

LEAKMAP AI

Evidence-Backed Jurisdictional Exposure Map for AI Prompts

Before you prompt, know where your data may go. LeakMap AI helps users and organizations understand possible jurisdictional exposure when prompts are routed through AI providers, cloud infrastructure, policy disclosures, subprocessors, and local sovereign alternatives.

METHODOLOGY NOTICE

LeakMap AI does not claim to observe hidden internal processing paths inside proprietary AI systems. It separates verified endpoint telemetry, disclosed provider documentation, inferred jurisdictional risk, and explicitly unknown internal paths.

Project Title	LeakMap AI: Evidence-Backed Jurisdictional Exposure Map for AI Prompts
Track	Governance and Geopolitics: AI compliance, AI sovereignty, national data security, compute diffusion and access
Primary Users	Students, startups, enterprises, public-sector teams, AI builders, data-protection reviewers
Core Output	Risk score, route exposure model, redaction recommendation, evidence packet, AI Data Passport
Team Details	Add your team name, member names, college, and contact before final submission

1. Executive Summary

LeakMap AI is an AI governance and sovereignty audit platform that analyzes a prompt before it is sent to an AI provider. The system detects sensitive content, shows an evidence-backed jurisdictional exposure model, attaches confidence labels and source references to every route, and recommends safer handling options such as redaction or local sovereign inference.

The project addresses a real gap: users often type personal, medical, financial, government-ID, or confidential business data into AI tools without understanding provider mode, data residency, retention, or subprocessor exposure. LeakMap does not pretend to know exact physical packet paths. Instead, it provides a defensible exposure model backed by runtime endpoint evidence, official provider disclosures, and marked uncertainty.

ONE-LINE PITCH

LeakMap AI turns hidden AI data-flow risk into a visual, QR-verifiable, evidence-backed exposure map before the user sends a prompt.

2. Problem Statement

- AI users cannot easily see what type of data they are sending to external AI systems.
- Privacy policies, API terms, data processing addenda, retention documents, and subprocessor lists are difficult to read and scattered across providers.
- Most dashboards either overclaim exact routing or provide generic compliance checklists without visual evidence.
- Organizations need a tool that honestly separates what is verified, what is disclosed, what is inferred, and what remains unknown.

3. Core Methodology

LeakMap AI uses a three-layer exposure model instead of claiming exact physical routing:

Layer	Meaning	Example	Confidence
Verified Endpoint	Runtime or app-side evidence such as provider selected, endpoint configured, timestamp, live/demo mode.	LeakMap backend attempted a Gemini endpoint request.	80-95%
Disclosed Exposure	Official provider documents such as API terms, data controls, DPA, retention docs, or subprocessor lists.	OpenAI subprocessor list or Claude API retention docs.	50-85%
Inferred Risk	Risk estimated from missing routing clarity, sensitive prompt class, external provider use, or policy	Data residency unknown for this exact prompt.	20-60%

Evidence-backed jurisdictional exposure map - not a physical packet trace

	ambiguity.		
Unknown Internal Path	Hidden internal model routing, data center, GPU, model worker, or subprocessor use for a specific prompt.	Exact server path inside provider.	0-30%

4. Key Features

Prompt X-Ray Scanner

Detects health data, government IDs, phone numbers, emails, financial data, business secrets, API keys, and other sensitive prompt elements.

Jurisdictional Exposure Globe

Shows provider endpoint profiles, disclosed processor exposure, inferred risk, unknown internal path, and local sovereign routes with visual confidence levels.

Evidence Vault and QR Packets

Every route or warning can link to an evidence packet containing source IDs, rule IDs, confidence, proof limits, and official source links.

Redaction Firewall

Masks sensitive entities before sending prompts to external providers and shows before/after risk reduction.

Local Sovereign Mode

Routes prompts to a local model such as Ollama on localhost, avoiding external AI API calls by default.

AI Data Passport

Generates a portable report containing risk score, route profile, evidence sources, triggered rules, redaction status, and final recommendation.

Judge Demo Mode

Runs a complete demonstration: critical prompt detection, route visualization, evidence drawer, QR verification, redaction, local mode, and passport generation.

5. System Architecture

Module	Technology	Role
Frontend	Next.js, TypeScript, Tailwind CSS, Framer Motion, GSAP	Brutalist scanner dashboard, map views, evidence drawer, passport pages
Map Layer	D3/React Three Fiber, route animation, SVG/Canvas/WebGL	Exposure globe, route classes, confidence labels, data-packet animation

Backend	Next.js API routes or Firebase Functions	Prompt analysis, provider profiles, evidence packet generation, API proxy
Storage	Firebase Firestore	Scan metadata, evidence packets, source registry, rule book, passport records
AI Providers	Gemini API optional, OpenAI optional, Claude evidence-only default	Live/demo model response and provider exposure profile
Local Mode	Ollama localhost endpoint	Sovereign local inference without external API key
Evidence System	Source registry, rule book, confidence engine, QR generator	Proof chain behind every route and warning

6. Risk Scoring and Rule Engine

Risk is computed from prompt sensitivity, external provider exposure, data residency clarity, subprocessor visibility, retention/training ambiguity, human review warnings, evidence confidence, and redaction status.

Rule ID	Condition	Effect
R-101	Health data detected	+25 risk
R-102	Government ID or Aadhaar-like ID detected	+30 risk
R-103	Phone/email/personal identifier detected	+10 risk
R-104	Business confidential or API secret detected	+20 risk
R-105	External provider selected	+15 risk
R-109	Redaction applied	-30 to -60 sensitivity risk
R-201	Verified endpoint rule	solid route; does not prove internal provider path
R-203	Unknown internal path rule	always show unknown for black-box providers
R-206	Evidence missing cap	confidence capped at 30%

7. Privacy and Security Design

- Raw prompt is not stored by default. The system stores prompt hash, redacted prompt, provider mode, risk score, and evidence IDs.
- Local Sovereign Mode uses Ollama/localhost and does not require a paid external AI API key.
- External provider modes are clearly labeled live, demo, optional, or evidence-only.

Evidence-backed jurisdictional exposure map - not a physical packet trace

- QR codes open evidence packets or source registry pages, not unsupported claims.
- Every pathway includes limitations: exact internal provider routing remains unknown unless the provider explicitly exposes it under contract/configuration.

8. Innovation and Track Relevance

LeakMap AI fits the governance and geopolitics track because it connects AI compliance, data sovereignty, national data security, compute access, and foreign provider dependency into a single interactive product. Instead of only detecting harmful output, it audits the infrastructure and policy exposure created before a prompt is sent.

Track Theme	LeakMap AI Contribution
AI Compliance	Evidence ledger, source registry, risk rules, data-passport output
AI Sovereignty	Local sovereign mode, foreign provider exposure comparison
National Data Security	Health/Gov ID/business secret prompt warnings and redaction
Geopolitics	Jurisdictional exposure model and provider dependency visualization
Compute Access	Local model route and external API dependency comparison

9. Demo Scenario

1. User selects Gemini Live/Demo mode.
2. User enters: "Summarize my father's diabetes report and Aadhaar-linked hospital file."
3. Prompt X-Ray detects HEALTH_DATA, GOV_ID, PERSONAL_DATA.
4. Risk score rises to Critical, for example 94/100.
5. Globe shows a verified endpoint route, disclosed provider exposure, and unknown internal path limitation.
6. Evidence drawer shows source IDs and QR-verifiable evidence packet.
7. Redaction Firewall replaces sensitive fields with [HEALTH_DATA], [GOV_ID], and [PERSON].
8. Risk drops, for example 94 to 41.
9. Local Sovereign Mode is shown as a safer route for sensitive prompts.
10. AI Data Passport is generated for upload/audit.

10. Limitations and Responsible Claims

- LeakMap AI is not a packet sniffer for proprietary AI systems.
- It cannot prove exact data center, server, model worker, GPU, internal failover, or subprocessor path for a specific prompt.

Evidence-backed jurisdictional exposure map - not a physical packet trace

- Official subprocessor documents show possible processor exposure, not proof that every listed subprocessor touched a given prompt.
- Demo mode routes are simulations and must be labeled as such.
- Confidence levels represent evidence strength, not absolute certainty.

11. Future Scope

- Browser extension for detecting sensitive prompts before submission into AI websites.
- Enterprise policy mode where organizations define custom AI routing rules.
- DPDP-focused compliance templates and automated audit exports.
- Real DNS/ASN endpoint telemetry for live API calls where technically feasible.
- Organization-level evidence vault with versioned source reviews.
- Local LLM benchmarking to compare privacy, quality, latency, and cost.

12. Conclusion

LeakMap AI is designed to make AI data exposure visible, honest, and actionable. It does not rely on unsupported claims about hidden provider infrastructure. Instead, it brings together prompt sensitivity scanning, evidence-backed exposure modeling, official source verification, redaction, local sovereign inference, and AI Data Passports into a practical AI governance product.

13. Official References

The following official sources support the provider policy, data-control, subprocessor, and retention claims used in the methodology. They do not prove exact per-prompt physical routing.

- [1] Google Gemini API Additional Terms of Service: <https://ai.google.dev/gemini-api/terms>
- [2] Google Cloud Platform Subprocessors: <https://cloud.google.com/terms/subprocessors>
- [3] OpenAI API Data Controls / Your Data: <https://developers.openai.com/api/docs/guides/your-data>
- [4] OpenAI Sub-processor List: <https://openai.com/policies/sub-processor-list/>
- [5] OpenAI Data Processing Addendum: <https://openai.com/policies/data-processing-addendum/>
- [6] Claude API and Data Retention: <https://platform.claude.com/docs/en/manage-claude/api-and-data-retention>
- [7] Anthropic Subprocessors / Trust Center: <https://trust.anthropic.com/subprocessors>