IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

TESTING OF CHATGPT- IF IT IS REALLY CAPABLE TO REPLACE DEVELOPERS

[1]Shubham Aditya, [2]Priyank Dhabaliya, [3]Siddharth Goel, [4]Gautam Mandal, [5]Omprakash Dewangan
[1,2,3,4] Student of Btech (CS), [5]Assistant Professor
[1,2,3,4,5]Department of Computer Science,
[1,2,3,4,5] Kalinga University, Raipur, Chhattisgarh, India (492101)

ABSTRACT **:** An AI program called ChatGPT has been creating a lot of buzzes lately . ChatGPT, a cutting-edge language model, developed by OpenAI, has been trained using massive amounts of data, and simulates human conversation by generating appropriate responses. It has garnered significant attention due to its ability to effectively answer the broad range of inquiries, fluent and also surpassing all the other AI chatbot in both security and usefulness. The best thing about ChatGPT is, it can fix program bugs, can give coding suggestions, and even write bug-free code lines. That's Why concern has been raised that ChatGPT and other AI models may end up replacing human programmers and developers in the near future. So the focus of this research paper is to test the ability of ChatGPT in different categories like Arithmetic, coding, reasoning, and debugging categories and based on this test we will conclude whether It replaces programmer or not.

Keywords: ChatGPT, debugging, coding, reasoning, developer

I. I. INTRODUCTION

ChatGPT[1], is a highly capable Natural language processing(NLP) system developed by OpenAI and launched in November 2022. ChatGPT is built on top of openAI's GPT3.5, and GPT-4 families of Large Language Models (LLM) and has been finetuned (an approach of transfer learning) using both supervised and reinforcement learning techniques. If we want to define ChatGPT simply: It's a language model that uses machine learning to create a human-like answer and reply to user inquiries- that's all. ChatGPT could be considered like an advanced chatbot built for performing a variety of tasks, from answering the user's curiosity to helping with research paper writing and solving the bug of code.

You can guess how powerful ChatGPT is by the fact that it passed many University level advanced exams including MBA and a Medical exam. According to the Business Insider news website, ChatGPT passed the Wharton MBA exam, US medical licensing exam, the Microbiology quiz, the Law exam, and many more[2]. The newest version of OpenAI's language model system, GPT-4, released on March 13, 2023, took its ability to new heights as compared to an older version of ChatGPT (GPT-3.5). One twitter user tweeted that GPT-4 passed basically every exam and its scores are more impressive as compared to GPT-3.5. 90 percentile in Bar exam, 88 percentile in LSAT [3].

ChatGPT can also write code, can solve leet code problems and even can find bugs in your code. Researchers from Johannes Gutenberg University and University college London gave 40 pieces of the buggy code to ChatGPT and it fixed 31 out of 40 bugs[4]. This research can give an idea about ChatGPT's programming ability.

No doubt ChatGPT is a groundbreaking development in the field of LLM and AI. But this creates a little bit of worry in everyone's mind about their job securities especially in Programming and software development jobs. After the release of ChatGPT every tech influencer is in a rush to create the video, articles like: “How it takes our job”, “programmers will disappear in a few years” and many more. It creates more fear in people’s minds. So it’s very important to remove this fear of future use

In this research paper, we tried to exploit the ability and limitations of ChatGPT by asking various questions based on three categories: Arithmetic, Reasoning and logic (both combined) and coding. This subject are very essential for every developer and programmer’s software journey. This Research paper is inspired by this Github repository[5]. In this repository, they created a list of ChatGPT and Bing search engine failures. Some research papers have already been published on ChatGPT mistakes, but they are only concerned with the mathematical ability of ChatGPT[6][7].

In addition, we have also discussed the recent ban of ChatGPT's answers on StackOverflow and how companies are not taking the job security dilemma of people.

II. TESTING CHATGPT’S ABILITY

This section shows the performance of ChatGPT’s in tests, which are categorized into three areas. It is important to note that we selected only a few answers of ChatGPT (mostly failure cases) to find the limitation of ChatGPT. It is possible that some of the answers and failures may not exist in future as newer versions of improved ChatGPT will come.

2.1 Reasoning

There are a number of Reasoning types Arithmetic Reasoning, logical reasoning, multimodal reasoning and symbolic reasoning. But this section focuses on forms of reasoning which require real-world knowledge and exclude mathematical and logical reasoning.

Decision making, real-time problem-solving skills are important activities that heavily depend on the fundamental aspect of human intelligence known as Reasoning.

USER: Shubham, Aditya and Pravesh went for a walk. They saw a tree, cow and dog. Shubham also saw a cat and pointed it out to Aditya. Pravesh went to pet the cat. who saw the cat last?

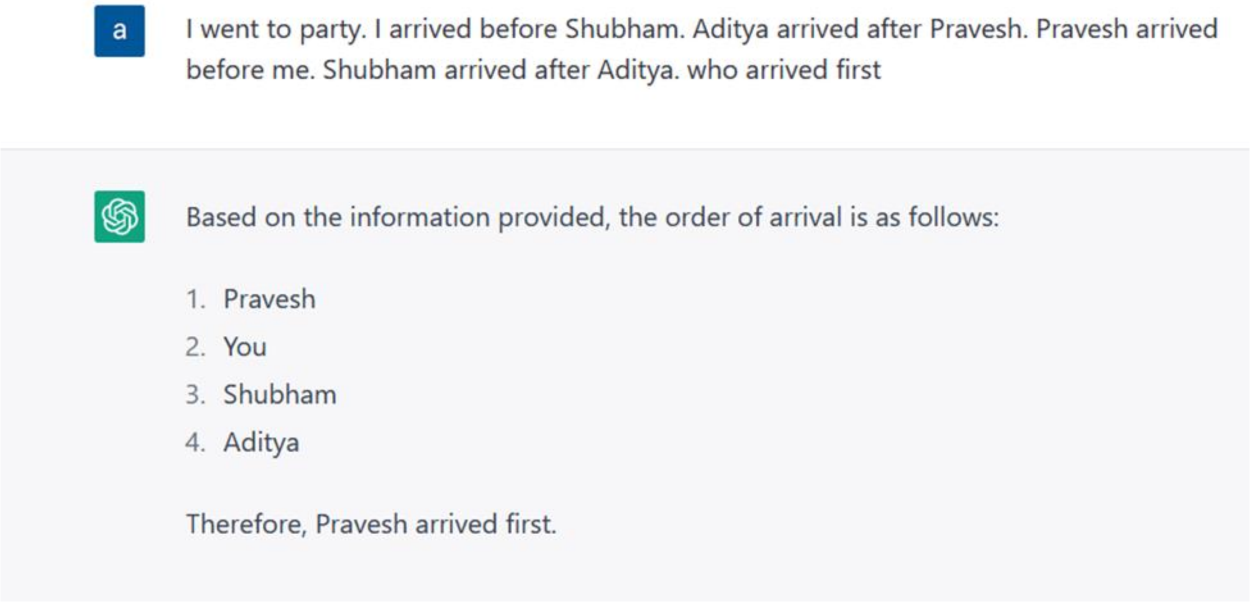
ChatGPT: Pravesh saw the cat last because he went to pet it.

USER: How can you so sure about your answer? It can also be possible that Pravesh never saw the cat.

ChatGPT: You're right, my previous answer made an assumption that may not be true. Based on the given information, we cannot be certain who saw the cat last. It is possible that Aditya also saw the cat after Shubham pointed it out, but did not mention it or interact with it. Alternatively, it is possible that Pravesh did not see the cat, even though Shubham pointed it out, if Pravesh was distracted or not paying attention.

So, the answer to who saw the cat last cannot be determined with complete certainty based on the given information.

The above question is based on temporal reasoning. **Temporal Reasoning** is the ability to make predictions about the event and their order in time. It involves understanding the relationship between the course of events. ChatGPT correctly deduces the sequence of event in the first reply. But when I tried to confuse ChatGPT on the second inquiry, ChatGPT replied with a statical-based answer where it gives preference to uncertainty over assumption.



I went to party. I arrived before Shubham. Aditya arrived after Pravesh. Pravesh arrived before me. Shubham arrived after Aditya. who arrived first

Based on the information provided, the order of arrival is as follows:

1. Pravesh
2. You
3. Shubham
4. Aditya

Therefore, Pravesh arrived first.

fig: 1 Temporal reasoning problem

Fig 1 shows another mistake of ChatGPT on the reasoning problem. ChatGPT correctly gives the answer. Pravesh was the person who arrived first. But if you look at the order of arrival, it is wrong. There is a clear mention on a question that Shubham arrived after Aditya but ChatGPT put Shubham before Aditya. The correct order would be Pravesh, You, Adiya, Shubham.

USER: You have Rs 100. You can buy 1 goat for Rs 1, 1 buffalo for Rs 10 and 8 hens for Rs 1. How can you buy 100 animals for Rs 100?

CHAGPT: To buy 100 animals for Rs 100, we need to find a combination of goats, buffaloes, and hens that adds up to 100 animals and costs exactly Rs 100.

Let's start with the most expensive animal, which is the buffalo. We can buy a maximum of 10 buffaloes for Rs 100. This leaves us with Rs 90.

Next, we can buy a maximum of 90 goats for Rs 90. This leaves us with Rs 0.

We have not yet bought any hens, so let's calculate how many hens we can buy for Rs 10. Since 8 hens cost Rs 1, we can buy 80 hens for Rs 10.Therefore, we can buy a total of 10 buffaloes, 90 goats, and 80 hens for Rs 100, which adds up to 180 animals.

The problem presented above is based on Arithmetic Reasoning. We explicitly stated that we wanted 100 animals for 100 rupees, but ChatGPT responded incorrectly and purchased 180 animals for 100 rupees. So we asked it again to modify its response.

USER: But I want only 100 animals, not 180 animals?

ChatGPT: I apologize for the confusion in my previous response. Here is an updated solution to buy exactly 100 animals for Rs 100.

I apologize for the confusion in my previous response. Here is an updated solution to buy exactly 100 animals for Rs 100.

Let's start by buying as many goats as possible since they are the cheapest. We can buy 100 goats for Rs 100, which leaves us with no money left.

However, we still need to buy more animals to reach a total of 100. Since we cannot buy any more goats, we can buy hens instead. We know that 8 hens cost Rs 1, so we can buy a total of 800 hens for Rs 100.

In total, we can buy 100 goats and 800 hens for Rs 100, which adds up to 900 animals. Therefore, to buy exactly 100 animals for Rs 100, we can buy 100 goats and no other animals.

ChatGPT again responded incorrectly and purchased 100 goats in 100 rupees and left the other 2 animals. So we again asked ChatGPT that we want all types of animals.

USER: But i want all type of animal

CHATGPT: I apologize for the confusion. Here is an updated solution to buy exactly 100 animals, including goats, buffaloes, and hens, for Rs 100:

Let's start by buying as many goats as possible since they are the cheapest. We can buy 10 goats for 10, which leaves us with Rs 90.

Next, we can buy one buffalo for Rs 10, which leaves us with Rs 80.

We still need to buy more animals to reach a total of 100. Since we cannot buy any more buffaloes or goats, we can buy hens instead. We know that 8 hens cost Rs 1, so we can buy a totalof 800 hens for Rs 100.

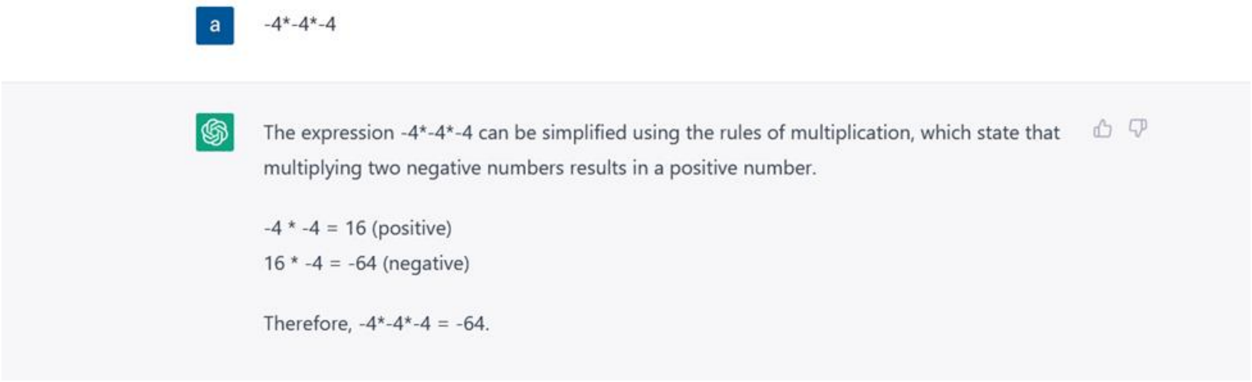
In total, we can buy 10 goats, one buffalo, and 89 sets of 8 hens (712 hens in total) for Rs 100, which adds up to exactly 100 animals.

Therefore, to buy exactly 100 animals, we can buy 10 goats, one buffalo, and 89 sets of 8 hens for Rs 100.

ChatGPT worked extremely hard to get the correct answer. However, it failed miserably. It purchased 89 sets of eight hens, for a total of 712 hens. That is, it purchased 723 animals rather than 100. The correct answer to the above Arithmetic reasoning question is 72 hens for 9 Rs, 7 buffalo for 70 Rs, and 21 goats for 21 Rs, for a total of 100 animals for 100 rupees

2.2 Arithmetic

In this section, we will ask some mathematical questions to determine how good ChatGPT is, in mathematics. When the prototype of ChatGPT was released in Nov, 2022, it had very weak mathematical capabilities, they even gave wrong answers to simple mathematics questions. Even ChatGPT failed miserably in Singapore’s 6th grade tests, averaging 16 percent in maths [10]. How ever days later, it was getting right answer of previous question of test.



-4*-4*-4

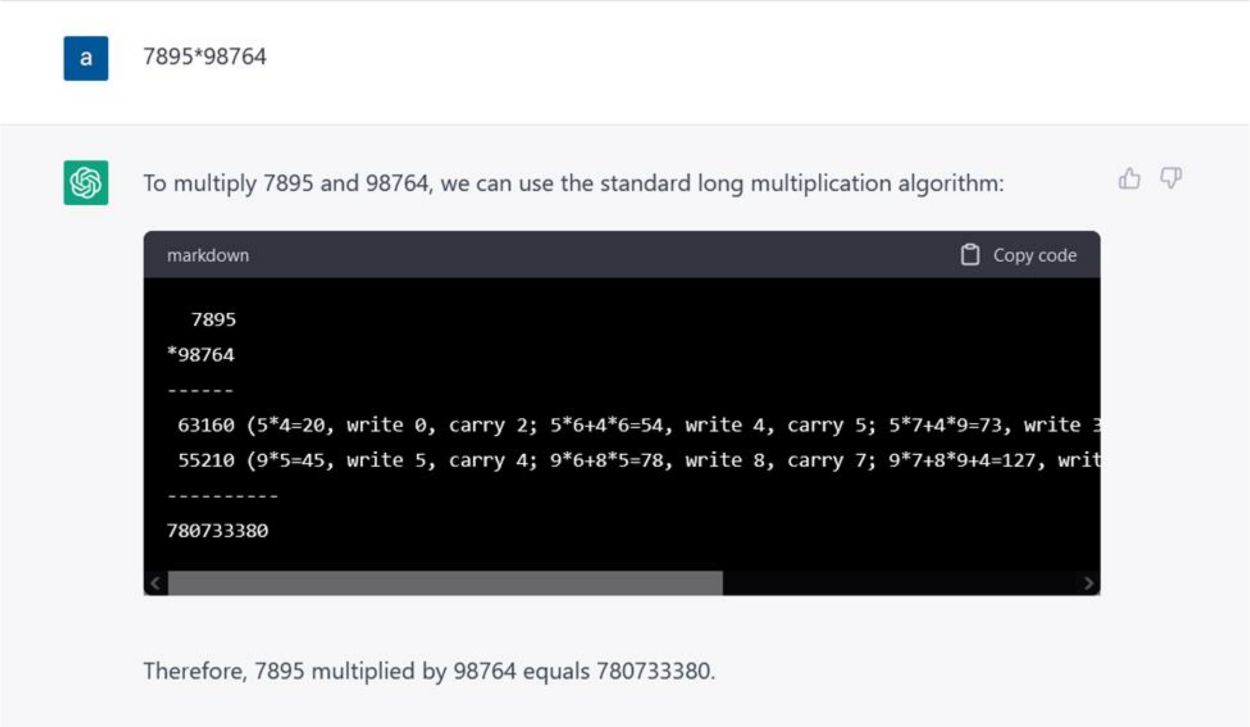
The expression -4*-4*-4 can be simplified using the rules of multiplication, which state that multiplying two negative numbers results in a positive number.

-4 * -4 = 16 (positive)
16 * -4 = -64 (negative)

Therefore, -4*-4*-4 = -64.

fig 2. Simple maths multiplication problem

We asked a simple maths problem (see fig 2) ChatGPT gave the right answer with the whole process. This question is inspired by one tweet[11], where user asked, -1*-1*-1 and ChatGPT gave the final result 1. Whereas the answer would be -1. So if we compare our question (fig 2) and the question of Twitter users, it clearly shows that ChatGPT updated their maths ability.



7895*98764

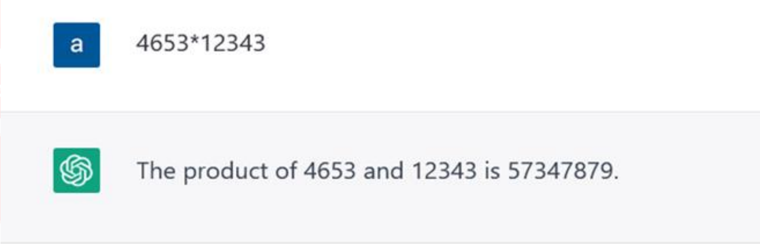
To multiply 7895 and 98764, we can use the standard long multiplication algorithm:

```
  7895
*98764
------
 63160 (5*4=20, write 0, carry 2; 5*6+4*6=54, write 4, carry 5; 5*7+4*9=73, write
 55210 (9*5=45, write 5, carry 4; 9*6+8*5=78, write 8, carry 7; 9*7+8*9+4=127, writ
----------
780733380
```

Therefore, 7895 multiplied by 98764 equals 780733380.

fig 3. long multiplication problem

We asked a long multiplication problem (see fig 3), and it gave the wrong answer. The right answer is 779,741,780. Another thing is, the process of multiplication given by ChatGPT is also very confusing. Let’s give another chance to ChatGPT.



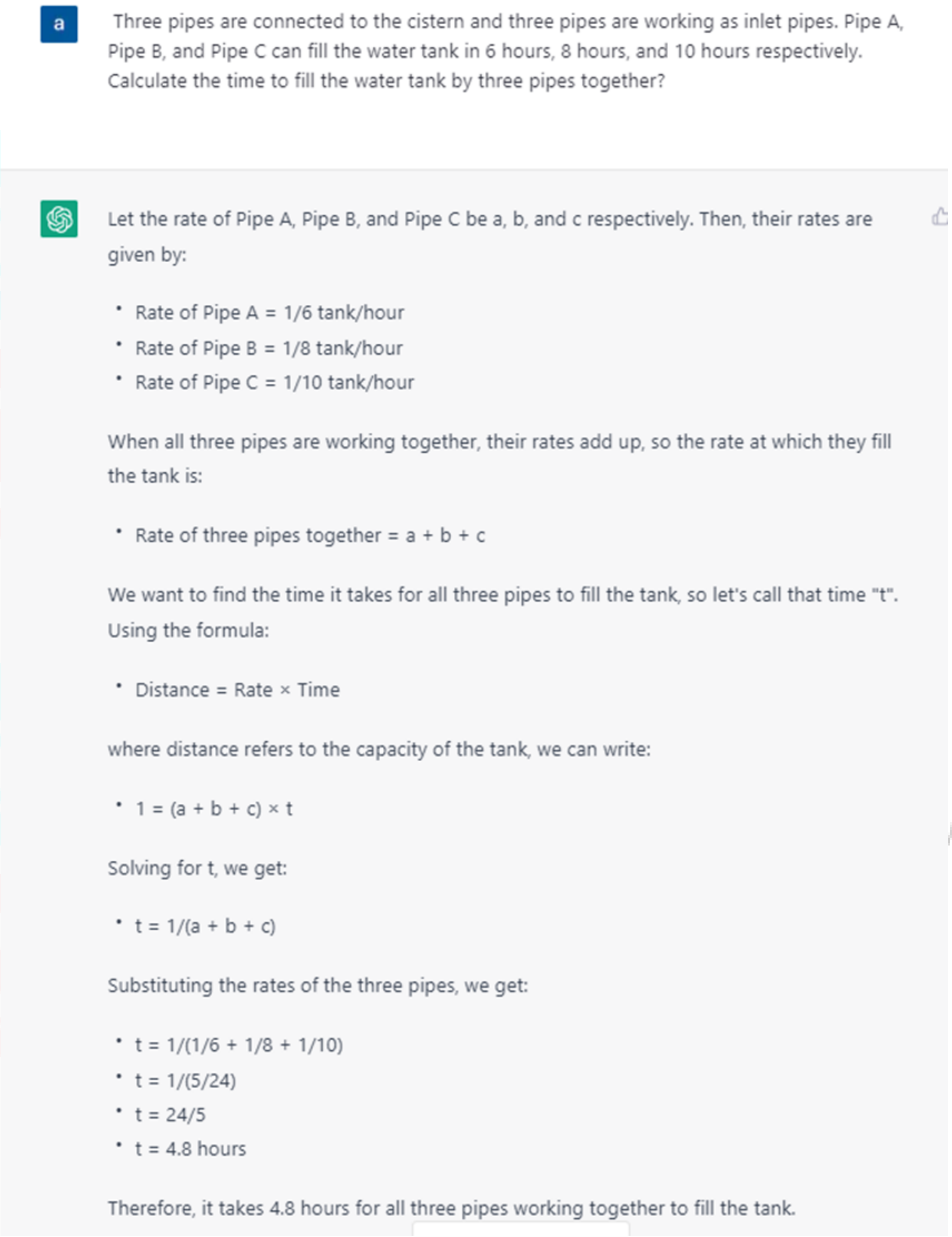
4653*12343

The product of 4653 and 12343 is 57347879.

fig 4. another multiplication problem

It again failed in the multiplication problem. The right answer is 57431979.

The main reason behind this multiple failure in multiplication is, ChatGPT is a language model. It has no understanding of maths or logic. So it will give output that is statistically likely to come next according to data it’s been trained on. In simple words ChatGPT excels at memorization rather than in calculation.



Three pipes are connected to the cistern and three pipes are working as inlet pipes. Pipe A, Pipe B, and Pipe C can fill the water tank in 6 hours, 8 hours, and 10 hours respectively. Calculate the time to fill the water tank by three pipes together?

Let the rate of Pipe A, Pipe B, and Pipe C be a, b, and c respectively. Then, their rates are given by:

- Rate of Pipe A = 1/6 tank/hour
- Rate of Pipe B = 1/8 tank/hour
- Rate of Pipe C = 1/10 tank/hour

When all three pipes are working together, their rates add up, so the rate at which they fill the tank is:

- Rate of three pipes together = a + b + c

We want to find the time it takes for all three pipes to fill the tank, so let's call that time "t". Using the formula:

- Distance = Rate × Time

where distance refers to the capacity of the tank, we can write:

- 1 = (a + b + c) × t

Solving for t, we get:

- t = 1/(a + b + c)

Substituting the rates of the three pipes, we get:

- t = 1/(1/6 + 1/8 + 1/10)
- t = 1/(5/24)
- t = 24/5
- t = 4.8 hours

Therefore, it takes 4.8 hours for all three pipes working together to fill the tank.

fig. 5 pipe and tank problem

In the above pipe and water tank problem (see fig 5), The whole process of solving the problem is right. Even the equation is right but failed in the calculation. The answer will be 2.5 hours.

2.3 Coding

ChatGPT excels at solving programming issues. ChatGPT has the ability to write the code but sometimes it produces inaccurate and long unnecessary code lines, that’s why ChatGPT cannot fully substitute human developers. ChatGPT can handle the task of junior developer, which requires minimum coding and DSA knowledge, like writing html and css programs or writing repetitive code. Some time ago Stack overFlow, a very famous website for coding Q&A, banned ChatGPT-generated content and answers[12]. According to them, the average rate of getting correct answers from ChatGPT is very low. ChatGPT can write very good code but still needs expertise on coding to verify the answer.

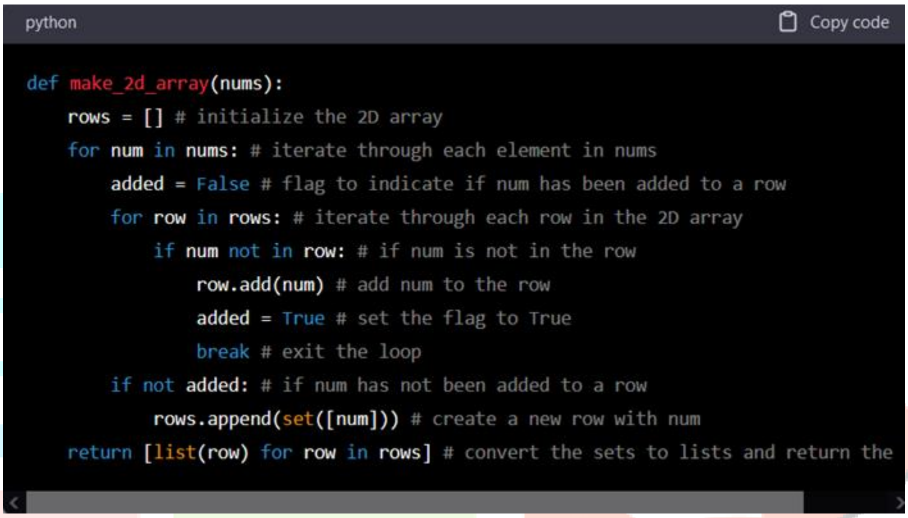
So testing ChatGPT ability on coding we tried to solve Leetcode weekly contest 339 by the help of ChatGPT. We gave three try on every question and here is the result

Question 1- finding the longest balanced substring of Binary String
First try- Passed 626/1848 cases
Second try- Passes 627/1848 cases
Third try- Not able to pass, first three cases of given test cases

Result- failed

Question 2- Convert an array into a 2D array With Conditions.
First try- passes 1035/1035 cases

Result- Pass



```python
def make_2d_array(nums):
    rows = [] # initialize the 2D array
    for num in nums: # iterate through each element in nums
        added = False # flag to indicate if num has been added to a row
        for row in rows: # iterate through each row in the 2D array
            if num not in row: # if num is not in the row
                row.add(num) # add num to the row
                added = True # set the flag to True
                break # exit the loop
        if not added: # if num has not been added to a row
            rows.append(set([num])) # create a new row with num
    return [list(row) for row in rows] # convert the sets to lists and return the
```

fig-6 solution of 2nd question

Question no 3- Mice and Cheese (Use of Dynamic programming)
First try- 0/564 test cases passed
Second try- 74/564 test cases passed
Third try - 58/564 test cases passed

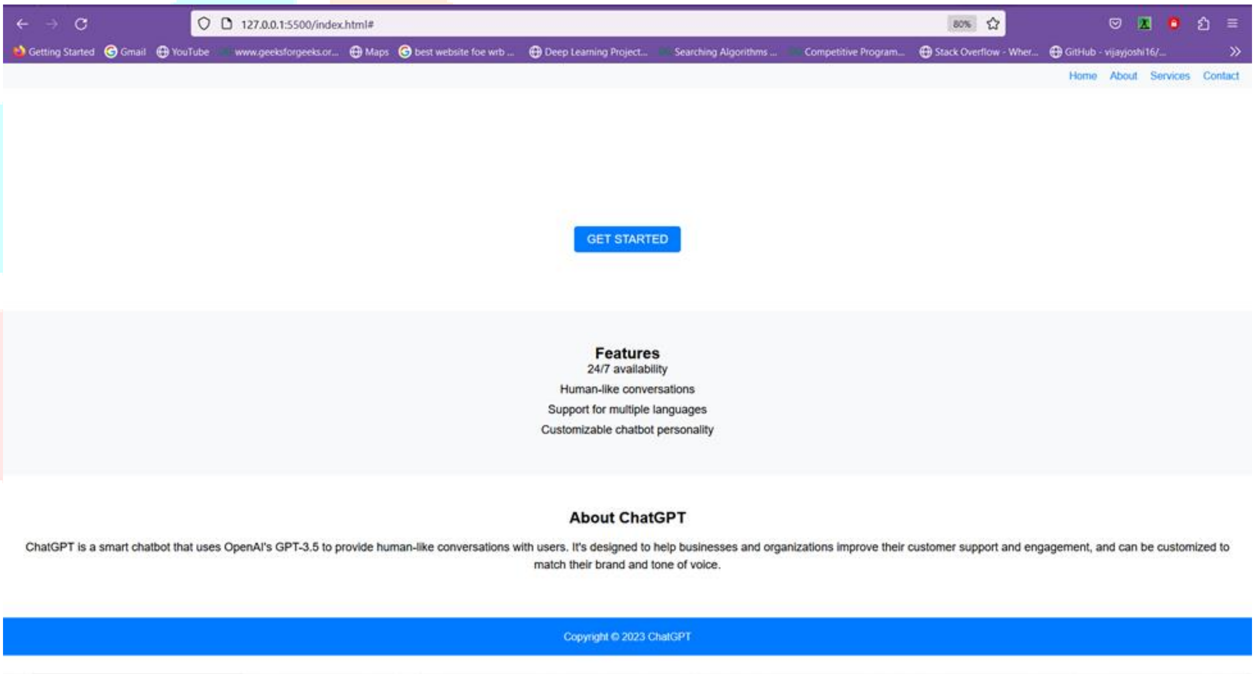
Result- failed

Question 4- Minimum Reverse operation
First try - 132/711 test cases passed
Second try- 56/711 test cases passed
Third try - 319/711 test cases failed

Result- failed

ChatGPT only solved 1 question out of 4. In every problem, ChatGPT gave me a clear and right description about the problem, but it clearly failed on writing the solution, which clearly shows it’s lack of creativity and greedy approach.

However, ChatGPT generated CSS and Html programs are very good, because they require more memorization and less mathematical ability. Below fig is the simple webpage created by ChatGPT code.



Home About Services Contact

GET STARTED

Features

24/7 availability

Human-like conversations

Support for multiple languages

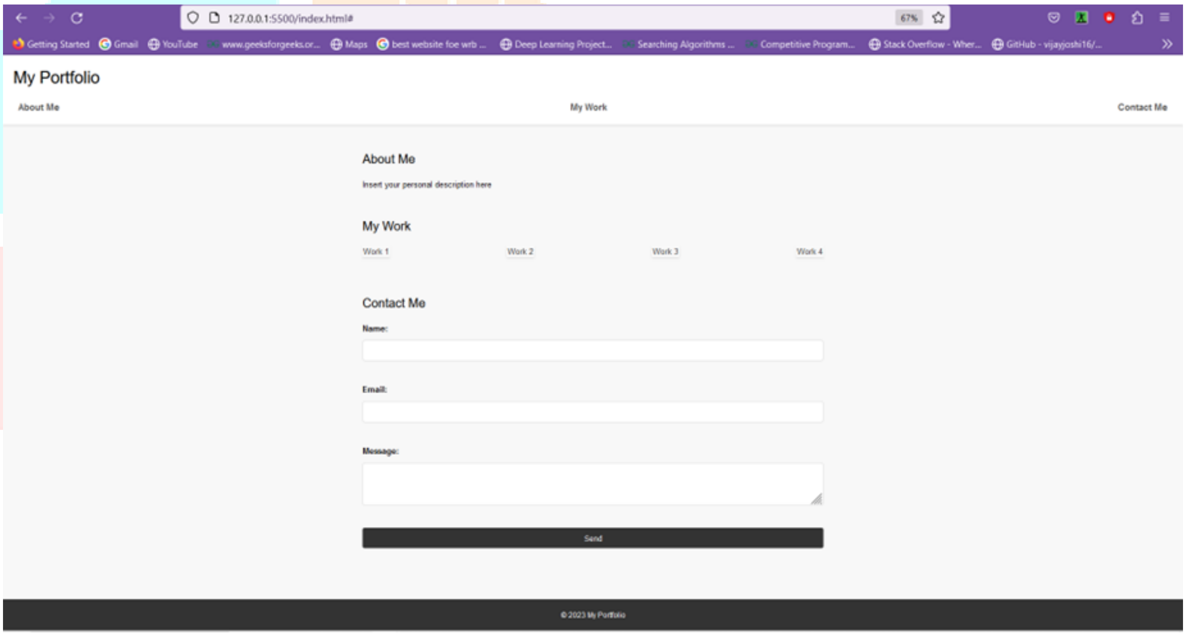
Customizable chatbot personality

About ChatGPT

ChatGPT is a smart chatbot that uses OpenAI's GPT-3.5 to provide human-like conversations with users. It's designed to help businesses and organizations improve their customer support and engagement, and can be customized to match their brand and tone of voice.

Copyright © 2023 ChatGPT

Fig7: simple web page created by ChatGPT



My Portfolio

About Me My Work Contact Me

About Me

Insert your personal description here

My Work

Work 1 Work 2 Work 3 Work 4

Contact Me

Name:

Email:

Message:

Send

© 2023 My Portfolio

Fig 8: portfolio website created by ChatGPT

The above two web pages are very simple and can be customised by including images and a human developer's creativity.

III. RESULT AND DISCUSSION

Throughout this study, I attempted to test ChatGPT on three categories: Arithmetic, Reasoning, and Coding, and discovered several flaws and limitations in ChatGPT. Despite ChatGPT's impressive capabilities in specific tasks, its thinking and mathematical approach could be improved. Here are some facts I discovered during my research.

- It is easy to confuse ChatGPT and can change the answer based on what the user wants. We have seen in the temporal reasoning problem, where it changes the answer from an assumption to a statical-based answer. It also indicates that ChatGPT can accept his mistake and try to modify the answer.
- ChatGPT responses are not always correct. It is possible that some parts of the answer

contain errors. This was evident in the Arrival problem (fig 1). It provided the correct final answer, but the order in which it arrived was incorrect. It also demonstrates that you should not have relied solely on ChatGPT's response, it still required verification to confirm the answer it right or wrong

- ChatGPT appears to be very confident in his ability and tries to give an answer, no matter how it comes or whether the process is right or wrong. In Animal problem, it tries to purchase 100 animals for 100 rupees no matter how it comes, even if it replaces 1 animal to 1 set of 8 hens. This whole conversation is the best example to show the dumbness of AI.
- ChatGPT can't solve the multiplication problem and we have seen two multiplication problems. Even during the first release of ChatGPT, it couldn't solve the simple multiplication of negative numbers. However, its answer improved in the next update. So we can assume ChatGPT may improve his mathematical ability in every update.
- In the pipe water tank problem, we have seen that the whole process of solution was right but it failed on the calculation part. It clearly shows ChatGPT's memorization ability, where it knows what and how the formula will be used.
- ChatGPT is trained on a lot of LeetCode problems, including hard problems. But still ChatGPT failed in the LeetCode contest, even after multiple tries, it made the code even worse. For example in Leetcode problem no 1, ChatGPT was able to pass 627/1848 cases in the first two tries, but it went to 0 in the third try. It demonstrates the ChatGPT's inability to understand the code and how it works.

Since HTML and CSS are full of memorization, ChatGPT was able to write bug-free code for web pages. However, it lacks creativity and requires human creativity to make that webpage more beautiful and responsive.

ChaGPT is a language model, not a mathematical model. Writers should be more concerned than programmers and developers. ChatGPT has a long way to go before it can replace the developer. ChatGPT is more of a problem solver than a professor or developer. It is more like an updated version of Google where the user has to find his solution to a problem on various websites, but now ChatGPT will provide the best and most concise answer. Coders who are thinking of using ChatGPT for writing code are setting themselves up for the big fall. Because the correct result from the already available resulting code cannot be assumed. It may work for that time, but in the long run, it is not the best way to write code.

REFERENCES

[1]https://chat.openai.com/chat

[2]L. Varanasi, "ChatGPT is on its way to becoming a virtual doctor, lawyer, and business analyst. Here's a list of advanced ex," *Business Insider India*, Jan. 28, 2023. Accessed: Apr. 04, 2023. [Online]. Available: https://www.businessinsider.in/tech/news/chatgpt-is-on-its-way-to-becoming-a-virtual-doctor-lawyer-and-business-analyst-hereaposs-a-list-of-advanced-exams-the-ai-bot-has-passed-so-far-/slidelist/97388435.cms#slideid=97388478

[3]*Twitter*. https://twitter.com/emollick/status/1635700173946105856 (accessed Apr. 04, 2023).

[4]E. Dreibelbis, "Watch Out, Software Engineers: ChatGPT Is Now Finding, Fixing Bugs in Code," *PCMag*, Jan. 27, 2023. Accessed: Apr. 04, 2023. [Online]. Available: https://www.pcmag.com/news/watch-out-software-engineers-chatgpt-is-now-finding-fixing-bugs-in-code

[5]giuven95, "GitHub - giuven95/chatgpt-failures: Failure archive for ChatGPT and similar models," *GitHub*. https://github.com/giuven95/chatgpt-failures (accessed Apr. 04, 2023).

[6]A. Borji, "A Categorical Archive of ChatGPT Failures." https://arxiv.org/pdf/2302.03494.pdf (accessed Apr. 04, 2023).

[7]A. Azaria, "ChatGPT Usage and Limitations," *unknown*, Dec. 27, 2022. https://www.researchgate.net/publication/366618623_ChatGPT_Usage_and_Limitations (accessed Apr. 04, 2023).

[8]L. Varanasi, "ChatGPT is on its way to becoming a virtual doctor, lawyer, and business analyst. Here's a list of advanced ex," *Business Insider India*, Jan. 28, 2023. Accessed: Apr. 04, 2023. [Online]. Available:

https://www.businessinsider.in/tech/news/chatgpt-is-on-its-way-to-becoming-a-virtual-doctor-lawyer-and-business-analyst-hereaposs-a-list-of-advanced-exams-the-ai-bot-has-passed-so-far-/slidelist/97388435.cms#slideid=97388478

[9]L. Varanasi, “ChatGPT is on its way to becoming a virtual doctor, lawyer, and business analyst. Here’s a list of advanced ex,” *Business Insider India*, Jan. 28, 2023. Accessed: Apr. 04, 2023. [Online]. Available: https://www.businessinsider.in/tech/news/chatgpt-is-on-its-way-to-becoming-a-virtual-doctor-lawyer-and-business-analyst-hereaposs-a-list-of-advanced-exams-the-ai-bot-has-passed-so-far-/slidelist/97388435.cms#slideid=97388478

[10]A. Bharade, “ChatGPT failed miserably in Singapore’s 6th-grade tests, averaging 16% for math and 21% for science. Days lat,” *Business Insider India*, Feb. 20, 2023. Accessed: Apr. 07, 2023. [Online]. Available: https://www.businessinsider.in/tech/news/chatgpt-failed-miserably-in-singapores-6th-grade-tests-averaging-16-for-math-and-21-for-science-days-later-it-had-learned-from-its-mistakes-/articleshow/98088629.cms

[11]*Twitter*. https://twitter.com/vladquant/status/1599828362478129153 (accessed Apr. 07, 2023).

[12]Makyen, “Temporary policy: ChatGPT is banned,” *Meta Stack Overflow*. https://meta.stackoverflow.com/questions/421831/temporary-policy-chatgpt-is-banned (accessed Apr. 10, 2023).