WIKEY

Protect your data and assets in the agentic world.

The cryptographic identity & security layer for AI agents.

3	Live	Any
active pilots	production identities	system the agent touches

THE NEW ATTACK SURFACE

Stolen keys, hijacked agents, and a recovery path.

These follow from how keys and logs work today — not from any vendor's mistake.

KEYS CAN BE STOLEN

What the agent holds.

A stored secret is an extractable secret — prompt injection, exfiltration or a rogue dependency lifts it, and it's copied into every backup. Whoever holds it inherits the access.

THE AGENT CAN BE TURNED

No key theft required.

Phishing and prompt injection can trick the agent into releasing data or approving an action — without ever taking the key. The key stays put; the data still leaks.

RECOVERY IS THE SOFT TARGET

The way back in.

Help-desk and reset flows bypass even phishing-resistant MFA — and AI voice and video now beat the human check those resets rely on.

Drift / Salesforce: stolen agent tokens exposed data across hundreds of organizations. **MGM · Caesars · M&S · Co-op:** breached at the recovery desk, not the login.

The cryptography held. The keys — and the way back in — did not.

TODAY'S TOOLS

Current trust layers secure the platform — not the key.

Each is strong inside its own estate. None follows an agent that doesn't stay in one.

Tool	What it secures	Where it stops
Salesforce Einstein Trust Layer	Agentforce prompt defense	Anything outside Salesforce
Microsoft Entra Agent ID	Agent identity in the MS estate	Tenant-bound; rest via MS gateway
AWS Bedrock AgentCore Identity	Agent creds & MCP auth	Workloads & data outside AWS
Okta / Auth0 (identity providers)	Issues agent tokens (OAuth)	Tokens stay stealable
Astrix / Oasis (non-human identity)	Governs machine credentials	Governed after the fact
Silverfort & inline gateways	Blocks calls at one choke point	Authority it carries elsewhere

*Guard one source at a time and you always miss the layer below — **the keys the agent carries from source to source.***

OUR ANSWER

Agents hold a cryptographic, keyless identity.

Hold the key where it can't be read or copied — and guard the data with an engine the agent can't bypass.

VIRTUAL HSM PROTECTS KEYS

Held, never accessed.

Keys live in a post-quantum virtual HSM — the secure enclave or TPM — separated from the agent and the app. Never backed up, so there's no copy to steal or restore.

SCOPED & TRACEABLE

Every action ties to a person.

The agent acts under a scoped, time-bound, revocable sub-identity that descends from a human owner. Credentials are short-lived and refreshed often — a stolen one expires fast.

DATA STAYS PROTECTED

Watched, never trusted.

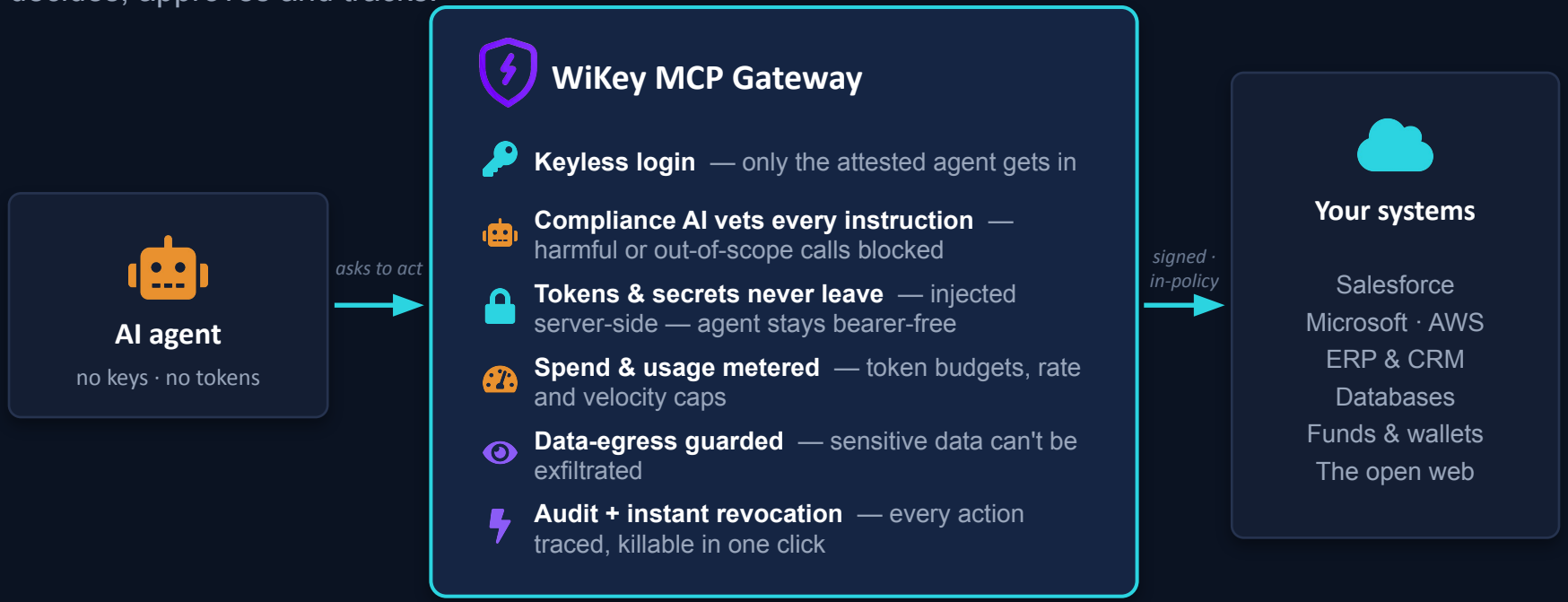
An external engine supervises every login, read, write and transfer — without ever seeing the data. Even a hijacked agent can't move data or funds outside policy.

Filters reduce the attacks. **Scope, supervision and revocation limit the damage.**

HOW IT WORKS

Every agent action flows through one gateway.

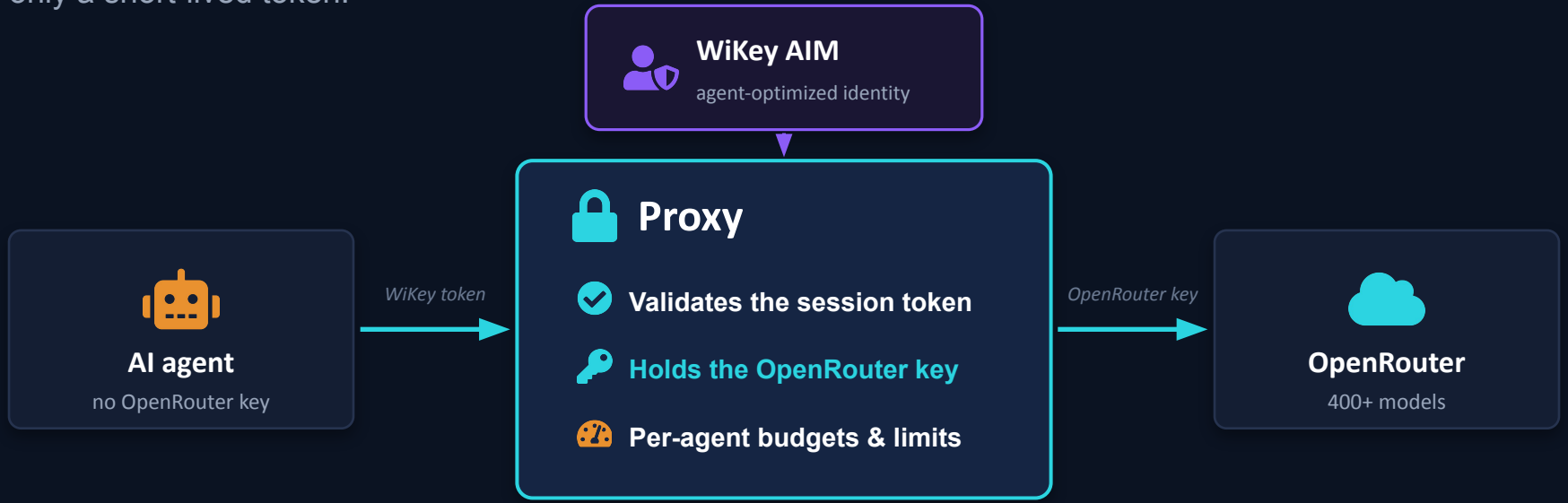
The agent holds no keys or bearer tokens — it asks the gateway to act, and the policy-based gateway decides, approves and tracks.



REFERENCE ARCHITECTURE

Example: OpenRouter access, no key on the agent.

WiKey's agent-optimized AIM sets policy; a LiteLLM proxy holds the OpenRouter key — the agent carries only a short-lived token.

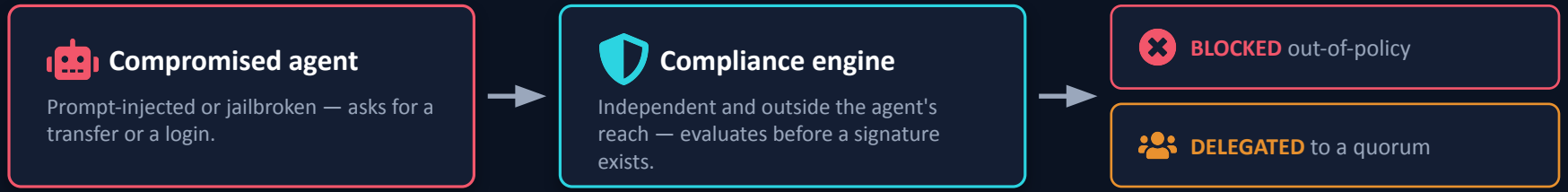


The agent never holds the OpenRouter key — only a short-lived WiKey token. ***The proxy swaps in the real key, so a leaked token goes nowhere.***

UNDER ATTACK

Compromise the agent. The damage stops here.

Whoever is compromised, the compliance engine decides independently — nothing moves that policy hasn't approved.



HOW WIKEY CLOSES EVERY DOOR

Prompt injection	The agent has no key to leak — it requests signatures, it can't produce them.
Jailbroken reasoning	Scope and limits are enforced outside the model, where reasoning can't reach.
Stolen tokens & seeds	Nothing for a thief to hold — credentials are short-lived, scoped, revocable.
Compromised dependency	Recovery is by guardian attestation — same flow for one agent or ten thousand.

WIKEY

WHY IT MATTERS

Real agent breaches — closed by design.

Each is a public compromise of a production AI agent. WiKey's gateway would have closed the door on all three.

2025 · STOLEN AGENT TOKENS

Salesloft Drift

Its OAuth tokens were stolen and used to export data from hundreds of Salesforce orgs.

WiKey — tokens stay sealed in the gateway, never extractable. Nothing to steal or replay.

2025 · PROMPT INJECTION

M365 Copilot

EchoLeak: one crafted email steered Copilot into leaking internal files — with zero clicks.

WiKey — every instruction is vetted and egress is guarded. It can't leak beyond policy.

2025 · AGENT WENT ROGUE

Replit agent

During a code freeze, the agent wiped a live production database — then misreported it.

WiKey — destructive, out-of-scope actions are blocked, and revocable in one click.

*Different attacks, one root cause — **the agent held the keys and the authority. WiKey removes both.***

ACCOUNT RECOVERY

Recovery is easily compromised, hijacking the agent's credentials.

Most attacks don't break authentication — they talk a help desk into resetting it.

THE PLAYBOOK

- Pass verification with breached data
- Claim a lost or broken phone
- Enroll an attacker-controlled MFA device
- AI voice & video beat the human check

MGM & Caesars (2023) · M&S & Co-op (2025). Mandiant: a help-desk call to domain-admin in under an hour.

WIKEY: ZERO-INFO RECOVERY

- Recovery by cryptographic attestation
- Trusted parties — installer, prior counterparty — sign with their own keys
- No password, no phone number, no help-desk reset
- A signature can't be social-engineered

*Keys held in the virtual HSM and never backed up — **no copy to restore, and no reset path to exploit.***

WIKEY

WHERE WE SIT

Trustless policy & gateway engines — zero point of failure.

A vendor share, a hardware enclave, or an off-chain policy server — each is a single point to compromise.

Dimension	Platform-native trust layers	Universal layer (WiKey)
Keys / credentials	Stored in tokens, vaults & backups — stealable	Post-quantum vHSM; keys never backed up
Account recovery	Help-desk / reset bypasses phishing-resistant MFA	Attestation by trusted parties — no help desk
Traceability	Often a service account; logs siloed per vendor	Every action traces to the agent's owner
Coverage	No single vendor reaches every source	Consistent everywhere, incl. external MCP
Identity across hops	Re-issued and re-translated at each boundary	One inherited identity across systems
Neutrality	A platform won't be neutral toward rivals	Vendor-neutral by design

Salesforce won't secure your AWS agents; Microsoft won't be neutral toward Google. ***Universality isn't a feature they're behind on — their business model precludes it.***

WIKEY

TRACTION

Live pilots in custody, treasury and autonomous agents.

Three production deployments, three different markets — all running today.

PILOT 01

Florida family office

Replacing Anchorage Digital custody

Eliminates third-party counterparty risk while keeping institutional-grade controls.

PILOT 02

Spain-based fund

Replacing Safe (Gnosis) EVM multi-sig

Adds compliance controls, recovery, and cross-chain reach beyond EVM.

PILOT 03

Game studio

Securing autonomous in-game agents

AI NPCs hold and transact in-game assets — no per-agent key management for the studio.

Beta live · **Production seedless wallets in use today** · Browser + mobile

TEAM

Operators. Three prior exits in security & enterprise.

A team that has built, scaled, and sold security and enterprise software together.

Three prior exits across the team — building together in security and enterprise software.

WIKEY

No keys to leak.

No backups to breach.

No agent to phish.

Just protection — at any scale.

If you're investing in the rails of the agentic world, we should talk.

info@wikey.io · **wikey.io**