# Learning to Learn Faster from Human Feedback with Language Model Predictive Control

Group Members:

Ibad Zaidi(20K-0152)

Rayyan Naser(20K-0194)

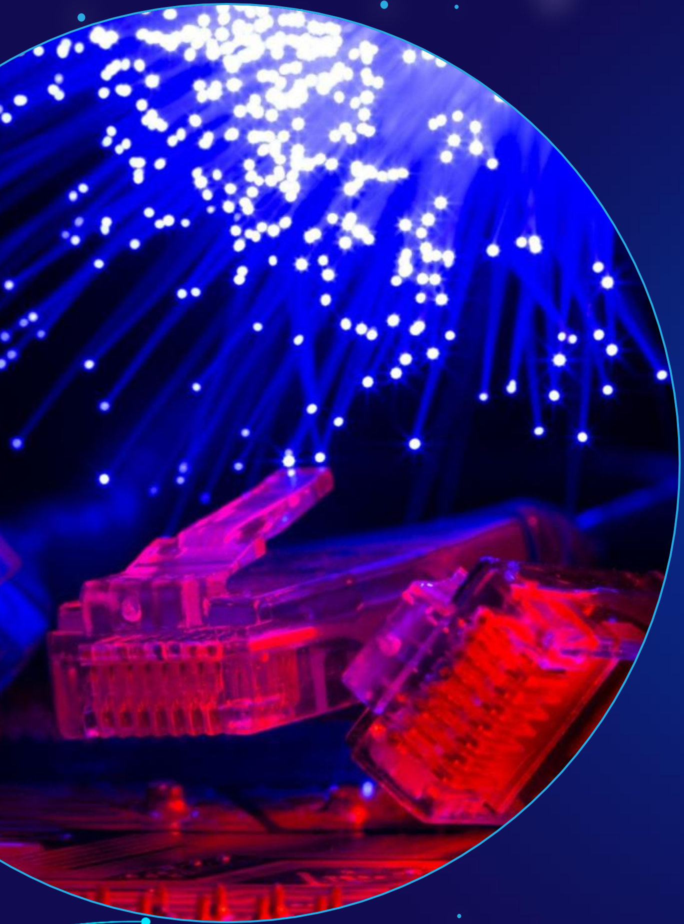
Muhammad Ali(20K-0334)

# Introduction

- Challenge: Teaching robots using natural language poses a significant challenge due to the complexity of translating human commands into actionable robot code.
- Importance for Non-Experts: Enhancing teachability allows individuals without specialized knowledge in robotics or programming to instruct and interact with robots effectively.
- Role of LLMs: Large Language Models (LLMs) show promising capabilities in generating robot code from natural language commands, bridging the communication gap between humans and robots.

# Problem Statement

- Short-term Learning Limitations: Current in-context learning methods are limited by their inability to retain information over extended interactions, leading to the forgetting of previous instructions.
- Need for Memory and Improvement: There's a crucial need for LLMs to remember past interactions and improve over time, ensuring long-term teachability.
- LMPC Introduction: Language Model Predictive Control (LMPC) emerges as a solution to enhance teachability by combining the strengths of in-context learning with the benefits of model fine-tuning.

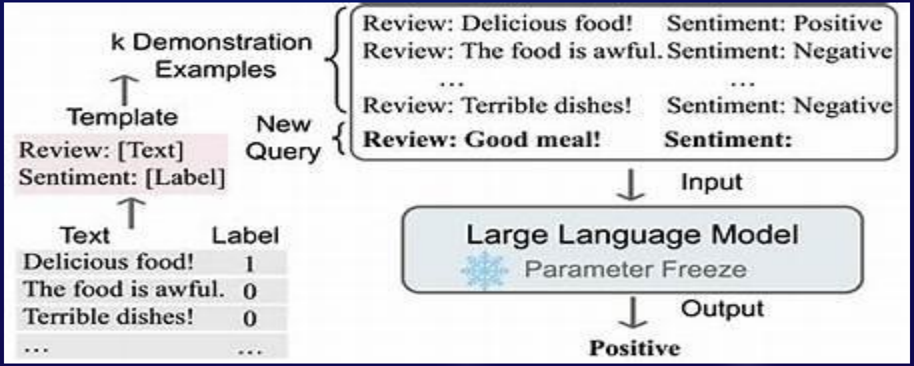


01

# Proposed Solution

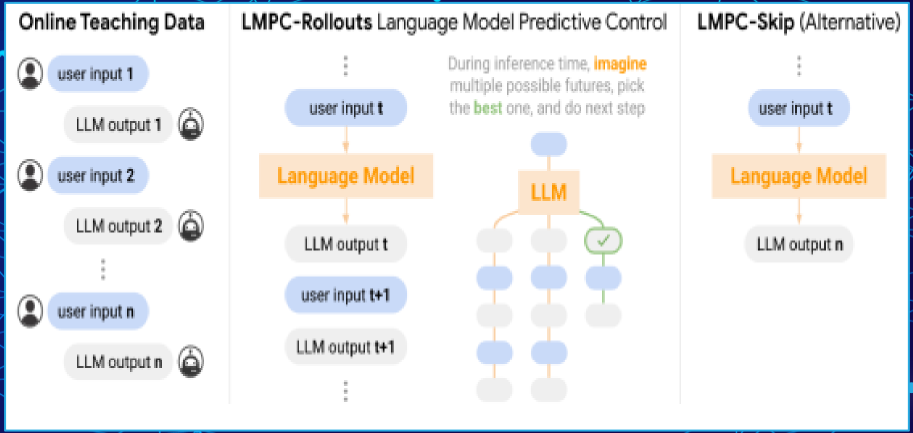
# What is in-context learning?

- In-context learning is the process of  training the model  to understand and generate text by considering the  context of the surrounding words or tokens.
- With respect to this research paper our main objective is to increase the capacity of LLMs to retain information in order to generate robot code with normal language given as input



# LMPC Framework

1. LMPC Concept: LMPC is a novel framework that integrates fast, in-context learning with slow, deliberate model fine-tuning to enhance robot teachability.
2. POMDP Formulation: It conceptualizes human-robot interactions as a Partially Observable Markov Decision Process (POMDP), optimizing the robot's actions based on observed human feedback.
3. Dual Adaptation Approach: This framework advocates for a balanced approach of immediate, in-context adaptation and gradual improvement through model fine-tuning.



# Brilliance of using MDP

- Researchers modelled human robot interactions as  Partially Observable Markov Decision processes (POMDPs).
- Human language serve as input and robot code as output representing actions.
- They trained an LLM to complete previous human robot interactions based on this approach
- They then integrated this model with current existing robotic techniques such as MPCs.
- Resulting in a boost of non expert teaching success rates on unseen tasks by 26.9 percent.

# Data Collection and Evaluation Setup

## User Participation

Engaged 35 non-expert users for data collection, emphasizing the diversity of user interactions.

## Teaching Sessions

Structured sessions focused on natural language feedback, avoiding technical critiques to mimic real-world scenarios

## Task Division

Separated tasks into training and testing sets to evaluate model generalization and teachability on unseen tasks

# Methodology

## Top-User Conditioning

Employs supervised fine-tuning to enhance the model's underlying capabilities based on accumulated interactions

## Slow Adaptation

Employs supervised fine-tuning to enhance the model's underlying capabilities based on accumulated interactions

## LMPC Variants

Introduces two LMPC strategies - LMPC-Rollouts, predicting future interactions, and LMPC-Skip, focusing on predicting the final action directly.

## Fast Adaptation

Utilizes in-context learning for quick adjustments to user feedback, enabling immediate improvements in task execution.

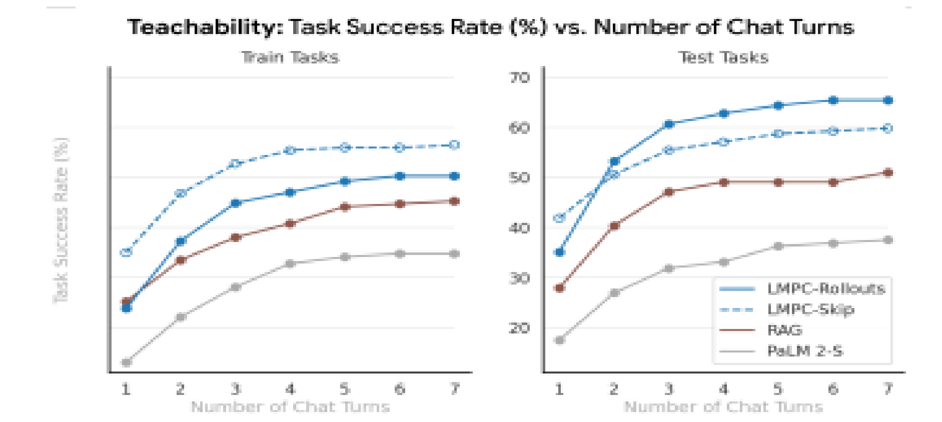
# Experiment Results

## Teachability Improvements:

Demonstrated significant improvements in teachability for unseen tasks and new robot embodiments, validating the effectiveness of LMPC

## Model Comparison

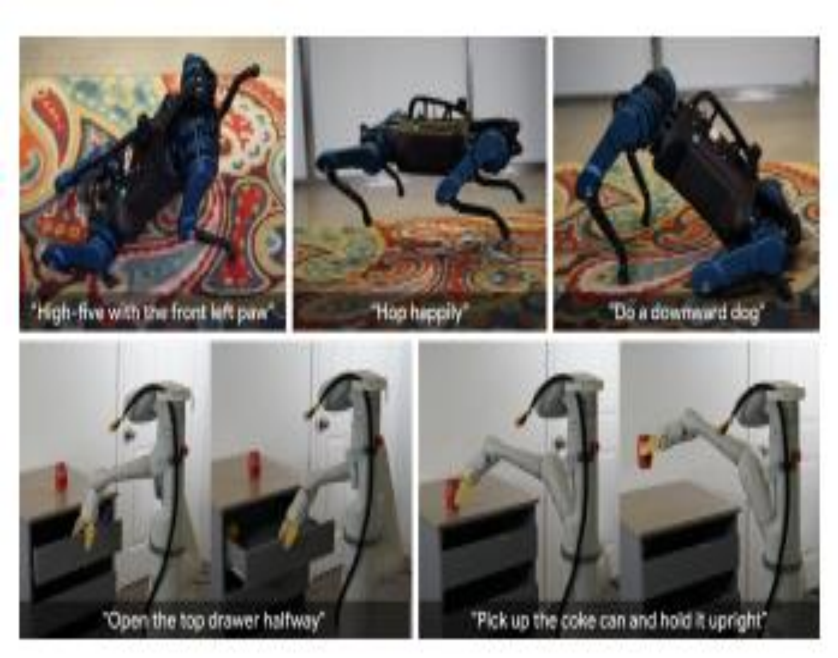
Highlighted the superior performance of LMPC-Rollouts and LMPC-Skip over base models, showcasing the benefits of the LMPC approach



# Real-world Evaluations

## Real-world Application

Successfully applied LMPC to control Mobile Manipulators and Robot Dogs in real-world tasks, proving the framework's practical viability




## User Feedback and Success Rates

Recorded positive user feedback and higher success rates in real-world tasks, emphasizing the practical impact of LMPC on robot teachability

# Limitations of Further Fine Tuning

- Iteration 2 Results: In the second iteration of fine tuning, both LMPC-Skip and LMPC-Rollouts models did not show improved performance, which might indicate a saturation point in learning capabilities under current methodologies.
- Data Distribution Concerns: The similar performance outcomes suggest that the data used in the second round of fine tuning was too similar in distribution to the initial training set. This could prevent the models from learning new patterns or improving upon their existing capabilities.
- Implications: This stagnation emphasizes the need to innovate in how training datasets are compiled and utilized, ensuring that each round of training introduces sufficiently diverse challenges to push the model's learning boundaries.

# Future Research Directions

- Iterative Self-Improvement: Future studies could focus on developing mechanisms that allow LLMs to iteratively improve through cycles of feedback and adjustment based on human interaction, further refining their performance and applicability.
- Efficient Fine Tuning Techniques: Investigating more efficient fine tuning methodologies, such as LoRA, could reduce resource demands and improve scalability and adaptability of the models.
- Synthetic Data Generation: To overcome the limitations of data diversity in training cycles, employing synthetic data generation could be key. This approach would allow the creation of varied and complex datasets, tailored to continually challenge and advance the model's capabilities.