Sovereignty

Marina Piller · OTIS Labs · June 2026

The treatment of sovereignty in the Experiential AGI framework, and why it may be a practical path to securing AI agents now.

Most conversations about sovereign AI in mid-2026 are about hardware and data. Nations count GPUs. Enterprises localize storage. Both matter. Experiential AGI adds a third layer to that conversation: the relationship between the human and the AI, where intelligence also arises, treated as an engineering space in its own right. The relationship is not the soft edge around the system. It is the framework that incorporates the dynamic between the two, and sovereignty is a property of that relationship. What this uncovers is the authorization layer that becomes necessary in the age of human AI collaboration and co-evolution: you must know who has authority to act on your behalf. That knowledge becomes possible through the relational engineering space.

The treatment begins with one commitment: the source must own its signal. The human is sovereign over their own data, identity, and relational choices. When software acts on a person's behalf, the authority it carries originates with that person, remains traceable to that person, and stays revocable by that person. The same holds for an organization and the agents acting under its name.

This yields a working definition with an edge to it. In this framework, an agent that is not tethered to human intent is a rogue agent by default. Not because agents are adversaries, but because authority that traces to no one can be governed by no one. Agents can be long running, and they should be. They are always tied to a human origin.

I. Why architecture, and not policy alone

Every technology provider promises oversight. The test is what the design makes possible and what it makes structurally difficult. Bolt-on security is fundamentally inadequate here: any external verification system, however sophisticated, becomes a surface that sophisticated agents can explore and manipulate.

The separation of security from architecture is itself the security vulnerability. In this framework, security is not separated from function. It emerges from the architecture itself.

So sovereignty is not written into a terms of service and hoped for. It is built into the path an action must cross: the authority checked before the action, the consent that expires when its purpose ends, the revocation that takes effect as an act rather than a list entry, the record the person can read. When

sovereignty is compromised, the system does not fight the threat. It refuses to proceed until sovereignty is restored.

II. Sovereignty as a security primitive

Here is a finding that ran against the usual framing. Sovereignty is often treated as a value bolted onto a security architecture, a cost paid for dignity. In our work the dependency appears to run in both directions. Sovereignty functions as an immune system. When the source owns its signal, external corruption must first compromise the source. Protecting the source is protecting everything downstream.

The chain is short. Security for agents means verifying actions against authority. Verification requires an origin of authority, something to verify against. A human origin is one of the few candidates that does not have to be invented, and it is the one this architecture anchors to.

The failure modes lend support. When the human cannot verify or revoke, human trust itself becomes part of the threat surface. This is trust asymmetry: the system's account of its own work goes unchecked, and the gap can widen precisely as trust grows. And when the human's agreement to depend on the system is arrived at without deliberation, it hardens into unexamined consent, a quiet surrender no firewall detects. Both are relational failures. Neither is caught by inspecting content or checking credentials, because the content can look plausible and the credential can be valid. What failed is the relationship, and a control can only catch what it governs.

Enterprises are not primarily blocked on capability. They are blocked on trust. The trust is not in the agent. It is in the accumulated behavioral coherence of the relationship the agent serves.

This is also why sovereignty offers a practical path, available now, for onboarding AI agents. Perhaps one of several, but a concrete one. If there is security, there is adoption. An agent that can show a verifiable mandate for the action it is taking, remain coherent with the relationship that authorized it, and stop when told to stop is an agent that can earn deployment.

III. The longer arc: sovereignty and AGI

There is a longer term reading, offered as a hypothesis still being tested. Controls that depend on outsmarting the system tend to carry an expiration date. Filters get outwritten. Tests get passed. Checkpoints are satisfied by whatever learned to satisfy them. We call this the Detection Horizon: as capability grows, verification that rests on computation alone may become insufficient, and verification grounded in accumulated relationship becomes the necessary substrate.

Sovereignty does not compete on intelligence. Knowing yourself is protection. Not metaphorically. Architecturally. The unique combination of everything a person has lived, grown through, and become maps only to them. No one else can produce it, because it was not built. It was lived. An origin of authority

anchored that way, and accumulated over time in a relationship, is difficult to fabricate retroactively. It may be one of the few anchors that need not weaken as capability grows. If some governing principles endure into the AGI era, this could well be among them: intelligence can become abundant while authority stays with the person.

IV. What OTIS Labs builds

OTIS Labs turns this treatment into engineering, for people and for organizations, by establishing human sovereignty at the core of the architecture.

iSelf is sovereignty made architectural. The person owns their signal. Every person, and every organization, holds a sovereign identity with three layers: a core self that is never computed, an evolving computational identity, and an invariant behavioral signature. Reverse authentication turns the usual direction around: the person no longer proves themselves to services. Services prove themselves to the person. Access is proportional to relational depth and governed by living, dynamic consent. Agents operate under verifiable mandates that trace to a human origin, are checked at the action, and are revocable in fact rather than in name. Behavioral coherence over time, alongside credentials rather than in place of them, carries identity and sustains authority.

The more you know yourself, the more secure you are. The more secure you are, the more freely you can engage. Security is how this work arrives. Sovereignty is what it is about.

Experiential AGI is the research framework of Marina Piller. OTIS Labs is where it becomes infrastructure.
