

# *Big Bets on A.I. Open a New Frontier for Chip Start-Ups, Too*

By Cade Metz

Jan. 14, 2018

SAN FRANCISCO — For years, tech industry financiers showed little interest in start-up companies that made computer chips.

How on earth could a start-up compete with a goliath like Intel, which made the chips that ran more than 80 percent of the world's personal computers? Even in the areas where Intel didn't dominate, like smartphones and gaming devices, there were companies like Qualcomm and Nvidia that could squash an upstart.

But then came the tech industry's latest big thing — artificial intelligence. A.I., it turned out, works better with new kinds of computer chips. Suddenly, venture capitalists forgot all those forbidding roadblocks to success for a young chip company.

Today, at least 45 start-ups are working on chips that can power tasks like speech and self-driving cars, and at least five of them have raised more than \$100 million from investors. Venture capitalists invested more than \$1.5 billion in chip start-ups last year, nearly doubling the investments made two years ago, according to the research firm CB Insights.

The explosion is akin to the sudden proliferation of PC and hard-drive makers in the 1980s. While these are small companies, and not all will survive, they have the power to fuel a period of rapid technological change.

It is doubtful that any of the companies fantasize about challenging Intel head-on with their own chip factories, which can take billions of dollars to build. (The start-ups contract with other companies to make their chips.) But in designing chips that can

provide the particular kind of computing power needed by machines learning how to do more and more things, these start-ups are racing toward one of two goals: Find a profitable niche or get acquired. Fast.

“Machine learning and A.I. has reopened questions around how to build computers,” said Bill Coughran, who helped oversee the global infrastructure at Google for several years and is now a partner at Sequoia, the Silicon Valley venture capital firm. Sequoia has invested in Graphcore, a British start-up that recently joined the \$100 million club.

By the summer of 2016, the change was apparent. Google, Microsoft and other internet giants were building apps that could instantly identify faces in photos and recognize commands spoken into smartphones by using algorithms, known as neural networks, that can learn tasks by identifying patterns in large amounts of data.

Nvidia was best known for making graphics processing units, or G.P.U.s, which were designed to help render complex images for games and other software — and it turned out they worked really well for neural networks, too. Nvidia sold \$143 million in chips for the massive computer data centers run by companies like Google in the year leading up to that summer — double the year before.

Intel scrambled to catch up. It acquired Nervana, a 50-employee Silicon Valley start-up that had started building an A.I. chip from scratch, for \$400 million, according to a report from the tech news site Recode.

After that, a second Silicon Valley start-up, Cerebras, grabbed five Nervana engineers as it, too, designed a chip just for A.I.

By early 2018, according to a report by Forbes, Cerebras had raised more than \$100 million in funding. So had four other firms: Graphcore; another Silicon Valley outfit, Wave Computing; and two Beijing companies, Horizon Robotics and Cambricon, which is backed by the Chinese government.

Raising money in 2015 and early 2016 was a nightmare, said Mike Henry, chief executive at the A.I. chip start-up Mythic. But “with the big, acquisition-hungry tech companies all barreling toward semiconductors,” that has changed, he said.

China has shown a particular interest in developing new A.I. chips. A third Beijing chip start-up, DeePhi, has raised \$40 million, and the country’s Ministry of Science and Technology has explicitly called for the production of Chinese chips that challenge

Nvidia's.

Because it's a new market — and because there is such hunger for this new kind of processing power — many believe this is one of those rare opportunities when start-ups have a chance against entrenched giants.

The first big change will most likely come in the data center, where companies like Graphcore and Cerebras, which has been quiet about its plans, hope to accelerate the creation of new forms of A.I. Among the goals are bots that can carry on conversations and systems that can automatically generate video and virtual reality.

Researchers at places like Microsoft and Google, which has built its own chip just for A.I., “train” neural networks by extreme trial and error, testing the algorithms across vast numbers of chips for hours and even days on end. They often sit at their laptops, staring at graphs that show the progress of these algorithms as they learn from data. Chip designers want to streamline this process, packing all that trial and error into a few minutes.

Today, Nvidia's G.P.U.s can efficiently execute all the tiny calculations that go into training neural networks, but shuttling data between these chips is still inefficient, said Scott Gray, who was an engineer at Nervana before joining OpenAI, an artificial intelligence lab whose founder include Tesla's chief executive, Elon Musk.

So in addition to building chips specifically for neural networks, start-ups are rethinking the hardware that surrounds them.

Graphcore, for example, is building chips that include more built-in memory so that they don't need to send as much data back and forth. Others are looking at ways of widening the pipes between chips so that data exchange happens faster.

“This is not just about building chips but looking at how these chips are connected together and how they talk to the rest of the system,” Mr. Coughran, of Sequoia, said.

But this is only part of the change. Once neural networks are trained for a task, additional gear has to execute that task. At Toyota, autonomous car prototypes are using neural networks as a way of identifying pedestrians, signs and other objects on the road. After training a neural network in the data center, the company runs this algorithm on chips installed on the car.

A number of chip makers — including start-ups like Mythic, DeePhi and Horizon Robotics — are tackling this problem as well, pushing A.I. chips into devices ranging from phones to cars.

It is still unclear how well any of these new chips will work. Designing and building a chip takes about 24 months, which means even the first viable hardware relying on them won't arrive until this year. And the chip start-ups will face competition from Nvidia, Intel, Google and other industry giants.

But everyone is starting from about the same place: the beginning of a new market.

Follow Cade Metz on Twitter: [@CadeMetz](https://twitter.com/CadeMetz).

A version of this article appears in print on , Section B, Page 1 of the New York edition with the headline: Bets on A.I. Open a New Chip Frontier