outfitted with the standard medley of mugs, tables, chairs, and computers—as a Google conference room. "I can give a zillion reasons." But say you slowly dim the lights. "When the lights become very dim, only the biggest reasons stand out." Those transitions from a blank reference allow Sundararajan to capture more of the network's decisions than Ribeiro's variations do. But deeper, unanswered questions are always there, Sundararajan says—a state of mind familiar to him as a parent. "I have a 4-year-old who continually reminds me of the infinite regress of 'Why?'"

GUPTA HAS A DIFFERENT TACTIC for coping with black boxes: She avoids them. Several years ago Gupta, who moonlights as a designer of intricate physical puzzles, began a project called GlassBox. Her goal is to tame neural networks by engineering predictability into them. Her guiding principle is monotonicity—a relationship between variables in which, all else being equal, increasing one variable directly increases another, as with the square footage of a house and its price.

Gupta embeds those monotonic relationships in sprawling databases called interpolated lookup tables. In essence, they're like the tables in the back of a high school trigonometry textbook where you'd look up the sine of 0.5. But rather than dozens of entries across one dimension, her tables have millions across multiple dimensions. She wires those tables into neural networks, effectively adding an extra, predictable layer of computation—baked-in knowledge that she says will ultimately make the network more controllable.

Caruana, meanwhile, has kept his pneumonia lesson in mind. To develop a model that would match deep learning in accuracy but avoid its opacity, he turned to a community that hasn't always gotten along with machine learning and its loosey-goosey ways: statisticians.

In the 1980s, statisticians pioneered a technique called a generalized additive model (GAM). It built on linear regression, a way to find a linear trend in a set of data. But GAMs can also handle trickier relationships by finding multiple operations that together can massage data to fit on a regression line: squaring a set of numbers while taking the logarithm for another group of variables, for example. Caruana has supercharged the process, using machine learning to discover those operations—which can then be used as a powerful pattern-detecting model. "To our great surprise, on many problems, this is very accurate," he says. And crucially, each operation's influence on the underlying data is transparent.

Combing the genome for the roots of autism

For geneticists, autism is a vexing challenge. Inheritance patterns suggest it has a strong genetic component. But variants in scores of genes known to play some role in autism can explain only about 20% of all cases. Finding other variants that might contribute requires looking for clues in data on the 25,000 other human genes and their surrounding DNA—an overwhelming task for human investigators. So computational biologist Olga Troyanskaya of Princeton University and the Simons Foundation in New York City enlisted the tools of artificial intelligence (AI).

"We can only do so much as biologists to show what underlies diseases like autism," explains collaborator Robert Darnell, founding director of the New York Genome Center and a physician scientist at The Rockefeller University in New York City. "The power of machines to ask a trillion questions where a scientist can ask just 10 is a game-changer."

Troyanskaya combined hundreds of data sets on which genes are active in specific human cells, how proteins interact, and where transcription factor binding sites and other key genome features are located. Then her team used machine learning to build a map of gene interactions and compared those of the few well-established autism risk genes with those of thousands of other unknown genes, looking for similarities. That flagged another 2500 genes likely to be involved in autism, they reported last year in *Nature Neuroscience*.

But genes don't act in isolation, as geneticists have recently realized. Their behavior is shaped by the millions of nearby noncoding bases, which interact with DNA-binding proteins and other factors. Identifying which noncoding variants might affect nearby autism genes is an even tougher problem than finding the genes in the first place, and graduate student Jian Zhou in Troyanskaya's Princeton lab is deploying AI to solve it.

To train the program—a deep-learning system—Zhou exposed it to data collected by the Encyclopedia of DNA Elements and Roadmap Epigenomics, two projects that cataloged how tens of thousands of noncoding DNA sites affect neighboring genes. The system in effect learned which features to look for as it evaluates unknown stretches of noncoding DNA for potential activity.

When Zhou and Troyanskaya described their program, called DeepSEA, in *Nature Methods* in October 2015, Xiaohui Xie, a computer scientist at the University of California, Irvine, called it "a milestone in applying deep learning to genomics." Now, the Princeton team is running the genomes of autism patients through DeepSEA, hoping to rank the impacts of noncoding bases.

Xie is also applying AI to the genome, though with a broader focus than autism. He, too, hopes to classify any mutations by the odds they are harmful. But he cautions that in genomics, deep learning systems are only as good as the data sets on which they are trained. "Right now I think people are skeptical" that such systems can reliably parse the genome, he says. "But I think down the road more and more people will embrace deep learning."
—Elizabeth Pennisi

Artificial intelligence tools are helping reveal thousands of genes that may contribute to autism.

Science

AI in Action: Combing the genome for the roots of autism

Elizabeth Pennisi

ARTICLE TOOLS	http://science.sciencemag.org/content/357/6346/25
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/9/393/eaag2882.full http://stm.sciencemag.org/content/scitransmed/8/321/321fs1.full http://stm.sciencemag.org/content/scitransmed/7/271/271ra8.full http://stm.sciencemag.org/content/scitransmed/5/209/209ed18.full http://stm.sciencemag.org/content/scitransmed/4/155/155cm11.full
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service