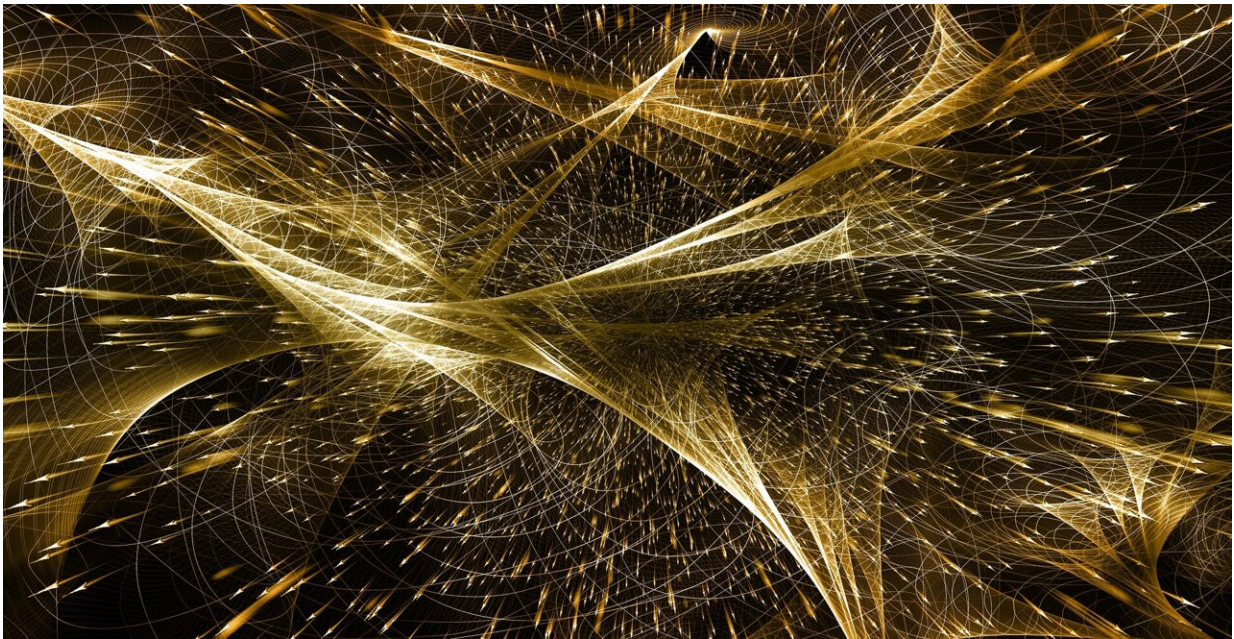# Scientists begin building AI for scientific discovery using tech behind ChatGPT

October 13 2023

Credit: Pixabay/CC0 Public Domain

An international team of scientists, including from the University of Cambridge, have launched a new research collaboration that will leverage the same technology behind ChatGPT to build an AI-powered tool for scientific discovery.

While ChatGPT deals in words and sentences, the team's AI will learn from numerical data and physics simulations from across scientific

fields to aid scientists in modeling everything from supergiant stars to the Earth's climate.

The team launched the initiative, called Polymathic AI earlier this week, alongside the publication of a series of related papers on the *arXiv* open access repository.

"This will completely change how people use AI and machine learning in science," said Polymathic AI principal investigator Shirley Ho, a group leader at the Flatiron Institute's Center for Computational Astrophysics in New York City.

The idea behind Polymathic AI "is similar to how it's easier to learn a new language when you already know five languages," said Ho.

Starting with a large, pre-trained model, known as a foundation model, can be both faster and more accurate than building a scientific model from scratch. That can be true even if the training data isn't obviously relevant to the problem at hand.

"It's been difficult to carry out academic research on full-scale foundation models due to the scale of computing power required," said co-investigator Miles Cranmer, from Cambridge's Department of Applied Mathematics and Theoretical Physics and Institute of Astronomy. "Our collaboration with Simons Foundation has provided us with unique resources to start prototyping these models for use in basic science, which researchers around the world will be able to build from—it's exciting."

"Polymathic AI can show us commonalities and connections between different fields that might have been missed," said co-investigator Siavash Golkar, a guest researcher at the Flatiron Institute's Center for Computational Astrophysics.

"In previous centuries, some of the most influential scientists were polymaths with a wide-ranging grasp of different fields. This allowed them to see connections that helped them get inspiration for their work. With each scientific domain becoming more and more specialized, it is increasingly challenging to stay at the forefront of multiple fields. I think this is a place where AI can help us by aggregating information from many disciplines."

The Polymathic AI team includes researchers from the Simons Foundation and its Flatiron Institute, New York University, the University of Cambridge, Princeton University and the Lawrence Berkeley National Laboratory. The team includes experts in physics, astrophysics, mathematics, artificial intelligence and neuroscience.

Scientists have used AI tools before, but they've primarily been purpose-built and trained using relevant data.

"Despite rapid progress of machine learning in recent years in various [scientific fields](), in almost all cases, machine learning solutions are developed for specific use cases and trained on some very specific data," said co-investigator Francois Lanusse, a cosmologist at the Center national de la recherche scientifique (CNRS) in France.

"This creates boundaries both within and between disciplines, meaning that scientists using AI for their research do not benefit from information that may exist, but in a different format, or in a different field entirely."

Polymathic AI's project will learn using data from diverse sources across physics and astrophysics (and eventually fields such as chemistry and genomics, its creators say) and apply that multidisciplinary savvy to a wide range of scientific problems. The project will "connect many seemingly disparate subfields into something greater than the sum of

their parts," said project member Mariel Pettee, a postdoctoral researcher at Lawrence Berkeley National Laboratory.

"How far we can make these jumps between disciplines is unclear," said Ho. "That's what we want to do—to try and make it happen."

ChatGPT has well-known limitations when it comes to accuracy (for instance, the chatbot says 2,023 times 1,234 is 2,497,582 rather than the correct answer of 2,496,382). Polymathic AI's project will avoid many of those pitfalls, Ho said, by treating numbers as actual numbers, not just characters on the same level as letters and punctuation. The training data will also use real scientific datasets that capture the physics underlying the cosmos.

Transparency and openness are a big part of the project, Ho said. "We want to make everything public. We want to democratize AI for science in such a way that, in a few years, we'll be able to serve a pre-trained model to the community that can help improve scientific analyses across a wide variety of problems and domains."