AI training — Almost certainly not.

**Assessment: Data subjects would NOT reasonably expect their procurement data to be used for AI model training.** This significantly weighs against relying on legitimate interest.

### *(d) Mitigating Measures*

[VERIFIED] [Grade A] EDPB Opinion 28/2024 (paras. 95-105) provides a non-exhaustive list of mitigating measures:

**Development phase:**

- Selection of sources: exclude inappropriate data sources

- Data preparation: pseudonymisation, data minimisation, filtering of personal data before training

- Methodological choices: regularisation, differential privacy

- Output controls: measures to prevent regurgitation

**For your scenario, recommended mitigating measures include:**

1. **Pseudonymisation** of all personal identifiers before training

2. **Differential privacy** techniques during model training

3. **Data minimisation** — strip business contact details, retain only transactional patterns

4. **Access controls** — limit model access, no public deployment of the base model

5. **Testing** against membership inference and extraction attacks

6. **Right to object** — implement Art. 21 mechanism allowing data subjects to opt out

7. **Transparency** — update privacy notices to inform data subjects about AI training

8. **DPIA** — mandatory under Art. 35 for high-risk AI processing

**Overall Balancing Assessment:**

| Factor | Weight | Direction |
|---|---|---|
| Commercial interest in AI innovation | Moderate | Favours controller |
| B2B professional context (not consumers) | Moderate | Favours controller |
| Benefit to clients (cost optimisation) | Moderate | Favours controller |
| Lack of reasonable expectations | Strong | Favours data subject |
| Indirect relationship with data subjects | Moderate | Favours data subject |
| Risk of data extraction from model | Moderate | Favours data subject |
| Commercially sensitive data | Moderate | Favours data subject |
| Mitigating measures available | Moderate | Favours controller (if implemented) |

**Conclusion on balancing:** Without robust mitigating measures, the balancing test **tips against** the controller. With comprehensive mitigation (pseudonymisation, differential privacy, transparency, opt-out), the balance may shift — but remains a close call requiring thorough documentation and a DPIA.

## D. EDPB's Position on AI and Personal Data

### 1. EDPB Opinion 28/2024 (17 December 2024)

**[VERIFIED]** **[Grade A]** This is the **definitive EDPB position** on AI model training with personal data. Key holdings:

**(a) AI models trained on personal data are not automatically anonymous:**

> *The EDPB considers that *"AI models trained on personal data cannot, in all cases, be considered anonymous. Instead, the determination of whether an AI model is anonymous should be assessed, based on specific criteria, on a case-by-case basis."* (para. 34)*

**(b) Personal data may remain "absorbed" in model parameters:**

> *"even when an AI model has not been intentionally designed to produce information relating to an identified or identifiable natural person from the training data, information from the training dataset, including personal data, may still remain 'absorbed' in the parameters of the model"* (para. 31)

**(c) Legitimate interest is available but requires rigorous assessment:**

The three-step test applies with particular attention to:

- Volume and proportionality of personal data processed

- Whether anonymised or synthetic alternatives exist

- Reasonable expectations of data subjects

- Privacy-preserving techniques (differential privacy, pseudonymisation)

- Right to object under Art. 21

**(d) Consequences of unlawful processing — Three scenarios:**

- **Scenario 1** (same controller, data retained): Lawfulness of subsequent processing assessed case-by-case.

- **Scenario 2** (different controller deploys): Deploying controller must conduct due diligence on whether development was lawful.

- **Scenario 3** (model anonymised after unlawful training): If model is genuinely anonymous, GDPR does not apply to subsequent operation — unlawfulness of initial processing does not impact subsequent operation.

### 2. EDPB ChatGPT Taskforce Report (23 May 2024)

**[VERIFIED]** **[Grade A]** The ChatGPT Taskforce addressed web scraping for AI training and confirmed:

- Each processing activity (collection, training, deployment) may require its own legal basis assessment.

- The fairness principle requires that processing is not *"unjustifiably detrimental, unlawfully discriminatory, unexpected or misleading to the data subject."*

- Publicly available data is not exempt from GDPR — the mere fact data is online does not mean it has been "manifestly made public" under Art. 9(2)(e).

### 3. Garante v OpenAI Enforcement (EUR 15M, December 2024)

**[VERIFIED]** **[Grade B]** The Italian DPA fined OpenAI EUR 15 million for:

- Training ChatGPT on personal data **without a valid legal basis** (Art. 6)

- **Lack of transparency** about data use for AI training (Art. 13)

- **No age verification** for users under 13 (Art. 8)

- Failure to notify a data breach (Art. 33)

**Relevance to your case:** This enforcement action demonstrates that DPAs are actively scrutinising AI training practices. Your scenario is distinguishable (B2B data, not web-scraped), but the principle that AI training requires a clear legal basis is firmly established.

## E. Impact of the Digital Omnibus Package (COM(2025) 837)

### 1. Overview

**[VERIFIED]** **[Grade B]** The **Digital Legislation Omnibus** (COM(2025) 837, published 19 November 2025) proposes significant GDPR amendments relevant to AI development. As a legislative proposal, it is **not yet law** and may undergo substantial changes during the legislative process.

### 2. Key Proposed Changes Affecting This Analysis

#### *(a) AI Development as Explicit Legitimate Interest*

**[VERIFIED]** **[Grade B]** **Proposed Recital 30:**

> *"The development and use of AI systems and the underlying models such as large language models and generative video models rely on data, including personal data, in various phases in the AI lifecycle, such as the training, testing and validation phase... The processing of personal data in this context may therefore be carried out for purposes of a legitimate interest within the meaning of Article 6 of Regulation (EU) 2016/679, where appropriate."*

This does **not** create an automatic legal basis — it merely **confirms** that AI development can constitute a legitimate interest. The full three-step test still applies:

> *"This does not affect its obligation to ensure that all other conditions of Article 6(1)(f) of Regulation (EU) 2016/679 as well as all other requirements and principles of that Regulation are met."*

#### *(b) Balancing Factors for AI Processing*

**[VERIFIED]** **[Grade B]** **Proposed Recital 31** provides guidance on the balancing test:

- Whether the interest is **beneficial for the data subject and society** (e.g., bias detection, accessibility)

- **Reasonable expectations** of data subjects

- **Safeguards** including enhanced transparency, unconditional right to object, respecting technical indications limiting AI use, privacy-preserving techniques, measures against regurgitation and data leakage

### *(c) Purpose Limitation Amendment*

**[VERIFIED] [Grade B] Proposed amendment to Art. 5(1)(b):**

> *"further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), be considered to be compatible with the initial purposes, independent of the conditions of Article 6(4) of this Regulation"*

This clarifies that the research exemption is **independent** of the Art. 6(4) compatibility assessment — but the exemption itself is not expanded to commercial AI training.

### *(d) Broader Definition of Scientific Research*

**[VERIFIED] [Grade B] Proposed new Art. 4(38):**

> *"'scientific research' means any research which can also support innovation, such as technological development and demonstration. These actions shall contribute to existing scientific knowledge or apply existing knowledge in novel ways... This does not exclude that the research may also aim to further a commercial interest."*

If adopted, this broader definition could potentially bring AI model development within the "scientific research" purpose — but only where it genuinely contributes to scientific knowledge, not mere commercial product improvement.

### *(e) Special Categories of Data Exception for AI*

**[VERIFIED] [Grade B] Proposed Art. 9(2)(k):**

> *"processing in the context of the development and operation of an AI system... or an AI model, subject to the conditions referred to in paragraph 5."*

New Art. 9(5) requires:

- Technical and organisational measures to avoid processing special categories

- Removal of identified special category data where practicable

- If removal requires disproportionate effort: effectively protect data from being used in outputs or disclosed

### *(f) Personal Data Definition — Relative Approach*

The proposed amendments adopt a **relative/subjective approach** to pseudonymised data: information is not personal data when the entity lacks *"means reasonably likely"* to identify the natural person. This could benefit AI developers who process properly pseudonymised data, as the pseudonymised data may not constitute personal data in their hands.

### 3. Impact Assessment: Current Law vs. Proposed Omnibus

| Issue | Current GDPR | If COM(2025) 837 Adopted |
|---|---|---|
| AI training as legitimate interest | Not explicitly listed; arguable | Explicitly confirmed in recital |
| Balancing test requirements | Full three-step test | Same, but with AI-specific guidance |
| Purpose limitation for research | Narrow "scientific research" | Broader definition, can include commercial aims |
| Special category data in AI | Very limited exceptions | New Art. 9(2)(k) exception for residual processing |
| Pseudonymised data | Personal data in all hands | Relative approach — may be non-personal for recipient |

### 4. Caveat

The Digital Omnibus is at **proposal stage**. The European Parliament and Council have not yet begun trilogue negotiations. The final text may differ substantially. Legislative proposals are classified as **Grade B** sources — they cannot be relied upon for current compliance decisions.

## IV. RECOMMENDATIONS

### Immediate Actions (Current Law)

1. **Do NOT proceed with AI model training in your current processor capacity.** This would constitute an unlawful determination of purposes under Art. 28(10).

2. **Restructure the legal relationship** with your clients:

- **Option A: Seek explicit contractual authorisation** from each client to use their data for AI model training (amending DPAs to reflect a joint controller or independent controller relationship for this purpose).

- **Option B: Obtain data subject consent** via your clients — technically complex and operationally burdensome in a B2B context.

- **Option C: Anonymise data first** — if procurement data can be fully anonymised before training (removing all personal identifiers, aggregating data), GDPR does not apply.

3. **Conduct a Data Minimisation Assessment:** Determine whether personal data is genuinely necessary for model training, or whether anonymised/pseudonymised transaction data suffices.

4. **If personal data is necessary, conduct a full DPIA** (Art. 35 GDPR) before any processing begins.

5. **Implement robust mitigating measures:**

- Pseudonymise all personal identifiers before training

- Apply differential privacy techniques

- Test the model against extraction and membership inference attacks

- Implement Art. 21 right to object mechanism

- Update privacy notices

6. **Document the legitimate interest assessment** thoroughly — this is required under the accountability principle (Art. 5(2)).

### Medium-Term Actions (Preparing for Digital Omnibus)

7. **Monitor legislative developments** on COM(2025) 837. If adopted as proposed, it will ease — but not eliminate — the legal basis analysis.

8. **Consider anonymisation-first architecture:** Design AI training pipelines that anonymise data by default, using personal data only where demonstrably necessary. This approach is robust under both current and proposed law.

9. **Engage with your DPO** and consider seeking guidance from your lead supervisory authority on the proposed AI training activities.

## V. CROSS-REFERENCE LIST

| Provision | Relevance |
|---|---|
| GDPR Art. 4(1) | Definition of personal data |
| GDPR Art. 4(7) | Definition of controller |
| GDPR Art. 5(1)(a) | Lawfulness, fairness, transparency |
| GDPR Art. 5(1)(b) | Purpose limitation |
| GDPR Art. 5(1)(c) | Data minimisation |
| GDPR Art. 5(2) | Accountability |
| GDPR Art. 6(1)(f) | Legitimate interest |
| GDPR Art. 6(4) | Compatible purpose test |
| GDPR Art. 9 | Special categories of data |
| GDPR Art. 21 | Right to object |
| GDPR Art. 22 | Automated decision-making |
| GDPR Art. 25 | Data protection by design |
| GDPR Art. 28 | Processor obligations |
| GDPR Art. 29 | Processing under authority |
| GDPR Art. 35 | DPIA |
| GDPR Art. 82 | Right to compensation |
| GDPR Art. 83(4)(a) | Administrative fines for processor violations |
| GDPR Art. 89 | Safeguards for research/statistics |
| GDPR Recital 47 | Legitimate interest / reasonable expectations |
| GDPR Recital 50 | Compatible purpose |
| GDPR Recital 81 | Processor obligations |

| | |
|---|---|
| EU AI Act Art. 10 | Data governance for high-risk AI |
| COM(2025) 837 | Digital Omnibus GDPR amendments (proposed) |

## VI. KEY SOURCES CITED

### EDPB Documents (Grade A)

- EDPB Opinion 28/2024 on AI models and personal data (17 December 2024)

- EDPB Guidelines 1/2024 on legitimate interest under Art. 6(1)(f) (8 October 2024)

- EDPB Guidelines 07/2020 on controller and processor concepts

- EDPB ChatGPT Taskforce Report (23 May 2024)

### CJEU Case Law (Grade A)

- C-621/22 — KNLTB (legitimate interest, commercial purposes)

- C-252/21 — Meta v Bundeskartellamt (legitimate interest, tracking)

- C-683/21 — Nacionalinis visuomenės sveikatos centras (joint controllership, processor liability)

- C-40/17 — Fashion ID (joint controllership without data access)

- C-210/16 — Wirtschaftsakademie (joint controllership, Facebook fan pages)

- C-634/21 — SCHUFA Scoring (automated credit scoring under Art. 22)

- C-413/23 — EDPS v SRB (relative approach to personal data/pseudonymisation)

- C-582/14 — Breyer (dynamic IP addresses, legitimate interest)

### Enforcement Decisions (Grade B)

- Garante v OpenAI — EUR 15M (20 December 2024)

### Legislative Proposals (Grade B)

- COM(2025) 837 — Digital Legislation Omnibus

## VII. DISCLAIMER

This response is for informational purposes only and does not constitute legal advice. The analysis is based on the GDPR as currently in force and publicly available EDPB guidance as of March 2026. The Digital Omnibus Package (COM(2025) 837) is a legislative proposal that may change substantially before adoption. For specific matters, implementation decisions, or individual DPA engagement, please consult qualified legal counsel in the relevant jurisdiction.

*Prepared by Kim De Bruyne (김덕배)*

*GDPR Expert — EU Data Protection Practice*

*26 March 2026*