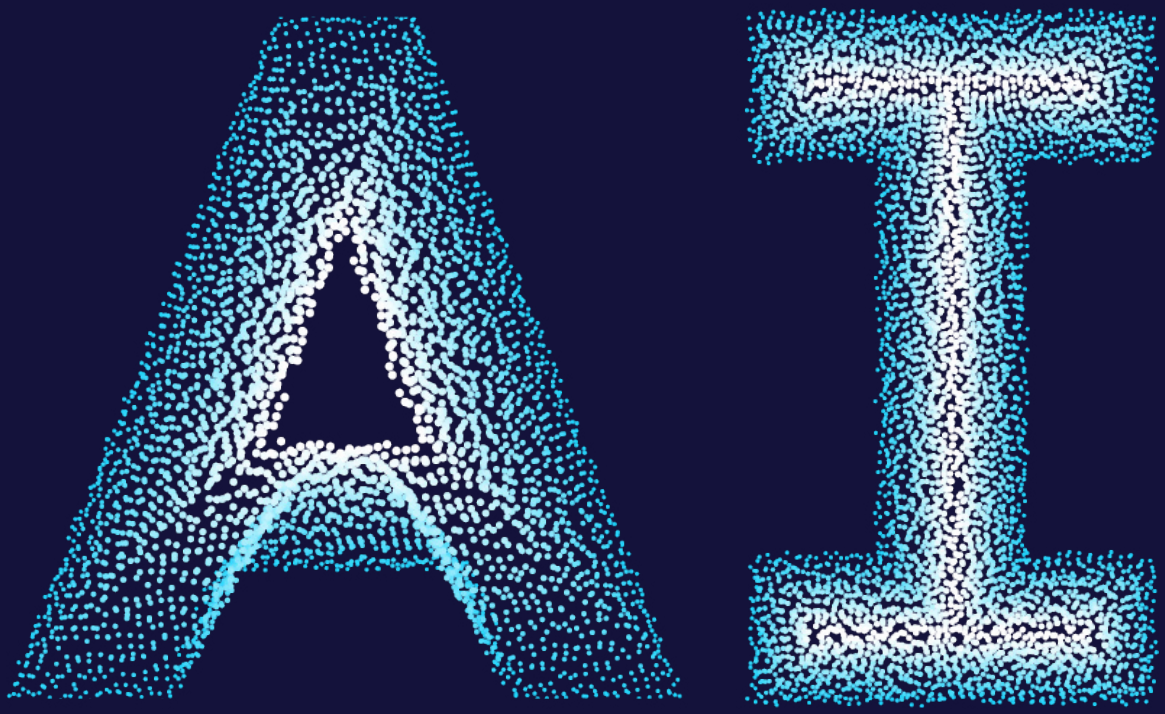
CITELINE SMARTSOLUTIONS
CITELINE CLINICAL

CITELINE

White Paper

AI in Pharma: Benefits, Risks, and the Road Ahead

by Luca Parisi



September 2024

Introduction

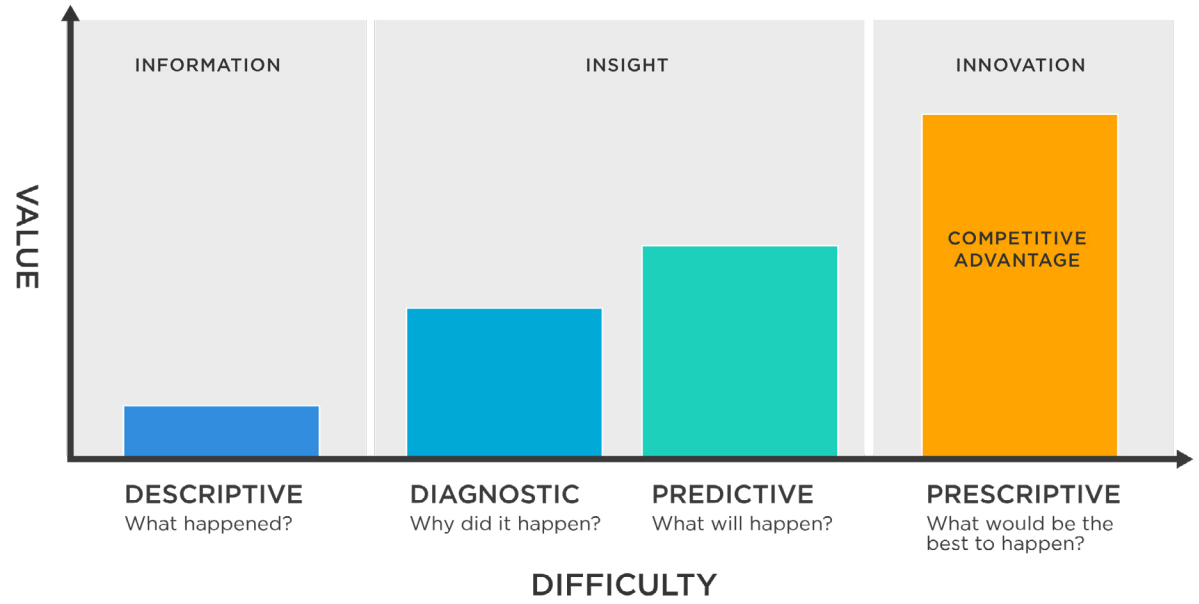
Use of artificial intelligence (AI) now permeates every industry, and pharma is no exception. Whether machine learning (ML) models that make predictions based on existing data or generative AI (GenAI) models that create new data based on the data they were trained on, AI is being used to streamline and accelerate each step of the drug development process from research through approval and marketing.

According to McKinsey & Co., generative AI alone could produce $60 billion to $110 billion a year in economic value across the pharma industry value chain. And $13 billion to $25 billion of that annual value alone would be for clinical development.[1]

AI is able to handle both structured and unstructured data, including multimodal data such as tabular, text, images, and videos. At its most basic level, AI can automate mundane tasks such as structured document and image analyses, enabling experts to spend more time on tasks that require their attention and proficiency.

On a deeper level, AI can unveil insights from historical data to inform operations and provide a lens into the future through predictive analytics, supplementing traditional descriptive and diagnostic analytics that solely provide analytical information anchored on historical patterns. It can also enable and accelerate expertise through prescriptive analytics, advising experts on the next best action to take to maximize added value.[2]

Figure 1. Value-difficulty trade-off from traditional descriptive analytics to prescriptive analytics via human-AI collaboration for sustainable competitive edge



Source: Jaspersoft

How AI specifically benefits pharma

AI can accelerate drug discovery and development by supporting the analysis of vast and differing datasets, including comprehensive drug databases, biochemical data, clinical trial data, and electronic health records (EHR). AI analysis is much faster and cheaper than traditional methods at identifying potential drug candidates, reducing the time required for drug discovery and maximizing the quality of the novel compound.

For example, AI-driven drug discovery platforms have significantly reduced time to identify drug candidates. What used to take four to five years can now take as little as eight months.[3]

AI can also empower drug repurposing. It can identify drug compounds already approved for other indications and help to predict their probability of success when repurposed to treat different diseases. There are AI-enabled systems that help prioritize the most promising candidates and estimate the safety profile and efficacy of existing drugs for other, similar diseases.[4,5]

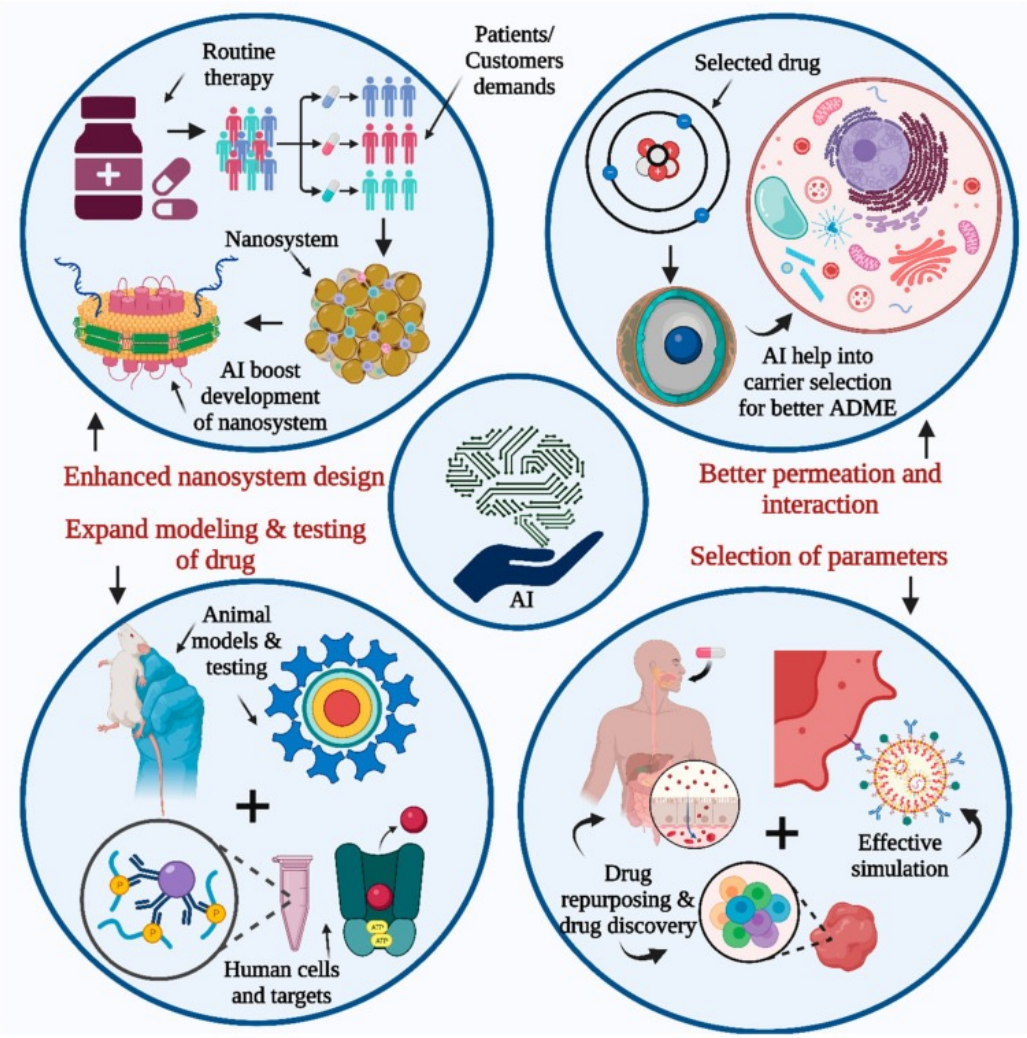
AI enables both the development of personalized treatments and titration of treatment pathways. It can also help advise on medication switching and tailoring dosage to an individual's specific needs. It does so in part by analyzing multimodal patient data to predict how an individual would respond to a treatment.

AI can support clinical trial planning and optimize clinical trial design through tailored protocol designs and investigator and site selection. Furthermore, AI can help monitor and rescue studies. This ensures on-time patient recruitment and trial delivery so drugs can reach the market and the patients that need them on time and on budget.

When it comes to ensuring clinical trial diversity, AI systems can be applied to help mitigate human biases in clinical trial design and optimize the trade-off between increasing diversity across various characteristics (e.g., race, ethnicity, demographics) and delivering the trial on time. These tools can ensure clinical trial diversity requirements are met and even exceeded.

AI can automate and streamline processes such as drug manufacturing, supply chain management, clinical trial planning and execution, and pharmacovigilance. This can lead to increased operational efficiencies for pharmaceutical sponsors and the contract research organizations (CROs) that run trials on their behalf.

Figure 2. Applications of AI in drug development



ADME: Drug absorption, distribution, metabolism, and excretion

Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10385763/

Risks of employing AI tools

Applying AI in pharma involves the potential use of sensitive patient data, including personal identifiable information (PII), which raises concerns about data privacy and security. This requires leveraging data as directed under the applicable regulations and standards, such as HIPAA in the US and GDPR in the European Union.

With this in mind, it is best practice to leverage the minimum amount of data required for a business application. For instance, when utilizing real-world data (RWD), use age instead of date of birth where possible, or postal code area instead of full address.

It is also crucial to apply the required encryption to such patient-level data at rest and in transit, and to models consuming them (e.g., encrypted model endpoints).

In light of technological advances such as GenAI, integrating AI in pharma requires revisiting existing regulations and standards to ensure its ethical application. Guiding principles such as the Good Machine Learning Practice (GMLP), a joint effort of the US Food and Drug Administration (FDA), Health Canada, and the UK's Medicines and Healthcare products Regulatory Agency (MHRA), can help in this regard.

While addressing ethical and regulatory concerns is crucial, redefining such regulations and standards, as well as understanding how to fully comply with them, may slow the adoption of AI in pharma.

From a software development perspective, good practice (GxP) compliance can be guaranteed by ensuring data and model reproducibility through versioning with tools like Data Version Control and appropriately logged experiments and artifacts, as well as documentations of methodological and evaluation steps followed for both technical and non-technical audiences.

Outcomes from AI-driven systems rely heavily on the quality of the training data used. Some factors affecting this quality include presence and extent of outliers and missing values, lack of or limited representativeness, and noisy or incorrect data. The term "noisy data" refers to data that contain irrelevant or erroneous data points.[6] As the saying goes, "garbage in, garbage out" — training with data that are suboptimal in quality will yield suboptimal results.

Ways to improve data quality for AI model development include:

- Imputation of missing values: filling in missing data points to achieve complete datasets via either statistical or ML-driven approaches[7]
- Data augmentation: creating new data points from existing ones to increase the amount of data[8]
- Data standardization: ensuring the data follow a consistent format and structure

There is also a risk of AI perpetrating human biases. Biases in the training data result from human processes; they need to be analyzed and mitigated so that AI systems do not continue to perpetuate these biases in their predictions. Some methods for addressing these biases include diverse data collection; adversarial training, which involves training a neural network to evaluate AI-generated content for bias[9]; and data augmentation to enhance participation of underrepresented populations.

The high complexity, and therefore limited transparency, of certain AI-driven systems like large language models (LLMs) with billions

of parameters can hinder trust and slow their adoption. This is especially true in highly regulated industries such as life sciences, where the outcomes from these AI systems contribute to the design and delivery of treatments to patients.

For example, it is important to review and assess the evidence used by LLMs to generate outputs. Grounding and tailoring these outputs with proprietary data via retrieval-augmented generation (RAG) and fine tuning can lead to more explainable and accurate outputs when using LLMs.

Where pharma is heading with AI

The next phase of AI adoption is already under way in pharma. Companies are hard at work devising the next generation of AI-enabled technologies to support the drug development process from discovery to launch and beyond.

For instance, GenAI platforms such as NVIDIA BioNeMo are accelerating drug discovery by optimizing molecular designs.[10] IBM employed GenAI in efforts to repurpose drugs for insomnia and Parkinson's disease to treat the dementia that often accompanies Parkinson's (PDD).[11]

EdgeAI is helping support more decentralized trials with remote patient monitoring through the use of wearable devices with sensors to monitor vital signs in real time. These data are analyzed locally and can alert doctors or patients and their caregivers of any pathophysiological activities, which can help with timely intervention and enhance quality of life.[12]

Multi-modal GenAI is being used to analyze RWD to identify subjects more holistically and for precision medicine. For example, using GenAI to simultaneously consider data such as imaging, EHR, and multiomics (a type of biological analysis) helps develop personalized treatments by understanding patients' conditions more holistically through the entire patient journey.[13]

AI can also be employed to develop "digital twins" for further personalized medicine. In one instance, a digital twin can be used to simulate a patient's glucose-insulin dynamics, contributing to personalized insulin delivery patterns. This can help patients with type II diabetes manage their condition more effectively.[14] And in the cardiovascular arena, digital twins empowered by multi-modal data from imaging, wearables, and electrocardiograms can help predict how the heart reacts to different treatments and dosages.[15]

Quantum ML can predict alternative treatment pathways and their outcomes. Quantum support vector machines have been investigated to detect schizophrenia from electroencephalography (EEG) signals[16], and quantum neural networks have been explored to predict treatment outcomes in depression and anxiety and improve the diagnosis of patients with Parkinson's.[17]

Quantum AI has also been employed in enhancing encryption of subject-level data such as RWD and EHR. Notably, IBM developed quantum-safe cryptographic solutions that are particularly relevant for healthcare, where sensitive patient data require robust protection against quantum attacks.[18]

Citeline SmartSolutions

Citeline is incorporating AI, advanced ML, and LLMs in its SmartSolutions suite of products, which help study sponsors reduce costly protocol amendments, increase predictability in clinical trial planning, and accelerate clinical development.[19]

Protocol SmartDesign combines industry-leading data from Citeline's Trialtrove and Sitetrove solutions with real-world and proprietary performance data assets to build and deliver more reliable clinical trials. And Investigator SmartSelect leverages AI-enabled technology built on Sitetrove and Trialtrove data to deliver a list of high-performing investigators with the experience and capacity to deliver a clinical trial on time and on budget.

References

1 McKinsey & Co. (2024) Generative AI in the pharmaceutical industry: Moving from hype to reality. Available from https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality#/ (Accessed Aug. 21, 2024)

2 Jaspersoft (2024) What is Prescriptive Analytics? Available from https://www.jaspersoft.com/articles/what-is-prescriptive-analytics (Accessed Aug. 21, 2024)

3 Savage N (2021) Tapping into drug discovery potential of AI. Biopharma Dealmakers. Available from https://www.nature.com/articles/d43747-021-00045-7 (Accessed Aug. 21, 2024)

4 Cyclica (2017) Cyclica Launches Ligand ExpressTM, a Disruptive Cloud-Based Platform to Revolutionize Drug Discovery. Available from https://cyclicarx.com/press-releases/cyclica-launches-ligand-express-a-disruptive-cloud-based-platform-to-revolutionize-drug-discovery/ (Accessed Aug. 21, 2024)

5 Kamya P, Ozerov IV, Pun FW, et al (2024) PandaOmics: An AI-drived Platform for Therapeutic Target and Biomarker Discovery. Journal of Chemical Information and Modeling, 64(10). Available from https://pubs.acs.org/doi/10.1021/acs.jcim.3c01619 (Accessed Aug. 21, 2024)

6 Otten NV (2023) Data Quality In Machine Learning – Explained, Issues, How To Fix Them & Python Tools. Spot Intelligence. Available from https://spotintelligence.com/2023/04/07/data-quality-machine-learning/ (Accessed Aug. 22, 2024)

7 Varsheni S (2024) Your guide to missing values imputation. Train In Data. Available from https://www.blog.traininindata.com/your-guide-to-missing-values-imputation/ (Accessed Aug. 22, 2024)

8 Awan AA (2022) A Complete Guide to Data Augumentation. Datacamp. Available from https://www.datacamp.com/tutorial/complete-guide-data-augmentation (Accessed Aug. 22, 2024)

9 Guvvala S (2023) Bias Mitigation in Generative AI. Analytics Vidhya. Available from https://www.analyticsvidhya.com/blog/2023/09/bias-mitigation-in-generative-ai/ (Accessed Aug. 22, 2024)

10 Powell K (2024) NVIDIA Generative AI is Opening the Next Era of Drug Discovery and Design. NVIDIA. Available from https://blogs.nvidia.com/blog/drug-discovery-bionemo-generative-ai/ (Accessed Aug. 22, 2024)

11 Rosen-Zvi M (2021) Finding New Uses for Drugs with Generative AI. IBM. Available from https://research.ibm.com/blog/generative-ai-new-drugs (Accessed Aug. 22, 2024)

12 National Edge AI Hub. Available from https://edgeaihub.co.uk/edgeai-health/ (Accessed Aug. 22, 2024)

13 Poon H (2024) Multimodal Generative AI: the Next Frontier in Precision Health. Microsoft. Available from https://www.microsoft.com/en-us/research/quarterly-brief/mar-2024-brief/articles/multimodal-generative-ai-the-next-frontier-in-precision-health/ (Accessed Aug. 22, 2024)

14 Mosquera-Lopez C, Jacobs PG (2024) Digital twins and artificial intelligence in metabolic disease research. Trends in Endocrinology & Metabolism, 35(6), 549–557. Available from https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760(24)00113-9 (Accessed Aug. 22, 2024)

15 Corral-Acero J, Margara F, Marciniak M, et al (2020) The 'Digital Twin' to enable the vision of precision cardiology. European Hearth Journal, 41(48), 4556–4564. Available from https://academic.oup.com/eurheartj/article/41/48/4556/5775673?login=false (Accessed Aug. 22, 2024)

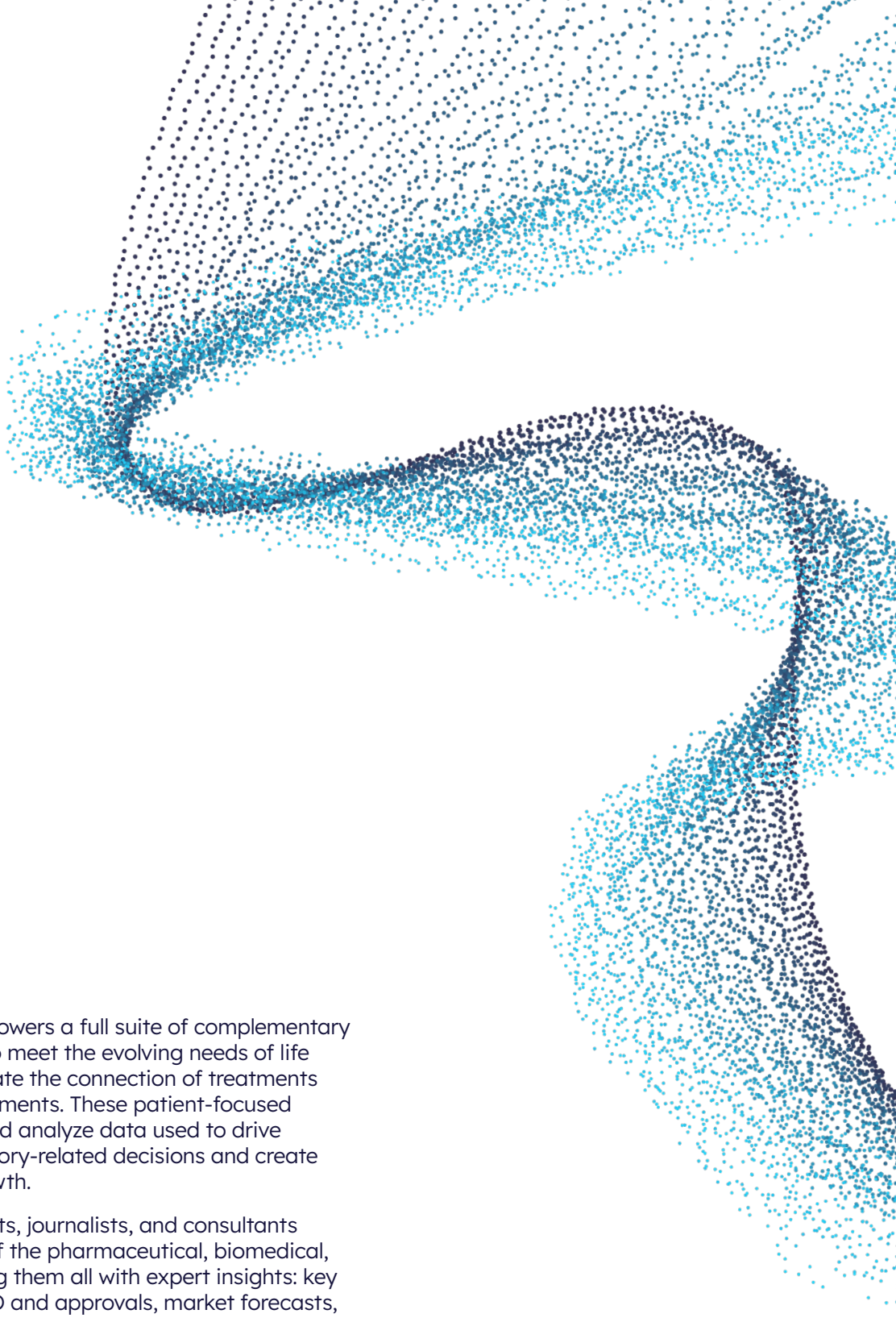
16 Aksoy G, Cattan G, Chakraborty S, et al (2024) Quantum Machine-Based Decision Support System for the Detection of Schizophrenia from EEG Records. Journal of Medical Systems, 48(29). Available from https://link.springer.com/article/10.1007/s10916-024-02048-0 (Accessed Aug. 22, 2024)

17 Stefano GB (2024) Quantum Computing and the Future of Neurodegeneration and Mental Health Research. Brain Sciences, 14(1), 93. Available from https://www.mdpi.com/2076-3425/14/1/93 (Accessed Aug. 22, 2024)

18 IBM (2024) IBM-Developed Algorithms Announced as NIST's First Published Post-Quantum Cryptography Standards. Available from https://newsroom.ibm.com/2024-08-13-ibm-developed-algorithms-announced-as-worlds-first-post-quantum-cryptography-standards (Accessed Aug. 22, 2024)

19 Citeline (2024) In a First for Pharma, Citeline SmartSolutions Take AI to New Levels, Optimizing Clinical Trial Planning and Site Selection. Citeline. Available from https://www.citeline.com/en/resources/introducing-smartsolutions (Accessed Aug. 23, 2024)

CITELINE
CLINICAL



Citeline, a Norstella company, powers a full suite of complementary business intelligence offerings to meet the evolving needs of life science professionals to accelerate the connection of treatments to patients and patients to treatments. These patient-focused solutions and services deliver and analyze data used to drive clinical, commercial, and regulatory-related decisions and create real-world opportunities for growth.

Citeline's global teams of analysts, journalists, and consultants keep their fingers on the pulse of the pharmaceutical, biomedical, and medtech industries, covering them all with expert insights: key diseases, clinical trials, drug R&D and approvals, market forecasts, and more. For more information on one of the world's most trusted health science partners, visit Citeline and follow on LinkedIn and X.
