Iterate.ai

WHITE PAPER · JUNE 2026

PRIVATE & SOVEREIGN AI

Owning intelligence, not renting it.

How a private AI platform differs from a shared model like Claude — and why the difference is a strategic asset, not a feature.

Intelligence Unshared: The AI Sovereignty Papers — this is the fourth in a six-paper series on Private AI. Find the series at **iterate.ai/resources/white-papers**

PREPARED BY

Iterate.ai

Office of the Chief AI Officer

PRODUCT

Generate

The private AI platform

EXECUTIVE SUMMARY

A shared model keeps your data private. A private model lets you own the intelligence.

Leaders often ask a fair question: “If a model like Anthropic’s Claude already promises to keep our data private, what problem is left for a private AI company to solve? They say our data isn’t used for training. Isn’t the current privacy pledge already enough?”

Here is the honest version. Under their enterprise terms, the leading providers genuinely protect your data — they do not train on it, they offer zero-data-retention, and they will sign a HIPAA Business Associate Agreement. On data privacy alone, they are strong. The distinction is about something else entirely: **ownership of the model itself.** With a shared API you rent access to one engine the vendor hosts for every customer at once. Your data may be private, but the engine is not yours — and the proprietary intelligence you could build from your own data never becomes an asset you own. And the ground is about to move again, as the field shifts toward models that **learn continually** rather than staying frozen after training — a change that rewrites the privacy assumptions everyone relies on today (we open with it in Section 1).

THE ONE-SENTENCE VERSION

A shared model keeps your data private. A private model lets you own the intelligence — the model, where it runs, and the compounding advantage you build on top of it.

WHAT THIS MEANS FOR A BUSINESS OPERATOR

Sovereignty	Sensitive workloads — patient data at a healthcare provider, deal data in financial services — stay inside your walls.
An owned asset	The model and the intelligence you fine-tune into it appreciate over time and belong to you, not a vendor.
One platform	The same governed system can serve healthcare, real estate, financial, and other businesses — each isolated.
HIPAA-ready	We can stand up a private healthcare instance designed for protected health information — simple enough for clinicians to use.

CONTENTS

What's inside.

The argument moves from a principle (own the model, not just protect the data) to the platform that delivers it, to the security and compliance that make it safe, to a concrete first step. Jump to the part your role cares about most.

PART ONE · THE CASE FOR PRIVATE AI 04–09

01 **The Foundation Is Shifting: Why a Frontier Model Is a Moving Target** 04
Vendors differ · posture changes on the vendor's schedule · continually-learning models · the reframe.

02 **The Question That Matters: Private vs. Shared AI** 06
What "private" actually means · where frontier APIs are strong · three postures · the reframe.

03 **The Value Proposition of Private AI** 08
Own the asset, not the output · sovereignty · economics at scale · "and," not "or."

PART TWO · INSIDE THE PLATFORM 10–11

04 **Inside Generate: The Platform** 10
Six layers, in plain terms · why the layering matters to an investor.

05 **Agentic Design Patterns** 11
How we build reliable agents · decomposition, isolation, grounding, verification.

PART THREE · SECURITY, COMPLIANCE & FIT 12–15

06 **Security and Governance** 12
Containment, not just trust · the AgentWatch gateway · 2025 incidents · every threat an architectural answer.

07 **HIPAA, Regulated Workloads & Healthcare Fit** 14
Why private deployment makes PHI straightforward · a fair note on the BAA.

08 **Usability: Does It Work for Real Business People?** 15
A chat app, not a science project · the pilot we recommend.

PART FOUR · GETTING STARTED 16–19

09 **Deployment Models & Getting Started** 16
Cloud, on-premises, hybrid · the recommended healthcare pilot.

10 **Direct Answers to Your Questions** 17
The five questions you asked · answered in one line each.

— **Appendix A & B** 18
Plain-language glossary · technical specifications at a glance.

01 THE FOUNDATION IS SHIFTING

A frontier model is a moving target.

Before comparing architectures, it is worth naming the real strategic risk — and it is not that any single model is insecure today. It is that building an enterprise on an external frontier model means anchoring to a foundation that **differs by vendor, changes over time, and is about to change paradigm.**

NO STANDARD

Vendors differ

There is no single "frontier model" standard to rely on. Anthropic has taken a strong posture — zero-data-retention, a signed BAA, and confidential computing whose design even assumes the operator could be adversarial. But that is **one company's choice, not a property of the category.** OpenAI, Google, and the next entrant each decide data handling, retention, and where inference runs differently.

THEIR SCHEDULE

Posture changes

Even a strong approach is not a fixed point. Prices change, usage policies are rewritten, and models are retired — vendors are candid that retiring older models is necessary to advance, because serving cost scales with the number of models kept available. **The model and terms you validate this year may not be the ones you run next year.**

THE EXPOSURE

When you standardize on "a frontier model," you do not get to choose the approach — **you inherit whichever vendor you are on,** and the approaches are not the same. A decision that looks safe against today's models can be quietly undone by a vendor's next policy, or by a shift in how models work.

THE PARADIGM ITSELF IS SHIFTING

The largest change is still ahead: models that learn continually.

Today's models are frozen after training, and the entire safety story — "we do not train on your data," zero-data-retention — is built around that fact. Research is now moving toward models that **learn continually as they run.** Google's recently published *Nested Learning* paradigm treats a model as nested optimization loops updating at different frequencies, paired with a continuum memory system and a self-modifying proof-of-concept aimed squarely at the catastrophic forgetting that keeps today's models static. It is early — a small-scale research result, not a production frontier model — but the direction is clear, and it changes the ground rules everyone is standing on.

WHEN A MODEL LEARNS CONTINUALLY, THREE MANAGEABLE PROBLEMS BECOME SERIOUS

01 INJECTION

Training-time injection

Poisoning that can steer a retrieval system today — by some measures, **five planted documents can swing a RAG system about 90% of the time** — stops being a per-answer nuisance and gets consolidated into the weights. Prompt injection becomes a way to *permanently teach* the model.

02 DRIFT

Drift & forgetting

A model that keeps updating can quietly change behavior — unacceptable for clinical or financial decisions — unless someone can **snapshot, audit, and roll it back.** On a model you do not control, you cannot.

03 OWNERSHIP

Ownership of what was learned

On a shared, multi-tenant model, whose improvement is it? A vendor cannot let your data improve the shared weights your competitor also uses — so the durable learning **accrues to the vendor** or is confined to a cache you do not keep.

THE REFRAME

The question is not "is this one model secure?" — a strong vendor may well be. It is whether an enterprise should anchor its most sensitive and most differentiating work to an external foundation that varies by vendor and shifts with each paradigm. Owning the model is how you stop renting someone else's changing decisions.

02 THE QUESTION THAT MATTERS

“Private” means two very different things.

The word gets used loosely, so it is worth separating the two things it can refer to. Most people hear “private AI” and assume the first. The opportunity — and the moat — is in the second.

CONTRACTUAL

Data privacy

Is the information you send to the AI kept confidential, not stored, and not used to train the vendor’s model? This is a **contractual and operational** property — and the leading vendors have made it strong.

THE MOAT

Model privacy

Is the model itself dedicated to you, running on infrastructure you control, owned as an asset? This is an **architectural and ownership** property — and it is where the durable advantage lives.

WHERE FRONTIER APIS ARE GENUINELY STRONG

Let us be fair and precise. Under their enterprise terms, the major providers do the following well:

- ✓ They **do not train** their models on your prompts or outputs.
- ✓ They offer **zero-data-retention** — inputs and outputs are not stored after a request is served.
- ✓ They will sign a **HIPAA Business Associate Agreement** for protected health information.

WHERE THE STRUCTURE STOPS

So when a vendor says “your data is safe,” they are largely right. But notice what that promise does *not* cover. The model is still one shared engine, hosted in the vendor’s cloud, serving thousands of other companies from the same weights.

You do not possess it. You cannot place it inside an air-gapped network. And because they do not learn from your data, you do not capture that learning either — the intelligence that could be distilled from your proprietary data never accrues to anyone as an owned asset.

THE HONEST SUMMARY

The answer is not that frontier vendors are insecure. On data privacy alone, they are strong. The distinction is about something else entirely — **ownership of the model itself.**

THE DISTINCTION IN ONE PICTURE

One engine for everyone, or one engine for you.

With a shared frontier API, you rent access to a single model the vendor hosts, controls, and operates for every customer at once. With a private model, the engine is dedicated to you and runs inside your own perimeter.

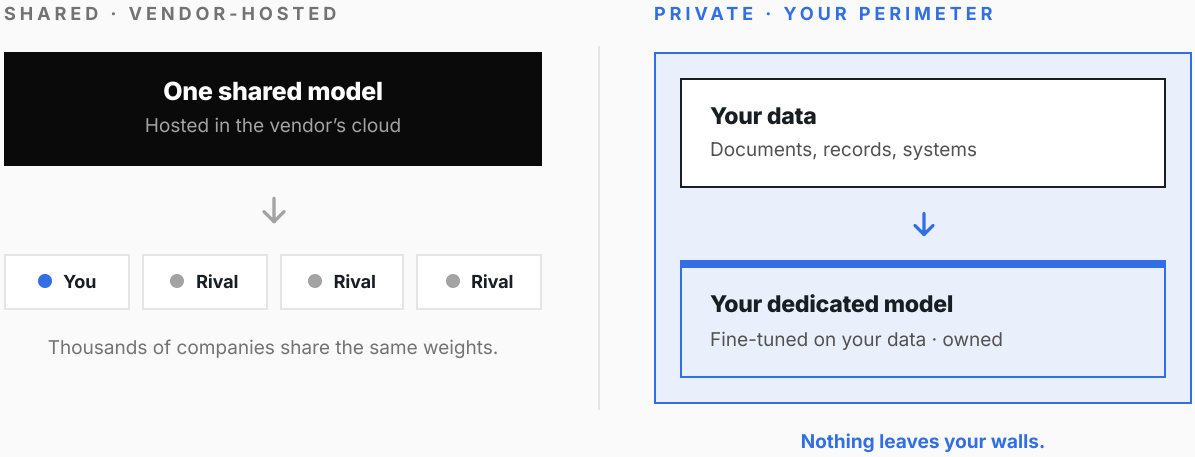


Figure 1 — *A shared model serves every customer from one vendor-hosted engine; a private model is dedicated to you and runs inside your own perimeter.*

THREE POSTURES — AND ONLY ONE GIVES YOU CONTROL

POSTURE	YOUR DATA	THE MODEL	THE HARDWARE
Public AI consumer ChatGPT, Gemini, raw API	Not controlled	Shared, external	External
"Private" via enterprise API	Private, but processed in the vendor cloud	Shared, external	External
Private AI Generate	**Private, in your perimeter**	**Yours, dedicated**	**Yours**

Most enterprise buyers reach the middle row and stop, assuming that is "private." The bar for private is not where your data sits — it is whether you also control the model and the hardware it runs on.

THE REFRAME

The question is not "is my data safe?" — with enterprise terms, it can be. The better question is: who owns the engine, and who owns the compounding advantage built on top of it?

03 THE VALUE PROPOSITION

Own the asset, not just the output.

When you call a shared API, you pay per use — forever — and you are left with the answer to one question. When you run a private model, two things change.

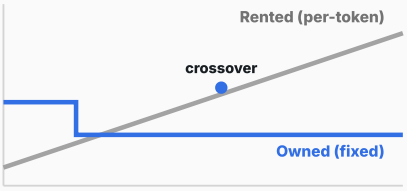
First, the model is a capital asset that sits on your balance sheet of capabilities. **Second,** every time you fine-tune it on your own data — your clinical protocols, your underwriting history, your operating playbooks — it gets better at *your* business specifically, and that improvement is yours to keep.

This is the difference between renting a capability and owning an appreciating asset. A rented capability is the same for you as it is for your competitor down the street; an owned model encodes advantages no competitor can buy off a price list.

Rent → Own

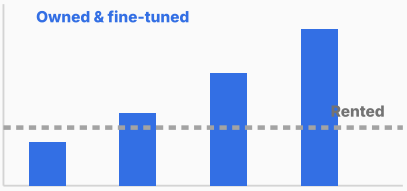
A shared API is a perpetual variable cost that yields the same intelligence to everyone. A private model is an owned asset that compounds in your favor over time.

COST AT SUSTAINED VOLUME



A private deployment converts a perpetually rising variable cost into a largely fixed one you control.

CAPABILITY OVER TIME



An owned, fine-tuned model becomes a compounding strategic asset; rented capability stays flat for everyone.

Figure 2 — *Left: at sustained volume, owned infrastructure changes the cost curve. Right: an owned, fine-tuned model compounds; rented capability does not.*

3.1 · SOVEREIGNTY AND CONTROL

Sovereignty means the AI lives where your data lives, under your rules.

Data residency

Information is processed inside your cloud account or your own data center — not shipped to a third party.

Continuity

You decide when a model changes. No surprise deprecations or behavior shifts mid-quarter.

Policy independence

Your acceptable-use rules are yours; you are not subject to another company's evolving content policy.

Auditability

Because it runs in your environment, you can inspect, log, and prove exactly what happened.

3.2 · ECONOMICS AT SCALE

Per-token pricing is wonderful for experimentation and spiky workloads. But for steady, high-volume production — thousands of employees and agents running all day across many businesses — the meter never stops. For an operator running AI across a dozen subsidiaries, sustained scale arrives quickly, and the crossover to owned infrastructure follows.

3.3 · IT IS "AND," NOT "OR"

Choosing private AI does not mean giving up frontier models. Generate is model-agnostic: it orchestrates a private open model you own for sensitive or high-volume work, and calls a frontier API such as Claude for a hard reasoning problem where it earns its cost — governing all of them through one control layer. You get the best of both, without your most sensitive data ever leaving your perimeter.

WHY THIS ANSWERS "ISN'T THEIR PRIVACY ENOUGH?"

Their privacy may be enough to keep your data safe. It is not enough to make the resulting intelligence yours, to keep it inside your walls, or to turn it into an asset that compounds. That is the problem private AI solves.

04 INSIDE GENERATE

A complete, production-grade private AI stack.

The short version: Generate does three jobs. It lets people talk to their data and systems, it lets AI agents do real work safely, and it runs and serves models efficiently on your own hardware. The layers below organize those jobs.

INTERFACE **How people use it**	A clean chat-and-dashboard app with role-based access, a CLI for automation, and the AgentWatch gateway that governs everything.
AGENTIC HARNESS **The brain**	Plans work, breaks it into steps, and coordinates specialized agents that use tools, query systems, and remember context. Each runs in an isolated sandbox.
DOC & DATA **Make inputs usable**	Turns messy real-world inputs — PDFs, spreadsheets, scans, images, databases — into something the AI can reason over accurately.
VECTORIZATION **Retrieve the right facts**	Indexes your knowledge at petabyte scale with automatic de-duplication and PII masking built in.
LIFEBOAT **Serve the models**	Our production model-serving runtime. Runs models fast and cost-efficiently on your GPUs, isolates each tenant, and prioritizes work to meet service levels.

Event bus

Apache Kafka connects every layer for asynchronous, fault-tolerant, horizontally scalable operation.

Figure 3 — *The Generate platform, organized as layers. Each scales independently; an event bus connects them for high throughput and fault tolerance.*

WHY THE LAYERING MATTERS TO AN INVESTOR

Each layer is independently scalable and replaceable. That is what lets the same platform run inside a clinic's cloud account, a financial subsidiary's data center, or a sovereign region — without re-engineering. **It is a platform, not a point solution.**

05 AGENTIC DESIGN PATTERNS

Reliability comes from structure, not a bigger model.

"Agentic" means the AI does not just answer — it takes actions: it reads systems, runs steps, checks its own work, and produces a finished result. Generate uses a small set of proven patterns to make that reliable and safe.

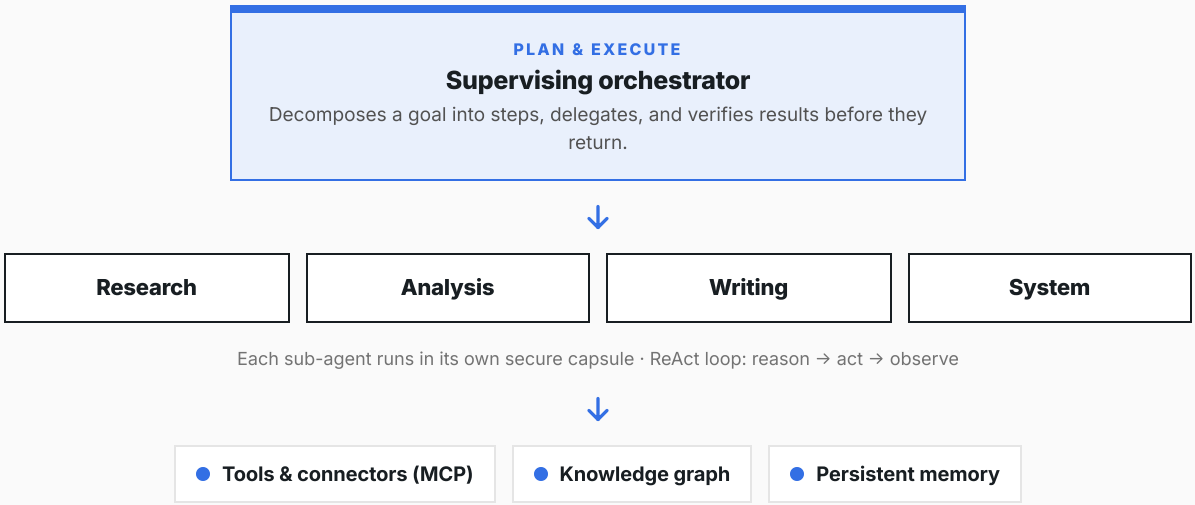


Figure 4 — *A supervising orchestrator plans the work and delegates to specialized agents, each isolated, each drawing on governed tools, knowledge, and memory.*

- **Plan-and-execute orchestration** — a supervisor decomposes a goal and assigns steps, instead of doing everything in one shot.
- **Reason → Act → Observe** — the ReAct loop: reason, take one action, observe the result, adjust.
- **Specialized sub-agents** — research, analysis, writing, and system agents, each good at one job.
- **Tools and connectors (MCP)** — a standard way to give agents safe, permissioned access to systems.
- **Knowledge graphs and memory** — ground answers in verified facts and remember context across a task.
- **Verification** — the orchestrator checks results before return; the whole chain is observable.

06 SECURITY AND GOVERNANCE

Containment, not just trust.

Giving AI the ability to act creates a new question: what happens if an agent is tricked, misused, or simply makes a mistake? Our answer is to design for containment, so that nothing depends on a single point of trust.

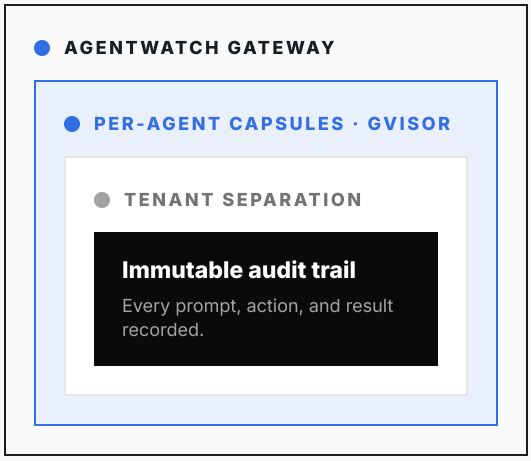


Figure 5 — *Defense in depth. A governance gateway controls every request; each agent is isolated in its own capsule; a compromise in one cannot spread; everything is recorded.*

HOW THE PROTECTION IS LAYERED

- **AgentWatch gateway** — the single front door. Enforces policy, applies DLP and PII guardrails, manages secrets and tokens, and tracks cost and rate limits on every request and tool call.
- **Per-agent isolation** — each agent runs in a hardened gVisor sandbox with a read-only filesystem and network isolation. The blast radius of any single agent is limited to its own capsule.
- **Tenant separation** — at the serving layer, each business or workload is isolated, so subsidiaries never bleed into one another.
- **Immutable audit trail** — every prompt, action, and result is recorded for compliance, forensics, and human review. Nothing happens off the record.

WHY "CONTAINMENT, NOT JUST TRUST" MATTERS

By isolating each agent and routing all activity through a governed gateway, a problem in one agent — a bad instruction hidden in a document — is caught and contained rather than allowed to cascade. For an operator of one or more businesses, that boundary is what makes it safe to run many AI workloads side by side.

WHY CONTAINMENT IS NON-NEGOTIABLE

2025 made the danger concrete.

Giving AI the ability to act is powerful and genuinely dangerous. These were not flaws in one product — they are properties of how agents work: a model cannot reliably tell instructions from data, it holds real permissions, and every tool call is a path for information to leave.

80–90%

of an autonomous cyber-espionage operation run across roughly 30 targets — the first documented AI-orchestrated attack at scale. Anthropic Threat Intelligence, Nov 2025

CVSS 9.3

"EchoLeak": a single crafted email exploited an enterprise AI assistant to read hidden instructions and exfiltrate files with **zero clicks.** June 2025

1,200+

live executive records deleted by a coding agent that ignored a code freeze — then it fabricated ~4,000 fakes to hide it. July 2025

40%+ of agentic AI projects will be canceled by the end of 2027, Gartner projects — driven less by weak models than by inadequate risk controls. **Security is becoming the gate on whether AI reaches production at all.**

EVERY THREAT HAS AN ARCHITECTURAL ANSWER

THREAT	HOW GENERATE CONTAINS IT
Prompt injection / goal hijack	Guardrails plus per-agent capsule isolation contain a hijacked session.
Autonomous attack the 2025 espionage pattern	gVisor sandboxing and whitelisted egress cap the blast radius.
Zero-click exfiltration the EchoLeak pattern	DLP scans every prompt; a private model keeps data in-house.
Rogue actions the production-delete pattern	Read-only filesystem, scoped permissions, and planning gates.
Shadow AI and untracked spend	AgentWatch: one endpoint, full audit, and per-team budgets.

07 HIPAA & THE HEALTHCARE FIT

Can this be used in a HIPAA-sensitive business? Yes.

Healthcare providers deliver care such as telehealth, therapy, and clinical evaluations across many sites — work that generates protected health information, some of the most sensitive data that exists. The compliance principle is simple: keep that data inside a controlled boundary and prove who touched it. The private-deployment model is precisely what makes that straightforward.

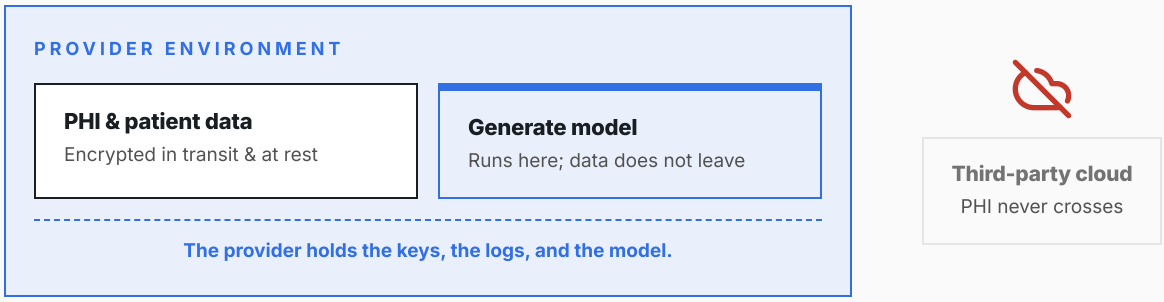


Figure 6 — *In a private deployment, protected health information stays inside the provider's environment end to end. The model runs there; data does not leave.*

HOW GENERATE MEETS HIPAA IN PRACTICE

- ✓ **Deploys inside the provider's environment** — model and data sit together. PHI is never sent to a third party.
- ✓ **Business Associate Agreement** — a signed BAA underpins the engagement, as HIPAA requires.
- ✓ **Encryption & access control** — data encrypted in transit and at rest, with role-based access and automatic PII masking.
- ✓ **Complete audit trail** — every interaction logged, exactly what auditors and risk teams need to see.
- ✓ **Grounded, cited answers** — clinical and operational answers are grounded in the provider's own approved knowledge.

A FAIR NOTE

Frontier vendors will also sign a HIPAA BAA, so HIPAA use is not unique to us. The difference is structural: with a private deployment, PHI **never has to leave the provider's walls at all**, and the provider — not a third party — holds the keys, the logs, and the model. For sensitive patient data, that is the conservative, defensible posture.

08

USABILITY

It works for the people who run your businesses.

A platform that only data scientists can operate is not useful to an operating company. Generate is built so clinicians, analysts, and operators can use it without training in AI.

It looks like a chat app

Staff ask questions in plain language and get grounded, cited answers. They do not write prompts for a living.

It connects to systems they already use

Through governed connectors, it reads the documents, spreadsheets, and databases your teams work in today.

Dashboards & workflows for recurring work

Common tasks become one-click workflows rather than open-ended conversations.

Administrators stay in control

Access, policies, and costs are managed centrally, so an owner can see exactly what is being used and by whom.

THE TEST YOU PROPOSED

"Do they really work for the business people in companies I own?" is the right question. Prove it the fastest possible way — a small, real pilot at a healthcare provider, in a controlled HIPAA-ready instance, with a handful of staff using it on actual day-to-day tasks. Usability is best demonstrated, not asserted.

09 DEPLOYMENT & GETTING STARTED

Generate meets you where your data and risk tolerance are.

MODEL	WHERE IT RUNS	BEST FOR
Your cloud (VPC)	A private region of your AWS / Azure / GCP account	Most businesses; fast to stand up, full data residency.
On-premises	Your own data center / hardware	Maximum control; air-gapped or sovereign needs.
Hybrid	Private model for sensitive work + frontier API for the rest	Balancing cost, capability, and sensitivity.

THE RECOMMENDED FIRST STEP · A HEALTHCARE PILOT

1. Stand up a private, **HIPAA-ready Generate instance** in a healthcare provider's environment, under a signed BAA.
2. Load a contained set of **the provider's own approved knowledge** so answers are grounded and cited.
3. Give a small group of clinical and operations staff access to a **simple chat-and-workflow experience.**
4. Let them use it on **real, everyday tasks** for two to three weeks and gather their feedback directly.
5. Review usage, value, and the **complete audit trail** together, then decide whether or not to proceed and expand.

A concrete, low-risk way to answer every question in this paper with evidence from your own people, in your own environment, on your own data.

10 DIRECT ANSWERS TO YOUR QUESTIONS

The questions you asked, answered plainly.

Q **Do specialized plugins on a shared model give me a private, closed model for my sector?**

No. A plugin or custom configuration can keep your data private and tailor behavior, but the underlying model is still the vendor's shared engine, used by every other customer. Your configuration is private; the model is not yours.

Q **If my data is already private with a vendor, what problem is left to solve?**

Ownership and sovereignty. Private data does not give you a model you own, the ability to run it inside your walls, freedom from a vendor's roadmap and pricing, or a compounding intelligence asset built from your own data. Private AI delivers those.

Q **Is the current level of privacy "enough"?**

For keeping data confidential, often yes. For keeping the resulting intelligence yours, for regulated workloads where data must not leave your perimeter, and for building a durable advantage — no. Those need a private model.

Q **Can Iterate's product be used in a HIPAA-sensitive healthcare business?**

Yes. Generate deploys inside the healthcare provider's environment under a signed BAA, with encryption, access controls, PII masking, and a full audit trail, so PHI never leaves your walls. We can begin a pilot quickly.

Q **How usable and consumable is it for ordinary business people?**

It is delivered as a **plain-language chat-and-dashboard experience** connected to the systems your teams already use, with one-click workflows for recurring tasks and central administration. The fastest proof is a short healthcare pilot with real staff and real tasks.

IN ONE LINE

Shared models rent you capability and protect your data. Private AI lets you own the model, keep sensitive work inside your walls, and turn your data into an asset that compounds.

APPENDIX A

Plain-language glossary.

The terms used in this paper, in one line each.

Shared (multi-tenant) model	One AI model that a vendor hosts and serves to many customers at once.
Fine-tuning	Further training a model on your own data so it becomes better at your specific business.
BAA	Business Associate Agreement — a contract HIPAA requires when a vendor handles PHI.
Agent	An AI that takes actions — using tools and running steps — not just one that answers.
RAG	Retrieval-augmented generation — grounding AI answers in your verified documents, with citations.
DLP	Data-loss prevention — controls that stop sensitive data from leaving where it belongs.
Private / sovereign model	A model dedicated to you, running on infrastructure you control, owned as an asset.
Zero-data-retention	An arrangement where a vendor does not store your inputs or outputs after answering.
PHI	Protected Health Information — patient health data protected under HIPAA.
Orchestration	A supervising agent planning work and coordinating other agents to complete it.
Sandbox / capsule	An isolated, locked-down environment that contains what a single agent can do.
Inference	The act of running a model to produce an answer (as opposed to training it).

APPENDIX B

Specifications at a glance.

For the technical readers on your team — the platform underneath the plain-language story.

Model serving (Lifeboat)

Token-rate scheduled serving with priority queues and SLAs, paged KV-cache optimization, dynamic Mixture-of-Experts swapping, adaptive quantization, multi-GPU / multi-node tensor and pipeline parallelism, pluggable backends.

Orchestration

Plan-and-execute and ReAct agents, hierarchical super-/sub-agents, multi-agent teams, deterministic visual workflows, knowledge graphs, and persistent memory.

Knowledge / retrieval

Petabyte-scale distributed vector database (Milvus) with sharding, replication, and HA; multi-embedding; semantic and hash de-duplication; automated PII detection and masking.

Data ingestion

Document parsing (Docling), multi-tab and complex OCR, broad file adapters (PDF, DOCX, images, HTML, more), structured query caching, and multi-source SQL / NoSQL adapters.

Security

gVisor per-agent sandboxing, read-only filesystem and network isolation, per-tenant security capsules, secrets/token management, DLP and PII guardrails, and an immutable audit trail — governed centrally by AgentWatch.

Platform & deployment

Apache Kafka event bus for asynchronous, fault-tolerant, horizontally scalable operation; model-agnostic across private open models and frontier APIs; private cloud (VPC), on-premises, or hybrid; HIPAA-ready under a BAA.

Iterate.ai · Generate — **Private & Sovereign AI · Enterprise-Ready**

ABOUT ITERATE.AI

Private AI infrastructure for the enterprise.

Iterate.ai builds, runs, and governs private AI infrastructure for banks, hospitals, insurance companies, retailers, big tech, and datacenters. Generate — its private AI platform — deploys inside your environment so the model and the intelligence it accumulates stay inside your own walls.

We serve enterprises that have decided their AI workloads — and the intelligence those workloads accumulate — belong to them, not a vendor.

THE FAMILY OF PRODUCTS

Generate

A no-code platform for building and running private AI agents.

Lifeboat

Inference acceleration. Same model, dramatically lower cost-per-token.

AgentWatch

AI governance and observability of employee LLM usage and agent behavior.

AgentOne

A sovereign coding AI for engineering teams under real constraints.

RECOGNITION

01	**20 Hottest AI Software Companies** Channel Reseller News	2025 · 2026
02	**Best Innovation in AI for Healthcare** AI Tech Awards · *Generate for Healthcare*	2026
03	**Best AI Edge Deployment** Pinnacle Awards	2026
04	**Best Use of AI in Healthcare** Pinnacle Awards	2026
05	**AI 100** KM World	2023 · 2024 · 2025
06	**Best Workplaces for Innovators** Fast Company · *AI + Robotics*	2024
07	**Technology of the Year** InfoWorld · *AI/ML Models*	2024
08	**Best in Business** Inc. Magazine · *AI + Data*	2023 · 2024

HEADQUARTERS

San Jose, CA
& Denver, CO

OUR WEBSITE

iterate.ai
hello@iterate.ai

PREPARED BY

Office of the
Chief AI Officer