IMS Health & Quintiles are now IQVIA™

White Paper

AI IN CLINICAL DEVELOPMENT

Improving safety and accelerating results

LUCAS GLASS, Global Head of Analytics Center of Excellence, IQVIA
GARY SHORTER, Head of Artificial Intelligence, R&D, IQVIA
RAJNEESH PATIL, Head of Process Design & Analytics, Centralized Monitoring, Services, IQVIA

TABLE OF CONTENTS

Introduction 3

Understanding AI and ML 3

AI and ML in clinical R&D 5

The benefits of AI: Enhancing operational efficiencies 5

Study design 5

Site identification and patient recruitment 6

Pharmacovigilance (PV) 6

Clinical monitoring 8

Patient care 9

AI and ML: Augmenting the future of clinical development 10

Talent, technology and expertise 10

About the authors 11

INTRODUCTION

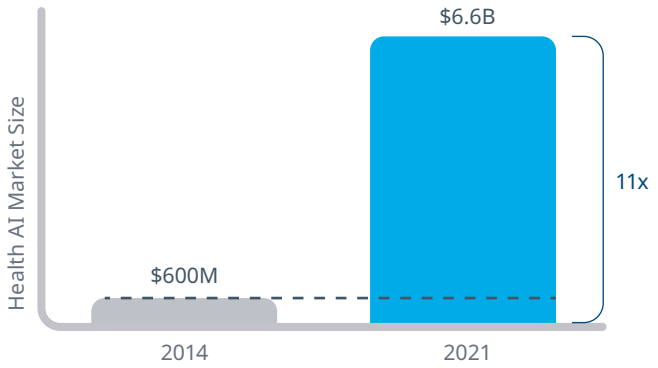
Artificial intelligence (AI) and machine learning (ML) tools are being leveraged across the clinical development landscape, delivering time and cost savings while reducing risks.

AI and ML tools are transforming how clinical development occurs, delivering significant time and cost efficiencies while providing better faster insights to inform decision making.

Advances in analytics technology coupled with the availability and integration of vast amounts of healthcare data have already helped automate processes and improve data quality across dozens of clinical development efforts.

As these tools evolve, new opportunities will continue to emerge that drive further benefits to the clinical research landscape. Applications of AI and ML in healthcare are expected to grow nearly $8 billion by 2022, up from $667.1 million in 2016; and almost half of global life science professionals say they are either using or interested in using AI in their research.[1]

Figure 1: Explosive growth in the AI health market[1]



1. Source: Accenture, "Artificial Intelligence: Healthcare's New Nervous System," https://www.accenture.com/us-en/insight-artificial-intelligence-healthcare

However, many people in the industry are still uncertain about what these technologies are and how they work. And many are unsure how to surmount the challenges required to leverage these technologies.

UNDERSTANDING AI AND ML

Artificial intelligence (AI) is not plug-and-play software. It is a sophisticated set of smart technologies that are customized to seek out and learn from data or experience; and then perform tasks based on what they learned while improving through experience. While AI can be used to mimic human intelligence, it can also uncover nuanced and complicated patterns that are far beyond human capabilities to process and identify.

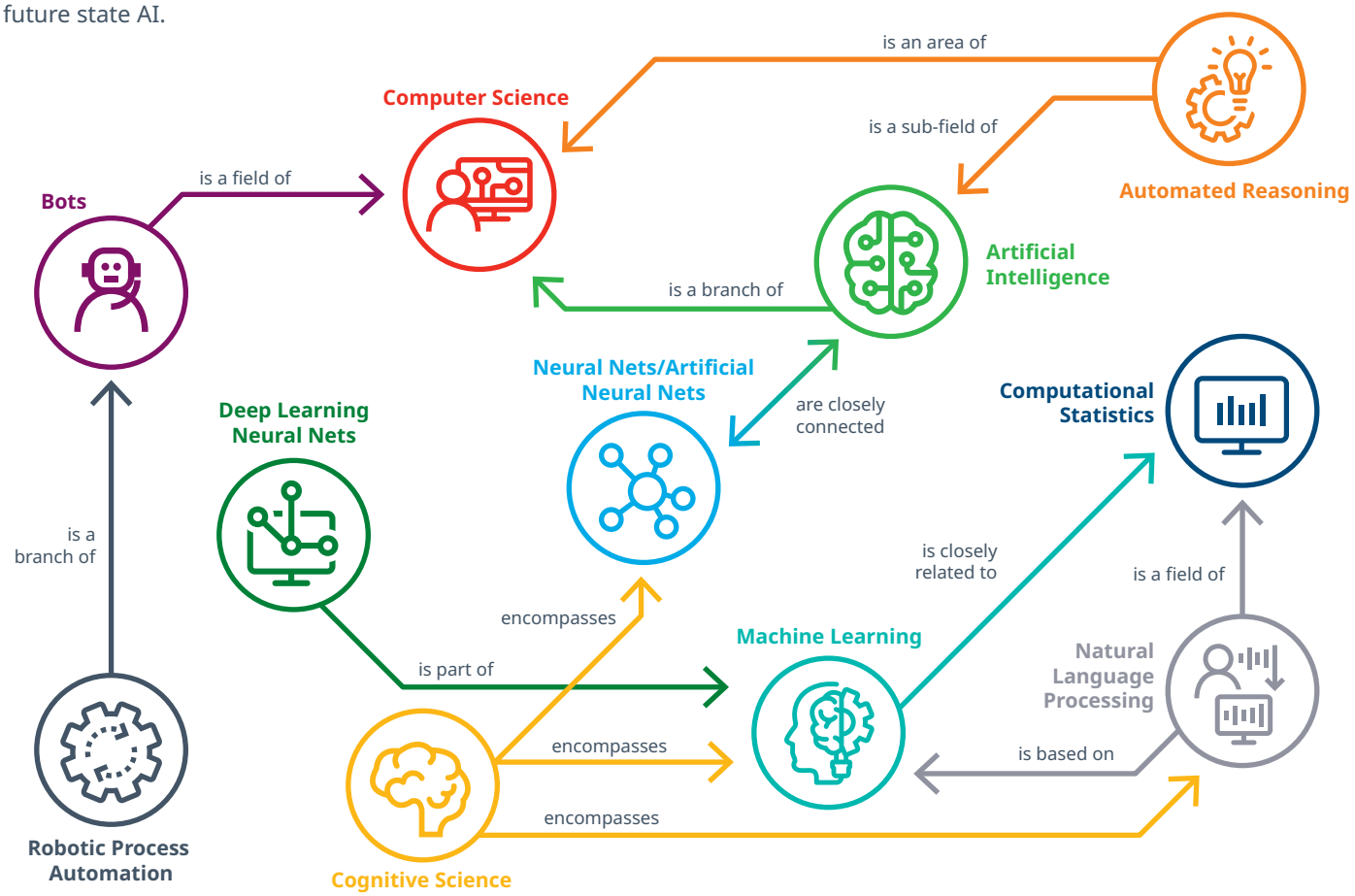
AI applications range from basic to incredibly complex. Simple AI applications can include tools like chatbots and automated phone screeners, which are taught proper response to different voice or text commands. On the more complex end of the spectrum, AI can analyze huge data sets to uncover trends and answer important questions – like which patients are most likely to respond to a certain treatment – then make predictions on how to respond or interpret what might be happening.

Applications of AI and ML in healthcare are expected to grow nearly $8 billion by 2022

Figure 2: AI and associated sciences

11 sciences constitute the full landscape of AI.

A mix of connections build upon one another, with a blend of data, tech, analytics and knowledge allowing the evolution into a more complex future state AI.

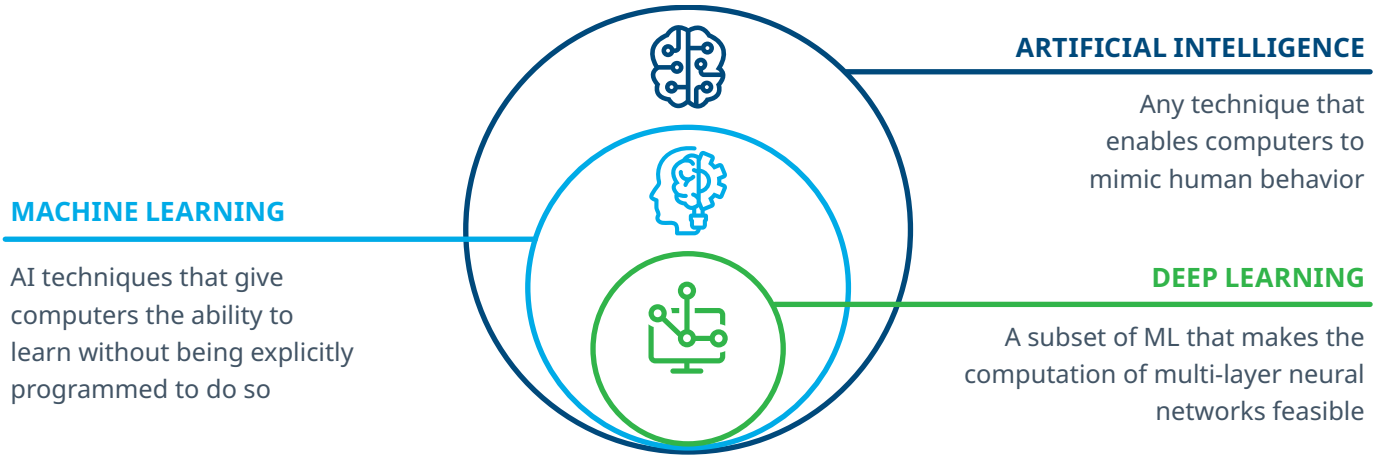


Machine learning (ML) is a branch of AI in which algorithms and statistical models are used to enable computers to perform specific tasks. Going beyond classic rules-based approaches, ML relies heavily on inference and patterns to address sophisticated challenges.

ML algorithms then leverage deep learning and complex mathematical techniques to analyze large data sets to find answers and predict outcomes. Over time the algorithm "learns" what patterns and information deliver the most useful results, then adapts future searches accordingly.

In general, more data helps ML algorithms improve at finding the right answers and predicting outcomes. Recent technology advances in big data have made it more practical to apply ML to massive data sets to derive insights which were previously impossible to detect.

As an example, AI and ML are already used in many common consumer applications. When Amazon® suggests a product you might like based on a past purchase, or Netflix® recommends a new show for you to watch, that is AI and ML at work.



AI AND ML IN CLINICAL R&D

These same techniques have a broad range of applications in clinical research and development. Applications of ML today include identifying molecules that have the most potential, finding patient populations that fit inclusion/exclusion criteria, and reviewing scans, claims reports and other healthcare data to interpret trends to improve and accelerate decision making.

These technologies promise to transform the way we plan and implement clinical research.

However, these incredible technological capabilities can only be leveraged by organizations that have access to the tools, expertise and expansive healthcare data sets to design and build effective algorithms and accurately interpret the results. But healthcare data is often unstructured and difficult to access, requiring extreme protection to ensure privacy. And, it only tells a portion of the story. In fact, we need to look beyond healthcare data and realize that predictors lie in human data sets.

Understanding, and using, human data science can create insights previously unattainable from employing only healthcare data. In short, we need to think differently about how data needs to be **collected, studied, combined** and **protected** to answer questions, solve problems and make decisions in healthcare.

When sponsors collaborate with partners that have the necessary technical and pharmaceutical expertise, they can achieve incredible time and cost savings, while reducing risks and improving the quality of their research.

THE BENEFITS OF AI: ENHANCING OPERATIONAL EFFICIENCIES

AI and ML are already delivering meaningful time, cost and quality benefits across various aspects of the clinical research process. Here are several examples of how it works:

Study design

Poor study design has catastrophic impact on the cost, efficiency and success potential of clinical trials. Leveraging vast healthcare data sets, we can use AI, ML and natural language processing (NLP) tools to assess and select optimal primary and secondary endpoints in study design, to ensure the most relevant protocols for regulators, payers and patients are defined. This helps to optimize the study design by informing ideal country and site strategies, enrollment models, patient recruitment and start-up plans.

Better study design leads to more predictable results, reduced cycle time for protocol development, fewer protocol amendments, and higher efficiencies throughout a study. It also results in improved recruitment rates, fewer non-enrolling sites, and fewer protocol amendments. These improvements increase chances of success and facilitate realistic and accurate planning.

Site identification and patient recruitment

Identifying trial sites to successfully perform clinical research with access to enough patients who meet inclusion/exclusion criteria is an ongoing challenge. As studies target more specific populations, recruiting goals become even harder to achieve, driving costs up, increasing timelines, and raising the risk of failure. According to Tufts Center for the Study of Drug Development (CSDD), nearly half of all sites miss enrollment targets.

One of the ways AI and ML can mitigate these risks is by identifying the sites with the highest recruitment potential and suggesting appropriate recruitment strategies. This involves mapping patient populations and proactively targeting sites with high predicted potential to deliver the most patients – before a single site is opened – and identifying the best avenues to recruit them. This means sponsors can open fewer sites, accelerate recruiting, and reduce the risk of under-enrollment.

Pharmacovigilance (PV)

In PV, massive amounts of structured and unstructured data must be integrated and reviewed in order to ensure quality and oversight. AI and ML technologies are addressing many of the challenges that PV units face in in harnessing the power of this data via new levels of insight and proactive analytics to enhance quality and oversight. These tools can be used to automate highly manual processing tasks, and translate and digitize safety case processing and adverse drug reaction (ADR) documents to make them more usable. They can also perform data listening tasks to monitor conversations on social media and other platforms, ensuring adverse events are promptly identified.

The insights derived from using AI for PV tasks lead to faster assessment of subject, site and study risks and overall study performance by domain experts, enabling project leads to drive efficiencies while improving patient safety.

Optical character recognition (OCR), natural language processing (NLP) and deep neural networks are being used to analyze and format structured and unstructured data for faster more efficient safety reviews.

+20.6%
increase in enrollment rate

IQVIA's data science team analyzes massive amounts of disease and region-specific healthcare data to locate targeted patient pools, as well as the research sites, healthcare facilities and physicians who treat them – even if those sites have never participated in clinical research.

Leveraging machine learning to synthesize these disparate data elements and to uncover meaningful insights enables targeted site identification to find the sites with ample access to the patient population. In ongoing oncology studies at IQVIA, this translates to a 20.6 percent increase in enrollment rate globally when comparing the ML-driven recommendations to the traditional experience-based site identification approach.

Figure 3: AI Safety and Translations

IQVIA SAFETY CASE PROCESSING

800,000+ cases processed annually

68,000+ safety regulatory reports

1/3 input data is structured or semi-structured

Auto Ingest of Adverse Events
Using optical character recognition to convert AE e-mails/PDFs digitally and importing to Safety System

Enhanced Coding Descriptions
Enriching the safety information by adding 3rd party ontology information

Expert Narrative Production
Machine learning from huge amounts of previous cases allowing more accurate resulting narratives

- Converting the majority of adverse event reporting media from various formats (often paper and some "dead" image scans) into a digitized and standard format. Full auditability
- Improving efficiency and productivity
- Leveraging semantic ontology data to enrich the information
- Producing a new IQVIA XML digital library for learning
- 20 years of history to learn from

70%
automated processing of ADR case reports

IQVIA's team is using AI and ML tools to help a large pharmaceutical firm process roughly 8,000 case lifecycles per week, allowing up to 70 percent of ADR case reports to be processed with no human intervention.

IQVIA TRANSLATION CAPABILITIES

130M+ words translated annually

93,000+ documents fed through Neural Machine Translation annually

50% global coverage through 11 language pairs

Auto-Identify Document Type
Using optical character recognition to identify, strip format, images, etc. and tag into structured definitions

Enhanced Language Translation
Enrich the Translator with pre- and post-processing specific clinical trial engines that augment translation requirements

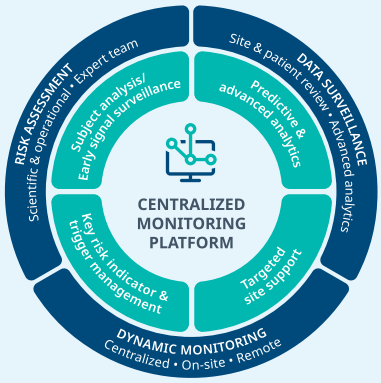
Integrated Translation into Digital Workflow
Collating segments back into formatted document and integrating into workflow management for follow-up

- Spanish, French, Korean, German, Japanese, Chinese, Russian, Polish and other languages
- Linkage into workflow systems (such as regulatory start-up)
- Bleu Score from 55% (Google) to 85% with our pre- and post-processing
- Documents include: informed consent, clinical protocol, ethics committee letters, investigator brochures, etc.
- Generic micro service capability across many value streams

Clinical monitoring

Tremendous manual effort is spent analyzing site risks and generating "action items" to mitigate those risks. AL and ML concepts can alleviate these pressures by manually assessing the risk environment and delivering predictive analytics to generate more effective clinical monitoring insights.

Advanced analytics provide composite site rankings for holistic risk assessments of sites, allowing for more specific identification of risks and removal of false positives. Using the composite evaluation of site risks across the study to produce a simple output quickly shows high risk sites, key risk indicators (KRIs) and site risk rank. They can also be used to proactively identify which sites are more likely to have recruitment and performance issues, or which patients are at higher risk for potential AEs. With these insights, we can take actions faster and avoid potential issues.



Using IQVIA's AI-driven centralized monitoring platform, clients have achieved:

37% reduction in query resolution times

31% lower source data verification (SDV) backlog

13 DAYS faster time to database lock vs. non-risk-based monitoring studies

IQVIA is currently using automation of action-decision frameworks where identification of risk, analysis of site performance, and communication of action items is conducted through AI and ML models. The ML algorithms are being used to auto-identify subjects with abnormal test results and detect vital sign outliers in lab analyses.

Figure 4: Machine Learning – Monitoring

IQVIA CENTRALIZED MONITORING

>10,000 records, running time 37 sec. 1,230 duplicates found

>85% accuracy in identification of outliers (small sample), with scale we would expect ~ 95%

89% accuracy to detect the vital signs outliers effectively

Duplicate Subject Detection
Using machine learning to come up with the best rules to automatically find similar records in dataset

Laboratory Tests Outlier Detection
Propose a dissimilarity measure to quantify the differences between subject records for flagging as outliers

Vital Signs Outlier Detection
Identify subjects with large or small variations from baseline using Boxplot, Normal Distribution, and linear regression

- New algorithms built to assist in the day-to-day monitoring of subjects on clinical trials
- Successful in 3 initials areas of focus chosen:
  - » Use of machine learning to detect duplicate subject registrations
  - » To auto-identify subjects with abnormal lab test results
  - » To detect any vital sign outliers

Patient care

Disease detection algorithms are now being designed to leverage medical information such as symptoms and procedures that typically precede a diagnosis in order to identify patients who are very likely to develop a disease. This allows for proactive care, as well as new recruiting insights for prodromal or early stage disease studies, and those that require treatment naive patients.

One area where this type of model is having a profound impact is Alzheimer's disease (AD) research. Due to the exponential effects of delays of the diagnosis of AD, much clinical research activity in AD is focused on the prodromal stage. Yet traditional screenings for prodromal AD patients delivers only a 20 percent precision rate.

79%

precision rate for predictive algorithm

IQVIA is partnering with hospitals and healthcare organizations to conduct data-driven disease detection for patient referrals across a wide portfolio of indications. These algorithms can quickly analyze the symptoms and patient features in the patient file, suggest a trial, and trigger a referral to the doctor in real time. Significant insights about drivers for these predictions can also generate ML-enabled screening tools to facilitate the providers' ability to leverage at the point-of-care.

To support recruiting for a pre-symptomatic AD study, IQVIA's data science team created a prodromal AD predictive algorithm and assessed data sets of 73 million de-identified US residents. The result was a 79 percent precision rate in screening – almost four times better than traditional methods. Roughly 223,000 potential prodromal AD patients were isolated within the data, 75 percent of whom were in primary care settings. These insights were then operationalized through referral networks to accelerate recruiting.

From 72,670,283 U.S. residents, ~223K prodromal AD patients were identified

76% of the identified patients are in the primary care setting

These data insights can then be operationalized through referral networks

AI AND ML: AUGMENTING THE FUTURE OF CLINICAL DEVELOPMENT

The clinical development landscape is changing quickly, and sponsors need to have the best insights to make the right decisions to increase predictability, reduce time to market and improve efficiencies. By using these methods in clinical development, we can introduce change to prove – and improve – how AI-derived insights can be applied to many aspects of development.

These technologies aren't going to replace human expertise. But they are going to accelerate our ability to analyze the data and take meaningful action in response to create a safer, more streamlined research environment. As the volume and complexity of clinical data continue to increase, and we bring in more real-world evidence, AI applications possess potential value to the clinical trial process and the pharmaceutical drug discovery industry.

TALENT, TECHNOLOGY AND EXPERTISE

These technologies hold huge promise for clinical development, but that promise can only be achieved by organizations that have the tools, talent and partners to leverage these technologies in relevant ways. To achieve the full potential of AI and ML in clinical development, pharma companies need partners who can provide:

DEEP PHARMACEUTICAL KNOWLEDGE AND DOMAIN EXPERTISE

This expertise should include in-depth understanding of healthcare data for insightful analyses, understanding of global regulatory environments, payer expectations, physician and patient behavior, and therapeutic knowledge.

STATE-OF-THE-SCIENCE AI AND ML TECHNOLOGIES

The solution should be capable of mining multiple data sets with speed and scale to identify high-level global and regional trends, as well as detailed physician and patient-level insights while applying rigor through GXP Software Development Life Cycle (SDLC) approach to produce top quality models.

DATA ANALYTICS AND MACHINE LEARNING EXPERTS

The customer team should include technical experts who are skilled in crafting machine learning algorithms relevant to the clinical development process.

ACCESS TO VAST HEALTHCARE DATA SETS

Machine learning algorithms are only as good as the data they can access. The most effective platforms will provide access to multiple global healthcare databases, including prescription data, EMRs, pharma sales data, and patient and disease trend data that are updated regularly.

THE ABILITY TO INTEGRATE THESE DISPARATE DATA SETS

Many healthcare data sets are unstructured, or inconsistent in form and formatting, making them difficult to analyze. A good partner will have strategies in place to "clean" the data so the algorithms can interpret results and translate them into actionable insights that drive better results.

With changes in the life sciences industry, a higher level of analytics is required to succeed. The goal is to advance our understanding of human health through better, more insightful decisions. Integrating human science with breakthroughs in data science and technology provides more relevancy and precision to decision makers. It can transform how we diagnose patients and treat their conditions, with minimal errors. How we can find patients faster – maybe even before they are patients.

The proven promise and huge potential of using AI and ML to accelerate drug discovery while cutting costs and risks is a tremendous catalyst for innovation. By continuing to leverage data, intelligence, analytics and domain expertise, we have the ability to transform the clinical development landscape in ways that will benefit patients, sponsors, payers and physicians alike.

ABOUT THE AUTHORS

LUCAS GLASS
Global Head of Analytics Center of Excellence, IQVIA

Lucas Glass is the global head of IQVIA's Analytics Center of Excellence. The Center is responsible for researching, developing and operationalizing machine learning and data science solutions in the R&D business.

Lucas started his career in pharmaceutical data science 15 years ago at Center (then owned by Galt Associates), working on pharmacovigilance data mining algorithms. Since then he has worked at the U.S. Department of Justice in healthcare fraud, several small startups, and TTC, LLC, which was acquired by IMS Health in 2012.

Lucas holds a BA in physics from Boston University, and an MS in biostatistics from Drexel University, and is a Ph.D. candidate at Temple University, where he is researching deep learning embedding techniques on large-scale healthcare data.

GARY SHORTER
Head of Artificial Intelligence, R&D, IQVIA

Gary Shorter holds an MSc and has served as a global biostatistics lead for multiple compounds in clinical trials. His 25-plus years of experience allows him to bring the same level of quality and domain expertise to the realm of AI, to ensure that quality AI tools are built and validated to the rigor of regulatory agencies' expectations. His recent products include Auto-eTMF and Auto-Translation specifically trained to clinical operations needs.

RAJNEESH PATIL
Head of Process Design & Analytics, Centralized Monitoring, Services, IQVIA

Rajneesh Patil's expertise spans process design, project portfolio management, risk-based monitoring and advanced analytics for clinical trial applications. He is a dental surgeon by training, and a Six Sigma Greenbelt with an MBA in operations management.

CONTACT US

iqvia.com/contactus

IMS Health & Quintiles are now