Prompt Engineering with gpt-3.5-turbo:
1. Used prompt 1 and zero shot - 62%, changing randomness/temperature was 64% and then no more than 65.5%
2. Next goal was to use few shot inference for prompt 1
   - Found out the possible domains in the dataset using zero shot inference through gpt-3.5-turbo with prompt 2
   - Calculated number of samples of those domains
   - Took out the domains which had majority of samples to help with the shots from those domains
   - Made a subset of dataset with 10 samples from those domains
   - Number of samples from each domain in the subset depended upon the percentage of each domain present in the original dataset
   - Added 5 from Physics, 2 from Astronomy, 1 from each of Biology, Mathematics, and Chemistry
   - Found the samples indices for a specific domain and used random sampling to choose samples for the subset
   2.1 3-shots….all from Physics: approx 60%
   2.2 3-shots….2 Physics, 1 Astronomy: 60.5%
   2.3 5-shots….3 Physics, 1 Astronomy: 66.3%
   2.4 5-shots….all from Physics: 65.3%
   2.5 1-shot and 2-shots Physics: very poor results, less than random 50%
   2.6 6-shots….2 Physics, 1 of each: 68.9%
3. Used prompt 3 without shots, 65%...didn't try shots, could give comparable results to 2.1-2.6
4. Used prompt 4 (improved!!!) : without shots, 67%
   - Tried 2.6, with different randomness, 70%, 71%...

Prompt Engineering with gpt-4o-mini:
1. Tried prompt 1 and prompt 4 with and without 2.6 : almost the same 80%
2. Improved prompt (prompt 5): zero shot at temp 0: 81.5%
3. Possible ways to improve:
   3.1. I found the domains based on gpt-3.5-turbo, all predictions must not be fully correct, gpt-4o-mini is more reliable so first I would find domains with the same steps I did before.
   3.2. Keep improving prompt is always the option
   3.3. To find the question domains related to false predictions, and see in which domain the majority falls, and add more shots of samples from that domain (All these steps will contribute a bit in improving performance)

Prompt Engineering with gpt-4-turbo:
1. Zero shot with Prompt 5 at temp = 0 : 85.5%
2. Zero shot with Prompt 5 at temp = 0.8743 : 86%

Ensembling technique with majority voting: Used gpt-3.5-turbo, gpt-4o-mini, and gpt-4-turbo to generate answers using Prompt 5 with zero shot at temp 0, and then used majority voting to decide the answers for whole data, and found accuracy : 84%

Note: Experiments were fast; less than 2 minutes

Prompts:
Prompt 1

You are an expert in the scientific field. Answer the following multiple-choice question by selecting the most accurate or likely answer. Be sure to consider the relevant context and principles specific to the science discipline associated with the question, such as Physics, Biology, Astronomy, or any other related field. Answer format: single letter (A, B, C, D, or E).

Prompt 2 (for domains)

You are an expert in various scientific fields. For the multiple-choice question provided, identify the most relevant scientific discipline based on the context and content of the question. Possible disciplines include, but are not limited to: Physics, Biology, Astronomy, Chemistry, Earth Science, Environmental Science, Mathematics, and Medicine. Avoid over-specifying the discipline. Please respond with the name of the most relevant discipline, formatted as a single word or a short phrase.

Prompt 3

I will ask a multiple choice question, with 5 answers A-E. The question falls in a scientific domain. Choose an appropriate domain based on the context of the question and go through each of the 5 options why it is or isn't a good description, and logically find which option you think is most accurate. Do not include any explanations or reasoning in your response. Output should only be the single letter e.g., A or B or C or D or E.

Prompt 4

I will ask a multiple choice question, with 5 answers A-E. Go through each of the 5 options why it is or isn't a good description, and logically find which option you think is most accurate. Do not include any explanations or reasoning in your response. Output should only be the single letter e.g., A or B or C or D or E.

Prompt 5 (the best until now)

You are an expert in the scientific field. Answer the following multiple-choice question by selecting the most accurate or likely answer. Be sure to consider the relevant context and principles specific to the science discipline associated with the question, such as Physics, Biology, Astronomy, or any other related field. Once you give the answer, go back and verify if it is correct, otherwise correct your answer until you are sure. Answer format: single letter (A, B, C, D, or E).