



Mark Kiel, MD, PhD

# ASK THE EXPERT

## Information Overload: When Your Data Is Too Much to Process, Turn to AI/ML

by Marnie Willman, BSc

**Mark Kiel**, MD, PhD, was a molecular genetic pathology fellow at the University of Michigan and is the founder and CSO of Genomenon, where he oversees the company's scientific direction and product development. Kiel's passion is to power the practice of precision medicine by organizing the world's genomic knowledge. To that end, he created Genomenon and the Mastermind suite of genomic tools.

**Q:** What are the clinical diagnostic applications of artificial intelligence (AI) and machine learning (ML)?

**A:** AI and ML have many applications in the clinical space. For instance, in clinical diagnostics, they can prioritize genetic variants for review. A patient's sequencing results typically contain numerous genetic variants that may be contributing to disease. Sometimes, in well-established disease-gene relationships (such as *CFTR* variants and cystic fibrosis), the causative variants are obvious. In other cases, especially for more clinically heterogeneous diseases or those with less specific clinical findings, the causative variant is not so obvious. In these situations, AI/ML models trained on historical and empirical data can be useful to nominate or prioritize potential diagnostic variants.

AI and ML can also be used to predict the consequences of novel variants. When a patient presents with a variant that has not been seen before, its role in causing disease is unknown. Even if the gene in question is associated with the disease, questions may remain if the specific variant lacks sufficient evidence to support causation. In this case, improvements in predictive

models of pathogenicity can leverage advanced AI/ML techniques.

Finally, AI and ML can organize, annotate, and prioritize evidence from biomedical literature. The day-to-day activities of variant analysts require frequent in-depth consultation with

**“Combining AI/ML with expert review saves time and improves accuracy.”**

empirical evidence in previous scientific and clinical publications to identify clinical cases and/or functional studies of genetic variants' effects on protein function. Given the daunting challenge of searching through this vast amount of complex information, AI/ML tools are ideally suited to help streamline this work.

Large language models (LLMs) like ChatGPT can also increase the speed of interpretation by summarizing the available evidence and extracting crucial information. Instead of requiring human eyes to review all of the literature, the software can index all of the material provided and prioritize only those that are most relevant.

**Q:** What are some unique challenges—and solutions—when incorporating AI/ML in a clinical lab?

**A:** Accuracy is paramount because it influences a patient's diagnosis and, subsequently, their treatment. Answers and predictions must be fully explained

and backed by evidence. Combining AI/ML with expert review saves time and improves accuracy. The AI output must be verified and supported by evidence or citations; the absence of this support creates more work for curators, who may pursue false leads or spend time correcting inaccurate or unsubstantiated information. Explaining the origin of the result and providing verifiable sources for assertions not only reduces the work required, but also increases the accuracy of the results.

**Q:** What role does human review play in AI/ML systems?

**A:** AI/ML can augment curators' abilities to make variant interpretation faster and more comprehensive, but human

review is still necessary to ensure the accuracy of the final classification. AI/ML is subject to false-positive artifacts and occasional faulty reasoning. In clinical care, where accuracy is paramount, it is necessary to have a human reviewer to ensure that no erroneous information is conveyed to the patient or used to inform their care. Using AI/ML, the data can be organized, annotated, prioritized, and presented to the human end user. It may even have a provisional summary of the AI/ML's determination, but it remains the human reviewer's duty to carefully and thoroughly review this work so that no errors are propagated.

**Q:** How can AI/ML be applied to drug discovery pipelines? What are the benefits?

**A:** AI/ML can be a solution to the information overload that often results from basic research in the drug discovery pipeline. Large amounts of data dispersed across multiple locations can be difficult to aggregate and assess in a timely or complete manner. AI/ML can rapidly ingest vast amounts of information, summarize it, and identify novel associations that can lead to new drug targets.

Laboratorians must fully recognize not just the benefits, but also the limitations of AI in drug discovery. Although it can be incredibly useful for rapidly assessing large amounts of information, it is important to remember that AI is not a trained scientist and may fail to appreciate nuances. Combining AI/ML results with review by a trained expert will make the most of its benefits while compensating for its limitations.

**Q:** What does explainability of AI output mean, and how does this influence Genomenon's work?



▲ Genomenon founder and CSO Mark Kiel at company headquarters. Credit: Genomenon

**A:** Explainability is a major point of focus for Genomenon because we work in an industry that impacts patients' lives. We put emphasis on combining AI with expert review to ensure that all decisions are backed by sufficient evidence. “Explainability” means that you can understand how an AI model reached a certain decision, which is where that vital expert review comes into play.

Genomenon's work is centered on indexing and annotating the evidence, followed by expert human review. During the human review, our curators' work is greatly expedited by the pre-annotation, prioritization, and organization of evidence that has been indexed by the AI systems we have in place. Without the annotations and the explanations of their origins, tracing back the information to confirm (or reject) the AI's conclusions would be much more challenging, time-consuming, and error-prone.

**Q:** What technological developments in AI/ML are you most excited about and why?

**A:** I am excited about the recent developments in large language models (LLMs), especially their application to biomedical information, as their training becomes more precise. There is a great deal of optimism and enthusiasm around LLMs, especially the models most people know as ChatGPT, BARD, or similar. I like to conceptualize these as simply “calculators for words,” but with the added benefit of having access to vast amounts of information (usually through the internet or other large repositories of training data) and the ability to iteratively learn and refine the quality of their output. I have often thought that, although language is not the only aspect of intelligence, it is a substantial component. I believe that there is a great promise in these language models, once refined, to contribute to the emergence of truly intelligent machines. For the time being, their use in genomics will be supportive, with appropriate checks and balances coming from expert (human) curators.

*Marnie Willman, BSc, is the clinical writer at Today's Clinical Lab.*