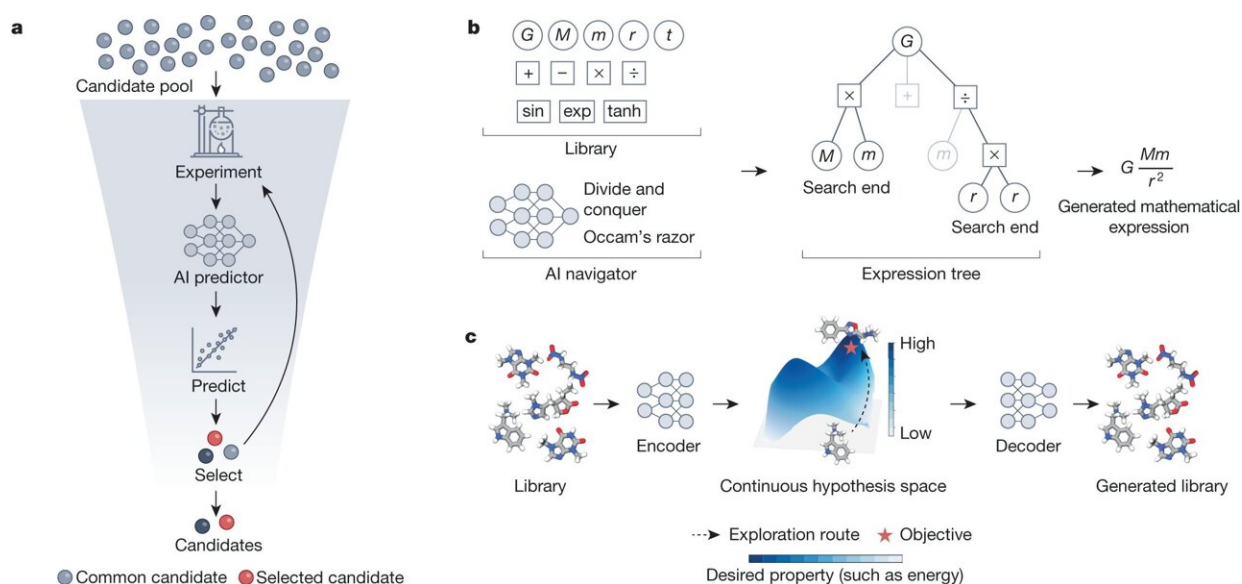# Q&A: Scientist discusses artificial intelligence beyond the clinic

August 16 2023, by Ekaterina Pesheva

AI-guided generation of scientific hypotheses. Credit: *Nature* (2023). DOI: 10.1038/s41586-023-06221-2

With the advent of ChatGPT4, the use of artificial intelligence in medicine has absorbed the public's attention, dominated news headlines, and sparked vigorous debates about the promise and peril of medical AI.

But the potential of AI reaches far beyond the frontlines of medicine.

AI is already changing the way scientists discover and design drugs. It is

predicting how molecules interact and proteins fold with never-before-seen speed and accuracy. One day, AI may even be used routinely to safeguard the function of nuclear reactors.

These are but a few of the exciting applications of AI in the [natural sciences](#), according to a commentary in Nature authored by Marinka Zitnik, assistant professor of biomedical informatics at Harvard Medical School. Zitnik led a team of author-researchers from 36 academic and industry labs from around the globe.

Zitnik, who is also an associate faculty member at the Kempner Institute for the Study of Natural & Artificial Intelligence at Harvard University, discussed the growing role of AI in science and discovery. The research is published in the journal *Nature*.

Harvard Medicine News: We've been deluged with news and commentaries about the use of AI in medicine, but we are not hearing as much about AI in science and discovery beyond medicine. Why is that?

Zitnik: I think it's because the realization of the vast opportunity that AI represents for the [life sciences](#) and the natural sciences more broadly has not happened yet. The practice of science may vary across disciplines, but the scientific method that helps us explain the natural world constitutes a universal, fundamental principle across all disciplines.

The [scientific method](#) has been around since the 17th century, but the techniques used to generate hypotheses, gather data, perform experiments, and collect measurements can now all be enhanced and accelerated through the thoughtful and responsible use of AI.

HMNews: Where do you see the most immediate

impact of AI in scientific discovery?

Zitnik: Discoveries made with the combined use of human expertise and artificial intelligence already affect our everyday lives. AI is used to synthesize novel drugs. It is used to design new materials with properties that make them robust and stiff to support the construction of bridges and buildings.

AI algorithms have been used to provide real-time feedback and control of stratospheric balloons for weather forecasting. In physics, which can seem so far away from everyday life, recently developed AI algorithms were used to control a tokamak simulator—a nuclear fusion reactor in development—to make its safe operation less dependent on human intuition and experience alone.

HMNews: What are you excited about in the long term?

Zitnik: I'm very excited about the potential of AI to not only contribute to scientific understanding, but to acquire it autonomously to generate knowledge on its own. It's been shown that AI models can capture complex scientific concepts, such as the periodic table of elements, from the literature without any guidance.

The capacity to develop autonomous knowledge can guide future discoveries embedded in past publications. For example, this could be the discovery of a molecule to treat Alzheimer's disease. Such a discovery would require identifying indirect relationships across publications and across disciplines—chemistry, biology, medicine—connecting chemical properties of molecules to biologic behavior of molecular pathways implicated in Alzheimer's disease and then to clinical phenotypes and patients' symptoms.

Connecting all these disciplines and publications to identify shared principles and generate a novel hypothesis would be impossible for a human. AI "co-pilots" could read not only scientific publications but also raw research data, images, and experimental laboratory data and then extract latent knowledge and present it as a hypothesis for evaluation by human experts. This requires AI models to formulate hypotheses that are neither written down nor directly implied or suggested in existing scientific literature.

These are the challenges that consume most of a scientist's time and often differentiate very good scientists from exceptional ones. We hope that in the future scientists would spend less time doing routine laboratory work and more time guiding, accessing, and evaluating AI hypotheses and steering AI models toward the research questions they're interested in.

Another exciting possibility is the idea of human-in-the-loop AI-driven design, discovery, and evaluation. It would be possible to automate routine scientific workflows and combine actual experimentation in the physical world with virtual AI models and robotics.

This would allow us to leverage predictions and conduct experiments in a high-throughput manner. It would create self-driving laboratories where some of the experiments would be directly guided by predictions and outputs made by AI models.

HMNews: What are some of the pitfalls you foresee? Where should we tread extra carefully?

Zitnik: One challenge relates to practical considerations. Implementing and integrating a model with laboratory equipment requires lots of work and [complex software](#) and hardware engineering, the curation of the

data, and better user interfaces. Currently, minor variations in software and hardware can lead to considerable changes in AI performance.

Thus, it becomes risky to couple virtual AI tools with actual physical devices that can operate in the real world. Data and models need to be standardized. Ultimately, if done properly, I would expect to see the emergence of self-driving labs and semi-autonomous discovery engines.

Another challenge relates to machine learning foundations. There are gaps in what algorithms currently can do versus what we need them to do to be used in a routine manner. Scientific data are multi-modal, such as black holes in cosmology, natural language in scientific literature, biological sequences like amino acids, and 3D molecular and atomic structures. Integrating these data is challenging but necessary because looking at any data set in isolation cannot give a holistic view of the problem.

Another important challenge is that most AI models today still operate as black boxes. This means that scientists, the users, cannot fully understand or explain how these models operate. That's a challenge because scientific understanding is at the heart of advancing science. How to develop more transparent, deep learning models? This remains elusive.

The misapplication and misuse of AI is yet another challenge. Algorithms can be developed for one purpose but used for another. This can create vulnerabilities to manipulation. For example, in the molecular sciences, we've seen increasing use of generative AI to design molecular structures. AI can generate structures that have drug-like properties, representing molecules that would be delivered to specific tissues, which makes them promising drug candidates. However, one could take the exact same algorithm and tweak the criteria.

Thus, instead of optimizing molecules to behave like medicines, the algorithm could generate molecules that resemble bioweapons. There should be a critical conversation around what is responsible use of AI in science. We need to think about establishing ethics review processes and implementation guidelines that currently do not exist.

HMNews: What do you see as some of the solutions?

Zitnik: Addressing the challenges will require new modes of thinking and collaboration. Moving forward, we have to change how research teams are formed. We expect to see more AI specialists and software and hardware engineers become critical members of scientific research teams.

We expect novel forms of collaboration involving government at all levels, corporations, and educational institutions. Involving corporations is important because as AI models continue to grow in size, training these models will require resources that generally exist only in a handful of big tech companies.

Universities, on the other hand, are better integrated across disciplines. Only at universities do we have departments of chemistry, biology, physics, and sociology, and so forth. Thus, academia is better positioned to understand and study how to prevent the various risks and misuses of AI.