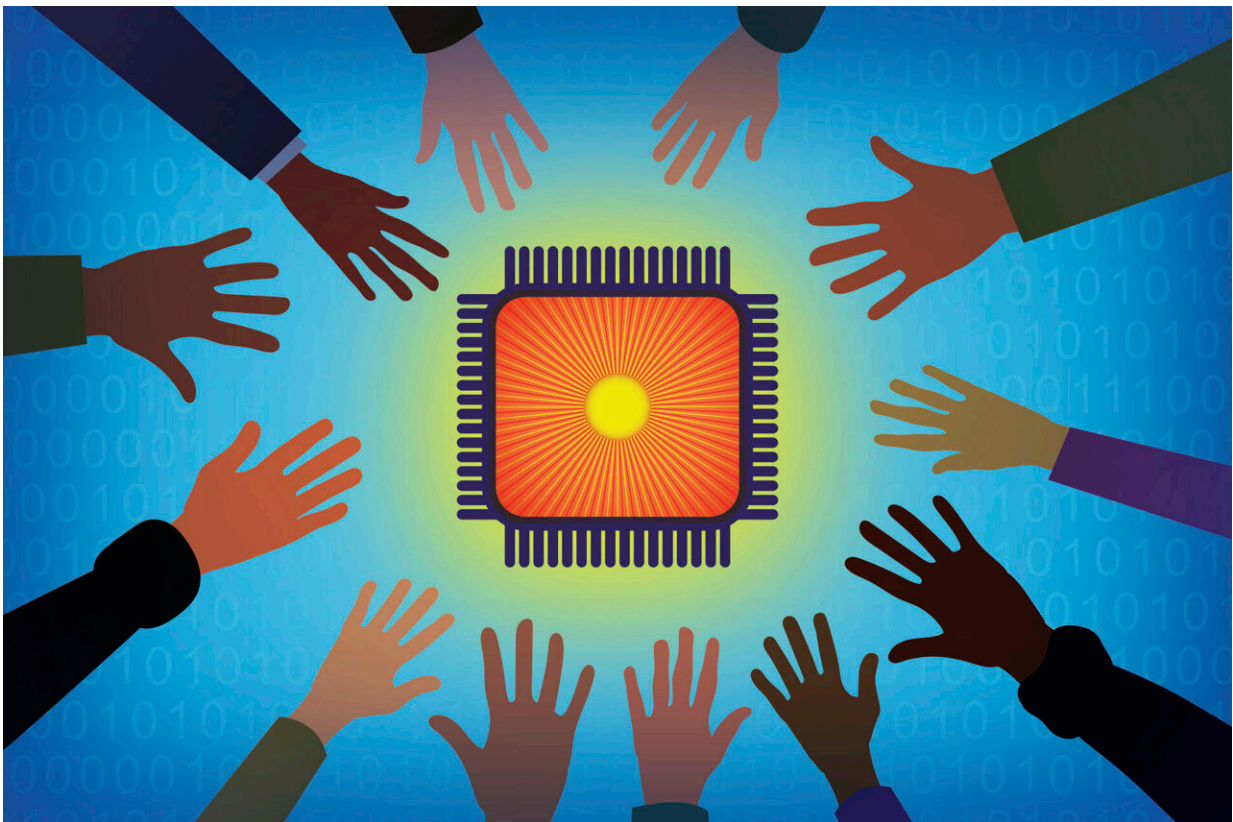# Tired of AI doomsday tropes, Cohere CEO says his goal is technology that's 'additive to humanity'

March 25 2024, by Matt O'brien

Credit: AP Illustration/Peter Hamlin

Aidan Gomez can take some credit for the 'T' at the end of ChatGPT. He was part of a group of Google engineers who first introduced a new

artificial intelligence model called a transformer.

That helped set a foundation for today's generative AI boom that ChatGPT-maker OpenAI and others built upon. Gomez, one of eight co-authors of Google's 2017 paper, was a 20-year-old intern at the time.

He's now the CEO and co-founder of Cohere, a Toronto-based startup competing with other leading AI companies in supplying large language models and the chatbots they power to big businesses and organizations.

Gomez spoke about the future of generative AI with The Associated Press. The interview has been edited for length and clarity.

Q: What's a transformer?

A: A transformer is an architecture of a neural network—the structure to the computation that happens inside of the model. The reason that transformers are special relative to their peers—other competing architectures, other ways of structuring neural networks—is essentially that they scale very well. They can be trained across not just thousands, but tens of thousands of chips. They can be trained extremely quickly. They use many different operations that these GPUs (graphics chips) are tailored for. Compared to what existed before the transformer, they do that processing faster and more efficiently.

Q: How important are they to what you're doing at Cohere?

A: Massively important. We use the transformer architecture as does everyone else in building large language models. For Cohere, a huge focus is scalability and production readiness for enterprises. Some of the other models that we compete against are huge and super inefficient. You can't actually put that into production, because as soon as you're faced with real users, costs blow up and the economics break.

Q: What's a specific example of how a customer is using a Cohere model?

A: I have a favorite example in the health care space. It stems from the surprising fact that 40% of a doctor's working day is spent writing patient notes. So what if we could have doctors attach a little passive listening device to follow along with them throughout the day, between their patient visits, listening into the conversation and pre-populating those notes so that instead of having to write it from scratch, there's a first draft in there. They can read through it and just make edits. Suddenly, the capacity of doctors boosts by a massive proportion.

Q: How do you address customer concerns about AI language models being prone to 'hallucinations' (errors) and bias?

A: Customers are always concerned about hallucinations and bias. It leads to a bad product experience. So it's something we focus on heavily. For hallucinations, we have a core focus on RAG, which is retrieval-augmented generation. We just released a new model called Command R which is targeted explicitly at RAG. It lets you connect the model to private sources of trusted knowledge. That might be your organization's internal documents or a specific employee's emails. You're giving the model access to information that it just otherwise hasn't seen out in the web when it was learning. What's important is that it also allows you to fact check the model, because now instead of just text in, text out, the model is actually making reference to documents. It can cite back to where it got that information. You can check its work and gain a lot more confidence working with the tool. It reduces hallucination massively.

Q: What are the biggest public misconceptions about generative AI?

A: The fear that certain individuals and organizations espouse about this

technology being a terminator, an existential risk. Those are stories humanity has been telling itself for decades. Technology coming and taking over and displacing us, rendering us subservient. They're very deeply embedded in the public's cultural brain stem. It's a very salient narrative. It's easier to capture people's imagination and fear when you tell them that. So we pay a lot of attention to it because it's so gripping as a story. But the reality is I think this technology is going to be profoundly good. A lot of the arguments for how it might go bad, those of us developing the technology are very aware of and working to mitigate those risks. We all want this to go well. We all want the technology to be additive to humanity, not a threat to it.

Q: Not only OpenAI but a number of major technology companies are now explicitly saying they're trying to build artificial general intelligence (a term for broadly better-than-human AI). Is AGI part of your mission?

A: No, I don't see it as part of my mission. For me, AGI isn't the end goal. The end goal is profound positive impact for the world with this technology. It's a very general technology. It's reasoning, it's intelligence. So it applies all over the place. And we want to make sure it's the most effective form of the technology it possibly can be, as early as it possibly can be. It's not some pseudo-religious pursuit of AGI, which we don't even really know the definition of.

Q: What's coming next?

A: I think everyone should keep their eyes on tool use and more agent-like behavior. Models that you can present them for the first time with a tool you've built. Maybe it's a software program or an API (application programming interface). And you can say, 'Hey model, I just built this. Here's what it does. Here's how you interact with it. This is part of your toolkit of stuff you can do.' That general principle of being able to give a model a tool it's never seen before and it can adopt it effectively, I think

is going to be very powerful. In order to do a lot of stuff, you need access to external tools. The current status quo is models can just write (text) characters back at you. If you give them access to tools, they can actually take action out in the real world on your behalf.

Citation: Tired of AI doomsday tropes, Cohere CEO says his goal is technology that's 'additive to humanity' (2024, March 25) retrieved 9 May 2024 from https://techxplore.com/news/2024-03-ai-doomsday-tropes-cohere-ceo.html