COMPUTER SCIENCE

# Core progress in AI has stalled in some fields

## When tuned up, old algorithms can match the abilities of their successors

*By* **Matthew Hutson**

Artificial intelligence (AI) just seems to get smarter and smarter. Each iPhone learns your face, voice, and habits better than the last, and the threats AI poses to privacy and jobs continue to grow. The surge reflects faster chips, more data, and better algorithms. But some of the improvement comes from tweaks rather than the core innovations their inventors claim—and some of the gains may not exist at all, says Davis Blalock, a computer science graduate student at the Massachusetts Institute of Technology (MIT). Blalock and his colleagues compared dozens of approaches to improving neural networks—software architectures that loosely mimic the brain. "Fifty papers in," he says, "it became clear that it wasn't obvious what the state of the art even was."

The researchers evaluated 81 pruning algorithms, programs that make neural networks more efficient by trimming unneeded connections. All claimed superiority in slightly different ways. But they were rarely compared properly—and when the researchers tried to evaluate them side by side, there was no clear evidence of performance improvements over a 10-year period. The result, presented in March at the Machine Learning and Systems conference, surprised Blalock's Ph.D. adviser, MIT computer scientist John Guttag, who says the uneven comparisons themselves may explain the stagnation. "It's the old saw, right?" Guttag said. "If you can't measure something, it's hard to make it better."

Researchers are waking up to the signs of shaky progress across many subfields of AI. A 2019 meta-analysis of information retrieval algorithms used in search engines concluded the "high-water mark … was actually set in 2009." Another study in 2019 reproduced seven neural network recommendation systems, of the kind used by media streaming services. It found that six failed to outperform much simpler, nonneural algorithms developed years before, when the earlier techniques were fine-tuned, revealing "phantom progress" in the field. In another paper posted on arXiv in March, Kevin Musgrave, a computer scientist at Cornell University, took a look at loss functions, the part of an algorithm that mathematically specifies its objective. Musgrave compared a dozen of them on equal footing, in a task involving image retrieval, and found that, contrary to their developers' claims, accuracy had not improved since 2006 (see chart, below). "There's always been these waves of hype," Musgrave says.

Gains in machine-learning algorithms can come from fundamental changes in their architecture, loss function, or optimization strategy—how they use feedback to improve. But subtle tweaks to any of these can also boost performance, says Zico Kolter, a computer scientist at Carnegie Mellon University who studies image-recognition models trained to be immune to "adversarial attacks" by a hacker. An early adversarial training method known as projected gradient descent (PGD), in which a model is simply trained on both real and deceptive examples, seemed to have been surpassed by more complex methods. But in a February arXiv paper, Kolter and his colleagues found that all of the methods performed about the same when a simple trick was used to enhance them.

"That was very surprising, that this hadn't been discovered before," says Leslie Rice, Kolter's Ph.D. student. Kolter says his findings suggest innovations such as PGD are hard to come by, and are rarely improved in a substantial way. "It's pretty clear that PGD is actually just the right algorithm," he says. "It's the obvious thing, and people want to find overly complex solutions."

Other major algorithmic advances also seem to have stood the test of time. A big breakthrough came in 1997 with an architecture called long short-term memory (LSTM), used in language translation. When properly trained, LSTMs matched the performance of supposedly more advanced architectures developed 2 decades later. Another machine-learning breakthrough came in 2014 with generative adversarial networks (GANs), which pair networks in a create-and-critique cycle to sharpen their ability to produce images, for example. A 2018 paper reported that with enough computation, the original GAN method matches the abilities of methods from later years.

Kolter says researchers are more motivated to produce a new algorithm and tweak it until it's state-of-the-art than to tune an existing one. The latter can appear less novel, he notes, making it "much harder to get a paper from."

Guttag says there's also a disincentive for inventors of an algorithm to thoroughly compare its performance with others—only to find that their breakthrough is not what they thought it was. "There's a risk to comparing too carefully." It's also hard work: AI researchers use different data sets, tuning methods, performance metrics, and baselines. "It's just not really feasible to do all the apples-to-apples comparisons."

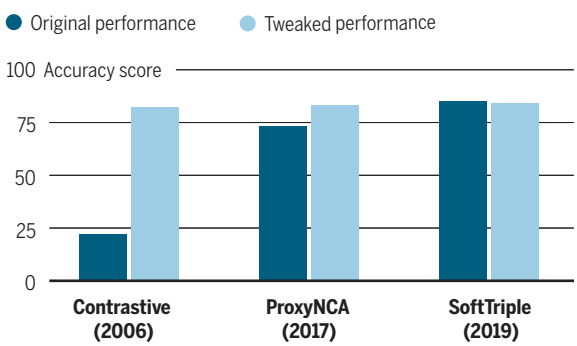
Some of the overstated performance claims can be chalked up to the explosive growth of the field, where papers outnumber experienced reviewers. "A lot of this seems to be growing pains," Blalock says. He urges reviewers to insist on better comparisons to benchmarks and says better tools will help. Earlier this year, Blalock's co-author, MIT researcher Jose Gonzalez Ortiz, released software called ShrinkBench that makes it easier to compare pruning algorithms.

Researchers point out that even if new methods aren't fundamentally better than old ones, the tweaks they implement can be applied to their forebears. And every once in a while, a new algorithm will be an actual breakthrough. "It's almost like a venture capital portfolio," Blalock says, "where some of the businesses are not really working, but some are working spectacularly well." ■

Matthew Hutson is a journalist in New York City.

### Old dogs, new tricks

After modest tweaks, old image-retrieval algorithms perform as well as new ones, suggesting little actual innovation.




CREDITS: (GRAPHIC) X. LIU/*SCIENCE*; (DATA) MUSGRAVE *ET AL.*, ARXIV: 2003.08505

# Core progress in AI has stalled in some fields

Matthew Hutson

*Science* **368** (6494), 927.
DOI: 10.1126/science.368.6494.927