Southwestern Business Administration Journal

Volume 21 | Issue 1 Article 2

2024

ChatGPT: What is it? How does it work? Can it be a teaching tool for an introductory programming course in higher education?

Dale Fontenot
Lamar University

Feroz Ahmed
Lamar University

Kaniz S. Chy
Lamar University

ChatGPT: What is it? How does it work? Can it be a teaching tool for an introductory programming course in higher education?

Dale Fontenot (Principal)
Feroz Ahmed
Kaniz S Chy
Lamar University

ChatGPT: What is it? How does it work? Can it be a teaching tool for an introductory programming course in higher education?

Dale Fontenot (Principal)
Feroz Ahmed
Kaniz S Chy
Lamar University
dale.fontenot@lamar.edu

Keywords: AI, ChatGPT, Coding, Education, Programming Course

Abstract:
ChatGPT is an interactive chatbot that uses natural language processing (NLP) to understand user input and generate responses. (Ruby, 2023) OpenAI.com released it in November of 2022, creating a buzz in many areas, including higher education. The focus of higher education on this new tool has primarily been academic dishonesty. Although this is a significant concern with ChatGPT, this paper focuses on its usage as a learning tool to help teach programming to management information systems students.

The paper's brief literature review focuses on ChatGPT's background and history. A study of how ChatGPT works at a high level is conducted, along with a walk-through of the process it uses to perform conversational modeling. (Ruby, 2023) Large language modeling is explained in terms that can be easily understood by management information systems students. After reviewing what ChatGPT is and how it works, the paper outlines and conducts an experiment using ChatGPT as a tool to assist students new to programming, in completing a simple programming assignment.

The experiment aims to see if a new programming student can use ChatGPT's responses to complete the assignment. The experiment uses conversational prompts to describe the steps in the coding assignment and has ChatGPT return responses in the form of Python code blocks.

The paper describes each step in the process, including the prompts written and responses received from ChatGPT. This experimental assignment is to create a console application that will take input from the end user in the form of the user's first name, last name, and year born. The input feature of the program must validate each piece of end-user data. This validation will be performed in conditional while loops. The program will then write the end-user data into a tuple that will also be stored in a list. An outer conditional while loop should terminate when the index

reaches five (5) total tuples with individual end-user data in the tuples. Once all the data is entered, the program will process the data into unique usernames and evaluate the usernames for duplicates.

The preliminary results of this experiment will determine whether ChatGPT can be successfully used as a learning tool in assisting students in completing programming assignments. The prompts written for and the responses received from ChatGPT will be evaluated as to their usefulness and repeatability.

Literature Review

AI text generators in academia have caused a stir among professors. Some are worried that students are using the tools for completing academic assignments, while others are open-minded and view it as a tool that can accelerate the learning process. Anthony Picciano, an education professor at Hunter College, is among those less concerned with AI technology, believing that if it helps students formulate ideas, it could be a valuable tool in the classroom (Aumann, A. 2023). However, ChatGPT may extend beyond what is expected. Xiaoming Zhai's paper "ChatGPT User Experience: Implications for Education" emphasizes the need to recognize both the advantages and challenges of ChatGPT and to identify the skill sets that will continue to be essential for students to succeed in the future. In the paper, a study was conducted to evaluate users' experience of ChatGPT. The results showed that participants found it user-friendly and capable of generating human-like responses but expressed concerns about its reliability and accuracy (Zhai, 2022).

Beth McMurtrie suggests that despite these concerns, ChatGPT and similar tools are powerful teaching aids. She considers them essential but understands that students must learn to use them aptly while considering the tools' limitations and benefits (McMurtrie, B. 2023).

Overall, the literature suggests that ChatGPT can be an effective tool in an introductory programming course in higher education, particularly for personalized feedback and guidance and for generating code examples. However, it is crucial to exercise caution in using ChatGPT as a standalone teaching tool and to be aware of its limitations.

Introduction

ChatGPT is an AI chatbot model based on natural language processing that can generate text to respond to human questions. It can also generate workable programming code based on the user's prompt. It can understand human language and respond to questions and prompts. Regarding its capabilities, ChatGPT can be

applied to various tasks, such as chatbot services, automated customer assistance, language translation, and more.

Natural language processing, or NLP, is an area of AI that enables computer systems to understand human speech in its natural setting. The communication gap between coded machines and complicated human language is effortlessly closed by NLP. NLP can be used in the environment of bots to determine what the user is saying or asking. Imagine if chatbots respond like humans and crack jokes; developers are trying to implement this feature for the chatbot, and that is only possible with NLP development (Devlin et al., 2019).

The earliest language models were based on n-gram models, representing a word's probability given the previous n-1 words. These models were widely used in speech recognition and machine translation in the 1980s and 1990s (Jelinek, 1990). In the early 2000s, neural network-based language models were developed, which were shown to outperform n-gram models on various tasks (Bengio et al., 2003). There has been a rapid growth in the size and complexity of language models, culminating in the recent development of models such as GPT-3, which has over 175 billion parameters (Brown et al., 2020). In 2015, Elon Musk, Sam Altman, Peter Thiel, and Reid Hoffman founded OpenAI, a nonprofit organization focused on developing artificial intelligence to benefit humanity. The Gym, which was launched in 2016, allowed researchers to develop and compare reinforcement learning systems (Kay, 2023). On March 23, 2016, Microsoft released Tay, an AI Chatbot designed to engage people in dialogue through tweets or direct messages followed by Zo, released in December 2016 on the Kik messenger app.

Large language models require massive amounts of training data, which can be challenging to obtain for some languages and domains. Training and using large language models require significant computational resources, including specialized hardware such as graphics processing units (GPUs) and tensor processing units (TPUs). Large language models are prone to overfitting the training data, making generalizing new data or tasks challenging. Large language models can be used maliciously, such as generating fake news or deep fakes. Training large language models requires significant energy, which can negatively impact the environment.

How does ChatGPT work?

This section of the paper will review, at a very high level, how ChatGPT works, starting with a description provided by ChatGPT.

According to ChatGPT,

> "ChatGPT is a natural language processing (NLP) model developed by OpenAI. It is based on the GPT-3 model, which was trained on a large corpus of text data. The model is designed to generate human-like responses to user inputs in natural language.
>
> The ChatGPT model takes user input and encodes it into a vector representation. This vector representation is then fed into the GPT-3 model, which uses its large corpus of text data to generate a response. The response is then decoded back into natural language and returned to the user.
>
> The ChatGPT model has been trained on various conversational topics, including sports, politics, music, movies, and more. This allows it to generate more accurate responses tailored to the user's interests and context. Additionally, the model can be further customized with additional training data for specific use cases or domains.
>
> ChatGPT also includes an interactive mode that allows users to interact with the model in real-time. In this mode, users can provide feedback on the generated responses and help train the model further. This helps ensure that the responses are accurate and relevant to the conversation. ChatGPT is powered by OpenAI's GPT-3 language model, which is trained on a large dataset of web text. This dataset includes webpages, books, and other sources of text from the internet."

ChatGPT stands for generative pre-trained transformer. The generative in GPT means it can generate new text based on the input it receives, while pre-training is the process of training the model on a large corpus of text data before fine-tuning it for specific downstream tasks. Finally, the transformer is a type of neural network architecture used for language modeling and natural language processing tasks.

Generative Pre-training Transformers, or GPT models, were first launched in 2018 by OpenAI, with version one being GPT-1. From that release, the models continued to evolve quickly, with GPT-2 being released in 2019 and GPT-3 in 2020. In 2022, the most recent version was released as InstructGPT and ChatGPT. With this release in 2022, human feedback was integrated into the system using supervised learning. This fast advancement in GPT model evolution has been driven by increased computational efficiency, cheaper processing power, and the plethora of available content on the internet. All these events helped

create GPT-3 and allowed it to be trained on significantly more data than GPT-2. This allowed GPT-3 (and now GPT-3.5 and GPT 4) to create a more diverse knowledge base capable of performing a more comprehensive range of tasks. (Ruby, 2023).

ChatGPT is a large language model based on natural language processing and deep learning modeling. It uses transformer techniques based on neural network technologies to generate human-like text. ChatGPT was designed and fine-tuned to support conversational modeling and can generate responses to human prompts that are both coherent and relevant. (Meyer, 2023) Large language models like ChatGPT are based on a type of machine learning known as natural language processing. Large language models are designed to process vast amounts of text data and then infer some relationship among the words. These language models have grown in popularity in recent years due to the increase in computing power, which allows the model to increase its capacity by increasing the size of its input datasets and parameter space. (Ruby, 2023).

Despite the increased computing power and the massive increase in text data due to the internet, challenges in developing large language models remained in building conversational AI systems. As mentioned earlier, the ability to generate relevant and coherent responses to human prompts was a crucial challenge in this process. The ability to generate responses that are on topic and make sense to the user in the context of the conversation is crucial. This was a struggle in traditional language models, but ChatGPT can generate responses that can contribute to the conversation meaningfully. It does this due to its specialized training and fine-tuning with a massive dataset of conversational text. Using this large dataset in training, ChatGPT has learned the patterns and nuances of natural conversation while generating relevant and coherent responses. (Kelk, 2022).

In traditional language models, the training typically involves predicting the next word in a sequence of words. This is typically done using a Long Short Term Memory (LSTM) model that fills in the blanks with the most statistically probable word given the context of the prompt. This is a sequential type modeling structure, and because of that, it has two significant limitations. The first limitation is that the model has difficulty ranking some words over others. The second limitation is that it reads in the prompt as individual words in sequential order and not as a whole corpus. This causes the LSTM to be trained with a fixed context that extends beyond the original prompt by several steps. According to Ruby (2023), this causes limitations in the complexity of the relationships between words and the meanings that can be derived.

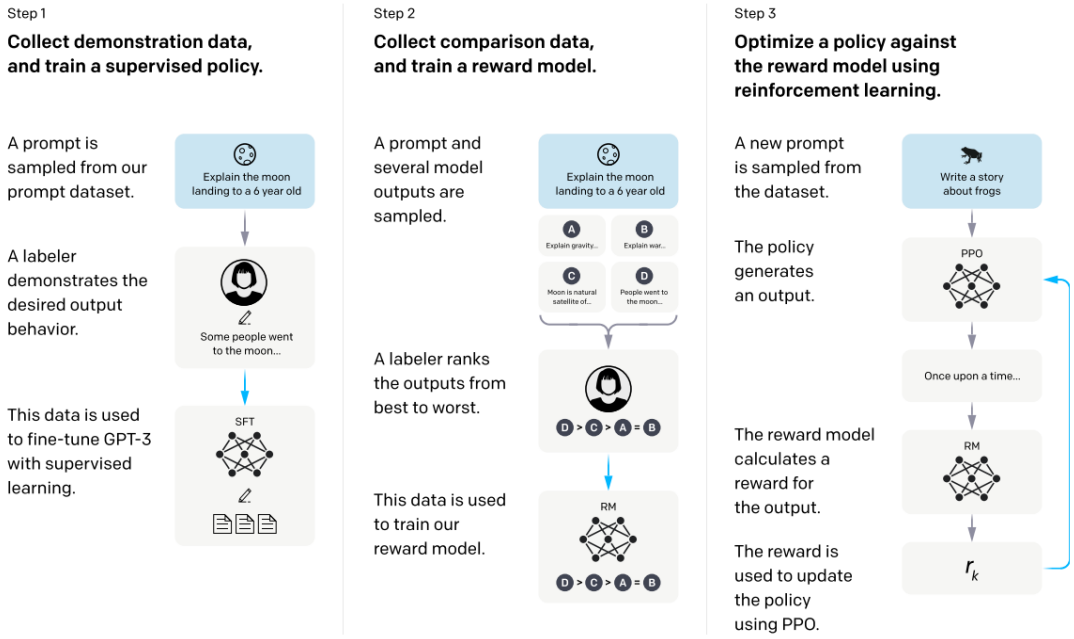
To overcome these limitations, in 2017, Google Brain introduced transformers. These transformers could process a complete prompt as input simultaneously and not sequentially. Transformer technology is used in newer large language models like GPT-3 and up. It is an algorithm used in training neural networks and allows substantial portions of the computations to be processed in parallel. This increases the training speeds of large language models significantly. These increases in processing speed have allowed pre-trained models to process larger and larger datasets with higher performance and more human-like responses. The transformer has significantly impacted natural language processing, but as the model grows ever larger, some argue that these models might become monopolized. (Luitse & Denkena, 2021) In addition to the use of transformer technology, ChatGPT also uses a technique known as self-attention which led to the multi-head attention model.

According to Luitse & Denkena (2021), the transformer model relies on an attention mechanism to compute the representation of ChatGPT's inputs (prompts) and outputs (responses). The standard attention-based architecture allows the model to focus on specific parts of the prompt when processing the text prediction task. However, instead of trying to encode the entire prompt into a fixed-length vector, it encodes the prompt into a sequence of memory vectors. This allows the training of the model by optimizing the information contained in these memory vectors, allowing the mode to focus on only the vectors that contain relevant information needed for the prediction when a prediction is being made. Attention allows a way of determining which parts of the prompt are essential and which are less critical, allowing for pique performance when processing longer sequences from a prompt. The transformer allows for calculating these attention vectors for more than a single word at a time in a process called multi-head attention. Using a multi-head attention module, encoding can be done in parallel processing, acting on all the words in a prompt at one time while the decoding is still done sequentially, generating the final response one word at a time. Larger datasets can be used as the transformer uses more parallel processing in training. In addition, the more parallel processing is used in training, the more the processing load can be spread out over larger clusters of processing units shortening the training time.

The transformer uses a method of encoding and decoding to process the input prompt and returns the output response. It uses layers of nodes to analyze the input to generate the response one word at a time, with each word being determined by the neural network analysis of previous words in the sequence. The process of generating a response is a two-step process. The first step is an encoding which takes the input and represents it in a numerical method that the neural network can understand. This is the process of tokenization which breaks

down each word into a token numerical representation to feed the neural network in the input process and then outputs the probability of what the next word will be in the segment. This token is then used to determine the following word in the segment and then repeats until it gets the highest probability for the correct word. The second step is decoding, which takes neural network output converting it back into human text that it returns as a response.

In addition to using transformer technology, ChatGPT was trained using a concept called Reinforced Learning with Human Feedback (RLHF). This involved three steps in the supervised learning process. The first step used supervised fine-tuning (SFT), in which 40 contractors created a supervised training data set in which the inputs (prompts) lead to known and established outputs (responses). The second step implemented a reward model where the responses generated in step 1 are scored, and the output is a scalar value that the model attempts to maximize. In the final step, the model is given a random prompt and returns a response generated using a policy created in step 2. This policy is based on what the machine learned in step 2 to maximize the reward it received. Based on the prompt/response pairing, that reward is then fed back into the model to help evolve this policy. (Meyer, 2023) Below is a visual representation of this process.



(Ouyang et al.,2022)

The experiment in using ChatGPT to teach programming.
We teach an introductory programming class to students in our Master of management information systems degree. The projects that are used in the course are customed designed by the instructor and are not standard programming projects such as Tic Tac Toe or Hangman. We use real-world examples such as creating usernames and passwords with data structures, input functions, and text files to input data for the programs. For this experiment, we copied our python project three's instructions directly from Blackboard with minor editing and pasted them into the ChatGPT prompt.

The expected results of the experiment were an incomplete program that would not satisfy the assignment requirements. This hypothesis was based on the fact that the program had multiple requirements to complete. The response for ChatGPT returned in a manner of seconds. The code was an almost workable solution. Although missing some of the required output, the program ran without bugs. It is worth noting that this prompt came after several other attempts of python coding prompts, so we may have "primed the pump" before giving ChatGPT the project three instructions. ChatGPT does appear to build on previous prompts, so if a student ran this experiment with no prior prompts and responses, we are unsure if the results would be the same.

Although the code was workable and satisfied most of the assignment's requirements, it should be noted that we wanted our code with "while" condition that used a "len" function to test for the number of employees entered. The ChatGPT program used a "for" loop with an "in range" function instead. We also wanted the username to be all lower case, but the ChatGPT usernames did not result in that format. In addition to creating a program with the correct code and output, we also needed to have commented code, which is a requirement for a good grade. The first program ChatGPT created was missing the comments because we forgot to tell it that we desired them. We then copied the ChatGPT code back into a prompt and asked it to comment the code, which it did without an issue.

The result of the experiment was that ChatGPT could, in fact, code the complete assignment in a manner that would have given the student an "A" grade. Because of this, we decided that we would need to change our teaching strategy to include the use of ChatGPT.

Recommendations
We are open to allowing the use of ChatGPT in the following situations:

- Use as a Code Assistant.

- Use to create simple lessons for basic logic.
- Use to assist in debugging code.
- Use creative scenarios and let students create programs by any method (beating plagiarism).
- Set the assignment parameters, such as using a while loop for data entry and validation.
- Determine all the required logic and data structures needed to make a good grade but leave the overall coding strategy to the student.

Conclusion

ChatGPT can be used to complete more than simple coding assignments. Plagiarism will be a significant unintended consequence of ChatGPT. Its unwarranted use by programming students might be thwarted by creating very complex coding assignments but will result in out-of-control grading. A better alternative might be "if you can't beat them, then join them" approach.

This could be done by creating assignments outlining the coding methods, functions, or attributes that must be used in the assignment but letting the students use any means they choose to code it. ChatGPT could also be used as a debugger, helping the students debug their written code. The focus on programming in a world with ChatGPT type of AI will need to shift from the mechanics of programming to the creative programming process. As such, we will be shifting our focus as programming instructors from teaching syntax to teaching critical thinking, allowing the programming student to design and plan software responses to business problems. They will then create those solutions by getting ChatGPT to do the rote coding work while they focus on implementing the solutions.

References:

Aumann, A. (2023,January 17). Professors Are Scrambling as Students Turn in AI. New York Observer, The (NY), 1.

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. Journal of Machine Learning Research, 3(Feb), 1137-1155.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for

Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171-4186.

Jelinek, F. (1990). Self-organizing statistical models for speech recognition. Readings in speech recognition, 623-659.

Kay, G. (2023, Feb 1). The history of ChatGPT creator OpenAI, which Elon Musk helped found before parting ways and criticizing. Retrieved from https://www.businessinsider.com/history-of-openai-company-chatgpt-elon-musk-founded-2022-12

Kelk, I. (2023, January 11). How does chatgpt work? Clarifai. Retrieved February 6, 2023, from https://www.clarifai.com/blog/how-does-chatgpt-work

Luitse, D., & Denkena, W. (2021). The Great Transformer: Examining the role of large language models in the political economy of ai. Big Data & Society, 8(2), 205395172110477. https://doi.org/10.1177/20539517211047734

McMurtrie, B. (2023). Teaching: Will ChatGPT Change the Way You Teach? The Chronicle of Higher Education, retrieved March 27, 2023 from https://www.chronicle.com/newsletter/teaching/2023-01-05?cid=gen_sign_in

Meyer, P. (2023, March 7). ChatGPT: How Does It Work Internally? - Towards AI. Medium. https://pub.towardsai.net/chatgpt-how-does-it-work-internally-e0b3e23601a1

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.

Ruby, M. (2023, January 31). How CHATGPT works: The models behind the bot. Medium. Retrieved February 6, 2023, from https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286

Zhai, Xiaoming, ChatGPT User Experience: Implications for Education (December 27, 2022). Available at SSRN: https://ssrn.com/abstract=4312418 or http://dx.doi.org/10.2139/ssrn.4312418