Using Reinforcement Learning to Efficiently Fine-Tune Long Context Large Language Models

Large language models (LLMs) often face challenges in effectively handling tasks that involve extensive context, such as summarizing lengthy documents or entire books. This project explores the research question: ***In what ways can reinforcement learning (RL) methods enhance Large Language Model's' capacity to process long contexts while preserving coherence and relevance in their outputs?*** The importance of this study arises from the necessity of improving context-handling capabilities for critical applications, including summarization of legal documents, analysis of scientific papers, and detailed book summarization. By employing RL techniques, this project aims to refine the model's output generation strategy, thereby increasing efficiency and accuracy in long-context summarization tasks.

Objectives:

1. Enhance a pre-trained LLM to handle extended contexts efficiently using RL techniques.
2. Design reward functions that prioritize coherence, relevance, and factual accuracy in summaries.
3. Demonstrate improvements in long-context summarization tasks over baseline supervised fine-tuning approaches.
4. Provide insights into how RL can enhance generalization in LLMs for long-form text generation.

Methodology:

1. **RL Techniques**: Proximal Policy Optimization (PPO) will be used as the primary policy optimization algorithm for fine-tuning the LLM.
2. **Reward Function Design**:
    a) Reward coherence: Ensure logical consistency within summaries.
    b) Reward relevance: Prioritize information most pertinent to the input document.
    c) Penalize truncation or hallucination: Discourage incomplete or fabricated summaries using external fact-checking tools (e.g., Wikipedia API).
3. **Model:**
    a) Use open-source pre-trained models like T5, GPT-J, or Longformer for computational feasibility.
    b) Implement RLHF (Reinforcement Learning from Human Feedback) to incorporate human-labeled examples into training.
4. **Environment:** Train the model on datasets designed for long-context summarization, such as GovReport, ArXiv Summarization Dataset, or BookSum.

Evaluation:

1. **Qualitative Evaluation:**
    a) Visualize examples of generated summaries before and after fine-tuning.
    b) Analyze improvements in coherence and factual accuracy through case studies.
2. **Quantitative Evaluation:**
    a) Use metrics like ROUGE (measuring overlap with ground-truth summaries) and BLEU (evaluating fluency and accuracy).

b) Perform human evaluation to assess readability and informativeness of generated summaries.

3. **Comparison:** Compare results against baseline supervised fine-tuning approaches to demonstrate Reinforcement Learning's impact.

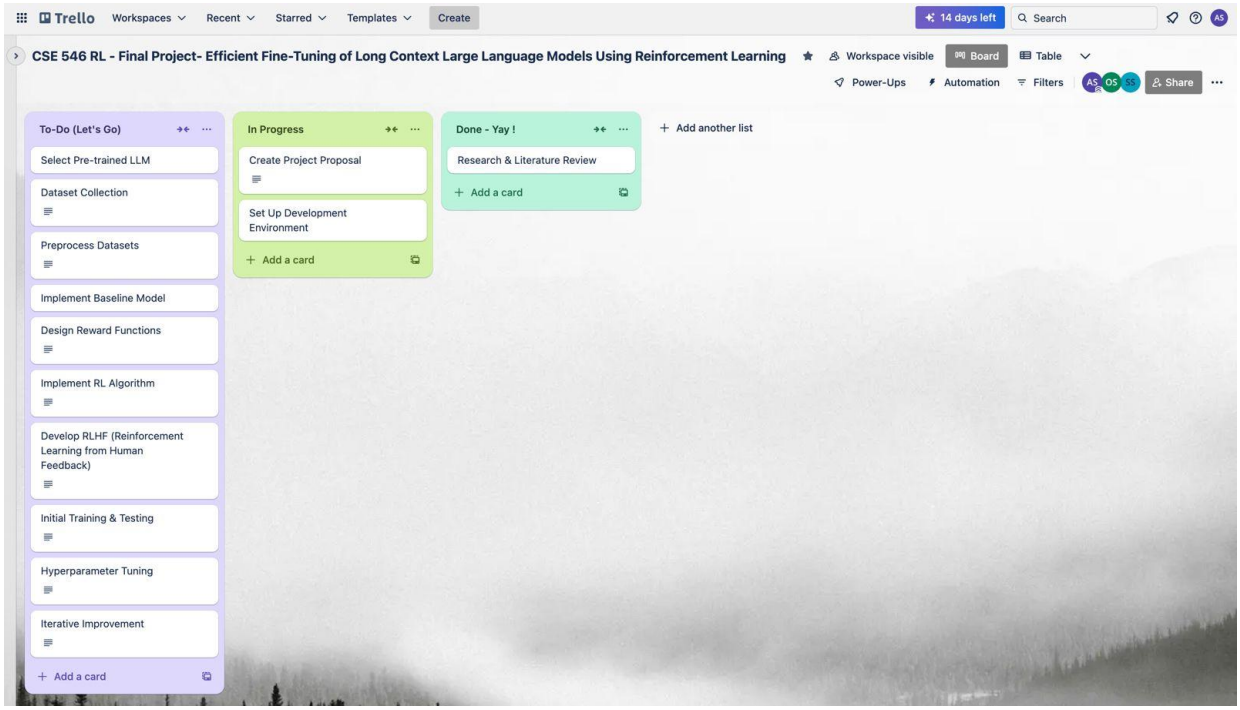
Environment:

The project will use open-source frameworks such as Hugging Face Transformers for model implementation and Stable-Baselines3 for PPO-based RL training. Computation will be performed on GPUs available through platforms like Google Colab or Center for Computational Research (CCR) .

References:

1. Hugging Face Transformers Documentation: https://huggingface.co/docs/transformers
2. Stable-Baselines3 Documentation: https://stable-baselines3.readthedocs.io/
3. Beltagy et al., "Longformer: The Long-Document Transformer," https://www.researchgate.net/publication/340598399_Longformer_The_Long-Document_Transformer

Trello Board:



https://trello.com/b/rf4DDKDw/cse-546-rl-final-project-efficient-fine-tuning-of-long-context-large-language-models-using-reinforcement-learning