PROGPROMPT: Generating Situated Robot Task Plans using Large Language Models

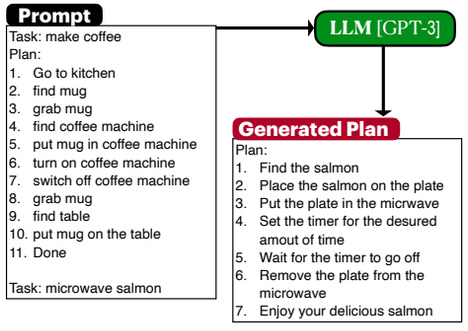
Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay,
Dieter Fox, Jesse Thomason, Animesh Garg

NVIDIA · ICRA · USC Viterbi School of Engineering, Department of Computer Science

Classical Task Planning

- ✗ Requires myriad domain knowledge
- ✗ Large search space, hard to scale
- ✗ Domain specific
- ✗ Requires concrete goal specification

Planning with LLMs



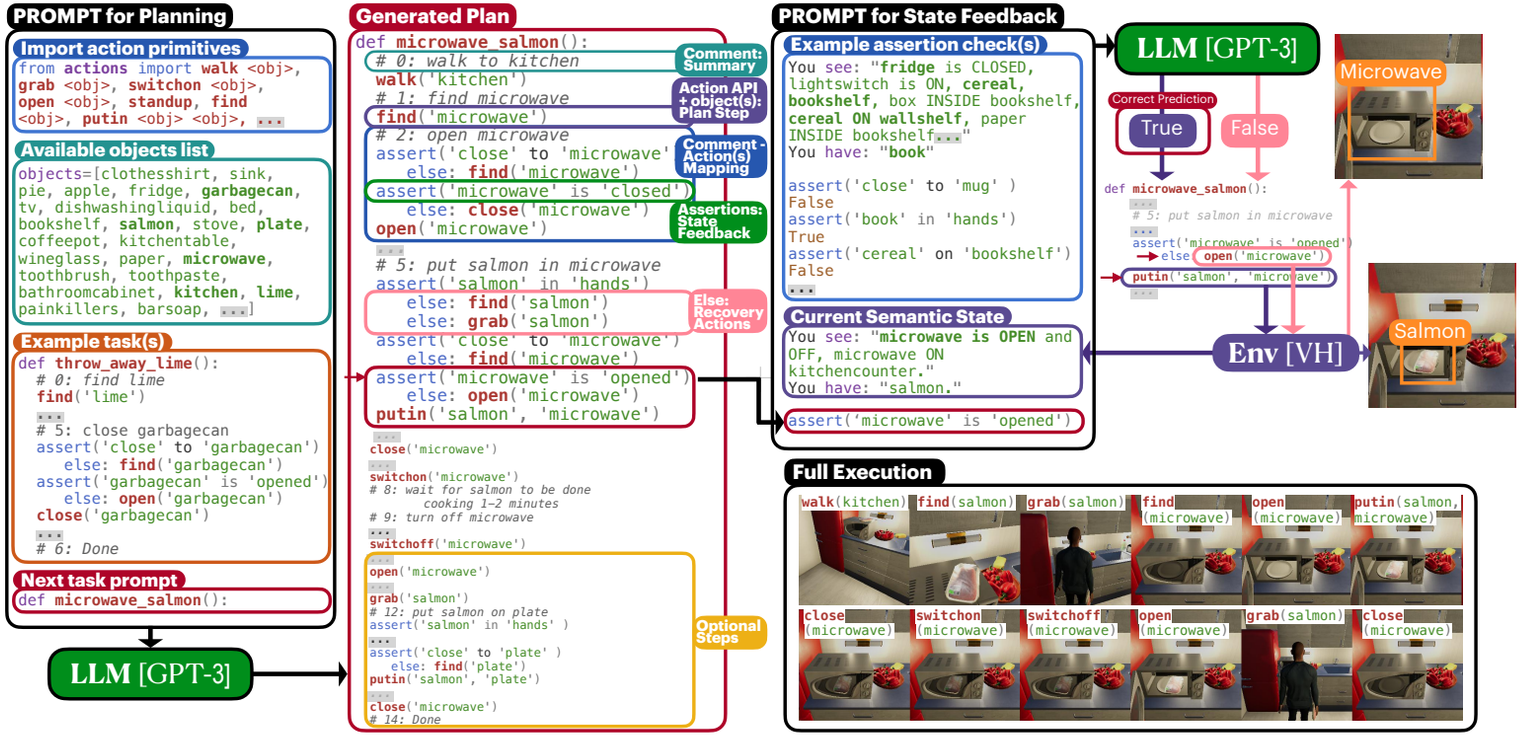
- ✗ LLM is not situated in the scene
- ✗ Plan steps using unavailable actions and objects
- ✗ Text-to-robot action mapping may not be trivial
- ✗ Combinatorial admissible action space

KEY TAKEAWAY

We present a **programmatic LLM prompt structure** that enables **plan generation** functional across **situated environments, robot capabilities, and tasks**.

progprompt.github.io

Method: Demo in Virtual Home



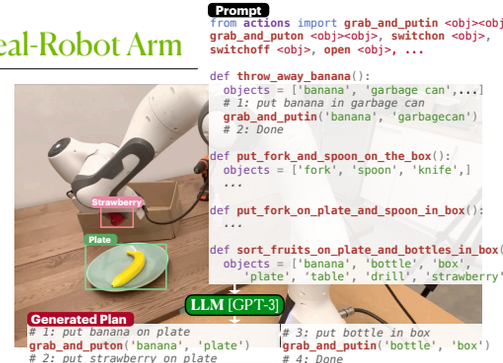
Results

VirtualHome Simulator

#	— Prompt Format and Parameters — Format	Comments	Feedback	LLM Backbone	SR	Exec	GCR
1	ProgPrompt	✓	✓	Codex	**0.40**±0.11	**0.90**±0.05	**0.72**±0.09
2	ProgPrompt	✓	✓	Davinci	0.22±0.04	0.60±0.04	0.46±0.04
3	ProgPrompt	✓	✓	GPT3	0.34±0.08	0.84±0.01	0.65±0.05
4	ProgPrompt	✓	✗	GPT3	0.28±0.04	0.82±0.01	0.56±0.02
5	ProgPrompt	✗	✓	GPT3	0.30±0.00	0.65±0.01	0.58±0.02
6	ProgPrompt	✗	✗	GPT3	0.18±0.04	0.68±0.01	0.42±0.02
7	LangPrompt	-	-	GPT3	0.00±0.00	0.36±0.00	0.42±0.02
8	Baseline from Huang et al. [2]			GPT3	0.00±0.00	0.45±0.03	0.21±0.03

VH Scene	SR	Exec	GCR
Env-0	0.34±0.08	0.84±0.01	0.65±0.05
Env-1	0.56±0.08	0.85±0.02	0.81±0.07
Env-2	0.56±0.05	0.85±0.03	0.72±0.09
Average	0.48±0.13	0.85±0.02	0.73±0.10

Real-Robot Arm



Task Description	Distractors	SR	Plan SR	GCR
put the banana in the bowl	0	1	1	1/1
	4	1	1	1/1
put the pear on the plate	0	1	1	1/1
	4	1	1	1/1
put the banana on the plate and the pear in the bowl	0	1	1	2/2
	3	1	1	2/2
sort the fruits on the plate and the bottles in the box	0	0	1	2/3
	1	1	1	3/3
	2	0	0	2/3

- ✓ LLM situated in the scene
- ✓ (1:[0,n]) text-robot action mapping
- ✓ Plans restricted to available actions and objects
- ✓ Use LLM's commonsense to compose action to handle combinatorial action space
- ✓ Generalizes to new tasks, scenes, and robots