The State of AI in the Biopharma Industry

Bio
Pharma
Trend

www.BiopharmaTrend.com | December 2023

Table of Contents

Executive Summary

AI in Pharma and Biotech at a Glance

- Artificial Intelligence (AI) is being adopted at every stage of drug development value chain. Impact varies dramatically from one use case to another, with the some of the quickest ROI expected in the area of biomarker-led clinical trials and clinical trial design and management.

- The latest focus is on the incorporation of generalized foundation models, with further fine tuning to domain-specific applications.

- End-to-end AI platforms provide a clear reduction in cost and time of preclinical research. Improvements in the success rate of drug approvals is still to be seen in the coming years.

AI Across R&D Value Chain

- AI enables a shift towards holistic drug discovery (e.g. target-agnostic phenotypic strategies, etc). The key areas of AI adoption include disease modeling and hypothesis creation, target discovery, drug design, ADME and toxicity prediction, in vivo modeling and improving alternative in vitro models (e.g. organs-on-a-chip). AI enables quite accurate prediction of clinical trial successes, in some specific cases reaching up to 79-86% accuracy. AI integration into translational and clinical research allows to improve clinical trial design, select likely responders, and improve diversity of clinical trial participants.

- AI is starting to make an impact in the biomanufacturing process, allowing to go from automation to intelligence process control. However, we are in early days of this transition. The industries shift towards complex biologic modalities will dictate the increasing need to move to AI-supported biomanufacturing processes.

Executive Summary

AI Across Therapeutic Modalities: SMs, Proximity Inducers, Biologics

- While AI-startups were traditionally focused on small molecules (still more than 40% of the total pipeline-focused companies in our database), there is a major shift towards biologics design. This includes notable technological advances in the design of antibodies, mRNA vaccines, other protein modalities, RNA/DNA-based therapeutics, and other modalities. AI is also extensively applied to model and design proximity inducing modalities, including PROTACs, molecular ‘glues,’ LYTACs, and other similar modalities.
-
- AI is extensively used in synthetic biology research, including CRISPR-based gene editing, gene sequence optimization, protein design, and metabolic pathway engineering.
-

AI Trends, Convergence with Other Tech

- AI appears to be a foundational technology, improving not only core processes of biology research and drug design, but pretty much every data-rich area of life sciences, engineering, and instrumental methods. For instance, AI-enabled Cryo EM allows to gain unprecedented insight into structural aspects of biological systems.

Executive Summary

Beyond Hype: AI Reality Check in Drug Discovery

- While AI progress in drug discovery, development, and clinical research is truly transformative, we are still in early days of the full technology adoption. While there are dozens of AI-inspired drug candidates in clinical trials as of today, some of the early clinical candidates have failed this year.

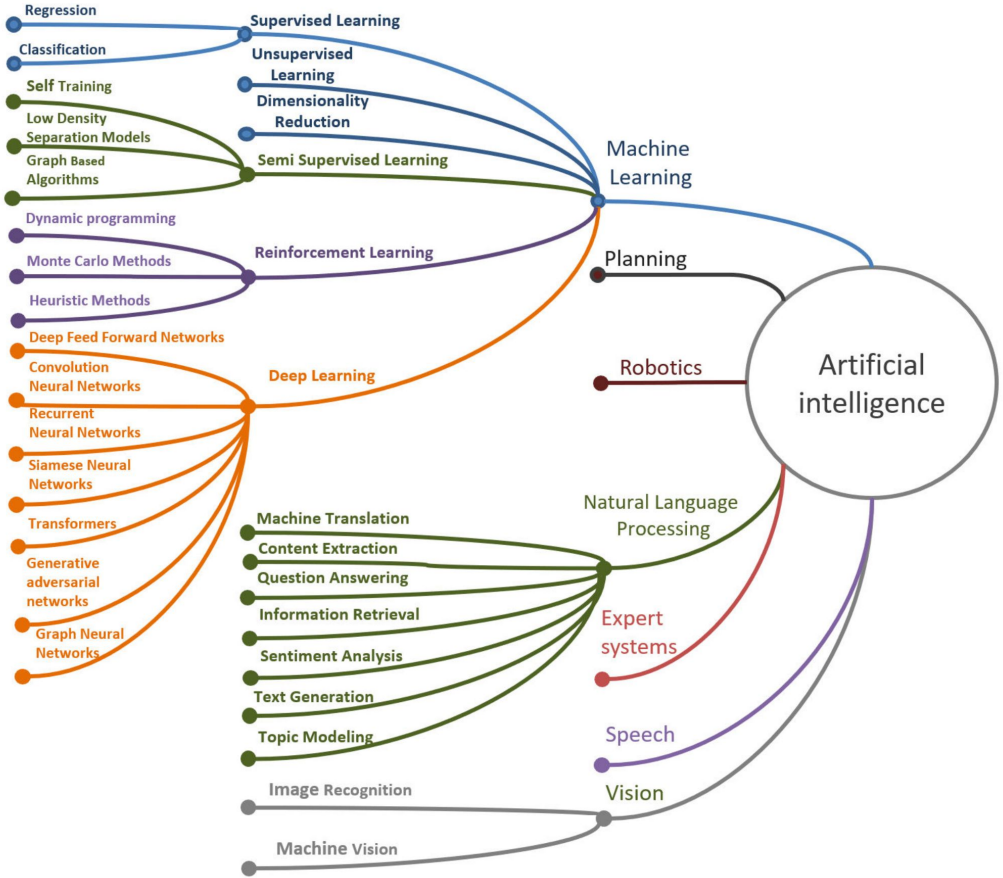
- This raises many questions, including: ‘how to determine the role of AI in the actual drug candidate selection process,’ how to assess the AI value over legacy approaches in quantitative terms, how to figure out new business models which would be aligned with an AI-centric research paradigm.

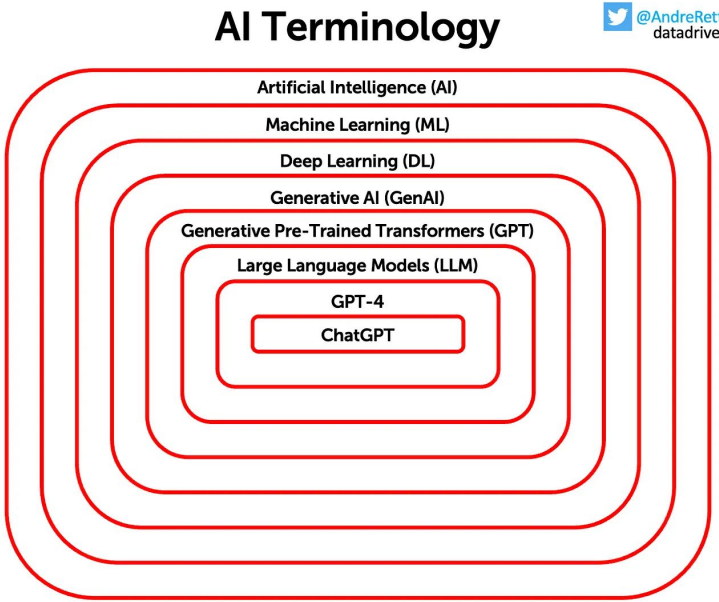
Conclusions

- AI is one of the most transformative technologies which is already reshaping the pharmaceutical industry and biotech. However, we are in early days, and a lot of work still remains in terms of education, cultural shift, organizational transformations, and most importantly, generating better data at scale to enable truly precise deep learning models of tomorrow.

INTRODUCTION

What is Artificial Intelligence?





In the context of the pharmaceutical and biotechnology industries, Artificial Intelligence (AI) refers to advanced computational technologies that aid in drug discovery, personalized medicine, patient data analysis, and the optimization of research and development processes by mimicking human cognition and decision-making [1]

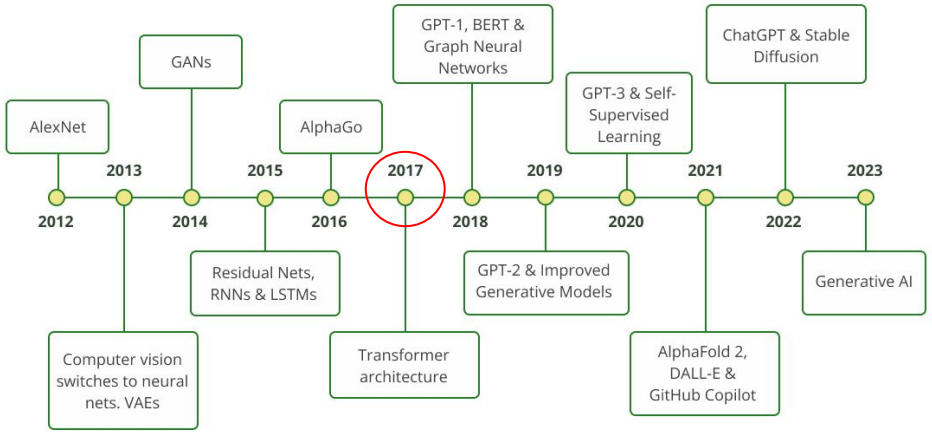
What is Artificial Intelligence?

Artificial intelligence is a relatively old concept, formalized at a famous Dartmouth College conference in 1956. The AI technologies in drug discovery have evolved from earlier machine learning (ML), cheminformatics, and bioinformatics concepts and approaches. For example, the application of machine learning to developing quantitative structure-activity relationship (QSAR) models and expert systems for toxicity prediction has a long history.

However, the rapid (in some cases -- “exponential”) advent of big data, advanced analytics, minimizing the cost of computation, GPU acceleration, cloud computing, algorithm development (e.g., deep neural nets and large language models), and the “democratization” of AI technology -- all led to a synergistic “boom” in commercializing and industrializing artificial intelligence, in particular, in the pharmaceutical and biotech industries.

The high sophistication and automation of some AI platforms led to their “commoditization” to the point they have trademarked commercial names. At the same time, some of them are offered as software-as-a-service to other companies. Examples include Pharma.AI by **Insilico Medicine**, mRNA DESIGN STUDIO™ by **Moderna**, CentaurAI® by **Exscientia**, Guardian Angel™ by **AI Therapeutics**, ConVERGE™ by **Verge Genomics**, Taxonomy3® by **C4X Discovery**, and many others.

Before 2017: the era of predictive AI. After 2017: the era of generative AI

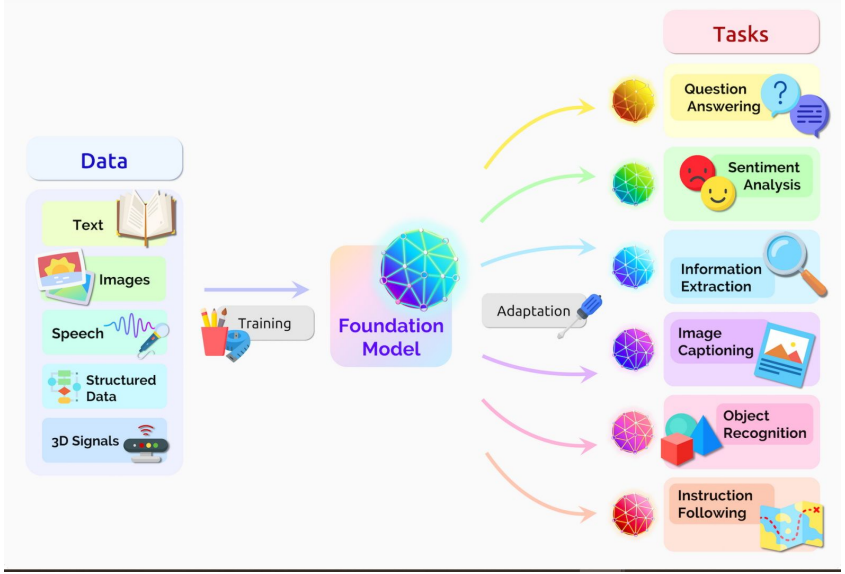


2017 was a significant year for the progress of AI, especially in the domain of deep learning and natural language processing. The introduction of the Transformer architecture and the attention mechanism, as detailed in the "Attention Is All You Need" paper by Vaswani et al., marked a paradigm shift.

Shift from predictive AI and relatively small models trained on labeled data, to large and ultra-large models (foundation models) trained mostly on unlabelled data via self-supervised learning, and other unsupervised methods.

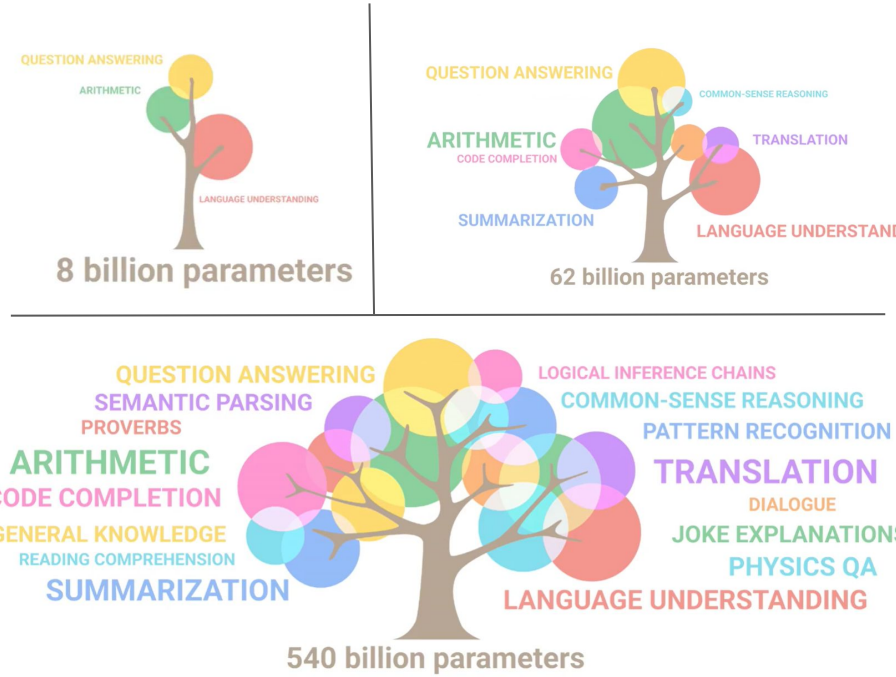
Generative AI, Foundation Models

Generative AI is a type of artificial intelligence focused on creating new content, such as text, images, and music, based on learned patterns from existing data.
It is closely connected to **foundation models**, which are large, pre-trained models (like GPT-3 or DALL-E) that provide a broad understanding and ability to generate diverse content, serving as the underlying structure for more specialized generative AI applications.



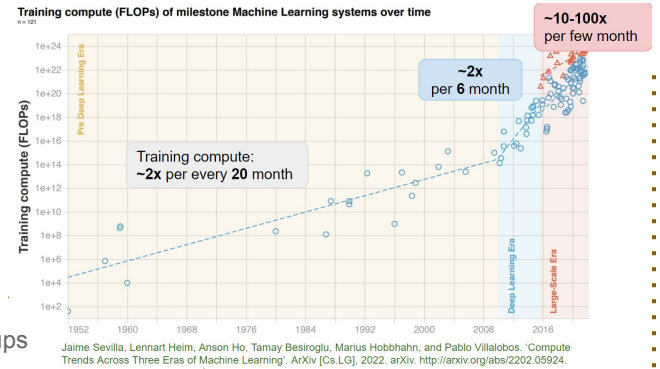
OpenAI Google ANTHROP\C Meta

The Phenomenon of 'Emerging Abilities'



Emerging abilities in AI refer to the phenomenon where an AI system develops new, often unexpected, skills or behaviors as it processes more data or interacts with its environment, going beyond its initial programming.

The Five Pillars of AI



Synthetic data startups

EHR vendors, and other healthcare data providers

Databases (protein database, ChembL, etc)

Biobanks (specimens, tissues, blood samples)

Robotized labs to generate data at scale (Arctoris, Recursion Pharma, Insilico Medicine, Cultura, Strateos Labs, Synthace

Consumer gene testing (23andMe, etc)

Big data

Computing

Cloud Infrastructure and Frameworks

Microsoft Azure,

AWS

Google Cloud

Palantir , etc

KNIME

Frameworks for building AI models.

Building Blocks of Modern AI

Specialized AI Hardware

NVIDIA - GPUs for molecular modeling, genomic data processing, and image analysis.

Google - TPUs (custom-built hardware accelerators) to speed up and scale up machine learning workloads

Intel - Xeon CPUs and FPGAs (Field-Programmable Gate Arrays) for HPC

Algorithms

Alpha Fold (DeepMind)

Foundation models for RNA (Deep Genomics)

ML algorithms for generating novel sequences for proteins (Generate Biomedicines)

ML algorithm for ligand-based de novo drug design: Diffusion Models (Iktos)

RNA modeling (RosseTTA)

DL Guardian Angel™ Algorithm (AI Therapeutics)

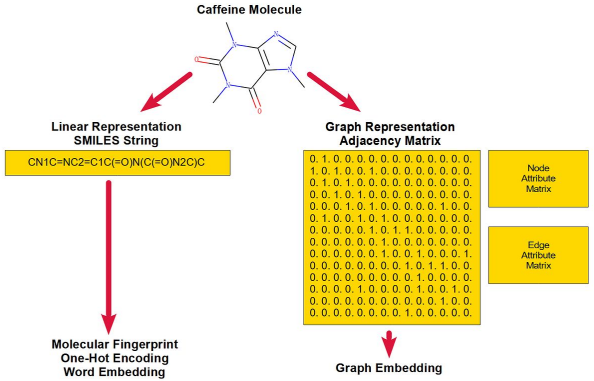
LLM model for protein structure prediction (META, facebook)

GAN generative models (InSilico Medicine)

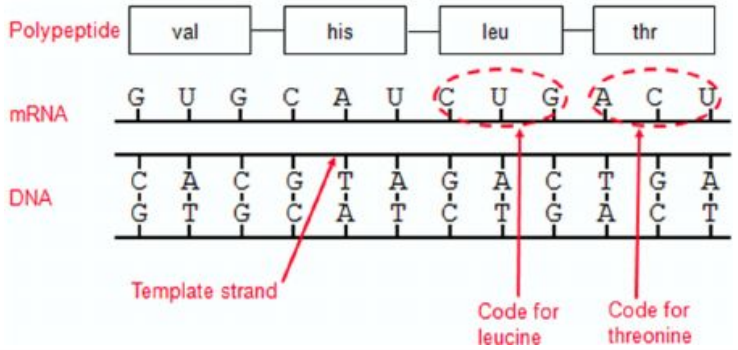
Transformers and attention algorithms, LSTM, etc.

AI in Life Sciences: Understanding Language of Chem and Bio

Language of Chemistry [2]



Language of Biology



Chemoinformatics and Molecular Modeling

Predicting drug-likeness, solubility, permeability, and toxicity.
Predicting how drug molecules will interact with biological targets.

Generative Models

GANs and VAEs - to design novel drug molecules Reinforcement learning is sometimes combined with generative models to optimize molecules for specific objectives.

Quantum Mechanics and Molecular Dynamics Simulations

Provide detailed insights into molecular interactions.
Offers more efficient outcomes than traditional methods.

Genomics, Transcriptomics, Proteomics and Protein Engineering

Genetic basis of diseases. Predicting protein folding, protein-protein interactions, and identifying potential sites for drug binding or modification.

Cell and Histology Image Analysis

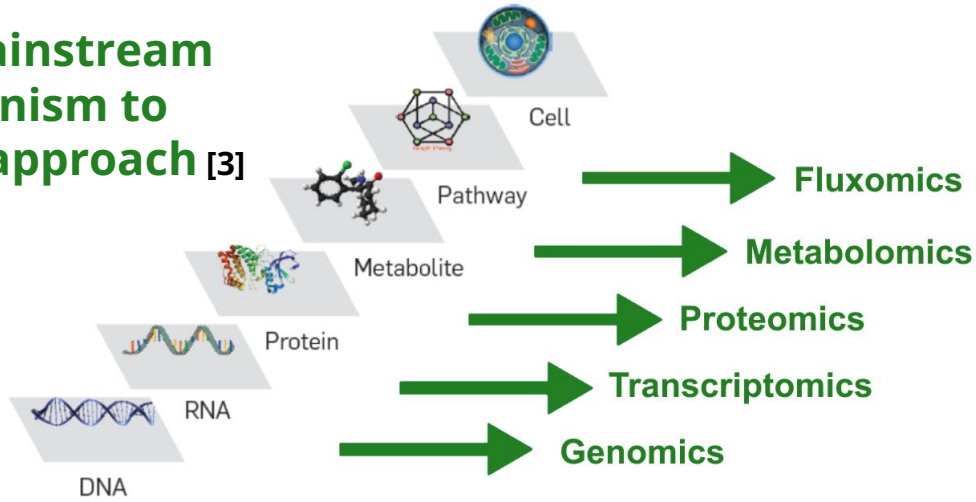
Identifying phenotypic changes in response to drugs from the high-content imaging data. Facilitating the understanding of drug mechanisms and effects.

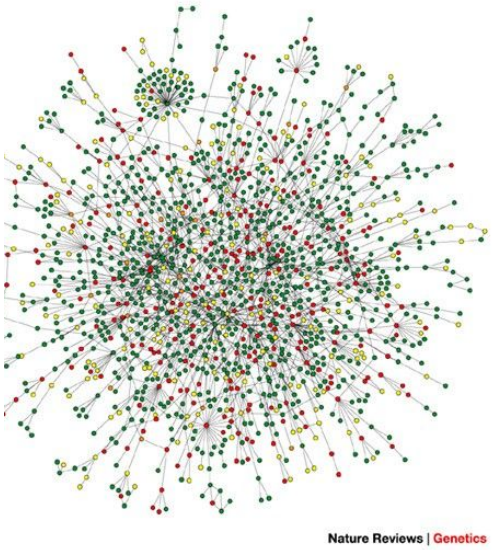
Systems Biology and Network Analysis

Modeling complex biological systems and networks.
Understanding regulatory pathways and side effects.

Modeling Biology at Scale

From mainstream reductionism to holistic approach [3]





Machine Learning (ML) and Deep Learning (DL) algorithms have been proven instrumental in the discovery of novel drug candidates, applying **predictive and generative modeling,** being a "game-changer" in **early-stage drug discovery.** Moreover, AI has expedited target identification by discerning intricate biological relationships within vast genomics and proteomics datasets, unveiling **novel druggable entities.**

AI holds immense potential for optimizing **clinical trial design** and **patient stratification,** enabling more efficient trials that bring therapies to market faster.
Natural Language Processing (NLP) and data mining techniques extract critical insights from an expanding corpus of medical literature and clinical trial data, enriching understanding of disease mechanisms and aiding in **drug repurposing.**

In clinical research, AI is used for the interpretation of multimodal patient data, facilitating **personalized treatment** strategies and **predictive diagnostics.** Furthermore, AI-driven tools have found applications in **pharmacovigilance,** automating the detection and evaluation of adverse drug reactions, thereby enhancing post-market surveillance.

Notable AI in Drug Development Milestones 2023

May 2023
FDA addresses AI/ML in drug development

FDA noted in an update that it acknowledged the **increasing use of AI/ML throughout the drug development life cycle,** with more than 100 submissions using AI/ML components reported in 2021.**[4]**

Why it matters: FDA has intensified its efforts to develop a **flexible regulatory framework for AI, focusing on fostering innovation while ensuring public health safety.** Collaborative efforts between the major regulatory institutions have resulted in the issuance of a preliminary discussion paper. This paper examines the **application of AI/ML in drug, biological product, and device development.** The agency plans to continuously seek input to refine its guidelines in this rapidly evolving field.

June 2023
Sanofi plans to put AI at the center of its operations

Sanofi is planning a **large-scale implementation of an AI-driven app - Plai, developed with Aily Labs.** The app incorporates internal company data to offer real-time insights and personalized scenarios to support data-driven decision-making.

Why it matters: Sanofi has leveraged AI to **expedite research, enhance clinical trial designs, streamline manufacturing, and boost supply chain efficiency.** The company has also partnered with various firms to fast-track treatments for cancer and immune diseases.

Similarly, other major pharmaceutical companies like AstraZeneca and Moderna have also embarked on extensive AI and data science endeavors.

Notable AI in Drug Development Milestones 2023

May 2023
FDA clears IND application for AI-discovered TACC3 PPI inhibitor

AI enables a team of four people develop a clinical pipeline

Why it matters: The FDA's approval of A2A Pharma's Investigational New Drug (IND) application for its TACC3 PPI inhibitor, A2A-252, exemplifies **the efficiency of small, AI-powered teams in advancing clinical programs. With minimal funding and a team of just four employees, A2A Pharma's generative AI platform has successfully developed two clinical-stage programs, including the TACC3 inhibitor.**

The upcoming phase 1 trial aims to enroll adults with ovarian, triple-negative breast, and endometrial cancers. Sridhar Vempati, the company's head of R&D, highlighted their use of AI in their SCULPT computational platform for drug development. SCULPT analyzes extensive small molecule libraries, ranging from thousands to millions of compounds, to design therapeutics targeting protein-protein interactions in cancer treatments.

June 2023
Insilico Medicine's AI drug enters phase 2 study

Generative AI powered with NVIDIA's GPUs for end-to-end drug development

Why it matters: Insilico employed generative AI to identify INS018_055, a small molecule now in **phase 2 study for treating idiopathic pulmonary fibrosis**.

Insilico utilizes NVIDIA Tensor Core GPUs within its Chemistry42 generative AI engine to create new molecular structures. Their Pharma.AI platform integrates various AI models for tasks such as **identifying target molecules, generating potential drug candidates, and evaluating their binding efficacy.** Additionally, Insilico has effectively used generative AI to f**orecast clinical trial results,** completing the portfolio of end-to-end platform.

AI Tackling Most Money-Draining Stage: Clinical Trials

Could AI shorten clinical trials by 2 years by 2030?

According to a recent Reuters report co-authored by Natalie Grover, Martin Coulter, and Julie Steenhuysen, **Amgen's AI tool, ATOMIC, now scans vast data to rank clinics and doctors** based on recruitment history, cutting enrollment time for some mid-stage trials by half. By utilizing ATOMIC, Amgen aims to shorten the typical drug development timeline by two years by 2030 **[5]**

Better clinical trials, but only with a good data to train AI

Novartis also leverages AI to **expedite patient enrollment in trials,** making the process faster, cheaper, and more efficient.

However, AI is only as good as the data it is trained on. With only about 25% of healthcare data available globally for research purposes, there are still limitations.

Bridging real-world data with clinical trial forecasting

Bayer utilized AI to **decrease participant numbers in a late-stage trial** for Asundexian. Specifically, it used AI to bridge mid-stage trial findings with extensive real-world data from millions of patients across Finland and the US, facilitating the **forecasting of long-term risks among a population analogous to the trial.** Bayer plans to use real-world data for an external control arm in a pediatric study of the same drug.

Millions of patients' data points processed in a few days

According to the Reuters report, Blythe Adamson, PhD, MPH, a senior principal scientist at Roche subsidiary Flatiron Health, emphasized how AI enables **rapid and large-scale analysis of real-world patient data**, contrasting it with traditional methods, which could take months to analyze data from 5,000 patients, whereas now millions of patients' data can be analyzed in just a few days.

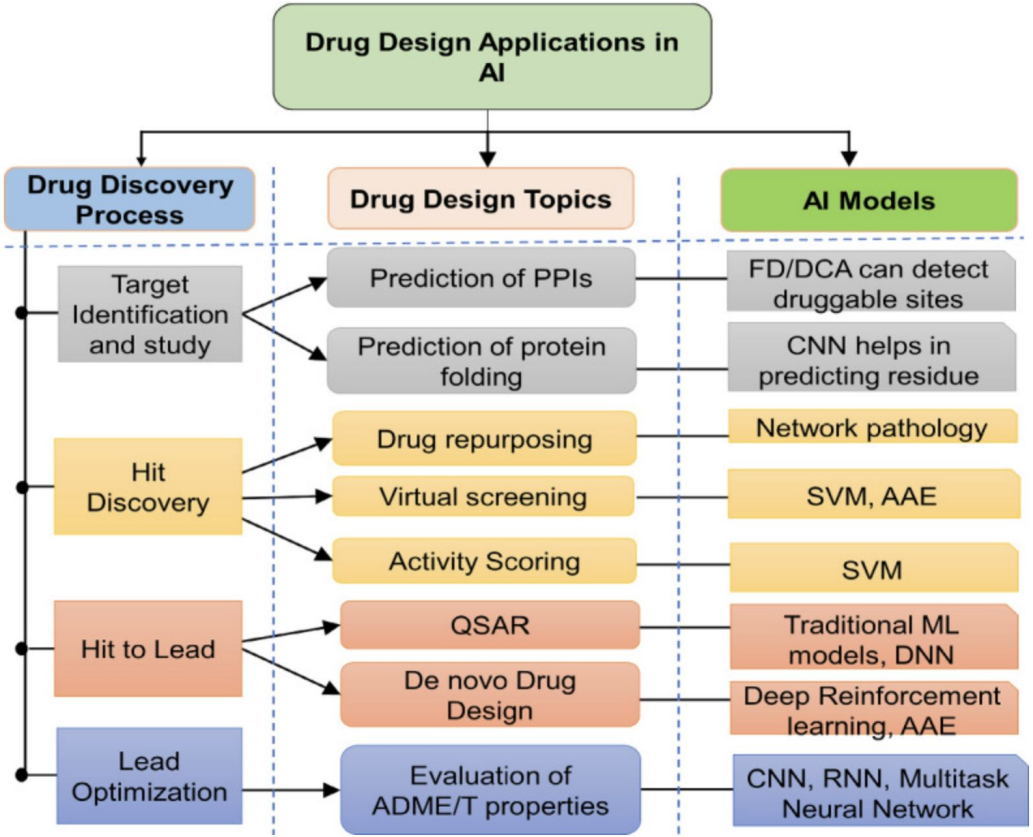
AMGEN

NOVARTIS

BAYER

Roche

AI Models Taking Over Drug Design



***See Appendix for more applications of deep learning models**

Deep learning is at the forefront of revolutionizing drug design by providing sophisticated computational models that can **analyze large datasets, predict molecular interactions, and optimize drug candidates.** This approach leverages **multi-layered neural networks,** capable of learning complex patterns and relationships within biological data.**[2]**

Key examples include:

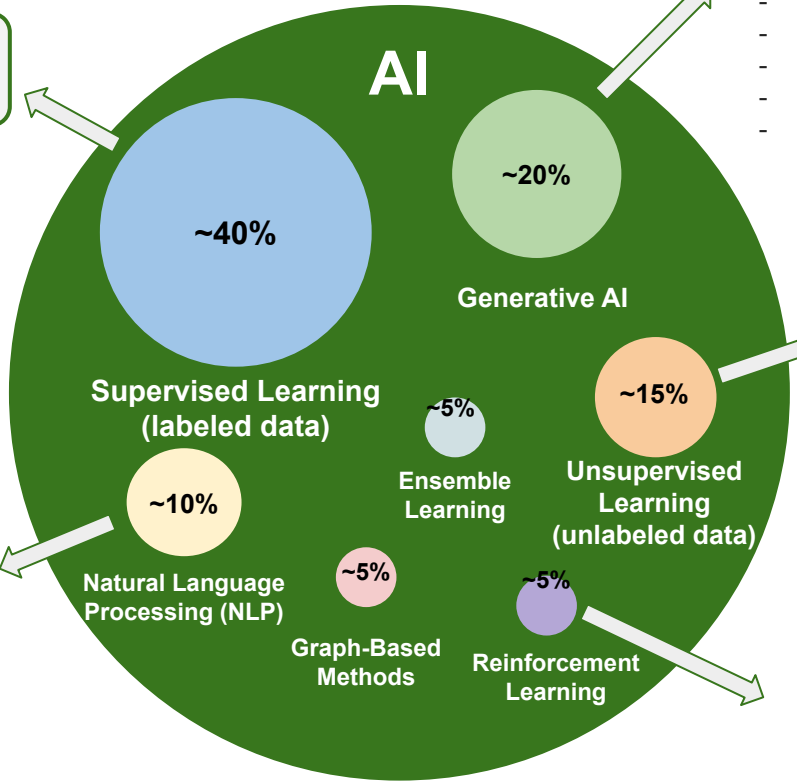
- ➔ Deep learning algorithms efficiently process vast amounts of biological and chemical data, including genomic sequences, protein structures, and pharmacological properties.
- ➔ These models predict physicochemical and pharmacokinetic properties of molecules, such as solubility and bioavailability.
- ➔ Deep learning excels in predicting how potential drug molecules interact with specific biological targets, a crucial step in identifying effective therapeutic agents.
- ➔ Advanced algorithms can design new drug-like molecules from scratch, tailored to specific biological targets, by learning from existing molecular data.

Strong and weak sides of AI approaches in DD

AI approach	Strengths	Challenges to overcome
Machine Learning	Predictive power, established & well-understood methods, versatility of application	Data dependance (+ may struggle with very large or complex datasets), prone to interpretability biases
Deep Learning	Feature extraction from complex data, scalability, state-of-the-art performance in many tasks	Data intensity, computational demands, “black box” functioning
Generative Models	Generation of novel structures and entities	Demanding validation, unknown risks of fully novel molecules
Natural Language Processing	Text mining, powerful for drug repurposing exploration	Quality and relevance, semantic understanding
Reinforcement Learning	Serves with optimization of de novo drugs and personalized treatments	Complexity, requirement for high-quality data
Graph-based models	Can capture complex relationships in drug-target interactions, suits molecular graph data	Graph construction can influence outcomes
Ensemble Learning	Combines multiple models to improve prediction accuracy and robustness	Increases computational complexity, may reduce interpretability.

Applications and trends of AI in DD

Classification is based on the learning paradigms within AI; it categorizes AI methods based on the nature of the learning process. The percentages are approximate and are based on industry reports and literature trends.



Will continue expanding and will remain as the overwhelming majority of used tools

- Compound screening (prediction of properties of novel compounds)
- Toxicity assessment
- Pharmacokinetics/ Pharmacodynamics (PK/PD)
- Clinical trial optimization
- Disease biomarker discovery
- Drug-drug and drug-target interactions

Likely to especially contribute to the clinical trials improvement

- Literature mining
- Clinical trial analysis
- Electronic health record analysis
- Knowledge graph construction

Will significantly grow, can create novel data by learning existing patterns and structures

- De novo drug design
- Compound libraries for HTS
- Deep insights into complex data
- Synthetic data generation
- Molecular property (lead) optimization

Expected to have one of the biggest scale-ups, aimed to reveal patterns, structures and relationships within the data

- Clustering of compounds
- Patient stratification
- Chemoinformatics
- Adverse effect detection
- Pattern Recognition in high-dimensional data

Expected moderate growth, but possibly more significant than in other AI application sectors

- Personalized medicine
- Optimizing drugs/ drug delivery
- Chemical synthesis planning and experiments optimization
- Adaptive clinical trial design

The AI Ecosystem in Biopharma

400+ AI-driven companies (startups/scaleups), offering a wide array of AI-driven platforms and services - from classical Software as a Service model to custom data science services, drug discovery ("Drug candidate-as as service"), and clinical trial support/management resources.

Top-tier pharmaceutical and biotech companies developing in-house AI expertise as part of their R&D strategy. Such players often collaborate with external AI vendors and AI-driven biotech startups to explore pilot programs in drug discovery/basic biology/clinical trial analytics.

Domain-specific software providers (e.g., KNIME, ChemAxon, Dotmatics, MolSoft, and others) primarily focus on cheminformatics/bioinformatics software but also provide machine learning-powered tools.

Academic labs in pharma/biotech space, conducting AI research and developing specialized frameworks and tools relevant to the industry (usually a cradle for future AI startups/spin-outs)

Top-tier technology companies like Google, Amazon, and Tencent entering the pharmaceutical space, leveraging cutting-edge AI technologies and big data infrastructures.

Non-domain-specific software providers developing AI-as-a-service packages and models suitable for application in pharmaceutical research (e.g., "out of the box AI")

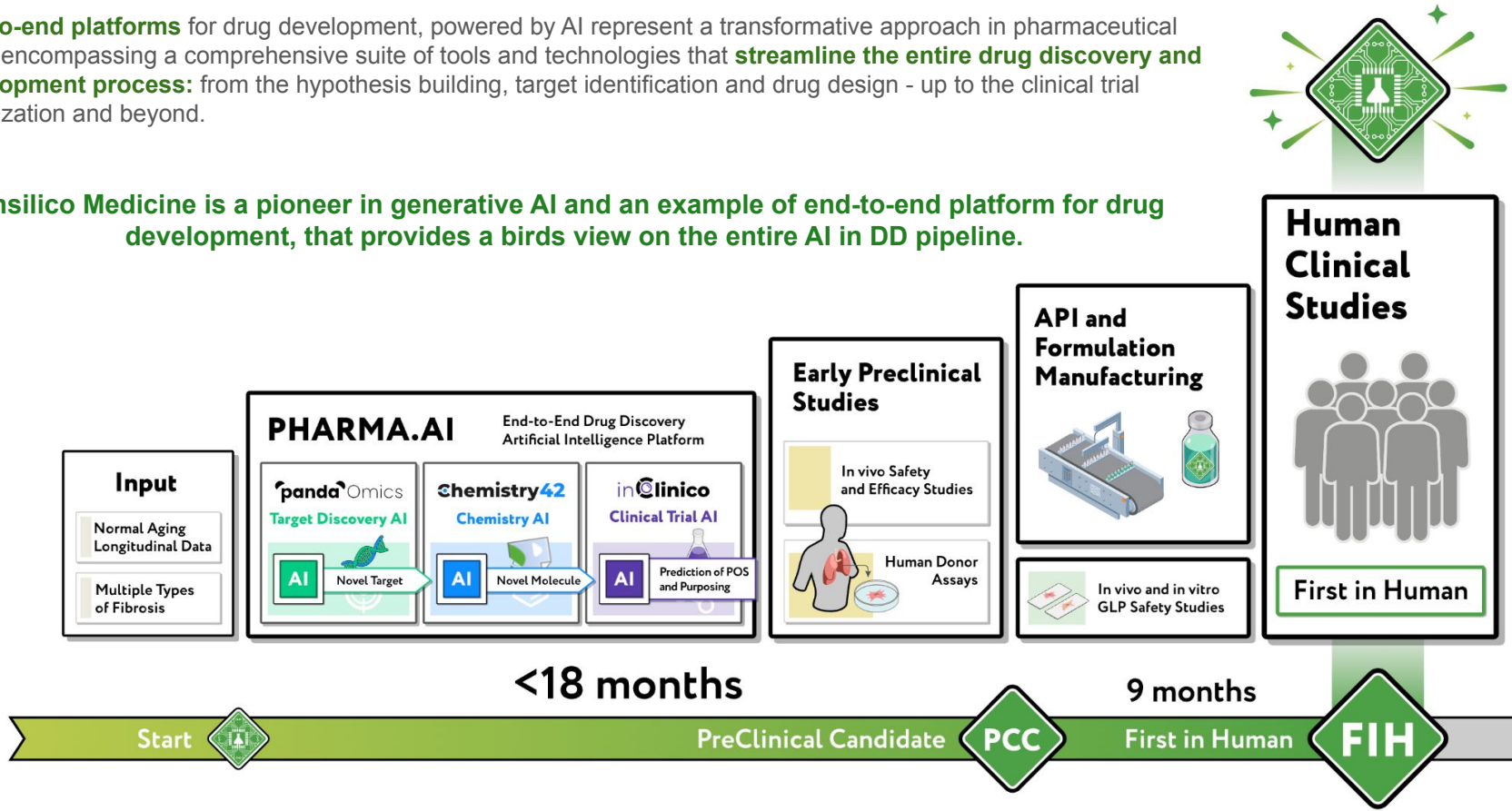
Contract research organizations (CROs) developing expertise in AI to augment their value offering to pharma/biotech customers

Open-source machine learning tools and frameworks, widely exploited by life science professionals in their research projects **[6]**

AI-enabled ‘Conveyor Line’ for Small Molecules

End-to-end platforms for drug development, powered by AI represent a transformative approach in pharmaceutical R&D, encompassing a comprehensive suite of tools and technologies that **streamline the entire drug discovery and development process:** from the hypothesis building, target identification and drug design - up to the clinical trial optimization and beyond.

Insilico Medicine is a pioneer in generative AI and an example of end-to-end platform for drug development, that provides a birds view on the entire AI in DD pipeline.



AI across chemical modalities

AI-enabled ‘Conveyor Line’ for Small Molecules

By offering a **unified and AI-integrated workflow,** end-to-end platforms aim to enhance the efficiency, reduce the time and cost, and increase the success rate of drug development, ultimately leading to more rapid delivery of novel therapeutics to the market.

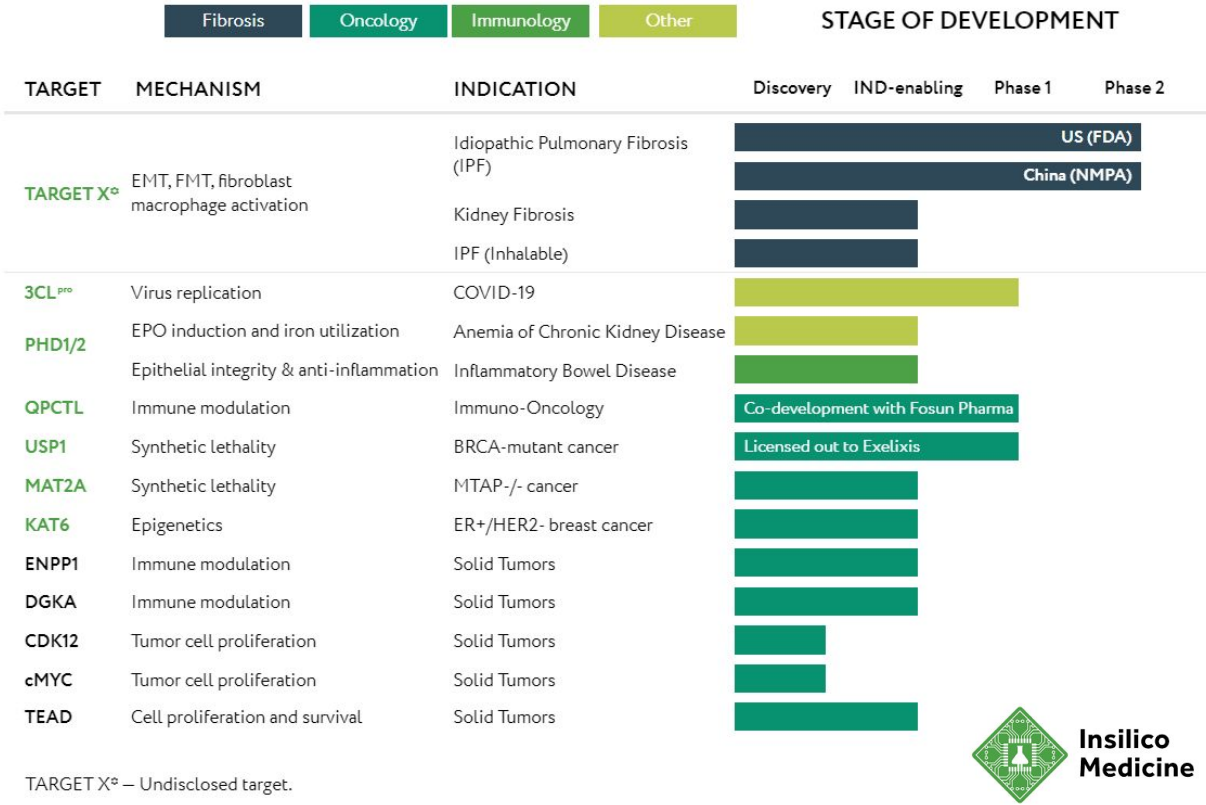
Traditional pharma approach usually allows to focus only on one drug development direction at a time, while AI assistance adds the exponential growth to the opportunities.

As an example, Insilico’s drug candidates suit currently includes

31 programs and 29 targets

out of which 11 preclinical candidates were nominated in 2021 and 2022 (2 and 9 correspondingly).

Drug candidates indications also vary tremendously in the mechanism of action and disease groups, while the majority of them were developed in a span of last several years.



AI IN DRUG DEVELOPMENT VALUE CHAIN STEP BY STEP

Modeling Biology, Target Discovery

Disease modeling and hypothesis building are crucial foundational stages in the drug development process. Before even identifying specific drug targets or designing potential therapeutic compounds, scientists must understand the intricacies of disease mechanisms and develop hypotheses regarding intervention points.

Disease modeling

Molecular pathway analysis

Using AI, one can map the complex interaction networks of proteins and genes, elucidating disease pathways. For instance, **Graph Neural Networks** can capture the intricate relationships in molecular networks.

Multimodal disease signature identification

Machine learning models can integrate diverse data types (genomics, proteomics, imaging, etc.) to identify a disease signature, providing a multidimensional understanding of disease pathology.

Disease subtype discovery

Unsupervised learning, especially clustering algorithms, can dissect heterogeneous diseases into distinct subtypes based on molecular or clinical characteristics, paving the way for specialized treatments

Disease trajectory modeling

Deep learning models, especially RNNs, can predict disease progression from patient health records, identifying critical transition points and potential intervention windows.

Hypothesis Building

Mechanistic hypothesis generation

Deep learning models, trained on vast biological datasets, can suggest possible mechanistic pathways of diseases, providing novel intervention points.

Disease association mapping

AI (especially **NLP models**) can mine clinical databases to identify unexpected associations between diseases, shedding light on shared underlying mechanisms.

Systematic literature review

NLP models can extract and synthesize insights from vast biomedical literature, identifying gaps in current understanding and suggesting novel hypotheses.

Biomarker discovery

Machine learning can sift through vast omics datasets to identify potential biomarkers, which can serve as indicators of disease presence, severity, or treatment response.

Disease Modeling & Hypothesis Building

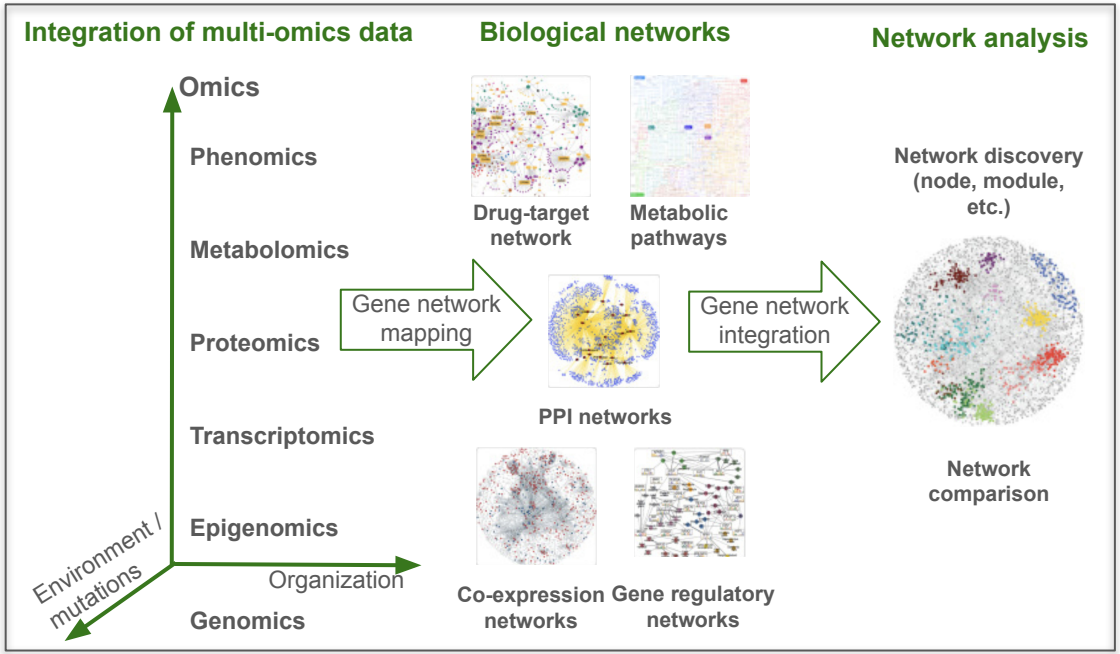
Transformation of Drug Discovery

Traditional Approach

Mostly based on **singular hypothesis-driven approaches,** dependent on trial-and-error in the lab.

Limited by human capacity to analyze multi-dimensional data.

Requires **manual literature reviews,** which are time-intensive and can miss connections.



AI-Enhanced Approach

Can reduce experimental failures by providing **predictive tools.**

Integrates disparate data sources for a **holistic understanding.**

Automated literature reviews and **data mining** expedite hypothesis generation.

AI enables transition from mainstream reductionism to holistic approach in drug discovery [3]

Unlock Report

The State of AI in the Biopharma Industry

Bio
Pharma
Trend

www.BiopharmaTrend.com | December 2023