IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Enterprise AI Infrastructure: The Eli Lilly Case Study

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-03-30 · enterprise ai · ai infrastructure · on-premises ai · data pipelines · supercomputing · pharma ai · machine learning · nvidia

Original article: https://intuitionlabs.ai/articles/enterprise-ai-infrastructure-eli-lilly



In this article

1. Executive Summary
2. Introduction & Background
3. Why “Chat Tools” Aren’t Enough
4. Historical Context: Computing in Pharma
5. Lilly’s AI Infrastructure Strategy
6. Key Components of Lilly’s AI Infrastructure
7. Case Studies and Industry Comparisons
8. Data and Collaboration in Enterprise AI
9. Enterprise AI Platforms and Tools
10. Data Analysis and Expert Perspectives
11. Implications and Future Directions
12. Conclusion

Executive Summary

Eli Lilly’s recent announcements illustrate a fundamental shift in enterprise AI strategy: major companies are **building their own AI infrastructure** rather than relying solely on packaged “AI apps” or cloud chat services. In October 2025 Lilly revealed a partnership with NVIDIA to construct the world’s most powerful AI supercomputer **dedicated to pharmaceuticals**, along with an “AI factory” that manages every stage of model development (www.prnewswire.com) (blogs.nvidia.com). Simultaneously, Lilly launched **Lilly TuneLab**, a federated AI/ML platform exposing Lilly’s proprietary drug-discovery models (trained on over $1 billion of R&D data) to biotech partners in exchange for their data contributions (investor.lilly.com) (blogs.nvidia.com). These initiatives underscore the enterprise need for **compute capacity, data access, and specialized platforms** that generic tools (such as public chatbots) cannot provide.

Alongside consumer AI tools such as ChatGPT, Lilly’s approach emphasizes **custom infrastructure**: thousands of GPUs networked with high-speed fabric, federated learning to safeguard data, and integrated software (e.g. NVIDIA’s BioNeMo, MONAI, Mission Control). This allows Lilly to train models on **millions of experiments**, apply AI at every step of drug discovery, and embed “digital assistants” into lab workflows (www.prnewswire.com) (www.lilly.com). Industry analysts note this is part of a broader enterprise trend: organizations see “cloud-based AI” (pay-per-call APIs) becoming **cost-prohibitive** at scale, commanding millions of dollars per month (www.faf.ae) (www.techtarget.com). AI platform selection should account for the audience, workload, data foundation and deployment model (Microsoft). For related background, see on-premises solutions.

This report explores **why enterprises need full AI stacks** to truly capitalize on AI, with Lilly as a case study. We examine the **compute hardware, networking, data platforms, and AI software** required – far beyond the scope of simple chat tools – and review other leading examples (e.g. Samsung’s “AI Megafactory,” DOE’s 110k-GPU supercomputer, Microsoft’s multi-datacenter AI cloud). Through data, expert analysis, and specific examples, we show that custom infrastructure is essential for mission-critical domains. While chat interfaces

can boost productivity in narrow tasks, only a dedicated AI infrastructure (compute, data pipelines, internal ML platforms) can deliver the scale, customization and integration required in fields like drug discovery. The Lilly-NVIDIA supercomputer exemplifies this principle: it is not an "app" but a **foundation** for next-generation R&D and operations (www.prnewswire.com) (www.axios.com).

Introduction & Background

Artificial intelligence has become ubiquitous in consumer applications – from virtual assistants to language chatbots – but its role in mission-critical enterprise and scientific settings remains fundamentally different. Enterprises **cannot rely solely on public chat tools or one-size-fits-all AI services**, because their challenges demand specialized data, compliance, and scale. In regulated fields like pharmaceuticals, proprietary data (clinical trials, lab experiments, patents) must remain confidential, and AI models must be trained or fine-tuned on **domain-specific knowledge** that generic systems lack. As one Lilly executive notes, a 150-year legacy means the company's *"most powerful asset is decades of data."* Leveraging that data for AI requires custom models and hardware, not just off-the-shelf apps (www.prnewswire.com) (www.lilly.com).

Over recent years, many companies experimented with "Demo AI" pilots using cloud-based APIs. These pilots often yield quick wins, but analysts warn they seldom scale. For complex customer-service or B2B workflows, evaluate chat interfaces for individual tasks (e.g. drafting text), then assess the internal data and backend integrations the application needs (www.techtarget.com). A 2026 analysis underscores this "enterprise AI trap": pilot projects cheer initial results but then stall when there is no infrastructure to handle compliance, data integration, or volume (foundationallm.ai). In short, enterprise AI success **depends on robust infrastructure and platform readiness**, not just flashy front-end demos.

Why "Chat Tools" Aren't Enough

Chat-based AI tools (like public chatbots and generative AI APIs) have driven enthusiasm for AI in the enterprise. They provide intuitive interfaces and can boost productivity on discrete tasks (e.g. summarizing reports, drafting emails). However, experts emphasize their limitations for strategic, large-scale applications. Chatbots are typically built on **general-purpose models** with broad training data, and they do not inherently incorporate a company's proprietary knowledge. They can hallucinate or give non-deterministic outputs without proper grounding. Crucially, **security and compliance** can be problematic: enterprise data sent to external services may violate privacy rules or IP protection.

In contrast, enterprise AI applications often require heavy customization and **integration with internal systems**. For example, a drug discovery AI needs to ingest lab instruments' data, molecular simulation results, and patient outcomes from Lilly's own research. A customer-service AI needs access to private CRM and historical support logs. These require pipelines that secure and preprocess data, and models that understand domain-specific terminology. Chat tools alone cannot handle this.

Industry commentators point out that **smaller, domain-specific AI models** often make more sense for companies than one massive generic LLM. Tech analyst Mike Vizard notes that most enterprises don't need an all-purpose 175-billion-parameter model – *"they need a model that 'knows the ropes' of [their field] and doesn't blow up the cloud bill every time someone hits Enter"* (techstrong.ai). Specialized models (or "agent" systems)

can run faster, be fine-tuned more easily, and keep data on-site. In practice, companies are moving toward **modular agent architectures**, where multiple lightweight models handle subtasks (e.g. one agent for data extraction, another for reasoning, another for execution), orchestrated by a controller (techstrong.ai). This modular approach demands **internal platforms and infrastructure**: model training environments, APIs, data storage, and orchestration tools.

Stanford's 2025 AI Index reports that the price of querying a model at a GPT-3.5-equivalent MMLU score of 64.8 fell from $20.00 per million tokens in November 2022 to $0.07 in October 2024—a reduction of more than 280-fold for that performance threshold (Stanford HAI). An analyst warns that enterprises now face *"monthly AI bills approaching tens of millions of dollars,"* making naïve cloud-API usage unsustainable (www.faf.ae). TechTarget reported that large enterprises using generative AI can incur **> $1 million per month in cloud costs**, driving them to in-house deployments (www.techtarget.com). Menlo Ventures' 2024 report found that **47% of solutions were developed in-house and 53% were sourced from vendors** (Menlo Ventures). At scale, enterprises should place workloads and data according to business and technical requirements, including consideration of hybrid architectures (Microsoft).

In summary, while chat tools can illustrate AI's potential, they **cannot replace the comprehensive stack** needed for enterprise applications. As Lilly's CIO Diogo Rau puts it, achieving a new standard of innovation demands *"excellence not just in science but also in technology,"* leveraging Lilly's deep R&D data with **purpose-built AI models** (www.prnewswire.com). Lilly's strategy exemplifies this principle by combining proprietary data, custom models, and vast compute resources into an "AI factory" – a far cry from a standalone chatbot.

Historical Context: Computing in Pharma

Enterprise computing has a long history of enabling complex science. Even decades ago, pharmaceutical companies invested in supercomputers for simulation and data analysis. For instance, Lilly's leadership proudly remembers that the company was *"the first adopter of the Cray-2 supercomputer among any commercial institution"* in the 1980s (www.lilly.com). That pioneering spirit is echoed today: Lilly calls its new system **LillyPod**, indicating it's the first of its kind in pharma.

In recent years, other industries have similarly begun building in-house AI infrastructure. For example, NVIDIA and HPE recently unveiled turnkey "AI factory" solutions for enterprise, combining on-premises GPU clusters with managed software stacks (nvidianews.nvidia.com). In manufacturing, Samsung announced an *"AI Megafactory"* with more than 50,000 NVIDIA GPUs and digital-twin simulation to automate chip production (www.techradar.com) (www.techspot.com). Government labs like the U.S. Department of Energy are commissioning exascale AI supercomputers (110,000+ GPUs) to accelerate science in fields ranging from materials to healthcare (nvidianews.nvidia.com). AI startups also use infrastructure partners: Anthropic selected Fluidstack to help deliver computing capacity (Anthropic).

These developments underscore a new reality: **AI at industrial scale requires industrial-scale infrastructure**. Demanding scientific workloads require suitable compute and data pipelines, but hardware ownership is a deployment choice: cloud, hybrid and on-premises infrastructure should be evaluated against

each company's needs (Microsoft). Lilly's initiative to construct its own supercomputer and collaborative platform is part of this broader trend of enterprise-grade AI engineering across sectors.

Lilly's AI Infrastructure Strategy

Eli Lilly's announcements in late 2025 and early 2026 provide a concrete case study of what enterprise-scale AI infrastructure looks like. The centerpiece is a partnership with NVIDIA to build **"the most powerful AI supercomputer owned and operated by a pharmaceutical company"** (www.prnewswire.com). This on-premises GPU cluster – dubbed **LillyPod** – is a first-of-its-kind NVIDIA DGX SuperPOD B300 system with **1,016 NVIDIA Blackwell GPUs** (www.lilly.com) (www.prnewswire.com). All GPUs and storage are connected via NVIDIA's high-speed Spectrum-X Ethernet fabric, meaning the entire cluster (compute, storage, networking) effectively shares one unified, ultra-fast network (www.prnewswire.com). This allows data to move instantly between thousands of cores during training or inference. NVIDIA's *Mission Control* management software will orchestrate workloads across the full pod, monitoring performance and automating operations securely across all 1,000+ GPUs (blogs.nvidia.com). For this system, Lilly aims to use **100% renewable electricity by 2030**, alongside efficient liquid cooling (NVIDIA) (investor.lilly.com).

Importantly, Lilly does not view this as "just another computer," but as an **AI factory** – a term Lilly uses to emphasize that the system will manage the **entire AI lifecycle**. As the official press release states, the supercomputer *"will power an AI factory, a specialized computing infrastructure that manages the entire AI lifecycle from data ingestion and training to fine-tuning and high-volume inference."* (www.prnewswire.com). In other words, rather than using external AI services or small proof-of-concept models, Lilly is building an integrated pipeline: raw R&D data goes in, models are trained or retrained on in-house and partner data, then those models are tuned and deployed at scale on Lilly's own inference infrastructure. This closed-loop cycle allows rapid iteration: scientists can take results from experiments, quickly update models, and refine hypotheses in days instead of years.

Lilly's executives stress that this infrastructure is needed to transform their science. Chief information and digital officer Diogo Rau emphasizes that Lilly combines *"scientific expertise with NVIDIA's leadership in AI computing"* to "turn data into actionable insights" (www.lilly.com). He explains that by having this capacity, the company can *"unlock deeper biological knowledge down to the individual cell and molecule levels."* For instance, one goal is to **go beyond fine-tuning public AI models** with widely available data: instead, Lilly can train *"powerful models using our own internal data"*, including the "lessons learned from millions of molecules that didn't succeed" (www.lilly.com). This deep reservoir of internal pharmaceutical data – decades' worth – becomes fuel for AI, enabling Lilly to search chemical space far more intelligently than any generic model could.

Chief AI Officer Thomas Fuchs puts it this way: *"Lilly is shifting from using AI as a tool to embracing it as a scientific collaborator."* (www.fiercebiotech.com) In practice, the shift means that every step of drug R&D is augmented by AI agents and models. Lilly plans to develop *"scientific AI agents"* – essentially autonomous assistants that help researchers plan experiments, interpret results, and connect digital simulations to lab work (www.lilly.com). For example, generative models can propose novel molecular structures grounded in biochemical rules, or predict patient responses to therapies. AI will also be embedded in advanced medical

imaging workflows: by using deep learning frameworks (e.g. NVIDIA MONAI), Lilly expects to analyze massive 3D imaging datasets *"in days instead of months"*, enabling more precise biomarker discovery (blogs.nvidia.com).

Beyond discovery, Lilly foresees AI in manufacturing and supply chains. Digital twins of production lines (using platforms like NVIDIA Omniverse and RTX-powered servers) will allow Lilly to simulate and stress-test manufacturing processes virtually (blogs.nvidia.com). This can improve yield, reduce downtime, and scale capacity to meet demand for blockbuster drugs. Robotic automation – already used for quality control in Lilly factories – can be further enhanced by AI for tasks like visual inspection or material transport. In short, Lilly expects its AI infrastructure to touch **discovery, development, and delivery** of medicines, effectively accelerating the entire pipeline.

A key element of Lilly's strategy is **collaboration** via a controlled ecosystem. Many of the AI models trained on LillyPod will be made available through **Lilly TuneLab**, an AI platform unveiled in September 2025 (investor.lilly.com). TuneLab lets external biotech firms access Lilly's proprietary drug-discovery models (built on Lilly's datasets) in a privacy-preserving manner. Partners contribute to model improvement without transmitting their underlying structures, sequences or individual data points (CDD). The initial TuneLab release includes models based on *"proprietary data obtained at a cost of over $1 billion,"* reflecting Lilly's decades of drug-discovery dataset development (investor.lilly.com). This federated approach (built on NVIDIA FLARE technology) means Lilly can *share intelligence without sharing raw data* (blogs.nvidia.com) (investor.lilly.com). As Nisha Nanda of Lilly's Catalyze360 program notes, TuneLab 'compresses decades of learning into instantly accessible intelligence' for biotechs (investor.lilly.com).

The goal is to make Lilly itself an "AI-native" company where intelligence is woven into every layer of R&D and operations (investor.lilly.com) (www.lilly.com). This enterprise strategy – massive on-prem compute plus rich data and software – is what Lilly believes will **supercharge medicine discovery and delivery** faster than conventional methods (www.prnewswire.com) (www.fiercebiotech.com).

Key Components of Lilly's AI Infrastructure

Lilly's announcements highlight several critical components of enterprise AI infrastructure. These are summarized below, with references to Lilly's statements and industry context:

- **High-Performance Compute (HPC) Hardware:** LillyPod is built on NVIDIA DGX SuperPOD architecture. A SuperPOD is a turnkey cluster combining multiple DGX systems with high-speed interconnect. LillyPod is now online as the world's first DGX SuperPOD with **DGX B300** systems, powered by **1,016 NVIDIA Blackwell Ultra GPUs** (NVIDIA, February 26, 2026). The GPUs are connected on a unified Spectrum-X fabric (blogs.nvidia.com) (www.prnewswire.com). Such scale means Lilly can train models on extremely large datasets (e.g. millions of chemical simulations) in parallel. For context, NVIDIA's announcement of DOE systems (Solstice/Equinox) shows that 110,000 GPUs yield ~2,200 exaflops of performance (nvidianews.nvidia.com) – Lilly's 1,000-GPU cluster, while smaller, is similarly unprecedented in pharma and delivers multi-exaflop *AI performance*. The cluster will be

managed for **energy efficiency and sustainability**: NVIDIA describes efficient liquid cooling and Lilly's aim to run the infrastructure on 100% renewable electricity by 2030, rather than a verified current renewable-power result (NVIDIA) (investor.lilly.com).

- **Networking and Orchestration:** To harness so many GPUs, Lilly employs advanced networking and management software. The entire SuperPOD uses NVIDIA's **Spectrum-X Ethernet**, ensuring a high-speed, unified interconnect among all compute and storage nodes. This means data flows seamlessly between GPUs during distributed training or inference (www.prnewswire.com). Orchestration is handled by NVIDIA's **Mission Control** software, which Lilly will use to monitor and manage the HPC cluster. Mission Control provides a central dashboard for scheduling jobs, automating workflows, and ensuring security across the multi-node system (blogs.nvidia.com). In effect, Mission Control treats the thousands of GPUs as a single compute continuum optimized for AI workloads, a critical capability for enterprise-scale usage. (Analogously, Microsoft's AI WAN interconnects datacenters to act as one giant supercomputer (news.microsoft.com).)
- **Software Stack and AI Platforms:** Lilly's AI factory comes with a full software ecosystem. The NVIDIA "full-stack" includes both low-level libraries (like cuDNN) and high-level toolkits. Notably, Lilly plans to use platforms like **NVIDIA BioNeMo** to train generative models for molecules. BioNeMo allows combining Lilly's experimental datasets with public scientific knowledge to propose novel antibodies, proteins, and other biotherapeutics (blogs.nvidia.com). For medical imaging, Lilly will leverage the open-source **MONAI** framework, which is optimized for healthcare AI; this can reduce image analysis from months to days (blogs.nvidia.com). These are examples of how an enterprise AI infrastructure integrates specialized ML frameworks tailored to its domain. Additionally, Lilly's TuneLab itself is an AI/ML platform built on federated learning (using NVIDIA FLARE) so that multiple organizations can use and improve shared models without exposing data (blogs.nvidia.com). This kind of internally-managed AI platform is a key part of infrastructure, beyond mere hardware: it governs data access, model deployment, and cross-company collaboration.
- **Data Pipelines and Proprietary Data:** A cornerstone of Lilly's plan is its vast, proprietary datasets from decades of drug R&D. The HPC will ingest pre-processed data from discovery, preclinical, and clinical experiments (e.g. assay results, chemical libraries, safety studies). Lilly explicitly notes it has "decades of data" at its disposal (www.prnewswire.com). For example, TuneLab's models were trained on Lilly's full "drug disposition, safety, and preclinical" datasets covering **hundreds of thousands of unique molecules** (investor.lilly.com). Building an AI infrastructure means creating automated pipelines to curate, label, and feed these data into model training. Lilly's system must also integrate new data continuously: as scientists run new assays or adopt the digital twins, the AI pipeline will capture those results to retrain models. We see a sign of this continuous learning vision in the LillyPod announcement: it will enable "rapid learning and iteration", letting scientists train models on "millions of experiments" to dramatically widen the scope of discovery (www.prnewswire.com). In short, enterprise AI infrastructure must ingest **both existing and emerging data** across the company, far beyond what a standalone chatbot could access.

- **Federated and Collaborative Learning:** To extend infrastructure beyond Lilly's walls while respecting privacy, Lilly's TuneLab platform employs **federated learning**. TuneLab's federated training is hosted by a third party. Each participating node trains on its own proprietary data, with raw data remaining in that node; only model updates are sent to the central server (Lilly TuneLab). The press release explains that curated biotech data enhances the models "for the benefit of others in the ecosystem and ultimately patients" (investor.lilly.com). Such federated platforms are becoming an important part of enterprise AI infrastructure, allowing cross-organization learning. TuneLab's design (federated AI/ML platform) enables Lilly to "expand access to advanced discovery tools" without compromising any party's data (www.prnewswire.com). This collaborative model is itself an infrastructure component – akin to how some large tech consortiums share models, but with secure data governance built-in.
- **Governance, Security, and Compliance:** Enterprise AI governance should be planned alongside the technical platform. NVIDIA describes Lilly's AI factory as a secure, scalable, purpose-built platform for healthcare and life sciences (NVIDIA). For general background on ChatGPT Enterprise, the publisher also provides a guide (intuitionlabs.ai).

Lilly's HPC can handle **high-volume training** of large **foundation models**, then immediately switch to **high-throughput inference** for internal apps, all within its own secure network (www.prnewswire.com) (blogs.nvidia.com). This end-to-end integration is precisely what enterprise experts say is lacking in "lighter" AI solutions (foundationallm.ai) (techstrong.ai). Lilly's program is an archetype of moving from reading science papers to *doing* AI-enabled science at industrial scale.

Case Studies and Industry Comparisons

Lilly's approach is one of several recent examples showing how major enterprises are building AI infrastructures. Below we compare Lilly with other high-profile AI initiatives:

Organization	AI Infrastructure Initiative	Components & Scale	Primary Goal
Eli Lilly (Pharma)	Formation of the **Lilly AI Factory** (announced Oct 2025) and **Lilly TuneLab** (Sept 2025) (www.prnewswire.com) (investor.lilly.com).	On-prem **NVIDIA DGX SuperPOD B300**: >1,000 Blackwell GPUs with unified Spectrum-X network (www.prnewswire.com). Managed by Nvidia Mission Control. Dedicated HPC in Lilly data center (100% renewable electricity is a 2030 aim; NVIDIA). TuneLab platform with federated learning.	Accelerate drug discovery and development with AI: train models on millions of internal experiments, share models with biotech partners via TuneLab (www.prnewswire.com) (investor.lilly.com).
U.S. DOE / Argonne (Science)	**Solstice & Equinox** AI Supercomputers (announced Oct 2025) (nvidianews.nvidia.com).	Solstice: 100,000 NVIDIA Blackwell GPUs; Equinox: 10,000 GPUs; high-speed interconnect. Combined ~2,200 exaflops (nvidianews.nvidia.com). Built by NVIDIA and Oracle.	Enable "agentic AI" for scientific discovery across research areas (healthcare, energy, materials) (nvidianews.nvidia.com).

Organization	AI Infrastructure Initiative	Components & Scale	Primary Goal
Samsung (Tech/Manufacturing)	**AI Megafactory** (announced late 2025) (www.techradar.com) (www.techspot.com).	>50,000 NVIDIA GPUs integrated across semiconductor fab operations. Uses NVIDIA Omniverse for digital twins and robotics.	Embed AI across chipmaking, robotics, and mobile manufacturing. AI-managed processes (e.g. lithography, quality control) to speed up production and optimize yield (www.techradar.com).
Microsoft (Cloud)	**Fairwater AI Superfactory** (announced 2025-2026) (news.microsoft.com).	Networked datacenters in Wisconsin and Atlanta (and others), linked by ultra-fast fiber. “Hundreds of thousands” of GPUs, exabytes of storage, millions of CPU cores across sites (news.microsoft.com).	Pool distributed resources as a unified platform to train large AI models (e.g. OpenAI’s models, Copilot). Reduce training time from months to weeks through parallelism (news.microsoft.com).
Anthropic (AI Startup)	**Claude Data Centers** (announced Nov 2025) (abcnews.go.com).	New AI datacenters planned in Texas and New York; $50 billion investment announced; working with Fluidstack. Large capacity for Claude chatbot (Cloud TPU/GPUs not specified). Industry-wide datacenter leasing surged by 7.4 GW in Q3 2025 for AI (abcnews.go.com).	Scale up computing to meet enterprise demand for Anthropic’s AI services. The investment reflects need for owned infrastructure as API usage grows (abcnews.go.com).
HPE & NVIDIA (Enterprise)	**NVIDIA AI Computing by HPE** turnkey solution (announced June 18, 2024) (nvidianews.nvidia.com).	Private cloud for AI: integrated NVIDIA AI stack on HPE hardware. Supports inference, fine-tuning, retrieval-augmented generation on private data (nvidianews.nvidia.com). Features HPE GreenLake service, sustainability focus.	Give enterprises a plug-and-play AI platform to run generative AI on-premises. Emphasizes data privacy/control while accelerating use-case deployment (nvidianews.nvidia.com).
(Emerging) Other Sectors	**Enterprise AI deployment choices**	IBM documents hybrid cloud infrastructure for its generative AI model development (IBM researchers).	Cost-control and customization: assess total cost of ownership, including infrastructure, support and implementation (Microsoft). Evaluate costs against workload needs and applicable data requirements.

This table illustrates that Lilly’s initiative is part of a **wider movement**: leading organizations are investing in bespoke AI infrastructure tailored to their domain. For instance, while Lilly focuses on molecular data and lab automation, Samsung focuses on manufacturing processes, and Microsoft on general AI model training. In each case, the solution is not an off-the-shelf AI app, but rather a **computational platform** combining hardware, software, and data pipelines.

Key Takeaway: Chatbots and generic AI services have democratized AI access but cannot solve the unique challenges of industries like pharma. Vendor software is also part of enterprise AI infrastructure: NVIDIA says its AI Enterprise platform streamlines data science pipelines and production deployment (NVIDIA). Enterprises that understand this are building “AI factories” (a turnkey concept also popularized by NVIDIA and HPE (nvidianews.nvidia.com)) to maintain competitive advantage and control over their data.

Data and Collaboration in Enterprise AI

A recurring theme in Lilly's planning (and in enterprise AI at large) is that **data is as important as compute**. Lilly emphasizes the value of its proprietary datasets, and plans to harness decades of R&D knowledge. As Diogo Rau stated, Lilly has *"decades of data"* from its 150-year history, and the goal is to use "purpose-built models" on that data to accelerate discovery (www.prnewswire.com). The initial set of TuneLab models alone was built on Lilly data whose acquisition cost exceeded $1 billion (investor.lilly.com).

Building AI infrastructure means establishing data pipelines that can feed this precious data into AI models continuously. In practice, Lilly will need to aggregate data from chemistry, biology, preclinical assays, learned synthetic routes, clinical trials, etc., into a coherent "AI-ready" format. This requires ETL processes, metadata tagging, quality control, and a fast storage backend. It also includes data from digital twins and sensors in manufacturing. For example, Lilly's supercomputer will support **digital-twin simulations** of supply chains and production lines, generating new data on process dynamics. Integrating this with R&D data (e.g. correlating synthesis failures in the lab with manufacturing bottlenecks) could reveal new insights. Rich, cross-domain data is a hallmark of "AI infrastructure" as opposed to siloed apps.

Lilly's strategy for data is twofold: **internal use** and **ecosystem sharing**. Internally, the company will use all available data to train its models. For example, the TuneLab is backed by Lilly's complete preclinical dataset covering "hundreds of thousands of unique molecules" (investor.lilly.com). In effect, Lilly compresses its historical research into AI models: these models implicitly encode lessons from failed compounds, side-effect profiles, and large screening campaigns. By continually retraining on new experiments, the models become ever more attuned to Lilly's specific drug-discovery space.

Externally, Lilly's approach is collaborative. Recognizing that innovation thrives in an ecosystem, TuneLab invites biotech partners to contribute data. The PR notes that partner contributions "fuel continuous improvement" of the shared AI models (investor.lilly.com). In this federated learning setup, Lilly infrastructure enables smaller companies to benefit from Lilly's AI while giving back their learnings. This symbiotic data collaboration is possible only because Lilly built a dedicated platform (with privacy-first design); a generic AI product would not so easily allow such controlled data combinatorics.

Industry-wide, this pattern is emerging: organizations are finding new ways to pool and reuse data for AI. For example, in healthcare some consortia use federated networks to train models on patient data across hospitals without exchanging records. Lilly's TuneLab is an analogous model for pharma R&D. This shows how **data infrastructure** in enterprise AI includes not just in-house databases but also protocols and platforms for data collaboration. TuneLab's federated training uses third-party hosting (Lilly TuneLab).

Finally, enterprise data infrastructure must deal with **privacy, IP, and compliance**. TuneLab's federated training is hosted by a third party (Lilly TuneLab). According to NVIDIA's description, *"data stays private and separate"* for each user (blogs.nvidia.com). Thus, the AI infrastructure is built with data sovereignty in mind. In industries like pharmaceuticals, this is crucial: sensitive data can never be carelessly placed on a public cloud. Lilly explicitly highlights that TuneLab operates on a third-party host *"employ [ing] federated learning, a privacy-preserving approach"* (investor.lilly.com), underscoring that data governance was a core design goal of their AI platform.

In summary, Lilly's example shows that enterprise AI infrastructure must provide **comprehensive data management**: massive storage and processing capacity on-premises, pipelines to collect lab/manufacturing data, and secure collaboration mechanisms for sharing models. General-purpose chat tools and Lilly's custom research infrastructure serve different roles. The Lilly case demonstrates that the true infrastructure includes every link in the chain from raw data to model to end-user application.

Enterprise AI Platforms and Tools

Beyond hardware and raw data, Lilly's plan includes specialized software and platforms – effectively treating AI as an integrated system. Several concrete examples emerge from Lilly's announcements:

- **Lilly TuneLab Platform:** This AI/ML platform (launched Sept 2025 (investor.lilly.com)) is perhaps the clearest example of an in-house tool. TuneLab is hosted off-site (to handle scale) but tightly controlled by Lilly. It offers a catalog of Lilly-trained models for various discovery tasks (chemistry predictions, ADME/toxicity, etc.) and, crucially, it implements **federated AI/ML**. Lilly describes TuneLab as "the first drug discovery platform to offer Lilly models and NVIDIA Clara [open models] in a federated environment" (blogs.nvidia.com). This means biotech firms can run Lilly's models against their own data without sending that data to Lilly. Rhino also identifies AWS cloud and data analytics services as supporting TuneLab (Rhino). This illustrates how a custom application can use external infrastructure services.
- **AI Lifecycle Orchestration:** The concept of an "AI factory" implies handling end-to-end workflows. Lilly explicitly says the new system will manage **fine-tuning and high-volume inference** in addition to training (www.prnewswire.com). In practice, this means Lilly is investing in MLOps tools: software for continuous integration/continuous deployment of models, experiment tracking, monitoring performance, and versioning. NVIDIA's Mission Control provides part of this, but Lilly likely also needs platforms akin to Databricks or Kubernetes for ML. They will have pipelines that take new lab data, retrain models, validate them, and push updates to researcher interfaces. For example, if a model suggests a novel molecule, Lilly researchers can synthesize it, test it, and the results would flow back into the model retraining cycle. This closed-loop demands a robust software pipeline – something that out-of-the-box chat systems simply cannot replicate. It is akin to building a mini "open-source Lab-to-Model pipeline" across the enterprise network.
- **Integration with Scientific Tools:** Lilly's AI infrastructure is not an island; it links to existing scientific software. For instance, Lilly will integrate AI with its **digital twin** initiatives for manufacturing, using NVIDIA Omniverse and RTX servers (blogs.nvidia.com). Omniverse allows detailed simulation of Lilly's production lines and supply chains, which can then feed into AI-driven optimization. Likewise, Lilly can use AI for real-time monitoring: sensor data from bioreactors or lab instruments could be analyzed by models trained on historical process data. Integrations with lab information management systems (LIMS), electronic lab notebooks, and clinical data systems are all part of making this infrastructure work. This all-encompassing approach is necessary for enterprise – it's not just building a new app, but weaving AI into the digital fabric of the company.

- **Productivity and Collaboration Tools:** Internally, Lilly will likely deploy AI-enhanced applications for employees across functions. The press mentions using large language models to accelerate **clinical writing and documentation**, and to power internal GPT-based agents for workflows (blogs.nvidia.com). For example, one use-case is speedier medical writing: automatically drafting sections of regulatory documents or summarizing research findings. Lilly scientists may get custom LLM-based assistants trained on Lilly's own medical and scientific corpus, again requiring fine-tuning on Lilly data. Traditional productivity suites may incorporate these AI copilot features on-premises. These workflows rely on a secure, bespoke platform and middleware aligned with Lilly's legal and IP requirements.
- **Governance and Monitoring Tools:** Governance teams can use the NIST AI Risk Management Framework to support their AI risk management efforts. Its use is a recommendation here, rather than a description of Lilly's implementation. For general background on ChatGPT Enterprise, the publisher also provides a guide (intuitionlabs.ai).

Analyst commentary on enterprise AI often echoes this: simple plug-and-play solutions give way to heavy engineering. For example, one analysis of IT trends predicts that enterprises will *"rebuild everything"* – from network architecture to talent structure – around AI (www.faf.ae). Lilly's investments mirror that thesis: they are effectively *architecting the organization* around this AI platform, not just running a pilot project.

Data Analysis and Expert Perspectives

Multiple industry experts and studies corroborate Lilly's strategy. Several salient points arise from the research:

- **ROI and Scale of Investments:** Building such infrastructure is expensive, but the payoff can be huge. Lilly estimates that the first TuneLab models include proprietary research data obtained at a cost of over $1 billion (Lilly). IBM, Microsoft, and Dell note that "enterprise control" over AI – including running on-prem – is a major driver in purchasing decisions (www.techtarget.com) (nvidianews.nvidia.com). A LinkedIn publication by technology analysts emphasizes that while chat tools encourage adoption, *"the cracks show"* at scale if underlying infrastructure is weak, potentially leading to wasted investment (foundationallm.ai). When companies moved AI from pilots to real products, unexpected costs appeared: e.g. a consumer brand might find their "pay-per-use" cloud bills skyrocketing to millions per month – making a capital investment in dedicated servers economically sensible (www.faf.ae) (www.techtarget.com). In other words, the math often favors owning hardware for steady, heavy workloads.
- **Talent and Cultural Challenges:** Analysts stress that technology is only half the battle – the other half is people and process. The World Economic Forum reports growing demand for generative AI skills among businesses and individuals (World Economic Forum). Techstrong's analysis similarly warns of a talent gap: *"skilled practitioners are scarce"*, so companies must commit to upskilling rather than just buying tools (techstrong.ai). Lilly's move implicitly acknowledges this: it is hiring AI scientists and engineers (the presentation at NVIDIA GTC, others) and involving researchers in iterative AI work. To truly leverage an AI factory, Lilly will need biologists and chemists who can use AI outputs, data scientists to maintain models, and IT staff to run the infrastructure. This holistic development – building

not just an app but an ecosystem of skilled users and engineers – underpins success. Microsoft's infrastructure report identifies an immediate need for AI experience and talent (Microsoft, The State of AI Infrastructure). Lilly's approach tries to do both.

- **Empirical Gains:** While Lilly's announcements focus on future capability, some evidence from other projects illustrates potential gains. Microsoft's Materials Science collaboration, for instance, demonstrates that AI+HPC can **compress years of lab work into weeks**. In that DOE partnership, researchers used Azure Quantum and AI to sift through 32 million candidate materials, isolating 18 promising compounds for batteries in just 80 hours (news.microsoft.com). This kind of speedup – finding useful outputs in days instead of decades – typifies what Lilly hopes to achieve in drug discovery. If LillyPod enables even a fraction of that acceleration for pharmaceutical R&D, the investment pays dividends in faster drug candidates and reduced costs per discovery. Microsoft also cautions that AI's speed does not guarantee accuracy (Microsoft). Performance and financial returns should be evaluated separately for each project.
- **Multi-stakeholder Benefits:** Lilly explicitly frames its platform as benefiting a broad ecosystem. By sharing models via TuneLab, Lilly creates a **collective learning environment**: each participant's data makes the shared AI models better. Analysts note that such federated platforms could become standard in pharma and biotech, ensuring that innovation isn't locked in silos. Lilly's TuneLab news page emphasizes improving access and pooling training data to translate AI's drug-discovery potential into tangible results (Lilly TuneLab). This collaborative goal should be distinguished from an assumed revenue or cost-recovery model.
- **Long-term Industry Shift:** Finally, thought leaders see Lilly's strategy as indicative of a broader era where **AI is the new lab instrument**. Huang (NVIDIA's CEO) famously predicted that drug research would move away from traditional wet labs towards AI-driven platforms (www.axios.com). He specifically cited Lilly's supercomputer as evidence: three years ago "all of their R&D budget was probably wet labs," but now they're pivoting to AI (www.axios.com). This endorsement from an outside leader reinforces that Lilly's approach is not a one-off but part of a structural change. In effect, companies in science-intensive industries will increasingly allocate R&D dollars to computing estates and AI engineers, just as manufacturing companies invest in robotics and automation.

Implications and Future Directions

Lilly's investments foreshadow several implications for enterprise AI:

- **Competitive Edge in Pharma:** By building an AI infrastructure, Lilly aims to shorten drug cycles and bring therapies to market faster. If successful, this could translate to blockbuster drugs reaching patients years earlier than competitors. Other pharma companies are likely to follow suit or partner with AI-specialists to avoid falling behind. Lilly's stated infrastructure goal is rapid learning and iteration (Lilly). For Lilly, the gamble is that the upfront cost of GPUs and data platforms will pay off in innovation leadership. Early results (e.g. new targets found, faster trial designs) will determine whether this becomes a standard model across the industry.

- **Organizational Transformation:** Lilly is not just buying hardware; it is restructuring how it does research. Expect Lilly to develop new roles (AI scientists, data engineers, ML ops managers) and possibly even change team structures (e.g. “AI centers of excellence”). The Foreign Affairs analysis suggests companies will evolve AI-centric governance and federated teams (www.faf.ae). Lilly’s case encourages enterprises to rethink job designs – for example, pairing biologists with AI specialists, or funding internal tools innovation rather than only external vendors.
- **Vendor Ecosystems:** Lilly’s partnership with NVIDIA and the involvement of multiple tech providers indicates the rise of AI infrastructure vendors. Beyond NVIDIA, companies like HPE, Cisco, Dell, et al. will offer “AI stack” solutions (often co-developed with partners). Lilly’s strategic choice of partners suggests it wants cutting-edge tech, but also robust enterprise support. We may see more pharma/biotech co-innovation labs (as Lilly has one with NV) and specialized services for life sciences AI. Vendors will compete to provide the secure, compliant AI platforms that Lilly needs.
- **Data as Strategic Asset:** Lilly’s emphasis on decades of research data highlights a broader point: companies with rich data histories have a huge advantage in this new AI race. This could place greater value on data governance and curation roles. The fact that Lilly shares some data only through AI models (not raw data) may spark new norms of how to monetize or share proprietary knowledge. Other industries might similarly build AI platforms to leverage their transactional or sensor data.
- **Ethical and Regulatory Considerations:** The move to enterprise AI infrastructure also raises regulatory questions. If AI influences drug design or patient treatment, oversight bodies (FDA, EMA, etc.) will scrutinize the models. Lilly will need to align its infrastructure with guidelines for AI in healthcare. For other enterprises, Lilly’s example will underline the need to embed compliance checks and ethical review into the AI pipeline from day one.
- **Beyond Chat – Toward Autonomy:** Looking forward, companies like Lilly may extend this infrastructure to *agentic AI* – systems that autonomously perform research loops. NVIDIA has already hinted that such supercomputers could enable *“agentic scientists”* (nvidianews.nvidia.com) that propose hypotheses and experiments. While dangerous to anthropomorphize, the vision is clear: Lilly is laying the groundwork for AI that can actively suggest new science. This is a far cry from passive chat tools, and it requires guardrails (as even Techstrong warns, *“clear guardrails”* are needed to avoid catastrophic errors) (techstrong.ai). Enterprise infrastructure will need to include safety layers as AI agents become more independent.

Deployment choices should follow workload requirements. For example, Microsoft offers Azure Local as an on-premises foundation for specialized workloads with governance and policy controls (Microsoft).

Conclusion

Enterprise decision-makers should define how they will judge infrastructure investments: research results that withstand validation, operational improvements, and useful collaboration. Set milestones for those outcomes alongside the technical deployment plan.

AI infrastructure capabilities should be selected for the business model and workloads (Microsoft, The State of AI Infrastructure). Hardware ownership should be evaluated as an architecture choice for each use case.

Looking ahead, we anticipate more enterprises following suit: developing private AI clouds, linking centers with fiber (as Microsoft is doing (news.microsoft.com)), and co-investing with tech partners. This shift has profound implications: IT budgets will increasingly include AI hardware and services; talent development will focus on AI engineering; and cross-industry collaborations (like Lilly with biotechs) will become common as data-sharing platforms.

References:

- Lilly press releases and web articles (www.prnewswire.com) (www.lilly.com) (investor.lilly.com)
- Nvidia blog posts and press releases (blogs.nvidia.com) (nvidianews.nvidia.com)
- FierceBiotech analysis of Lilly/NVIDIA partnership (www.fiercebiotech.com)
- Axios and AP News coverage of Nvidia and Anthropic statements (www.axios.com) (abcnews.go.com)
- TechTarget and TechRadar industry reports (news.microsoft.com) (www.techtarget.com)
- Microsoft and NVIDIA publications on AI infrastructure use cases (news.microsoft.com) (blogs.nvidia.com)
- Expert commentary on enterprise AI trends (techstrong.ai) (www.faf.ae).

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.prnewswire.com/news-releases/lilly-partners-with-nvidia-to-build-the-industrys-most-powerful-ai-supercomputer-supercharging-medicine-discovery-and-delivery-for-patients-302597285.html
2. https://blogs.nvidia.com/blog/lilly-ai-factory-nvidia-blackwell-dgx-superpod
3. https://investor.lilly.com/news-releases/news-release-details/lilly-launches-tunelab-platform-give-biotechnology-companies
4. https://www.lilly.com/news/stories/new-supercomputer-could-change-future-medicine
5. https://www.faf.ae/home/2026/1/4/the-infrastructure-reckoningai-cost-crisis-forces-enterprises-to-rebuild-everything-or-fall-behind-permanently
6. https://www.techtarget.com/searchEnterpriseAI/news/366617361/Enterprises-shift-to-on-premises-AI-to-control-costs
7. https://learn.microsoft.com/en-us/azure/architecture/ai-ml/guide/data-science-and-machine-learning
8. https://www.axios.com/2026/01/21/nvidia-jensen-huang-davos-eli-lilly
9. https://foundationallm.ai/blogs-why-ai-infrastructure-is-the-real-enterprise-battleground
10. https://techstrong.ai/videos/the-rise-of-domain-specific-ai-for-real-world-roi
11. https://hai.stanford.edu/ai-index/2025-ai-index-report/research-and-development?stream=top
12. https://menlovc.com/2024-the-state-of-generative-ai-in-the-enterprise
13. https://learn.microsoft.com/en-us/azure/architecture/hybrid/hybrid-start-here
14. https://nvidianews.nvidia.com/news/hpe-nvidia-ai-computing-generative-ai
15. https://www.techradar.com/pro/nvidia-is-building-an-ai-megafactory-for-samsung-with-50-000-gpus-could-this-be-the-start-of-a-new-ai-dawn
16. https://www.techspot.com/news/110114-samsung-nvidia-join-forces-ai-megafactory-50000-gpus.html

17. https://nvidianews.nvidia.com/news/nvidia-oracle-us-department-of-energy-ai-supercomputer-scientific-discovery
18. https://www.anthropic.com/news/anthropic-invests-50-billion-in-american-ai-infrastructure
19. https://marketingassets.microsoft.com/gdc/gdcBkY6mR/original
20. https://blogs.nvidia.com/blog/lilly-ai-factory-live
21. https://investor.lilly.com/news-releases/news-release-details/lilly-partners-nvidia-build-industrys-most-powerful-ai
22. https://www.fiercebiotech.com/biotech/eli-lilly-and-nvidia-set-build-pharmas-largest-ever-supercomputer
23. https://www.collaborativedrug.com/cdd-blog/cdd-vault-update-integrating-lilly-tunelab-small-molecule-and-antibody-predictive-models-in-cdd-vault
24. https://news.microsoft.com/source/features/ai/from-wisconsin-to-atlanta-microsoft-connects-datacenters-to-build-its-first-ai-superfactory?msockid=12a5375e44b46441039021be456a6544
25. https://tunelab.lilly.com
26. https://abcnews.go.com/US/wireStory/anthropic-announces-50b-investment-new-us-data-centers-127455028
27. https://arxiv.org/abs/2407.05467
28. https://learn.microsoft.com/en-us/azure/well-architected/cost-optimization/principles
29. https://www.rhinofcp.com/blog/federated-computing-provides-decentralized-intelligence-for-drug-development-at-scale
30. https://www.nist.gov/itl/ai-risk-management-framework
31. https://www.weforum.org/publications/the-future-of-jobs-report-2025/1-drivers-of-labour-market-transformation
32. https://news.microsoft.com/source/features/innovation/how-ai-and-hpc-are-speeding-up-scientific-discovery
33. https://tunelab.lilly.com/news
34. https://blogs.microsoft.com/blog/2026/02/24/microsoft-sovereign-cloud-adds-governance-productivity-and-support-for-large-ai-models-securely-running-even-when-completely-disconnected

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.