WHITE PAPER

AI-Augmented Molecular Discovery: From In Silico Design to Federated Intelligence

By Revvity Signals

June, 2026

How Predictive Models, Experimental Data, and Platform Innovations Are Redefining Scientific Discovery

Summary

It is clear that artificial intelligence (AI) and machine learning (ML) are primed to transform scientific research. Yet much of the discussion remains rather abstract, focusing on algorithms rather than model outcomes. In practice, the most-immediate and high-impact application of AI in molecular discovery is predictive design: the ability to generate, evaluate, and refine small molecules, sequence based constructs, and materials virtually, before committing experimental resources in the wet lab.

Over the past decade, a small number of specialist organizations, such as Insilico Medicine, BenevolentAI, and Recursion, have shown that predictive modeling can dramatically reduce discovery timelines. However, for most pharmaceutical, biotechnology, and materials science companies, scaling these approaches to be of practical value has proven difficult. Operational barriers need to be hurdled, because models must be tightly coupled to experimental data, governed appropriately, continuously improved, and delivered directly into the workflows that scientists already use.

This white paper outlines:

- A pragmatic framework for AI-augmented molecular discovery that protects proprietary data while maximizing the learning capacity of AI models
- Platform requirements needed to link AI models to operations
- Federated learning networks, and how these can accelerate discovery
- Signals Xynthetica™, the emerging Models-as-a-Service (MaaS) capability from Revvity Signals
- How collaborations such as Lilly TuneLab™ at Revvity open up new possibilities by bringing together predictive models trained on proprietary data, and trusted infrastructure

The Shift from Experiment Driven to Predictive Discovery

Scientific discovery has always been iterative: design, make, test, analyze, repeat. What has changed is the cost and speed of iteration.[1] In many domains, including drug discovery, advanced materials, and biologics, experimental cycles are slow, expensive, and constrained by physical throughput.[2]

AI-driven approaches fundamentally alter this dynamic by augmenting the design phase and moving much of the iteration into a virtual world.[2,3] Instead of synthesizing and testing thousands of candidates in the wet lab, scientists can now harness AI models to address specific use cases, such as[4]

- Generating novel structures or sequences (e.g. antibodies) that meet a targeted profile
- Predicting key properties and behaviors including physicochemical properties, ADMET profiles, antibody developability, and immunogenicity risk
- Virtually screening against multiple objectives such as potency, selectivity, synthesizability, diversity, solubility, stability, and safety liabilities
- Shortlisting the best candidates for synthesis and experimental validation

The result is not the replacement of wet lab science, but its amplification. Experiments become more informed, more targeted, and more likely to succeed. In turn, AI models are tuned with fresh experimental data, enabling improved predictions.

Why Algorithms Alone Are Not Enough

The recent proliferation of generative models, such as diffusion models, graph neural networks, transformers, has created the impression that discovery breakthroughs are primarily down to algorithm development. In reality, predictive performance depends far more on how and when you integrate these algorithms with data.[5]

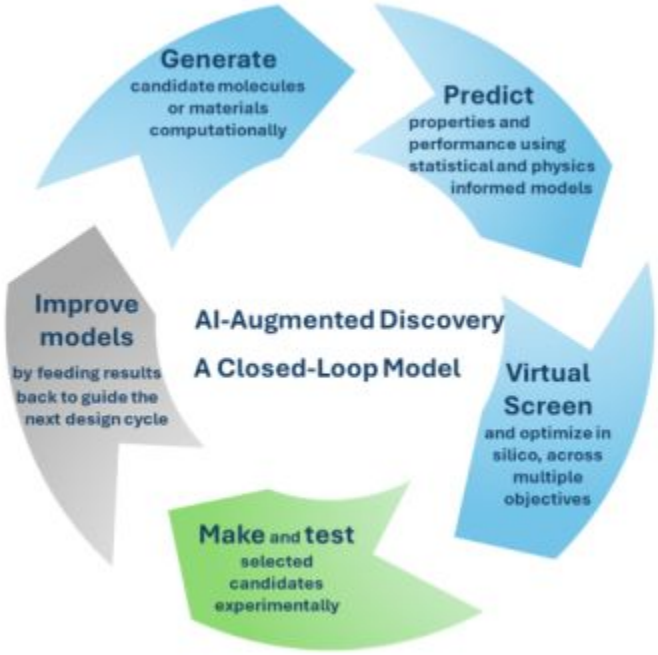
Three practical challenges typically limit the impact of AI:

1. Data fragmentation: High-value scientific data is largely private, distributed across instruments, teams, and collaborators, and often captured in different formats. Yet without consistent data context, even the most advanced models degrade rapidly.
2. Workflow disconnects: Many AI tools are not integrated into the daily workflows of bench scientists. AI predictions generated in isolation without experimental context, provenance, or traceability can be difficult to trust and act upon.
3. Model operationalization: Applying models at scale means building infrastructure for version control and governance, secure execution, and continuous retraining of models with fresh data across diverse modalities including structure-based, sequence-based,

and formulation-based representations. Setting up all these aspects internally is impractical for most organizations.

AI-Augmented Discovery: A Closed-Loop Model

AI-augmented discovery works best when virtual and wet-lab science are tightly coupled in a continuous learning loop, with models regularly learning from real experimental outcomes such as in vitro ADMET assays, antibody developability screens, or physical characterization. This loop needs to be embedded in governed, auditable scientific workflows.[6]



MaaS: Removing Infrastructure Barriers

Two possible broad approaches to infrastructure are currently being used, one relying on public solutions and the other taking the private, proprietary route. The use of public AI services inherently runs the risk of exposing IP-rich, business critical enterprise data to third parties. In addition, with token-based pricing, accessing models through external infrastructures is becoming increasingly expensive as the use of AI solutions accelerates.[4] On the other end of the spectrum, AI models can be self-hosted within an organization, but this approach can lead to duplicated effort across business groups, increasing both costs and time to market.[4]

The newest alternative solution to scale AI-augmented discovery is to deliver a Model-as-a-Service (MaaS).

The MaaS approach enables organizations to access AI/ML models (with the associated AI

technology to use them) that can be used as a shared resource across the company.[4]

- **Centralized access** to trusted models (public, commercial, or private) covering small molecule, biologics and materials
- **Consistent governance** and version control, with monitoring of model usage
- **Direct integration** with experimental data systems including assay platforms for ADME, bioactivity, and developability characterization
- **Updateable** models permitting improvements without disrupting users

MaaS enables scientists to focus on scientific questions, while offloading execution, security, and lifecycle management to the service provider.

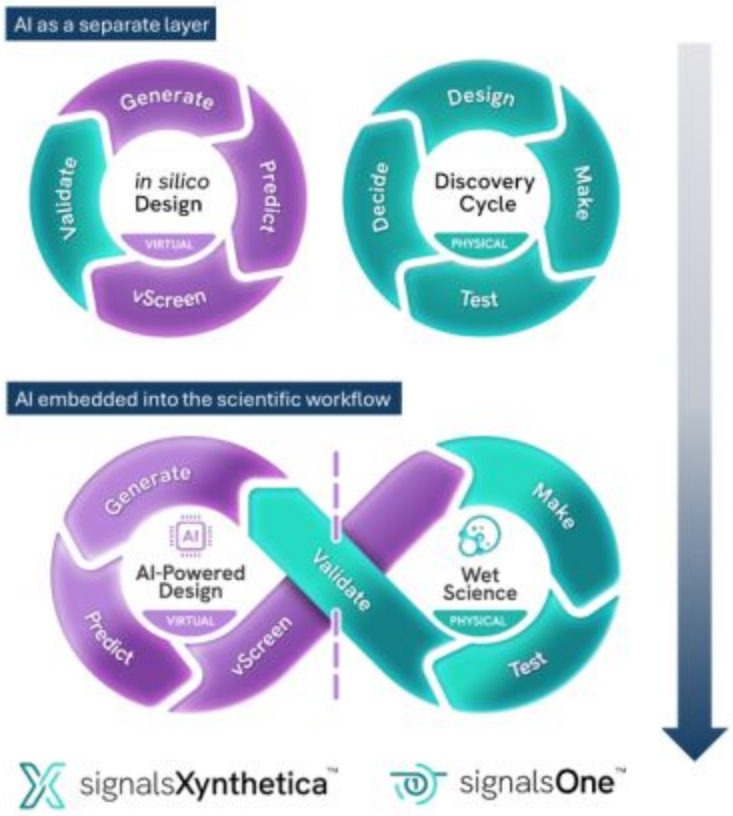
Delivering AI through MaaS offers considerable advantages to enterprise organizations, including reduced complexity, lower costs, increased security, faster innovation, non-duplication, and freedom of choice. [4]

Signals Xynthetica : Operationalizing AI for Molecular Design

Signals Xynthetica™ is Revvity Signals' emerging capability designed to make AI augmented discovery practical at scale.[7] Built within the Signals platform, Xynthetica connects virtual design, predictive modeling, and experimental validation in a governed, secured environment. Key principles of Signals Xynthetica include:

- Workflow-native AI: Models are applied directly where scientists design (Signals ChemDraw™ or Signals Notebook), analyze, and interpret experiments (Signals One)
- Continuous learning: Experimental outcomes feed back to the models, improving predictive performance over time for endpoints such as physicochemical properties, ADMET profile, developability of biologics and metrics specific to materials
- Governance by design: Models, data, and results remain auditable and compliant
- Cross-domain applicability: The same framework applies to pharmaceuticals, biologics, and materials science

Rather than positioning AI as a separate layer, Xynthetica embeds prediction into the fabric of scientific work.

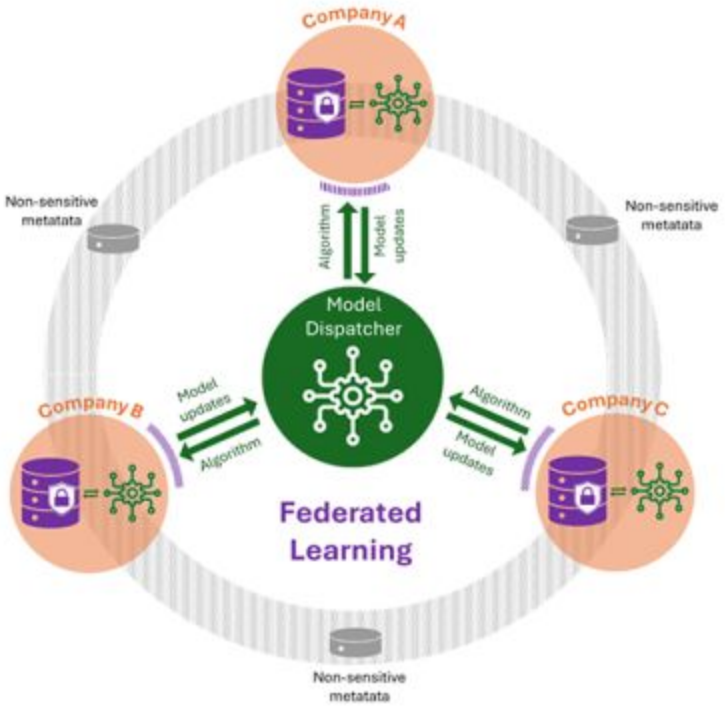


The Data Reality: Why Federated Learning Is Inevitable

The performance of modern AI models scales with data quantity, quality, and diversity. Unlike public text or images, scientific data is scarce and highly protected. It is critical that ML in areas such as pharmaceutical research are trained with ample data in order to build reliable, consistent models with a "virtual knowledge" vast enough to allow generalization to a large array of data points.[8] Yet no single organization, no matter how large, possesses sufficient data to fully realize the potential of predictive discovery models.[9]

To solve this challenge, the recent and very active research into federated learning techniques shows how ML can access multiple sources of data while eliminating the risk of leaking proprietary information to third parties.[10] For example, domains that involve molecular informatics, such as Drug Discovery, are increasingly adopting federated learning approaches.[10]

The landmark 3-year MELLODDY project successfully demonstrated that federated learning could be adopted in highly data-protective organizations such as drug discovery, without compromising proprietary information.[11] Through federated learning, models can be shared across organizations, while training occurs on private data. Updated models, but not raw data, are then aggregated to fuel collective improvement without compromising confidentiality.



Adapted from Simm et al. 202112

From Concept to Reality: Federated Discovery Networks

While federated learning has been discussed academically for years, operationalizing it in a real discovery environment demands a wholistic approach:

A trusted execution platform

Secure data stewardship

Harmonized assay metadata and ontology-driven annotation

Standardized workflows for experimental data capture and analysis

Incentives aligned across participants

The Revvity Signals platform already serves as the system of record for experimental data across hundreds of organizations, making it a natural substrate for federated discovery, where models can improve continuously without compromising intellectual property.

Lilly TuneLab at Revvity: A Proof Point for Federated AI

To develop these new capabilities, Eli Lilly and Company has created Lilly TuneLab,™ a first-of-its-kind collaborative platform that offers AI/ML drug discovery models trained on Lilly's proprietary data.13 Tunelab can be used for in silico property predictions across both Small Molecules and Antibody therapeutics, and benefit from decades of amassed research data.

Building on a federated learning platform, TuneLab allows individual companies to train models while retaining data privacy. Federation enables companies to securely share training data to teach the models to become more accurate and generalizable, contributing to the acceleration of future discoveries for all.

However, as discussed, the real-world predictive performance of models depends heavily on how and when you integrate these algorithms with your data, and this is where a MaaS comes in.

Revvity has partnered with Lilly to make TuneLab's models available through the Revvity Signals platform, expanding on the framework of Signals Xynthetica.[14] Existing Revvity Signals software, already integrated into an organization's workflow such as Signals One™, is able to seamlessly link scientists to the models they need through TuneLab's federated platform.

Collaborations such as TuneLab at Revvity illustrate how federated learning can move from concept to practice. By combining high-quality predictive models with a trusted scientific data platform, TuneLab at Revvity can:

- Extend advanced AI capabilities to biotechs
- Improve model performance through diverse experimental contributions spanning both small molecules and antibodies
- Shorten discovery timelines across the ecosystem

Importantly, these networks shift the industry from isolated optimization toward shared progress, without putting their proprietary data at risk.

Thanks to Signals Xynthetica, scientists will be able to invoke TuneLab predictive models directly from their ideation, registration, and structure-activity relationship workflows.[15] No data exports, file uploads, or context-switching will be required, and predictions will be generated at the point of scientific decision-making. The results from applying the AI models will be automatically linked to structures, assays, and experimental metadata, and outputs will be traceable, comparable, and ready for downstream analysis.

Trust, Governance, and the Role of the Platform

At the heart of AI-augmented discovery is trust, and scientists must be confident that:

- Data remains private and secure
- Models behave consistently and transparently
- Results are reproducible and traceable

Platforms that already manage experimental data, collaboration, and compliance are uniquely positioned to extend that trust into AI and federated learning contexts. By accessing TuneLab through Signals Xynthetica MaaS, scientists can be assured that model delivery, versioning, and lifecycle will be fully managed. In addition, the secure governance offered by Signals software will ensure organizational control over data and usage.[15]

The Road Ahead: From Tools to Infrastructure

The future of AI in science will not be defined by standalone algorithms, but by solutions that make predictive power accessible and governable, and which continuously improve. AI-augmented discovery represents a shift from episodic experimentation to learning systems, where every experiment strengthens the next prediction whether optimizing a small molecule's ADMET profile, improving an antibody's developability or refining a formulation. Platforms that connect models, data, and scientists will define this next era.

Signals Xynthetica and federated initiatives such as TuneLab at Revvity point toward that future: one where predictive intelligence becomes a shared capability, accelerating discovery while preserving the integrity and independence of scientific organizations.

How does the TuneLab Early Adopter program work?

Are you ready to join scientists at the cutting-edge of development and to accelerate your discoveries?

Selected existing Signals customers will receive one year of complimentary access to Lilly's TuneLab models, delivered through Signals Xynthetica MaaS, along with the corresponding Signals software components. In exchange, participating customers will provide relevant experimental data to help fine-tune the models directly within Signals Xynthetica.

The updated models—not the experimental data—will then be securely transmitted to Lilly's federation partner, where they will be aggregated with models from other contributing companies to generate improved, collective versions. This federated learning process will enable the models to increase in accuracy and generalizability while ensuring that all participants' intellectual property remains protected within the secure Signals cloud environment.

If you're ready to join the TuneLab at Revvity program,
you can pre-register through the Revvity Signals website.

Visit these dedicated websites to learn more

about Signals Xynthetica, TuneLab at Revvity, or Lilly TuneLab.

Revvity Signals provides software and scientific informatics solutions that enable discovery organizations to capture, analyze, and act on complex experimental data. Through continued innovation in AI augmented discovery and platform infrastructure, *Revvity Signals* is helping scientists move faster, from insight to impact.

References

1. Green DV. AI in drug design: evolution or revolution? *Emerg Top Life Sci.* 2025; ETLS20240005
2. Gil G et al. Accelerating science with AI. *Science*. 2025;390(6777):965
3. Hitachi Ventures.AI is Powering the future of material science: from lab to real-world breakthroughs. 2024. Available at: https://medium.com/@HitachiVentures/ai-is-powering-the-future-of-material-science-from-lab-to-real-world-breakthroughs-

2f92cf56ed90 (last accessed 30 January 2026)

4. Verma I et al. AI at scale, without the price tag: Why enterprises are turning to Models-as-a-Service. 2025. Available at: https://www.redhat.com/en/blog/ai-scale-without-price-tag-why-enterprises-are-turning-models-service (last accessed 30 January 2026)
5. Suri GS et al. Beyond boundaries: exploring the transformative power of AI in pharmaceuticals.
Discov Artific Intell. 2023;4:82
6. Zenil H, et al. The future of fundamental science led by generative closed-loop artificial intelligence. *Front Artif Intell.* 202326 Jan 2026. Provisionally accepted. Available at: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2026.1678539/abstract (last accessed 30 Jan 2025)
7. Revvity Signals. Signals Xynthetica. Available at: https://revvitysignals.com/products/research/signals-xynthetica (last accessed 30 January 2026)
8. Bassani D et al. Federated learning in computational toxicology: an industrial perspective on the Effiris Hackathon. *Chem Res Toxicol*. 2023;36:1503-17
9. Marshall A. Drugmakers share data to feed voracious foundation models. *Nat Biotechnol*. 2025;43(11):1743-6
10. Hanser T. Federated learning for molecular discovery. *Curr Opin Struct Biol*. 2023;79:102545
11. Heyndrickx W et al. MELLODDY: cross-pharma federated learning at unprecedented scale unlocks benefits in QSAR without compromising proprietary information. *J. Chem. Inf.* Model. 2024;64:2331-44
12. Simm J et al. Splitting chemical structure data sets for federated privacy□preserving machine learning. *J Cheminformat.* 2021;13:96
13. Lilly. Lilly TuneLab. Available at: https://tunelab.lilly.com/ (last accessed 30 Jan 2025)
14. Zaks Equity Research. RVTY & Eli Lilly partner to expand access to AI drug discovery models. Available at: https://finance.yahoo.com/news/rvty-eli-lilly-partner-expand-135700194.html (last accessed 30 Jan 2025)
15. Revvity Signals. TuneLab at Revvity. Available at: https://revvitysignals.com/ai-design-lilly-tunelab-xynthetica (last accessed 30 Jan 2025)

revvity signals

77 4th Avenue
Waltham, MA 02451 USA
(800) 762-4000
(+1) 203-925-4602
revvitysignals.com