



ISO 9001

NASSCOM

Clutch Top B2B

DPIIT

ENTERPRISE FIELD GUIDE · 2026

# Your Own Private AI Lab. Your Data. Your Control.

We design, deploy, and operationalize private AI Labs for enterprises — complete with GPU infrastructure, local LLMs, data pipelines, and AI workflow orchestration. Zero cloud dependency. 100% data sovereignty. Full enterprise control.

**On-Premise** LLMs · Mistral, Llama 3, Gemma

GPU Infrastructure · **NVIDIA A100/H100**

**Zero** Data Leakage

ROI in **6-12 Months**

**Book AI Labs Consultation →**

● **12-18 min read · 7 sections**

# What you'll walk away with

- 1 The Case for a Private AI Lab** p.03  
Why the cloud-API default is becoming the most expensive and least defensible choice.
- 2 What We Build in Your AI Lab** p.04  
Six integrated layers — from GPU hardware to analytics — forming a complete enterprise AI platform.
- 3 Industry-Specific AI Lab Solutions** p.05  
Pre-built use-case libraries and model configurations for Oil & Gas, Manufacturing, and Healthcare.
- 4 Common Pitfalls & How We Solve Them** p.07  
The failure modes that sink in-house AI efforts — and how a structured build avoids each trap.
- 5 20-Week AI Lab Setup Timeline + Next Steps** p.08  
From kickoff to a fully operational AI Lab — and how to start.

## WHO THIS IS FOR

CTOs, Heads of IT, CISOs, and digital-transformation leaders at enterprises that handle **sensitive, regulated, or competitively critical data** — and who can't afford to send it to a third-party cloud API. If **data sovereignty, predictable AI costs, and zero vendor lock-in** matter to you, this guide is built for you.

## Ready to explore a private AI Lab?

Book a free AI Lab design consultation. We'll assess your infrastructure and define the architecture.

[Book AI Labs Consultation →](#)

[Explore AI Labs](#)

# The case for a **Private AI Lab**

In the last 18 months the question stopped being "should we use AI?" and became "where should our AI actually live?" For data-sensitive enterprises, the cloud-API default is quietly becoming the most expensive — and least defensible — choice.

## 100%

Data sovereignty. With on-premise or air-gapped deployment, sensitive prompts and documents never leave your perimeter.

## 80%

Lower AI OpEx vs cloud. A private lab cuts ongoing AI operating costs versus per-token cloud API billing.

## 20 wks

Lab live timeline. A full lab — GPU infrastructure, local LLMs, RAG pipelines and orchestration — goes operational.

### MYTH VS REALITY

#### MYTH

"On-prem AI is too expensive to justify."

#### REALITY

Per-call API pricing scales with usage forever. A lab is a **one-time capital base** that pays for itself in 6-12 months. The cloud bill is the real expense.

#### MYTH

"You need a research team to run local LLMs."

#### REALITY

Open models on **Ollama/vLLM** plus a 90-day hypercare handover means your existing team operates the lab independently.

#### MYTH

"Open-source models can't match the big APIs."

#### REALITY

Mistral, Llama 3, and Gemma, **fine-tuned on your data**, beat generic frontier APIs on the tasks that actually matter to you.

[Talk to Our Team →](#)

# What We Build in Your AI Lab

Six integrated layers — from hardware to analytics — forming a complete enterprise AI platform.

## LAYER 1

### Edge AI Infrastructure

On-premise GPU servers (NVIDIA A100/H100) or edge-optimized hardware for local AI inference — no cloud dependency.

- > NVIDIA GPU server configuration
- > Edge inference optimization
- > Thermal & power management
- > Redundancy & failover design

## LAYER 2

### Private LLM Deployment

Deploy and fine-tune open-source LLMs (Mistral, Llama 3, Gemma) entirely within your infrastructure.

- > Ollama / vLLM deployment
- > Domain-specific fine-tuning
- > Quantization for edge hardware
- > Model versioning & rollback

## LAYER 3

### Data Pipeline & Vector Store

Ingest, process, and embed your enterprise data for retrieval-augmented generation (RAG) and analytics.

- > ETL pipeline design
- > Vector database (Qdrant/Chroma)
- > Document ingestion & chunking
- > Real-time data sync

## LAYER 4

### AI Workflow Orchestration

N8N-based workflow automation connecting your AI models to enterprise systems and business processes.

- > N8N self-hosted deployment
- > Agent workflow design
- > API integration layer
- > Monitoring & alerting

## LAYER 5

### Security & Compliance

Enterprise-grade security controls ensuring your AI lab meets regulatory and data governance requirements.

- > RBAC & access management
- > Full audit trail
- > Data encryption at rest & transit
- > DPDP Act / GDPR compliance

## LAYER 6

### Analytics & Observability

Real-time dashboards tracking model performance, usage, cost, and business impact of your AI investments.

- > Model performance monitoring
- > Usage & cost analytics
- > Business KPI dashboards
- > Automated ROI reporting

## See all 6 layers in a live walkthrough

Book a consultation and we'll map these layers to your existing infrastructure.

[Book a Consultation →](#)

[AI Labs Overview](#)

# Industry-Specific AI Lab Solutions

Pre-built use-case libraries and model configurations for your industry.



OIL & GAS

## Oil & Gas AI Lab

### PREDICTIVE MAINTENANCE

Anomaly detection using RNNs and time-series analysis to predict equipment failures before they occur.

### RESERVOIR MANAGEMENT

CNNs for seismic data interpretation and reinforcement learning for optimal reservoir extraction strategies.

### DRILLING OPTIMIZATION

Real-time data analytics and reinforcement learning to optimize drilling parameters and reduce NPT.

### SAFETY & COMPLIANCE

Computer vision for PPE compliance monitoring and NLP for automated safety incident reporting.

Mistral 7B

Gemini Pro

NVIDIA GPU Servers

Custom CNNs



MANUFACTURING

## Manufacturing AI Lab

### QUALITY CONTROL VISION

Computer vision models for real-time defect detection on production lines, replacing manual inspection.

### SUPPLY CHAIN OPTIMIZATION

Demand forecasting and inventory optimization agents that reduce waste and prevent stockouts.

### PREDICTIVE MAINTENANCE

Sensor-driven anomaly detection for CNC machines, motors, and assembly lines to prevent unplanned downtime.

### PROCESS OPTIMIZATION

AI-driven parameter tuning for manufacturing processes, reducing energy consumption and improving yield.

[Get Your Industry AI Lab Proposal →](#)



HEALTHCARE

## Healthcare AI Lab

### MEDICAL DOCUMENT INTELLIGENCE

RAG-powered agents that process patient records, lab reports, and clinical notes with full data privacy.

### DIAGNOSTIC ASSISTANCE

Medical imaging analysis using custom CNNs for radiology, pathology, and dermatology screening.

### CLINICAL WORKFLOW AUTOMATION

AI agents that automate appointment scheduling, insurance verification, and discharge summaries.

### REGULATORY COMPLIANCE

Automated compliance monitoring for HIPAA, DPDP Act, and hospital accreditation standards.

# 350+

**Global clients. 25+ years of delivery.** Swaran Soft builds these AI Labs on a 100% open-source stack — Ollama, N8N, Qdrant, Mistral, Llama 3 — deployed entirely in *your* infrastructure. From Gurugram to the world. ISO 9001 Certified. NASSCOM Member. DPIIT Recognised.

## AI LABS FAQs

### What hardware do we need to set up an AI Lab?

It depends on your workload. We conduct a 2-week infrastructure assessment to define exact GPU/edge requirements before any procurement.

### How is a private AI Lab different from using cloud AI APIs?

A private lab gives you 100% data sovereignty, predictable costs (no per-token billing), zero vendor lock-in, and the ability to fine-tune models on proprietary data.

### What ongoing support do you provide after the lab is set up?

Full documentation, team training, and 90-day hypercare support to ensure your team can operate independently. Optional managed AI services available.

## Need an AI Lab tailored to your industry?

We'll assess your infrastructure, define the architecture, and provide a detailed proposal within 5 business days.

[Get a Custom Proposal →](#)

[View Case Studies](#)

# Common pitfalls — and how we solve them

Most failed in-house AI efforts fail for the same handful of reasons. Here's what goes wrong, and how a structured build avoids each trap.

## ⚠️ Buying GPUs before knowing the workload

### WHAT GOES WRONG

Teams over- or under-spend on hardware because no one mapped the actual inference and training demand.

### HOW WE PREVENT IT

Every build opens with a 2-week infrastructure assessment that defines requirements before a single server is procured.

## ⚠️ Treating security as an afterthought

### WHAT GOES WRONG

A lab goes live without RBAC, audit trails or encryption — then fails its first compliance review.

### HOW WE PREVENT IT

A dedicated security & compliance layer (RBAC, full audit trail, encryption at rest and in transit, DPDP/GDPR alignment) is part of the core stack, not a bolt-on.

## ⚠️ Deploying a model with no path to improve it

### WHAT GOES WRONG

A fine-tuned model degrades or misbehaves and there's no way to compare versions or revert.

### HOW WE PREVENT IT

Model versioning & rollback plus performance monitoring are built in, so every change is measurable and reversible.

## ⚠️ An island that doesn't connect to the business

### WHAT GOES WRONG

A technically impressive lab delivers no value because it never connects to the systems people actually use.

### HOW WE PREVENT IT

An N8N orchestration and API integration layer wires the lab into enterprise systems, validated with a real production pilot before handover.

## ⚠️ No one left who can run it

### WHAT GOES WRONG

The build partner leaves and the internal team can't operate or extend the lab.

### HOW WE PREVENT IT

Full documentation, team training and 90-day hypercare ensure your team operates independently.

## De-risk your AI Lab build from day one

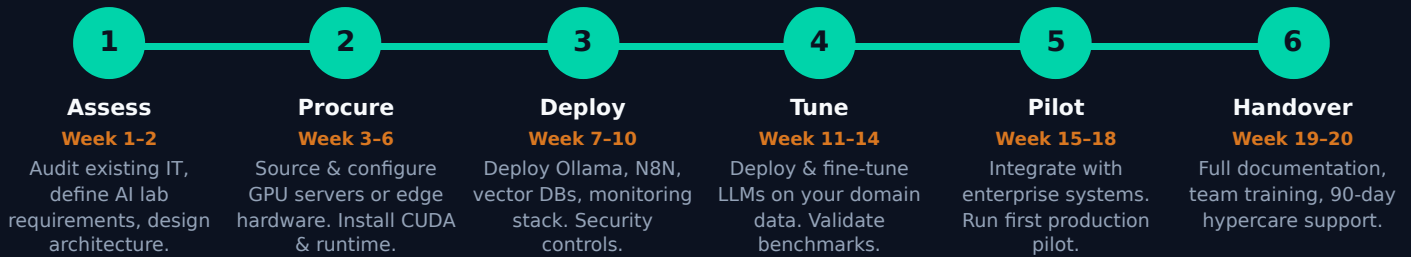
Our structured 20-week approach prevents every one of these pitfalls. Let's scope your lab.

[Start Your Lab Build →](#)

[AI Strategy & Roadmap](#)

# 20-Week AI Lab Setup Timeline

From kickoff to a fully operational AI Lab in 20 weeks.



"We design, deploy, and operationalize private AI Labs for enterprises — complete with GPU infrastructure, local LLMs, data pipelines, and AI workflow orchestration. Zero cloud dependency. 100% data sovereignty."

[Positioning drawn from the AI Labs page]

**Swaran Soft · India's Premier Agentic AI Enterprise Partner**



## Book a Consultation

A free AI Lab design consultation. We'll assess your infrastructure and define the architecture.

[Book AI Labs Consultation](#)



## Submit Your Use Case

Tell us the workflow you'd start with. We'll return a detailed proposal within 5 business days.

[Submit Your Use Case](#)



## Get Monthly Insights

Enterprise AI trends, real case studies and playbooks — curated for senior IT leaders.

[Read the Blog](#)

## Ready to Build Your Private AI Lab?

Book a free AI Lab design consultation — detailed proposal within 5 business days.

[Book Free Consultation →](#)

[Explore AI Labs](#)