DOCUMENT A **Difficulty: ***

A few weeks ago, Wharton professor Ethan Mollick told his students to play around with GPT, an artificial intelligence model, and see if the technology could write an essay based on one of the topics in his course. The algorithmically generated essays were at least reasonable, Mollick recalled. They also passed another critical test: a screening by Turnitin, a popular anti-plagiarism software. AI, it seems, had suddenly gotten pretty good.

It certainly feels that way right now. Over the past week or so, screenshots of conversations with ChatGPT, the newest iteration of the AI model developed by the research firm OpenAI, have gone viral on social media. People have directed the tool, which is freely available online, to make jokes, write TV episodes, compose music, and even debug computer code. Along with the recent updates to DALL-E, OpenAI's art-generation software, and Lensa AI, which can produce digital portraits with the help of machine learning, GPT is a stark wakeup call that artificial intelligence is starting to rival human ability.

How does GPT even work? At its core, the technology is based on a type of artificial intelligence called a language model, a prediction system that essentially guesses what it should write, based on previous texts it has processed. GPT was built by training its AI with an extraordinarily large amount of data, much of which comes from the vast supply of data on the Internet.

OpenAI is trying to commercialize its technology, but this current release is supposed to allow the public to test it. The technology certainly has its flaws. While the system is theoretically designed not to cross some moral red lines — it's adamant that Hitler was bad — it's not difficult to trick the AI into sharing advice on how to engage in all sorts of evil and nefarious activities, particularly if you tell the chatbot that it's writing fiction.

Both GPT's impressive capabilities and its limitations reflect the fact that the technology operates by generating ideas based on what it has read and processed before. For this reason, the AI can sound extremely confident while not displaying a particularly deep understanding of the subject it's writing about. This is also why it's easier for GPT to write about commonly discussed topics, like a Shakespeare play or the importance of mitochondria.

"It wants to produce texts that it deemed to be likely, given everything that it has seen before," explains Vincent Conitzer, a computer science professor at Carnegie Mellon. "Maybe it sounds a little bit generic at times, but it writes very clearly. It will probably rehash points that have often been made on that particular topic because it has, in effect, learned what kinds of things people say."

So for now, we're not dealing with an all-knowing bot. Overall, the system is perfectly comfortable making stuff up, which obviously makes no sense upon human scrutiny. These limitations might be comforting to people worried that the AI could take their jobs, or eventually pose a safety threat to humans. But AI is getting better and better, and even this current version of GPT can already do extremely well at certain tasks.

Right now, it's not clear how many enterprising students might start using GPT, or if teachers and professors will figure out a way to catch them. Still, these forms of AI are already forcing us to wrestle with what kinds of things we want humans to continue to do, and what we'd prefer to have technology figure out instead.

"My eighth grade math teacher told me not to rely on a calculator since I won't have one in my pocket all the time when I grow up," Phillip Dawson, an expert who studies exam cheating, told Recode. "We all know how that turned out."

"AI is finally good at stuff, and that's a problem,"
By Rebecca Heilweil, VOX, 2022

> Take notes about what GPT can do and how.

> Take notes about its flaws and limitations.

DOCUMENT B **Difficulty: ****

During 2023, the shape of politics to come appeared in a video. In it, Hillary Clinton — the former Democratic party presidential candidate and secretary of state — says: "You know, people might be surprised to hear me saying this, but I actually like Ron DeSantis a lot. Yeah, I know. I'd say he's just the kind of guy this country needs."

Further investigations found the video was produced using generative artificial intelligence (AI).

The Clinton video is only one small example of how generative AI could profoundly reshape politics in the near future. Experts have pointed out the consequences for elections. These include the possibility of false information being created at little or no cost and highly personalised advertising being produced to manipulate voters. The results could be so-called "October surprises" —ie a piece of news that breaks just before the US elections in November, where misinformation is circulated and there is insufficient time to refute it — and the generation of misleading information about electoral administration, such as where polling stations are.

Concerns about the impact of generative AI on elections have become urgent as we enter a year in which billions of people across the planet will vote. During 2024, it is projected that there will be elections in Taiwan, India, Russia, South Africa, Mexico, Iran, Pakistan, Indonesia, the European Union, the US and the UK. It is likely that each of these elections will be influenced by new generative AI technologies in the same way the elections of the 2010s were shaped by social media.

While politicians spent millions harnessing the power of social media to shape elections during the 2010s, generative AI effectively reduces the cost of producing empty and misleading information to zero. This is particularly concerning because during the past decade, we have witnessed the role that so-called "bullshit" can play in politics. In a short book on the topic, the late Princeton philosopher Harry Frankfurt defined bullshit specifically as speech intended to persuade without regard to the truth. Throughout the 2010s this appeared to become an increasingly common practice among political leaders. With the rise of generative AI and technologies such as ChatGPT, we could see the rise of a phenomenon my colleagues and I label "botshit".

Humans might use untrue material created by generative AI in an uncritical and thoughtless way. And that could make it harder for people to know what is true and false in the world. If AI-produced hallucinations are used to answer important but difficult to verify questions, such as the state of the economy or the war in Ukraine, there is a real danger it could create an environment where some people start to make important voting decisions based on an entirely illusory universe of information. There is a danger that voters could end up living in generated online realities that are based on a toxic mixture of AI hallucinations and political expediency.

Although AI technologies pose dangers, there are measures that could be taken to limit them. Technology companies could continue to use watermarking, which allows users to easily identify AI-generated content. They could also ensure AIs are trained on authoritative information sources. Journalists could take extra precautions to avoid covering AI-generated stories during an election cycle. Political parties could develop policies to prevent the use of deceptive AI-generated information. Most importantly, voters could exercise their critical judgment by reality-checking important pieces of information they are unsure about.

The rise of generative AI has already started to fundamentally change many professions and industries. Politics is likely to be at the forefront of this change.

"Beware the 'botshit': why generative AI is such a real and imminent threat to the way we live" By Andre Spicer, The Guardian, January 3rd, 2024

> Take notes about what generative AI is and what it can do.

> Take notes about the potentially harmful consequences.

DOCUMENT C **Difficulty: ***

A scientific discipline, AI officially began in 1956, during a summer workshop organized by American researchers at Dartmouth College in New Hampshire. Since then, the term "artificial intelligence", probably first coined to create a striking impact, has become so popular that today everyone has heard of it.

However, the success of the term AI is sometimes based on a misunderstanding, when it is used to refer to an artificial entity endowed with intelligence and which, as a result, would compete with human beings. This idea, which refers to ancient myths and legends, like that of the golem [from Jewish folklore, an image endowed with life], have recently been revived by contemporary personalities including the British physicist Stephen Hawking (1942-2018), American entrepreneur Elon Musk, American futurist Ray Kurzweil, and proponents of what we now call Strong AI or Artificial General Intelligence (AGI). We will not discuss this second meaning here, because at least for now, it can only be ascribed to a fertile imagination, inspired more by science fiction than by any tangible scientific reality confirmed by experiments and empirical observations.

For the researchers who coined the expression, AI was initially intended to simulate each of the different faculties of intelligence using machines. More precisely, this scientific discipline was based on the conjecture that all cognitive functions — especially learning, reasoning, computation, perception, memorization, and even scientific discovery or artistic creativity — can be described with such precision that it would be possible to programme a computer to reproduce them. [...]

Many achievements using AI techniques surpass human capabilities [...]. Computers are proving, or helping to prove, mathematical theorems; knowledge is being automatically constructed from huge masses of data using machine learning techniques.

As a result, machines can recognize speech and transcribe it —just like typists did in the past. Computers can accurately identify faces or fingerprints from among tens of millions, or understand texts written in natural languages. Using machine learning techniques, cars drive themselves; machines are better than dermatologists at diagnosing melanomas using photographs of skin moles taken with mobile phone cameras; robots are fighting wars instead of humans; and factory production lines are becoming increasingly automated. Today, AI applications affect almost all fields of activity[...], transforming many trades and eventually eliminating some.

With AI, most dimensions of intelligence - except perhaps humour - are subject to rational analysis and reconstruction, using computers. Moreover, machines are exceeding our cognitive faculties in most fields, raising fears of ethical risks.

These risks fall into three categories —the scarcity of work, because it can be carried out by machines instead of humans; the consequences for the autonomy of the individual, particularly in terms of freedom and security; and the overtaking of humanity, which would be replaced by more "intelligent" machines.

However, if we examine the reality, we see that work (done by humans) is not disappearing, but it is changing and calling for new skills. Similarly, an individual's autonomy and freedom are not inevitably undermined by the development of AI —so long as we remain vigilant in the face of technological intrusions into our private lives.

Finally, contrary to what some people claim, machines pose no existential threat to humanity. Their autonomy is purely technological, in that it corresponds only to material chains of causality that go from the taking of information to decision-making. On the other hand, machines have no moral autonomy, because even if they do confuse and mislead us in the process of making decisions, they do not have a will of their own and remain subjugated to the objectives that we have assigned to them.

"Artificial Intelligence: Between Myth and Reality,"
by Jean-Gabriel Ganascia, *unesci.org*, June 25th, 2018

> Explain how the author defines AI and its applications.
> Identify the fears associated with AI. Say if the author thinks they are founded.