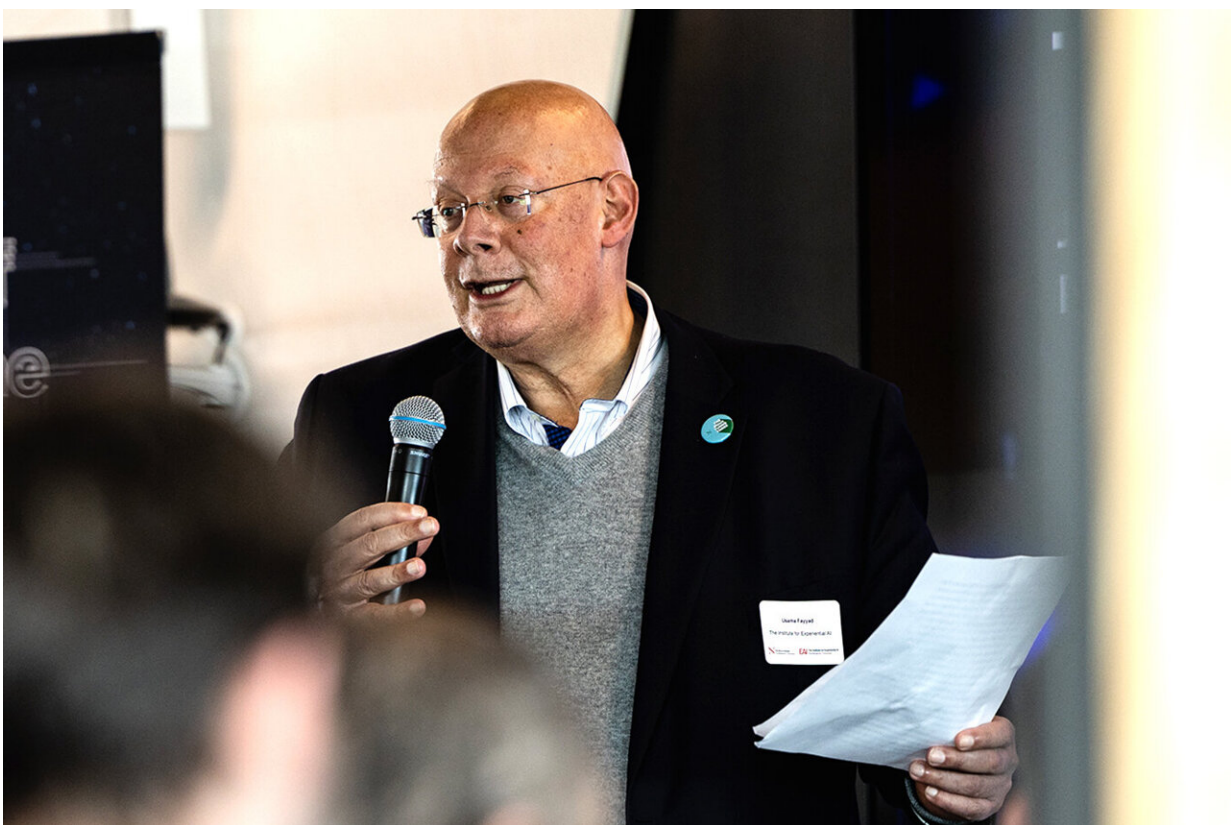


Is 'generative' AI the way of the future? Expert explains new models, the need for human involvement

April 20 2023, by Tanner Stening



Credit: Andy Gagne for Northeastern University

Ever since the 20th century's earliest theories of artificial intelligence set the world on an apparently irreversible track toward the technology, the

great promise of AI—one that's been used to justify that march forward—is that it can help usher in social transformation and lead to human betterment.

With the arrival of so-called generative AI, such as OpenAI's endlessly amusing and problem-riddled ChatGPT, the decades long slow-roll of AI advancement has felt more like a quantum leap forward. That perceptive jump has some experts worried about the consequences of moving too quickly toward a world in which machine intelligence they say could become an all-powerful, humanity-destroying force à la The Terminator.

But Northeastern experts, including Usama Fayyad, executive director for the Institute for Experiential Artificial Intelligence, maintain that those concerns don't reflect reality. That, in fact, AI is being integrated in ways that promote and necessitate human involvement—what experts have coined "human-in-the-loop."

On Tuesday, April 25, Northeastern will host a symposium of AI experts to discuss a range of topics related to the pace of AI development, and how progress is reshaping the workplace, education, health care and many other sectors. Northeastern Global News sat down with Fayyad to learn more about what next week's conference will take up; the upside of generative AI; as well as broader developments in the space. The conversation has been edited for brevity and clarity.

First things first, there's been a lot of talk about generative AI. What is it, and how is it different from other forms of AI?

Generative AI refers to the kind of AI that can, quite simply, generate outputs. Those outputs could be in the form of text like you see in what

we call the large [language models](#), such as ChatGPT (a chatbot on top of a large language model), or images, etc. If you are training [the AI] on text, text is what you will get out of it. If you are training it on images, you get images, or modifications of images, out of it. If you are training it on sounds or music, you get music out of it. If you train it on programming code, you get programs out, and so on.

It's also called generative AI because the algorithms have the ability to generate examples on their own. It's part of their training. Researchers would do things like have the algorithm challenge itself through generative adversarial networks, or algorithms that generate adversarial examples that could confuse the system to help strengthen its training. But since their development, researchers quickly realized that they needed human intervention. So most of these systems, including ChaptGPT, actually use and require human intervention. Human beings facilitate a lot of these challenges as part of the training through something called reinforcement learning, a machine learning technique designed to basically improve the system's performance.

ChatGPT has received a lot of press recently, particularly on social media where users have been posting their individual interactions with the AI. What are some of the practical applications of this new technology?

We are seeing it applied in education—in [higher education](#) in particular. Higher education has taken note—including Northeastern, in a very big way—of the fact that these technologies have challenged the way we conduct, for example, standardized testing. Educators have realized that this is just another tool. At Northeastern we have many examples that we will cover in this upcoming workshop of people using it in the classroom. Be it in the College of Arts, Media and Design for things like

[Salvador] Dalí and LensaAI for images; or be it in writing classes, English classes, or in engineering.

Just like we transitioned from the slide ruler to the calculator, to the computer and then the whole web on your mobile phone—this is another tool, and the proper way to train our students to be ready for the new world is to figure out ways to utilize this technology as a tool.

It's too early to see real world applications at large scale. The technology is too new. But there are estimates that anywhere from 50%–80%—I'm more in the 80% camp—of the tasks done by a knowledge worker can be accelerated by this technology. Not automated, accelerated. If you're a lawyer and drafting an agreement, you can have a first draft customized very quickly; but then you have to go in and edit or make changes. If you're a programmer, you can turn out an initial program. But it typically won't work well; it will have errors; it's not customized to the target. Again, a human being, provided they understand what they're doing, can go in and modify it, and save themselves 50%–80% of the effort.

Can you say more about what you mean by acceleration versus automation?

It's acceleration, not automation because we know the technology can hallucinate in horrible ways, in fact. It can make up stuff; it can try to defend points of view that you ask it to defend; you can make it lie, and you can lie to it and have it believe you.

They call this specific class of technology stochastic parrots, meaning parrots that have—let's say—random variation. And I like the term parrots because it correctly describes the fact that they don't understand what they're saying. So they say stuff, and the stuff may sound eloquent, or fluid. That's one of the big points that we try to make: somehow we

have learned in society to associate intelligence with eloquence and fluidity—basically someone who says things nicely. But in reality these algorithms are far from intelligent; they are basically doing autocomplete; they are repeating things they've seen before, and often they are repeated incorrectly.

Why do I say all of this? Because it means you have a human-in-the-loop needed in doing this work, because you need to check all of this work. You remove a lot of the repetitive monotonous work—that's great. You can accelerate it—that's productive. You now can spend your time adding value instead of repeating the boring tasks. All of that I consider positive.

I like to use accounting as a good analogy. What did accounting look like 60–70 years ago? Well, you had to deal with these big ledgers; you had to have nice handwriting; you had to have good addition skills in your head; you had to manually verify numbers and go over sums and apply ratios. Guess what? None of those tasks—none, zero—are relevant today. Now, have we replaced accountants because we've now replaced everything they used to do with something that is faster, better, cheaper, repeatable? No. We actually have more accountants today than in the history of humanity.

What topics will this upcoming workshop cover as it relates to AI?

What we're doing with this workshop is we're trying to cover the three areas that matter. What is the impact of ChatGPT and generative AI in the classroom, and how should we use it? We bring in folks who are doing this work at Northeastern to provide examples in one panel.

Second, how is the nature of work changing because of these technologies? That will be addressed during another panel where we

think about different business applications. We will use the law and health care as the two running examples here.

The third panel is all about responsible use. How does one look out for the ethical traps, and how does one use this technology properly? We start the whole workshop by having one of our faculty members give an overview of what this technology is to help demystify the backbox, if you will.

The idea, basically, is to show that not only are we (Northeastern) aware of the technological developments taking place, but that we have some of the top experts in the world leading the way. And we are already using this stuff in the classroom as of last semester. Additionally, we want to communicate that we're here and ready to work with companies, with organizations to learn ways to best utilize this technology—and to do so properly and responsibly.

There's plenty of evidence now known that ChatGPT has a human-in-the-loop component. Sometimes humans are answering questions, especially when the algorithm gets in trouble. They review the answers and intervene. By the way, this is run-of-the-mill stuff for even Google Search engine. Many people don't know that when they use the Google Search engine, that the MLR, or the machine learning relevance algorithm that decides which page is relevant to which query—that gets retrained three or four times a day based primarily on human editorial input. There's a lot of stuff that an algorithm cannot capture—that the stochastic parrot will never understand.

Recently, there's been a lot of concern about the threats posed by, in particular, how AI development may be outpacing our safety controls and protocols. What do you make of those concerns?

Those concerns are focusing on the wrong things. Let me say a few things. We did go through a bit of a phase transition around 2015 or 2016 with these kinds of technologies. Take handwriting recognition, for example. It had jumps over the years, but it took about 15 years to get there, with many revisions along the way. Speech recognition: the same thing. It took a long time, then it started accelerating; but it still took some time.

With these large language models, like reading comprehension and language compilation, we see major jumps that happened with the development of these large language models that are trained on these large bodies of literature or text. And by the way, what is not talked about a lot is that OpenAI had to spend a lot of money curating that text; making sure it's balanced. If you train a large language model on two documents that have the same content by two different outcomes, how does the algorithm know which one is right? It doesn't. Either a human has to tell it, or it basically defaults to saying, "Whatever I see more frequently must be right." That creates fertile ground for misinformation.

Now, to answer your question about this proposed moratorium. In my mind, it's a little bit silly in its motivations. Many of the proponents of this come from a camp where they believe we're at the risk of an artificial "general" intelligence—that is very far from true. We're very, very far from even getting close to that. Again, these algorithms don't know what they are doing. Now, we are in this risky zone of misusing it. There was a recent example from Belgium where someone committed suicide after six months of talking to a chatbot that, in the end, was encouraging him to do it.

So there are a lot of dangers that we need to contend with. We know there are issues. However, stopping isn't going to make any difference. In fact, if people agreed to stop, only the good actors will; the bad actors

continue on. What we need to start to do, again, is emphasize the fact that fluency, eloquence is not intelligence. This technology has limitations; let's demystify them. Let's put it to good use so we can realize what the bad uses are. That way we can learn how they should be controlled.

Provided by Northeastern University

Citation: Is 'generative' AI the way of the future? Expert explains new models, the need for human involvement (2023, April 20) retrieved 4 December 2023 from <https://techxplore.com/news/2023-04-generative-ai-future-expert-human.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.