sapio sciences

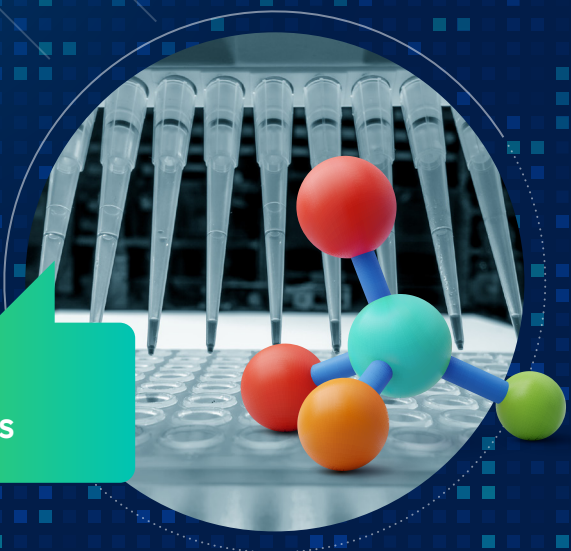
What is Science-Aware AI?

Learn how science-aware AI is being used to support drug discovery, clinical development, and pharmaceutical manufacturing.

Configure my workflow to support more samples

AI

Visualize data from my latest experiment



Set up a plate with 10 samples

Capturing Value in Life Sciences with AI

AI is evolving rapidly. The 2010s experienced a boom in the development of machine-learning algorithms to help discern patterns in large, complex scientific data sets. In the 2020s, new, generative AI tools pre-trained on vast data sets exhibit even greater transformative potential.[1] As a result, AI is providing substantial potential for scientists to adopt more complex techniques that outperform existing methods.[2]

Three functional areas of life sciences — drug discovery, clinical development, and pharmaceutical manufacturing — are likely to capture the most value from new AI workflows and processes.

A significant opportunity for AI value capture is available in life sciences. Presently, the average time to bring a new medication to market is about 10 to 12 years, with costs exceeding $2.6 billion.[3] Deloitte, a leading multinational strategy and technology consulting firm, estimates that the top 10 biopharmaceutical companies could capture between $5 billion and $7 billion of value by scaling the use of AI over 5 years. In addition, smaller biopharma companies could achieve between $2.5 billion and $3.5 billion.[4] Three functional areas of life sciences — drug discovery, clinical development, and pharmaceutical manufacturing — are likely to capture the most value from new AI workflows and processes (see Table 1).[5]

What We'll Cover

Table 1 – Estimated Value Capture from New AI Workflows and Processes

Functional Area	Drug Discovery Pre-discovery and discovery research	Clinical Development Preclinical, translational, clinical, regulatory, and safety	Pharmaceutical Manufacturing Product development, manufacturing, quality, and supply chain
Benefit	Discover better drug candidates and accelerate timelines by almost 3 years per successful drug	Accelerate timelines by 1.5 years per successful drug	Lower supply chain risks and get critical medicines into the hands of patients faster
Revenue	$0.3 to 1.5 billion upside per successful drug	$200 to 800 million upside per successful drug	1 to 3% uplift
Cost	$600 to 800 million decrease per successful drug	$300 to 400 million decrease per successful drug	3 to 5% decrease in production and fulfillment 10 to 15% working capital reduction

Source: Reinventing life sciences in the age of generative AI, Accenture[6]

Science-Aware AI Will Produce More Effective Medications

AI enables computers to understand, translate, see, and interpret spoken and written languages, analyze data, make recommendations, and more. Core technologies, such as machine learning (ML), natural language processing (NLP), computer vision, and generative models, are initially learning mechanisms for AI. Domain-specific knowledge is used to train an interactive, reliable, computer-based AI decision-making system that can solve complex problems within a specific domain.[7]

Science-aware™ AI refers to a new generation of core technologies specifically designed and trained to navigate the complexities of science-based criteria specific to laboratory discovery, development, and manufacturing.

The heart of a new medication pipeline is the laboratory where experiments are performed to identify novel compounds, conduct clinical development, and assure manufacturing quality and compliance.

Science-aware AI capabilities can be seamlessly integrated into a laboratory informatics platform empowering scientists to collect, manage, and analyze data from experiments that outperform existing methods. By understanding the unique needs of laboratory scientists, science-aware AI assists in crafting tailored workflows, conducting cutting-edge research, and leveraging specialized tools to decipher intricate biological, chemical, and multimodal data.

Nine Characteristics of Science-Aware AI

Domain-Specific Knowledge

Science-aware AI systems can be deeply integrated with domain-specific knowledge bases (i.e., a collection of facts and heuristics organized about the system's domain).[8] These AI systems can understand, interpret, and apply scientific principles, terminologies, and methodologies specific to disciplines such as biology, chemistry, and pharmacology. This allows them to provide highly contextual insights and recommendations relevant to the specific scientific area where they are deployed.

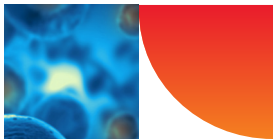
For example, the prediction of drug-protein interactions (DPIs), a cornerstone of effective drug development, can benefit significantly from the integration of AI. By applying sophisticated computational techniques, such as machine learning and deep learning, AI can analyze large-scale biological and chemical datasets, encompassing molecular structures, chemical properties, and experimental binding affinities. These domain-specific knowledge bases provide a rich training environment for AI algorithms to discern and learn the intricate patterns and relationships that govern the interactions between drug compounds and protein targets.[9]

Lab Workflow Automation

Science-Aware AI introduces transformative automation to laboratory workflows by leveraging its ability to understand and respond to science-based search, data entry, experiment design, plate layouts, informatics support, software documentation queries, and acting as a proxy to leverage many scientific analysis tools – all built natively into the LIMS, ELN or Scientific Data Cloud informatics environment.

Through intelligent automation, repetitive tasks such as data entry, sample management, and result interpretation can be streamlined, reducing manual errors and increasing efficiency. By integrating Science-Aware AI within informatics software, laboratories can more easily configure and automate complex workflows while ensuring that compliance and quality control standards are maintained. This not only speeds up processes but also enhances precision, allowing researchers to focus on higher-value tasks like hypothesis generation and data analysis. Ultimately, Science-Aware AI empowers laboratories to run smarter and faster, delivering better outcomes in drug development and clinical testing.

Sapio Sciences' approach to Science-Aware AI for automating and optimizing lab workflows is embodied in their product, ELaiN (Electronic Laboratory AI-Powered Notebook).



Advanced-Data Interpretation and Contextualization

Science-aware AI excels at interpreting complex, multi-dimensional data sets, such as genomic sequences, proteomic data, and high-throughput screening results, along with its associated metadata (contextual, informative data such as its origin, location, quality, processes, etc.).[10] The combination of complex data sets and contextual metadata helps scientists interpret the advanced data and derive meaningful conclusions and hypotheses.

For example, AI can be used to develop predictive models for treatment response during clinical development. By analyzing patient data, including genetic profiles, AI can identify risk factors and the most effective treatment options to help improve clinical outcomes and reduce the risk of adverse effects.[11]

In Silico Experimentation and Predictive Modeling

Science-aware AI systems can conduct in silico experiments to enable the simulation of biological processes, chemical reactions, and other critical characteristics of drug candidates. Modeling based on the in silico experimental results can facilitate highly accurate predictions of likely outcomes to better target and reduce the time and cost of in vitro and in vivo experiments.

90% of drug failures in clinical development are mainly due to a lack of efficacy and uncontrolled toxicity. Drug distribution is an important process in pharmacokinetics because it has the potential to influence both the amount of medicine reaching the active sites and its effectiveness and safety. Traditionally, predicting distribution properties has relied heavily on in vitro and in vivo studies, which are costly and time-consuming. Furthermore, in vitro screening of compounds is typically limited to a few properties, with emphasis placed on only a few of the most promising candidates.[12]

AI systems can efficiently and cheaply screen thousands or millions of candidate molecules rather than limiting the examination of distribution features to just a few. In silico distribution-related models can be employed for quick and early screening of drug distribution properties before the most promising are further explored in vitro to reduce the cost and time of clinical development.[13]

Process Automation and Optimization

Science-aware AI systems are designed to automate and optimize scientific workflows by tailoring processes to specific drug discovery, development, and delivery goals. This includes automating data collection, analysis, and reporting, as well as optimizing experimental designs based on previous outcomes and existing data.

Drug formulation spans the discovery, development, and delivery phases of new medicines. The traditional approach to drug formulation relies on iterative trial-and-error methods, requiring many resource-intensive and time-consuming in vitro and in vivo experiments. AI- and ML-directed workflows can fast-track chemical, biology, and multimodal-based recipe development and generate new knowledge in formulation science.[14]

AI models can employ ML and process development data to more quickly design optimal process parameters and scale-up strategies. Pharmaceutical manufacturers can adopt advanced process control approaches that combine AI techniques with scientific knowledge to improve manufacturing efficiency and output. And AI methods integrated with process performance metrics can offer better trend monitoring to address manufacturing discrepancies before they impact the supply chain or cause a drug shortage.[15]

Continual Learning for Performance Improvement

Science-aware AI systems are capable of continual learning — they improve their understanding and performance based on new data and evolving scientific knowledge. This characteristic ensures that science-aware AI remains up to date with the latest scientific advancements and that it can incorporate new knowledge into analysis and recommendations.

Continual learning ML models can be used for predictive analytics in situations where clinical outcomes can be automatically obtained and used to update an algorithm. If a model were to predict a clinical outcome, such as all-cause mortality within three months, at three months the model could be automatically updated using the actual clinical outcome, continuously improving the standard of care.[16]

The long-term goal is to use continual learning models of AI to optimize and support clinical decisions – subject to FDA guidance. For example, AI models could be combined with therapeutics to provide the optimal medication dosing for individual patients. In this situation, the model would be making active clinical decisions — attempting to optimize the eventual clinical outcome for patients — and would require additional evaluation and safeguards before updating the algorithm.[17]

Cross-Domain Connectivity and Collaboration

Science-aware AI systems can facilitate multimodal research by connecting data, tools, and insights across different scientific domains. It supports collaborative efforts by enabling seamless data integration and communication between researchers from various fields, fostering innovation at the intersections of disciplines.

New modalities, such as recombinant proteins, peptides and engineered antibodies are key drivers of biopharmaceutical industry growth. According to Boston Consulting Group (BCG), over the past few years, revenues from new-modality products increased by $60 billion, while revenues from conventional products declined by $10 billion. Furthermore, BCG projects that the percentage value of new modalities in the five-year forward pipeline between 2019 and 2023 will increase from 41% to 56%, far outpacing conventional ones.[18]

Transformer architecture, initially proposed and developed by researchers at Google in 2017,[19] is the foundation for generative AI tools, such as ChatGPT. Researchers in biology and chemistry realized that transformer models could be applied to design protein-based biotechnologies.[20] This insight led the Google DeepMind team to develop AlphaFold, an AI system that accurately predicts 3D models of protein structures based on their sequence of amino acids,[21] which is having a significant impact on research areas that need protein structure information, including drug discovery.[22]

GxP Guidelines and Regulatory Compliance

Science-aware AI systems can be designed in accordance with good practice (GxP) guidelines and regulations, such as good laboratory practice (GLP), which help ensure quality and safety. The U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified 10 guiding principles to inform the development of good machine learning practice (GMLP).[23] Concepts outlined in draft regulatory guidance regarding GMLP are generally analogous to good laboratory practice.[24]

AI/ML techniques can strengthen traditional GLP. For example, errors in the preanalytical phase of testing are thought to account for the highest frequency of laboratory errors. However, they are difficult to prevent because many of the relevant processes are beyond laboratory oversight. ML techniques can be used to identify preanalytical errors, including the use of optical character recognition to identify mislabeled samples, the detection of spuriously increased glucose results due to intravenous fluid contamination, and logistic regression and support vector machines to detect “wrong blood in tube” errors.[25]

As part of quality assurance in the analytic phase, result verification processes are used to identify test errors before result release. This is commonly implemented using a nonadaptive, rule-based system to accept or reject results. An adaptive, ML-based approach using artificial neural networks to develop a model could be used for auto-verification purposes.[26]

Decision Support and Actionable Insight

Science-aware AI systems provide enhanced decision support by generating actionable insights from complex data sets. It goes beyond mere data analysis by offering scientifically sound recommendations, hypothesis generation, and decision-making support — empowering scientists and researchers to make more informed and impactful decisions.

For example, decision intelligence (DI), derived from Google Research and Development, combines behavioral research with AI/ML-based decision-making for data-driven technology applications. It enables the structured discovery of decision points and their decision context. As a result, DI can help identify where AI/ML enables users to make better decisions.[27]

Google Research has also developed Articulate Medical Intelligence Explorer (AMIE), a research AI system based on a large language model (LLM) optimized for diagnostic reasoning and conversations. Recognizing that many aspects of good diagnostic dialogue are unique to the medical domain, AMIE was trained on real-world datasets comprised of medical reasoning, medical summarization, and real-world clinical conversations.[28] DI and AMIE are just two examples of how science-aware AI is going beyond traditional data analysis to help scientists make more informed and impactful decisions.

A Science-Aware™ Laboratory Informatics Platform

Over the past half-century, the experimental and data tools of laboratory scientists have evolved from glass beakers, bunsen burners, and paper notebooks to sophisticated computerized test instruments and specialized informatics software. Today, a new generation of AI core technologies specifically designed and trained to navigate the complexities of science-based criteria is revolutionizing laboratory informatics. The foundation of this laboratory revolution is a science-aware informatics platform.

Generative AI accessed through a simple, natural language chat interface built into a laboratory informatics platform can leverage specialized, science-aware Large Language Models (LLMs) to rapidly create experiments, search and visualize data, and generate code as requested.

A science-aware, AI-powered laboratory informatics platform combines a Laboratory Information Management System (LIMS), Electronic Laboratory Notebook (ELN), and Scientific Data Management System (SDMS) on a unified foundation for drug discovery, clinical development, and pharmaceutical manufacturing. It incorporates core AI technologies, such as machine learning (ML), natural language processing (NLP), and generative models into the platform to collect, manage, and analyze data from experiments in a manner that outperforms existing methods. For example, generative AI accessed through a simple chat interface built into a laboratory informatics platform can leverage specialized, science-aware LLMs to rapidly create experiments, search and visualize data, and generate code as requested.

A science-aware, AI-powered laboratory informatics platform supports the nine characteristics previously described to improve laboratory operations beyond existing methods and capture value in drug discovery, clinical development, and pharmaceutical manufacturing now and for years to come.

Sapio Sciences

For 14 years, Sapio Sciences has been developing and delivering informatics software to make laboratory research easier for scientists to accelerate scientific progress and reduce suffering in the world. Sapio has always recognized the fundamental role of a comprehensive platform, a scalable data architecture, and a unified user experience to empower scientists. The Sapio Science-Aware™ Informatics Platform combines essential products, such as an ELN, LIMS, and SDMS, and critical capabilities, including advanced AI technology, configurability, extensibility, and interoperability, into a unified laboratory informatics platform that delivers a seamless end-to-end user experience now and for years to come.

1. AI will transform science — now researchers must learn to tame it, Nature 621, September 28, 2023, p 658. 2. Science in the age of AI: How artificial intelligence is changing the nature and method of scientific research, The Royal Society, May 2024, p 5. 3. Petra Jantzer, Selen Karaca-Griffin, Kailash Swarna, Tracy Ring, Jen Spada, Reinventing life sciences in the age of generative AI, Accenture, 2024, p 8. 4. Aditya Kudumala, Adam Isreal, Sai Lella, Johnathan Fan, and Wendell Miranda, Realizing Transformative Value from AI & Generative AI in Life Sciences, Deliotte, 2024, p 2. 5. Jantzer, p 17. 6. Ibid. 7. From data to decisions: a guide to the core AI technologies, Neuroni.com, accessed September 19, 2024, https://neuroni.co/en/articles/key-ai-technologies. 8. Ibid. 9. M. K. G. Abbas, Abrar Rassam, Fatima Karamshahi, Rehab Abunors, and Maha Abouseada, "The Role of AI in Drug Discovery," ChemBioChem, Volume 25, Issue 15, July 15, 2024. 10.O'Leary, N.A., Cox, E., Holmes, J.B. et al. "Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets," Sci Data 11, 732, July 5, 2024. 11.Manlio Vinciguerra, "The Potential for Artificial Intelligence Applied to Epigenetics," Mayo Clinic Proceedings: Digital Health, Volume 1, Issue 4, December 2023. 12.Thi Tuyet Van Tran, Hilal Tayara, and Kil To Chong, "Recent Studies of Artificial Intelligence on In Silico Drug Distribution Prediction," International Journal of Molecular Sciences, January 17, 2023. 13.Ibid. 14. Pauric Bannigan, Matteo Aldeghi, Zequing Bao, et al., "Machine learning directed drug formulation," Advanced Drug Delivery Reviews, Volume 175, August 2021. 15.Arthur C. Fisher, "The Future is the Present: Artificial Intelligence in Pharmaceutical Manufacturing," Pharmaceutical Technology, Volume 47, Issue 9, September 2023. 16.Cecilia S. Lee and Aaron Y. Lee, "Applications of continual learning machine learning in clinical practice," Lancet Digit Health, Volume 2, Issue 6, June 2020. 17.Ibid. 18.Mike Brochu, Lu Chen, and Brian Bush, "New Drug Modalities 2023," BCG, June 30, 2023. 19.Jakob Uszkoreit, "Transformer: A Novel Neural Network Architecture for Language Understanding," Google Research, August 31, 2017. 20.Marc Airhart and Anita Shiva, "Turbocharging Protein Engineering with AI," University of Texas College of Natural Sciences, September 25, 2024. 21. AlphaFold, Google DeepMind, https://deepmind.google/technologies/alphafold/, accessed October 3, 2024. 22. Zhenyu Yang, Xiaoxi Zeng, Yi Zhao, and Runsheng Chen, "AlphaFold2 and its applications in the fields of biology and medicine, Signal Transduction and Targeted Therapy, March 14, 2023.23. Good Machine Learning Practice for Medical Device Development: Guiding Principles, U.S. FDA, Health Canada, and United Kingdom MHRA, October 2021, https://www.fda.gov/media/153486/download, accessed October 3, 2024. 24. Daniel S. Herman, Daniel D. Rhoads, Wade L. Schulz, and Thomas J. S. Durant, "Artificial Intelligence and Mapping a New Direction in Laboratory Medicine: A Review," Clinical Chemistry, 67:11 (2021), p. 1476. 25. Herman, p. 1474. 26. Ibid. 27. Lucas Glass, "Decision Intelligence is the Key to AI/ML Adoption in Life Sciences," TDWI (Transforming Data With Intelligence), https://tdwi.org/articles/2021/11/17/adv-all-decision-intelligence-and-ai-ml-adoption-in-life-sciences.aspx, accessed October 4, 2024. 28. Alan Karthikesalingam and Vivek Natarajan, "AMIE: A research AI system for diagnostic medical reasoning and conversations," Google Research, January 12, 2024, https://research.google/blog/amie-a-research-ai-system-for-diagnostic-medical-reasoning-and-conversations/, accessed October 4, 2024.