

The Business Case for Sovereign AI

Control of data, models and the production path. Continuum as the operating model. Observable Governance as the spine.

GOVERNANCE FLOWS DOWN · OBSERVABLE FLOWS UP

Executive summary

Boards already buy models and cloud. Staff already use unofficial tools. Public and regulated buyers now ask where the file lives, who holds the keys, and which AI statute applies.

Two clocks are running at once. The EU AI Act is already live for bans, literacy, general-purpose models and — from 2 August 2026 — transparency and Commission enforcement. High-risk product rules were pushed (Annex III to 2 December 2027; product-embedded Annex I to 2 August 2028) but they were not cancelled.

At the same time there is no single “EU residency law”: GDPR allows transfers with safeguards, while France (HDS / SecNumCloud), Germany (social-code cloud rule plus BSI C5), DORA, NIS2 and the European Health Data Space add place-and-operator rules that a US-operated region often cannot meet.

McKinsey finds 44 in 100 European technology leaders citing security and 31 in 100 citing in-country or in-region storage as reasons they will not use public cloud — the same freeze that keeps pilots in the sandbox.

The question is what control is worth paying for, on which work, so a live assistant can touch an official file without handing the estate to a stack and a statute you do not accept?

A popular approach is to classify the work. Assign a sovereignty tier (place, operator, stack, law). Map the AI Act role — provider, deployer, or both — and the risk tier. Put that tier on one Continuum path: named stop owner, attested door including a governed MCP for agents, fail-closed if the rules engine is silent, searchable record of allow or deny with the rule version. Do not buy a second sovereign-AI tower. Measure success as live work under a query you can defend to a board and a supervisor.

To support this, McKinsey sizes Europe’s high-sovereignty path at up to €480 billion of GDP a year by 2030, mostly from adoption that today does not start. They size a \$500–600 billion sovereignty-shaped AI market by 2030. None of that arrives if the official tool cannot connect and cannot be halted. Governance flows down. Observable flows up.

Governing thought

Sovereign AI is not a national flag on a data centre. It is the ability to decide who may touch which work, under which law, on which stack — and to prove it. Without that, “sovereign” is a procurement label. With it, AI can leave the sandbox in regulated work.

KanTec Continuum is how that decision travels down the path. **Observable Governance** is how you prove it travelled. Sovereignty that cannot be queried is theatre — the same defect as a pilot with no guardrails and no door.

Words used here

- Sovereign AI — enough control over data, compute, models and decisions that a foreign law, outage or vendor change cannot silently take the work.
- Minimum sufficient sovereignty — only as much control as that workload needs. Not “build every model in-house.”
- Territorial — where the data and computers sit. Operational — who runs them. Technological — who owns the stack and the IP. Legal — which court and statute apply.
- Production path — the live process, live records, live customer. Not a trial.
- Continuum / Observable Governance — shared path and searchable record of allow or deny, including which rule version applied.

1. Why this is a business case, not a slogan

Situation

Boards already fund cloud, models and copilots. Staff already use unofficial tools. Public buyers and regulated firms are being asked where the data lives and who can compel it.

Complication

- Many pilots never leave the trial because there is no named door into the official system and no rule that can halt an agent. Sovereignty questions freeze the same programmes.
- A US or other extra-territorial statute can reach data you thought was “in region” if the operator or the keys sit elsewhere.
- McKinsey finds European technology leaders naming security (44 in 100) and the need to keep data in-country or in-region (31 in 100) as reasons they will not use public cloud — which is also why official AI stalls.

Question

What level of control is worth paying for, on which work, so AI can run live without handing the estate to a stack you do not govern?

Answer

Classify the work. Assign a sovereignty tier. Put that tier on Continuum: dual-form rules, attested door (including a governed MCP for agents), searchable record. Do not buy a second “sovereign AI tower.”

2. What McKinsey puts on the table

- 71 in 100 executives, investors and officials in a global McKinsey survey called sovereign AI an existential concern or a strategic imperative.
- They estimate 30 to 40 percent of AI spend could be shaped by sovereignty rules — a market on the order of \$500 to \$600 billion a year by 2030.
- For Europe, a high-sovereignty path is sized at up to €480 billion a year of GDP by 2030: about €416 billion from wider use of AI (takers) and about €63 billion from European firms keeping a share of the supply (makers).
- They are explicit: CEOs delay scale because a sovereign offering is missing. The constraint is adoption, not model quality.

- Global AI spend is put in the \$1.3–1.5 trillion range by 2030; gen AI value at stake remains in the \$2.6–4.4 trillion-a-year band — most of it unrealised while work stays in trial.

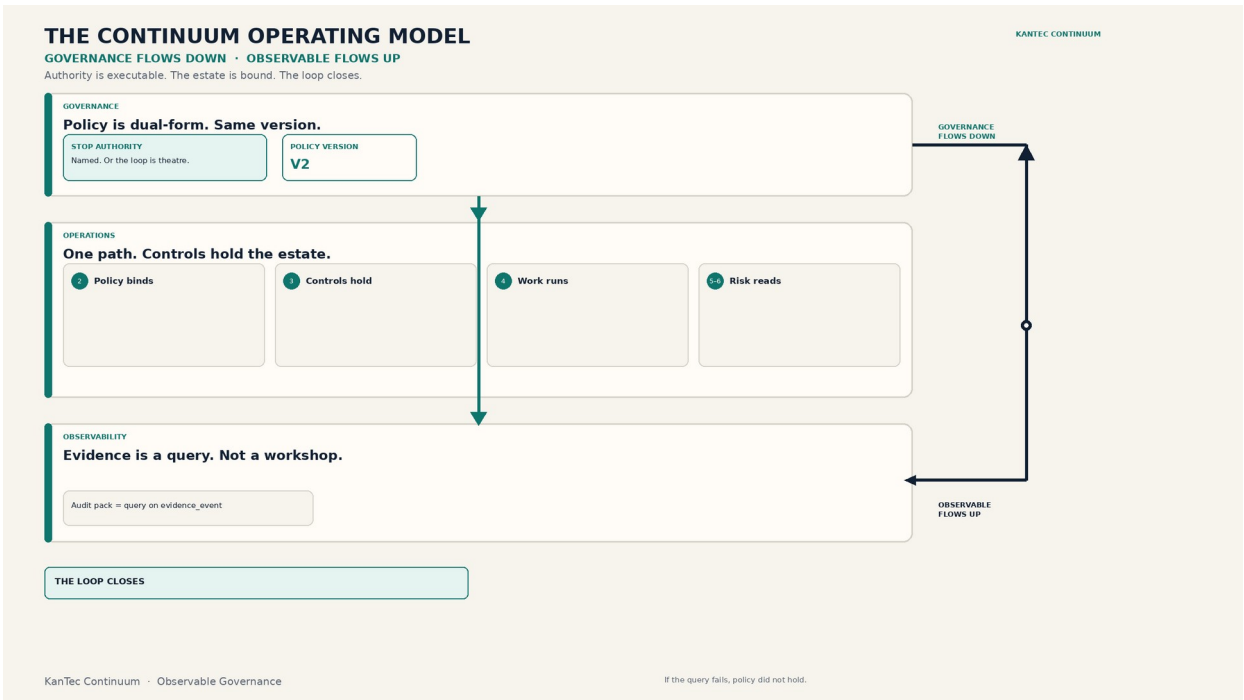
McKinsey does not invoice “the cost of not being sovereign” per firm. They price delayed adoption and missed GDP. That is the same shape as the pilot-trap numbers: use at the desk, nothing in operating profit.

3. Four kinds of control — pick per workload

Figure. Sovereignty is not a sticker on the top box. It is whether policy can travel down this path and whether the record can travel back up — in a jurisdiction you accept.

Kind	Business question
Place	Do the files and the computers sit where our law expects?
Operator	Who holds the keys, the admin rights and the support contract?
Stack	Can we change model, cloud or tool without losing the work?
Law	Which court can compel a copy? Which AI statute applies?

McKinsey’s working rule: minimum sufficient sovereignty. Classify the work by how sensitive the regulation is and how exposed you are to a third party. Then pick a tier. A public chatbot drafting a blog is not a claims file. A claims file is not a national identity system.



4. Why sovereignty fails without a production path

- A “sovereign region” with no checkpoint in front of the agent is still shadow AI, just hosted nearer home.
- A national model that cannot write to the official system is still a sandbox.
- A residency clause in a contract, with no searchable record of which rule version allowed the last write, will not survive a regulator or a board.

The business case therefore has two layers. Layer one: where and under whom the stack runs. Layer two: Continuum — stop owner, attested door, fail-closed if the rules engine is silent, evidence you can query. Layer two is what lets layer one pay off.

5. Payoff for the buyer

Unstick live work

Regulated teams stop waiting for a pack that never comes. The same path that satisfies residency is the path that can halt an agent and show last month's yes or no. That is how unofficial tools become named tools, and how trials reach the official file.

Keep value that would otherwise leave

If 30 to 40 percent of AI spend will follow sovereignty rules, the firm that cannot show control loses that work to a competitor who can — or never starts it. Europe's €480 billion figure is mostly taker productivity. You capture it only if adoption actually happens.

Avoid a second plane

Do not stand up a "sovereign AI office" beside the AI office beside the security tower. One control plane. Sovereignty is a tier on the workload, not a new department.

What this is not

- Not "we must train our own frontier model." Most work needs a governed door and a residency tier, not a science project.
- Not a ban on foreign models. It is a rule about which work may use them, with which keys, with which record.
- Not cheaper compute. Sovereign options often cost more. You pay for control and for adoption that would otherwise never start.

6. EU AI Act — what is live, what is deferred

The Act (Regulation 2024/1689) is product-and-use law, not a management system. Provider means you put the system on the market or into service under your name. Deployer means you use it. One firm is often both. Fines can reach €35 million or 7 percent of worldwide turnover for the worst breaches.

Already in force

- From 2 February 2025 — banned practices (Article 5) and AI literacy (Article 4). Do not run social scoring or the listed prohibited biometric uses. People who operate the systems need to understand them.
- From 2 August 2025 — duties on providers of general-purpose AI models: documentation, a public summary of training content, copyright policy, and extra duties if the model is treated as systemic-risk. Models already on the market before that date have until 2 August 2027.
- From 2 August 2026 — Article 50 transparency: tell a person they are talking to an AI; label deepfakes and other synthetic media; machine-readable marking of AI-generated content. Systems already on the market have until 2 December 2026 for the marking rule. The Commission's power to enforce general-purpose-model duties, including fines, also starts 2 August 2026.

Deferred — not cancelled

- Stand-alone high-risk uses in Annex III (work, education, essential private services, some public uses) apply from 2 December 2027.
- AI built into already-regulated products (Annex I) applies from 2 August 2028.

Those dates moved under the 2026 Digital Omnibus on AI. Use the extra time to put the path in place. Do not read delay as “high-risk does not apply to us.” Classify every live assistant now: banned, general-purpose model, transparency-only, or likely high-risk when Annex III bites.

What the Act does not do

It does not, by itself, keep data in France or Germany. It does not replace GDPR, DORA or sector hosting rules. It does not create a searchable production record. That record is the control plane.

7. European data residency — a patchwork, not one law

GDPR does not say “all personal data must stay in the EU.” It says transfers need an adequacy decision or proper safeguards. Boards keep missing that distinction — and then colliding with sector and national rules that do require place and operator.

- GDPR Chapter V — transfers out of the EEA need adequacy or safeguards (standard clauses, binding rules). Not a blanket lock-in.
- Schrems / extra-territorial access — even an “EU region” can fail if the operator or the keys can be compelled under a non-EU statute. Sovereignty is operator and keys, not only the pin on the map.
- France health data (HDS) — hosting personal health data collected in care requires a certified host. The updated rules require hosting in the EEA and honesty about access from outside. SecNumCloud for French public cloud use is stricter: store and run in the EU, administer in the EU, and resist third-country public-authority demands.
- Germany health and social data (§393 SGB V) — cloud processing only in Germany, the EEA, or an adequacy country, with a German establishment and BSI C5 assurance. Standard clauses alone are not enough for that slice of work.
- European Health Data Space (Regulation 2025/327) — secondary use of health data is expected to be stored and processed in the Union, with narrow adequacy-based exceptions.
- DORA (financial digital operational resilience) — not a general residency ban. Critical ICT providers need an EU presence; contracts must make concentration and exit real. NIS2 adds incident and supply-chain duties for essential and important entities.
- Public procurement — many Member State tenders now ask for SecNumCloud, C5 Type 2, or an equivalent “immune to third-country law” clause. That is how residency shows up in a buying decision.

Working rule: GDPR is the floor. Sector plus Member State rules are the lock. Minimum sufficient sovereignty means you apply the lock only to the workloads that carry it — health file, claims file, public identity — and you write that tier onto the Continuum path so an agent cannot silently use a global endpoint.

8. Schrems II — why an “EU region” is not the end of the argument

In July 2020 the EU Court of Justice struck down the old Privacy Shield (Schrems II). Standard contractual clauses survived, but only if the exporter can show, in a transfer impact assessment, that the law of the destination country does not give public authorities access that undercuts EU essential guarantees — or that extra technical and organisational measures close that gap.

- US surveillance law (in particular FISA 702 and equivalent executive powers) was the fact pattern. An adequacy deal (the Data Privacy Framework) later returned for certified US organisations. It is a political instrument. It can be challenged again. It does not cover every processor or every sub-processor.
- Practical effect for AI: inference and logs are personal data when they carry a customer or staff identifier. A model endpoint in an “EU region” still fails the test if the operator, the support staff or the key-management service can be compelled from outside the EEA, or if the fallback model is global.
- What a board should demand: where processing happens; who can administer; where keys live; which sub-processors see prompts and outputs; what happens on a lawful-access request; whether the same workload can fail closed onto an EEA-only path.
- Observable Governance is how you prove the last point. If the record cannot show that last month’s writes stayed on the tiered path, the transfer impact assessment is a memo.

Schrems II is why sovereignty is operator and keys, not a pin on a map. Continuum writes that choice onto the live path so an agent cannot silently hop to a global endpoint.

9. DORA — digital operational resilience for financial work

DORA (Regulation 2022/2554) has applied since 17 January 2025 to banks, insurers, investment firms, some crypto and other financial entities, and to the ICT providers they use. It is a cyber and operational-resilience rulebook, not a residency statute. It still decides whether an AI stack can run live on a claims, payments or trading path.

What the entity must run

- ICT risk management — protect, detect, recover. The management body owns it. You cannot outsource the duty.
- Incident handling and reporting — classify, notify, learn. An agent that writes a bad payment is an ICT incident, not “just a model.”
- Digital operational resilience testing — regular tests; some firms must run threat-led penetration testing at least every three years, and ICT providers must cooperate.
- Information-sharing on cyber threats — permitted and expected among peers, under controls.

Third-party ICT — the part that hits AI vendors

- Article 28 — a written third-party ICT strategy, a register of every ICT contract (the Register of Information), a risk assessment before you sign, and a real exit plan for important functions.
- Article 29 — concentration risk. One model host, one cloud, one MCP farm for every critical function is a DORA finding waiting to happen.
- Article 30 — mandatory contract terms: access, security, integrity of data, help on incidents, audit and inspection, exit and return of data, cooperation in tests. Deeper terms when the service supports a critical or important function.
- Critical ICT third-party providers can be designated by the European Supervisory Authorities. If designated, they need a Union subsidiary and come under a lead overseer — including groups based outside the EU.

DORA does not ban a non-European cloud. It does ban “we have a contract, so we are fine.” The financial entity stays responsible. An AI assistant on a payments path is an ICT service supporting a function. It belongs in the register, in the concentration view, and on a Continuum path with an exit — fail-closed onto a named alternative, not a hope.

How DORA and Schrems II meet on this plane

- Schrems II asks: can a non-EU power read this file? DORA asks: can we still run the book if that provider fails, is compelled, or is concentrated?
- Both are answered by the same architecture: tier the workload, name the door, record the decision, keep an exit.

10. 90-day move

- Day 0 — name the stop owner and the workloads that cannot sit on a foreign-operated stack.
- Days 1–30 — tier those workloads (place, operator, stack, law). List unofficial tools already touching them.
- Days 31–60 — one beachhead path: attested door (API and, for agents, governed MCP), rules live, residency and key-ownership explicit.
- Days 61–90 — searchable record works. First rule change cites live evidence. Board pack is that query, plus the tier list.

Scale by cloning the path onto the next regulated process. Do not scale by announcing a national model.

11. Decision

Treat sovereign AI as a workload tier on Continuum, not as a poster or a second cloud contract. Fund minimum sufficient control on the processes that are blocked today. Measure success as live work under a queryable rule — in a place and under an operator the board will defend.

Governance flows down. Observable flows up. Sovereignty that cannot travel that loop is not sovereignty.

— KanTec Platform Engineering · Business case —