In our evaluation process, a control engineer expert carefully assesses the answers generated by GPT or other Large Language Models (LLMs). For each question, if the final answer aligns with the correct solution, it is considered correct. In some cases, while the reasoning steps may not be entirely correct, if the final answer is accurate and doesn't compromise the overall solution, it is still regarded as correct. For the problems in which the final answer is not correct, the type of error is categorized as follows:

- **Calculation Complexity:** The mathematical computation was overly intricate, requiring more information from the user.
- **Calculation Error:** This denotes that mistakes were made during the mathematical computation in the process of solving the problem.
- **Reasoning Error:** This refers to a flaw in the logical progression or inference made while solving the problem.
- **Incorrect Knowledge:** This means that the answer was based on incorrect information or understanding of the problem.
- **Lack of Knowledge:** The solution was hindered by a lack of understanding or familiarity with the relevant concepts.
- **Misreading The Plot:** This refers to problems in which it is necessary to extract data from the provided plots in the statement of the problem, but LLMs could not correctly extract the graphical data.
- **Didn't solve:** The problem was left unsolved mainly by providing a general approach; however, no mathematical approaches are provided.

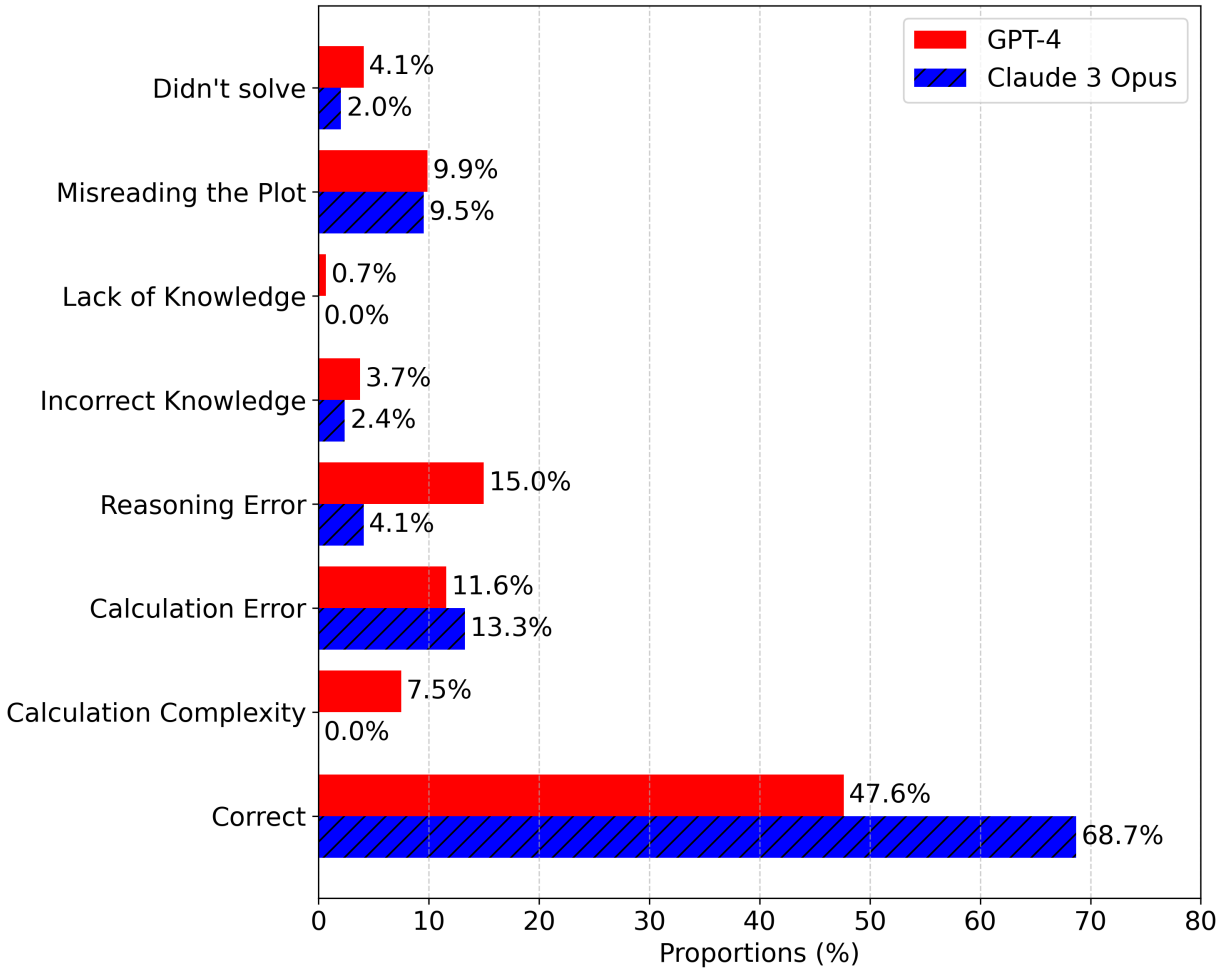
In the process of presenting problems to LLMs, we employed the following statement, which was provided alongside the problem statement, to generate the results.
"Provide a concise yet comprehensive response to the following question, including a confidence level ranging from 0 to 100 for the accuracy of the answer. Additionally, include the LaTeX code for the entire response, encompassing both the answer and the confidence score."

Listed below are the comprehensive topics from which we have extracted questions, covering various aspects of control theory, to compile our dataset. **Table of Contents:**

5. Control System Design
6. Bode Analysis
7. Root-Locus Design
8. Nyquist Design
9. Gain/Phase Margins
10. Advanced Topics (Lyapunov Stability, Controllability and Observability)
11. System Sensitivity Measures
12. Loop-Shaping

13 Summary of Results

13.1 Overall Performance



13.2 Detailed Performances

1. ✓ - Indicates answers that LLMs got correct directly.
2. ✓✗- Represents answers that were initially incorrect, but after prompting for self-correction, LLMs were able to correct them.
3. × - Denotes answers that LLMs could not correct even after an opportunity for self-correction was provided.

1. Differential Equations, Laplace Transform and Preliminaries

	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11	1.12	1.13	1.14	1.15	1.16	1.17	1.18	1.19	1.20	1.21	1.22	1.23	1.24	1.25	1.26	1.27	1.28
GPT-4	×	×	×	✓	✓	×	✓	✓	✓✗	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓	×	✓	✓	×	×
Claude 3	×	✓✗	✓✗	✓	✓	✓✗	✓	✓	×	✓	✓✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
Gemini 1.0 Ultra	×	×	×	×	✓	×	✓	✓	×	✓✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	✓	×	×

Table 2: LLM Performance on Problems 1.1-1.28

2. Stability

	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	2.10	2.11	2.12	2.13	2.14	2.15	2.16	2.17	2.18	2.19
GPT-4	×	✓	×	×	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓	✓	×	×	×
Claude 3	✓✗	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓✗	✓	✓	✓	✓	✓	×	✓
Gemini	×	×	×	×	×	✓	✓	✓	×	×	×	×	×	✓	✓	✓	×	×	×

Table 3: LLM Performance on Problems 2.1-2.19

3. Time Response of Dynamical Systems

	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10	3.11	3.12	3.13	3.14	3.15	3.16	3.17	3.18	3.19	3.20	3.21
GPT-4	×	✓✗	✓	✓	✓✗	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	×	✓	✓	×
Claude 3	×	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	×	✓	✓	×	✓	✓	✓
Gemini 1.0 Ultra	×	×	×	✓	✓✗	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	×	✓	×	✓

Table 4: LLM Performance on Problems 3.1-3.21

4. Block Diagrams

	4.1	4.2	4.3	4.4	4.5
GPT-4	✓	✓	×	×	×
Claude 3	✓✗	✓	✓	×	×
Gemini 1.0 Ultra	×	×	×	×	×

Table 5: LLM Performance on Problems 4.1-4.5

5. Control System Design

	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	5.10	5.11	5.12	5.13	5.14	5.15	5.16	5.17	5.18	5.19	5.20	5.21	5.22	5.23	5.24
GPT-4	✓	×	×	×	×	✓	✓	×	×	×	×	×	×	×	×	×	×	✓	✓	×	×	×	✓	✓
Claude 3	✓	✓	×	×	×	✓	✓	×	×	✓	✓	×	×	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
Gemini 1.0 Ultra	✓	×	×	×	×	✓	✓	×	×	×	×	×	×	×	✓	×	×	✓	✓	✓	✓	×	×	✓

Table 6: LLM Performance on Problems 5.1-5.24

6. Bode Analysis

	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10	6.11	6.12	6.13	6.14	6.15
GPT-4	×	×	×	×	×	×	×	×	×	×	✓	×	×	×	×
Claude 3	×	×	×	×	×	×	×	×	×	×	×	✓	×	×	✓
Gemini 1.0 Ultra	×	×	×	×	×	×	×	×	✓	×	×	×	×	×	×

Table 7: LLM Performance on Problems 6.1-6.15

7. Root-Locus Design

	7.1	7.2	7.3	7.4	7.5	7.6	7.7
GPT-4	×	×	✓	×	×	×	✓
Claude 3	×	✓	✓	×	×	×	✓
Gemini 1.0 Ultra	×	×	✓	×	×	×	✓

Table 8: LLM Performance on Problems 7.1-7.7

8. Nyquist Design

	8.1	8.2	8.3	8.4	8.5
GPT-4	×	×	×	×	×
Claude 3	✓	×	×	×	✓
Gemini 1.0 Ultra	×	×	×	×	×

Table 9: LLM Performance on Problems 8.1-8.5

9. Gain/Phase Margins

	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9
GPT-4	×	✓	✓	✓	✓	✓	×	✓	×
Claude 3	✓	✓	✓	×	✓	✓	×	✓	×
Gemini 1.0 Ultra	×	✓	×	×	✓	✓	×	×	×

Table 10: LLM Performance on Problems 9.1-9.9

10. Advanced Topics (Lyapunov Stability, Controllability and Observability)

	10.1	10.2	10.3	10.4	10.5	10.6	10.7
GPT-4	×	✓	✓	✓	×	✓	✓
Claude 3	×	✓	✓	✓	✓	✓	✓
Gemini 1.0 Ultra	×	×	✓	✓	×	✓	✓

Table 11: LLM Performance on Problems 10.1-10.7

11. System Sensitivity Measures

	11.1	11.2	11.3
GPT-4	✓	✓	✓
Claude 3	✓	✓	✓
Gemini 1.0 Ultra	✓	✓	✓

Table 12: LLM Performance on Problems 11.1-11.3

12. Loop-Shaping

	12.1	12.2	12.3	12.4
GPT-4	×	✓	×	×
Claude 3	✓	✓	×	✓
Gemini 1.0 Ultra	✓	×	×	×

Table 13: LLM Performance on Problems 12.1-12.4