

What is the intended purpose and function of your computing innovation?

ChatGPT is an artificial intelligence based language model that uses recent advances in transformers in order to create generative content based on massive amounts of online data. Given the large data set that it is trained on, the model is able to create generative content about almost anything, and in some cases can replace typical search engines such as Google. It is also being used in order to replace customer service agents by having conversations with users and being able to respond back meaningfully in an automated fashion. It can automate the brainstorming for a news article, summarize wikis, or recently Khan Academy's use of the software to create interactive learning assistants [1].

Describe the development that went into the creation of your computing innovation - the who, what when where and how.

OpenAI originally began as a non profit that was created by an elite team of AI researchers, and famous entrepreneur Sam Altman, who has strong affiliations with ycombinator and the startup ecosystem. Their first main product was ChatGPT, which was launched in November 2022 and it reached 100 million users faster than any other product ever. It was created in the San Francisco area [3]. The company used a lot of the advancements on research of transformers, which are specialized algorithms that are able to recognize patterns in bits of data [2].

What is a beneficial effect of your computing innovation?

ChatGPT and the platforms that have been built on top of it have created a spur of companies over the past year that have had some serious impacts. One of which is the increased productivity of knowledge workers. Due to its quick ability to create quite reliable content, many professions have started to use this reliable content to increase the output of their workers. Programmers now have to write a fraction of the code they once had to [3], as ChatGPT is able to generate a lot of the code itself. Recently released products like github copilot have been using OpenAI's api in order to power their code helper tool that is built into modern development environments like VS Code. This will lead to smaller teams of people or developers to create more with less. We could see teams of 10 people being able to output as much code as a 100 person company. This means there will be a lot more opportunities for startups.

What is a harmful effect of your computing innovation?

One harmful effect is the potential of the removal of certain more monotonous jobs or knowledge work in favor of ChatGPT or other platforms. For example, many law firms have lower level associates that take care of lower level functions and perform less decision heavy tasks. Many of these things can now be done by a language model. Given a case and the details of such a case and free access to internet data up until 2021, it is quite reasonable to assume that ChatGPT can perform very similar tasks to current lawyers. This could lead to higher levels of unemployment as the model continues to get better. Even now we are seeing layoffs due to this technology which is still so new [4].

What is the data your computing innovation uses and how is that data consumed, produced or transformed?

Data is the main driving force behind ChatGPT and more generally, most AI driven technologies. Similarly to how humans learn, these large language models need to consume large amounts of data. This specific model uses a deep neural network, which can be thought of as a series of neurons that parallel human brains. This neural network allows ChatGPT to recognize patterns in the text in order to predict what's next [5]. To the model, all of our text can be boiled down to tokens, which are just a series of integers [6], but by recognizing patterns in these tokens it is able to generate coherent responses. GPT 3's data set consists of 60% filtered web data that is known as "common crawl" data, which contains articles, blog posts, and other information stored in web pages over 8 years of web crawling. GPT 3 also uses data from common hubs of information such as reddit and Wikipedia, making up 22% and 3% of the model's dataset respectively. 16% of the dataset comes from books and academic journals [7].

Provide an example of how this computing innovation may be used beyond the intended purpose and function.

Although ChatGPT has opened up an entirely new realm of automation, it has already had some unintended consequences due to its mass popularity. One of these complications comes in education and plagiarism. Students are now able to prompt ChatGPT with an essay prompt for example such as “explain the historical significance of 9/11 and how it has affected the greater western world?”, and it will generate a well thought out response with research that it has already done due to its very large data set from the open internet. Even teachers are saying that the model is capable of writing quality work [8]. This has forced some universities to step in, Villanova University recently added to their academic integrity policy to remove the right of students to use GPT generated text [9]. As the technology gets better, and as it becomes more integrated into our lives, it may become easier for students to remove the burden of coming up with original ideas and content. This could have massive implications down the line if students become reliant on an Artificial Intelligence model that is far more thoughtful than them. Perhaps students will get progressively less capable of coming up with ideas, which could impact the progress of humans all together.

References:

- [1] [Khan Academy](#)
- [2] [TechTarget](#)
- [3] [Who created OpenAI?](#)
- [4] [CNBC](#)
- [5] [Zapier Blog](#)
- [6] [OpenAI Tokenizer](#)
- [7] [ChatGPT Stats](#)
- [8] [Teacher review of GPT](#)
- [9] [Villanova Integrity policy change](#)

