EU AI SOVEREIGNTY

Sovereignty is four things. Not one.

Sovereignty is not a flag on the box; it is the ability to keep operating when a provider, a price, or a government changes its mind. An honest assessment of where European AI stands in 2026, what data residency and "sovereign cloud" actually buy, and how to build a stack that is portable across providers without becoming hostage to any of them: the four dimensions, the five cloud gradients costed, and the swap architecture. Portability over chauvinism, compliance over slogans.

FOR

Public-sector CDOs, EU banks, ministries, defence-adjacent

STANCE

Portability over chauvinism. Compliance over slogans.

COMPANION

Briefing 04 (Risk & Compliance), CISO & vendor briefing

DIRECT LINE

charafeddine@cohorte.co

02 THE HONESTY CONTRACT

Sovereignty is a real engineering problem. It is also a political slogan.

This briefing separates the two. The political slogan is what gets pitched in conference talks; the engineering problem is what the architect has to design. A document that conflates them sells one to people who came for the other, which is how Europe ended up with a decade of sovereign-cloud headlines and very few sovereign systems in production.

Three positions on this page that the rest of the briefing defends.

European LLMs in 2026 are not at parity with US frontier models on capability. The gap is narrowing, the trajectory is real, and several European labs (Mistral, Aleph Alpha, DeepL on translation, the Dutch and Spanish national projects) are producing serious work. A buyer in 2026 who needs frontier capability on a hard task and chooses a European model on capability terms alone is making a sub-optimal decision. A buyer who chooses one on portability, compliance, and political-risk terms is making a defensible decision.

Sovereign cloud is a real category and it is also marketed at much greater volume than it is purchased. Microsoft, AWS and Google operate EU-region deployments with European personnel and European law-of-the-land guarantees. OVHcloud, Scaleway, Aruba and others operate genuinely European infrastructure. The decision is rarely between sovereign and non-sovereign; it is between gradients of operational sovereignty, each with a cost in either capability or convenience.

Portability beats sovereignty as the practical engineering goal. An architecture that can swap from US-frontier to European-mid-tier without rewriting half the application is one that has the option to move when politics, pricing, or regulation makes it appropriate. An architecture that is hard-wired to a single provider, sovereign or not, is one that has surrendered the option. The hard work is portability; the sovereignty question collapses into "where is the system today" once portability is in place.

> An architecture that can move is an architecture that did not have to.

03 THE FOUR DIMENSIONS

Decompose "sovereignty" before deciding which one matters.

When a public-sector buyer asks for sovereignty, they usually mean one or two of these four dimensions. When a European software vendor markets sovereignty, they often mean a third. When a regulator references it, they often mean a fourth. The decomposition below makes the conversation tractable.

DIMENSION 01

Data residency

The physical location where data is stored and processed. The question with the clearest legal answer (GDPR, sectoral rules, AI Act Article 10). Often what "sovereignty" actually means to the legal team.

DIMENSION 02

Model provenance

Who trained the model, on which data, under which jurisdiction. Matters for political-risk exposure and for AI Act foundation-model obligations. The dimension with the least clean answer for most buyers in 2026.

DIMENSION 03

Infrastructure control

Whose servers, whose personnel, whose chain of authority. The "sovereign cloud" question. Gradients exist: hyperscaler EU-region; hyperscaler EU-region with European personnel; European hyperscaler; European cloud at scale; on-premises.

DIMENSION 04

Regulatory alignment

Whose courts adjudicate, whose laws apply, whose enforcement is available. CLOUD Act exposure for US-controlled entities; EU-US Data Privacy Framework adequacy; sectoral regulators' reach. The most legally consequential dimension.

Most "sovereignty" failures are decomposition failures. A team that pursued infrastructure-control sovereignty (dimension 03) when the actual problem was model-provenance exposure (dimension 02) spends a year migrating to a European cloud and remains exposed on the question that mattered. Naming the dimension is the cheapest part of the engagement; it is also the one most often skipped.

04 DATA RESIDENCY

01 What "European data" actually buys.

And what it does not.

What it buys. Clean GDPR compliance for personal-data flows. Reduction of CLOUD Act exposure (qualified, depending on the controller's corporate structure). Defensibility for the public-sector buyer whose procurement rules require EU storage and processing. Alignment with sectoral regimes (eIDAS, NIS2 for critical entities, sectoral health and finance rules). A clearer story to the press if a question is asked.

What it does not buy. Immunity from extraterritorial legal process when the operator's parent is US-incorporated. Independence from the foundation-model training distribution. Protection from supply-chain risk in the underlying compute (GPUs sourced from non-European suppliers). A different reliability level on the AI system itself. Most of the technical risk of an AI system is unaffected by where the data sits.

The practical design. Data residency is achieved at the storage layer (object stores, vector databases, log archives) and the processing layer (inference endpoints). The architecture choice that matters is whether the model invocation traverses non-EU infrastructure at any point in the chain. For most buyers, a hyperscaler EU-region deployment with the appropriate processor and sub-processor configuration is sufficient and faster to ship than a full European-only stack.

THE DATA-RESIDENCY CHECKLIST

Storage. Object store, vector DB, log archive all in EU regions, with no replicas outside.

Processing. Inference endpoint EU-region; pre-processing pipeline EU-region; embeddings computed and stored EU-region.

Logs and traces. Telemetry endpoints EU-region. Where the vendor's default is US, configure the EU endpoint or accept the gap in writing.

Sub-processors. The DPA lists each. Each one's region is named. Substitutions notified.

Vendor controls. Access by vendor personnel to your data requires either EU residency for the personnel, audited break-glass procedures, or explicit consent. Most enterprise contracts can support either of the first two.

05 MODEL PROVENANCE

02 European LLMs in 2026.

The honest assessment.

The European LLM landscape in mid-2026 has three credible players for general-purpose work (Mistral with its Large family, the EU Open-EuroLLM consortium, and a third entrant whose name shifts quarter to quarter), a few strong domain-specific labs (DeepL on translation, several legal-tech and code labs), and an interesting set of openly-licensed weights that European integrators can fine-tune. A buyer who needs frontier capability will not, in 2026, choose a European model on capability terms.

This is not a permanent state. The trajectory is real. The trajectory is also not the buying decision. The buying decision is what gets deployed in the next six months, against a published bar, in a regulated environment, with a published procurement timeline. For that decision, the relevant question is not "is there a European model at parity" but "what does the gap cost and is the cost worth the political-risk reduction it buys."

For most enterprise workflows, the gap is not where it is reported to be. The frontier-vs-mid-tier difference dominates the conversation in press coverage; in practice the gap on workflow-specific tasks closes once a calibration set and the reliability-level method (Briefing 02) are applied. A mid-tier European model on a workflow-specific task often earns a defensible reliability level. The bar is lower than the frontier-on-benchmark headline would suggest, and lower than a 2024-era assessment would have predicted.

What this implies for the architecture: design for portability, install with the model that earns the bar at the time of deployment, plan to swap as the European landscape evolves. The system's verification layer (Briefing 02) does not care which model is underneath; it cares about the reliability level the model earns on the workflow's calibration set. The team that builds for portability discovers that the sovereignty question, in the technical sense, has become a tractable swap rather than a binding constraint.

> The capability gap is real and shrinking. The capability gap on the workflow's calibration set is often smaller than the press coverage suggests.

06 INFRASTRUCTURE CONTROL

03 Sovereign cloud, in five gradients.

Each with a cost.

"Sovereign cloud" is not one thing. The market has five distinguishable gradients. Each has a real engineering offer; each has a real cost in capability, convenience, or expense. A buyer who treats them as interchangeable will pay the cost of the most restrictive without buying the benefit it implies.

GRADIENT	WHAT IT OFFERS	WHAT IT COSTS
G1: Hyperscaler EU-region	EU data residency, standard contractual clauses, EU regulatory compliance.	CLOUD Act exposure (US parent). Most expedient. Lowest cost in capability.
G2: EU-region + EU personnel	Above, plus operational control by EU-resident personnel. Audited break-glass for US support.	Limited offer; specific configurations only. Slightly higher cost; significantly reduced CLOUD Act risk in practice.
G3: Hyperscaler-EU JV	Joint venture with European partner; sovereign data trustee structure; specific qualifications (e.g. France SecNumCloud).	Premium pricing. Slower feature rollout. Strong story for public sector and regulated banks.
G4: European hyperscaler	OVHcloud, Scaleway, Aruba, etc. EU-incorporated, EU-operated, EU-regulated.	Smaller AI service catalogue. Mid-tier model availability via partnership rather than first-party. Often the right answer for buyers prioritising regulatory alignment.
G5: On-premises	Full infrastructure control. Compatible with classified / restricted workloads.	Highest cost in convenience and absolute spend. Requires in-house operations capability for the entire stack. Defensible only for the workloads that need it.

The right gradient is not the highest one. Most enterprise workloads, even regulated ones, are satisfied by G1 or G2. Public-sector and defence-adjacent workloads often require G3 or G4. Classified work requires G5. A buyer who chooses G5 for a non-classified workload pays for sovereignty they did not need and forgoes the capability they did.

07 REGULATORY ALIGNMENT

04 Whose law applies.

Whose courts decide.

The regulatory dimension is the one that survives changes in technology. Frontier capability shifts; EU-US data adequacy decisions shift; cloud-region availability shifts. The legal-enforcement perimeter is durable. A system designed against the AI Act, GDPR, eIDAS, NIS2, DORA where applicable, and the sectoral rules of the deploying entity is a system that survives several news cycles of geopolitical noise.

CLOUD Act exposure. The US Clarifying Lawful Overseas Use of Data Act grants US courts the ability to compel disclosure from US-incorporated entities regardless of where the data is stored. The practical exposure depends on the contracting entity's corporate structure, the sub-processors' structures, and the specific dataset. For most enterprise data, the exposure is manageable through contractual structures and the EU-US Data Privacy Framework adequacy decision (where it remains in force). For the most sensitive data (national security, classified, certain healthcare records), the exposure is material and is the reason gradients G3-G5 of the cloud spectrum exist.

EU-US Data Privacy Framework. The successor to Privacy Shield. Adequacy decision in force; subject to ongoing Court of Justice review. A prudent architecture does not depend on the framework remaining in force for the lifetime of the system. The portability story (page 08) is also the contingency plan for an adequacy reversal.

Sectoral regulators. Finance (ECB, EBA, national regulators under DORA). Health (sectoral data rules per Member State, EHDS where applicable). Critical entities (NIS2). Defence-adjacent (national rules; defence procurement codes). Each sectoral regulator has supplementary requirements that the AI Act does not pre-empt. The mapping in Briefing 04 covers the major ones.

> Regulatory alignment is the dimension that does not shift overnight. Build to it; the others adjust to fit.

08 THE PORTABILITY ARCHITECTURE

Build for swap. Not for chauvinism.

A system that can swap its underlying model, its retrieval index, and its hosting region without rewriting the application is a system that has the option to move. The option is the asset. The decision to exercise it is downstream of business, legal, and political conditions that the engineering team does not control. The engineering team's job is to make the option exist.

YOUR APPLICATION	INTERFACE · THE REGISTRY	RE-VERIFICATION GATE
Calls a **capability**, not a vendor: chat, embedding, agent. It never names a model.	Maps the capability to a backing model. US frontier today, European mid-tier tomorrow: a registry edit and a re-verification, not a refactor.	Every swap re-runs the bar on your calibration set. It holds, it ships. It fails, the gate holds or rolls back.

The model layer behind an interface. The application calls a model interface, not a vendor. The interface declares a capability profile (chat, embedding, vision, agent), not a model. The mapping from interface to backing model lives in the registry. Swapping the backing model from a US frontier to a European mid-tier is a registry edit and a re-verification, not a refactor.

The retrieval layer behind a contract. The retrieval system implements a contract (search this corpus, return chunks with provenance, respect these permissions). The implementation can be Pinecone, Weaviate, pgvector, or an EU-region equivalent. The application does not care; the contract holds. The contract is enforceable at the gate; the registry tracks which implementation is currently behind the contract.

The verification layer everywhere. Every swap re-runs the verification on the workflow's calibration set. The reliability level either holds or it does not. The deployment decision after the swap is data-driven: if the new model meets the bar, the system continues; if not, the gate holds and the operator either renegotiates the bar or stays on the previous backing model. The decision is not made on capability anecdotes from the press.

The contract terms that preserve the option. Procurement clauses on data portability, no lock-in for proprietary file formats, exit-assistance commitments, and source-code escrow for critical custom components. None of these are AI-specific; they are decades-old enterprise procurement discipline that the AI gold rush sometimes encourages teams to skip.

THE PORTABILITY TEST

"Can we run the same workflow with a different backing model in two weeks, with the verification re-run on our calibration set?"

If **yes**, the architecture is portable. The sovereignty question is a swap decision.

If **no**, the architecture is locked. The sovereignty question is a rebuild.

THE REFERENCE IMPLEMENTATION

Context Kubernetes (github.com/Cohorte-ai/context-kubernetes) implements the manifest pattern that makes the swap a configuration change rather than a code change. The architecture is documented in the source paper (Mouzouni, 2026).

WHERE THIS FAILS

Workflows that depend on a vendor-specific capability (a specific tool the vendor provides, a specific RAG variant, a specific agent framework). The decision to use a vendor-specific capability is the decision to forgo the option. Make it consciously, with the trade-off named.

09 THE PUBLIC-SECTOR BUYER'S CHECKLIST

The questions the procurement officer actually has to defend.

A public-sector buyer (ministry, regional government, public hospital group, defence-adjacent agency) is accountable for choices the private-sector buyer is not. The checklist below is the one Cohorte uses when the buyer is on the public side of the desk.

QUESTION	WHY IT MATTERS	THE DEFENSIBLE ANSWER PATTERN
Where does the data live	GDPR, sectoral rules, sovereignty doctrine.	EU region with named processor and sub-processors. DPA on file. Region change requires notification.
Who processes the data	CLOUD Act and similar extraterritorial exposure.	EU-resident personnel for break-glass; audited access; documented contractual structure with sovereign trustee where required by the sector.
Which model is the system using	AI Act transparency, foundation-model obligations, political-risk exposure.	Named in the agent passport. Provider's AI Act technical documentation referenced. Open-weight option available as fallback.
What if the vendor changes the model	Continuity, conformity, predictability.	Re-verification on the calibration set before traffic shifts; rollback to the previous backing model if the bar is not met. The portability architecture from page 08.
Can we move providers if we have to	Political risk, regulatory change, lock-in.	The portability test from page 08. Two-week swap rehearsal at the start of the engagement, documented.
What is the appeal route for a citizen	AI Act Article 86 (right to explanation for high-risk individual decisions); sectoral administrative law.	Documented appeal route, with the supporting evidence (logs, reliability profile, gate-firing record) accessible by the named appeal officer.

> A public-sector procurement that cannot answer these six questions in writing should not sign the contract.

10 TRADE-OFFS HONESTLY

No clean answer. Three honest ones.

The sovereignty conversation is dominated by two unhelpful positions: "use anything American" and "use only European". Neither is supported by the engineering details for most regulated buyers. The three positions below are honest defaults Cohorte deploys depending on the buyer's binding constraint.

Position A: Capability first, with sovereign portability designed in. Use the best-performing model for the workflow, EU-region hosted, with the portability architecture (page 08) installed so the swap to a European backing model is a two-week exercise when the time comes. This is the right default for most private-sector regulated buyers (banks, insurers, hospitals not handling national-security data). The current advantage in capability is consumed for the workflow's verification; the option to move is preserved.

Position B: European-first, with the capability gap measured and accepted. Choose a European model and a European cloud (G3 or G4 from page 06), with the gap to frontier capability measured against the workflow's calibration set. This is the right default for public-sector buyers, defence-adjacent work, and entities whose procurement rules require it. The trade-off is the capability gap; the trade-off is real and is the price of the regulatory alignment the buyer needs.

Position C: Fully air-gapped, with capability reduced to what fits on-premises. Open-weight models on G5 infrastructure. The right default for classified workloads and a small set of national-security-adjacent buyers. The capability is dramatically reduced and the operational cost is dramatically higher; both are justified for the workloads that require it.

> **The decision rule.** The binding constraint should choose the position. If the binding constraint is the procurement rule, choose position B. If the binding constraint is the workload classification, choose position C. If the binding constraint is the workflow's reliability bar at scale, choose position A. The teams that get sovereignty wrong tend to apply the position they prefer to a constraint that requires a different one.

11 WHAT THIS IS NOT

Three statements this briefing does not make.

A sovereignty briefing that does not name the politics it is staying out of has not done its work. The three statements below are the lines this document does not cross, with the reasoning attached.

Not "use European because European is European." Chauvinism is not a procurement criterion. The European AI ecosystem deserves support on the merits and on the structural argument that a continent that does not produce its own foundation models surrenders strategic autonomy. Both arguments are real. Neither obliges a hospital to deploy a sub-optimal system on Tuesday because it is European. The hospital's binding constraint is patient care; the system that supports that constraint best, within the regulatory perimeter, is the one to deploy.

Not "the sovereign cloud question is solved." It is not. The market has gradients; the gradients have costs; the legal landscape is in motion. A buyer who reads this briefing and believes the sovereign cloud question is settled has read the briefing wrong. The briefing argues that decomposing the question makes it tractable, not that someone has solved it.

Not legal advice. The references to CLOUD Act, GDPR, Data Privacy Framework adequacy, sectoral regulators, and AI Act provisions are the practitioner's working understanding. The legal team for the deploying entity is the authoritative source. Where this briefing's text conflicts with the legal team's view, the legal team is right.

> Sovereignty is too important to be left to a slogan. It is also too contested for a single document to settle.

12 HOW THIS LANDS

Cohorte's role. Architect, not advocate.

Cohorte does not take a side on whether European AI should win. Cohorte teaches the architecture that makes the sovereignty question a deliberate decision rather than an emergent property of vendor choices. The engagement pattern below is what the AI Readiness Program installs when sovereignty is on the agenda.

STEP 01 · DECOMPOSITION	The four-dimension decomposition (page 03) applied to the buyer's specific situation. **Which dimension is the binding constraint?** Written, signed by the executive sponsor.
STEP 02 · INVENTORY	Existing AI systems audited against the four dimensions. The gap to the buyer's chosen position (A, B or C from page 10) named in writing.
STEP 03 · PORTABILITY INSTALL	The model interface, retrieval contract and verification layer installed across the portfolio. **The portability test (page 08) passes by end of step 03.**
STEP 04 · SOVEREIGN SWAP REHEARSAL	A two-week rehearsal: one production system swapped to a European backing model, re-verified, re-deployed. **The rehearsal is the evidence the option is real.**
STEP 05 · CONFORMITY BASELINE	The conformity reporting (Briefing 04) wired to the chosen sovereignty position. Procurement-facing documentation produced. Audit-facing documentation produced.

> The sovereignty position is a decision. The portability is the architecture that makes the decision honest.

13 REFERENCES

References. For each claim, the source.

Reading list. The legal texts, the architecture references, and the operational artefacts behind the briefing's claims.

European Union (2024). Regulation 2024/1689 on AI. Articles 9-17 (high-risk obligations), Article 86 (right to explanation), Annex III (high-risk categories).

European Union (2016/2018). Regulation 2016/679 (GDPR). Articles 28 (processor), 32 (security), 44-50 (international transfers).

European Commission (2023). Adequacy decision on the EU-US Data Privacy Framework. Status reviewed periodically by the Court of Justice.

US Congress (2018). Clarifying Lawful Overseas Use of Data Act (CLOUD Act). 18 U.S.C. § 2713.

ANSSI / SecNumCloud (ongoing). French Cyber Security Agency qualification for sovereign cloud. The reference for G3 in France.

European Union (2022). Regulation 2022/2554 (DORA). Articles 28-30 on third-party ICT risk and the register of contractual arrangements.

European Union (2022). Directive 2022/2555 (NIS2). Sectoral cybersecurity baseline for critical entities.

Mistral AI (ongoing publications). Model cards and technical reports for the Mistral Large family and the open-weight Mistral / Mixtral releases. Source for the 2026 capability landscape on the European side.

EU Open-EuroLLM consortium (2025-2026). Public model releases and supporting documentation. The consortium's stated mission and current cadence.

OVHcloud, Scaleway, Aruba (ongoing). The major European hyperscaler offerings. Capability and pricing reviewed quarterly inside Cohorte engagements.

Mouzouni, C. (2026). Context Kubernetes paper. Source for the portability architecture (page 08). github.com/Cohorte-ai/context-kubernetes.

Cohorte (2026). Briefings 01-04 in this series; the CISO & vendor briefing. The cross-references throughout this document. Full record at **teams.cohorte.co/research**.

FOR MINISTRIES, PUBLIC HOSPITALS AND EU BANKS

One discovery call. Bring the binding constraint.

Sixty minutes. We decompose the sovereignty question against your situation, identify the binding constraint, name the position (A, B or C) your situation actually requires, and you leave with a one-page architecture brief.

charafeddine@cohorte.co

Cohorte SAS · Société par actions simplifiée, registered in France · founded September 2022 · Paris & Rabat

Reference architecture, portability test, sovereign-swap rehearsal protocol · all at **teams.cohorte.co/research**