ARTIFICIAL INTELLIGENCE

Good old-fashioned engineering can close the 100,000-year "data gap" in robotics

Ken Goldberg is the president of the Robot Learning Foundation, Mill Valley, CA, USA; chair of the Berkeley AI Research (BAIR) Lab Steering Committee, Berkeley, CA, USA; a cofounder of Ambi Robotics, Berkeley, CA, USA and Jacobi Robotics, Emeryville, CA, USA; and the William S. Floyd distinguished chair of engineering at UC Berkeley, Berkeley, CA, USA. Email: goldberg@berkeley.edu

Large vision-language models (VLMs) based on internet-scale data can now pass the Turing test for intelligence. In this sense, data have "solved" language, and many claim that data have solved speech recognition and computer vision.

Will data also solve robotics? Rich Sutton points out in "The Bitter Lesson" (*1*) that data and black-box "end-to-end" models have surpassed all the best-laid analytic work in artificial intelligence (AI). I accept that this trend will eventually produce general-purpose robots. But the question is… when?

Using commonly accepted metrics for converting word and image tokens into time, the amount of internet-scale data (texts and images) used to train contemporary VLMs is on the order of 100,000 years—it would take a human that long to read or view these data (*2*). However, the data needed to train robots are a combination of video inputs with robot motion commands: Those data do not exist on the internet.

One way to collect robot data is teleoperation, where human "trainers" use remotely controlled devices to painstakingly choose every motion of a robot as it performs a task, like folding a towel, over and over again. Many companies are gearing up with fleets of robots and humans to collect data this way.

However, the largest such dataset reported so far is on the order of 1 year of data (it was collected in less than 1 year by many human-robot systems). These data have been used to train large models, and initial results are intriguing. However, this suggests that at current data collection rates, a general-purpose robot, based on a ChatGPT-sized set of robot data, will be available in …100,000 years (*3*). So, how can we close this 100,000-year "data gap"?

Researchers are actively pursuing two additional methods for generating robot data: simulation and three-dimensional analysis of internet videos. Digital simulation today looks incredibly lifelike—consider the special effects in action movies and the deepfakes generated by AI. Simulation data work well for robots that fly or walk, or even for doing backflips, but simulation is notoriously unreliable for robot manipulation.

This sim2real "gap" arises because physical manipulation involves precise and changing contacts between the edges and surfaces of objects and grippers, very small but important material deformations, and very nuanced and changing frictional forces due to microscopic surface variations. Robots trained on simulation data can work well in simulation, but they often fail when manipulating physical objects.

The third potential source of robot data is videos on the internet. YouTube includes about 35,000 years of videos. However, accurately "lifting" a video image back into three dimensions to recover precise finger and object motions is a grand challenge for computer vision that is not expected to be solved in the foreseeable future.

There is a fourth option. Robot data can be collected from real robots working with real objects in real environments. Today, real commercial robots are performing specific tasks like driving or e-commerce package sorting.

One example is Waymo, which has robot taxis operating in several US cities. These taxis have "level 4" autonomy—they are largely autonomous, but whenever they need help, they can connect with human operators who log into the car remotely.

Another example is Ambi Robotics, which has package-sorting and -stacking machines operating in postal and warehouse facilities. Both Waymo and Ambi have created "data flywheels," where commercial robots collect data that are used to improve robot performance and to enable adjacent robot skills, like highway merging for Waymo and package stacking (very different from sorting) for Ambi.

Another feature that Waymo and Ambi have in common is that they do not use pure end-to-end AI methods. These companies combine advances in AI with rigorous engineering methods like Kalman filters, Gaussian processes, inverse kinematics, and motion planning. Let us use the term GOFE (good old-fashioned engineering) to describe rigorous engineering methods based on the math and physics developed over the past 200 years.

Whereas end-to-end AI methods are "model free," GOFE is "model based." GOFE segments problems into modules so that each module can be tested, fixed, and fine-tuned independently. Model-based methods can be combined with model-free methods to "kickstart" robots to achieve the levels of reliability required for adoption in real commercial environments, where they can then begin generating real robot data. Over the past 4 years, Ambi Robotics has collected 22 years of real robot data using a combination of model-based and model-free methods to sort more than 100 million real packages. It is now using those data to train the next generation of model-free systems (*4*).

As noted at the beginning, I do not disagree with Rich Sutton—I believe that model-free AI will eventually surpass GOFE to enable fully general-purpose robots in the future. I look forward to that future and hope I get to see it.

But when will the general-purpose robots arrive? I'm not sure that the public (or investors) are willing to wait very long. We can start closing the 100,000-year data gap by combining GOFE with model-free methods so that real robots can collect data as they perform useful work, such as driving taxis and sorting packages. Those data can be used to improve performance and enable robots to learn adjacent skills, spinning up real data flywheels until they collect enough data to enable general-purpose robots.

—Ken Goldberg

REFERENCES

1. R. Sutton, The bitter lesson (13 March 2019); https://cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf.
2. K. Black (@kvablack), "My favorite slide that I made for my talk last weekend—A very silly thought experiment in which we compare language datasets to robotics datasets (in the most shallow way possible). Yes it is to scale; I learned that the maximum shape size in Keynote is 20,000pts," X, 12 November 2024; https://x.com/kvablack/status/1856373781603987655.
3. K. Goldberg, J. Zhang, Comparing the size of robot data to LLM data (13 March 2024); https://drive.google.com/file/d/16DzKxYvRutTN7GBflRZj57WgsFNHr29t/view.
4. V. Satish, J. Mahler, K. Goldberg, PRIME-1: Scaling large robot data for industrial reliability (30 January 2025); https://ambirobotics.com/blog/prime-1-scaling-large-robot-data-for-industrial-reliability/.

10.1126/scirobotics.aea7390