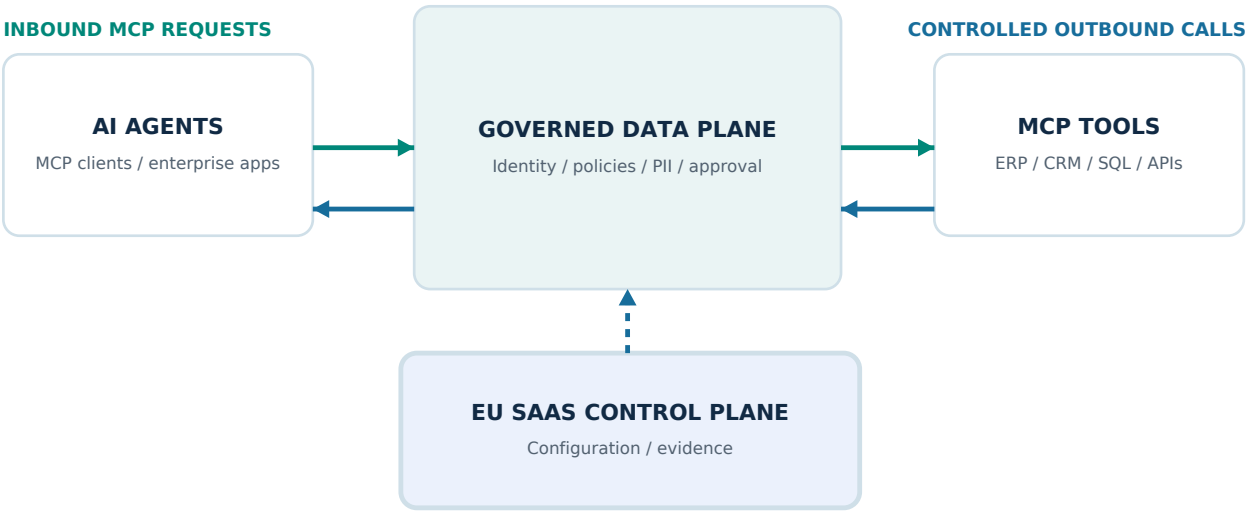
PUBLIC CONCEPT WHITEPAPER · VERSION 0.1

Sovereign AI Governance

An enterprise control plane for what AI agents are allowed to do.

Govern MCP-connected AI agents through enterprise authorization, privacy-aware execution, human approval, and verifiable activity evidence—while retaining control over where sensitive tool traffic runs.



Status: Product concept and planned architecture. No production availability, certification, security assurance, or regulatory compliance is implied.

Prepared for prospective design partners, enterprise security leaders, AI platform teams, and early ecosystem collaborators.

01 Executive summary

Enterprise AI is shifting from answering questions to taking actions: retrieving customer records, creating purchase requests, updating applications, and invoking business APIs. These actions turn integration design into a question of operational authority: who permitted an agent to act, under which conditions, on which data, and with what accountability?

Sovereign AI Governance proposes an EU-oriented governance platform that sits between MCP-capable agents and enterprise tools. The proposed service combines a centrally managed control plane with policy enforcement at the execution boundary.

From connecting AI agents to governing enterprise actions.

The proposed foundation

- Action authorization: assess agent identity, delegated authority, tool scope, validated arguments, and business conditions before execution.
- Privacy guardrails: inspect relevant tool data for PII and secrets, with explicit allow, redact, tokenize, or block outcomes.
- Human oversight: require authorized approval of the exact proposed action before sensitive writes.
- Audit evidence: generate attributable, minimized records of decisions and outcomes.
- Deployment choice: EU-managed gateway or customer-hosted edge data plane under a common control plane.

The product is planned, not launched. This paper describes a design direction and testable commitments rather than current customer deployments or guaranteed regulatory outcomes.

02 The enterprise problem

MCP standardizes how AI applications interact with tools. It does not by itself resolve enterprise ownership of every delegated business operation. Conventional identity and API infrastructure can provide important controls, but agent workflows introduce dynamic tool selection, chained actions, variable arguments, and approval requirements that span multiple systems.

Four practical questions

Question	Enterprise risk if unanswered
Who is acting?	An agent may use a broad service credential without an attributable human owner or limited mandate.
What may it do?	A tool might be technically accessible even when a specific transaction violates an internal policy.
What data leaves?	Tool results can contain personal data, identifiers, confidential fields, or credentials.
Who authorized it?	Sensitive actions may be executed without an approved decision or reconstructable evidence.

Illustrative workflow

A procurement agent can draft a purchase order, but releasing an €8,000 order may require a finance manager. A simple tool allowlist cannot necessarily distinguish an acceptable draft from an unauthorized commitment. An enterprise governance layer must evaluate the action semantics and preserve approval integrity.

The proposed wedge

Focus first on organizations moving from read-only AI demonstrations to governed workflows that write to business systems. The initial target is German and DACH mid-market enterprises with existing identity platforms, defined approval controls, and MCP-enabled agent pilots.

03 Proposed solution and value

The platform is designed around governed autonomous execution, not just centralized tool discovery.

Product pillar	Intended customer outcome
AI Action Firewall	Decisions of ALLOW, DENY, or REQUIRE_APPROVAL based on registered tools and trusted business parameters.
Agent identity & delegation	Task-limited agent permissions tied to an accountable owner; expiry and revocation.
Guardrail engine	Policy-defined inspection of selected request and response fields for PII, secrets, and unsafe data flows.
Human-in-the-loop	Authorized approval, bounded to canonical action arguments with expiry and replay protection.
Evidence & traceability	Searchable audit metadata, policy versions, decision traces, and controlled exports.
Sovereign deployment options	Managed EU gateway for simpler onboarding; local customer edge for sensitive execution paths.

Differentiation thesis

Generic gateway capabilities are increasingly available. The intended advantage is the integration of business-specific authorization, customer-controlled execution, and evidence-rich oversight into one configurable enterprise workflow. This advantage remains a market hypothesis to validate with design partners.

What the product is not

It is not an AI model, a replacement for enterprise identity providers, an automatic legal-compliance certificate, or a universal guarantee against prompt injection or sensitive-data leakage.

04 Architecture: one control plane, flexible data planes

Two execution models share the same conceptual Go gateway and governance contracts. The control plane distributes approved configurations and receives sanitized operational evidence. The data plane handles business tool calls and enforces decisions at the point of execution.

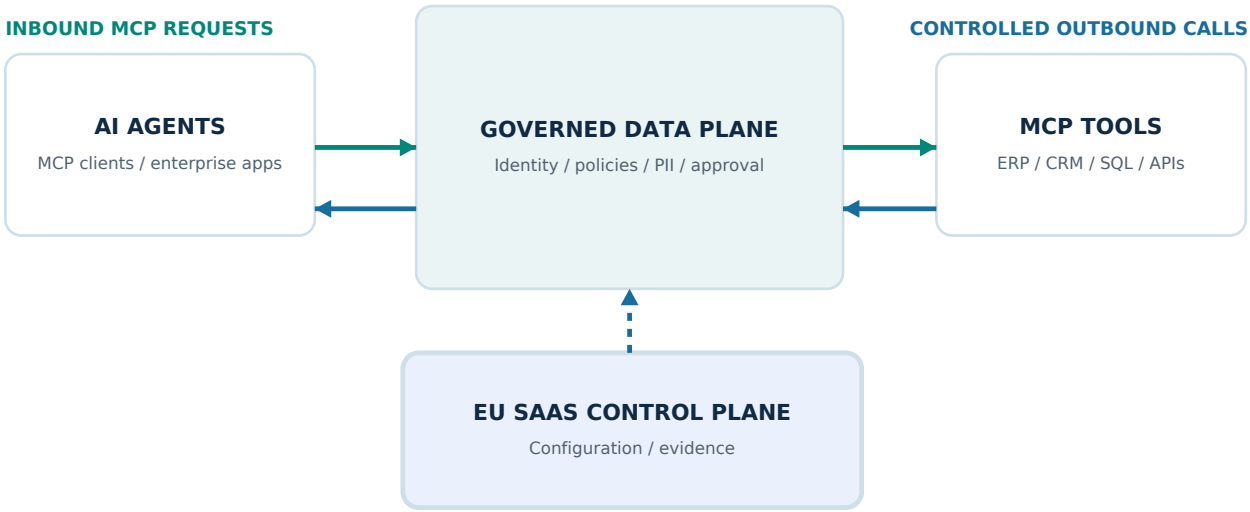


Figure 1. Managed SaaS: inbound MCP requests are authenticated and governed before approved outbound calls reach registered tools; results return through the same enforcement boundary.

Enterprise edge execution

In the customer-hosted mode, the same gateway design runs inside the customer network. Agent requests travel from local applications through local enforcement to private MCP servers. A separate outbound-initiated management channel synchronizes signed policies and approved, minimized evidence with the EU control plane.

In both modes, the gateway revalidates the action and its exact approved arguments before forwarding the call. Return traffic is governed before reaching the originating agent.

Critical design principle: in the customer-edge mode, no raw business payload is sent to the central SaaS by default. Some customer-selected approvals or evidence fields can still be personal data and require a documented lawful processing arrangement.

05 How governed action execution works

Consider a procurement agent calling submit_purchase_request. The proposed gateway does more than authenticate the HTTP request.

Step	Control	Outcome
01	Identify agent, tenant and accountable owner	Reject unknown or expired authority.
02	Resolve registered tool and validate arguments	Reject unregistered targets or malformed data.
03	Evaluate local published policy	ALLOW, DENY, or REQUIRE_APPROVAL.
04	Inspect relevant request fields	Apply configured DLP and sensitivity controls.
05	Obtain approval if required	Bind approver to exact action and argument digest.
06	Revalidate and execute once	Protect against changed arguments and retries.
07	Inspect result and return to agent	Apply response field policies and schema checks.
08	Record minimized evidence	Attribute decision, policy version, and execution outcome.

An agent’s ability to call a tool is not the same as authority to commit a business action.

Security invariants

- Authorization decisions come from trusted server-side context, never from the agent’s own claim of authority.
- High-impact actions are fail-closed when mandatory policy or approval components are unavailable.
- Approvals expire, bind to the exact transaction, and cannot safely be replayed.
- Audit records minimize personal data and must not contain raw access tokens or API secrets.

06 Privacy, sovereignty, and regulatory support

The design seeks to reduce unnecessary data movement and help enterprise customers demonstrate governance controls. It does not establish GDPR or EU AI Act compliance simply by being EU-hosted.

Privacy by design

- Inspect only fields needed for the configured policy; avoid raw payload retention by default.
- Apply schema-aware PII and secrets detection with documented detection limits.
- Use tenant isolation, bounded retention, role-controlled evidence access, and protected secrets.
- Keep private-network tool traffic local when customer-hosted execution is selected.

Regulatory support boundaries

For GDPR, a hosted service may still be a personal-data processor even when inspection happens only in memory. Customer/provider responsibilities, appropriate agreements, records of processing, breach response, transfer analysis, and security measures need separate consideration.

For the EU AI Act, policy, audit, and approval evidence may support applicable oversight and record-keeping obligations. Duties depend on the system, risk category, provider/deployer role, and use case; the platform cannot make an entire AI system compliant on its own.

Proposed evidence exports

Policy version history, tool authorization decisions, action-approval records, data-protection events, agent ownership, and correlation IDs. Export templates are intended as operational evidence—not a legal determination.

Reference sources and applicability notes appear at the end of this paper.

07 Initial use cases and target customers

Workflow	Proposed control	Why a buyer cares
Procurement agent	Allow draft creation; require approval above spend threshold; audit exact execution.	Reduce the risk of unauthorized commercial commitments.
Customer support agent	Permit approved customer lookups; mask sensitive fields; block unsafe exports.	Protect customer records while enabling agent assistance.
HR operations agent	Limit employee-data access to approved roles and purpose-bound agents.	Preserve personnel-data boundaries and accountable access.

Target buyer group

Primary: CTO, Head of AI Platform, and enterprise application owners. Governance stakeholders: CISO, DPO, internal audit, and business process approvers. Early ecosystem partners: German IT consultancies and integrators deploying AI agents on behalf of Mittelstand customers.

Design-partner engagement

- Choose one real business workflow involving sensitive reads or controlled writes.
- Map existing identity, tool permissions, approval rules, and data boundaries.
- Validate whether gateway-based enforcement can apply to that workflow without disrupting operations.
- Use a pilot to measure policy accuracy, integration effort, latency, evidence completeness, and operator confidence.

No pilot customers, partner logos, commercial commitments, or market adoption are asserted in this paper.

08 Proposed delivery roadmap

Stage	Planned capabilities	Validation gate
Foundation	Multi-tenant control plane, identity, tool registry, protocol-compatible MCP gateway.	Secure client-to-tool routing with tenant isolation.
Governance MVP	OPA business policies, basic PII rules, exact-action approvals, immutable-style audit evidence.	End-to-end controlled procurement workflow.
Enterprise pilot	Customer edge gateway, signed policy distribution, local DLP, audit buffering, reference connectors.	Safe offline behavior and data-boundary validation.
Expansion	Additional domain policy packs, enterprise connectors, deeper evidence and security assurance.	Repeatable customer onboarding and validated demand.

What is planned versus available

All product capabilities described in this whitepaper are proposed or planned unless subsequently announced as available in a release note. The architecture is a design baseline. No production SLA, independent certification, or specific release date is promised.

Success criteria for early validation

- Demonstrate unauthorized write prevention under realistic agent behavior.
- Demonstrate auditable authorized approval of an unchanged action.
- Demonstrate minimized exposure of PII in controlled tool responses.
- Demonstrate that edge payload traffic remains within the defined customer boundary.

This staged approach prioritizes measurable security outcomes over a broad connector catalog.

09 Invitation to design partners

The proposed platform is intended for enterprises that want to deploy capable agents without surrendering control of business operations.

Turn agent prototypes into governed business workflows.

We are seeking discussions with enterprise AI platform teams, security leaders, and technology partners who are evaluating production MCP-enabled agent deployments.

A useful first conversation would cover: a single target workflow; which tools and data it accesses; current identity and approval mechanisms; desired execution location; and evidence that security teams need to authorize production use.

Potential collaboration formats

- Discovery interview: understand blocked or delayed agent use cases.
- Architecture review: test feasibility of managed SaaS or customer edge deployment.
- Design-partner pilot: co-define measurable security and operational acceptance criteria.

Publication note: Add the official company name, website URL, and a verified contact address before public release. Avoid publishing a public waitlist form until its notice, retention, and consent/other legal-basis arrangements are configured.

The information in this whitepaper is a forward-looking product concept and may change. It is not legal advice, a binding offer, or a warranty of product capabilities.

References and terminology

Source material

Internal product concept: Sovereign EU MCP SaaS & Enterprise Agent Governance Control Plane — Product Blueprint (4 pages). Supporting design baseline: Product Strategy & Requirements, System & Traffic Architecture, Security/Privacy/Compliance, and Implementation & Codex Delivery Plan (20 November 2026).

External references (checked 20 November 2026)

1. Model Context Protocol, 2026-07-28 specification and Streamable HTTP transport: https://modelcontextprotocol.io/specification/2026-07-28

2. European Union AI Act, consolidated Regulation (EU) 2024/1689 as amended: https://eur-lex.europa.eu/eli/reg/2024/1689/2026-07-27/eng

3. European Commission, AI Act implementation timeline: https://ai-act-service-desk.ec.europa.eu/en/ai-act/timeline/timeline-implementation-eu-ai-act

4. GDPR, Regulation (EU) 2016/679: https://eur-lex.europa.eu/eli/reg/2016/679/oj

Terminology

Control plane: service for governance configuration, policy lifecycle, identity mappings, administration, and evidence access.

Data plane: execution path through which authenticated MCP requests, policy decisions, tool invocations, and responses pass.

AI Action Firewall: proposed policy enforcement that evaluates the meaning and permitted authority of an agent’s tool call.

HITL: human-in-the-loop approval before a sensitive action is permitted.

PII: personally identifiable information; used here informally for personal-data detection controls, not as a replacement for GDPR terminology.

Version 0.1 · Public concept edition · 20 November 2026