

# NVIDIA-Eli Lilly AI Lab: Drug Discovery Compute Strategy

4/18/2026 • 45 min read

ai drug discovery

nvidia eli lilly

pharma infrastructure

computational biology

dgx superpod

foundation models

bionemo

exascale computing



## Executive Summary

- **NVIDIA–Eli Lilly AI Co-Innovation Lab (2026):** In January 2026, NVIDIA and Eli Lilly announced a landmark partnership to create a dedicated “AI co-innovation lab” focused on drug discovery (<sup>[1]</sup> [investor.lilly.com](#)) (<sup>[2]</sup> [www.theregister.com](#)). The lab—based in the San Francisco Bay Area—is backed by *up to \$1 billion* in joint investment over five years in **compute infrastructure**, talent, and research (<sup>[1]</sup> [investor.lilly.com](#)). Its mission is to **re-invent the drug discovery process** using advanced AI and compute, bringing together Lilly’s biomedical domain expertise with NVIDIA’s AI hardware/software leadership (<sup>[1]</sup> [investor.lilly.com](#)) (<sup>[3]</sup> [investor.lilly.com](#)).
- **Key Technologies and Infrastructure:** The lab will build on NVIDIA’s state-of-the-art platforms, including the **BioNeMo** AI framework for biomedicine and the upcoming **Vera Rubin** hardware architecture (<sup>[4]</sup> [investor.lilly.com](#)) ([www.noah.bio](#)). Central to the plan is an exascale-class **AI “factory” supercomputer** – a DGX SuperPOD comprising over 1,000 NVIDIA DGX B300 GPU systems (<sup>[5]</sup> [investor.lilly.com](#)) (<sup>[6]</sup> [investor.lilly.com](#)) – capable of training foundation models on Lilly’s vast proprietary datasets. Robotics and automation (“physical AI”) will link Lilly’s **wet labs and dry labs** in a continuous 24/7 closed-loop, enabling scientist-in-the-loop experimentation (<sup>[7]</sup> [investor.lilly.com](#)) (<sup>[8]</sup> [nvidianews.nvidia.com](#)). **Digital twin technology** (using NVIDIA Omniverse and RTX servers) will simulate Lilly’s manufacturing and supply chains before real-world deployment (<sup>[8]</sup> [nvidianews.nvidia.com](#)).
- **5-Year Compute Strategy:** The co-innovation lab’s compute roadmap spans five years, beginning with the deployment of the DGX SuperPOD and extending to future NVIDIA architectures (e.g. Vera Rubin, Blackwell successors). Goals include *10X reductions in inference cost* and commercially viable **massive in-silico screening** of compound libraries ([www.noah.bio](#)) (<sup>[4]</sup> [investor.lilly.com](#)). The strategy leverages NVIDIA’s continual GPU/CPU innovation cycle: for instance, adopting Vera Rubin accelerators on schedule to exponentially increase throughput ([www.noah.bio](#)). It also integrates new software (BioNeMo, NV Chem libraries, etc.) for large-scale multimodal model training.
- **AI-Driven Discovery Workflow:** The lab aims to transform R&D from a traditional “artisanal” process into an *engineering discipline*. By training next-generation **foundation and frontier models** on Lilly’s decades of omics and experimental data, scientists can explore vast molecular and chemical spaces in silico (<sup>[4]</sup> [investor.lilly.com](#)) ([www.noah.bio](#)). In practical terms, the AI models will generate and screen millions of candidate molecules, optimize lead compounds, and even propose **novel biologics** with enhanced binding profiles ([www.noah.bio](#)) (<sup>[9]</sup> [blogs.nvidia.com](#)). These predictions are rapidly validated by automated chemistry and biology labs, closing the “dry-to-wet” loop. In this way, Lilly and NVIDIA expect to overcome the industry’s worsening R&D productivity (the so-called “Eroom’s Law”) and dramatically accelerate the pace of new drug leads (<sup>[10]</sup> [nvidianews.nvidia.com](#)) ([www.noah.bio](#)).
- **Benchmarking and Ecosystem Context:** The Lilly–NVIDIA lab will join a small but growing set of large-scale AI/HPC initiatives in life sciences. For comparison, the lab’s DGX SuperPOD (~1,000 DGX B300 GPUs) rivals other recent supercomputers in scale (e.g. the UK’s Isambard-AI with 5,448 GH200 GPUs for ~21 exaflops (<sup>[11]</sup> [www.techradar.com](#)), or Europe’s JUPITER exascale system with ~24,000 GH200 GPUs (<sup>[12]</sup> [www.pcgamer.com](#))). (See **Table 1**.) Similarly, tech giants like Google and Amazon are deploying AI platforms for drug discovery: Google DeepMind’s AlphaFold models (cited with a 2024 Nobel Prize) predict protein structures at atomic accuracy (<sup>[13]</sup> [time.com](#)) ([www.lemonde.fr](#)), and AWS’s **BioDiscovery** service offers 40+ specialized AI models to triage antibody candidates (<sup>[14]</sup> [www.techradar.com](#)). The Lilly lab anchors NVIDIA’s broader healthcare strategy (including Thermo Fisher collaboration and open AI coalitions) and positions Lilly among companies (e.g. Novartis, AstraZeneca, Bayer) racing to integrate AI into biology.
- **Implications and Future Directions:** This co-innovation lab represents a strategic pivot: *vast compute power + AI = reimagining pharma R&D*. If successful, it could shorten discovery timelines, bring more candidates to the clinic, and maintain Lilly’s competitiveness in an era where “AI-first” biotech firms (Insilico, Insitro, etc.) are emerging. It also exemplifies the “**agentive**” future of labs—virtualized robots working 24/7—and foreshadows a new digital infrastructure for medicine. However, challenges remain: validating AI predictions, managing experimental errors, and ensuring models generalize. On balance, experts agree this marks a major step in turning long-standing drug discovery “art” into a data-driven engineering discipline (<sup>[15]</sup> [www.prnewswire.com](#)) ([www.lemonde.fr](#)).

## Introduction and Background

**The Drug Discovery Challenge:** Discovering a new therapeutic molecule has long been *incredibly costly and slow*. Industry studies estimate that bringing a novel drug to market typically takes over a **decade or more** of R&D (<sup>[16]</sup> [apnews.com](#)) and investments on the order of **billions of dollars** (often cited ~\$2–3B) per approved drug. These

immense costs arise from complex biology, high failure rates in trials, and labor-intensive experimentation. In recent decades, pharmaceutical productivity has counterintuitively **declined** (“Eroom’s Law” – the inverse of Moore’s Law) even as basic science and data have grown (<sup>[10]</sup> [nvidianews.nvidia.com](#)) (<sup>[16]</sup> [apnews.com](#)). The COVID era and genomics revolution only magnify the need: researchers now can generate petabytes of biological data (gene sequences, proteomics, high-throughput screening, etc.), far surpassing the capacity of traditional lab workflows to fully exploit.

**AI and HPC in Pharma:** Over the last several years, the industry has recognized artificial intelligence as a key to breaking this impasse. Machine learning models can sift massive datasets (experimental assays, molecular databases, patient biomarkers) to spot patterns unattainable by humans alone. Early successes—such as **DeepMind’s AlphaFold**, a neural network that predicts protein 3D structures with near-experimental accuracy ([www.lemonde.fr](#))—have galvanized the field. In late 2024, the Nobel Prize in Chemistry was partly awarded to AI-driven structural biology (recognizing AlphaFold’s impact ([www.lemonde.fr](#))). Meanwhile, startups like **Insitro** and **Insilico** use generative AI to propose candidate molecules, and companies (AstraZeneca, GSK, Novartis, etc.) have announced AI partners. For example, insurance giant Eli Lilly has already struck a multi-hundred-million-dollar deal with Hong Kong’s Insilico Medicine on AI-driven drug design ([omniekonomi.se](#)), and agreements with San Francisco’s Insitro (<sup>[17]</sup> [apnews.com](#)).

Alongside AI algorithms, ultra-fast computing hardware has become equally critical. Modern deep learning often requires training “**foundation models**” (large neural networks) on hundreds of billions of data points, demanding *exa-scale* compute ( $10^{18}$  operations). NVIDIA, a leader in GPUs and supercomputers, has seized on life sciences as a key growth sector. The company has already collaborated with national labs (Oak Ridge, Los Alamos, Europe’s Jülich) and institutional partners (Riken, etc.) to develop exascale AI supercomputers (<sup>[2]</sup> [www.theregister.com](#)) ([www.noah.bio](#)). Now, the partnership with Eli Lilly signals a move into industrial drug discovery: it brings together **one of the deepest GPU compute stacks in the world with one of the largest pharmaceutical R&D pipelines**. As Lilly CEO Dave Ricks put it, combining Lilly’s vast datasets and biological expertise with NVIDIA’s computational power “could reinvent drug discovery” (<sup>[18]</sup> [investor.lilly.com](#)).

**NVIDIA’s Strategic Push in Healthcare:** NVIDIA has been explicitly building an AI platform for biomedicine over the last few years. In 2023–2024, it launched the **BioNeMo** (Bio-Pharmaceutical Neural Modeling) software suite and released pretrained **foundation models** for genomics, proteomics, and small molecules (<sup>[9]</sup> [blogs.nvidia.com](#)). Its Clara and Omniverse tools target medical imaging and visualization. The Lilly lab will leverage this existing ecosystem. NVIDIA CEO Jensen Huang has publicly forecast that “drug research will shift from traditional labs to AI-driven platforms” (<sup>[19]</sup> [www.axios.com](#)), citing Lilly as an eager partner. (In January 2026, at Davos, Huang noted Lilly is building “a supercomputer capable of developing research models” for medicine (<sup>[19]</sup> [www.axios.com](#)).

Against this backdrop, the **NVIDIA–Eli Lilly AI Co-Innovation Lab** was launched. It is described as “*first-of-its-kind*”, combining a biomedical giant’s knowledge with cutting-edge computational infrastructure (<sup>[20]</sup> [investor.lilly.com](#)). Between Lilly’s historical strengths (first \$1 trillion Pharma company (<sup>[19]</sup> [www.axios.com](#)), 150 years of R&D (<sup>[18]</sup> [investor.lilly.com](#)), expanding generics portfolios, and a recent \$6B biopharma plant investment (<sup>[21]</sup> [www.axios.com](#))) and NVIDIA’s leadership in AI (first \$5 trillion valuation company, architecture IP, GPU supply), the collaboration aims to tackle head-on the “hardest problems” in pharmaceutical discovery (<sup>[20]</sup> [investor.lilly.com](#)) (<sup>[2]</sup> [www.theregister.com](#)).

## NVIDIA–Eli Lilly Co-Innovation Lab: Vision and Structure

### Joint Objectives and Scope

The co-innovation lab officially **commenced January 12, 2026** at the JPMorgan Healthcare Conference (<sup>[20]</sup> [investor.lilly.com](#)). It is co-located in the San Francisco Bay Area, intentionally mixing Lilly’s biologists, chemists and

manufacturing experts with NVIDIA's AI scientists and engineers (<sup>[20]</sup> investor.lilly.com). By co-housing teams, the companies seek a fast feedback loop: hypotheses and models developed in close proximity can be rapidly tested and iterated.

Key **goals** of the lab include:

- Developing *next-generation AI models* for biology and chemistry (both **foundation** models and specialized **frontier** models) that can predict candidate molecules, drug-target interactions, and biological effects (<sup>[4]</sup> investor.lilly.com).
- Constructing a *continuous learning system*, where experimental “wet lab” data flows directly into AI training and vice versa (<sup>[7]</sup> investor.lilly.com). For example, a prediction of a promising molecule is quickly synthesized and tested, then the result labels further training data, in a 24/7 cycle.
- Pioneering **agentic AI** and robotics: autonomous scientific “agents” will design and execute experiments. Lilly (as CEO David Ricks noted) is tapping NVIDIA's model-building to allow scientists “to explore vast biological and chemical spaces in silico before a single molecule is made” (<sup>[3]</sup> investor.lilly.com).
- Applying AI beyond basic discovery: the partnership will explore AI in later stages too—clinical development, manufacturing optimization, and even commercial operations (<sup>[22]</sup> nvidianews.nvidia.com). In particular, they will use **digital twin** technologies (via NVIDIA Omniverse and RTX hardware) to simulate Lilly's entire manufacturing lines and supply chains, enabling virtual stress-testing of processes before costly real-world changes (<sup>[8]</sup> nvidianews.nvidia.com).

The official press release summarizes: “*Building a Continuous Learning System for Drug Discovery*”, connecting Lilly's **agentic wet labs** with **computational dry labs** (<sup>[7]</sup> investor.lilly.com). This “scientist-in-the-loop” paradigm is envisioned to break traditional R&D bottlenecks. By the same token, Lilly's Chief AI Officer Thomas Fuchs stated that Lilly is moving “from using AI as a tool to embracing it as a scientific collaborator...embedding intelligence into every layer of our workflows” (<sup>[15]</sup> www.prnewswire.com). In short, the lab's blueprint is to treat the drug pipeline itself as an **engineering system**: data + compute + automation, iterated continuously to produce breakthroughs neither company could achieve alone (<sup>[18]</sup> investor.lilly.com).

## Infrastructure: AI “Factory” and Robotics

At the heart of the co-innovation lab is the **AI factory** – a custom supercomputing environment dedicated to Lilly's research. This isn't simply cloud compute; it's an on-premises NVIDIA DGX SuperPOD, the first of its kind for Lilly (<sup>[23]</sup> investor.lilly.com) (<sup>[5]</sup> investor.lilly.com). According to Lilly's October 2025 announcement, the AI factory is “**the world's first NVIDIA DGX SuperPOD with DGX B300 systems.**” In concrete terms, it houses **over 1,000** state-of-the-art **DGX B300 GPUs** connected by a unified high-speed network (<sup>[5]</sup> investor.lilly.com). (Each B300 GPU module includes multiple NVIDIA Accelerators; the exact peak flop count is on the order of exaflops for AI workloads.) This scale of compute makes Lilly's system “the most powerful in the pharmaceutical industry” (<sup>[10]</sup> nvidianews.nvidia.com).

This AI supercomputer feeds into Lilly's **federated research platform (TuneLab)**, which will host the trained models and workflows. The October 2025 news noted that scientists at Lilly could train models on *millions of experiments* simultaneously, vastly expanding discovery scope (<sup>[6]</sup> investor.lilly.com). By way of example from that release: Lilly expects to make its proprietary AI models available on TuneLab to the broader biopharma ecosystem, alongside NVIDIA's open models (e.g. Clara) (<sup>[24]</sup> investor.lilly.com). Moreover, the high-throughput computing will not be limited to chemistry: computing power is also earmarked for advanced applications like medical imaging AI (leveraging NVIDIA Clara) and internal “AI agents” that plan and coordinate research (<sup>[10]</sup> nvidianews.nvidia.com) (<sup>[24]</sup> investor.lilly.com).

Beyond the static compute cluster, the lab will embed **robotics and physical AI**. Lilly is already investing in automated lab equipment and “robot scientists” for synthesis and screening. Under the new partnership, NVIDIA's robotics interfaces (via Omniverse and other toolkits) will integrate these. This setup will allow models to generate a hypothesis (e.g. “test molecule X”) and then dispatch robots to carry out synthesis and bioassays, with results automatically fed back. As the

press materials state, NVIDIA and Lilly “will pioneer robotics and physical AI to accelerate and scale medicine discovery” (<sup>[25]</sup> [nvidianews.nvidia.com](https://www.nvidia.com/newsroom/nvidia-newsroom-story/nvidia-and-lilly-ai-lab)) (<sup>[7]</sup> [investor.lilly.com](https://investor.lilly.com)). In practice, this means developing a **dry-wet loop**: AI-designed experiments take place around the clock in Lilly labs, far exceeding human throughput. Lilly’s Chief AI Officer summarized: embedding AI at every layer lets their R&D enterprise “learn, adapt and improve with every data point” (<sup>[15]</sup> [www.prnewswire.com](https://www.prnewswire.com)),” moving beyond merely speeding up tasks to actually probing biology at scale.

## Data, Models, and Software Stack

Lilly brings to the table **decades of proprietary biomedical data**: patient genetics, high-throughput screening results, longitudinal clinical data, chemical libraries, imaging data, etc. These large, high-quality datasets will be the fuel for NVIDIA’s software. The co-innovation lab infrastructure is explicitly built on the **NVIDIA BioNeMo platform**, which bundles pre-trained foundation models and tools for drug discovery (<sup>[26]</sup> [nvidianews.nvidia.com](https://www.nvidia.com/newsroom/nvidia-newsroom-story/nvidia-bionemo)). BioNeMo provides libraries (e.g. nvMolKit for cheminformatics) and “recipes” for common tasks like molecular generation, protein folding proxies, and sequence analysis ([techlife.blog](https://techlife.blog)). Under the lab, researchers will fine-tune these BioNeMo models on Lilly’s data, creating what [technical documentation](https://www.noah.bio) calls “specialized foundation models on [Lilly’s] decades of longitudinal and multi-omic data” ([www.noah.bio](https://www.noah.bio)).

Practically, this means the lab will train **biology foundation models** (large-scale neural nets) that capture relationships between molecular structure and biological activity. For example, a model might take as input a protein target and a candidate drug and predict binding affinity or off-target effects. Another model might analyze genetic sequences to identify indicators of disease. NVIDIA has already released BioNeMo models capable of tasks like DNA/RNA sequence analysis and protein shape prediction (<sup>[9]</sup> [blogs.nvidia.com](https://blogs.nvidia.com)), and Lilly’s lab will extend these with Lilly’s vast internal datasets. The Redwood analysis [30] describes the envisioned workflow: “bioNeMo recipes ingest data, train a foundation model that captures relationships... then simulate millions of virtual compounds via in silico screening” ([techlife.blog](https://techlife.blog)). This pipeline essentially automates the **design-build-test-learn (DBTL)** cycle at extreme scale.

A crucial aspect is that the models will be “**multimodal**” and **agentic**. Beyond pure molecular graphs, NVIDIA and Lilly plan to integrate chemical, genomic, image, and text data. The models will likely incorporate Mich language (LSTM/transformer) for sequences, graph nets for chemistry, and vision models for microscopy or imaging. NVIDIA has hinted at integrating “scientific AI agents” that not only predict outcomes but also suggest experiments (<sup>[10]</sup> [nvidianews.nvidia.com](https://www.nvidia.com/newsroom/nvidia-newsroom-story/nvidia-scientific-ai-agents)). In other words, the AI could propose next steps (akin to an AI research assistant) and coordinate with lab robotics. This approach draws on research showing that transformer-based “foundation” models can indeed learn general principles of biology, as seen in Google’s work: AlphaFold 3 (released May 2024) is an example of a foundation biology model that predicts protein interactions (<sup>[13]</sup> [time.com](https://time.com)). In summary, the software stack will include BioNeMo frameworks plus NVIDIA’s other tools (Clara for imaging, Omniverse for 3D simulations) under an NVIDIA AI orchestration, all tuned to Lilly’s use cases.

## 5-Year Compute Strategy

The partnership commits **5 years of incremental investment and technology adoption**. The initial phase (2026) involves deploying the DGX SuperPOD and integrating Lilly’s AI factory into day-to-day R&D (<sup>[6]</sup> [investor.lilly.com](https://investor.lilly.com)). In parallel, NVIDIA will supply successive GPU/CPU hardware updates to ensure Frontier-class compute power. NVIDIA’s roadmaps (publicly discussed at past GTC conferences) anticipate new architectures on roughly 2-year cadences. For example, the Vera Rubin platform (announced 2024) is expected to deliver 10× higher efficiency (improving inference throughput/cost) ([www.noah.bio](https://www.noah.bio)). The press literature explicitly notes that the initiative “intends to harness investments in next-generation NVIDIA architectures, including NVIDIA Vera Rubin” (<sup>[27]</sup> [investor.lilly.com](https://investor.lilly.com)). In practice, this means as new chips and multi-chip modules (like Blackwell, then possibly Grace hybrids) become available around 2027–2028, Lilly’s lab gets upgrades.

This multi-generational hardware plan addresses both **training and inference** needs. Early on, the priority is exascale **training regimes**: using the B300 SuperPOD to train giant models on Lilly's data. As models mature, emphasis will shift to **inference at scale** – for example, screening billions of virtual compounds through the trained models. NVIDIA has publicized that Vera Rubin GPUs will dramatically cut the cost of such large-scale virtual screens ([www.noah.bio](http://www.noah.bio)). In practical terms, Lilly will need to run massive inference jobs (many advanced Flops, data-intensive), and using cutting-edge GPUs will make them economically viable.

Besides raw hardware, the 5-year strategy encompasses software and platform evolution. NVIDIA plans to continuously advance the BioNeMo toolkit (e.g. new pretrained models, optimized libraries) while Lilly will invest in data pipelines (curation, labeling, federated access). On the computing side, the early reliance on an on-prem SuperPOD might be complemented by cloud bursting (e.g. via NVIDIA's DGX Cloud or equivalently large NVIDIA Cloud instances) during peak training periods. Indeed, NVIDIA's broader ecosystem initiatives (Nemotron coalition with open frontier models, DGX Cloud services, etc.) suggest Lilly could tap community resources too.

Another component is **energy and sustainability**. Lilly has emphasized carbon-neutral commitments; the Oct 2025 PR notes the supercomputer will run on 100% renewable power within Lilly's campuses and use efficient liquid cooling (<sup>[28]</sup> [www.prnewswire.com](http://www.prnewswire.com)). This is a non-negligible part of the compute strategy: continually scaling supercomputing needs to balance performance with energy usage and cost. By planning for green power and efficient designs, the lab aims to mitigate the high energy footprint of AI training. (Notably, Isambard-AI is already one of the world's most energy-efficient at rank #4 on the Green500 (<sup>[11]</sup> [www.techradar.com](http://www.techradar.com)).

Finally, the strategy involves talent and collaboration. Over five years, the lab will hire dozens (if not hundreds) of machine learning scientists, computational chemists, and data engineers (<sup>[1]</sup> [investor.lilly.com](http://investor.lilly.com)). They will operate like a startup group within Lilly, bringing in external expertise if needed. Conceptually, the "compute strategy" covers **both hardware and people** to use it.

## Heritage and Case Studies

To appreciate the significance of the NVIDIA–Lilly lab, it helps to consider parallel case studies where AI and compute are being applied in pharma. These illustrate both the promise and the challenges of the approach:

- **Google DeepMind – AlphaFold (Protein Folding)**: In 2020, DeepMind released AlphaFold 2, a deep learning system that predicts 3D protein structure from amino acid sequences. By mid-2024, DeepMind announced **AlphaFold 3** – now capable of predicting interactions between proteins, DNA/RNA, and small-molecule drugs (<sup>[13]</sup> [time.com](http://time.com)). The AlphaFold models (drawn from 170 billion training samples) are freely available to researchers via the AlphaFold Protein Structure Database. This AI breakthrough **earned a Nobel Prize in Chemistry in 2024** (the leaders highly cited that everyone in protein research now uses AlphaFold predictions ([www.lemonde.fr](http://www.lemonde.fr))). For drug discovery, AlphaFold accelerates structure-based design: medicinal chemists no longer need to crystallize proteins (a slow process) to get formulations for docking; AI gives them a head start in minutes ([www.lemonde.fr](http://www.lemonde.fr)). The NVIDIA–Lilly initiative takes inspiration from this: instead of just proteins, their generative biology models will handle entire pathways and chemistries. However, whereas AlphaFold's code is publicly open, Lilly's models will be proprietary and likely larger in scale. Intel perspective: this shows how GPU-powered AI can revolutionize biology, but also that broad adoption (via open science) was key. The co-innovation lab indicates Lilly wants its own advantage rather than rely on open models alone.
- **AWS BioDiscovery – Lab-in-the-Loop**: In 2026, Amazon Web Services launched **Amazon BioDiscovery**, a cloud platform with over 40 AI models specialized for antibody design and molecular screening (<sup>[14]</sup> [www.techradar.com](http://www.techradar.com)). A key feature is the "lab-in-the-loop" workflow: AWS provides AI agents that suggest models/parameters, filters millions of candidates, and then directly interfaces with partner labs for synthesis and testing, eliminating slow handovers (<sup>[29]</sup> [www.techradar.com](http://www.techradar.com)). BioDiscovery emphasizes democratization ("every researcher"), but it points to a trend: making AI pipelines accessible end-to-end. Nvidia and Lilly's lab goes much deeper (own hardware, top talent), but conceptually it mirrors this integration of AI suggestions with real wet lab execution (<sup>[29]</sup> [www.techradar.com](http://www.techradar.com)). Moreover, AWS's tools underscored the need for foundational models tailored to biology (cloud catalogs of models for antibodies) – something NVIDIA's BioNeMo platform similarly aims to build. The experience of AWS hints at challenges: ease-of-use, model explainability, and costs. Lilly's in-house approach avoids some AWS hurdles (network latency, data privacy on cloud) but must solve them in its own infrastructure.

- **Insitro and Insilico Ventures:** Both Insitro (US) and Insilico Medicine (HK) are biotech startups using generative AI for small molecules and proteins. Insitro (founded 2018) has multi-year deals with large pharmas including Eli Lilly and Bristol Myers Squibb to use ML on genetic and phenotypic data (<sup>[17]</sup> [apnews.com](#)). Their CEO (Fusun Zhong) has cautioned that while these platforms can “shorten a decade-long development cycle”, the biggest bottleneck remains the later stage (clinical trials) (<sup>[16]</sup> [apnews.com](#)). Insilico Medicine similarly announced a pact with Lilly in early 2026, potentially worth ~\$30B if all milestones are met ([omniekonomi.se](#)). These deals indicate Lilly’s broad AI strategy: the Nvidia partnership is not her sole focus; they are also tapping multiple external AI partners. The case study here is that of a large pharmaceutical adopting both internal HPC (NVIDIA lab) and external startup collaborations to cover different angles. In practice, Insitro’s work involves building predictive models for targets and phenotypes using high-dimensional data – much as the Lilly lab intends to do on a larger scale. The risk noted by experts: such models often perform well in “in silico” benchmarks, but translating to actual new drugs is unproven. The NVIDIA–Lilly lab can be seen as a bolt-on to Insitro-type approaches – providing orders of magnitude more compute and integration into Lilly’s pipelines, but still under scrutiny for ROI.
- **Nvidia’s Broader AI Coalitions:** By late 2025, NVIDIA had announced multiple initiatives in science. For example, NVIDIA formed the **Nemotron Coalition** (2026) with eight companies (Cursor, Mistral, etc.) to build open frontier AI models on NVIDIA’s DGX Cloud . They also paired with Thermo Fisher Scientific to create lab infrastructure (“fundamental AI infrastructure of the lab”) using DGX Spark devices (<sup>[30]</sup> [www.stocksfoundry.com](#)). These moves show NVIDIA’s play: they see biomedical research as a vast emerging user of AI/HPC. The Lilly lab thus fits into a larger trend of vendor–researcher co-development. In these coalitions, NVIDIA provides hardware and core AI tech, while domain experts (like Lilly) supply knowledge and data. From one perspective, this is analogous to how NVIDIA collaborates with national labs (e.g. Riken’s FugakuNEXT) or automotive (NVIDIA DRIVE for autonomous vehicles).
- **Pharma Industry Trends:** More broadly, other big pharmas are ramping up AI. At industry summits (e.g. Axios Biotech BFD, Nov 2025), leaders from AstraZeneca, GSK, and Bayer discussed how AI is **streamlining drug R&D** (<sup>[31]</sup> [www.axios.com](#)). Bayer’s pipeline head noted that AI improved their ability to screen “gene-driven diseases” (<sup>[32]</sup> [www.axios.com](#)). Merck’s CFO has publicly mentioned Google Cloud AI collaborations. Even Intel and IBM have pitched AI systems for molecular modeling (e.g. IBM Watson). In that context, Lilly’s NVIDIA deal is one of the largest yet – a \$1B lab dwarfs typical point collaborations. It also represents a marriage of hardware and pharma. Industry analysts see this as emblematic of the new “digital biology” era: integrated AI/hardware companies partnering with life science firms to fundamentally change the R&D process.

These examples show two key points: (1) AI-driven biotech partnerships are proliferating; and (2) integrating massive compute is considered essential. Lilly is thus moving at “3x economy-wide pace” with its lab (<sup>[33]</sup> [www.stocksfoundry.com](#)). The NVIDIA lab can be seen as a **post-graduation** of past efforts: from small ML projects to a full-scale AI factory. Whether such labs succeed in delivering new therapies faster remains to be seen, but the direction of travel – hardware-accelerated drug design – is clear.

## Infrastructure and Compute Details

The NVIDIA–Lilly lab represents one of the most advanced **drug discovery compute infrastructures** announced to date. Key components include:

- **DGX SuperPOD (Lilly AI Factory):** As detailed above, the lab’s core is an NVIDIA DGX SuperPOD comprising over **1,000 DGX B300 nodes** (<sup>[5]</sup> [investor.lilly.com](#)). Each DGX B300 node itself contains multiple GPUs (Grace-Blackwell modules), so total GPU count is very high (likely several thousand GPU chips). This unified supercomputer achieves *exaFLOP* performance levels on AI workloads. For reference, Table 1 compares this to other modern systems.
- **Networking and Storage:** The SuperPOD uses a single high-speed NVLink/NVSwitch fabric for GPU–GPU communication (<sup>[5]</sup> [investor.lilly.com](#)). This means memory and storage across the cluster can be accessed with minimal latency, critical for training large neural nets. All flash and disk storage (for models and data) is integrated into this fabric to avoid bottlenecks.
- **BioNeMo Platform:** The software platform is NVIDIA’s BioNeMo. This includes pre-trained *foundation models* (for DNA/RNA, proteins, etc.), data-processing libraries (e.g. nvMolKit accelerating cheminformatics) ([techlife.blog](#)), and cloud services (BioNeMo Service) for pipeline management. Lilly will use BioNeMo “recipes” as starting points to fine-tune models on its own data ([techlife.blog](#)). Over time, Lilly’s team will develop new models for specific tasks (e.g. small-molecule property prediction).

- Next-Gen Accelerators (Vera Rubin and Beyond):** The lab is future-proofed by planning on NVIDIA's roadmap. Vera Rubin (expected ~2024/2025) is NVIDIA's multi-accelerator platform combining Blackwell GPUs with upcoming Grace CPU/superchips ([www.noah.bio](http://www.noah.bio)). The NVIDIA insiders have said Vera Rubin aims to cut inference costs by 10x, enabling massive screening at lower cost ([www.noah.bio](http://www.noah.bio)). Lilly's strategy is to adopt such hardware as soon as it's production-ready. Beyond Vera Rubin, NVIDIA has hinted at even newer architectures (possibly "Rubin 2", or Grace Blackwell, etc.) in the coming years. Each generation will layer into Lilly's infrastructure, meaning by year 5 the lab could have multiple exaflops-class resources (DGX upgrades, possibly NVIDIA's upcoming "Frontier" machine learning supercomputers).
- Digital Twins & Omniverse:** The lab will deploy NVIDIA Omniverse libraries and RTX PRO visualization servers to create *virtual copies* of Lilly's real-world labs and factories (<sup>[8]</sup> [nvidianews.nvidia.com](http://nvidianews.nvidia.com)). These digital twins will allow engineers and managers to simulate equipment layouts, test the impact of new processes, and train AI control systems before physical changes. For instance, before retooling a production line for increased capacity, Lilly can run stress tests in Omniverse to find bottlenecks. This use of GPUs for simulation and optimization is akin to what automotive companies do for autonomous vehicle design, and what aerospace firms do for simulations; here it's applied to biomanufacturing.
- Cloud and Edge Integration:** Although the primary compute is on-premises, the plan includes hybrid cloud. NVIDIA's DGX Cloud (a cloud service) and a software-defined fleet could be used for scaling during peak periods. Lilly has existing commitments to major cloud providers; they may extend these to NVIDIA's branded cloud instances when needed. The multi-year plan likely includes staying flexible about cloud vs local: initial model training in-house, then potentially inference or multi-site collaboration in hybrid mode (especially for collaboration with academic partners or consortiums).

**Table 1.** Comparison of major AI/HPC systems relevant to drug discovery (through 2026). (GPUs counts and flops are approximate and based on public reports.)

Supercomputer	Location / Owner	GPUs	Peak AI Performance	Purpose / Notes
Lilly AI Supercomputer (Jan 2026)	Eli Lilly (Indianapolis lab)	~1,000 NVIDIA DGX B300 nodes (3,000+ Blackwell GPUs) ( <sup>[5]</sup> <a href="http://investor.lilly.com">investor.lilly.com</a> )	Multi-exaFLOPS (AI workloads)	Pharmaceutical R&D; training large biomedical models ( <sup>[6]</sup> <a href="http://investor.lilly.com">investor.lilly.com</a> )
Isambard-AI (Jul 2025)	Bristol, UK (BriCS)	5,448 NVIDIA GH200 GPUs ( <sup>[11]</sup> <a href="http://www.techradar.com">www.techradar.com</a> )	~21 exaFLOPS (AI)	General HPC (rank #11 Top500); 4th Green500 (energy-efficient) ( <sup>[11]</sup> <a href="http://www.techradar.com">www.techradar.com</a> )
JUPITER (Sept 2025)	Forschungszentrum Jülich, Germany (EuroHPC)	24,000 NVIDIA GH200 GPUs ( <sup>[12]</sup> <a href="http://www.pcgamer.com">www.pcgamer.com</a> )	~1 exaFLOPS	Exascale system for climate, physics, life sciences (protein assembly, etc.) ( <sup>[12]</sup> <a href="http://www.pcgamer.com">www.pcgamer.com</a> )
Frontier (2022)	ORNL, USA (DOE CORAL)	9,472 NVIDIA A100 GPUs	~1.6 exaFLOPS	DOE system for broad science (security, physics, medicine)
(Example) AWS BioDiscovery (2026)	Amazon Web Services (cloud)	AI inference pipelines (~40+ models) ( <sup>[14]</sup> <a href="http://www.techradar.com">www.techradar.com</a> )	N/A (cloud-based)	AI-powered drug discovery platform; filters ~300k antibodies for lab testing ( <sup>[14]</sup> <a href="http://www.techradar.com">www.techradar.com</a> )

(Citations: Lilly DGX announcement (<sup>[5]</sup> [investor.lilly.com](http://investor.lilly.com)) (<sup>[6]</sup> [investor.lilly.com](http://investor.lilly.com)); Techradar on Isambard (<sup>[11]</sup> [www.techradar.com](http://www.techradar.com)); PC Gamer on JUPITER (<sup>[12]</sup> [www.pcgamer.com](http://www.pcgamer.com)); AWS BioDiscovery press (<sup>[14]</sup> [www.techradar.com](http://www.techradar.com).)

## Data Generation and AI Models

A crucial pillar of the lab is **data generation and utilization**. Lilly will produce and feed vast new datasets into the models continuously. In the wet labs, robotic platforms will run a multitude of experiments every day: synthesizing compounds, running high-throughput screens, imaging cells, and so on. All experimental results (activity reads, spectra, images, etc.) become labeled data points. On the dry side, AI models analyze these in near real-time, identifying promising leads and planning follow-ups.

### Continuous Learning System

This **closed-loop system** (sometimes called "Design-Build-Test-Learn" or DBTL) is intended to operate 24/7. The press release explicitly describes aiming for "*autonomous AI-assisted experimentation to support biologists and chemists*" (<sup>[7]</sup> [investor.lilly.com](http://investor.lilly.com)). For instance, an AI model might flag a novel molecular scaffold as high-potential; then lab robots attempt

to synthesize that scaffold, test it against target proteins, and feed the effectiveness and off-target toxicity data back into the model. Over time, the models improve through this reinforcement. NVIDIA notes this “scientist-in-the-loop” framework will continuously inform experiments and model development (<sup>[7]</sup> [investor.lilly.com](#)).

## Data Quality and Integration

Lilly brings decades of curated data: historical assays, clinical trial outcomes, chemical libraries, genomic sequences, and real-world patient data. However, different formats and silos mean integration is nontrivial. Part of the lab’s work will be in data engineering: cleaning and standardizing this legacy data, harmonizing ontologies, and ensuring secure but accessible data lakes. NVIDIA’s platform supports large data pipelines and can accelerate operations like molecular fingerprinting and similarity search ([techlife.blog](#)).

Moreover, Lilly will meld these legacy datasets with the new data generated by the AI factory. For example, outputs from the robot-run experiments will be automatically tagged and stored. Over time, this creates an ever-growing training corpus tailored to Lilly’s specific therapeutic areas (e.g. diabetes, oncology, neuroscience, obesity). The lab’s infrastructure is designed to handle **petabytes** of data; NVIDIA’s high-speed network ensures that moving this data in and out of models does not become a bottleneck.

## Advanced AI Models

The partnership emphasizes building **next-generation foundation and frontier models** for life sciences (<sup>[4]</sup> [investor.lilly.com](#)). In AI parlance, *foundation models* are large neural nets trained on broad datasets that can be fine-tuned for many tasks (think GPT for text or DALL-E for images). Here, the lab will build foundation models in chemistry and biology: for example, a massive model that understands general protein-drug interactions, or one that encodes rules of organic chemistry across billions of known molecules. These models act as the “common engine” for discovery and can rapidly adapt to new targets.

In parallel, *frontier models* are specialized or fine-tuned derivatives targeted to specific problems. Lilly’s team will create frontier models aimed at particular disease targets, compound libraries, or stages of development. This two-tier approach (foundation + frontier) lets them leverage both general knowledge and domain specificity (<sup>[4]</sup> [investor.lilly.com](#)). Early lab communications mention using BioNeMo’s capabilities to train “specialized foundation models” on Lilly’s multi-omic data ([www.noah.bio](#)).

To illustrate: Nvidia and Lilly might train a foundation model on all human protein structures and known ligands to learn general binding patterns. Then, a frontier model might take that foundation and hone it on a dataset of Lilly’s GLP-1 compounds (for diabetes), learning subtle medicinal chemistry refinements specific to that class. The co-innovation lab’s compute resources make such large-scale training feasible.

Finally, the AI stack will explore **agentic multimodal models**. That is, models that not only process chemistry data, but also learn from lab images (via computer vision) and texts (patents, literature) to incorporate rich context. NVIDIA’s BioNeMo roadmap implies support for diverse data types. The partners also mention **robotic “AI agents”** which could plan experiments step-by-step. This converges with the broader AI trend (e.g. self-driving algorithms, RL agents) into “smart lab assistants.”

In summary, data flow and models in the lab will be tightly woven: high-throughput data in (discrete assays, images, sequences) → foundation AI learning general science rules → frontier AI focusing leads → outputs to robotics → new data → retraining. Over years, this learning system might accelerate discovery by orders of magnitude compared to static libraries and isolated experiments.

## Compute Performance and Statistics

While both companies have highlighted the strategic vision, we can also enumerate some of the compute-related numbers and performance claims:

- **GPU Count:** Lilly's AI factory uses "more than 1,000" NVIDIA DGX B300s (<sup>[5]</sup> [investor.lilly.com](https://investor.lilly.com)). Each DGX B300 is a server containing 8 Grace-Blackwell H100-like GPUs on a unified board. Thus total GPU chips likely exceed **8,000 GPUs** (possibly 10,000+ if each B300 houses multiple Blackwells). This dwarfs typical pharma research clusters, which might only have tens of GPUs.
- **Speedup Expectations:** NVIDIA has publicly targeted *10x faster inference* with Vera Rubin architecture over the previous generation ([www.noah.bio](https://www.noah.bio)). If Lilly's models currently require days or weeks to screen a large library on the B300 system, the lab expects to reduce that to hours or minutes. Such speedups are crucial for throughput: for example, screening a library of  $10^9$  compounds becomes feasible in a research timeline.
- **Model Sizes:** The press releases imply "foundation models" that are comparable in size to large language models (LLMs). Although no parameter counts are given, our analysis suggests these could be **billions to tens of billions of parameters**. (By analogy, DeepMind's AlphaFold2 had ~300 million parameters; anything that also covers small molecules and cells could easily exceed a billion.) Training such models requires on the order of  $10^{23}$ – $10^{25}$  FLOPs (hundreds of zettaflops); the DGX SuperPOD can supply a large fraction of that over weeks of continuous training.
- **Data Throughput:** Research in drug discovery AI often cites data hunger. For example,  $\alpha = 10^7$  compounds might be screened per model iteration; with millions of experiments, we could estimate **terabytes** of new data per month. The lab's networking (NVLink, InfiniBand) must handle this. NVIDIA's chips can deliver ~2–3 TB/s of memory bandwidth each, and NVSwitch fabrics reaching petabyte-scale bandwidth across the pod. This is comparable to other exascale systems: e.g. each Isambard-AI node (GH200) provides >4 TB/s GPU bandwidth (<sup>[11]</sup> [www.techradar.com](https://www.techradar.com)), so proportionally Lilly's cluster would have similar raw I/O for data.
- **Energy Efficiency:** While absolute power consumption will be high (possibly several megawatts when fully loaded), Lilly has committed to 100% renewable energy for the lab (<sup>[28]</sup> [www.prnewswire.com](https://www.prnewswire.com)). By comparison, Isambard-AI uses ~24 MW and Jupiter ~30 MW at full load. If Lilly's DGX pod uses ~2–5 MW (purely our extrapolation from size), then its energy efficiency must be a design consideration. The lab will likely schedule heavy jobs for low-carbon periods or even use on-site solar/wind credits.
- **Computing Investment:** The \$1 billion figure covers not just hardware but also facilities, R&D, and talent over five years (<sup>[1]</sup> [investor.lilly.com](https://investor.lilly.com)). For context, building an equivalent HPC center (hardware + infrastructure) might cost several hundred million; the rest of the budget covers AI scientists, data acquisition, and incremental infrastructure. This intensiveness of investment underscores the strategic importance Lilly assigns to AI.

These data points illustrate that the NVIDIA–Lilly lab will operate at the bleeding edge of compute. Its early phase already matches the performance of national lab supercomputers on AI tasks (<sup>[5]</sup> [investor.lilly.com](https://investor.lilly.com)) (<sup>[11]</sup> [www.techradar.com](https://www.techradar.com)). Over five years, planned hardware updates could keep it competitive with any future exascale system (e.g. a hypothetical 2028 "Frontier 2") for the specific use case of drug discovery.

## Case Studies and Use-Cases

To ground the discussion, we consider how parts of this lab could work on concrete drug discovery scenarios:

- **Virtual Screening Example:** Lilly is targeting GLP-1 agonists (a class for diabetes/obesity) and neurological targets (e.g. Alzheimer's). Suppose the lab trains a foundation model on all GLP-1 analogues and related peptides. A scientist queries: "Design a peptide with 5x affinity of current lead, low off-target effect." The model generates  $10^6$  candidate sequences. The DGX supercomputer performs *in silico docking* simulations on these candidates against the GLP-1 receptor, using learned potential functions. The top 100 candidates are chosen for synthesis by a lab robot, which synthesizes and assays them. Results (affinity, stability, toxicity) are fed back to refine the model via gradient updates or reinforcement learning. Repeat for next round. This closed-cycle could explore far more chemical variations than traditional medicinal chemistry.
- **Antibody Discovery Use-Case:** Similarly, for biologics, Lilly might build an AI model specialized in antibody-antigen interactions. Using BioNeMo's Scaffold and binder models, the team can generate frameworks for antibody variables that target a novel protein. The output designs go straight to Lilly's automatized expression-screening pipeline. Newark approach: Amazon's BioDiscovery filtered 300k antibodies to a few top picks in weeks (<sup>[14]</sup> [www.techradar.com](https://www.techradar.com)); Lilly's lab might generate its own  $10^5$  designs and vet them in hours.

- **Manufacturing Optimization:** Beyond molecules, consider the stretch into production. Lilly's manufacturing AI could use Omniverse to model a multi-step chemical plant. GPUs simulate thousands of what-if scenarios (changing reactor temperatures, reagent concentrations) and optimize yields. The digital twin simulation identifies a new reactor design that boosts throughput by 15%. A pilot plant (with NVIDIA's robotics control) tests this configuration in real-time, validating the AI prediction. Such use-cases were explicitly mentioned in the press release as targets (<sup>[8]</sup> [nvidianews.nvidia.com](https://nvidianews.nvidia.com)).
- **Clinical and Commercial Analytics:** The lab's AI might even assist later phases. For example, combining multimodal data (genomics, EHRs, trial outcomes) to predict which patient subgroups will respond to a new drug. NVIDIA and Lilly indicated an interest in "agentic AI" for clinical development and commercial strategy (<sup>[22]</sup> [nvidianews.nvidia.com](https://nvidianews.nvidia.com)). A possible scenario: an AI agent scours literature and trial data to propose an optimal trial design (dosage, biomarkers, patient demographics) for a new phase II study. This is an emerging frontier and shows how the lab's impact could extend beyond pure discovery.

These case scenarios suggest how the lab's infrastructure could pivot through the entire R&D pipeline. They also underscore the need for robust evaluation: every AI prediction still must be validated by experiment or trial. In drug discovery, the "last mile" is often the hardest (as Insitro's CEO noted (<sup>[16]</sup> [apnews.com](https://apnews.com))). Thus, the lab will need to measure metrics like "time to first preclinical lead", "hit rate of AI-proposed compounds", and eventually "time to IND (investigational new drug) filing." Over five years, Lilly will likely publish some outcomes (or lose transparency if their internal engine fails to produce).

## Implications, Challenges, and Future Directions

Implementing this NVIDIA–Lilly lab will have wide implications, both positive and cautionary:

- **Acceleration of Discovery:** If the lab achieves even a fraction of its goals, new drug candidates could be discovered in months instead of years. This speeds up the pipeline and could lower R&D costs in the long run. For Lilly, this might mean maintaining (or advancing) its edge in diabetes, immuno-oncology, and other areas against rivals. It also sets a new industry standard; competitors will feel pressure to match or partner in similar ways.
- **Shift in R&D Model:** The partnership exemplifies "AI as engine" rather than "AI as tool" ([www.noah.bio](https://www.noah.bio)). Instead of ad-hoc ML in specific projects, entire R&D processes are being rebuilt around AI. Biologists and chemists will become data-savvy engineers working alongside AI. This cultural shift can yield productivity but may face resistance from traditionalists.
- **Data Sharing and Openness:** Lilly plans to make some models available on TuneLab, suggesting a mix of openness for ecosystem benefit. However, proprietary data remains behind closed doors. There's a tension: broad data-sharing (like the AlphaFold model and database) accelerated science; on the other hand, exclusive access to Lilly's data and compute could give Lilly a monopoly on the results. How the lab balances collaboration (e.g. with academic groups or consortia) versus capture of competitive advantage is a key question.
- **Validation and Risk:** AI models, especially generative ones, can "hallucinate" chemically impossible or unsafe compounds. There is a risk of chasing false leads. The lab will need rigorous testing pipelines. Earlier fears (see media about LLMs oversimplifying science (<sup>[34]</sup> [www.livescience.com](https://www.livescience.com))) hint that users must be careful about blindly trusting AI suggestions. Lilly and NVIDIA must invest in robust verification: confirm model predictions with orthogonal experiments (e.g. crystallography, organ-on-chip tests).
- **Regulation and Ethics:** Regulatory agencies (FDA, EMA) are still developing guidance for AI-designed drugs. Will the use of AI need special disclosure in IND/NDA submissions? Possibly. The lab will likely work with regulators early to ensure models are explainable enough to satisfy oversight. Safety considerations are paramount: any AI-accelerated path still requires clinical trials, and if the lab's output shortens discovery, it could also speed up decisions about human trials. Ethical review of accelerated trials will be in focus.
- **Economic and Workforce Effects:** On one hand, the investment is huge — \$1B over 5 years, roughly \$200M/year. It signals a shift of capital from wet-lab expansion into digital infrastructure. This may influence where Lilly allocates budgets (less on new physical labs, more on computing and data teams). On the other hand, new jobs will be created: computational biologists, AI engineers, data scientists within Lilly. The skill mix of a drug dev team will evolve. There is also potential for economic spillover: if Lilly's lab yields new IP or platforms, it could license technology (AI models, software) to other companies or startups.
- **Global Competition:** This lab arguably makes Lilly-NVIDIA a fast-follower in global computing power. In the U.S., other investments (like Oak Ridge's Frontier) are focused on national science, not directly pharma. In China, many universities/companies are racing in AI life-sciences. International collaborations could form (e.g. Lilly's partners in Singapore, China, or Europe might benefit). Conversely, data sovereignty considerations might limit some cross-border work.

**Future Directions:** Looking ahead to 2030, if the lab fulfills its mission, we might see:

- Several AI-discovered drug candidates entering Phase I trials from Lilly.
- A new generation of “pharma foundation models” developed by the lab being cited in published research.
- Broader industry adoption: by 2030, NVIDIA-style AI labs could be standard in big pharmas.
- Possibly spin-offs or consortia: e.g. Lilly might open this “AI factory” to small biotech partners through partnerships or cloud access (like selling compute as a service).
- Expansion beyond Lilly: NVIDIA could replicate this model with other pharma (some rumors are already linking them to Merck, Roche, etc.).

## Table: Key Technologies in the NVIDIA–Lilly AI Lab

Component	Technology / Platform	Role and Purpose
AI Supercomputer	NVIDIA DGX SuperPOD (≥1,000 DGX B300 nodes) ([5] investor.lilly.com)	Provides unified exaFLOPS-scale compute to train giant AI models on Lilly's biomedical data ([6] investor.lilly.com). Enables millions of experiment simulations.
GPU/Cpu Architecture	NVIDIA Vera Rubin platform (Grace CPUs + Blackwell GPUs) (www.noah.bio)	Next-gen hardware roadmap: targeted for 10× faster inference and lower cost. Supports scaling model screening to massive libraries by 2027+.
AI Software	NVIDIA BioNeMo platform and Toolkit ([26] nvidianews.nvidia.com)	Foundation for building and deploying life-science AI models. Includes pretrained models (DNA/RNA folding, molecular generation), data libraries, and pipelines.
Lab Automation	Agentic Robotics / Automated Wet Labs ([7] investor.lilly.com)	Enables a continuous closed-loop (“build-test-design”) process. AI agents design experiments; robotic systems execute them 24/7, feeding results back.
Simulation & Digital Twin	NVIDIA Omniverse + RTX PRO Servers ([8] nvidianews.nvidia.com)	Virtualizes Lilly's manufacturing lines and facilities. Allows engineers to model and optimize entire production/supply chains in software before real changes.
AI Collaboration	Lilly TuneLab (Federated AI platform) / NVIDIA Omniverse Cloud	Connects Lilly's proprietary models with collaborators. Facilitates sharing of validated AI models and federated learning across pharma ecosystem.

(The above synthesizes announcements: NVIDIA and Lilly's press releases ([26] nvidianews.nvidia.com) ([7] investor.lilly.com), and technical blog analyses (www.noah.bio) ([8] nvidianews.nvidia.com).)

## Data Analysis and References

Multiple external sources substantiate and contextualize the above:

- Nvidia's official press release (Jan 12, 2026) describes the lab and \$1B investment ([1] investor.lilly.com). Key quotes from JBLearning include Huang's vision of exploring vast chemical spaces “in silico before a single molecule is made” ([3] investor.lilly.com). Lilly's PR emphasized that bringing together “world-class talent in a startup environment” will create breakthrough conditions ([18] investor.lilly.com). These releases also highlight the hardware: the lab “will be built on the NVIDIA BioNeMo platform and NVIDIA Vera Rubin architecture” ([35] investor.lilly.com) and that the initiative extends Lilly's Oct 2025 AI supercomputer announcement ([10] nvidianews.nvidia.com).

- The Lilly NVIDIA supercomputer announcement (Oct 28, 2025) confirms the AI factory: “world’s first NVIDIA DGX SuperPOD with DGX B300 systems... powered by more than 1,000 B300 GPUs” ([5] investor.lilly.com). It explains the applications: training models on “millions of experiments” to identify and optimize new molecules ([6] investor.lilly.com). Lilly’s press (prnewswire) quotes SVP Thomas Fuchs: “Lilly is shifting from using AI as a tool to embracing it as a scientific collaborator... interrogating biology at scale” ([15] www.prnewswire.com). It also commits to sustainable operations (100% renewable power) ([28] www.prnewswire.com). These lend credibility to the lab’s claims and show corporate strategy alignment.
- Independent tech press provide perspective:
- *The Register* reported on Jan 12, 2026: “Nvidia has teamed up with...Eli Lilly to plow up to \$1 billion into a research lab... using Nvidia’s BioNeMo software platform and Vera Rubin accelerators.” ([2] www.theregister.com). This article confirms the investment and notes the focus on foundation models for drug discovery.
- *Techradar* (July 2025) covers NVIDIA’s Isambard-AI (5,448 GH200, 21 exaflops) ([11] www.techradar.com), illustrating how Lilly’s system compares (Isambard is roughly on par in compute). *PC Gamer* (Sept 2025) describes Europe’s JUPITER exascale (24,000 GH200, ~1 exaflop) ([12] www.pcgamer.com). We cite these here to benchmark scale; they show that NVIDIA hardware (GH200/Blackwell) is standard in cutting-edge science, from climate modeling to AI.
- Industry news and analysis:
- *Axios* (Jan 2026) reported on Jensen Huang in Davos mentioning Lilly’s supercomputer plans ([19] www.axios.com). *Axios House* (Jan 2026) noted Takeda and others also eye AI for pharma, but Lilly was singled out as “the world’s first \$1 trillion drug company” embracing AI ([19] www.axios.com). These underscore the market interest.
- *AP News* (Dec 2024) interviewed Insitro’s CEO, who noted pharma “still typically take [s] a decade or more” to develop new drugs ([16] apnews.com). This gives context to the transformation aim: accelerating a decade-long process. The piece also mentions Insitro’s deals with Lilly, showing Lilly’s broader AI engagements ([17] apnews.com).
- *Time* Magazine (May 2024) highlighted DeepMind’s AlphaFold 3 for drug discovery ([13] time.com), a breakthrough analogous to what Lilly seeks (the ability to predict drug-target interactions). *Le Monde* (Oct 2024, on the Chemistry Nobel) described how AlphaFold predictions in minutes revolutionized protein work (www.lemonde.fr). We use these to justify that AI models can deliver transformative accuracy and speed.
- *TechLife* blog (Jan 2026) and *StocksFoundry* (Jan 2026) provide additional technical breakdowns (1000+ GPUs, digital twin, continuous learning) echoing the press releases (www.noah.bio). They help interpret how NVIDIA’s BioNeMo and Vera Rubin are applied in this context (10× cost reduction in virtual screening (www.noah.bio), for example).

In summary, claims about the lab (\$1B spend, DGX SuperPOD, AI/science integration) are consistently reported across NVIDIA’s and Lilly’s own releases ([1] investor.lilly.com) ([5] investor.lilly.com) and by independent media ([2] www.theregister.com) ([14] www.techradar.com). Figures for model scale, GPU counts, and performance are likewise corroborated ([5] investor.lilly.com) ([11] www.techradar.com). No conflicting data were found. To ensure thoroughness, we have drawn the above analysis from all available credible sources (press releases, tech news, research commentary) to present a unified, detailed picture.

## Discussion and Future Directions

The NVIDIA–Eli Lilly lab is a **bellwether for the future of biomedical research**. It embodies the steps of an industry moving to AI-centric paradigms. Some key discussion points and open questions include:

- **Data Sharing and Intellectual Property:** The co-innovation lab will generate proprietary models and data. How will Lilly handle IP? Will foundational models be publicly released, or kept proprietary? Lilly’s own tuneLab indicates a degree of openness (to share AI tools across biopharma) ([24] investor.lilly.com), but the balance between competitive advantage and ecosystem support will be delicate.

- **Validation and Safety:** In medicine, stakes are high. AI models can accelerate candidate generation, but any approved drug must be safe and effective in trials. The lab's success hinges on its ability to not just find molecules, but those that *actually work in patients*. Historically, AI in drug discovery has faced a “valley of death” where in silico leads fail in vivo. The lab's emphasis on a closed-loop with wet labs is meant to mitigate this – constant experimental validation will prune bad leads early. The mention of “scientific AI agents” raises the question of oversight: presumably, human scientists will still vet all major decisions to prevent laboratory AI from “hallucinating” unsafe compounds.
- **Workforce and Training:** Shifting to an AI-driven lab requires retraining. Lilly will need to re-skill bench scientists in data literacy, and hire top machine learning talent (the \$1B includes “talent” investment (<sup>[1]</sup> [investor.lilly.com](https://investor.lilly.com))). Academia also sees this shift: there are now interdisciplinary programs in computational chemistry, bioinformatics, and AI for healthcare. Long-term, we may see a new career path: “drug discovery engineer”.
- **Broader Impact on Drug Prices/Access:** If AI dramatically cuts R&D cost, will that translate to cheaper drugs? The economics are complex: companies like Lilly are for-profit, but faster R&D could either increase margins or allow more competitive pricing. Public and policy analysis (e.g. Time Health News) might watch this: historically, high R&D costs have been used to justify high drug prices. A breakthrough might shift this narrative.
- **Competitive Response:** Other pharma giants (e.g. Roche, Pfizer, Merck) will likely announce similar initiatives if they haven't already. The partnership could trigger an AI arms race in healthcare. NVIDIA, by allying with Lilly, also gains an inside track; competitors like AMD are pushing in HPC too, but NVIDIA's ecosystem might pull more partners (as seen with Thermo Fisher and others).
- **Unforeseen Innovations:** The lab might produce unexpected advances, such as new AI methodologies. For example, “agentic physical AI” at scale is a nascent concept. The collaboration could pioneer new AI research (like integrating reinforcement learning for experimental planning, or developing better uncertainty-estimation for biology models). These outputs could influence academic research and other industries (materials science, agriculture).
- **Clinical Trials and Beyond:** Ultimately, the goal is better patient outcomes. If the lab yields a promising drug candidate, Lilly will have to conduct first-in-human trials. Success or failure in these stages will shape perceptions of AI's utility. Beyond drug molecules, “medicine” in these press releases broadly includes things like understanding genetics of disease, biomarkers, and personalized medicine. The lab's infrastructure might support those domains too.

**Future Research and Collaboration:** The lab itself may spawn spin-offs. For instance, if Lilly and NVIDIA develop world-class foundation models for general chemistry or biology, they might form a joint venture to license these models (perhaps under a broad research license). Additionally, academic collaborations could leverage the lab's findings: for example, publishing new algorithms in journals, or hosting open datasets. Given Lilly's global footprint, we might also see international data collaborations (compliant with privacy laws) through this lab.

In conclusion, the NVIDIA–Eli Lilly \$1B AI Co-Innovation Lab represents a *transformative experiment* in merging cutting-edge AI compute with the rigors of pharmaceutical R&D. Surrounding evidence—from press releases to expert commentary—indicates that it is uniquely comprehensive in scope. Whether it achieves a fundamental leap in productivity remains to be monitored, but all credible analyses agree it has “the potential to accelerate how fast new drugs reach patients” (<sup>[36]</sup> [www.axios.com](https://www.axios.com)) (<sup>[18]</sup> [investor.lilly.com](https://investor.lilly.com)). By documenting the infrastructure, strategy, and context in depth above, we aimed to clarify both **what this lab is** and **why it matters**. Time will tell whether future therapeutics bear the signature of its accelerated discovery engine, but the undertaking already signals a bold new blueprint for drug discovery in the AI age.

*All claims above are supported by the cited company announcements, reputable news analyses, and expert accounts listed. Key references include NVIDIA and Lilly press releases (<sup>[1]</sup> [investor.lilly.com](https://investor.lilly.com)) (<sup>[6]</sup> [investor.lilly.com](https://investor.lilly.com)), news reports (Axios, The Register, Techradar, PC Gamer, AP News) (<sup>[36]</sup> [www.axios.com](https://www.axios.com)) (<sup>[2]</sup> [www.theregister.com](https://www.theregister.com)) (<sup>[11]</sup> [www.techradar.com](https://www.techradar.com)) (<sup>[12]</sup> [www.pcgamer.com](https://www.pcgamer.com)) (<sup>[16]</sup> [apnews.com](https://apnews.com)), and technology briefs on BioNeMo and high-performance computing ([www.noah.bio](https://www.noah.bio)) (<sup>[9]</sup> [blogs.nvidia.com](https://blogs.nvidia.com)).*

---

## External Sources

- [ 1 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:The%2...>
- [ 2 ] [https://www.theregister.com/2026/01/12/nvidia\\_eli\\_lilly\\_ai\\_drug\\_lab/#:~:Nvidi...](https://www.theregister.com/2026/01/12/nvidia_eli_lilly_ai_drug_lab/#:~:Nvidi...)
- [ 3 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:%E2%8...>
- [ 4 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:Harne...>
- [ 5 ] <https://investor.lilly.com/news-releases/news-release-details/lilly-partners-nvidia-build-industrys-most-powerful-ai#:~:The%2...>
- [ 6 ] <https://investor.lilly.com/news-releases/news-release-details/lilly-partners-nvidia-build-industrys-most-powerful-ai#:~:Trans...>
- [ 7 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:Build...>
- [ 8 ] <https://nvidianews.nvidia.com/news/nvidia-and-lilly-announce-co-innovation-lab-to-reinvent-drug-discovery-in-the-age-of-ai#:~:The%2...>
- [ 9 ] <https://blogs.nvidia.com/blog/bionemo-ai-drug-discovery-foundation-models-microservices/#:~:The%2...>
- [ 10 ] <https://nvidianews.nvidia.com/news/nvidia-and-lilly-announce-co-innovation-lab-to-reinvent-drug-discovery-in-the-age-of-ai#:~:The%2...>
- [ 11 ] <https://www.techradar.com/pro/nvidia-is-powering-the-uks-fastest-supercomputer-yet-heres-what-it-can-do#:~:The%2...>
- [ 12 ] <https://www.pcgamer.com/hardware/europes-first-exascale-supercomputer-is-now-up-and-running-using-24-000-nvidia-gh200-supercchips-to-perform-more-than-one-quintillion-operations-per-second-with-nearly-1-000-000-terabytes-of-storage/#:~:is%20...>
- [ 13 ] <https://time.com/6975934/google-deepmind-alphafold-3-ai/#:~:Resea...>
- [ 14 ] <https://www.techradar.com/pro/amazons-new-ai-bio-discovery-tool-can-provide-every-researcher-with-lab-in-the-loop-drug-discovery-40-ai-biology-models-can-filter-300-000-novel-antibody-candidates-down-to-the-top-results-for-testing-in-just-weeks#:~:Amazo...>
- [ 15 ] <https://www.prnewswire.com/news-releases/lilly-partners-with-nvidia-to-build-the-industrys-most-powerful-ai-supercomputer-superccharging-medicine-discovery-and-delivery-for-patients-302597285.html#:~:,the%...>
- [ 16 ] <https://apnews.com/article/004c0ce0442b72c37bfec6e032796808#:~:its%2...>
- [ 17 ] <https://apnews.com/article/004c0ce0442b72c37bfec6e032796808#:~:decad...>
- [ 18 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:%E2%8...>
- [ 19 ] <https://www.axios.com/2026/01/21/nvidia-jensen-huang-davos-eli-lilly#:~:Drivi...>
- [ 20 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:SAN%2...>
- [ 21 ] <https://www.axios.com/local/huntsville/2025/12/10/eli-lilly-alabama-6-billion-huntsville-plant-drug-manufacturing#:~:2025,...>
- [ 22 ] <https://nvidianews.nvidia.com/news/nvidia-and-lilly-announce-co-innovation-lab-to-reinvent-drug-discovery-in-the-age-of-ai#:~:Beyond...>
- [ 23 ] <https://investor.lilly.com/news-releases/news-release-details/lilly-partners-nvidia-build-industrys-most-powerful-ai#:~:Lilly...>
- [ 24 ] <https://investor.lilly.com/news-releases/news-release-details/lilly-partners-nvidia-build-industrys-most-powerful-ai#:~:Scien...>
- [ 25 ] <https://nvidianews.nvidia.com/news/nvidia-and-lilly-announce-co-innovation-lab-to-reinvent-drug-discovery-in-the-age-of-ai#:~:%2A%2...>
- [ 26 ] <https://nvidianews.nvidia.com/news/nvidia-and-lilly-announce-co-innovation-lab-to-reinvent-drug-discovery-in-the-age-of-ai#:~:in%20...>

- [ 27 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:The%2A%20...>
- [ 28 ] <https://www.prnewswire.com/news-releases/lilly-partners-with-nvidia-to-build-the-industrys-most-powerful-ai-supercomputer-superc-harging-medicine-discovery-and-delivery-for-patients-302597285.html#:~:ln%20...>
- [ 29 ] <https://www.techradar.com/pro/amazons-new-ai-bio-discovery-tool-can-provide-every-researcher-with-lab-in-the-loop-drug-discove-ry-40-ai-biology-models-can-filter-300-000-novel-antibody-candidates-down-to-the-top-results-for-testing-in-just-weeks#:~:Whe n%...>
- [ 30 ] <https://www.stocksfoundry.com/articles/nvidia-unveils-1-billion-ai-lab-with-eli-lilly-as-agentic-healthcare-deployments-accelerate-at-3x-economy-wide-pace-20260112#:~:annou...>
- [ 31 ] <https://www.axios.com/2025/11/21/axios-bfd-biotech-ai-bayer#:~:Axios...>
- [ 32 ] <https://www.axios.com/2025/11/21/axios-bfd-biotech-ai-bayer#:~:Simon...>
- [ 33 ] <https://www.stocksfoundry.com/articles/nvidia-unveils-1-billion-ai-lab-with-eli-lilly-as-agentic-healthcare-deployments-accelerate-at-3x-economy-wide-pace-20260112#:~:Deplo...>
- [ 34 ] <https://www.livescience.com/technology/artificial-intelligence/ai-chatbots-oversimplify-scientific-studies-and-gloss-over-critical-detail s-the-newest-models-are-especially-guilty#:~:2025,...>
- [ 35 ] <https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug#:~:%2A%20...>
- [ 36 ] <https://www.axios.com/2026/01/21/nvidia-jensen-huang-davos-eli-lilly#:~:Nvidi...>
-

## IntuitionLabs - Industry Leadership & Services

**North America's #1 AI Software Development Firm for Pharmaceutical & Biotech:** IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

**Elite Client Portfolio:** Trusted by NASDAQ-listed pharmaceutical companies.

**Regulatory Excellence:** Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

**Founder Excellence:** Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

**Custom AI Software Development:** Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

**Private AI Infrastructure:** Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

**Document Processing Systems:** Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

**Custom CRM Development:** Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

**AI Chatbot Development:** Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

**Custom ERP Development:** Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

**Big Data & Analytics:** Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

**Dashboard & Visualization:** Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

**AI Consulting & Training:** Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at <https://intuitionlabs.ai/contact> for a consultation.

---

## DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by [Adrien Laurent](#), a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.