

Die 5-Schichten-Architektur des Legal Agent Data Harness: Maximale Datensicherheit und Halluzinationsfreiheit für die Rechtspraxis

Kommunikation der technischen Struktur und der Sicherheitsmechanismen des Legal Agent Data Harness, mit besonderem Augenmerk auf "Data Readiness" und der Eliminierung von RT-Halluzinationen in juristischen Workflows. Der Legal Agent Data Harness entkoppelt die logische Schlussfolgerung von der Infrastruktur, wobei das LLM als Engine für Sprachverständnis und Synthese fungiert. Der Harness kontrolliert Informationsfluss, Kontext und Verifizierung, um zustandslose Medialien in zuverlässige juristische Agenten zu transformieren und die Einhaltung des Berufsgeheimnisses gemäß § 203 StGB zu gewährleisten.

Schicht 1: Ingestion & Privacy Guard

Schutz des Berufsgeheimnisses (§ 203 StGB)

Bevor Daten verarbeitet werden, bereinigt ein PII-Scrubber alle personenbezogenen Informationen und stellt über Ethical Walls sicher, dass Mandantenkontexte strikt isoliert bleiben.



Tenant & Conflict Check

Automatische Auflösung von Mandatskontexten zur Vermeidung von Interessenkollisionen bereits auf Ingestion-Ebene.

Schicht 2: Multi-Agent Orchestration & Refusal Ladder

Die 6-stufige Gate-Agenten-Prüfung

Eine "Refusal Ladder" (Verweigerungsleiter) nutzt sechs parallele Prüfinstanzen, um unbegründete oder falsche Antworten sofort zu blockieren, bevor sie den Nutzer erreichen.

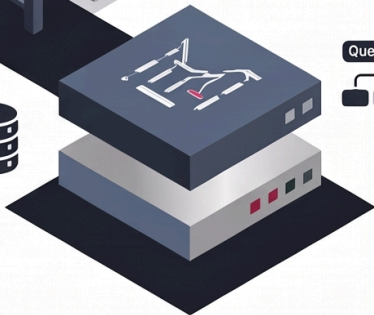


Anti-Context Rot & State Machine

Eine Durable State Machine verhindert "Context Rot", indem sie Arbeitsgedächtnis bereinigt und Zwischenschritte in Langzeitspeicher auslagert, um die Kohärenz über längere Aufgaben zu wahren.

Hybride Datenhaltung für Data Readiness

Trennung in eine Normative Ebene und eine Operative Ebene für Data Readiness.



Query Deconstruction & Semantic Infill

Anfragen werden in strukturierte Filter (Metadaten) und semantische Pfade zerlegt; die Vektorsuche wird nur zur qualitativen Textanreicherung über bereits gefilterten Daten genutzt.

Schicht 3: Hybride Retrieval Engine

Hybride Datenhaltung für Data Readiness

Normative Ebene (RDF/OWL)

Trennung in eine Normative Ebene und eine Operative Ebene für Data Readiness.



Operative Ebene (Neo4j Property Graph)

Mandatsdaten

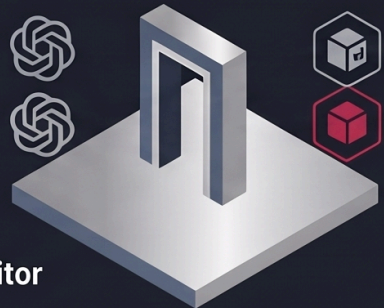
Query Deconstruction & Semantic Infill

Anfragen werden in strukturierte Filter (Metadaten) und semantische Pfade zerlegt; die Vektorsuche wird nur zur qualitativen Textanreicherung über bereits gefilterten Daten genutzt.

Schicht 4: Universal Inference Gateway & MCP-Sandboxes

Modell-Agnostik ohne Vendor-Lock-in

Das Gateway (LiteLLM/vLLM) ermöglicht den flexiblen Wechsel zwischen Modellen je nach Sensibilität der Daten.



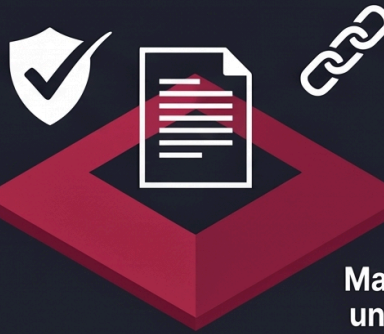
Sichere Werkzeugnutzung via MCP

Tool-Aufrufe erfolgen über das Model Context Protocol (MCP) in isolierten Sandboxes, wodurch das Risiko von schädlichem Code oder unbefugtem Systemzugriff minimiert wird.

Schicht 5: Citation Auditor & Audit Ledger

Garantierte Halluzinationsfreiheit

Ein Citation Auditor vergleicht jede Behauptung deterministisch mit den Knowledge-Graph-Knoten, bevor die Antwort gestreamt wird.



Immutable Audit Ledger

Jeder Schritt, jeder Abruf und jede Entscheidung wird manipulations-sicher protokolliert, um die Einhaltung des Anwaltsprivilegs und Compliance-Vorgaben nachzuweisen.

Maximale Datensicherheit und Halluzinationsfreiheit