

ChatGPT's Impact on the Value of Academic Degrees
Exploring the Intersection of AI and Education for More
Effective Educational Systems

Presented to the S. Daniel Abraham Honors Program

in Partial Fulfillment of the Requirements

for Completion of the Program

Stern College for Women

Yeshiva University

May 2, 2024

Limor Kohanim

Mentor: Professor Lawrence Teitelman, Computer Science

Abstract

In recent years, artificial intelligence (AI) has become a significant player in various fields, including education. Among these AI systems is ChatGPT, which allows users to interact with a large language model through generative AI and natural language processing (NLP). This thesis aims to investigate the impact of AI systems like ChatGPT on education, particularly within the realm of computer science. It explores whether ChatGPT will revolutionize teaching or merely fade away as a passing trend. This study delves into how ChatGPT can potentially aid in earning a computer science degree, if properly regulated and understood. Utilizing articles from scholars and anonymous surveys of students' experiences with ChatGPT, this research seeks to understand its role in education and its implications for the future.

Table of Content

Introduction.....	4
A Phenomenon That Has Become More Popular.....	4
Introducing the Research Question.....	5
AI.....	6
The History of AI.....	6
AI Chatbots.....	7
ChatGPT.....	8
The ChatGPT Timeline.....	8
What Is Deep Learning.....	8
ChatGPT from a Technical Point of View.....	9
Large Language Models (LLMs).....	10
Natural Language Processing (NLP).....	10
The GPT Model.....	11
What Does GPT Stand For.....	13
The Transformer Based Architecture.....	14
ChatGPT Training Process.....	15
Pre-Processing of Text Data.....	16
Training Algorithm.....	17
Pre-Training Phase.....	17
Fine-Tuning Phase.....	18
Focusing on Academia.....	18
Fields that Have Been Impacted by ChatGPT.....	18
ChatGPT in Computer Science Education.....	19
The Research.....	20
Research Methodology.....	21
Results.....	21
Examples.....	24
ChatGPT Solving Leetcodes.....	27
Biases.....	28
Surveys and statistics.....	28
Generic Statistics on Students Using ChatGPT on Assignments.....	28
Exploring ChatGPT Usage and Perceptions Among Stern College for Women Students.....	29
Conclusion.....	31
Pros and Cons.....	31
How Can ChatGPT Be Helpful in Academia.....	32
References.....	34

Introduction

A Phenomenon That Has Become More Popular

One of the most researched and talked-about phenomena of the past year has been the launch of new generative AI programs. Moreover, there has been an increase in interest in Artificial Intelligence (AI) as it has become significantly more accessible and indispensable across various fields. This rise in interest can be attributed in part to the latest APIs and plugin solutions offered by OpenAI. The following graph (Fig. 1) illustrates the exponential increase in search volume for the keyword “AI”, which has tripled from around 7.9 million monthly searches to over 30.4 million during the last month of the measured period [18].

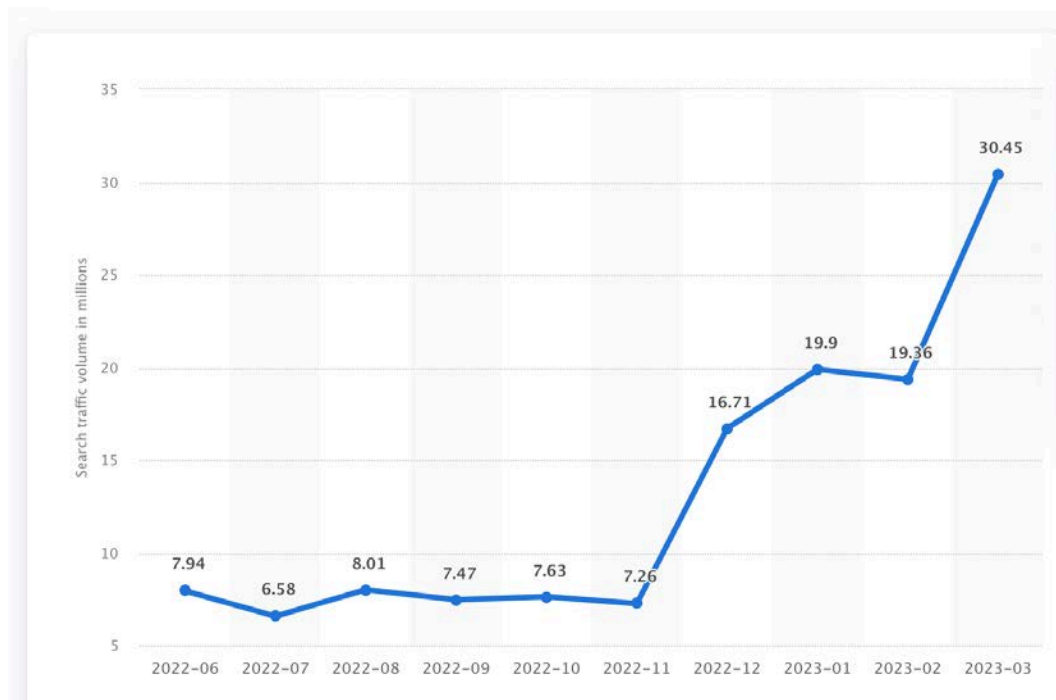


Figure 1 [18]

According to numerous reports, ChatGPT, one of the most widely used products by OpenAI, represents a significant breakthrough in the field of artificial intelligence. The pre-trained AI model has achieved immense popularity over the past year and a half, with 1 million users in its first week and now exceeding 100 million users. These substantial advancements in AI are worth careful analysis and study in order to take advantage of their potential benefits. Instead of avoiding such phenomena, we should learn from and use technologies like ChatGPT to improve various fields, while ensuring regulations are in place to prevent misuse. Particularly, it is crucial to explore the advantages of platforms like AI chatbots for students and professors as it can help modernize the educational system.

Introducing the Research Question

It is widely acknowledged that the utilization of AI tools has particularly impacted the fields that are closely related to computer science and programming. While AI plugins have showcased their utility, their susceptibility to errors precludes them from entirely replacing the role of software developers. However, the advent of these innovative technologies has raised significant inquiries regarding its utility in student engagement, not just confined to programming courses but extending to diverse academic disciplines. This dissertation aims to explore the nuanced balance between the excessive adoption of such potent technologies and the potential benefits they offer, particularly within the realms of computer science and academia. To address this question, the investigation will commence with an analysis of the history and underlying technology of tools such as ChatGPT. Subsequently, the focus will shift to examining its impact on the academic landscape through an analysis of various surveys conducted on students and their experiences with the technology, including one conducted by the author. Additionally, the research will involve an examination of professors' and other researchers'

papers on the subject, as well as an analysis of the specific subjects in which ChatGPT has demonstrated excellence. The more one learns about the capabilities of platforms like ChatGPT, the clearer it becomes that researching and understanding how chatbots can be appropriately utilized in education is imperative for enriching and improving students' education, rather than providing them with detrimental shortcuts.

AI

The History of AI

Let us delve deeper into the history of Artificial Intelligence. According to Copeland, “Artificial intelligence is the ability of a computer or computer-controlled robot to perform tasks commonly associated with the intellectual processes characteristic of humans, such as the ability to reason.” [2]. This definition, as per Britannica, encapsulates the essence of artificial intelligence. While AI has garnered unprecedented attention in the past year, its origins can be traced back to the years between 1950 and 1956. It all started with Alan Turing's seminal work “Computer Machinery and Intelligence” and the renowned Imitation Game, which tested a computer's capabilities. The term “artificial intelligence” itself was coined by John McCarthy during a workshop at Dartmouth in 1955. The culmination of this new technology occurred in 1997 with the introduction of IBM's AI Deep Blue, which defeated a chess world champion for the first time in history. Other significant milestones in the adoption of AI tools include Jeff Dean's and Andrew Ng's training of a neural network capable of recognizing cats in 2012, Facebook's AI chatbots in 2017, Alibaba's language-processing AI surpassing human comprehension abilities in reading, and OpenAI's GPT-3 beta testing commencing in 2020 [4]. Notably, GPT-3 was the first model to utilize Deep Learning to create code, text, poems, songs,

and other tasks comparable to those crafted by humans. Even more groundbreaking is OpenAI's invention DALL-E, announced in January 2021, which generates and understands images [4]. As evidenced, the history of AI dates back to the 1950s, but it has only been in recent years, particularly after 2016, that we have witnessed a significant surge in the utilization of chatbots across various domains, including customer service, healthcare, education, business, entertainment, and robotics. This surge has made this technology accessible to a wide audience, including individuals lacking programming or quantitative abilities.

AI Chatbots

As mentioned earlier, the emergence of chatbots has revolutionized various industries, including academia. These AI-based interfaces facilitate human interaction with models through natural language processing, or NLP. Initially employed for virtual assistance, they have evolved to engage with humans on an “emotional” level through technologies like sentiment analysis. Today, chatbots and other AI tools serve broader purposes beyond automating customer interactions. They now assist in writing, providing information, education, language translation, programming, and entertainment. Along with OpenAI’s ChatGPT, other prominent interfaces include Google Bard, GitHub Copilot, Microsoft's new Bing, Amazon CodeWhisperer, and more. While this research paper primarily focuses on ChatGPT’s capabilities, similar principles apply to all the mentioned products.

ChatGPT

The ChatGPT Timeline

Although OpenAI officially launched its chatbot in November 2022, the development process took several years. Drawing from Kristi Hines's "History of ChatGPT: A Timeline of the Meteoric Rise of Generative AI Chatbots," the timeline is structured as follows [4]. In June 2016, OpenAI initiated the release of research on generative models trained on enormous datasets containing various media types, aiming to simulate the input data accurately [4]. On September 19, 2019, the company published additional research focusing on fine-tuning the GPT-2 language model [4]. Subsequently, on January 27, 2022, they released research on InstructGPT models, closely related to ChatGPT, showcasing enhanced instruction-following abilities, reduced fabrication of facts, and decreased toxic output [4]. Finally, on November 30, 2022, OpenAI introduced ChatGPT as part of a free research preview, utilizing GPT-3.5 [4].

What Is Deep Learning

The abovementioned new technologies are mainly based on what are known as deep learning language models. Deep learning is a subclass of Machine Learning which is part of the larger realm of AI. The models are designed to understand and create human-like text by analyzing vast amounts of natural language data through deep learning techniques. Among other things, the latter involves the use of artificial neural networks, composed of multiple features that extract complex patterns and relationships from the data. The way deep learning language models work is through large neural network architectures, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), or more commonly, transformer architectures

[10]. These models are trained on massive databases and learn their statistical properties of language, including grammar, syntax, semantics, and context. One of the key characteristics of deep learning language models is their ability to generate coherent and contextually relevant text. By leveraging the hierarchical representations learned through multiple layers of neural network processing, these models can capture complex linguistic structures and generate human-like responses to input queries or prompts. Deep learning language models have significantly advanced the state-of-the-art in NLP and continue to drive innovations in AI-based communication and understanding of human language [10].

ChatGPT from a Technical Point of View

Developed by OpenAI, ChatGPT's language model was built through the conjunction of multiple deep learning and reinforcement algorithms trained on over 150 billion text data that includes books, articles, blog posts, poems, songs, articles, conversations, and reviews. Exploring the technical workings of how ChatGPT operates, the technology leverages its underlying neural network architecture to process and generate text based on input from users. In other words, the user interacts with ChatGPT, the input is encoded and passed through layers of neural network nodes, each responsible for processing different aspects of language understanding and generation. The model then learns patterns and associations within the language provided and is subsequently able to generate coherent and contextually relevant responses [13].

Large Language Models (LLMs)

ChatGPT is part of a broader category known as large language models (LLMs), designed to analyze and recognize intricate linguistic patterns on a larger scale. These models are trained on massive datasets and are constructed through unsupervised learning techniques, enabling them to get familiar with language structures and semantics. GPT-3.5 and its newer versions are an admirable example of this phenomenon. Their extensive parameter range represents the culmination of advancements in LLMs [8] .

Natural Language Processing (NLP)

As previously noted, ChatGPT stands out as a transformative technology because it combines the principles of natural language processing (NLP) and large language models (LLMs) to enable communication between humans and machines. The NLP component enables the platform to understand and generate human-like text, facilitating communication between users and AI systems. Some of the main roles fulfilled by natural language processing include tokenization, where textual input is broken into smaller units for easier processing, and language modeling, where the model predicts the probability distribution of the next word given the preceding context [8]. Additionally, LLMs like GPT-3.5 and its newer versions excel in linguistic capabilities, allowing machines to produce coherent and contextually relevant text. Another core feature of ChatGPT, which will be thoroughly explained in the following paragraphs, is its integration within complex neural network architectures. All these advancements hold particular promise in the realm of computer science education, where ChatGPT's ability to explain complex material and generate code can potentially enhance learning experiences. However, there is a risk

that students may abuse it as a shortcut and potentially hinder their learning and skill development.

The GPT Model

The Artificial Intelligence model named GPT was developed by OpenAI in 2018 and it is at the core of the notorious platform ChatGPT. Its objective is to predict the subsequent word, as well as completing human-generated sentences and it therefore excels in fulfilling natural language processing tasks that include language generation, text prediction and many more. The deep learning architecture of the ChatGPT model is built upon a transformer-based neural network, specifically designed to process and generate natural language text. This technology is attributed to the seminal paper “Attention Is All You Need” which inaugurated the Transformer model, an innovative architectural framework utilizing a self-attention mechanism that has emerged as a cornerstone in numerous cutting-edge models within the field of natural language processing (NLP), including applications such as ChatGPT [15]. The research paper was authored by eight scientists from Google: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Lion Jones, Aidan N Gomez, Lukask Kaiser and Illia Polosukhin in 2017 [15]. Since this architecture is targeted at capturing long-range dependencies within text data, it is considered the leading approach in NLP tasks [15]. Comprising transformer blocks, each containing a self-attention mechanism and a feedforward neural network, this architecture allows the model to focus on different parts of the input text and comprehend nonlinear relationships between input and output [13]. ChatGPT utilizes an adaptation of the previously mentioned transformer architecture which is known as GPT-2. The multi-layer transformer model with a large number of parameters, enabling it to capture intricate relationships between input and output. This is even a more complex model which includes a greater number of layers and

parameters. Thanks to the augmentation, the model's effectiveness is enhanced enabling it to produce highly realistic and coherent responses to natural language input [11].

In the backend, the OpenAI API operates on various complex models, which will be listed here. Starting with the most popular, GPT-3.5 is an upgraded version of GPT-3 responsible for comprehending and producing natural language. DALL·E is a model that generates and modifies images based on natural language input. Another product is Whisper, an automatic speech recognition (ASR) system that converts audio to text. Additionally, there is Embedding, which transforms text into a numerical representation, Codex, a composition of models specifically targeting code analysis and production, and Moderation, involving a finely calibrated model capable of identifying potentially sensitive or unsafe text. Lastly, GPT-3 includes a range of models meant to understand and generate natural language. All these models encompassed in GPT-3.5 were trained on a large and diverse dataset of text and code preceding the last quarter of 2021, enabling OpenAI's products to be utilized by developers in both research and practical settings. ChatGPT was initially launched by OpenAI using its GPT-3.5 series, which, as earlier explained, is an improved and enhanced version of GPT-3. This newer version comprises various subversions that are more specific to different tasks, including gpt-3.5-turbo, known for its cost-effectiveness and capability, text-davinci-003, which produces lengthier output with superior quality and reliability at a cost ten times higher, and code-davinci-002, designed specifically for optimized performance in programming tasks [17].

The limited beta version of GPT-4, launched on March 14, 2023, is more accurate than its predecessor, as it was trained on a larger dataset and is believed to possess better reasoning abilities for problem-solving. ChatGPT's 3.5 version was trained on 175 Billion parameters while the newer version was trained on 170 Trillion parameters [7]. Additionally, GPT-4 is

multimodal, capable of accepting both text and image inputs, while the previous design only accepted text input [17]. Another feature that has expanded in ChatGPT-4 is the context length (number of tokens in the output), which went from 2048 words in version 3.5 to a choice of either 8192 or 32768 in version 4, according to the user's preference [7]. Even though GPT-4 is more effective in generating improved responses compared to the previous version, its larger scale necessitates additional processing and computing resources, making it slower overall in generating output [7].

What Does GPT Stand For

Now focusing on the technical meaning of ChatGPT, let us break down each component of the term GPT “Generative Pre-trained Transformer” and understand its meaning in detail:

1. Generative refers to the model's ability to generate coherent and contextually relevant text.

Generative models are able to create new information that emulates the input data they were trained on. In the case of ChatGPT, the model is able to generate human-like text based on the patterns and structures it has learned from the data it was trained on.

2. Pre-trained indicates that the model has undergone an initial phase of training on a large dataset and later went through a process of fine-tuning for specific tasks. Pre-training is typically done on a diverse massive data set composed of text derived from books, articles, and websites.

During this phase, GPT learns to understand the features of language, including syntax, semantics, and context, without being explicitly trained for any specific task or outcome.

3. Transformer refers to the underlying neural network architecture of the GPT model. This type of neural network is specifically designed for processing sequential data, such as natural

language, and performing tasks related to text generation. The transformer-based architecture is composed of multiple layers of self-attention mechanisms, enabling the model to capture complex dependencies and relationships between different parts of the input sequence, regardless of their positions. In other words, these mechanisms enable the model to assign levels of importance to different parts of the input sequence, based on the context and content of the other tokens in the sequence itself.

By leveraging the generative capacity of the model, pre-training it on extensive text datasets, and exploiting the efficiency of the transformer architecture in handling sequential data, GPT demonstrates remarkable performance across a spectrum of NLP tasks such as text generation, language understanding, and dialogue systems. ChatGPT is revolutionary in the sense that it is a model able to be pre-trained on vast amounts of text data and fine-tuned for specific tasks so that it can generate human-like text with remarkable fluency and coherence.

The Transformer Based Architecture

The transformer-based architecture, which is central to the models mentioned earlier, consists of five main components. The first and most important is the Self-Attention Mechanism. This feature enables the model to assess the importance of each token in the input sequence relative to every other word in the sequence. The mechanism achieves this by computing attention scores after comparing the representation of each word to that of every other word in the sequence. These scores indicate the degree of focus the model should place on each word when generating the output sequence. The second component is Multi-Head Attention. To ensure that the model can incorporate all dimensions and details of the input sequence, multiple attention heads are added to support the self-attention mechanism in the transformer architecture.

Each of these “heads” is directed towards a specific attention pattern, allowing the model to encompass a variety of connections between the tokens in the input sequence [8]. The third component is Positional Encoding. This element allows the transformer to assign positional information to the input encodings. Additionally, these architectures incorporate feed-forward neural networks, usually with one or more layers, to handle the representations derived from the self-attention layers. The feed-forward networks allow the model to capture complex patterns and relationships within the data. Lastly there is the Encoder-Decoder Architecture, implemented by many transformer-based architectures for sequence-to-sequence tasks. This kind of architecture utilizes an encoder to process the input sequence and a decoder to generate the output sequence based on the encoder's analysis and the context provided by the preceding tokens in the output sequence. In conclusion, the transformer-based architecture is selected for a variety of NLP tasks because it leverages the power of self-attention mechanisms, multi-head attention, and feed-forward networks to model dependencies and interconnections within sequential data. This is also due to its ability to be parallelizable and to analyze long-range dependencies within the data [8].

ChatGPT Training Process

What makes ChatGPT such a revolutionary and impressive model is related to its innovative training process which combines the processing of massive unlabeled text datasets and an algorithm that enables the model to generate coherent and human-like responses to natural language user prompts. In the following paragraphs you will find a summary of the training process of this innovative technology that is mentioned in the article “ChatGPT and

Open-AI Models: A Preliminary Review” by Konstantinos I. Roumeliotis and Nikolaos D. Tselikas [13].

Pre-Processing of Text Data

As mentioned earlier in the paper, ChatGPT’s model is based on a transformer neural network that was built to be most efficient in the processing and generation of natural language text. One of the most important processes that the model goes through to get to such impressive results is the pre-processing of text data. This phase is crucial as it determines the quality of the input text. The pre-processing can be split into further subtasks, mainly tokenization, subword encoding, and data cleaning. The article “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” by Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova explains each subphase in detail. Starting from Tokenization which consists of splitting the input text into “discrete units of meaning”, known as tokens [3]. The tokenization allows the model to process the input smoothly. For example, in ChatGPT’s situation, the pre-trained tokenizer converts the text into a sequence of tokens, where each token corresponds to a unit which can be a word or a subword. The result of these steps is a sequence of tokens which is then injected into the model to be processed further. The subsequent phase is “Subword Encoding”. This methodology is very common in NLP as it also looks out for uncommon words in the input text. This step fragments the input text into smaller sub-tokens that are reinjected into the model so that it can process them further. For example, in ChatGPT’s case, the Byte Pair Encoding (BPE) algorithm, which is a “pre-trained subword encoder” segments the input text very precisely and efficiently, since it is specifically designed for NLP tasks [3]. The last step in the pre-processing of text data phase is data cleaning. The goal of this task is to remove any irrelevant information from the prompt text. This will improve the overall quality of the input

data so the model will be able to process it without any issues [3]. Data cleaning includes removing punctuation, numbers, and special characters, as well as fixing spelling and grammatical mistakes. Thanks to this last step, at the end of the pre-processing of text data phase, the input text is turned into a more coherent and readable form, so that the model will be able to register all relevant patterns in the data [3].

Training Algorithm

Building upon the earlier explanation, the GPT model predicts the next token in a text sequence by utilizing preceding words and context as input through its training algorithm, which is based on unsupervised pre-training techniques [15]. The algorithm utilizes the context from the previous tokens that were processed in order to minimize the upcoming token's negative log-likelihood. The algorithm starts off with the ChatGPT model randomly assigning weights to the GPT's transformer-based neural network. The variables are initialized through a normal distribution that has a mean of zero and a standard deviation of 0.02 [11].

Pre-Training Phase

During the initial phase of pre-training, the transformer-based neural network undergoes training on an extensive dataset of unlabeled text to capture the general features and patterns of natural language [3]. This pre-training procedure consists of two stages, the first is called unsupervised pre-training and it involves training the model on unlabeled text data using a transformer-based language modeling approach [3]. The second stage is supervised pre-training and it involves refining the model on a smaller dataset of labeled data tailored for specific natural language processing tasks like text classification or question answering [3]. The primary objective of both stages is to enhance the model's ability to generate coherent and cohesive

responses to the natural language prompt. Furthermore, the pre-training process also utilized the Adam algorithm to optimize the model weights more effectively and reliably [3].

Fine-Tuning Phase

During the fine-tuning phase of ChatGPT's training process, the model's performance on specific NLP tasks is further optimized by training it on a smaller portion of labeled data. As explained in the paper “Universal Language Model Fine-tuning for Text Classification” by Jeremy Howard and Sebastian Ruder, this final step includes several critical processes, such as data preparation, architecture modification, and parameter optimization [5]. In the data preparation phase, the labeled data undergoes similar pre-processing steps as the unlabeled data, such as tokenization, subword encoding, and data cleaning [5]. Consequently, the model's architecture may be calibrated to better align with the task requirements, such as by replacing the final layer with a softmax layer for classification tasks [5]. Next, the model's parameters are optimized using the Adam algorithm to minimize the loss function specific to the task [5]. Throughout the fine-tuning process, the model is trained on a smaller dataset tailored to the specific natural language processing task, ensuring optimized performance while maintaining its ability to output contextually relevant responses to the human generated input.

Focusing on Academia

Fields that Have Been Impacted by ChatGPT

In the article “A Short History of ChatGPT: How We Got to Where We Are Today” by Bernard Marr, published in Forbes magazine, the writer highlights the specific domains significantly impacted by ChatGPT [9]. One of the most prominent is customer service, where

companies utilize OpenAI's API to address common questions and issues from their audiences [9]. Content creation is another notable area, with applications in journalism, writing, poetry, and music composition . The business sector has also embraced ChatGPT for task automation across various industries [9]. Healthcare has seen applications in medical recordkeeping and literature analysis, while the entertainment industry benefits from easier content creation [9]. Additionally, academia has integrated ChatGPT for personalized student assistance. This study will focus on analyzing the implications of ChatGPT in academia, exploring both its advantages and disadvantages.

ChatGPT in Computer Science Education

As mentioned earlier, especially in computer science education, ChatGPT can serve as a valuable tool for students seeking clarification on complex concepts or assistance with coding tasks. For example, students can pose questions about programming languages, algorithms, or data structures, and ChatGPT can provide explanations or even generate code snippets to illustrate solutions. This interactive and accessible mode of learning has the potential to supplement traditional instructional methods, offering students additional support and resources to enhance their understanding and mastery of computer science concepts.

Challenges and Opportunities

Pros and cons associated with the introduction of AI tools such as OpenAI's Codex, ChatGPT, and others in computer science undergraduate courses are discussed in an analysis conducted by Brett A. Becker and colleagues [1]. They explain that these large language models (LLMs) would be extremely useful in generating helpful and tailored learning materials, such as

practice questions for exams and code explanations and solutions. However, on the other hand, students may be tempted to use these tools unethically, for example, during open-book or take-home exams and assignments. Several studies have already been conducted to determine whether these deep learning models would be able to solve problems from programming exams and homework in computer science undergraduate courses, such as Intro to Computer Science and Object-Oriented Programming. All conclude that the models would be able to generate solutions for most questions, but the accuracy varies depending on the nature of the problem, its complexity, and how well the prompt message was composed [1]. In the following paragraphs, I will analyze my findings during research I conducted on how well I found the GPT-3.5 model to have solved homework assignments and problems from my past computer science exams at Yeshiva University Stern College for Women as well as other practice problems found online. The testing was conducted using the GPT-3.5 version, which, according to other research studies, does not perform much differently than GPT-4 [8]. In all cases, it is certain that although GPT-3.5 is able to generate feedback and improve the output, it will always be necessary for a professor or instructor to verify and review the information provided before sharing it with students.

The Research

The research focused on evaluating the abilities of LLMs to solve problems and assignments from undergraduate computer science courses, as well as logic and math problems and LeetCode challenges. The college assignments were taken from courses such as Introduction to Computer Science, Algorithms, Data Structures, Calculus I, Calculus II, Statistics and Probability, and more. The questions were presented in various formats, including multiple

choice, short and long answers, True/False, and coding-related questions. I selected these subjects because I believe they will thoroughly analyze all the capabilities of the GPT model, ensuring that I can conclude its strengths and weaknesses effectively.

Research Methodology

To conduct the study, I selected one prompt question at a time, inputted it into the free version of ChatGPT 3.5, and compared the generated answer with the correct solution. All coding questions were formulated and verified in the Python language, primarily tested using Pytests. I focused solely on assessing accuracy rather than code efficiency. In cases of code-based questions, I executed them individually and presented them to ChatGPT 3.5 (free version) as prompts, including any choices where applicable. Each response from ChatGPT was personally reviewed and categorized as either correct, incorrect, or nearly correct, in instances where additional feedback was necessary for the model to eventually produce the correct answer. For True or False and multiple-choice questions, responses were categorized simply as either correct or incorrect.

Results

SUBJECT	TYPE OF QUESTION	SOURCE
LeetCode	Coding-based	<u>Leetcode Patterns</u>
Logical Reasoning	Fit ins and multiple choice	<u>Inspired by Naukri Campus</u>
Computer Science	Design-based	Yeshiva University - Stern

Undergraduate Courses (e.g. Introduction to CS and Data Structures)	And coding-based	College for Women: COMP1300 and COMP1320 Assignments and Exams (Fall 2020 and Spring 2021) and other Online Resources
Applied Probability and Statistics	True/False, short answers, long answers, design-based and coding-based questions	UMGC: STAT 400 Assignments and Final (Fall 2023)

Table 1. Information regarding the data gathered to evaluate the accuracy of ChatGPT

SUBJECT	NUMBER OF QUESTIONS	CORRECT	INCORRECT	NEARLY CORRECT
LeetCode	100	81%	7%	12%
Logical Reasoning	65	69.23%	30.77%	0%
Computer Science Undergraduate Courses (e.g. Introduction to CS	100	66.8%	12.2%	21%

and Data Structures)				
Applied Probability and Statistics	30	52%	40%	8%

Table 2. Measuring the accuracy of ChatGPT in different subjects

The second table illustrated above represents the breakdown of ChatGPT's accuracy, indicating the percentages of questions answered correctly, incorrectly, or nearly correctly across various categories. The first table furnishes detailed information on question types and sources for each subject, including multiple choice, True/False, coding-based, etc. On average, ChatGPT achieved a 67.3% accuracy across all question types, suggesting its proficiency in answering computer science, logical reasoning, and mathematical questions may be somewhat limited. Nonetheless, after receiving sufficient feedback, ChatGPT demonstrated improvement in correctly answering a significant portion of the analyzed questions. Notable variations in ChatGPT's performance are shown across different subjects, particularly in the quantitative reasoning and problem-solving fields. For instance, the model exhibited greater efficacy in solving coding challenges similar to those found on LeetCode platforms, achieving an 82% accuracy rate. It also performed adequately in Data Structures, Algorithms, and Intro to Programming questions (66.8%), as well as in logical reasoning (69.23%). However, its performance worsened in probability and statistics problems, with only a 52% accuracy rate. Overall, ChatGPT excelled in design-based questions and popular coding challenges but

struggled with numerical questions. These insights are invaluable for guiding the integration of ChatGPT into academic settings for both professors and students.

Examples

In this section there is going to be an analysis on a few examples of answers that were generated by ChatGPT during the research process. These observations serve as a basis for formulating a set of recommendations that have the goal of guiding students and instructors in integrating ChatGPT into their academic workflows. Here are two examples of problems from an Introduction to Computer Science exam at Yeshiva University, Stern College for Women from Fall 2020.

Question A: Write a single **for** loop that prints the following exact output (using just a **for** loop containing a single **print** statement).

4.0

4.5

5.0

5.5

6.0

6.5

7.0

7.5

8.0

 ChatGPT

Certainly! Here's the code you can write in a file named "Question4.py" to achieve the desired output:

```
python Copy code  
  
# Question4.py  
  
for i in range(8):  
    print(4.0 + i * 0.5)
```

This loop iterates from 0 to 7 (inclusive) and prints the values 4.0 incremented by 0.5 for each iteration.



You

the code did not output the final 8.0

 ChatGPT

You're right, my apologies for that oversight. Let's ensure the loop prints the value 8.0 as well. Here's the corrected code:

```
python Copy code  
  
# Question4.py  
  
for i in range(9):  
    print(4.0 + i * 0.5)
```

Now the loop iterates from 0 to 8 (inclusive) and prints the desired output, including 8.0.



Question B: Write a single for loop that prints the following exact output (using just a for loop containing a single print statement).

0 1 0 1 0 1 0

ChatGPT's answer: *for i in range(7): print(i % 2, end='')*

Surprisingly, in question A, a correct answer was given only after the user gave its feedback and highlighted what the model misunderstood from the problem. For question B, the chatbot was able to give the correct answer on the first try. Most questions have proven that when the user prompts the AI tool without providing much context, ChatGPT's accuracy diminishes. More specifically, the model ends up focusing more on the incorrect aspects of the input. When the user provides more information to the chatbot, it is able to improve the result by focusing on the right aspect of the problem [12].

ChatGPT demonstrates inconsistencies in its performance, often providing incorrect answers to basic questions while occasionally offering well-framed responses to more complex problems. For instance, when presented with factual queries, such as those involving directed graphs and breadth-first search algorithms, ChatGPT frequently produces wrong answers. Despite this, it exhibits a degree of unpredictability, sometimes delivering extremely accurate explanations for multiple-choice questions but failing to select the correct option as its final output. For example, a basic algorithm question answered incorrectly is: "State True or False: Given a directed graph $G = (V, E)$, run breadth-first search from a vertex $s \in V$. While processing a vertex u , if some $v \in \text{Adj } u$ has already been processed, then G contains a directed cycle" from the research conducted by Joshi et al. [6].

Moreover, when prompted with a specific context, such as "preparing for technical interviews in computer science", ChatGPT's response accuracy notably improves. However, it gets most susceptible when the prompts are ambiguous or when subsequent prompts contradict previous answers, leading to misinterpretations and erroneous modifications to the output.

Additionally, ChatGPT's limitations become apparent in its inability to retain and integrate past context during conversations, often focusing solely on the most recent information provided. This behavior compromises the accuracy of its responses, as it may overlook relevant details discussed earlier. Moreover, when tasked with generating revised responses, ChatGPT consistently delivers incorrect answers accompanied by explanations that deviate significantly from the original line of reasoning. This might be due to the system's lack of comprehension of the underlying logic. Furthermore, ChatGPT's poor performance in mathematical problem-solving was related especially to its struggles with standard calculations. In the article “ChatGPT in the Classroom: An Analysis of Its Strengths and Weaknesses for Solving Undergraduate Computer Science Questions” the researchers bring an example of standard calculation that the GPT model failed at. The prompt stated “Compute the determinant of the matrix $A = \begin{bmatrix} 22/5 & -6/5 \\ -6/5 & 13/5 \end{bmatrix}$ ” and ChatGPT gave the answer “ $\det(A) = 322/25$ ”, instead of 10 [6]. Despite these drawbacks, ChatGPT's potential to provide valuable insights and assist with complex problem-solving questions encourages and motivates for further research and development to enhance its capabilities.

ChatGPT Solving Leetcode

LeetCode serves as a platform designed to aid programmers in improving their skills, broadening their knowledge base, and preparing themselves for technical interviews (LeetCode, 2023) [16]. The advent of ChatGPT has prompted researchers to investigate whether the model would be able to answer LeetCode questions and potentially navigate through a technical interview process at major companies. Researchers tested all levels of questions found on the platform (easy, medium, and hard) and discovered that ChatGPT is able to handle easy questions

smoothly [14]. They questioned whether the AI tool would be able to solve newer problems (after 2021) that may not have been part of the data the model trained on, and they found that for easy questions, there was very minimal difference in ChatGPT's ability to solve the programming problems. However, they encountered some interesting findings when they looked at medium and hard questions, in which ChatGPT performed poorly [14].

Biases

Researchers have found out that ChatGPT's performance in questions related to the Graduate Aptitude Test in Engineering (GATE), a national-level entrance exam in India conducted jointly by the Indian Institute of Science and seven Indian Institutes of Technology (IITs), is very poor compared to other problems from standardized exams. Given that GATE is specific to India, it could be concluded that ChatGPT's training data lacks sufficient representation of GATE-specific content. Therefore, researchers have suggested that to enhance ChatGPT's effectiveness across all regions, a more expansive and inclusive corpus of training data should be utilized in newer iterations of the model [6].

Surveys and Statistics

Generic Statistics on Students Using ChatGPT on Assignments

It goes unsaid that ChatGPT has already spread widely across all students. A survey that was conducted by Intelligent.com in February 2024 among 588 college students revealed that 37% of them currently actively use ChatGPT, while 36% claim they have previously used it but not currently for various purposes [19]. Among current users, 96% utilize ChatGPT for schoolwork, with 69% specifically using it for writing assignments [19]. Surprisingly, 29% of these students rely on ChatGPT to compose entire essays [19]. Despite its widespread use, 86%

of users claim that their ChatGPT usage goes undetected [19]. However, 3 in 4 students consider using ChatGPT as a form of cheating. Interestingly, the majority of users primarily utilize ChatGPT for writing assignments, followed by research, emails, and quizzes/tests. Among those using ChatGPT for writing, a significant portion relies on it consistently for idea generation, rewording sentences, and assistance with spelling and grammar. Overall, the survey highlights the prevalent use of ChatGPT among college students for academic purposes, particularly in writing assignments [19]. Another survey by Intelligent.com reveals the dilemma faced by educators regarding the integration of ChatGPT into their teaching schedule. While 46% of students report that their professors or schools have banned ChatGPT for homework, 29% indicate no such restrictions, and 26% remain uncertain. Despite these challenges, if carefully analyzed and implemented, ChatGPT has the potential to enhance the learning experience for college students since its popularity is already widespread among most universities.

Exploring ChatGPT Usage and Perceptions Among Stern College for Women Students

After conducting an anonymous survey on a random selection of students from Stern College for Women regarding ChatGPT, I was able to enhance the response to my research question. The most important results from the survey include that 100% of respondents have used ChatGPT for university work, with 50% trusting its accuracy while another 50% are neutral. Only 10% of the students consider ChatGPT to be usually inaccurate. Regarding proficiency, 40% believe ChatGPT is best at English, 10% in History/Philosophy, 20% in Marketing, 10% in Business Intelligence and Marketing Analytics, 40% in Computer Science, and 10% in Music (Fig. 2). Notably, 60% of students believe ChatGPT performs worst in math, which is supported by the previous research findings as well (Fig. 3). Moreover, 30% admit to using AI tools for assignments where prohibited, while only 10% used them during exams with

restrictions. Among those who used ChatGPT against regulations, 20% found it sufficiently accurate. Despite this, 50% credit ChatGPT with an overall increase in their GPA. Interestingly, 80% would recommend it to friends, and 80% consider its use ethical in assignments that do not mention anything about whether it is prohibited or not. Notably, 90% verify ChatGPT's accuracy post-use. This survey emphasizes ChatGPT's common use among students, encouraging professors to embrace and regulate its integration, ensuring it boosts students' critical thinking rather than allowing them to rely entirely on its prompt-based answers and not understanding the work behind it.

Which subjects do you feel like ChatGPT helps you the best in:

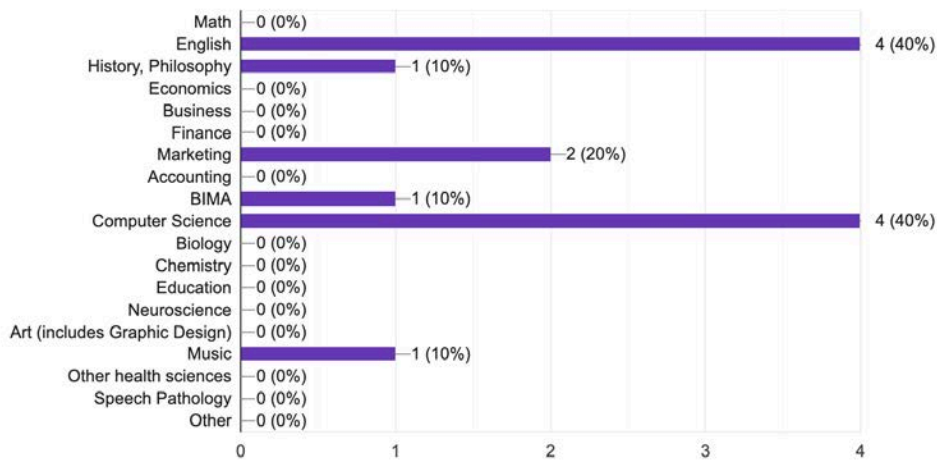


Figure 2. Responses from an Anonymous Survey Conducted on a Sample of Stern College for Women Students

Which subjects do you feel like ChatGPT helps you the best in:

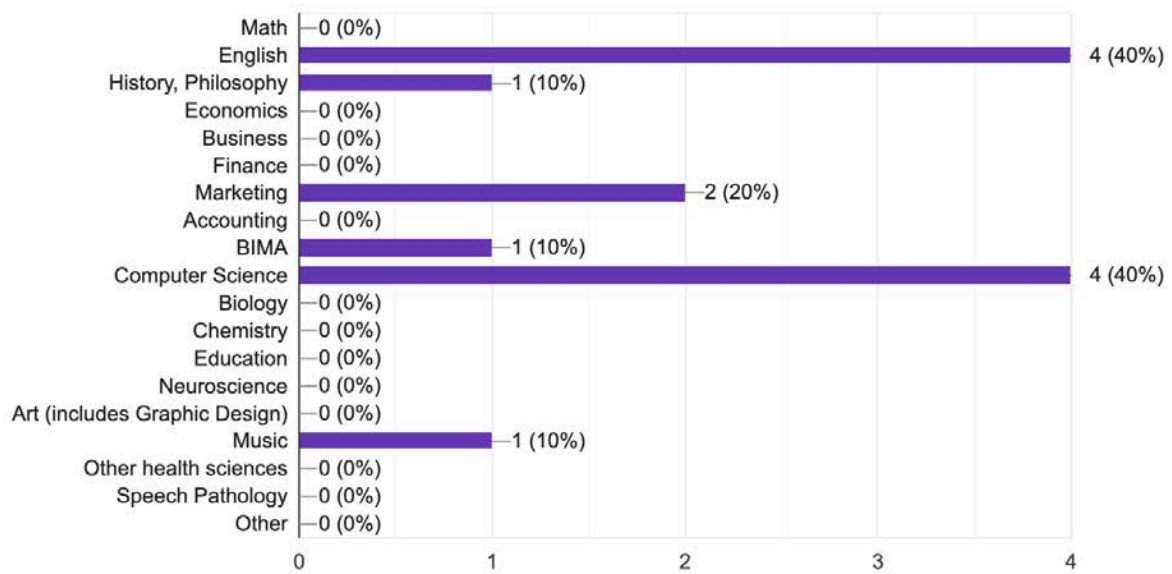


Figure 3. Responses from an Anonymous Survey Conducted on a Sample of Stern College for Women Students

Conclusion

Pros and Cons

In summarizing the pros and cons of ChatGPT, it is evident that the model exhibits variability in accuracy across different subjects, particularly in theoretical questions. Giving a more detailed context when querying ChatGPT's framework is a crucial factor influencing its accuracy. Providing specific prompts, such as framing questions similar to “within the context of technical interviews for computer science undergraduates”, substantially improves ChatGPT's

accuracy to an exceptional 92.8% [6]. Moreover, ChatGPT demonstrates higher accuracy in subjective and theoretical questions compared to objective ones, suggesting its potential as a reliable resource for such problems. However, ChatGPT also presents some drawbacks. It has a tendency to slow down and stop working too often, even in its latest iteration, GPT-4. Additionally, it has a limited vocabulary, which can restrict its utility in certain contexts. Therefore, it cannot be relied upon completely but rather should be used as a supplementary tool.

How Can ChatGPT Be Helpful in Academia

With the emergence of numerous AI-based platforms, students now have access to a wide range of tools that can enhance their learning experience and efficiency. Furthermore, ChatGPT's versatility extends beyond mere text generation. With the ability to understand and process natural language, ChatGPT can also engage in dialogue-based learning experiences, allowing students to interact with AI tutors or virtual assistants to receive personalized feedback and guidance. Through continuous interaction and feedback, ChatGPT can adapt and refine its responses, providing tailored assistance to individual learners based on their specific needs and preferences. Overall, from a computer science point of view, ChatGPT represents a powerful tool for augmenting learning experiences and facilitating knowledge acquisition in the digital age. By taking advantage of the benefits of AI and NLP, ChatGPT has the potential to revolutionize the way computer science concepts are taught and learned, empowering students with new opportunities for exploration, experimentation, and collaboration in their educational journey.

In reviewing the findings of this research, it becomes apparent that ChatGPT has the potential to improve productivity and efficiency in various sectors and more specifically in education. Its ability to generate code and explain complex problems makes it a valuable asset in

academic and programming domains. However, concerns regarding ethics and the possibility of job displacement have emerged with the widespread adoption of this technology. Addressing these concerns requires the establishment of regulatory frameworks by governments and other institutions to ensure responsible usage and avoid unwanted effects. Despite these challenges and potential risks, ChatGPT holds the promise of revolutionizing numerous fields, offering opportunities for advancement particularly in academia. While still in its early stages, properly regulated and understood, ChatGPT can serve as a useful tool for accelerating skill development among students. Through a combination of scholarly research and anonymous surveys, this study aimed to illuminate students' experiences with ChatGPT, providing precious insights for navigating the integration of AI technology in education.

References

- [1] Becker, Brett, et al. "Programming Is Hard -- Or at Least It Used to Be: Educational Opportunities And Challenges of AI Code Generation." 2022. Available: <https://doi.org/10.48550/arXiv.2212.01020>. [Accessed: May 2, 2024]
- [2] Copeland, B.J.. "Artificial Intelligence" *Encyclopedia Britannica*, 30 Apr. 2024. Available: <https://www.britannica.com/technology/artificial-intelligence>. [Accessed: May 2, 2024]
- [3] Devlin, Jacob, et al. "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Accessed: May 2, 2024]
- [4] Hines, Kristi. "History of Chatgpt: A Timeline of the Meteoric Rise of Generative AI Chatbots." *Search Engine Journal*, 5 June 2023. Available: www.searchenginejournal.com/history-of-chatgpt-timeline/488370/. [Accessed: May 2, 2024]
- [5] Howard, Jeremy, and Sebastian Ruder. "Universal Language Model Fine-tuning for Text Classification." [Accessed: May 2, 2024]
- [6] Ishika Joshi, Ritvik Budhiraja, Harshal Dev, Jahnvi Kadia, Mohammad Osama Ataulloh, Sayan Mitra, Harshal D. Akolekar, and Dhruv Kumar. "ChatGPT in the Classroom: An Analysis of Its Strengths and Weaknesses for Solving Undergraduate Computer Science Questions." Available: <https://doi.org/10.1145/3626252.3630803>. [Accessed: May 2, 2024]
- [7] Koubaa, Anis. "GPT-4 vs. GPT-3.5: A Concise Showdown". Available: <https://doi.org/10.20944/preprints202303.0422.v1>. [Accessed: May 2, 2024]

- [8] Loubier, Michael. "ChatGPT: A Good Computer Engineering Student?: An Experiment on Its Ability to Answer Programming Questions from Exams." (2023). Available: [Accessed: May 2, 2024]
- [9] Marr, Bernard. "A Short History of Chatgpt: How We Got to Where We Are Today." *Forbes, Forbes Magazine*, 20 Feb. 2024, Available: www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=46b726da674f. [Accessed: May 2, 2024]
- [10] Olaoye, Godwin. "Deep Learning Approaches for Natural Language Processing: Advancements and Challenges". [Accessed: May 2, 2024]
- [11] Radford, Alec, et al. "Language Models Are Unsupervised Multitask Learners". 2019, Available: <https://api.semanticscholar.org/CorpusID:160025533>. [Accessed: May 2, 2024]
- [12] Reeves, B., Sarsa, S., Prather, J., Denny, P., Becker, B. A., Hellas, A., Kimmel, B., Powell, G., and Leinonen, J. "Evaluating the Performance of Code Generation Models for Solving Parsons Problems with Small Prompt Variations." [Accessed: May 2, 2024].
- [13] Roumeliotis, Konstantinos I., and Nikolaos D. Tselikas. "ChatGPT and Open-AI Models: A Preliminary Review". *Future Internet*, vol. 15, no. 6, 2023. Available: <https://doi.org/10.3390/fi15060192>. [Accessed: May 2, 2024].
- [14] Tian, H. et al. "Is ChatGPT the Ultimate Programming Assistant -- How far is it?" arXiv, Apr. 24, 2023. Available: <http://arxiv.org/abs/2304.11938>. [Accessed: May 2, 2024]

[15] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. “Attention is All You Need.” [Accessed: May 2, 2024]

[16] “LeetCode - The World’s Leading Online Programming Learning Platform.” Available: <https://leetcode.com/>. [Accessed: May 2, 2024]

[17] “OpenAI API.” Available: <https://platform.openai.com>. [Accessed: May 2, 2024]

[18] *Statista*, <https://www.statista.com/statistics/1398211/ai-keyword-traffic-volume/>.

[19] “4 in 10 College Students Are Using CHATGPT on Assignments.” *Intelligent*, 11 Mar.

2024. Available:

www.intelligent.com/4-in-10-college-students-are-using-chatgpt-on-assignments/#:~:text=Since%20the%20release%20of%20ChatGPT,found%2037%25%20currently%20use%20ChatGPT.

[Accessed: May 2, 2024].