IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

AI Foundation Models in Pharma: GSK-Noetik Oncology Deal

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

Published by IntuitionLabs · 2026-04-17 · ai foundation models · spatial biology · drug discovery ai · virtual cell models · oncology research · pharma ai · nsclc · colorectal cancer

Original article: https://intuitionlabs.ai/articles/ai-foundation-models-pharma-gsk-noetik-oncology



In this article

1. Executive Summary
2. Introduction
3. Background: AI in Drug Discovery and Foundation Models
4. Report Overview
5. Foundation Models in Pharmaceutical Research
6. What Are Foundation Models?
7. How Foundation Models Differ from Traditional AI in Pharma
8. Growth of Foundation Models in Drug Discovery
9. The Noetik Platform and Cancer Foundation Models
10. Noetik: Origins and Data Assets
11. Technology and OCTO Models
12. The Noetik–GSK Collaboration: Deal Details and Innovation
13. Scope of Collaboration
14. Focus on NSCLC and Colorectal Cancer Models
15. Perspectives on the Deal
16. Disease Context: NSCLC and Colorectal Cancer
17. Non-Small Cell Lung Cancer (NSCLC)
18. Colorectal Cancer (CRC)
19. Noetik–GSK Deal in Context: Business and Industry Perspectives
20. Commercial Innovation: Licensing AI Infrastructure
21. Industry Trends and Comparisons
22. Critiques and Considerations
23. Technical and Data Perspectives
24. Case Studies and Real-World Analogues
25. Implications, Challenges, and Future Directions
26. Scientific and Clinical Implications
27. Risks and Validation
28. Future Directions
29. Conclusion

Executive Summary

The recent announcement of a **five-year strategic collaboration** between GlaxoSmithKline (GSK) and Noetik – an AI-driven biotech platform – represents a watershed moment in biotech R&D. Under the agreement dated January 8, 2026, GSK committed **$50 million in upfront capital and near-term milestones combined**, plus **annual licensing fees**, to **license Noetik's AI "foundation models"** for cancer research. The announcement does not disclose the upfront/milestone split (issuer announcement). Specifically, GSK will gain access to Noetik's OCTO-VC virtual cell models for **non-small cell lung cancer (NSCLC)** and **colorectal cancer (CRC)** (natlawreview.com) (www.biopharmatrend.com). These models, trained on hundreds of millions of spatially-

resolved human cells, simulate gene expression patterns, cell states, and tumor-immune interactions in virtual patient tissues (natlawreview.com) (www.biopharmatrend.com). The partnership combines Noetik's spatial-omics AI platform with GSK's expertise in tumor immunology and drug development, with joint efforts to generate bespoke human spatial datasets aligned to GSK's priorities (natlawreview.com) (www.biopharmatrend.com).

This deal exemplifies a **new paradigm in drug discovery**: licensing AI models – built on large-scale biological data – as enterprise infrastructure, rather than buying discrete drug candidates or one-off services (natlawreview.com) (www.biopharmatrend.com). GSK's global head of AI, Kim Branson, emphasizes that "foundation models are only as good as the underlying training data", and Noetik's high-quality spatial biology at scale is novel (www.fiercebiotech.com) (www.1stoncology.com). Noetik CEO Ron Alfa describes Noetik's ambition as moving the industry from "probabilistic 'shots on goal' to deterministic engineering of cancer drugs" (issuer announcement). The transaction – believed to be among the first of its kind – establishes biological AI as a *scalable asset class* rather than a bespoke service (natlawreview.com) (www.biopharmatrend.com).

This report provides a comprehensive analysis of the Noetik–GSK collaboration and its broader context. We begin with the technical background of **foundation models** in pharmaceutical R&D, contrasting them with conventional AI/ML approaches (www.sciencedirect.com). We then examine Noetik's technology and foundation models (including the OCTO-VC models for NSCLC and CRC), the specifics of the deal, and how these models fit into oncology research pipelines. The report surveys the biology of NSCLC and colorectal cancer and illustrates how multi-modal spatial data and AI models can improve target discovery, patient stratification, and therapeutic design (www.nature.com) (pmc.ncbi.nlm.nih.gov). We include case studies of similar AI-driven efforts in oncology, review reactions from industry experts, and critique the implications of "AI foundation models" as a new paradigm (www.biopharmatrend.com). Finally, we discuss potential challenges (data quality, model validation, regulatory acceptance) and future prospects (extension to other diseases, personalized " digital twins", etc.). The report includes three comparative tables and a milestones table to aid clarity.

Introduction

Background: AI in Drug Discovery and Foundation Models

Over the past decade, artificial intelligence (AI) and machine learning (ML) have made substantial inroads into drug discovery and development. Efforts range from leveraging deep learning to predict protein structures (e.g. DeepMind's AlphaFold) to using generative models to design novel small molecules or antibodies. These AI advances have helped tackle inherent challenges in pharma, such as high failure rates and slow development timelines (www.sciencedirect.com). For a concrete example of AI applied to biological data, Noetik's original OCTO model generates images of protein expression in tumor sections from visual prompts (Noetik's July 2024 technical report).

A particularly notable breakthrough is the emergence of **"foundation models"** – large, general-purpose AI models that are pre-trained on massive datasets with self-supervised learning, and later fine-tuned for specific tasks (www.sciencedirect.com) (www.nature.com). These models were first popularized in natural language

processing (e.g. GPT-4) and image generation (e.g. DALL•E), but the paradigm is spreading rapidly into biomedical domains. A *Nature* Perspective envisions pretraining multimodal foundation models across genomics, transcriptomics, proteomics and other omics datasets (Perspective). The review emphasizes that foundation models in pharma could become "transformative": they learn broad biological patterns from huge datasets and can be adapted (via fine-tuning or prompting) to many downstream tasks (www.sciencedirect.com).

Unlike traditional narrow AI models (e.g. a neural net trained on one type of assay to predict one target property), foundation models aim to capture holistic representations of biological systems. As one perspective notes, **foundation models** can encode comprehensive information about molecular and cellular biology across multiple contexts, enabling novel applications. For example, a recent *Drug Discovery Today* review highlights how "biological foundation models" could furnish a statistical understanding of entire molecular and cellular systems, thus reducing uncertainty in target and mechanism discovery (www.sciencedirect.com). In oncology specifically, experts envision that foundation models – trained on multi-omics and spatial biology data – could allow *in silico* experiments on "patient genes, proteins, cells, and tissues" to accelerate hypothesis testing (natlawreview.com) (www.genengnews.com).

This shift is sometimes described as a "fourth wave" of AI in biomedicine (www.biopharmatrend.com). In this wave, instead of AI assisting human-led tasks one-by-one, we use **global models of biology** as infrastructure. These AI-driven "virtual cells" or "digital twins" may be used not just to screen compounds but to **engineer drugs** with deterministic understanding of tumor microenvironments (www.genengnews.com) (www.biopharmatrend.com). Achieving this vision requires massive, high-quality training data – often patient-derived – and advanced model architectures. For example, Noetik's July 2024 report describes the original **OCTO** model as trained on a joint dataset of multiplex protein staining, spatial gene expression, DNA sequencing and H&E images (technical report). This spatial, multimodal data is crucial because it captures how cells interact in intact tissues, a dimension often missing from traditional single-cell sequencing.

To put this in context: foundation models are not needed (and not yet fully realized) in biomedical fields the way LLMs are in text. Spatial transcriptomics data are not structured like language, and building effective FMs in this domain remains very challenging (pmc.ncbi.nlm.nih.gov). One recent perspective emphasizes that spatial data combine gene expression with coordinates, which do not have an obvious sequential structure for standard transformer models (pmc.ncbi.nlm.nih.gov). Nevertheless, the potential payoff is immense: pre-trained multimodal AI models could unlock new insights from vast omics atlases. **Summarizing**, the GSK–Noetik deal fits into a larger trend where pharma is betting on AI-coded knowledge as a new class of asset. Establishing foundation models trained on patient data may revolutionize R&D by offering generalizable virtual models of tumor biology, rather than point solutions.

Report Overview

This report examines the GSK–Noetik AI partnership in depth. We begin by reviewing *foundation models in pharma* (technical definitions, trends, and relevant research), and then describe Noetik's technology platform (data, models, founders, funding). We analyze the deal's structure and what it grants each party, with an

emphasis on the non-traditional licensing/subscription model and the selection of NSCLC and CRC as target diseases. We also place NSCLC and colorectal cancer in medical context (incidence, genomics, existing therapies) and review how advanced AI approaches are already being applied in these fields.

Case studies of similar efforts (e.g. AI-driven spatial analysis in lung and colorectal cancers) will illustrate how multi-omics models are used in oncology research. We incorporate quantitative data and expert commentary at each stage. The report then discusses business and strategic implications – for example, how shifts to model licensing might reshape biotech R&D and investments. We consider limitations and challenges (data biases, model validation, regulatory approval) and end with predictions about future extensions (expanding to other cancers, integrating with clinical trials, etc.). Throughout, we use an academic tone, rigorous analysis, and extensive citations to ensure credibility (www.sciencedirect.com) (natlawreview.com). The report concludes with a summary of key findings and recommendations.

Foundation Models in Pharmaceutical Research

What Are Foundation Models?

“Foundation models” (FMs) are large AI models trained on broad datasets with a self-supervised objective, then adapted to specific tasks. In natural language, GPT-4 is a paradigmatic FM: it is pre-trained on vast text corpora and can perform myriad language tasks via prompting or fine-tuning. In image analysis, DALL•E or Stable Diffusion are FMs in that sense. In biomedicine, the FM concept applies similarly: a model ingesting extensive biomedical data and yielding versatile embeddings. A key property of FMs is their *generality*: instead of building a model to predict one particular assay or target, FMs encode latent knowledge that can be leveraged in multiple downstream contexts (www.sciencedirect.com).

Unlike typical supervised models (trained on labeled examples for a narrow task), FMs often use vast unlabeled or weakly-labeled data. They may also be multimodal (incorporating text, images, molecular graphs, etc.). After pre-training, FMs are fine-tuned with task-specific data, which can dramatically reduce the need for large labeled datasets. A *Drug Discovery Today* review notes that recent FM-driven projects cover “target discovery, molecular optimization, and preclinical research” (www.sciencedirect.com).

In the pharmaceutical arena, FMs are relatively new. Many recent AI initiatives in drug R&D have followed one of two paths: (i) domain-specific models developed for a particular chemical or genomics dataset, or (ii) purely generative AI pipelines for molecule design. By contrast, **foundation models** seek to unify disparate data (genomics, imaging, electronic health records, etc.) under a single framework. For example, AlphaFold can be viewed as the first protein “foundation model” – once trained on millions of protein sequences and structures, it can predict 3D structures of almost any protein, implicitly encoding general knowledge of protein folding (www.sciencedirect.com). Noetik’s models extend this paradigm to include **cellular and tissue anatomy**.

How Foundation Models Differ from Traditional AI in Pharma

A useful contrast is with conventional machine learning approaches (Table 1). Traditional AI tasks in pharma include knowledge graphs (for linking drugs to targets, mainly using curated databases) and task-specific deep learning models (e.g. neural nets trained on binding data, or CNNs for medical images) (www.sciencedirect.com). These usually rely on supervised training with labeled data. For instance, a model might be trained to predict mRNA expression from drug exposure, or to identify tumor cells in histology. Such models work well in well-defined domains but typically cannot generalize outside their narrow training scope.

By contrast, FMs are pretrained on very large, often *un*labeled datasets. They can incorporate **self-supervised learning**, where the model creates its own training signal from patterns in the data (e.g. predicting a hidden part of the input). For example, Noetik's OCTO-VC is trained to predict gene expression in spatially-resolved cells given neighboring cells – this self-supervision requires no manual labels (www.biopharmatrend.com). Because of this, FMs can be multimodal and "physics-like" in function: they learn statistical priors over complex biological data (www.sciencedirect.com).

Table 1 (below) outlines key differences:

Aspect	Conventional AI/ML	Foundation Models (Noetik's approach)
Data Requirements	Large labeled datasets for each task (e.g. annotated images, specific assay results).	Massive unlabeled/multimodal datasets (e.g. millions of cells with spatial omics) used for pretraining.
Training	Task-specific (each model trained anew for each output).	Pre-trained on broad datasets, then fine-tuned or prompted for various tasks.
Scope	Narrow: one model for one task (predict binding affinity, detect mutation).	Broad: a single model encodes knowledge usable for many downstream experiments.
Adaptability	Low generalization beyond training domain.	High: model can be repurposed to new related tasks with minimal additional training.
Examples	QSAR models for drug likeness, ML prognosis classifiers.	Alphafold (protein FM), GPT-like models in biology, Noetik's OCTO-VC (tumor FM).
IP/Deal Structure	Often per-project contracts or licensing of a single model for an asset.	Could be licensed as ongoing infrastructure (subscription to model use).

Table 1. Comparison of conventional AI/ML models versus large foundation models in pharmaceutical research (www.sciencedirect.com) (www.noetik.ai).

Growth of Foundation Models in Drug Discovery

A concrete example of foundation-model research is Noetik's OCTO-VC, a transformer trained with masked-token modeling to predict spatial patterns of gene expression (Noetik's technical report).

This surge reflects advances in computational power and data availability. Biotech companies and academic groups are mining large-scale biological datasets (genomic, imaging, clinical) to train these models. For example, a 2025 *Nature* perspective envisions "multimodal foundation models, pretrained on diverse omics datasets (genomics, transcriptomics, proteomics, spatial profiling, etc.)" that will have "unprecedented potential

for characterizing molecular states of cells" and facilitating applications from biomarker discovery to *in silico* perturbation studies (www.nature.com). In oncology, where clinical data and spatial biology are abundant but complex, such models could identify previously hidden patterns.

Foundation models also attract pharma investment. The Noetik–GSK agreement includes $50 million in upfront capital and near-term milestones combined, plus annual licensing fees (issuer announcement). Other biotech firms (e.g. Chai, Insitro, Massive Bio, Deep Genomics) are building multi-omics FMs for drug target discovery. One industry summary lists numerous startups (Deep Genomics's BigRNA, Helical for DNA/RNA, Chai-1 for multi-modal, etc.) all pioneering bio foundation models (www.biopharmatrend.com).

Overall, foundation models are viewed both as a cutting-edge technology and an emerging asset class in biotech (natlawreview.com) (www.biopharmatrend.com). By encoding vast biological knowledge, they promise to make drug discovery more deterministic. As Noetik's executives put it, the goal is to move from "probabilistic 'shots on goal' to deterministic engineering of cancer drugs" (www.genengnews.com).

The Noetik Platform and Cancer Foundation Models

Noetik: Origins and Data Assets

Noetik Inc. (San Francisco) is a private "AI-native biotech" co-founded by Drs. **Ron Alfa** and **Jacob Rinaldi** (both previously at Recursion Pharmaceuticals) (www.genengnews.com) (www.fiercebiotech.com). The name "Noetik" (from Greek *noetikos*, "intellectual") reflects the founders' vision of applying advanced AI to understand complex biology beyond what humans can easily parse (www.genengnews.com). The company's mission is to harness high-throughput **spatial biology** data together with self-supervised learning to revolutionize cancer therapy development (www.genengnews.com).

Since inception, Noetik has focused on generating and curating **large multimodal omics datasets** from patient tumors. According to its own statements and press coverage, Noetik operates advanced spatial biology platforms (both proteomic and transcriptomic profiling) to collect *whole-tissue* data continuously. Within months of starting, they generated **hundreds of patient samples' worth of data**, each with genomic, histology (H&E), proteomic and transcriptomic measurements (www.genengnews.com). This scale of data collection – operating spatial sequencers 24/7 and automating quality control – is atypical for academia or pharma, and Noetik says it is "not [our] normal customer" for spatial services (www.genengnews.com). In effect, Noetik is creating an internal *oncology tissue atlas* to serve as AI training data (www.genengnews.com) (www.biopharmatrend.com). According to a founding press release, Noetik's platform comprises *"hundreds of millions of spatially resolved human cells"*, the largest such dataset in oncology known (natlawreview.com).

Noetik's early fundraising reflects investor excitement. After a **$14M seed round led by DCVC, announced in September 2023** (DCVC's September 7, 2023 announcement), the company attracted a **$40M Series A in Aug 2024** (venturecapital.com). This Series A (led by Polaris and others) was described as *oversubscribed*, explicitly to fund expansion of the spatial atlas and the "training of its multi-modal cancer foundation models" (venturecapital.com). In that announcement, Noetik emphasized that its labs generate spatial, high-

dimensional, multimodal data from thousands of patients (venturecapital.com). A subsequent media interview similarly notes that Noetik's teams "stood up the largest cluster in South SF" and employed 18 staff by summer 2023, underscoring rapid scaling (www.genengnews.com).

Table 2 summarizes these milestones, highlighting how Noetik has transitioned from seed to Series A to strategic partnerships (including the GSK deal).

Date	Milestone	Details
Sep 2023	Seed Financing ($14M announced)	Led by DCVC; seed announcement dated September 7, 2023 (DCVC). Funds spatial biology atlas and ML platform development.
Aug 2024	Series A ($40M)	Led by Polaris Partners, etc. To expand spatial omics atlas, scale foundation model training (venturecapital.com).
2025	Agenus Collaboration	Partnership with immuno-oncology firm Agenus (platform integration details not public) (www.fiercebiotech.com).
Jan 2026	GSK Agreement ($50M in upfront capital and near-term milestones combined, plus annual licensing fees)	Five-year licensing of spatial AI foundation models in NSCLC and CRC (natlawreview.com) (www.biopharmatrend.com).

Table 2. Major milestones and funding events for Noetik (sources: investor releases and news (venturecapital.com) (www.biopharmatrend.com)).

Technology and OCTO Models

Central to Noetik's approach is its **OCTO** platform – short for "Oncology Counterfactual Therapeutics Oracle" (www.biopharmatrend.com). Within OCTO, Noetik has developed a series of AI models and tools. Noetik's first foundation model was **OCTO**, described in its July 16, 2024 technical report. That model was trained on multiplex protein staining, spatial gene expression, DNA sequencing and H&E data, and outputs protein images from visual prompts. The subsequent **OCTO-VC** model is described in a separate December 18, 2024 technical report as a transformer trained with masked-token modeling on spatial transcriptomics to predict spatial gene-expression patterns. That report specifies a training dataset of **nearly 40 million cells** across cancer indications. Its virtual-cell predictions depend on neighboring-cell gene expression and a prompt specifying genes expressed or not expressed in the virtual cell.

The foundation for OCTO-VC is Noetik's proprietary spatial atlas. According to press materials, OCTO-VC (at launch) was *"trained on approximately 40 million spatially resolved cells spanning multiple cancer types"* (www.biopharmatrend.com). This suggests the model has encountered a wide diversity of tumor microenvironments. Moreover, Noetik built **Celleporter**, a companion visualization tool, to explore these virtual tissues (www.biopharmatrend.com). Together, OCTO-VC and Celleporter form a kind of "digital twin" system for tumors: one can simulate alterations (e.g. a new immune infiltration) and see predicted effects on the tissue.

Importantly, Noetik describes OCTO-VC as a “spatial transcriptomics model” that uses prompts and local neighborhood context to predict gene expression patterns (www.biopharmatrend.com). This architecture is analogous to how large language models use context to predict missing text, but applied to the spatial grid of cells. The model thus captures **cell-cell interactions**, particularly immune-tumor and stromal-tumor dynamics, learned from real patient samples.

Upgrading from OCTO, Noetik also refers to the entire exoskeleton as a **“multimodal world-model”** for patient biology (www.biopharmatrend.com). In practice, Noetik’s platform integrates: (1) high-density spatial DNA/RNA data, (2) spatial proteomics (immunofluorescence, imaging mass cytometry, etc.), (3) histological images, and (4) clinical annotations. By aligning these modalities, Noetik’s foundation models can, in principle, cross-reference molecular, histologic, and clinical dimensions. For example, one could input a genomic mutation plus tissue images and get a prediction of immune cell states in situ.

In summary, Noetik’s technology represents the **frontier of spatial AI in oncology**. It differs from prior AI efforts (like Insilico’s molecule generators or Vision-based pathology tools) by explicitly modeling the multi-cellular tumor architecture. In the words of Noetik’s VP of AI, Daniel Bear, these “world models” allow researchers to go *“beyond the limited data available from any one patient to ask ‘what if?’ questions about patient genes, proteins, cells, and tissue”* (www.1stoncology.com). Noetik argues that such counterfactual models will drive the next wave of discovery and therapeutic development in cancer.

The Noetik–GSK Collaboration: Deal Details and Innovation

Scope of Collaboration

On January 8, 2026, Noetik announced a five-year strategic collaboration with GSK to license its virtual cell foundation models in oncology (natlawreview.com). The **financial terms** are $50 million in upfront funding and near-term milestones (natlawreview.com) (www.fiercebiotech.com). Further, GSK will pay **annual subscription fees** to maintain access to the models (www.fiercebiotech.com) (www.biopharmatrend.com). In practical terms, GSK’s AI and therapeutics teams will get **direct, non-exclusive license** to Noetik’s **OCTO-VC models** specifically for **NSCLC and CRC research** (natlawreview.com) (www.biopharmatrend.com). GSK will also collaborate with Noetik to generate additional **bespoke spatial biology datasets** aligned to GSK’s strategic priorities (for example, tissue samples of interest, treatment conditions, etc.) (natlawreview.com) (www.biopharmatrend.com).

It is worth noting that this deal is explicitly **model licensing**, not the development of a specific drug. Noetik CEO Ron Alfa and CBO Shafique Virani underline this as a *new paradigm*: GSK is buying a *platform* rather than fore-right to a single molecule. Shafique Virani says, “We are moving the industry from AI services collaborations to licensing AI infrastructure… monetizing a biological foundation model as a scalable enterprise asset” (natlawreview.com). The **non-exclusive** nature means Noetik can license similar models to others (GSK’s ecosystem, competitors, etc.). However, GSK’s early commitment of $50M plus ongoing fees is intended to make this a “strategic anchor partnership” (natlawreview.com).

Table 3 compares this structure with more conventional pharma-biotech AI deals:

Aspect	Traditional Pharma–Biotech Collaboration	Noetik–GSK Foundation Model Partnership
Asset/Deliverable	Rights to develop a specific drug candidate or technology (e.g., antibody, small molecule, diagnostic).	License to AI **foundation models** (OCTO-VC) and related data.
Scope of Work	Focused on one or a few targets/compounds; often experimental or clinical trials.	Broad integration into R&D: target discovery, cell simulation, etc.
IP Arrangement	Typically royalties on sales of eventual drug; patents on molecules.	Noetik retains model IP; GSK pays for access/subscription.
Payment Model	Milestone payments per project (e.g., upon clinical trial advances), plus royalties.	$50M in upfront capital and near-term milestones combined; plus **annual licensing fees**. (issuer announcement)
Duration	Often 3-5 years or project-limited; may lapse if targets failed.	Five-year term with renewal likely upon success (extended licensing).
Business Model	One-time deals per asset or per project.	Ongoing platform relationship; model as **recurring infrastructure** (natlawreview.com).
Risk	High risk on single project; discontinuous if project stops.	Diversified: model can support many projects; GSK committed to underlying tech.

Table 3. Comparison of the Noetik–GSK foundation model licensing agreement with a typical pharma–biotech partnership. The key innovation is treating the AI model as an ongoing infrastructure asset, rather than a one-off drug candidate (www.biopharmatrend.com) (natlawreview.com).

Focus on NSCLC and Colorectal Cancer Models

The GSK deal specifically covers **non-small cell lung cancer (NSCLC)** and **colorectal cancer (CRC)**. In the announcement, Noetik's OCTO-VC models in these two indications are identified by name. GSK will use the models to **simulate human tumor biology** in those cancers, shopping them into its drug discovery and development pipelines (natlawreview.com) (www.biopharmatrend.com).

Why these two cancers? NSCLC and CRC are among the most prevalent and deadly. NSCLC accounts for ~85% of lung cancers and remains a leading cause of cancer death globally (pmc.ncbi.nlm.nih.gov). Colorectal cancer is similarly common worldwide. Both have active research in immuno-oncology (e.g. checkpoint inhibitors), targeted therapies, and biomarkers. GSK has existing efforts in these areas (e.g. PD-L1/PD-1 pathways in NSCLC, and EGFR/BRAF in CRC). By integrating Noetik's models, GSK aims to **improve patient stratification, target identification, and combination strategies** in these indications.

Crucially, Noetik's models are built to **simulate single-cell gene expression patterns and cell interactions** within tumors of these types. They capture the tumor microenvironment (including stromal and immune cells) at unprecedented resolution. GSK's press release emphasizes that OCTO-VC can "go beyond the limited data available from any one patient to ask 'What if?' questions about patient genes, proteins, cells, and tissue"

(natlawreview.com). For NSCLC and CRC, where immunotherapy response is variable, such “counterfactual” modeling could predict which tumor phenotypes are likely to respond to new therapies or reveal novel immuno-oncology targets.

Under the agreement, GSK also gets access to **custom spatial datasets** aligned with its interests. For example, if GSK is developing a new immunotherapy, it could have Noetik generate spatial profiles of tissue samples treated with relevant cytokines. These bespoke data would then be used to fine-tune or query the models. This co-development aspect means GSK doesn’t just license static models, but participates in expanding Noetik’s atlas in lines important to GSK’s pipeline (natlawreview.com) (www.biopharmatrend.com).

The deal’s timeline and financials are significant (Table 2). The agreement includes $50 million in upfront capital and near-term milestones combined, plus annual licensing fees; the announcement does not disclose the upfront/milestone split or a schedule for these payments (issuer announcement). This structure effectively treats Noetik’s models as *continuously maintained software assets*, akin to a biotech version of enterprise software. It reflects a bet that such models will remain valuable and improve over time.

Perspectives on the Deal

GSK’s AI leader Dr. **Kim Branson** praised the approach, emphasizing data quality: “Foundation models are only as good as the underlying training data… Noetik’s approach to generating high-quality spatial data at scale to train foundation models is novel” (www.fiercebiotech.com). Branson views integrating these models into drug discovery as a way to “deepen our understanding of biology” and support novel medicine development (www.fiercebiotech.com).

From Noetik’s side, CEO **Ron Alfa** calls this licensing “a new paradigm in biotech” – essentially confirming that Noetik sees the deal as redefining how biotech products are commercialized (www.genengnews.com). He states that with this agreement, GSK gains “one of the most extensive oncology multimodal spatial training sets in existence” and the ability to query tumor biology at “previously impossible” resolution (www.1stoncology.com). The press quotes emphasize terms like “world models”, “deterministic engineering”, “what-if questions”, highlighting the conceptual shift from simple prediction to simulation and hypothesis generation (natlawreview.com) (www.1stoncology.com).

Industry commentators also took note. BiopharmaTrend noted that this move underscores a shift from service contracts to ongoing infrastructure licensing (www.biopharmatrend.com). Tech-LifeSci (a biotech industry analysis site) frames the GSK–Noetik alliance as part of a broader “fourth wave” powered by transformers and lab-in-loop workflows (www.biopharmatrend.com). Some LinkedIn posts and biotech blogs (not directly citable here) hailed GSK’s commitment as “huge news” and evidence that commercial pharma is now financially backing AI model development. In short, multiple sources paint this transaction as signaling that **“biological foundation models” are no longer sci-fi but a viable business model** (natlawreview.com) (www.biopharmatrend.com).

Disease Context: NSCLC and Colorectal Cancer

To understand the implications of this deal, it is useful to review the current status of NSCLC and CRC research, as well as why AI modeling could be particularly valuable in these cancer types.

Non-Small Cell Lung Cancer (NSCLC)

NSCLC comprises several subtypes of lung cancer (mostly adenocarcinoma and squamous cell carcinoma) and accounts for about 80-85% of lung cancer cases globally (pmc.ncbi.nlm.nih.gov). It is a leading cause of cancer mortality; one source notes lung cancer (mostly NSCLC) is *"the majority of lung cancer cases and a leading cause of cancer-related mortality globally"* (pmc.ncbi.nlm.nih.gov). Treatment has evolved significantly: targeted therapies against mutations (EGFR, ALK, ROS1, etc.) can yield dramatic responses in molecularly selected patients, and immune checkpoint inhibitors (PD-1/PD-L1 blockers) have become standard in advanced NSCLC. However, **response rates are still limited**, toxicities can be severe, and many patients do not benefit due to complex tumor-immune microenvironments.

A critical challenge in NSCLC is **patient stratification** and understanding the tumor microenvironment (TME). For example, PD-L1 expression alone is an imperfect biomarker; tumors with similar mutational profiles can have very different immune phenotypes (inflamed, desert, excluded). Researchers have applied various AI and multi-omics approaches to tease apart this heterogeneity. Notably, a 2025 *Nature Communications* study developed an **AI-powered spatial cell phenomics** pipeline for NSCLC: combining histology (H&E) and multiplex immunofluorescence imaging, they identified 10 distinct "cell niches" within lung tumors and built a predictor of patient survival (www.nature.com). This and similar studies illustrate that **spatial multi-modal analysis can yield prognostic and therapeutic insights in NSCLC** (www.nature.com) (pmc.ncbi.nlm.nih.gov).

From a drug development viewpoint, NSCLC is crowded but still ripe for innovation. GSK has active programs including multiple immuno-oncology agents. An AI model that can rapidly simulate NSCLC tumor biology could help GSK prioritize which mutations, immune pathways, or combination strategies to pursue in trials. For example, OCTO-VC could virtually test how changing expression of a target molecule or adding an immune cell infiltrate might affect tumor behavior. This is significant because clinical and animal model testing thousands of hypotheses is impossible.

AI in NSCLC: Case Study

The Schallenberg *et al.* (2025) study in *Nat. Comm.* (www.nature.com) is an illustrative example. They took a large cohort of NSCLC patient tissues, used multiplex IF to stain for immune markers, and applied deep learning to create spatial maps of immune cell neighborhoods (e.g. identifying regions where T cells cluster). They linked these "cell niches" to survival in independent cohorts (www.nature.com). Crucially, their approach integrated multiple data types (image plus molecular markers) with AI. This is similar in spirit to Noetik's

approach – though Noetik does it at a much larger scale and in silico. Such studies lend credence to the idea that high-dimensional AI models, trained on spatial biology, can uncover clinically actionable patterns in NSCLC.

Colorectal Cancer (CRC)

Colorectal cancer is among the most common cancers worldwide, with rising incidence in younger populations. While early-stage CRC can often be cured surgically, advanced/metastatic CRC still has limited options. Some patients benefit from therapies targeting KRAS, BRAF mutations or MSI-high tumors (e.g. pembrolizumab for MSI-high CRC), but again many patients' tumors are biologically complex. The tumor microenvironment in CRC includes not only tumor cells, but also cancer-associated fibroblasts, endothelial cells (blood vessels), and various immune cell subsets. Unlike NSCLC, CRC has a substantial known link to microbial factors and inflammatory signatures as well.

AI and multi-omics are also active in CRC research. For example, a 2024 *Cell Reports Medicine* study used integrated multi-omics and deep learning on CRC samples to identify a "macrophage-oriented immune module" that predicts chemotherapy response (pmc.ncbi.nlm.nih.gov). They constructed a CRC-CCIM (CRC immune module) of FOLR2^+ macrophages and exhausted T cells and showed that an AI model ("CCIM-Net") could predict which patients respond to standard chemo based on spatial interaction patterns (pmc.ncbi.nlm.nih.gov). This illustrates that data-driven spatial modeling can yield testable biological targets in CRC.

Another recent study (Trahearn *et al.*, 2025) developed a computational pathology pipeline for CRC tissues (pmc.ncbi.nlm.nih.gov). They trained a deep learning classifier on H&E images to identify 8 cell types (including tumor, immune, fibroblast) and then generated spatial maps of cell distributions. These maps revealed patterns (e.g. tumor–immune proximities) that were predictive of patient outcomes (pmc.ncbi.nlm.nih.gov). This shows how standard histology, when combined with AI, can indirectly capture spatial biology without specialized staining – a concept that Noetik generalizes with rich molecular data.

GSK's interest in CRC likely reflects the drug giant's role in colon cancer R&D. By licensing Noetik's CRC models, GSK could better understand which CRC subtypes might respond to certain therapies or identify novel biomarkers. Given the adaptive complexity of CRC, a holistic foundation model could theoretically unify data from genomics, histology and clinical outcomes to improve trial design.

Future Potential in NSCLC and CRC

For both NSCLC and CRC, the overarching promise is to move **beyond snapshots**. Noetik's models allow asking *"what if we moved a cell or changed a gene?"* across an in silico tumor, something intractable by lab experiments alone. If GSK can validate that OCTO-VC predictions correlate with clinical or experimental results (e.g. if model-suggested targets indeed influence outcomes in mouse models or patient subsets), it will set a precedent for using such FMs routinely.

However, caution is warranted: these models are novel and must be rigorously tested. The literature on LLMs warns that huge models can invent plausible-sounding but false statements. Similarly, a foundation model of biology could make confident predictions that need careful validation. The press release itself acknowledges

this shift: “foundations models narrow uncertainty” (www.sciencedirect.com), but only if built on quality data. Practitioners will want to see how well simulated gene expression matches real lab measurements, and whether virtual experiments hold up in vivo. We discuss these validation considerations later.

Noetik–GSK Deal in Context: Business and Industry Perspectives

Commercial Innovation: Licensing AI Infrastructure

The $50M Noetik–GSK deal stands out in current biotech landscape. Historically, large pharma–startup collaborations typically focus on **rights to drug candidates**. For instance, a Big Pharma might pay upfront for an option on a small molecule series or an antibody in preclinical development, with milestone payments tied to development progress, and royalties on sales (www.biopharmatrend.com). AI partnerships have been more modest: prior agreements often involved AI startups providing specific analytics or predictions on a project-by-project basis, rather than licensing a full platform.

Here, GSK is effectively **paying for the right to use an AI platform**. BiopharmaTrend noted that GSK’s structure emphasizes licensing the model rather than commissioning a single use case (www.biopharmatrend.com). GSK will pay annually to keep these models in-house, analogous to a software subscription. This suggests that GSK expects continuous value extraction – the models can be reused across multiple programs and updated over time.

This model of “AI-as-infrastructure” is being talked about as a sea-change. By treating the model as a persistent asset, Noetik and GSK hope to create a **long-term partnership**, not a one-off project. Noetik’s Chief Business Officer, Shafique Virani, explicitly contrasted this with older deals: previously, biotech might provide AI services (e.g. consultancy or per-project modeling), but now GSK is licensing the *infrastructure itself* (natlawreview.com). If the model proves valuable, GSK can use it to “validate Noetik’s platform as a scalable, revenue-generating engine” (natlawreview.com).

From GSK’s viewpoint, this is a bet that **early, deep integration of AI will yield strategic advantage**. The press release quotes suggest GSK sees this as a way to “deepen understanding of biology” in their pipeline (www.fiercebiotech.com). It aligns with reports that big pharma companies have re-opened their AI initiatives after a slight lull – they want to capitalize on GenAI momentum. For example, other pharma have made headlines with AI tie-ups (e.g. Novartis/Schrödinger, GSK/Insitro, AstraZeneca/Regeneron). The Noetik deal is distinguished by focusing on generative biological models rather than just software tools or chemical AI.

Investors also view this as positive signal. Noetik’s $50M plus subscription is large for a pre-clinical-stage startup (no actual drugs from the platform yet), implying confidence that foundation models are commercially legitimate. Market observers might note that deals like this could transform startup valuations; instead of “X-yield biotech with candidate pipelines”, “X-yield AI platform with scalable licenses” becomes a model. This is why some called it a “big signal moment for TechBio” (www.linkedin.com).

Industry Trends and Comparisons

Noetik's deal ties into several broader trends:

- **Data-centric pharma**: There is growing emphasis on using real-world data and complex biological data (e.g. spatial omics, multi-omics) to guide R&D. Large pharma have data science teams and partnerships with data platforms (like Tempus, Flatiron). Noetik contributes a cutting-edge example of this, aligned with that trend.
- **AI vendor ecosystem matures**: With hundreds of academic and startup models now, pharma is considering which platforms to trust. If Noetik's models outperform others in discovery, competitors (e.g. Roche, J&J) may seek similar deals. Already, Noetik's press release suggests this is one of the "first and largest" such deals, implying there may be more in future (natlawreview.com).
- **Digital twin / virtual patient**: Long-term, pharma has dreamed of "clinical trial in silico". Noetik's talk of patient-specific modeling edges in that direction. Other groups (like the Human Cell Atlas, or deep learning agents) are also working on *in silico* patient models. The concept of a virtual twin is becoming scientifically credible, and such foundation models are prototypes of that idea.
- **Regulatory interest**: FDA's January 2025 draft guidance already proposes a risk-based credibility assessment framework for AI-generated information supporting regulatory decisions about drug safety, effectiveness or quality. It is nonbinding draft guidance, not for implementation. Its scope excludes AI used solely for drug discovery. EMA's reflection paper, adopted in September 2024, addresses AI across the medicinal-product lifecycle. Regulatory relevance depends on the model's intended use and the role of its outputs in evidence submitted for review.

One can also contrast this with other high-value tech deals. For example, Graphcore and Intel have deals to supply AI hardware; Apple/Google license LLM engines; but in biotech it is novel. There is an environmental analogy: Noetik's models for biology are like climate model simulations for earth – and Big Pharma is sponsoring a "supercomputer" of cancer biology.

Critiques and Considerations

Not everyone is unreservedly optimistic. Some skeptics might wonder: *will the models truly work as promised?* The data science community is aware that large models can underperform if domain mismatch exists or if training data is insufficient. Key questions include: Are the Noetik models transparent or interpretable? How will GSK validate model outputs? Will predictions truly correlate with clinical outcomes? In Foundation-model hype elsewhere (e.g. text), issues of "hallucination" are known; analogously, "hallucinated biology" is a concern.

Furthermore, we must consider practical limits. Noetik's models are **non-exclusive**, meaning competitors (e.g. Merck, Bristol Myers, etc.) could on principle license the same models. GSK's competitive edge will depend on how effectively it integrates and validates these tools. The subscription model implies ongoing payments; if the scientific value does not materialize, those fees could become sunk costs.

However, Noetik and GSK appear mindful of this: the collaboration includes “bespoke data generation” (natlawreview.com). In effect, GSK will co-generate more data to fine-tune the models, increasing confidence that model predictions are relevant. It is a risk-sharing mechanism: GSK pays to shape the data space where the models are most needed, thus effectively guiding model development toward its pipeline needs.

Technical and Data Perspectives

While the business and clinical implications are paramount, it is also critical to examine the technical aspects of Noetik’s models and data:

- **Model Validation**: Has Noetik published validations of OCTO-VC? The press releases focus on capabilities, not quantitative performance. Independent validation would involve comparing model predictions (e.g. virtual gene expression profiles) against held-out lab data. For foundation models, one expects they are evaluated like any AI: e.g. correlation or error metrics on test data, and biological plausibility. These specifics are proprietary, but future publications or collaborations with GSK may yield results.
- **Data Quality and Biases**: The models’ utility depends on the underlying data representing diverse patient populations and tumor subtypes. Noetik’s dataset is large, but biases may exist (e.g. over-representation of certain ethnicities or cancer stages). GSK should ensure that spatial samples cover the diversity relevant to its therapy targets. If foundation model is skewed, predictions could favor one subgroup.
- **Case Study – Spatial Data Simulation**: A pertinent question is: can OCTO-VC simulate perturbations not present in the training data? For example, if a new drug reduces expression of PD-L1, can the model predict downstream effects on T-cell infiltration? Theoretically yes, if the model has learned underlying dependencies, but this is a frontier of inference. It is akin to asking an LLM a question outside its training; success may vary.

One might analogize to existing “virtual pathology” models: A recent *Nat Med* paper trained an AI to predict protein expression from standard histology in lung cancer (www.nature.com). These models can often replicate known markers under diversity. Similarly, OCTO-VC’s strength is multi-dimensional, not tinted by a single label.

- **Computational Infrastructure**: Training on ~40–100 million cells with transformers and multi-modal data is computationally intensive. Noetik must have deployed substantial GPU/TPU resources and memory. GSK will likely access the model via cloud or secure API. The collaboration may involve building integration pipelines: for instance, feeding GSK’s internal RNAseq or pathology data into OCTO-VC for analysis.
- **Data Analysis Integration**: GSK’s researchers will need new workflows. Interpreting model outputs (virtual tissue maps, counterfactual scenarios) requires bioinformatic skill. It may blur the line between computational biology and bench experiments. Possibly GSK will form joint teams or upskill scientists to "interrogate the model". This could lead to internal knowledge transfer, which itself is valuable.

Case Studies and Real-World Analogues

Several contemporary research efforts offer a preview of how Noetik-like models can be used in practice. These examples – while not Noetik's products – demonstrate similar principles of spatial AI in oncology:

1. **NSCLC Spatial Phenotype Mapping (Nature Comm, 2025)** (www.nature.com). As noted earlier, Schallenberg *et al.* used AI to classify the spatial organization of cells in lung tumors. They integrated histology and multiplex staining to identify prognostic "cell niches" and built a survival predictor. This project required co-registering multiple imaging modalities and applying deep nets; it illustrates the power of capturing patterns invisible to the eye. If OCTO-VC had been available, they might have run *in silico* experiments, e.g. "predict how the niche composition changes if immune cells are added."
2. **CRC Immune Module Modeling (Cell Reports Med, 2024)** (pmc.ncbi.nlm.nih.gov). Bao *et al.* integrated single-cell RNA-seq, spatial IHC, and clinical data to define a macrophage-centered immune signature in CRC. They constructed "CCIM-Net," a deep learning model that uses spatial features (distance statistics, cell counts) to predict chemo response (pmc.ncbi.nlm.nih.gov). Notably, they demonstrated that an AI-identified target (FOLR2+ macrophages) could be validated in vivo. This pipeline – multi-omics discovery → spatial mapping → AI model → experimental validation – is a prototype for how Noetik's model could generate hypotheses for GSK (e.g. highlighting new target subtypes).
3. **Computational Pathology in CRC (The Journal of Pathology, 2025)** (pmc.ncbi.nlm.nih.gov). The Trahearn *et al.* study showed that even standard H&E slides, when processed with AI, can yield biologically meaningful spatial maps of cell types (pmc.ncbi.nlm.nih.gov). They trained a convolutional neural network to label cells, then mapped where tumor, immune, fibroblast cells co-occur. This "ecological map" approach disclosed novel prognostic features. For GSK researchers, it highlights that even without multi-omics assays, AI can extract spatial biology from routine data. Noetik's models, by contrast, start from richer inputs. In a GSK workflow, one could imagine using OCTO-VC to predict how the spatial TME appears in various genomic subtypes (information that would take weeks to gather from tissue).
4. **Histology–Transcriptomics Integration (Nature Methods, May 29, 2025)**. Chen *et al.* introduced **OmiCLIP**, a foundation model linking H&E images with transcriptomic data from Visium tissue patches (original study).

These case studies indicate that (a) spatial and multi-modal models are scientifically productive and (b) AI can find clinically relevant patterns. The GSK–Noetik collaboration essentially brings this cutting-edge research capability in-house at GSK.

Implications, Challenges, and Future Directions

Scientific and Clinical Implications

If successful, the Noetik models could have several impacts on drug R&D:

- **Target Discovery and Prioritization**: By simulating thousands of tumor states, GSK could identify novel targets that emerge as key regulators in silico. For example, a gene that the model predicts will shift immune infiltration patterns might become a therapeutic candidate.
- **Patient Stratification and Biomarkers**: The models could suggest new biomarker signatures. GSK might find that virtual tumor profiles cluster into subtypes that correlate with outcomes, guiding clinical trial design (e.g. selecting patients likely to respond to a new drug).
- **Mechanistic Insight**: Researchers could perturb the virtual cell network to study mechanisms. For instance, "knock out" a cell receptor in the model and see predicted effect on cytokine gradients, generating hypotheses about resistance mechanisms.
- **Reduced Experimentation Cycle**: By answering "what if" questions computationally, the models could reduce the number of wet-lab experiments needed. This accelerates iteration: teams can run simulations overnight rather than months of cell culture or animal studies.
- **Generative Biology**: In the future, one might imagine using generative AI to propose entirely new therapeutic strategies. For instance, combining OCTO-VC with a generative molecular AI: ask the model what drug best shifts a tumor from one state to another. (This is speculative but illustrates potential synergy.)

Might some of these ideas make it into peer-reviewed science? Possibly. We expect GSK and Noetik may publish results if the collaboration yields noteworthy discoveries. The *Nature* Perspective on multimodal foundation models identifies biomarker discovery, gene regulation inference and *in silico* perturbations as potential applications of context-specific transfer learning (Perspective).

Risks and Validation

While promising, there are challenges:

- **Validation**: The performance of a foundation model must be vigorously tested. In conventional drug R&D, models are validated at each step (e.g. how well a candidate binds target). For AI models, analogues would include: holdout tests (predicting data not seen during training) and correlation with real experiments. For example, if OCTO-VC predicts a protein distribution in a hypothetical cell mix, one could verify by targeted imaging. GSK will likely design validation studies early on.

- **Generalizability**: Models trained on one dataset may not generalize to all contexts. If Noetik's training was mostly on, say, early-stage tumors, will it correctly model advanced metastatic sites? Part of GSK's control is the ability to augment data: if gaps are found, they can obtain relevant samples to retrain or fine-tune the model.
- **Interpretability**: Large AI models are often black boxes. Being able to interpret what the model has learned biologically will be important for adoption. Techniques like "attention mapping" (seeing which inputs influence an output) might be used. GSK scientists may demand explainable AI, especially if regulators later inquire how models were used.
- **Regulatory and Ethical**: Currently, no drug approval depends solely on an AI foundation model. However, if GSK uses model-derived biomarkers for patient selection, that could affect trial conduct. FDA has already issued nonbinding draft guidance on AI supporting drug and biological-product regulatory decisions, and EMA has adopted a medicinal-product-lifecycle reflection paper. Applicability depends on the intended use: FDA's draft scope excludes AI used solely for drug discovery. Ethical concerns include patient data privacy (though Noetik's data is de-identified spatial data) and transparency about AI involvement in decisions.

Future Directions

Beyond NSCLC and CRC, the model deals could extend to other areas: Noetik may train OCTO-VC variants for breast, prostate, or melanoma if capital and partnerships allow. The concept of foundation models could also apply to other disease domains (e.g. neurological diseases with single-cell brain atlases).

In 5–10 years, we may see an ecosystem of biomedical foundation models: tissue-specific, disease-specific, or even global models of human biology. Companies like Noetik might offer a portfolio of such models. Pharma might subscribe to platforms (like cloud-based AI services) rather than develop their own models from scratch. Already Noetik's site hints at expanding: their homepage mentions "one of the largest collections of multimodal tumor data anywhere on Earth" used to train models (www.noetik.ai).

The **economic implications** are profound. If foundation models prove capable, valuation of biotech companies may hinge on their data and models more than on drug pipelines. Business models will change: revenue from software licenses may come alongside or instead of product sales. Investors and executives will have to become literate in AI technology to manage these assets.

On the science side, a successful collaboration may spur more open-data initiatives: if training data size is critical, there may be pressure for sharing spatial omics datasets (subject to privacy) to benefit model training. Initiatives like the Human Tumor Atlas Network (NIH) may dovetail with companies like Noetik. Academic labs might increasingly collaborate, contributing patient samples in exchange for model access.

Finally, a "social implications" perspective: these models blur the line between virtual and real. As Noetik says, "the data we generate is not designed for human interpretation, but rather [for] training data for foundation models" (www.noetik.ai). We are outsourcing some of our biological intuition to machines. This shift will require new skill sets and mindsets among biologists and clinicians.

Conclusion

The Noetik–GSK deal, which includes $50M in upfront capital and near-term milestones combined plus annual licensing fees (issuer announcement), is a landmark in biotech: it is among the first major agreements to treat AI models trained on human biology as licensable infrastructure. By focusing on NSCLC and colorectal cancer, GSK is betting that sophisticated foundation models can unlock new insights in two of the largest oncology markets. The collaboration synthesizes several cutting-edge threads: spatial omics data, self-supervised learning, multi-modal foundation models, and AI-as-a-service business models.

Our analysis shows that this approach is grounded both in current science and in ambitious vision. Dozens of studies academically and industrially have demonstrated the power of AI on spatial oncology data (www.nature.com) (pmc.ncbi.nlm.nih.gov), and Noetik's scale brings unprecedented data and computational muscle. The deal structure – especially the recurring license fees – indicates an expectation that the models will continually improve and be widely used internally at GSK (www.biopharmatrend.com).

However, many unknowns remain. Will the models' predictions translate into real-world drug successes? Can biases or blind spots in the training data be identified and mitigated? Will other pharma companies replicate this model, and how will partnering terms evolve? These questions underscore that while the promise is great, proof lies in future outcomes.

What is clear is that **Noetik and GSK are trailblazing a new paradigm** in pharmaceutical R&D. If foundation models deliver on their potential, they could make the drug discovery process faster, more efficient, and more data-driven. This in turn could accelerate the development of novel therapies for cancers like NSCLC and CRC – leading to better outcomes for patients. Regardless of details, this deal signals that AI – particularly spatial, multi-omic foundation models – is now a core part of the strategic playbook for major pharma companies.

Table 4 below contrasts this approach with conventional AI/ML strategies, and **Table 2** (above) summarizes Noetik's key milestones. We conclude with the overarching view that GSK's commitment of $50M in upfront capital and near-term milestones combined, plus annual licensing fees, has likely set a new industry benchmark: transforming AI models into monetizable biotech infrastructure. The implications will unfold in coming years, with the potential to reshape how we understand and treat cancer.

Table 4. Comparison of AI Approaches in Oncology R&D (illustrative).

Approach	Description	Applications	Advantages	Limitations
Traditional ML	Models trained on labeled data for specific tasks (e.g. supervised CNN for image classification, QSAR for molecular properties).	Predicting target–drug affinity, image-based diagnosis, basic biomarker classification.	Well-understood, interpretable pipelines; effective on narrow tasks with good data.	Limited generalization, requires large labeled datasets, often single-modality.
Generative Models	Neural networks that generate new samples (e.g. GANs, VAEs) or optimize molecules.	Designing new small molecules, antibodies via reinforcement, simulating synthetic data.	Can propose novel structures or scenarios, reduce manual chemistry iteration.	Often ignore full biological context; results must be validated in wet lab.

Approach	Description	Applications	Advantages	Limitations
Knowledge Graph AI	Integrates structured biomedical databases into graph form; uses ML for inference.	Drug repurposing, target identification via network analysis.	Leverages existing curated knowledge, good for hypothesis generation.	Focuses on known networks; may miss novel interactions not in databases.
Foundation Models	Large AI models pre-trained on vast multimodal biomedical data (genomics, imaging, etc.) via self-supervised learning.	Holistic modeling of disease (e.g. virtual cell simulations), advanced biomarker discovery, in silico perturbation studies.	Highly generalizable; can answer diverse “what-if” questions; improves with more data.	Require enormous high-quality data, complex to train/interpret; novel risk of unpredictable outputs.
Noetik’s Spatial AI (OCTO)	A specialized foundation model trained on spatially-resolved tumor omics.	Simulating tumor microenvironment in NSCLC/CRC; matching targets to patient subgroups; predicting therapy responses.	Context-rich predictions (accounts for cell neighborhoods); built on real human tumor data.	Cutting-edge and experimental; performance still to be clinically validated.

Table 4. AI methodologies in oncology drug discovery and development, comparing traditional machine learning with the new generation of foundation models (adapted from reviews (www.sciencedirect.com) (pmc.ncbi.nlm.nih.gov)).

Source links

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

1. https://www.1stoncology.com/blog/gsk-licenses-noetik-s-ai-foundation-models-in-anchor-partnership-to-transform-cancer-therapeutic-researchdevelopment1234661874
2. https://natlawreview.com/press-releases/gsk-licenses-noetiks-ai-foundation-models-anchor-partnership-transform
3. https://www.biopharmatrend.com/news/gsk-licenses-ai-based-virtual-cell-models-from-noetik-in-five-year-oncology-research-agreement-1446
4. https://www.fiercebiotech.com/biotech/gsk-inks-model-deal-50m-bet-noetiks-cancer-ai-platform
5. https://www.sciencedirect.com/science/article/pii/S1359644625002314
6. https://www.nature.com/articles/s41467-025-65783-z
7. https://pmc.ncbi.nlm.nih.gov/articles/PMC10897549
8. https://www.noetik.ai/octo
9. https://www.nature.com/articles/s41586-025-08710-y
10. https://www.sciencedirect.com/science/article/pii/S1359644626000607
11. https://www.genengnews.com/topics/drug-discovery/spatial-biology-and-ai-combined-by-noetik-to-develop-immuno-oncology-drugs
12. https://pmc.ncbi.nlm.nih.gov/articles/PMC12806139
13. https://www.noetik.ai
14. https://www.noetik.ai/octo-vc

15. https://www.biopharmatrend.com/post/724-7-companies-building-foundation-models-in-biology-and-chemical-synthesis
16. https://www.dcvc.com/news-insights/dcvc-leads-14-million-seed-round-to-support-noetik
17. https://venturecapital.com/headlines/funding-events/noetik-secures-40m-series-a-funding-for-advancing-aidriven-cancer-therapeutics-v1
18. https://pmc.ncbi.nlm.nih.gov/articles/PMC13085305
19. https://pmc.ncbi.nlm.nih.gov/articles/PMC11717494
20. https://www.linkedin.com/posts/alexmerwin_techbio-aiinbiotech-foundationmodels-activity-7424529466397442048-UPk1
21. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
22. https://www.fda.gov/media/184830/download
23. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
24. https://www.nature.com/articles/s41591-025-04060-4
25. https://www.nature.com/articles/s41592-025-02707-1
26. https://www.noetik.ai/foundation-models

THE PUBLISHER

About IntuitionLabs

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

AI consulting and adoption

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

Software, data and life-science workflows

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

Enterprise platforms and regulated delivery

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

Work with IntuitionLabs

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

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

Website: https://intuitionlabs.ai

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