

Datenströme in KI-Workflows absichern. Datenhoheit souverän wahren.

Semantische Firewall

Sensible Daten schützen, Prompt Injections blockieren und KI-Innovationen sicher skalieren

AI-Workflows verarbeiten Geschäftsdaten autonom. Dabei werden Daten aus verschiedenen Quellen kombiniert, deren Vertraulichkeit erst zur Laufzeit bestimmbar ist. Ohne architektonisch verankerten Mechanismus gibt es keine Stelle, die Vertraulichkeit prüft und durchsetzt; weder beim Ausgang der Daten, noch beim Eingang der Antwort, noch bei der Weitergabe zwischen Anwendungen. Die Verantwortung liegt implizit bei Entwicklern und

Anwendern — ein fehleranfälliger, nicht skalierbarer Ansatz. Klassische Schutzmechanismen wie IAM, DLP, Verschlüsselung, etc. prüfen Metadaten und Zugriffsrechte, nicht die inhaltliche Vertraulichkeitsstufe eines Payloads zur Laufzeit. Wir schließen diese Lücke mit einer Toolbox zur semantischen Klassifikation der Datenströme und zur automatischen Durchsetzung von Policies.

Herausforderungen

- **Unkontrollierter Datenabfluss:** KI-Agenten und automatisierte Workflows leiten Daten ohne inhaltliche Prüfung an externe LLMs weiter. Ohne semantische Klassifikation bleibt unklar, welche vertraulichen Informationen das eigene Netzwerk verlassen.
- **Berechtigungsdiffusion:** KI-Anwendungen kombinieren Daten aus unterschiedlichen Quellen und können dadurch vertrauliche Informationen unbefugten Nutzern offenlegen. Klassische rollenbasierte Zugriffskonzepte berücksichtigen den Inhalt der Antworten nicht.
- **Prompt Injection und Datenexfiltration:** Manipulierte Inhalte in Datenquellen können KI-Agenten zur Preisgabe vertraulicher Informationen veranlassen. Klassische Sicherheitsmechanismen erkennen diesen Angriffsvektor nicht.
- **Unkontrollierter Mosaikeffekt:** Die Kombination unkritischer Informationen kann einen hochvertraulichen Gesamtkontext erzeugen, ohne dass dies erkannt wird.

Ganzheitliche Lösung

- **Semantische Datenkontrolle:** Datenströme werden zur Laufzeit inhaltsbasiert klassifiziert und anhand definierter Sicherheitsrichtlinien automatisch maskiert oder blockiert – noch bevor Payloads das Unternehmensnetzwerk verlassen.
- **Laufzeitprüfung und Zugriffsschutz:** Die semantische Bewertung erfolgt im jeweiligen Benutzerkontext und verhindert zuverlässig unberechtigte Dateneinsichten.
- **Schutz vor Prompt Injection:** Eingehende Modell-Responses werden auf eingeschleuste Anweisungen geprüft und kritische Inhalte blockiert, bevor sie Agenten oder Nutzer erreichen.
- **Intelligentes Routing:** Abhängig von der Klassifikation werden Anfragen automatisiert an geeignete Modelle und Ausführungsumgebungen (On-Premises, Private oder Public Cloud) weitergeleitet.
- **Revisionssicherheit:** Sämtliche Klassifikationsentscheidungen werden nachvollziehbar und revisions-sicher protokolliert.

Business Value

- ✓ **Innovationsfähigkeit und Risikominimierung:** Die Organisation wird zur sicheren Nutzung von KI- und Agententechnologien befähigt. Gleichzeitig minimiert eine präventive Systemarchitektur das Risiko kritischer Datenlecks nachhaltig.
- ✓ **Datenhoheit:** Die semantische Prüfung der Datenströme erfolgt vollständig auf der eigenen

Infrastruktur (On-Premises oder in einer Virtual Private Cloud). Dadurch ist sichergestellt, dass sensible Daten den Vertrauensbereich des Unternehmens nicht verlassen.

- ✓ **Nachweisbarkeit:** Jede sicherheitsrelevante Entscheidung wird lückenlos protokolliert und ist jederzeit auditierbar.

Warum S&N?

Leistungen & Expertise

Know-how

- **Praxiserprobte Toolbox** zur semantischen Klassifikation und zum semantischen Routing von Datenströmen mit praxistauglicher Latenz sowie nachgewiesener Trefferquote und Datenabschirmung.
- **Langjährige Erfahrung** in der Konzeption und Umsetzung von Sicherheitsarchitekturen für komplexe Softwaresysteme.
- **Branchenexpertise und Domänenwissen** sowie tiefgreifendes Verständnis regulatorischer Anforderungen und Sicherheitsstandards im Banken- und Behördenumfeld.

Unser Angebot

- **Beratung:** Analyse der bestehenden KI-Architektur, Identifikation kritischer Datenflüsse und Empfehlung einer souveränen Infrastruktur für die semantische Firewall.
- **Entwicklung:** Implementierung und kundenspezifische Anpassung von Klassifizierungspipelines.
- **Integration:** Anbindung an bestehende KI-Plattformen, Gateway-Infrastrukturen und Workflow-Engines – On-Premises oder in einer dedizierten VPC.
- **Migration:** Schrittweise Integration in bestehende Sicherheitsarchitekturen sowie bei Bedarf die Überführung bestehender DLP-Systeme in inhaltsbasierte Filtermechanismen.
- **Service:** Kontinuierliche Wartung sowie Erweiterung um zusätzliche Klassifikationsmodule.

Ansprechpartner



Maximilian Blume

Maximilian.Blume@sn-invent.de