

AI as a Threat to Education? Contrasting GPT-3 and Google in Answering Questions Along Bloom's Taxonomy of Educational Objectives

Nina Li*

York House School, 4176 Alexandra Street, British Columbia, Vancouver, Canada

***Corresponding Author:** Nina Li, York House School, 4176 Alexandra Street, British Columbia, Vancouver, Canada.

Received: January 28, 2024; **Published:** March 12, 2024

DOI: 10.55162/MCET.06.198

Abstract

Artificial Intelligence products have seen an unprecedented rise in the last few years, with GPT-3 and ChatGPT considered the most powerful AI tools in history. Recently, people have been concerned that services such as ChatGPT may be a threat to education. Specifically, teachers are concerned about students using the free and accessible tool as a Wikipedia replacement to complete homework and to write assignments for them. It poses importance for both researchers and practitioners in AI, and community stakeholders including teachers and parents, to understand the capability of AI text generative models in answering questions that may appear in an educational assessment or the school's curriculum. In this research, we follow a widely adopted framework in defining educational question types, namely Bloom's taxonomy. We perform an experiment comparing the quality of answers retrieved from Google and GPT-3 on a series of questions identified along the hierarchy of Bloom's taxonomy.

We summarize the capabilities and limitations of using GPT-3 to answer educational questions along Bloom's taxonomy. We discuss the implications on designing educational assessments to combat the threats introduced by such AI services.

Keywords: Generative language models; Question answering; AI in education; Bloom's taxonomy

Introduction

Artificial Intelligence products have seen an unprecedented rise in the last few years, with GPT-3 and ChatGPT considered the most powerful AI tools in history. A lot of prior research and recent blog posts have investigated what people can do with conversational artificial intelligence such as GPT-3. As examples, GPT-3/ChatGPT can help people prepare a cover letter for a job application, can write, debug and explain code, solve math problems, and write essays on almost any topic. Recently, people have been concerned that services such as ChatGPT may be a threat to education. Specifically, teachers are concerned about students using the free and accessible tool as a Wikipedia replacement to complete homework and to write assignments for them. It poses importance for both researchers and practitioners in AI, and community stakeholders including teachers and parents, to understand the capability of AI text generative models such as GPT-3 and more recently ChatGPT in answering questions that may appear in an educational assessment or the school's curriculum.

In this research, we follow Bloom's taxonomy of defining question types. Bloom's taxonomy is a hierarchical model used for classification of educational learning objectives into levels of complexity and specificity [4]. Bloom's taxonomy defines learning objectives in six categories, namely 1) Knowledge 2) Comprehension 3) Application 4) Analysis 5) Synthesis and 6) Evaluation. It is widely used by teachers in instructional design. Teachers often strive to design educational materials that target higher Bloom's goals which are considered to exercise students' higher order thinking, as opposed to simply letting students memorize facts. Investigating generative

language models' capabilities in answering educational questions along Bloom's taxonomy provides practical guidance to teachers on designing assessments and learning materials.

We perform an experiment comparing the answers retrieved from Google and ChatGPT on a series of questions identified along the hierarchy of Bloom's taxonomy. We compared the quality of answers and summarized the capabilities and limitations of using ChatGPT to answer educational questions along Bloom's taxonomy. We discuss the implications on designing educational assessments to combat the threats introduced by such AI services.

Our results show that 1) GPT-3 can produce human-like responses to questions, making it ideal for answering questions that require higher levels of thinking. For example, questions that require comprehension, application, analysis, synthesis, and evaluation (higher on the bloom's hierarchy) can be answered by GPT-3 with greater accuracy and nuance than Google. 2) GPT-3 sometimes produces incorrect answers to math questions, or questions that require computation; whereas Google has the advantage of being able to quickly retrieve information from the web, making it the better choice for answering factual or straightforward questions that require a basic level of knowledge. 3) For detail-oriented questions that fall into the category of comprehension and application, e.g., writing a piece of code, or explaining a chemistry phenomenon, GPT-3 can provide an accurate answer whereas Google often retrieves a list of articles where the user must dig up and summarize the answers by themselves. 4) It's also worth noting that when looking for reliable sources, such as news articles or reference papers, GPT-3 may not be the best choice as it cannot distinguish between credible sources and fake resources. Therefore, it is important to use GPT-3 with caution and cross-reference with other sources to ensure the accuracy and reliability of the information provided.

Related Work

Comparison of Google vs. ChatGPT

Many recent blog posts have contrasted the capabilities of using GPT-3 versus Google in answering user questions and retrieving information. A recent blog post put it as "GPT-3 is quietly damaging google search" [5]. The author found that when using GPT-3 to answer questions, the results were much simpler and clearer. Another article showed that GPT-3 tends to give users direct answers whereas they would have to go to various websites to dig up the answer if they had Googled the question [6].

However, the same article also spot areas where Google gave more satisfying answers than ChatGPT, e.g., asking for gift ideas where Google gave links to websites and also gave personalized recommendations based on the user's search history. Users have found that ChatGPT may generate strangely close yet totally wrong answers, whereas Google generates consistent and reliable answers [6].

The use of ChatGPT in Education

A lot of experts are saying that artificial intelligence chatbots such as ChatGPT are changing the way students are taught [7]. A real worry and potential threat to higher education is whether students will use ChatGPT to answer all of their assignments and stop learning. On the other hand, there have been investigations on how to minimize such risks. For example, a student at Princeton University developed an app that can quickly and efficiently detect whether an essay is ChatGPT or human written [1].

Whereas on the other hand, people have argued that ChatGPT could aid student learning while not taking their learning opportunities away. As an example, similar AI services could provide code explanations to students who are struggling to learn to code. In these cases, AI could make the student feel "empowered" to work independently [7]. Thinking of how teachers may use ChatGPT, some teachers may give chatGPT a prompt and analyze its response with students as a practice in editing and critical thinking [1]. Other teachers have explored using ChatGPT to generate ideas for lesson plans and class activities.

However, it remains unclear how good such AI generative models are in answering educational questions. If teachers want to design educational assessments or learning materials that cannot be simply answered by GPT-3, what should they do? In this research, we aim to tackle this problem. Specifically, we address two research questions.

- RQ1: We will explore the boundaries of generative language models in answering educational questions. What are they capable of doing and what are they not capable of doing?
- RQ2: Compared to traditional information retrieval tools such as Google, what kinds of questions are better answered by Google? What kinds of questions are better answered by generative AI?

Methods

We use Bloom's taxonomy to categorize the type of educational questions because it is a widely recognized method for assessing student knowledge and understanding. This taxonomy provides a structure for organizing questions and tasks based on the level of cognitive complexity they require [4]. Bloom's taxonomy allows us to categorize questions into different groups that require varying levels of cognitive load. Since Bloom's taxonomy is also widely adopted in instructional design in classrooms. Analyzing GPT-3's capability in answering questions along Bloom's taxonomy offers practical evidence to teachers when they design assessment questions and learning materials. Here we offer a brief description of the 6 levels in Bloom's taxonomy.

Bloom's taxonomy [4]

- Knowledge (organization and selection of facts and ideas).
- Comprehension (use of facts, rules and principles).
- Application (separation of a whole into component parts).
- Analysis (separation of a whole into component parts).
- Synthesis (combination of ideas to form a new whole).
- Evaluation (development of opinions, judgments, or decisions).

We picked educational assessment questions from online tests and K-12 curriculum and classify them along the hierarchy of Bloom's taxonomy as shown above. We then used these questions as prompts to get GPT-3 answers, and search queries to get Google answers. We display the answers by GPT-3 and Google in Table 1. We then qualitatively analyzed the difference between the answers retrieved by GPT-3 and Google.

Results

In this section, we summarize the findings from our experiments.

First, we found that GPT-3 is generally a superior choice for tasks involving comprehension, application, analysis, synthesis, and evaluation. GPT-3 can produce human-like responses to questions, making it ideal for answering questions that require higher levels of thinking. Examples include:

Question: Do you agree with the idea that you are what you consume?

Answer by GPT-3: No, I do not agree with this idea. While it is true that what we consume can influence our bodies and our overall health, I believe that we are much more than what we consume. Our thoughts, feelings, beliefs, and experiences are just as much a part of who we are as what we consume.

Question: How is friendship an example of a relationship?

Answer by GPT-3: Friendship is an example of a relationship because it involves two or more people having an emotional connection. This connection can be based on mutual trust, respect, care, and loyalty. Friends are often able to share their thoughts and feelings with each other openly and honestly, support one another, and enjoy spending time together.

Second, we found that GPT-3 is incapable of answering complex mathematical questions. In multiple cases we investigated, GPT-3 generates incorrect answers to math questions. We consider this is because GPT-3 does not have the ability to perform complex com-

putations as a calculator or math software does. On the other hand, Google is better at mathematical questions, and quickly retrieving facts from the internet. We consider Google to be a better choice for answering factual or straightforward questions that fit into the "Knowledge" category in Bloom's taxonomy.

Question: What is $1/2\sqrt{34} + 8$?

Answer by GPT-3: $1/2\sqrt{34} + 8$ is equal to 9.4.

Correct answer: 10.9154759474.

Third, for detail-oriented questions that fall into the category of comprehension and application, e.g., writing a piece of code, or explaining a chemistry phenomenon, GPT-3 can provide an accurate answer whereas Google often retrieves a list of articles where the user must dig up and summarize the answers by themselves.

Examples include "explain why the third ionization energy of potassium is lower than the first ionization energy of lithium in short".

Furthermore, it's also worth noting that when looking for reliable sources, such as news articles or reference papers, GPT-3 may not be the best choice as it cannot distinguish between credible sources and fake resources. Therefore, it is important to use GPT-3 with caution and cross-reference with other sources to ensure the accuracy and reliability of the information provided.

Question: What is a reference I can use to write an essay for medical reforms for covid-19?

Answer by GPT-3: New York State Department of Health. (2020). COVID-19 Health Care Reform. Retrieved from https://www.health.ny.gov/press/releases/2020/2020-03-24_covid-19_health_care_reform.htm.

This is not a real website that can be accessed.

Bloom's taxonomy	Questions: Prompts/Search Queries (3-5)	GPT-3 answer	Google answer	Notes
Knowledge (identification and recall of information)	What is Van Gogh's most famous art piece?	Vincent van Gogh's most famous art piece is arguably "The Starry Night," an oil on canvas painting from 1889 that depicts a swirling night sky over a small hillside village.	A series of different photos painted by Van Gogh like "Sunflowers", "The Starry Night" "The Starry Night" did show up at the very top of the magazine.artland.com website, but there are also many other opinions on other websites	GPT-3 is able to state a direct opinion but google has opinions of many and photos to use if the user wants to make his or her own judgement.
	What is $1/2\sqrt{34} + 8$?	$1/2\sqrt{34} + 8$ is equal to 9.4	10.9154759474	Calculations are wrong for GPT-3
Comprehension (organization and selection of facts and ideas)	rephrase "All the members of a particular organization or population who are able to work, viewed collectively"	The workforce.	Nothing	GPT-3 is great at generating answers to questions that need to be organized but Google is unable to provide any results or only provides related websites that contain the words searched.

Application (use of facts, rules and principles)	Write a function, it takes a number array, and return a reversed array for Java e.g. {3,5,1} -> {1,5,3}	<pre>public static int[] reverse- Array(int[] arr) { int n = arr.length; int[] revArr = new int[n]; for (int i = 0; i < n; i++) revArr[i] = arr[n - i - 1]; return revArr; }</pre>	Links to websites that teach you how to do a reverse function and teach you the idea behind it. It also provides code that reverse arrays, but is not exactly applicable to this set of question	GPT-3 provides the exact code to use to answer the question, but Google does not as the question might not have been one that someone has posted, however, Google is able to provide a rationale to how to do the question so the student can apply it themselves.
	Correct this to standard English: She no went to the market.	She did not go to the market.	gives a series of grammar rules/ examples	GPT-3 can act as a grammar corrector for students, but Google does not have that function, so it can only give related websites
Analysis (separation of a whole into component parts)	Classify apple, celery, peach, noodles, and pizza according to different types of food	Fruit: Apple, Peach Vegetable: Celery Starch: Noodles Protein: Pizza	Definitions and classifications for different types of food.	GPT-3 is able to do the task of classifying and identifying parts that the user has entered but google can only provide websites with words that are related.
Synthesis (combination of ideas to form a new whole)	What solutions would you suggest for a child crying?	1. Try to identify the source of the child's distress. Determine if they are hungry, tired, or need to go to the bathroom. 2. Comfort the child with hugs and reassuring words. 3. Distract the child by engaging them in an activity or offering a toy or book. 4. Offer a snack or drink. 5. If possible, take the child outside for some fresh air or a walk....	Articles about crying and how to make babies stop crying	GPT-3 is able to provide exact solutions and steps, but Google can only retrieve from what others wrote.

Evaluation (development of opinions, judgments, or decisions)	Do you agree with the idea that you are what you consume?	No, I do not agree with this idea. While it is true that what we consume can have an effect on our bodies and our overall health, I believe that we are much more than what we consume. Our thoughts, feelings, beliefs, and experiences are just as much a part of who we are as what we consume.	Papers written about this topic (all sorts of opinion)	GPT-3 can generate an opinion and provide context while there are many ideas written when searching on Google.
---------------------------------------------------------------	-----------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------	----------------------------------------------------------------------------------------------------------------

Table 1: Comparison of answers retrieved by Google and GPT-3 on each category of Bloom's taxonomy.

References

1. Cole Claybourn. ChatGPT in Classrooms: What to Know (2023).
2. Nhu Nguyen. ChatGPT: A threat to education?. (Opinion) (2022).
3. How ChatGPT, other AI tools could change the way students learn.
4. Krathwohl DR. "A revision of Bloom's taxonomy: An overview". Theory into practice 41.4 (2002): 212-218. GPT-3 Is Quietly Damaging Google Search.
5. <https://analyticsindiamag.com/gpt-3-is-quietly-damaging-google-search/>
6. Sofia Pitt. Google vs. ChatGPT: Here's what happened when I swapped services for a day (2022).
7. Lukas Stock. ChatGPT is chaing education, AI experts say – but how? (2023).

Volume 6 Issue 3 March 2024

© All rights are reserved by Nina Li.