THE WALL STREET JOURNAL.

https://www.wsj.com/articles/the-ultimate-learning-machines-11570806023

THE SATURDAY ESSAY

The Ultimate Learning Machines

The future of artificial intelligence depends on designing computers that can think and explore as resourcefully as babies do.

By Alison Gopnik
Oct. 11, 2019 11:00 am ET

Last July, I went to the Defense Advanced Research Projects Agency (DARPA), the blue-sky government research lab that helped to invent the computer and the internet. I was there, strange as it may seem, to talk about babies. The latest big DARPA research project, Machine Common Sense, is funding collaborations between child psychologists like me and computer scientists. This year I also talked about children's minds at Google, Facebook and Apple.

Why are quintessentially geeky places like DARPA and Google suddenly interested in talking about something as profoundly ungeeky as babies? It turns out that understanding babies and young children may be one key to ensuring that the current "AI spring" continues—despite some chilly autumnal winds in the air.

In the past, scientists unsuccessfully tried to create artificial intelligence by programming knowledge directly into a computer. Now they rely instead on "machine learning"—techniques that let the computers themselves work out what to do based on the data they see. These techniques have led to amazing breakthroughs. For example, you can give a machine learning system millions of animal pictures from the web, each one labeled as a cat or a dog. Without knowing anything else about animals, the system can extract the statistical patterns in the pictures and then use those patterns to recognize and classify new examples of cats and dogs.

With a machine learning system like Google Deep Mind's Alpha Zero, you can train a computer from scratch to play a videogame or even chess or Go. The computer gets a score, and after it plays many millions of games it can learn how to maximize that score, without explicitly being told about the strategies of chess or Go.

The problem is that these new algorithms are beginning to bump up against significant limitations. They need enormous amounts of data, only some kinds of data will do, and they're not very good at generalizing from that data. Babies seem to learn much more general and powerful kinds of knowledge than AIs do, from much less and much messier data. In fact, human babies are the best learners in the universe. How do they do it? And could we get an AI to do the same?

First, there's the issue of data. AIs need enormous amounts of it; they have to be trained on hundreds of millions of images or games. In fact, the big recent advances haven't primarily come about because of conceptual breakthroughs—the basic principles behind the machine learning algorithms were discovered back in the 1980s. The new development is that the internet now provides massive data sets for AIs to train with (everybody who has ever posted a LOLcats picture has contributed to the new AI), while Moore's law has led to equally massive increases in computational power.

Children, on the other hand, can learn new categories from just a small number of examples. A few storybook pictures can teach them not only about cats and dogs but jaguars and rhinos and unicorns.

A child participates in a study of head-mounted camera footage led by Dr. Jessica Sullivan at Skidmore College. **PHOTO:** ERICA WOJCIK, PH.D

The kind of data that children learn from is also very different from the data AI needs. The pictures that feed the AI algorithms have been curated by people, so they generally provide good examples and clear categories. (Nobody posts that messed-up smartphone shot where the cat ran halfway out of the picture.) Games like chess and Go provide curated data in another way, since people designed these games to have clearly defined rules and a restricted range of possibilities.

As psychologists have recently started to find out, the kind of data that children learn from is very different. Researchers like Linda Smith at Indiana University and Michael Frank at Stanford University have outfitted toddlers with super-light head-mounted cameras—a sort of baby Go Pro. This reveals that what babies see is very different from the millions of clear photographs in an internet data set. Instead, the cameras show a chaotic series of badly filmed videos of a few familiar things—balls and toys and parents and dogs—moving around at odd angles.

SHARE YOUR THOUGHTS

Do you think computers will ever be able learn as well as human babies do? Join the conversation below.

AIs also need what computer scientists call "supervision." In order to learn, they must be given a label for each image they "see" or a score for each move in a game. Baby data, by contrast, is largely unsupervised. Parents may occasionally tell a baby the name of the animal they're seeing or say "good job" when they perform a specific task. But parents are mostly just trying to keep their children alive and out of trouble. Most of a baby's learning is spontaneous and self-motivated.

Even with a lot of supervised data, AIs can't make the same kinds of generalizations that human children can. Their knowledge is much narrower and more limited, and they are easily fooled by what are called "adversarial examples." For instance, an AI image recognition system will confidently say that a mixed-up jumble of pixels is a dog if the jumble happens to fit the right statistical pattern—a mistake a baby would never make.

Current AIs are like children with super-helicopter-tiger moms—programs that hover over the learner dictating whether it is right or wrong at every step. Not unlike human children, those

helicoptered AI children can be very good at learning to do specific things well, but they fall apart when it comes to resilience and creativity. A small change in the learning problem means that they have to start all over again.

DARPA loves acronyms, so at UC Berkeley we're building a system that we call MESS (appropriately for babies), short for Model-Building, Exploratory, Social Learning System. These are the elements that are the secret of babies' success and that have largely been missing from current AIs.

A baby examines a ball as part of a Johns Hopkins study on learning from surprises. **PHOTO:** JOHNS HOPKINS UNIVERSITY

One of the secrets of children's learning is that they construct models or theories of the world. Toddlers may not learn how to play chess, but they develop common-sense ideas about physics and psychology. Psychologists like Elizaberth Spelke at Harvard have shown that even 1-year-old babies know a lot about objects: They are surprised if they see a toy car hover in midair or pass through a wall, even if they've never seen the car or the wall before. Babies know something about people, too. Felix Warneken at the University of Michigan has shown that if 1-year-olds see someone accidentally drop a pen on the floor and reach for it, they will pick up the pen and give it to them. But they won't do this if the person intentionally throws the pen to the floor.

The grand challenge of the new DARPA Machine Common Sense program is to design an AI that understands these basic features of the world as well as an 18-month-old. Some computer scientists are trying to build common sense models into the AIs, though this isn't easy. But it is even harder to design an AI that can actually learn those models the way that children do. Hybrid systems that combine models with machine learning are one of the most exciting developments at the cutting edge of current AI.

Another secret of children's learning is familiar to every parent—they are insatiably curious and active experimenters. Parents call this "getting into everything." AIs have mostly been stuck inside their mainframes, passively absorbing data. They haven't had much opportunity to get out there and gather the data themselves, or to select which data will teach them the most. Those chaotic baby-cam videos make more sense when you take the perspective of a baby who is exploring the world—picking things up and dropping them, putting them together and taking them apart.

> **AIs that are motivated by curiosity are more robust and resilient learners than those that are just motivated by immediate rewards.**

Recent studies also show just how intelligent this playful everyday experimentation can be. For example, Aimee Stahl and Lisa Feigenson from Johns Hopkins showed 1-year-olds toys that did surprising things, like the car that hovered in midair or seemed to go magically through a wall. The babies were surprised, like those in the earlier studies. But this time the researchers let the babies play with the cars. The babies displayed curiosity, playing more with the toys that did weird things than with those that behaved more predictably. But they also played differently—

dropping the gravity-defying car and banging the wall-dissolving one against the table. It's as if they were trying to figure out just why these objects were so weird.

In my lab at Berkeley, we're collaborating with computer scientists like Deepak Pathak and Pulkit Agrawal who try to make AIs that are similarly curious, active learners. Usually, machine learning systems reward the AI when they do something right, like bumping up their score in a game. But these AIs get a reward when they do something that leads to a surprising or unexpected result, and this makes them explore weird events, just like the babies. In fact, AIs that are motivated by curiosity are more robust and resilient learners than those that are just motivated by immediate rewards. This kind of active learning is another cutting-edge frontier in AI.

A final crucial factor that sets children apart from AIs is the way that they learn socially, from other people. Culture is our nature, and it makes our learning particularly powerful. Each new generation of children can take advantage of everything that earlier generations have discovered.

AIs can learn from very specific and controlled kinds of human supervision. But human children learn from the people around them in much more sophisticated ways. For example, our colleagues in computer vision at Berkeley, especially in Jitendra Malik's lab, are trying to design robots that can learn a new skill by imitating people. They show the robots a person accomplishing a goal and try to train it to replicate the results. Babies learn this way from the time they are 9 months old or so.

But imitation turns out to be very difficult to teach. Suppose you want to imitate how someone ties a knot—a really difficult task for robots but one that every sneaker-wearing child can master. Do you imitate all of the unnecessary details in the way they do it, replicating exactly the angle and speed of each step? Or do you figure out what the person is trying to do and do it as simply and efficiently as you can? Or do you add a refinement that will make the knot work even better?

MORE SATURDAY ESSAYS

- 'Cancer Is Still Beating Us—We Need a New Start October 4, 2019
- 'Can the New Afghanistan Survive America's Exit? September 27, 2019
- 'When the Slave Traders Were African September 20, 2019
- 'The Long History of American Slavery Reparations September 20, 2019

Studies in our lab and others show that children decide how to imitate intelligently, based on what they think the other person is trying to do and how the world works. So far, robots can sometimes learn to exactly replicate a particular action, but they can't imitate in the sophisticated way that children can.

There is another way that social life is a crucial part of babies' brilliance. Even very young babies already have a moral sense, rooted in their relationships with the people who care for them. Toddlers are already altruistic and empathetic and have basic ideas of fairness and compassion. For babies, learning and love, computation and care, are inextricably connected. Designing a truly intelligent AI, like raising a child, means instilling those ungeeky virtues. This might be a good direction for DARPA and Google too.

Is it possible for physical systems to solve all of these problems? In some sense, it must be, because those physical systems already exist: They're called babies. We even know how to make new ones, and it is a lot easier and more fun than programming.

But we are still very far from approaching that level of intelligence in machines. That's OK, because we don't really want AIs to replicate human intelligence; what we want is an AI that can help make us even smarter. To create more helpful machines, like curious AIs or imitative robots, the best way forward is to take our cues from babies.

Ms. Gopnik, a columnist for Review, is a professor of psychology at the University of California, Berkeley, and the author of "The Philosophical Baby" and "The Scientist in the Crib," among

other books.