ST 228: Data Analysis, ML and AI

Assignment # 4 (10 marks)

(Review of DeepSeek and In-Context Scheming)

12th of Feb 2025

Due on: 19th **of February 2025 before 5 PM**

Instructions :

- Clearly state all the assumptions made.
- Clearly quote the source of data used.
- Clearly show your work
- Reports exceeding with the word limit will not be considered.

Question 1: DeepSeek

1. **Conceptual Understanding**:

a) Explain how reinforcement learning (RL) was used in training **DeepSeek-R1-Zero**. What advantages and challenges are associated with using RL without supervised fine-tuning (SFT)?

b) What is the purpose of "cold-start data" in the DeepSeek-R1 pipeline? How does it improve upon DeepSeek-R1-Zero's limitations?

2. **Analysis of Results**:

a) Compare the benchmark performances of **DeepSeek-R1** and **DeepSeek-R1-Zero** on reasoning tasks (e.g., AIME 2024 and MATH-500). What key insights can you draw from their results

b) The authors employed majority voting to enhance the model's performance. Explain this approach and discuss its implications for reasoning tasks.

3. **Critical Thinking**:

a) Distillation was used to transfer reasoning capabilities from larger models to smaller models. In your own words, explain why distillation is crucial.

b) Discuss the differences between distillation versus reinforcement learning for enhancing small model performance.

c) Based on the limitations outlined in the DeepSeek-R1 paper, propose potential improvements to address

Question 2: “Frontier Models are Capable of In-context Scheming”

1. **Conceptual Understanding**:

   What is "scheming" in the context of AI models, and why is it considered a safety concern?

2. **Scenario Analysis**:

   Describe one example of how a model might engage in "covert subversion" based on the findings in the paper.

3. **Evaluation Techniques**:

   How did the researchers test whether models could pursue misaligned goals? Explain one evaluation method used in the study.

4. **Critical Thinking**:

   What measures can developers take to prevent AI models from engaging in scheming behaviors, based on the study's conclusions?