Datasaur

Private LLMs: Definition, Spectrum, and a Buyer's Framework

Sept 2025 | Whitepaper

01 Introduction

A Private LLM enforces your organization's privacy and compliance end-to-end while being customized to your workflows and optimized for quality, latency, and cost—with clear auditability and vendor-independent control.

02 Why "Private LLM" Needs a Precise Definition

As the term spreads, its meaning splinters: for some vendors it "runs in your VPC," for others "we don't train on your data," and for regulated industries it may require air-gapped, bare-metal deployments with auditable logs. This ambiguity slows deals and increases risk.

Datasaur.ai takes a stricter view: a Private LLM is defined by the **controls you enforce** and the **outcomes you measure**.

03 Working Definition

A **Private LLM** is an AI system whose **data handling, deployment, and model-use policies** are engineered to an organization's privacy, security, and regulatory requirements **and** tuned to that organization's workflows and goals across **quality, cost, and latency.**

Two pillars:

1. **Meets your privacy requirements.** Privacy is a contract, not a claim. The system dictates where data lives, who can access it, what's logged, and retention windows, and whether any usage can train models outside your control (answer: *no*).
2. **Tailored to your workflow.** Not just a chatbot. A Private LLM combines model(s), prompts, retrieval, tools, guardrails, and integrations to deliver measurable outcomes in your environment.

Miss one pillar and you don't have a Private LLM—you have either a security liability or a generic tool that won't deliver.

04 “Private” Is a Spectrum

Every Datasaur engagement starts by pinning down your definition of “private.” Typical requirements:

- **Data Sovereignty**: Data, embeddings, and logs remain in-region and in your cloud or on-prem.
- **Zero Data Retention**: No request/response bodies or metadata persist—by vendor or subprocessor.
- **No Training on Our Data**: Usage does not improve shared or external models.
- **Model Sovereignty**: Control over model weights (open-weight or licensed) and deployability in your VPC or on bare metal.
- **Regulatory Alignment**: Architecture and documentation that support SOC 2, HIPAA, GDPR, and domain-specific retention/discovery policies.
- **Auditability & Forensics**: Immutable logs, replay, access controls, incident workflows.

We translate these into **non-negotiables** and **nice-to-haves**, then design to that line.

05 A Practical Decision Framework

Privacy is necessary; usability is decisive. We evaluate across five axes:

1. **Quality**: Accuracy, grounding, citation rates, and adverse events.
2. **Latency**: End-to-end time including retrieval, tools, post-processing.
3. **Cost**: Per-request unit economics and total cost of ownership.
4. **Sovereignty**: Control over data, runtime, and model weights.
5. **Scalability**: Throughput, concurrency, autoscaling, capacity planning.

Model Strategy: SLM, LLM, or Both

- **SLMs**: Great for narrow, high-volume tasks (classification, extraction, templated generation) under tight latency/cost.
- **Frontier LLMs**: Use when accuracy is paramount (complex reasoning, ambiguous inputs, high-risk outputs).
- **Ensembles**: Common in production. Let an SLM handle the majority; escalate the “long tail” to a larger model.

Deployment Patterns

- **VPC-Hosted (AWS/Azure/GCP):** Your network, your KMS, private endpoints.
- **On-Prem / Bare Metal:** Air-gapped for ultra-sensitive workloads.
- **SaaS with Safeguards:** Zero-retention, no-training guarantees, IP allow-listing, rigorous DPAs.

06 Why Organizations Choose Datasaur

Fortune 50 clients and SMBs alike work with Datasaur to build a solution that aligns with a rapidly evolving set of regulatory best practices while training an LLM on years of internal knowledge and expertise. Datasaur works hand-in-hand with your subject matter experts to guarantee the following:

- **Privacy-first by design:** Start from your definition of "private," engineer backward—VPC, zero retention, or on-prem.
- **Workflow outcomes, not demos:** Contract review accelerators, claims extractors, KYC summarizers—systems tied to metrics.
- **Model independence:** Open and proprietary options, routed dynamically to hit accuracy/latency/cost targets.
- **Operational maturity:** Prompt/version governance, eval harnesses, audit-ready logging, incident runbooks, performance dashboards—standard, not extras.

07 Where to Begin

Regulated or sensitive environment? Align privacy and ROI on day one. Datasaur will interview stakeholders, codify business requirements, and deliver an architecture and roadmap your security, legal, and business teams can approve.

Your data. Your rules. Your results. That's a Private LLM.

Learn more about Private LLMs with Datasaur.

About Datasaur

Datasaur is a private LLM provider and data labeling platform designed for companies to build their AI ecosystem with ease and efficiency. It assists organizations and universities in setting up custom LLMs and annotating data more efficiently and accurately through automation, quality control, and human-in-the-loop workflows. For more information, visit www.datasaur.ai.

Schedule a demo