

# Cook2LTL: Translating Cooking Recipes to LTL Formulae using Large Language Models

Prompt

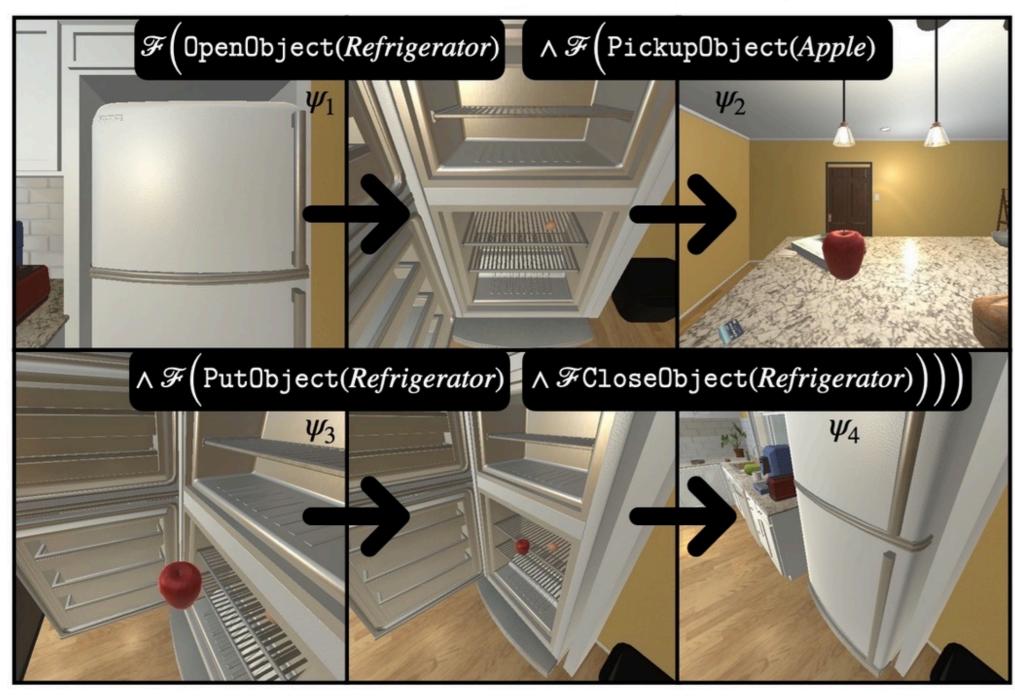


Angelos Mavrogiannis<sup>1</sup>, Christoforos Mavrogiannis<sup>2</sup>, Yiannis Aloimonos<sup>1</sup>

<sup>1</sup>Department of Computer Science, University of Maryland, College Park

<sup>2</sup>Department of Robotics, University of Michigan

 $\phi = \mathscr{F}\texttt{Refrigerate}(Apple) = \mathscr{F}(\psi_1 \wedge \mathscr{F}(\psi_2 \wedge \mathscr{F}(\psi_3 \wedge \mathscr{F}\psi_4)))$ 



**Goal**: Given a cooking recipe in the form of natural language, extract unambiguous robot-executable plans with actions that are admissible in a kitchen environment.

### LLM Action Reduction

- Following ProgPrompt [4], we prompt an LLM with a pythonic import of the admissible actions in the environment and two example task plans in the form of pythonic functions.
- Once acquiring the plan for a newly seen action, we add the action to the import to enable model to invoke it in subsequent executions.

boil <obj><obj>

LLM Output

**Primitive Actions** 

rom actions import ((stir <obj>, cut <obj>, pour <obj><obj>, turn <obj><obj>, shake <obj>, pick up <obj>,put <obj><obj>, remove <obj><obj>, open <obj>, close <obj>, turn on <obj>, turn off <obj>, taste <obj>, wait <time>)) def bake(cake: what, oven: where, 30 minutes: time) boil(eggs: what, pan: where): def cook(pasta: what): # put water in the pot # pick up the cake # put water in the pot pick up(cake) put(water, pot) put(water, pot) # pick up pot # put the cake on the baking pan # pick up the pot pick up(pot) put(cake, baking pan) pick up(pot) # put pot on stovetop # open the over # put the pot on the stovetop put(pot, stovetop) open (oven) put(pot, stovetop) # turn on stovetop # put the baking pan in the oven # turn on the stovetop turn on(stovetop) put(baking pan, oven) turn on(stovetop) # wait until the water boils # close the over # wait until the water is boiling wait(water==boiled) close(oven) wait(water\_is\_boiling) # pick up the pasta # turn on the over # put the eggs in the pot pick up(pasta) turn on(oven) put(eggs, pot) # put the pasta in the pot # wait for 30 minutes # wait until the eggs put(pasta, pot) wait(timer==30 minutes)

# Challenges

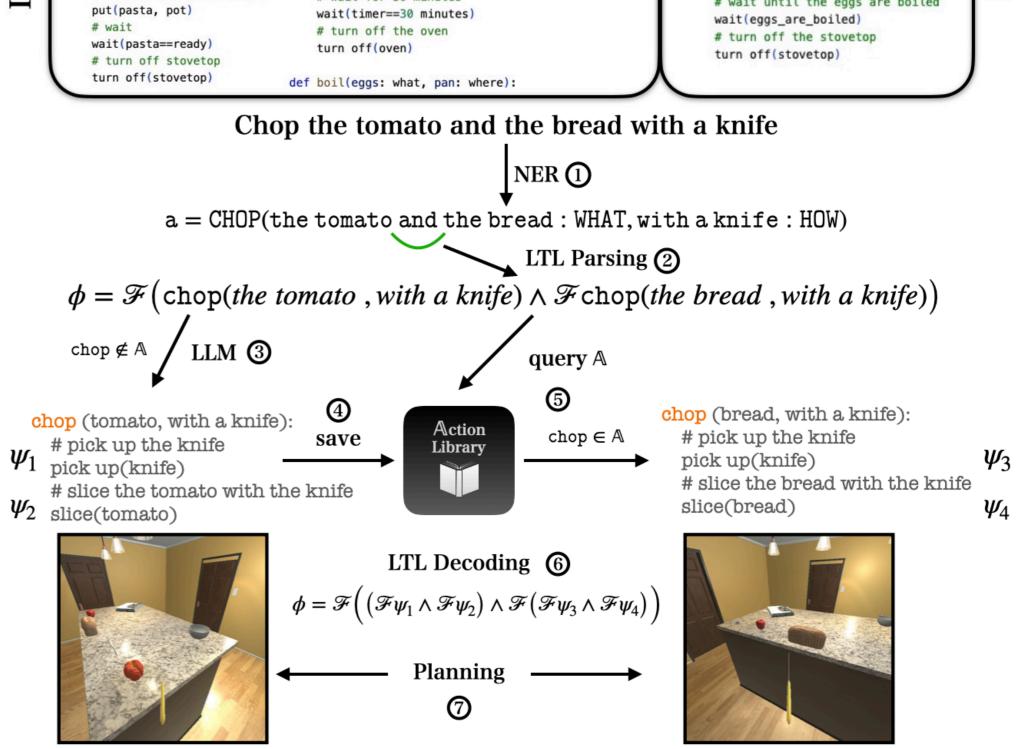
- Cooking poses a unique set of challenges to robots [1].
- Natural language has a practically infinite space of actions, while robots can only execute a small set of actions.
- The language of recipes is ambiguous, with contextimplicit parts of speech, underspecified tasks, and explicit sequencing language (e.g. until, before) [2].

### Approach

- Semantically parse a recipe  ${\cal r}$  into a function representation for every detected high-level action.
- Reduce each high-level action  $a \notin \mathscr{A}$  to a combination of primitive actions from  $\mathscr{A}$ .
- Cache the action reduction policy to an action library (A) for future use.
- Translate r into an LTL formula  $\phi$  with function representations as atomic propositions.

## Named Entity Recognition (NER)

• Annotate subset of Recipe1M+ dataset [3] with salient



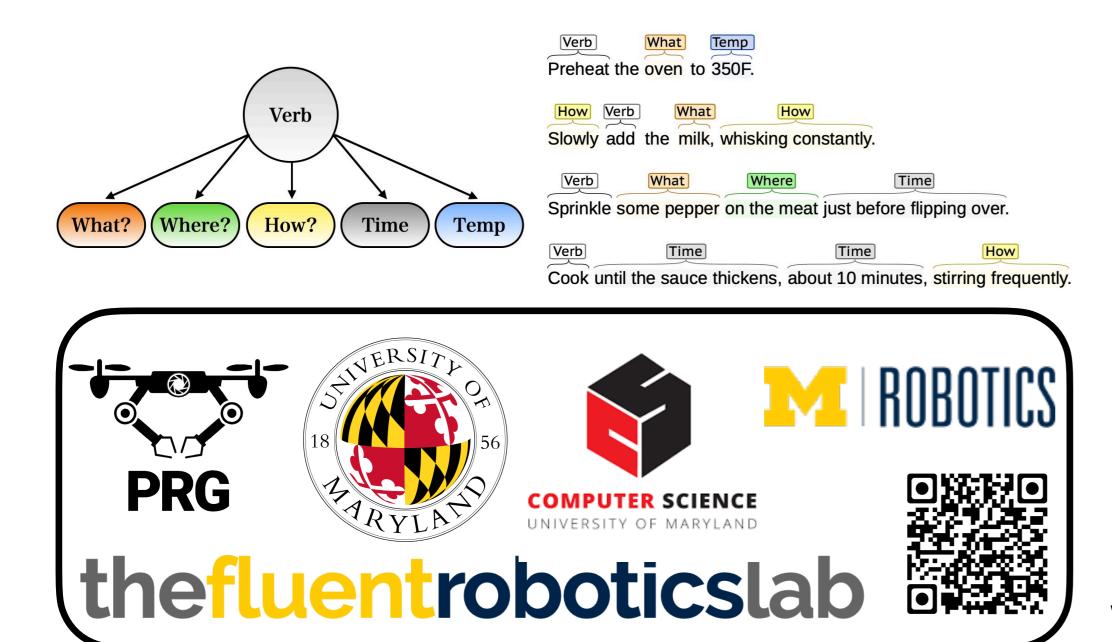
	Active Modules		
Metric	AR*	AR	Cook2LTL (AR+ $\mathbb{A}$ )
Executability (%)	$0.91 \pm 0.01$	$0.92 \pm 0.01$	$\boldsymbol{0.94 \pm 0.01}$
Time (min)	$14.85 \pm 1.05$	$9.89 \pm 0.46$	$6.05 \pm 0.12$
Cost (\$)	$0.19 \pm 0.01$	$0.16 \pm 0.00$	$0.11 \pm 0.00$
API calls (#)	$275\pm0.00$	$231\pm0.00$	$134 \pm 0.00$

#### Results

 We simulate Cook2LTL (AR+A) on held out Recipe1M+ recipes and observe that it decreases LLM API calls (-51%), Latency (-59%), and Cost (-42%) compared to a

categories  ${\mathscr C}$  of an action.

• Fine-tune a BERT NER model to predict  $\mathscr{C}$ .



baseline system (AR\*) that queries the LLM for every newly encountered action at runtime (See table above).

• Additional simulations on 4 simple cooking tasks in an AI2-THOR [5] kitchen show that Cook2LTL is still more time-efficient but fails when the 1st LLM-generated plan is incorrect.

#### <u>References</u>

[1] Bollini et al. Interpreting and executing recipes with a cooking robot. Experimental Robotics 2013.

[2] Malamud et al. Cooking with Semantics. ACL 2014.

[3] Marin et al. A dataset for learning cross-modal embeddings for cooking recipes and food images. IEEE TPAMI 2019.

[4] Singh et al. ProgPrompt: Generating Situated Robot Task Plans using Large Language Models. CoRL 2021.

[5] Kolve et al. Ai2-THOR: An Interactive 3D environment for visual AI. RSS 2021.

We acknowledge the support of ARL and the NSF-funded NeuroPAC program.