# Reinforcement Learning for Adaptive Context Length Optimization in Large Language Models

By

Akash Singh

Om Atul Salunke

Swarat Sarkar

**1. What is a problem that you are addressing using AI methods?**

**Ans.** While many Large Language Models today are capable of processing text like GPT-2, GPT-3 and BERT, these work on fixed context lengths which can lead to excessive resource usage or inefficient model performance. Using a fixed context size may lead to wastage of computational power on simpler tasks whereas taking shorter contexts may lead to loss of critical information.

Our Goal is to thus develop a method that adapts the context length according to the scenarios using RL.

**2. What are the existing solutions in this domain? What are the gaps and limitations in current approaches that your work can address?**

**Ans.** Existing Approaches

- Fixed Context Size – Models like GPT-2, BERT process inputs with a fixed number of tokens.
- Sliding Window Approaches – Models like Longformer extend context dynamically but at a fixed step size.
- Models like LongLoRA use sparse attention mechanisms but are not trained to select the optimal context size per input.

Our approach here tries to optimize both accuracy in the form of ROUGE score and efficiency in terms of tokens/sec.

**3. Share a High-Level Concept for Your Proposed Solution**

**Ans.**

- Train a PPO (Proximal Policy Optimization) RL agent to choose the optimal context length for each input text.
- Use ROUGE score as a performance metric and token efficiency as a reward signal.
- Compare the adaptive model vs. fixed context models to validate improvements.

**4. What are the key AI methods you plan to use?**

**Ans.**

To get preliminary results using the less computational resources we have at our disposal we used –

- LLM - DistilGPT-2 (for fast inference)
- Dataset - (1%) of the CNN/DailyMail dataset
- Tokenization - Hugging Face Tokenizers

- Metrics – ROUGE
- Algorithm - Proximal Policy Optimization (PPO) (from Stable-Baselines3)
- Used a step function that decide the context from a corpus of [50, 100, 150, 250]
- Environment - Custom Gym Environment for Context Length Optimization
- Reward Function – ROUGE Score, Token Efficiency

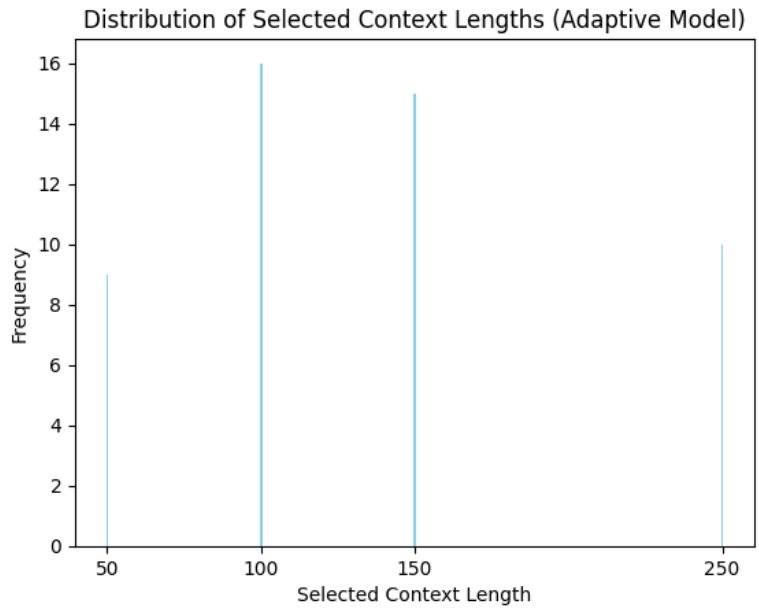
**5. Show any initial results that you have already obtained**

**Ans.**

Context Length and ROUGE Score

```
Average ROUGE Score: 23.7760
Average Selected Context Length: 136.00
```

Distribution of Context Lengths




Results for Fixed Context Lengths

```
Fixed 50 Tokens - ROUGE: 0.1993
Fixed 100 Tokens - ROUGE: 0.2059
Fixed 150 Tokens - ROUGE: 0.1763
Fixed 250 Tokens - ROUGE: 0.1369
```

Efficiency Results

```
Adaptive Model - Tokens/sec: 1318.60, Avg Context Length: 136.00
Fixed 50 Tokens - Tokens/sec: 840.21, Avg Context Length: 50.00
Fixed 100 Tokens - Tokens/sec: 1181.79, Avg Context Length: 100.00
Fixed 150 Tokens - Tokens/sec: 1295.13, Avg Context Length: 150.00
Fixed 250 Tokens - Tokens/sec: 1369.75, Avg Context Length: 249.42
```

### NOTE

These results were surprising for us, and we understand there may be some issues with the tokenization process or bias due to dataset subset selected. We will try to replicate and improve the results.