

Artificial intelligence

Facebook joins Amazon and Google in AI chip race

Company needs faster hardware for smart digital assistant and real-time video moderation



Facebook wants to use AI to help its army of human moderators decide what content should be allowed on the service © Bloomberg

Richard Waters in San Francisco FEBRUARY 18, 2019

Facebook is racing Amazon and Google to develop its own artificial intelligence chips, after recognising that it needs dramatically faster computing to deliver the next breakthrough in AI.

The company's goals include a digital assistant imbued with enough "common sense" to converse with a person on any subject, a significant step up from today's voice-controlled devices, said Yann LeCun, the company's chief AI scientist and one of the pioneers of modern AI.

It also wants to make AI a more practical tool in controlling its social network, such as monitoring video in real time and helping its army of human moderators decide what content should be allowed on the service.

In an interview with the FT, Mr LeCun said that Facebook hoped to work with a number of chip companies on new designs — it recently announced a project with Intel — but he also indicated that it was developing its own custom "ASIC" chips to support its AI programs.

"Facebook has been known to build its hardware when required — build its own ASIC, for instance. If there's any stone unturned, we're going to work on it," he said, in Facebook's first official comments confirming the scope of its chip ambitions.

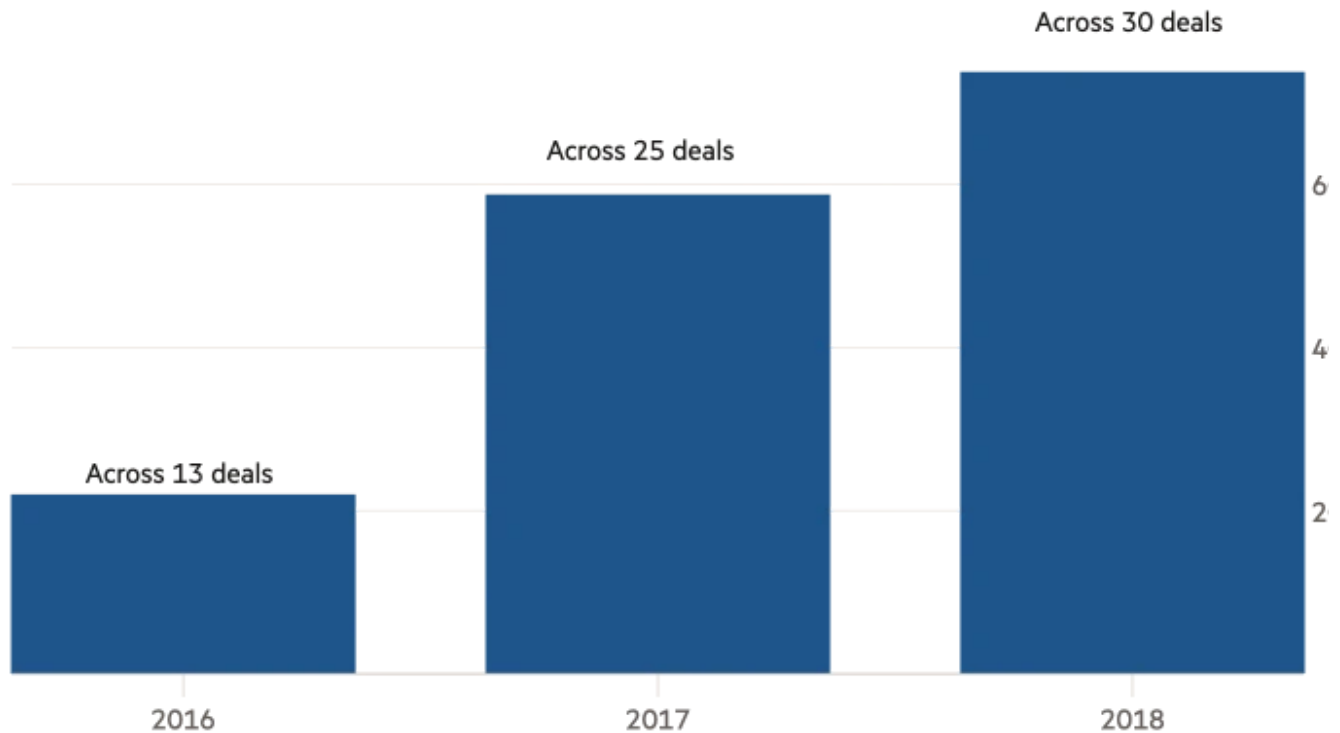
Referring to the chance for the company to make its own breakthroughs in chips, which comprise the foundation of computing systems, he added: "There's certainly a lot of room at the bottom."

Facebook's decision to create its own chips represents another long-term challenge to [Nvidia](#), the main producer of the graphics processors currently used for AI in data centres. Nvidia is facing a short-term squeeze from a pullback by large data centre customers.

The need for more specialised AI chips, designed to perform single tasks with lightning speed and with lower power consumption, rather than the general-purpose processors of the past, has seen a wave of investment not only by the likes of Google, Amazon and Apple, but also by dozens of start-ups.

Global equity funding figures for AI chip startups

\$m



Source: CB Insights

© FT

The focus on new silicon designs and hardware architectures points to a need for fundamental breakthroughs in basic computing to prevent today's AI from becoming a dead end.

Mr LeCun said that throughout the history of AI, it had often taken big improvements in hardware before researchers had come up with the insights that led to breakthroughs in the field.

“For a fairly long time, people didn't think about fairly obvious ideas,” he said, holding back AI development. These include back propagation, a core technique in today's deep learning systems which sees algorithms go back over their calculations to minimise errors. This was an obvious extension of earlier research, said Mr LeCun, but only became widely used in the 1990s after computing hardware had advanced.

Facebook has designed other types of hardware in the past, for instance coming up with new ideas for data centre equipment, before open-sourcing them for others to work on. The same approach would be applied to chip designs, Mr LeCun said, adding: “The objective is to give it away.”

The company is also focusing its research on new designs for neural networks, which are at the heart of the deep learning systems behind recent advances in things like image and language recognition.

Thirty years ago, while working at AT&T's Bell Labs on AI chips, Mr LeCun built the first “convolutional” neural network — a design based on how the visual cortex works in animals, and which is now common in deep learning systems.



Yann LeCun, Facebook's chief AI scientist

Today's neural networks, using a technique called supervised learning, require masses of data to train, while also consuming vast amounts of power when operating at the scale of a company like Facebook.

The company already carries out a number of instant analyses on all of the 2-3bn photos a day that are uploaded to its core service, including using facial recognition to identify people in them, creating captions to describe the scenes, and identifying things like nudity that are not allowed on the service.

Mr LeCun said that Facebook was working on “anything we can do to lower the power consumption [and] improve the latency” to speed up processing. But the huge demands posed by monitoring video flowing over the site in real time would require new neural network designs, he added.

Facebook is also looking to new neural network architectures to mimic more aspects of human intelligence and make its systems more natural to interact with.

Mr LeCun said the company was betting heavily on “self-supervised” systems, which are able to make more extensive predictions about the world around them rather than only being able to reach conclusions directly tied to the data they have been trained on.

That could enable them to develop the same kind of broad understanding of the world that enables humans to deal with new situations.

One thing Facebook would be interested in is offering smart digital assistants —

something that has a level of common sense. They have background knowledge and you can have a discussion with them on any topic

Yann LeCun

“In terms of new uses, one thing Facebook would be interested in is offering smart digital assistants — something that has a level of common sense,” he said. “They have background knowledge and you can have a discussion with them on any topic.”

The idea of [imbuing computers with common sense](#) is in its very early stages, and Mr LeCun said this deeper level of intelligence “won’t happen tomorrow”.

“You want machines like human beings or animals to understand what will happen when the world responds to your interactions with it. There’s a lot of work we’re doing in causality,” he said. “Being able to predict under uncertainty is one of the main challenges today.”

Facebook is part of a broader research effort to try to enhance today’s neural networks. Mr LeCun was due to outline the work at a speech at a [chip conference](#) in San Francisco on Monday.

These include networks that adapt their design in response to the data that pass through them, making them more flexible in the face of changes in the real world. Another avenue of research is into networks that only “fire” the neurons needed to solve a particular problem, an approach that echoes the way the human brain functions and could greatly reduce power consumption.

The research work also includes adding computer memory to neural networks, so that they can retain more information and develop a stronger sense of context when having a “conversation” with a person.

Advances in how neural networks function are likely to have knock-on effects on the design of the chips that power them, potentially opening up more competition to the companies that make today’s leading AI chips.

Google’s TPUs — which have emerged as the most powerful data centre chips for machine learning — are “still fairly generic”, Mr LeCun said. “They make assumptions that are not necessarily true for future neural network architectures.”

Flexibility in silicon design, on the other hand, can have other drawbacks. Microsoft, for instance, has plans to plant a different type of chip, known as a field programmable gate array, in all its data centre servers. These are more flexible in how they can be used but are less efficient in processing very large amounts of data, putting them at a disadvantage to chips that have been optimised for specialised tasks.

**FT/Oliver Wyman Executive
Dinner 2019**

Barcelona

Barcelona

25 February 2019

Trust: Telcos' Secret Weapon To Winning The Customer Of The Future



[Register now](#)

Presented by

[Copyright](#) The Financial Times Limited 2019. All rights reserved.

Topic Tracker : Artificial intelligence

Hide

Experimental Feature

0

1000

Topic Tracker measures how much of a topic you have covered and suggests the best articles to keep you one step ahead.

[Give feedback](#)

Follow the topics in this article

Technology sector

Artificial intelligence

Facebook Inc

Richard Waters

How easy or hard was it to use FT.com today?

[Leave feedback](#)

