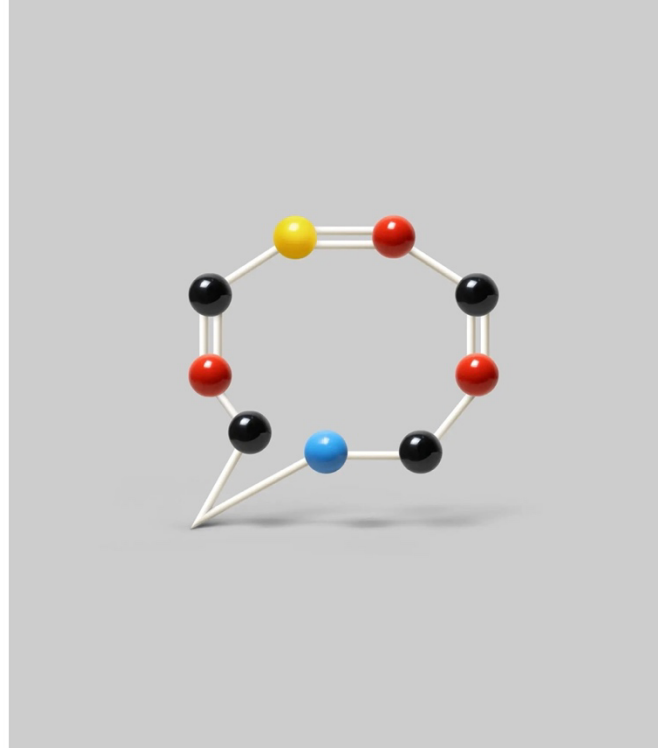


A.I. Is Learning What It Means to Be Alive

Given troves of data about genes and cells, A.I. models have made some surprising discoveries. What could they teach us someday?



By **[Carl Zimmer](#)**

- Published March 10, 2024 Updated March 12, 2024

In 1889, a French doctor named Francois-Gilbert Viault climbed down from a mountain in the Andes, drew blood from his arm and inspected it under a microscope. Dr. Viault's red blood cells, which ferry oxygen, had surged 42 percent. He had discovered a mysterious power of the human body: When it needs more of these crucial cells, it can make them on demand.

In the early 1900s, scientists theorized that a hormone was the cause. They called the theoretical hormone erythropoietin, or “red maker” in Greek. Seven decades later, researchers found actual erythropoietin after filtering [670 gallons of urine](#).

And about 50 years after that, biologists in Israel announced they had found a rare kidney cell that makes the hormone when oxygen drops too low. It's called [the Norn cell](#), named after the Norse deities who were believed to control human fate.

It took humans 134 years to discover Norn cells. Last summer, computers in California discovered them on their own in just six weeks.

The discovery came about when researchers at Stanford programmed the computers to teach themselves biology. The computers ran an artificial intelligence program similar to ChatGPT, the popular bot that became fluent with language after training on [billions of pieces of text](#) from the internet. But the Stanford researchers trained their computers on raw data about millions of real cells and their chemical and genetic makeup.

The researchers did not tell the computers what these measurements meant. They did not explain that different kinds of cells have different biochemical profiles. They did not define which cells catch light in our eyes, for example, or which ones make antibodies.

The computers crunched the data on their own, creating a model of all the cells based on their similarity to each other in a vast, multidimensional space. When the machines were done, they had learned [an astonishing amount](#). They could classify a cell they had never seen before as one of over 1,000 different types. One of those was the Norn cell.

“That’s remarkable, because nobody ever told the model that a Norn cell exists in the kidney,” said Jure Leskovec, a computer scientist at Stanford who trained the computers.

The software is one of several new A.I.-powered programs, known as foundation models, that are setting their sights on the fundamentals of biology. The models are not simply tidying up the information that biologists are collecting. They are making discoveries about how genes work and how cells develop.

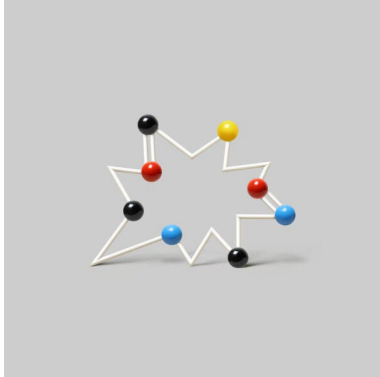
As the models scale up, with ever more laboratory data and computing power, scientists predict that they will start making more profound discoveries. They may reveal secrets about cancer and other diseases. They may figure out recipes for turning one kind of cell into another.

“A vital discovery about biology that otherwise would not have been made by the biologists — I think we’re going to see that at some point,” said Dr. Eric Topol, the director of the Scripps Research Translational Institute.

Just how far they will go is a matter of debate. While some skeptics think the models are going to hit a wall, more optimistic scientists believe that

foundation models will even tackle the biggest biological question of them all: What separates life from nonlife?

Heart Cells and Mole Rats



Credit...Doug Chayka

Biologists have long sought to understand how the different cells in our bodies use genes to do the many things we need to stay alive.

About a decade ago, researchers started industrial-scale experiments to fish out genetic bits from individual cells. They recorded what they found in catalogs, or “[cell atlases](#),” that swelled with billions of pieces of data.

Dr. Christina Theodoris, a medical resident at Boston Children’s Hospital, was reading about a [new kind of A.I. model](#) made by Google engineers in 2017 for language translations. The researchers provided the model with millions of sentences in English, along with their translations into German and French. The model developed the power to translate sentences it hadn’t seen before. Dr. Theodoris wondered if a similar model could teach itself to make sense of the data in cell atlases.

In 2021, she struggled to find a lab that might let her try to build one. “There was a lot of skepticism that this approach would work at all,” she said.

Shirley Liu, a computational biologist at the Dana-Farber Cancer Institute in Boston, gave her a shot. Dr. Theodoris pulled data from 106 published human studies, which collectively included 30 million cells, and fed it all into a program she created called GeneFormer.

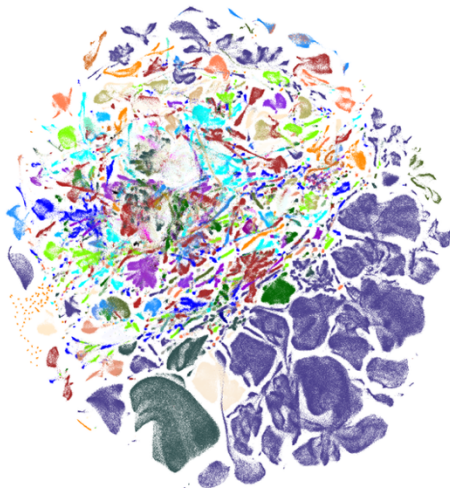
The model gained [a deep understanding](#) of how our genes behave in different cells. It predicted, for example, that shutting down a gene called TEAD4 in a

certain type of heart cell would severely disrupt it. When her team put the prediction to the test in real cells called cardiomyocytes, the beating of the heart cells grew weaker.

In another test, she and her colleagues showed GeneFormer heart cells from people with defective heartbeat rhythms as well as from healthy people. “Then we said, Now tell us what changes we need to happen to the unhealthy cells to make them healthy,” said Dr. Theodoris, who now works as a computational biologist at the Gladstone Institutes in San Francisco.

GeneFormer recommended reducing the activity of four genes that had never before been linked to heart disease. Dr. Theodoris’s team followed the model’s advice, knocking down each of the four genes. In two out of the four cases, the treatment improved how the cells contracted.

The Stanford team got into the foundation-model business after helping to build one of the biggest databases of cells in the world, known as [CellXGene](#). Beginning in August, the researchers trained their computers on the 33 million cells in the database, focusing on a type of genetic information called messenger RNA. They also fed the model the three-dimensional structures of proteins, which are the products of genes.



Computers learned how to classify over a thousand types of cells based on how their genes turn on and off. This map shows how they organized 36 million cells into clusters. Credit...Jure Leskovec

From this data, the model — known as Universal Cell Embedding, or U.C.E. — calculated the similarity among cells, grouping them into more than 1,000 clusters according to how they used their genes. The clusters corresponded to types of cells discovered by generations of biologists.

U.C.E. also taught itself some important things about how the cells develop from a single fertilized egg. For example, U.C.E. recognized that all the cells in the body can be grouped according to which of three layers they came from in the early embryo.

“It essentially rediscovered developmental biology,” said Stephen Quake, a biophysicist at Stanford who helped develop U.C.E.

The model was also able to transfer its knowledge to new species. Presented with the genetic profile of cells from an animal that it had never seen before — a naked mole rat, say — U.C.E. could identify many of its cell types.

“You can bring a completely new organism — chicken, frog, fish, whatever — you can put it in, and you will get something useful out,” Dr. Leskovec said.

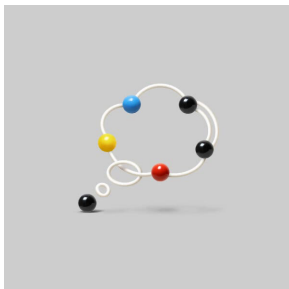
After U.C.E. discovered the Norn cells, Dr. Leskovec and his colleagues looked in the CellXGene database to see where they had come from. While many of the cells had been taken from kidneys, some had come from lungs or other organs. It was possible, the researchers speculated, that previously unknown Norn cells were scattered across the body.

Dr. Katalin Susztak, a physician-scientist at the University of Pennsylvania who studies Norn cells, said that the finding whetted her curiosity. “I want to check these cells,” she said.

She is skeptical that the model found true Norn cells outside the kidneys, since the erythropoietin hormone hasn’t been found in other places. But the new cells may sense oxygen as Norn cells do.

In other words, U.C.E. may have discovered a new type of cell before biologists did.

An ‘Internet of Cells’



Credit...Doug Chayka

Just [like ChatGPT](#), biological models sometimes get things wrong. Kasia Kedzierska, a computational biologist at the University of Oxford, and her colleagues recently gave GeneFormer and [another foundation model](#), scGPT, [a battery of tests](#). They presented the models with cell atlases they hadn't seen before and had them perform tasks such as classifying the cells into types. The models performed well on some tasks, but in other cases they fared poorly compared with simpler computer programs.

Dr. Kedzierska said she had great hopes for the models but that, for now, “they should not be used out of the box without a proper understanding of their limitations.”

Dr. Leskovec said that the models were improving as scientists trained them on more data. But compared with ChatGPT's training on the entire internet, the latest cell atlases offer only a modest amount of information. “I'd like an entire internet of cells,” he said.

More cells are on the way as bigger cell atlases come online. And scientists are gleaning different kinds of data from each of the cells in those atlases. Some scientists are cataloging the molecules that stick to genes, or taking photographs of cells to illuminate the precise location of their proteins. All of that information will allow foundation models to draw lessons about what makes cells work.

Scientists are also developing tools that let foundation models combine what they're learning on their own with what flesh-and-blood biologists have already discovered. The idea would be to connect the findings in thousands of published scientific papers to the databases of cell measurements.

With enough data and computing power, scientists say, they may eventually create a complete mathematical representation of a cell.

“That's going to be hugely revolutionary for the field of biology,” said Bo Wang, a computational biologist at the University of Toronto and the creator of scGPT. With this virtual cell, he speculated, it would be possible to predict what a real cell would do in any situation. Scientists could run entire experiments on their computers rather than in petri dishes.

Dr. Quake suspects that foundation models will learn not just about the kinds of cells that currently reside in our bodies but also about kinds of cells that *could* exist. He speculates that only certain combinations of biochemistry

can keep a cell alive. Dr. Quake dreams of using foundation models to make a map showing the realm of the possible, beyond which life cannot exist.

“I think these models are going to help us get some really fundamental understanding of the cell, which is going to provide some insight into what life really is,” Dr. Quake said.

Having a map of what’s possible and impossible to sustain life might also mean that scientists could actually create new cells that don’t yet exist in nature. The foundation model might be able to concoct chemical recipes that transform ordinary cells into new, extraordinary ones. Those new cells might devour plaque in blood vessels or explore a diseased organ to report back on its condition.

“It’s very ‘Fantastic Voyage’-ish,” Dr. Quake admitted. “But who knows what the future is going to hold?”

New Risks

If foundation models live up to Dr. Quake’s dreams, they will also raise a number of new risks. On Friday, more than 80 biologists and A.I. experts [signed a call](#) for the technology to be regulated so that it cannot be used to create new biological weapons. Such a concern might apply to new kinds of cells produced by the models.

Privacy breaches could happen even sooner. Researchers hope to program personalized foundation models that would look at an individual’s unique genome and the particular way that it works in cells. That new dimension of knowledge could reveal how different versions of genes affect the way cells work. But it could also give the owners of a foundation model some of the most intimate knowledge imaginable about the people who donated their DNA and cells to science.

Some scientists have their doubts about how far foundational models will make it down the road to “Fantastic Voyage,” however. The models are only as good as the data they are fed. Making an important new discovery about life may depend on having data on hand that we haven’t figured out how to collect. We might not even know what data the models need.

“They might make some new discoveries of interest,” said Sara Walker, a physicist at Arizona State University who studies the physical basis of life. “But ultimately they are limited when it comes to new fundamental advances.”

Still, the performance of foundation models has already led their creators to wonder about the role of human biologists in a world where computers make important insights on their own. Traditionally, biologists have been rewarded for creative and time-consuming experiments that uncover some of the workings of life. But computers may be able to see those workings in a matter of weeks, days or even hours by scanning billions of cells for patterns we can’t see.

“It’s going to force a complete rethink of what we consider creativity,” Dr. Quake said. “Professors should be very, very nervous.”

Carl Zimmer covers news about science for *The Times* and writes the [Origins column](#). [More about Carl Zimmer](#)

<https://www.nytimes.com/2024/03/10/science/ai-learning-biology.html>

Makes me think of the maxim “He knows the price of everything, and the value of nothing.” The cell is the form and model of everything living, metabolic and ideational. The human self, however nebulous and simplistic to consider it singular—though it can be if the unconscious/preconscious/can’t-be-conscious is included through adjusted conception—functions like a cell: membranes/ boundaries, energy sources, what is allowed in, what is expelled, etc. But to say that AI will figure out and “comprehend” life is to misrepresent and diminish what brains-minds-consciousness-experience-knowing constitute; even the nonhuman “personalities” and sensory/cognitive/ emotional “processes” of animals—and plants?—are more than “cells” metabolizing, however intricately connected and confounded the mentation and the metabolism, still distinctively different from atoms and molecules. (Whatever they actually are. What really are mitochondria and how and “where” do they all reside in the cell and ... beyond it?) Biological as life is, Life is more than “biology.”

TJB