From:	Satya Nadella </O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=MICROSOFT.ONMICROSOFT.COM-55760-SATYA NADELLA>
To:	Kevin Scott
CC:	Amy Hood
Sent:	6/12/2019 6:02:47 PM
Subject:	RE: Thoughts on OpenAI

DX0680

Very good email that explains, why I want us to do this… and also why we will then ensure our infra folks execute.

Amy - fyi

Sent from Mail for Windows 10

_____

From: Kevin Scott
Sent: Wednesday, June 12, 2019 7:16:11 AM
To: Satya Nadella; Bill Gates
Subject: Re: Thoughts on OpenAI

REDACTED

  The thing that's interesting about what Open AI and Deep Mind and Google Brain are doing is the scale of their ambition, and how that ambition is driving everything from datacenter design to compute silicon to networks and distributed systems architectures to numerical optimizers, compilers, programming frameworks, and the high level abstractions that model developers have at their disposal. When all these programs were doing was competing with one another to see which RL system could achieve the most impressive game-playing stunt, I has highly dismissive of their efforts. That was a mistake. When they took all of the infrastructure that they had built to build NLP models that we couldn't easily replicate, I started to take things more seriously. And as I dug in to try to understand where all of the capability gaps were between Google and us for model training, I got very, very worried.

  Turns out, just replicating BERT-large wasn't easy to do for us. Even though we had the template for the model, it took us ~6 months to get the model trained because our infrastructure wasn't up to the task. Google had BERT for at least six months prior to that, so in the time that it took us to hack together the capability to train a 340M parameter model, they had a year to figure out how to get it into production and to move on to larger scale, more interesting models. We are already seeing the results of that work in our competitive analysis of their products. One of the Q&A competitive metrics that we watch just jumped by 10 percentage points on Google Search because of BERT-like models. Their auto-complete in Gmail, which is especially useful in the mobile app, is getting scarily good.

REDACTED

REDACTED We have very smart ML people in Bing, in the vision team, and in the speech team. But the core deep learning teams within each of these bigger teams are very small, and their ambitions have also been constrained, which means that even as we start to feed them resources, they still have to go through a learning process to scale up. And we are multiple years behind the competition in terms of ML scale.

REDACTED

DX0680.003

REDACTED