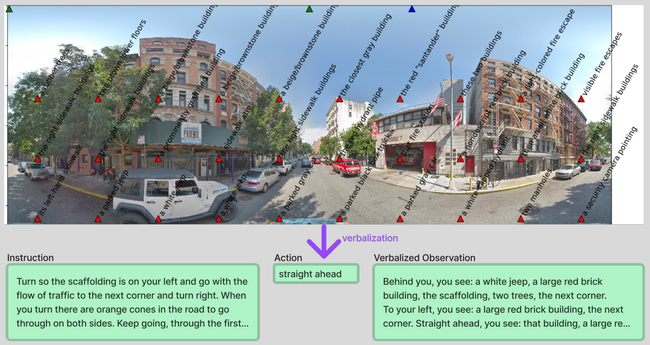
r2llab.com

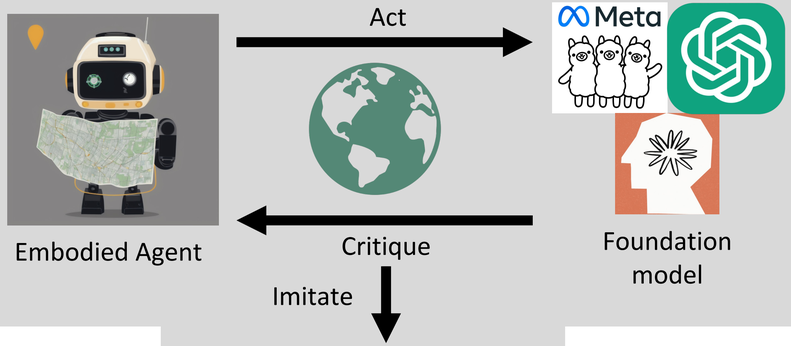
Learning to Plan from Language Feedback Models

Victor Zhong, Assistant Professor, University of Waterloo
Dipendra Misra, Microsoft Research
Eric Yuan, Microsoft Research
Marc-Alexandre Côté, Microsoft Research

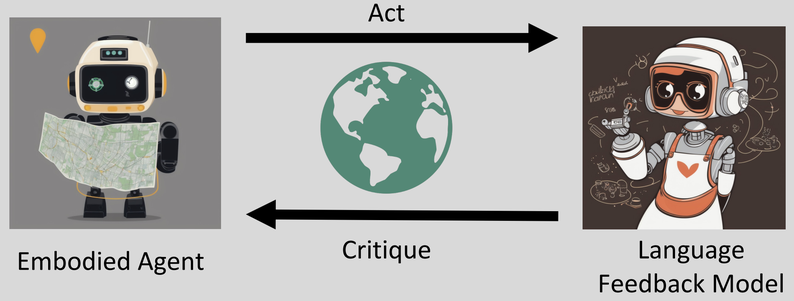
First: verbalize environment



Train time: train small feedback model by imitating foundation model using 1000s of samples



Test time: imitate productive behaviour identified by language feedback model



Agent Prompt
Task: Your task is to grow a apple. This will require growing several plants, and them being crosspollinated to produce fruit…
Obs: This room is called the green house. In it, you see:
  a flower pot 3, a bee hive. The bee hive door is closed…
T-1: You move to the green house.
T-2: The door is already open…
Action: move apple seed (in seed jar, in inventory, in agent, in green house) to flower pot 3 (in green house)

Example: identifying productive behaviour for imitation

Instruction: clean some lettuce and put them in the fridge

Comparison to other learning to plan

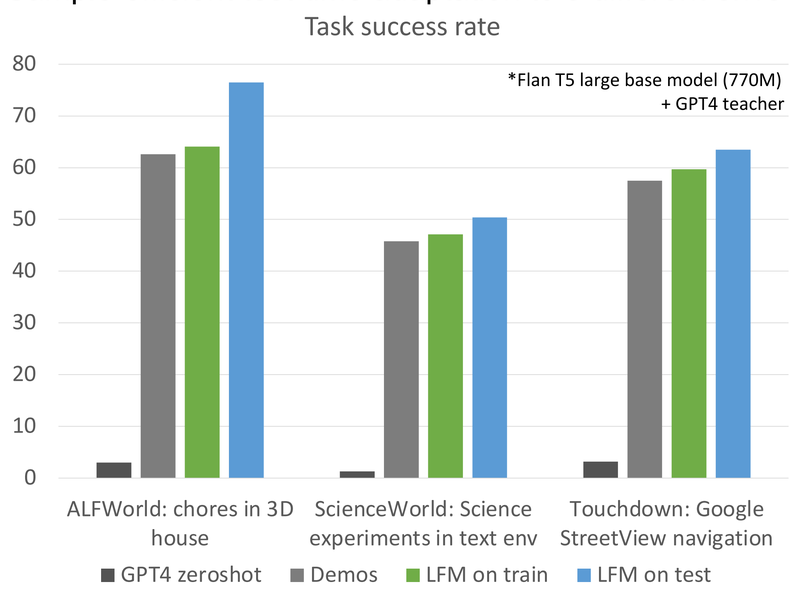
Imitation learning
+ good performance
- step-by-step annotation

Reinforcement learning
+ no step-wise annotation
- many trials for long horizon tasks w/ sparse rewards

Learning from human language feedback
+ easy to annotate
- expensive human intervention

Learning from language feedback models
+ no human annotation necessary (few LLM annotations)
+ few trials required (10ks instead of 1Ms)

Sample-efficient test-time adaptation to 3 different envs



Next: how to learn from open language feedback?

Feedback Prompt
You will be shown a playthrough for solving a task. Task: put two candles in the drawer.
Before: you open drawer 6. Drawer 6 is open…
Step 21 - Your action: close drawer 6. Result: You closed drawer…
Step 22 - Your action…
First, is the player on the right track to solving the task? If so, **list the helpful steps** and **why**… Next, under heading #Summary, summarize in one sentence **what the player is doing at a high level**. Finally, under heading #Improvement, **describe how the player can improve their strategy to solve the task.**

Language feedback
- **Step 28**: **The player locates the candles by arriving at countertop 1, where a candle is available.**
- **Step 29**: **The player successfully takes candle 3 from countertop 1, demonstrating the ability to collect items…**
#Summary
The player is searching the drawers, trying to find candles around…
#Improvement
The player can improve their strategy by:
- not closing drawers unnecessarily…