Safetronic AI Trust Gateway

Human-Authorised Trust for Enterprise AI Knowledge

Key Benefits

- **Human-Authorised AI Knowledge**
  AI systems reason only over explicitly approved material
- **Deterministic Protection Against Knowledge Corruption**
  Unauthorised content is excluded by design
- **Cryptographic Provenance and Auditability**
  Every knowledge decision is provable after the fact
- **Trusted AI Answers**
  Outputs are backed by a verifiable chain of trust
- **Clear Separation of Duties**
  Source ownership and AI knowledge admission are governed independently
- **Built on the Proven Safetronic Platform**
  PKI, HSMs, Mobile MFA, and audit trails trusted in regulated industries

Safetronic AI Trust Gateway is a trust enforcement capability for Retrieval-Augmented Generation (RAG) systems in regulated enterprise environments. Built on the Safetronic platform's proven trust model, leveraging PKI, remote authorisation, and HSM capabilities to extend trust into AI knowledge systems.

It ensures that AI systems reason only over knowledge that has been explicitly authorised by accountable individuals, and that every AI answer can be traced back to a verifiable chain of trust.

Safetronic AI Trust Gateway governs knowledge admission and was designed for sovereign, regulated environments where accountability, auditability, and non-repudiation are mandatory.

What Is Safetronic AI Trust Gateway

Safetronic AI Trust Gateway enforces explicit trust controls at the point where knowledge is admitted into an AI system, rather than relying on inference, relevance, or runtime monitoring.

It ensures that:

- AI systems reason only over explicitly authorised knowledge
- Every AI answer is traceable to approved source material
- Knowledge integrity is preserved across transformation and reuse
- Tamper-evident audit evidence is retained for governance and review

Why AI Knowledge Requires Governance

RAG systems are rapidly becoming a foundation of enterprise AI. They allow large language models to generate answers grounded in internal documentation, policies, and operational records. RAG systems transform documents into operational inputs for automated reasoning.

Once admitted into a knowledge base:

- Content is treated as authoritative
- Plausible but unapproved material can influence outcomes
- Traditional access control and encryption are insufficient

In regulated environments, organisations must be able to prove:

- Who authorised the knowledge
- When it was approved
- That it was not altered
- That only approved material influenced AI outputs

Safetronic AI Trust Gateway addresses this requirement directly.

Safetronic Approach

Safetronic AI Trust Gateway applies proven principles from payments, identity, and cryptographic governance to AI knowledge.

- ***Trust is explicit, not inferred.*** Knowledge becomes trusted only when an authorised individual approves its use.
- ***Authority is cryptographically enforced.*** Every approval is recorded as a PKI-backed signing event protected by hardware security modules (HSMs).
- ***Verification is deterministic.*** Only knowledge with a valid, continuous chain of trust is eligible to influence AI outputs.

This approach allows AI systems to remain fast and scalable, while ensuring that governance is enforced before reasoning occurs.

Trust Roles

Safetronic AI Trust Gateway introduces two complementary trust roles:

Knowledge Custodians

- Custodians are responsible for the integrity of original source material. They may represent internal departments, external partners, or supply-chain contributors.
- Custodians approve source datasets for downstream use and authorise cryptographic signing to establish provenance and integrity at origin. This creates a durable root of trust that persists beyond format changes or processing.

RAG Curator

- The RAG Curator governs the AI knowledge base. This role decides which authorised material is admitted into the knowledge domain used for retrieval and reasoning.
- The Curator verifies Custodian-approved material, authorises signing of transformed representations, and controls what knowledge elements are eligible for retrieval.

Together, these roles enforce clear separation of duties and preserve accountability across the AI knowledge lifecycle. This mirrors established enterprise governance models used in payments and identity systems.

Human-Authorised Signing

All trust decisions are enforced through ***explicit human approval***.

- Approval requests are delivered via Safetronic mobile MFA
- Private keys remain protected in HSMs
- Each event produces durable forensic evidence

Signing events are used to:

- Authorise source datasets
- Authorise AI-ready knowledge
- Preserve trust across transformation and reuse

Safetronic AI Trust Gateway is designed for organisations with strong governance requirements, including: Financial services, Government, Critical infrastructure, and Regulated enterprises. It allows AI systems to inherit the same trust standards already applied to other critical platforms.

Governance Without Friction

Safetronic AI Trust Gateway enforces governance without slowing AI systems down.

Human approval occurs only at defined trust boundaries. During retrieval and answer generation, the system performs verification only. There is no discretionary judgement, behavioural inference, or runtime policy evaluation.

This allows organisations to apply strong governance while preserving the performance and scalability expected of modern AI systems.

Safetronic AI Trust Gateway enforces:

- ***Provenance.*** Every knowledge element is traceable to its source and approving authority
- ***Authorisation.*** Only explicitly approved knowledge is eligible for AI use
- ***Non-Repudiation.*** Approval events are cryptographically bound to identities and time-stamped

Trusted Answers, Not Just References

Modern RAG systems often return document references alongside AI answers. While useful, references alone do not establish authority.

Safetronic AI Trust Gateway extends this model by attaching a cryptographic chain of trust to every referenced source. For each AI answer, organisations can verify:

- Which documents influenced the response
- Which system of record they originated from
- Who authorised the source material
- Who approved its inclusion into the AI knowledge base
- That the content was not altered after approval

AI outputs become defensible decisions, not just plausible responses.

Safetronic AI Trust Gateway operates alongside existing enterprise AI security and data protection controls, including runtime monitoring, data loss prevention, encryption, and confidential computing. It focuses specifically on governing knowledge authority: establishing who approved the information that AI systems are permitted to use, and preserving that authority through cryptographic proof.

Salt Group Pty Ltd
Level 30, 459 Collins Street
Melbourne VIC 3000
Australia

Australia, UK/EU & Asia Pacific
T: +61-3-9614-4416 (AU/AsiaPac)
T: +44-20-3966-1686 (UK/EU)
E: sales@saltgroup.com.au