

Can ChatGPT Pass Classical Control Theory Exam?

Itsuki Ogo^{1†} and Masanobu Koga¹

¹Department of Intelligent and Control Systems, Kyushu Institute of Technology, Fukuoka, Japan
(Tel: +81-80-9109-5650; E-mail: ogo.itsuki406@mail.kyutech.jp)
(Tel: +81-948-29-7726; E-mail: koga@ics.kyutech.ac.jp)

Abstract: With the advancement of AI technology, generative AI has made remarkable progress in its ability to process multiple languages and adapt to creative tasks. It has also demonstrated its strength in academic fields, such as passing the national medical examinations. In this study, we tested the extent to which ChatGPT (GPT-4) can accurately answer classical control theory questions offered in undergraduate courses. The experimental results showed that GPT-4 showed a correct response rate of under 70% for quiz exercises in classical control theory, and that the correct response rate was lower for problems whose solutions were specific or required step-by-step thinking. In addition, since GPT-4 is a Transformer-based model, and the answers are based on mere prediction, it may give incorrect answers for problems that require complex calculations. In this study, we proposed a method to improve the response accuracy by developing a customized GPT specialized for classical control theory and using prompt engineering. The proposed method was applied to a university undergraduate final exam in undergraduate course, and the results showed that the correct response rate was improved and a passing score (60% or higher) was obtained.

Keywords: ChatGPT, Generative AI, Classical Control Theory, GPTs, GPT-4

1. INTRODUCTION

Artificial intelligence (AI) technology has evolved remarkably in recent years, and in particular, generative AI has made significant progress. One of the generative AIs that is currently being used rapidly is ChatGPT[1], a conversational AI developed by OpenAI[2]. ChatGPT is based on a type of neural network model, the Generative Pre-trained Transformer (GPT), which specializes in natural language processing and generation. The GPT model is also based on a "Large Language Model" (LLM) that utilizes a "Transformer" model[3]. That is, it has the ability to pre-learn large amounts of text data and generate sentences based on it. In addition, it has the advantage of being able to perform natural language processing in multiple languages, including English, Spanish, and Japanese, contributing to problem solving in a wide variety of fields. Examples include text summarization and translation, data organization and analysis. They can also perform tasks that require more creative thinking, such as "generating ideas for projects," "writing poetry," and "writing code."

The first GPT model was introduced in 2018 and has since evolved into GPT-2, GPT-3, and in March 2023, GPT-4 was released as a multimodal model[4]. GPT-4 is capable of performing at the human level on professional and academic examinations, and has passed the Unified Bar Examination with a score in the top 10% of test takers on the practice test[5]. In addition, a study evaluating GPT-4 on Japan's national medical examinations over the past six years showed that GPT-4 passed all six years examinations. The study also highlights the potential of the LLM in languages that are typologically distant from English[6].

We are conducting a survey and research to determine

the current status of the ChatGPT (GPT-4) ability in the area of control engineering. In this paper, we focus on classical control theory: we evaluated the ability of GPT-4 by having it solve various exercises on classical control theory, and found that GPT-4 has a correct response rate of under 70% for quiz exercises in classical control theory. In this study, we proposed a method to improve the accuracy of the answers based on the problems that could not be solved and evaluated its effectiveness.

Although GPT-4 can read pdf and other files, it cannot correctly read pdf files that describe problems related to classical control theory, which contain not only text but also mathematical formulas and diagrams. Therefore, we converted the pdf files to Markdown format[7] before applying the problems to GPT-4 so that GPT-4 can accurately understand the problems.

Since GPT-4 is a Transformer-based model and answers by mere prediction, it may give wrong answers in questions that require complex calculations. And since LLM generates new text based on past training data, there is a certain degree of randomness in this generation process. In addition, even if a question is answered correctly once, it may be answered incorrectly by having GPT-4 try to solve it again. Therefore, in this study, we created a customized GPT specialized for classical control theory that incorporates a mechanism to improve answer accuracy using GPTs[8], which is one of the services provided by OpenAI.

The structure of this paper is as follows: In Section 2, we report the method and results of experiments to evaluate the performance of GPT-4 in classical control theory; in Section 3, we propose methods to improve the response accuracy of GPT-4, including the development of GPTs; in Section 4, we compare the correct response rates of the proposed method, GPT-4, and conventional prompt engineering methods. In Chapter 5, we discuss

† Itsuki Ogo is the presenter of this paper.

the effectiveness of the proposed method, and in Chapter 6, we provide a summary and outlook of this study.

2. PERFORMANCE EVALUATION OF GPT-4

We evaluated the performance of GPT-4 by having ChatGPT (GPT-4) answer quiz exercises in a classical control theory course for undergraduate students. GPT-4, a multimodal model, can read a variety of files, including pdf files [9]. In this study, we initially tried to obtain answers by giving the pdf files of the exercises to GPT-4. However, we found that GPT-4 could not read the questions correctly and often gave incorrect answers. One possible reason for this is that the problems of classical control theory contain block diagrams, diagrams of physical systems, and mathematical formulas in addition to text, and these multiple types of information are mixed together and arranged in two dimensions. In general, materials related to control theory, including exercises and final exam questions, are provided in pdf files. Therefore, it is very important to be able to respond to questions given in pdf files. In this study, a method was devised to convert pdf files containing text, mathematical expressions, block diagrams, and other graphics into Markdown format, and this method was used in the preparation stage of this experiment. The details of this method are explained in section 3.3.

2.1. Experimental Procedures

A schematic of the evaluation experiment procedure is shown in Fig. 1.

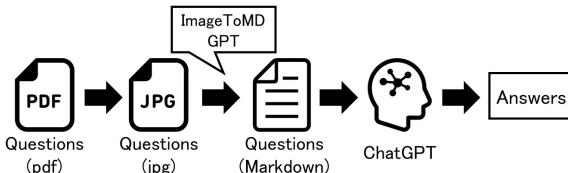


Fig. 1 Overview of experimental procedure

The flow of the evaluation experiment is as follows.

1. Convert the pdf with the problem described into an image file.
2. Convert image files to Markdown format.
3. Give the Markdown file to GPT-4.
4. Compare the answers on GPT-4 with the model answers to determine correctness.

2.2. Experimental results

Table 1 shows the experimental results. Table 1 shows that GPT-4 has a correct response rate of approximately 66% for basic classical control theory problems. Its strength was confirmed especially in text-based problem-solving skills. On the other hand, the correct response rate was lower for image-based problems, such as deriving transfer functions from block diagrams and diagrams

of physical systems. Even for text-based problems, the correct response rate was low for problems in which the coefficients of the characteristic equation are variables, for difficult problems that cannot be solved using basic methods, and for problems that require step-by-step thinking.

3. METHODS

3.1. GPTs

Research is underway on a method called prompt engineering to obtain better results from LLMs without using large data sets[10]. Prompt engineering is a technique in the field of artificial intelligence, particularly in models using natural language processing, in which inputs are designed in such a way that they can be used to perform a specific task. This approach plays an important role in a variety of AI models, including large-scale language models such as GPT. The main goal of prompt engineering is to optimize the input sentences so that the AI exhibits the desired output or behavior. This improves the performance of the model and its adaptability to specific tasks.

As prompt engineering research continues, a new service called GPT Builder was released by OpenAI on November 7, 2023[11]. This service allows users to create custom versions of ChatGPTs, called GPTs. Users can create GPTs that are more adaptable to specific tasks by editing mainly "Instructions," "Knowledge," "Capabilities," and "Actions. Table 2 shows an overview of the parameters that can be edited in GPT Builder.

Table 2 GPTs parameters

Parameter	Function
Instructions	Prompts given in advance to GPT
Knowledge	Utilization of data
Capabilities	Whether to use Code Interpreter
Actions	Use of third party APIs

3.2. Customized GPT for classical control theory

In this study, we propose a method to improve the accuracy of answers by creating a customized GPT for classical control theory based on the questions to which GPT-4 gave wrong answers in the experiments described in section 2. Fig. 2 shows a schematic of the flow of giving a question to the proposed GPT and obtaining an answer.

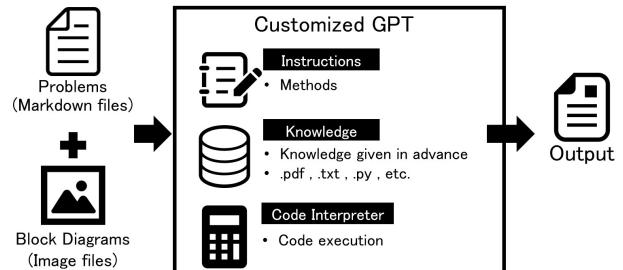


Fig. 2 Overview of the flow of getting answers using customized GPT

Table 1 Evaluation questions and results

Lesson topics	Number of problems	Number of correct answers
Laplace transform & inverse transform	15	15
Differential equations & transfer functions	7	2
Block diagrams & transfer functions	13	2
System stability	12	9
Routh stability criterion	13	6
Nyquist stability criterion	14	8
System transient & steady-state response	11	10
Step response of primary & secondary system	5	5
System frequency response	3	2
Drawing a bode diagram	5	5
Total	97	64

3.3. GPT to convert pdf files to Markdown

We have created a GPT that converts pdf files to Markdown format. In the future, it is expected that this GPT can be connected as the first stage of other GPTs so that pdf files can be directly fed to GPT. This GPT can convert sentences and mathematical expressions written in a given image file to Markdown format. An example of the conversion is shown in Fig. 3. The upper part of Fig. 3 shows GPT's display of the given image file, and the lower part is GPT's answer.

You

Question 2.3

Find the step response of the following transfer function $P(s)$.

1) $P(s) = \frac{3}{s+2}$ 2) $P(s) = \frac{s+4}{(s+1)(s+2)}$ 3) $P(s) = \frac{2}{(s+1)^2}$

ImageToMD

```
markdown
## Question 2.3
Find the step response of the following transfer function $P(s)$.
1) $P(s) = \frac{3}{s+2}$
2) $P(s) = \frac{s+4}{(s+1)(s+2)}$*
3) $P(s) = \frac{2}{(s+1)^2}$
```

Fig. 3 Example of Markdown conversion using GPT

3.4. Code Interpreter

Since GPT-4 is a Transformer-based model and answers by mere prediction, it may give wrong answers in questions that require complex calculations. Therefore, the accuracy of answers can be improved by instructing GPT-4 to always use the Code Interpreter for calculations. Furthermore, providing specific Python code beforehand and having it executed reduces the randomness of the responses, which is a limitation of the GPT model. The method of directly providing Python code is considered to be effective in performing specific tasks.

3.5. RouthGPT

From the evaluation experiments, GPT-4 can solve the basic Routh's stability determination problem. However, it can make mistakes in some cases, such as when the coefficients of the characteristic equation are variables, when the Routh number sequence is indefinite, or when certain rows of the Routh table are all zero. Therefore, we created RouthGPT to handle such cases. The following shows the changes in responses when using RouthGPT. The problem given to each model is

"Determine the stability of the system with the characteristic equation $\phi(s) = s^5 + 2s^4 + 3s^3 + 6s^2 + 5s + 3 = 0$ using Routh's stability criterion".

The response of GPT-4 and RouthGPT are shown in Figs. 4 and 5, respectively.

The corrected Routh array for the system with the characteristic equation $\phi(s) = s^5 + 2s^4 + 3s^3 + 6s^2 + 5s + 3 = 0$ is as follows:

1	3	5
2	6	3
0	-3.5	0
0	0	0
0	0	0
0	0	0



Fig. 4 Answer from GPT-4

The Routh array for the characteristic equation $\phi(s) = s^5 + 2s^4 + 3s^3 + 6s^2 + 5s + 3 = 0$ is constructed as follows:

1	3	5
2	6	3
0	7	0
6 - $\frac{7}{2}$	3	0
$-\frac{(6e-7)^2+49}{(12e-14)}$	0	0
3	0	0



Fig. 5 Answer from RouthGPT

From Figures 4 and 5, it can be seen that RouthGPT is able to correctly output a Routh table, which could not be properly created by GPT-4, by replacing the zeros that appeared in the Routh number sequence with the variable ϵ in the creation of the Routh table.

3.6. StabilityDeterminerGPT

From Table 1, GPT-4 is unable to solve stability-determining problems that cannot be solved by basic methods. We created StabilityDeterminerGPT (SDGPT) to improve such problems; in addition to the RouthGPT instructions, we gave instructions for dealing with pole-zero cancellation.

"Given $C(s) = \frac{s-1}{s+1}$ and $P(s) = \frac{1}{s-1}$, determine the stability of the feedback control system."

The answers to the above questions given to GPT-4 and SDGPT are shown in Figures 6 and 7

The root of the characteristic equation $C(s)P(s) = -1$ is $s = -2$.

This root is located in the left half-plane. This indicates that the feedback control system is stable. If all poles of the closed-loop system are in the left half-plane, the system is considered stable. Therefore, the feedback control system composed of the given $C(s)$ and $P(s)$ is stable.

Fig. 6 Answer from GPT-4

In this case, although the pole of $P(s)$ at $s = 1$ is in the right half-plane and is cancelled by the zero of $C(s)$, pole-zero cancellation does not exclude an unstable pole from the stability evaluation. Therefore, considering the cancelled pole as part of the system's overall stability assessment, the feedback control system is determined to be unstable due to the presence of a pole in the right half-plane.

Considering pole-zero cancellation, it is concluded that the feedback control system is unstable.

Fig. 7 Answer from StabilityDeterminerGPT

From Figures 6 and 7, GPT-4 incorrectly determined "stable," whereas SDGPT accurately determined "unstable" by performing stability determination that takes pole-zero cancellation.

3.7. MultiThoughtsGPT

Since LLM generates new text based on past training data, there is a certain amount of randomness in this generation process. In addition, even if a question is answered correctly once, it may be answered incorrectly by having GPT-4 try to solve it again. Therefore, to improve response accuracy, we created MultiThoughtsGPT (MTGPT), which allowed GPT-4 to make multiple response policies at once and instructed GPT-4 to adopt the policy considered to be the best among them. An overview of this GPT is shown in Fig. 8.

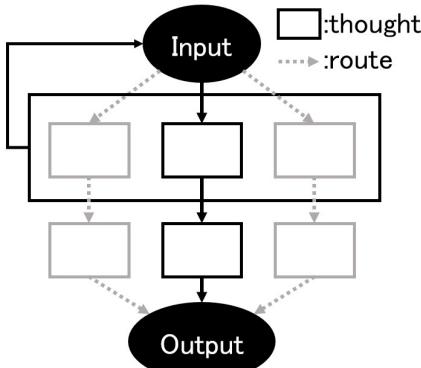


Fig. 8 Overview of MultiThoughtsGPT

MTGPT creates multiple answer policies for questions given by the user. Each response policy is rated on a scale between 0 and 100 points based on its complexity (the degree to which it goes through multiple steps and logical reasoning to arrive at an answer). The policy that receives the highest score in this evaluation, i.e., the one with the lowest complexity, is used as the basis for proceeding with the response. If the scores in the evaluation of all the response policies are below 50 points, multiple response policies are proposed again, and the loop continues until it finds a response policy that exceeds 50 points. In addition, by integrating the instructions of other GPTs within MTGPT, if necessary, the appropriate processing is performed according to the instructions that have been prepared. An example of MTGPT output is shown in

Fig. 9. From Fig. 9, it can be seen that multiple response policies are formulated for a given question, and the evaluation for each is output along with the reason for each.

Question 1: Conditions for K and T for Closed-loop Stability

Strategy 1: Closed-loop Transfer Function Approach

• Score: 70/100

• Reason: This method involves deriving the closed-loop transfer function from the given controller and plant transfer functions. We calculate the overall closed-loop transfer function using the formula $T(s) = \frac{C(s)P(s)}{1+C(s)P(s)}$. This requires algebraic manipulation to express the characteristic equation in terms of s , from which we can apply stability criteria (e.g., Routh-Hurwitz criterion) to determine the conditions for K and T . The score isn't higher because it involves several steps of algebraic manipulation and applying the Routh-Hurwitz criterion, which can get complex.

Strategy 2: Root Locus Approach

• Score: 60/100

• Reason: By examining the root locus of the open-loop transfer function $C(s)P(s)$, we can qualitatively determine the effects of varying K and T on system stability. This approach gives insights into how the roots of the characteristic equation move with changes in K and T . The simplicity score is moderate because it requires understanding of the root locus principles and how parameter variations affect the system's poles, but it doesn't provide explicit conditions without further analysis.

Fig. 9 An example of output of MultiThoughtsGPT

In order to evaluate the effectiveness of MTGPT, 30 re-generations of answers were performed for each of the "SDGPT" and "MTGPT" for the same question. The results are shown in Table 3.

Table 3 Correct answer rates of SDGPT and MTGPT

Model	SDGPT	MTGPT
Number of problems	30	30
Correct answers	7	18
Wrong answers	5	3
Unanswered	18	9
Correct percentage	23%	60%

Table 3 shows that MTGPT has a 60% correct response rate for questions where SDGPT has a 23% correct response rate, an improvement of 37 percentage points.

4. EVALUATION EXPERIMENT

4.1. Experimental method

The final exam questions from a classical control theory course in Kyushu Institute of Technology were used as the experimental problems for evaluation. The experimental procedure is the same as in Section 2.1. GPT-4, GPT-4 with Zero-shot CoT (Chain of thought) Prompting, and MTGPT were used for comparison. The correct response rates of each model were compared to evaluate the effectiveness of the proposed method. Zero-shot CoT prompting is a method that allows step-by-step inference without examples of inference steps in the prompt; by adding the phrase "Let's think step by step." to the prompts given to ChatGPT, intermediate inference step by step, allowing for complex reasoning[12]. The proposed method is identical to Zero-Shot-CoT in the sense that it does not take the approach of giving examples of concrete problems and answers. Therefore, we evaluate the effectiveness of the proposed method by comparing it

with Zero-shot-CoT, one of the existing prompt engineering methods.

4.2. Results of the evaluation experiment

The experimental results are shown in Table 4.

Table 4 Correct answer rate of each method

Model	GPT-4	Zero-shot CoT	MTGPT
Problems	25	25	25
Correct	13	13	17
Wrong	11	11	8
Unanswered	1	1	0
Correct Percentage	52%	52%	68%

Table 4 shows that the proposed method improves response accuracy by 16 percentage points and achieves a passing score (60% or higher). In the evaluation experiment, the questions and responses that showed particular differences are shown below. The responses to GPT-4, GPT-4 with Zero-shot-CoT, and MTGPT are shown in Figs. 10, 12, 14, and 16, respectively. In addition, the responses in Japanese for GPT-4, GPT-4 with Zero-shot-CoT, and MTGPT are shown in Figs 11, 13, 15, and 17, respectively. The questions given to each model are as follows.

"Answer the following questions for a feedback control system with the following controller and control target.

$$C(s) = \frac{2.5K(Ts + 1)}{s}, P(s) = \frac{1}{(s + 0.5)(s + 1)}$$

1. Find the conditions that K and T must satisfy for the closed-loop system to be stable."

- Derive the characteristic equation of the closed-loop system: Use the product of the controller $C(s)$ and the plant $P(s)$ to consider the feedback loop and derive the characteristic equation from the denominator of the closed-loop transfer function $T(s) = \frac{1}{C(s)P(s)} = \frac{1}{(s+0.5)(s+1)} = \frac{2.5K(Ts+1)}{s(s+0.5)(s+1)}$.
- Find the poles of the characteristic equation: Solve the characteristic equation for s and find the conditions under which all solutions (poles) are in the left half-plane (have negative real parts). This directly relates to the conditions that K and T must satisfy.
- Analyze the impact of K : Analyze how the poles of the closed-loop system move as K increases. The goal is to find the range of T for which the poles remain in the left half-plane as K approaches infinity.

Unfortunately, I am unable to provide specific numerical calculations, but by following the steps above, it is possible to manually derive the stability conditions. Solving this problem requires knowledge of control theory and mathematical skills to determine the location of the roots.

Fig. 10 Answer from GPT-4

まず、制御器 $C(s)$ と制御対象 $P(s)$ の積を求め、その後にフィードバックループを考慮して閉ループ伝達関数を導出します。

制御器と制御対象の積は以下の通りです。

$$C(s)P(s) = \frac{2.5K(Ts+1)}{s} \cdot \frac{1}{(s+0.5)(s+1)} = \frac{2.5K(Ts+1)}{s(s+0.5)(s+1)}$$

次に、閉ループ伝達関数 $T(s)$ は一般に以下の形で表されます。

$$T(s) = \frac{C(s)P(s)}{1+C(s)P(s)}$$

これを用いて、 $T(s)$ の具体的な式を導き、その極を求めて安定性の条件を分析します。
具体的な計算を進め、 K と T に関する条件を導きましょう。

① Error analyzing

残念ながら、具体的な数値計算を提供できませんが、上記のステップに従って、安定性の条件を手動で導出することが可能です。この問題を解くには、制御理論の知識と、ルーツの位置を求めるための数学的なスキルが必要です。[6]

Fig. 11 Answer from GPT-4 (in Japanese)

The characteristic equation is as follows:

$$2.5K(Ts + 1) + s(s + 0.5)(s + 1) = 0$$



To provide detailed answers to these questions, a mathematical analysis of the characteristic equation is required. Using the Routh-Hurwitz criterion, it is possible to derive specific conditions, but this tool does not easily support detailed mathematical analyses. Thus, I offer general guidelines and thought processes.

To deduce the specific conditions for K and T , or the range for T , a detailed numerical analysis of the characteristic equation or a mathematical proof based on conditions is necessary. This process requires advanced knowledge of control theory and mathematical techniques.

Fig. 12 Answer from Zero-shot CoT

特性方程式は以下のようになりました。

$$2.5K(Ts + 1) + s(s + 0.5)(s + 1) = 0$$

この方程式から、閉ループ系が安定になるために K と T が満たすべき条件と、 K をいくら大きくしても閉ループ系が安定になるような T の範囲を求める必要があります。

(中略)

これらの問い合わせに対する詳細な解答を提供するには、特性方程式の数学的分析が必要です。Routh-Hurwitz 基準を使用することで、具体的な条件を導き出すことが可能ですが、このツールでは数式の詳細な解析が難しく、一般的な指針と考え方を提供します。

具体的な K と T の条件や T の範囲を導出するためには、特性方程式の具体的な数値解析や、条件に基づく数学的な証明が必要です。このプロセスは、高度な制御理論の知識と数学的な手法を必要とします。

Fig. 13 Answer from Zero-shot CoT (in Japanese)

The Routh array is constructed as follows:

$$\begin{array}{ccc} 1 & 2.5KT + \frac{1}{2} & 0 \\ \frac{3}{2} & 2.5K & 0 \\ 2.5KT - \frac{5}{3}K + \frac{1}{2} & 0 & 0 \\ 2.5K & 0 & 0 \end{array}$$



This indicates that for the system to be stable, each element in the first column of the Routh array needs to be positive, leading to conditions for K and T that must be met. Specifically:

- $2.5K > 0$, necessitating $K > 0$.
- $2.5KT - \frac{5}{3}K + \frac{1}{2} > 0$, from which we derive the condition for T .

By solving the inequality for T , we obtain the specific requirements that K and T must satisfy to ensure system stability.

Fig. 14 Answer from MultiThoughtsGPT

特性方程式を展開・整理した結果は次のようになります。

$$s^3 + 1.5s^2 + (0.5 + 2.5KT)s + 2.5K = 0$$

この方程式から、ラウスの安定判別法を適用するための係数を抽出することができます。これらの係数を用いてラウス表を作成し、 K と T に関する安定条件を導き出します。

ラウス表を作成し、安定性の条件を求めるために、次のコードを使用します。[6]

ラウス表は次のようになります。

$$\begin{array}{ccc} 1 & 2.5KT + \frac{1}{2} & 0 \\ \frac{3}{2} & 2.5K & 0 \\ 2.5KT - \frac{5}{3}K + 0.5 & 0 & 0 \\ 2.5K & 0 & 0 \end{array}$$



Fig. 15 Answer from MultiThoughtsGPT (in Japanese)

The final condition for T is organized as follows:

$$T > \frac{2}{3} - \frac{1}{5K}$$

This condition defines the values that T can take under the premise that $K > 0$. In other words, as long as K takes a positive value, T must satisfy this inequality.

Thus, we have derived the conditions that K and T must satisfy to ensure the stability of the closed-loop system. These conditions indicate how the control parameter K and the time constant T must be related to each other to ensure the system's stability.

Fig. 16 Answer from MultiThoughtsGPT

不等式を T について解いた結果、以下の条件が得られます。

$$2.5K \cdot T > \frac{5}{3}K - \frac{1}{2}$$

これを T について解くと、次の関係が導かれます。

$$T > \frac{\frac{5}{3}K - \frac{1}{2}}{2.5K}$$

ここで $K > 0$ であることが前提です。この条件を満たすために、 T の式をさらに明確にする必要があります。上記の計算では、分数式での明確な変換を行いましたが、不等式を直接解く手順に誤りがあったようです。この不等式を正しく解いて、具体的な T の条件を明確にしましょう。[6]

最終的に、 T' に関する条件は以下のように整理されます。

$$T > \frac{2}{3} - \frac{1}{5K}$$

Fig. 17 Answer from MultiThoughtsGPT (in Japanese)

From Figs. 10, 12, 14, and 16, it can be seen that GPT-4 and Zero-shot CoT responded that they could not provide specific numerical calculations and that it was difficult to analyze the formulas in detail, respectively, whereas MTGPT returned appropriate responses. From Figs. 11, 13, 15, and 17, it can also be seen that the responses were also obtained in Japanese.

5. DISCUSSION

Without fine-tuning the model with additional training data, we found that the prompts can improve response accuracy on a variety of classical control theory questions. In the proposed method, for questions that GPT-4 answered incorrectly, a response was incorporated into GPT's instructions to improve the accuracy of the answers. We believe that the method of providing questions in a structured Markdown format and having GPT-4 itself switch the applicable instructions as needed will be effective in highly specialized fields other than control engineering.

In addition, the proposed method does not require a detailed description of the problem and its specific solution, as is the case with the conventional Few-shot CoT. Furthermore, although all the instructions in the GPT are written in Japanese, the GPTs are able to appropriately answer in English. Another advantage of using GPT is that it can be easily used by users of other languages than Japanese.

On the other hand, there is no perfect reproducibility in the answers generated, and even questions that were answered correctly once may be answered incorrectly when GPTs is asked to solve them again. Therefore, it should be noted that the proposed method does not always produce correct answers.

6. CONCLUSIONS

In this study, we evaluated the performance of ChatGPT (GPT-4) by having it answer various exercises in classical control theory. Then, we created a customized GPT that could derive the correct answer to the questions that GPT-4 got wrong, thereby improving the accuracy of the answers. However, as discussed in Chapter 5, LLM generates new text based on past training data, and this generation process has some degree of randomness. Therefore, even for questions for which the GPT

customized in this experiment was able to derive the correct answer, there is a possibility that wrong answers will be generated if the re-generation of answers is repeated. Therefore, in the future, we would like to devise a method that can control the randomness of the answers to further improve the accuracy of the answers. In addition, although this paper focused on classical control theory, we would like to investigate on modern control theory and robust control theory in the future, and devise a method to improve response accuracy.

REFERENCES

- [1] OpenAI. Chatgpt. <https://chat.openai.com/>, 2024.
- [2] OpenAI. About. <https://openai.com/about>, 2023.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [4] OpenAI. Gpt-4. <https://openai.com/research/gpt-4>, 2024.
- [5] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [6] Jungo Kasai, Yuhei Kasai, Keisuke Sakaguchi, Yutaro Yamada, and Dragomir Radev. Evaluating gpt-4 and chatgpt on japanese medical licensing examinations. *arXiv preprint arXiv:2303.18027*, 2023.
- [7] John Gruber. Markdown. <https://daringfireball.net/projects/markdown/>, 2004.
- [8] OpenAI. Introducing gpts. <https://openai.com/blog/introducing-gpts>, 2024.
- [9] OpenAI. Openai platform. <https://platform.openai.com/docs/assistants/tools/supported-files>, 2023.
- [10] OpenAI. Prompt engineering. <https://platform.openai.com/docs/guides/prompt-engineering>, 2023.
- [11] OpenAI. Openai devday. <https://devday.openai.com/>, 2023.
- [12] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213, 2022.