Large Language Models for Autonomous Robot Planning and Acting

Research Talk at Randolph College

Yue Cao

Purdue University, ECE

yuecao@purdue.edu

Overview

1. Real Autonomous Robots

2. LLM for Planning

3. LLM for Acting

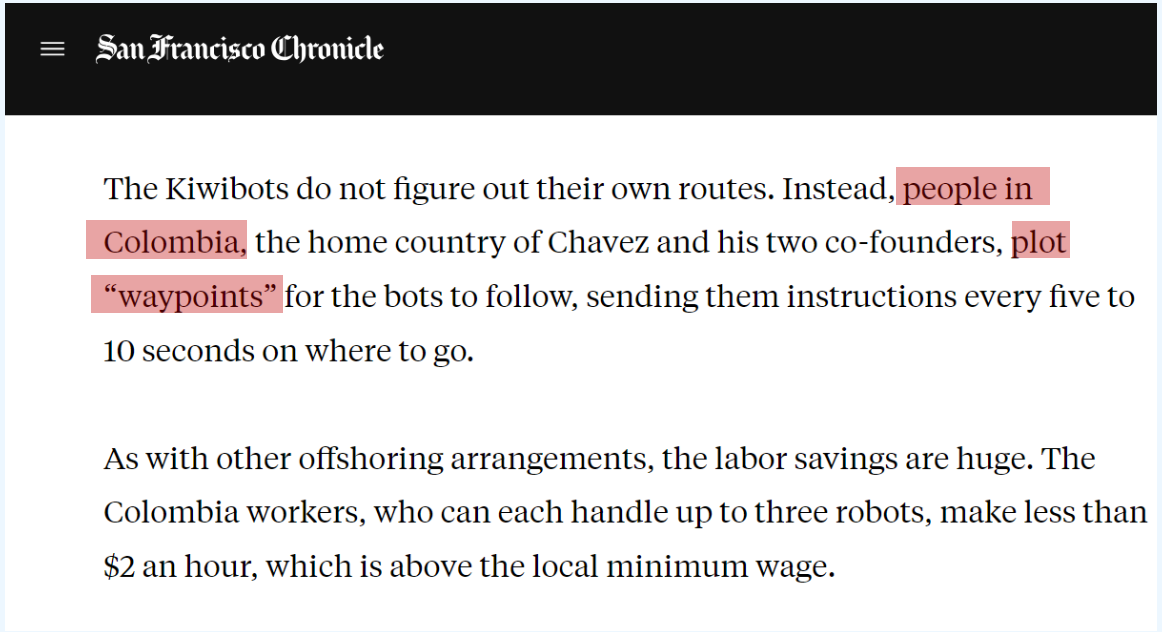
Fake Autonomous Robots

Kiwibot 2018: food delivery at UC Berkeley campus

Fake Autonomous Robots

Kiwibot 2018: food delivery at UC Berkeley campus

What is actually behind?



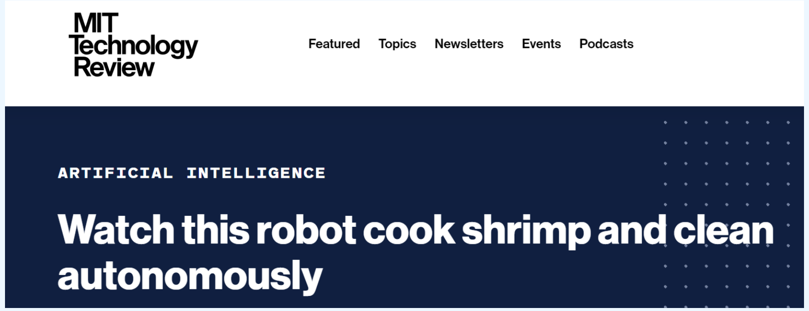
The Kiwibots do not figure out their own routes. Instead, people in Colombia, the home country of Chavez and his two co-founders, plot "waypoints" for the bots to follow, sending them instructions every five to 10 seconds on where to go.

As with other offshoring arrangements, the labor savings are huge. The Colombia workers, who can each handle up to three robots, make less than $2 an hour, which is above the local minimum wage.

"Kiwibots win fans at UC Berkeley as they deliver fast food at slow speeds", San Francisco Chronicle, 2019

Fake Autonomous Robots

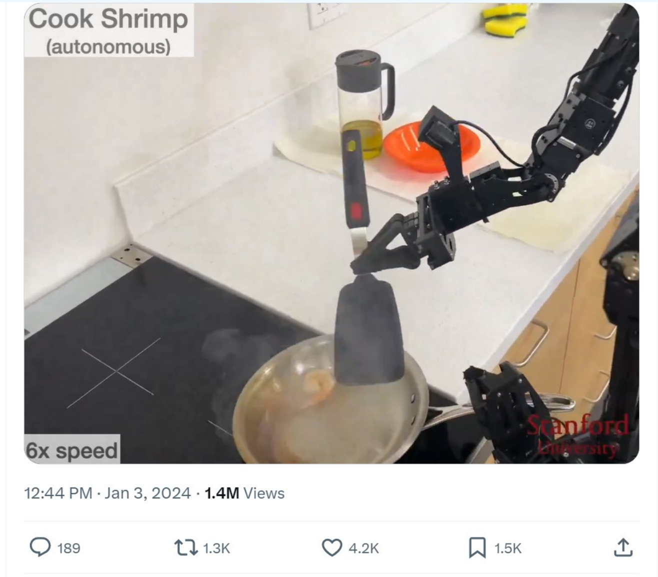
Mobile ALOHA 2024: Stanford


MIT Technology Review

Featured Topics Newsletters Events Podcasts

ARTIFICIAL INTELLIGENCE

Watch this robot cook shrimp and clean autonomously

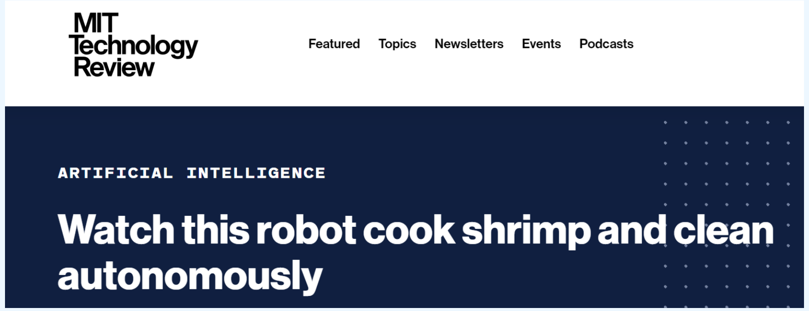


12:44 PM · Jan 3, 2024 · **1.4M** Views

189 1.3K 4.2K 1.5K

Fake Autonomous Robots

Mobile ALOHA 2024: Stanford

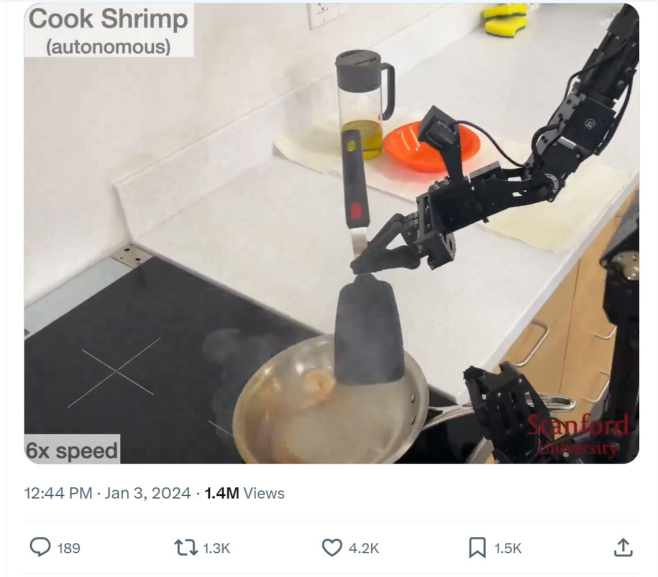


MIT Technology Review

Featured Topics Newsletters Events Podcasts

ARTIFICIAL INTELLIGENCE

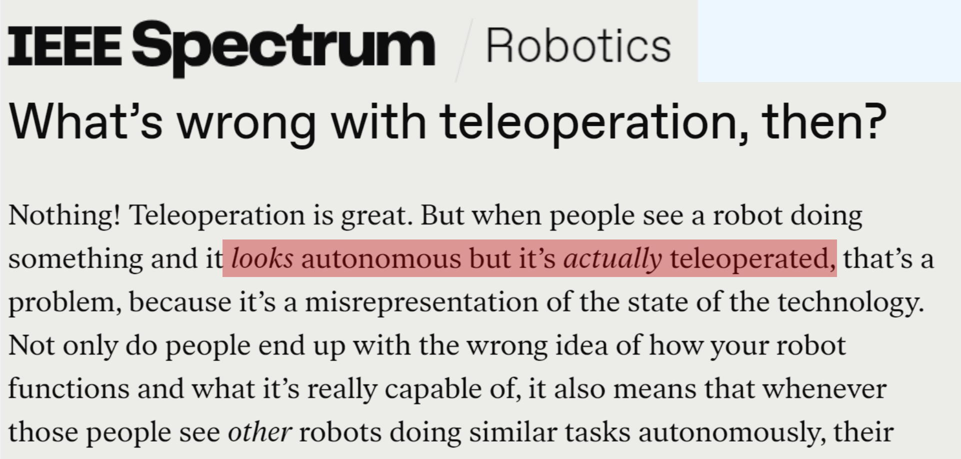
Watch this robot cook shrimp and clean autonomously



12:44 PM · Jan 3, 2024 · **1.4M** Views

189 1.3K 4.2K 1.5K

What is actually behind?



IEEE Spectrum Robotics

What's wrong with teleoperation, then?

Nothing! Teleoperation is great. But when people see a robot doing something and it *looks* autonomous but it's *actually* teleoperated, that's a problem, because it's a misrepresentation of the state of the technology. Not only do people end up with the wrong idea of how your robot functions and what it's really capable of, it also means that whenever those people see *other* robots doing similar tasks autonomously, their

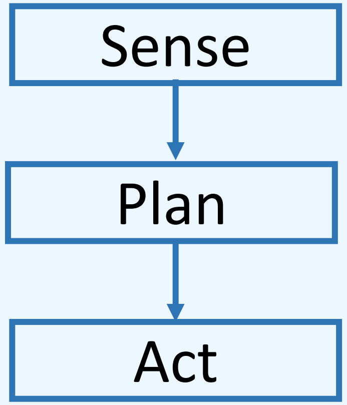
"That Awesome Robot Demo Could Have a Human in the Loop", IEEE Spectrum, 2024

Towards Real Autonomous Robots

Two fundamental questions

1. New technology?
   Artificial intelligence (AI)

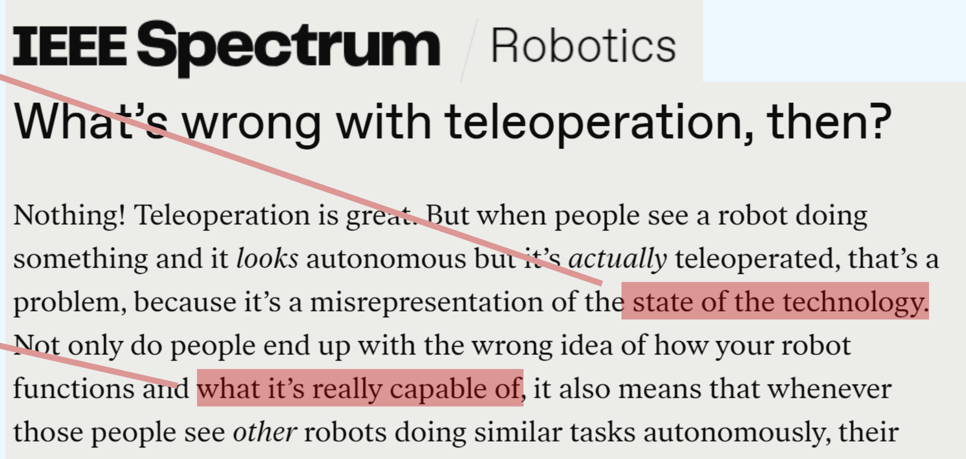
2. Robot capability?



Sense

Plan

Act



IEEE Spectrum / Robotics

What's wrong with teleoperation, then?

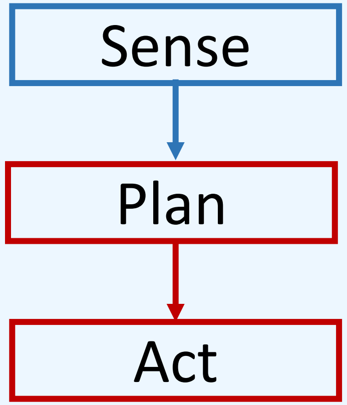
Nothing! Teleoperation is great. But when people see a robot doing something and it *looks* autonomous but it's *actually* teleoperated, that's a problem, because it's a misrepresentation of the state of the technology. Not only do people end up with the wrong idea of how your robot functions and what it's really capable of, it also means that whenever those people see *other* robots doing similar tasks autonomously, their

Towards Real Autonomous Robots

Two fundamental questions

1. New technology?
   Artificial intelligence (AI)

2. Robot capability?



Sense

Plan

Act

My Research Focus

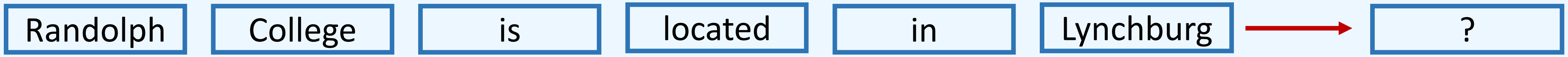
Large Language Models

For

Planning and Acting

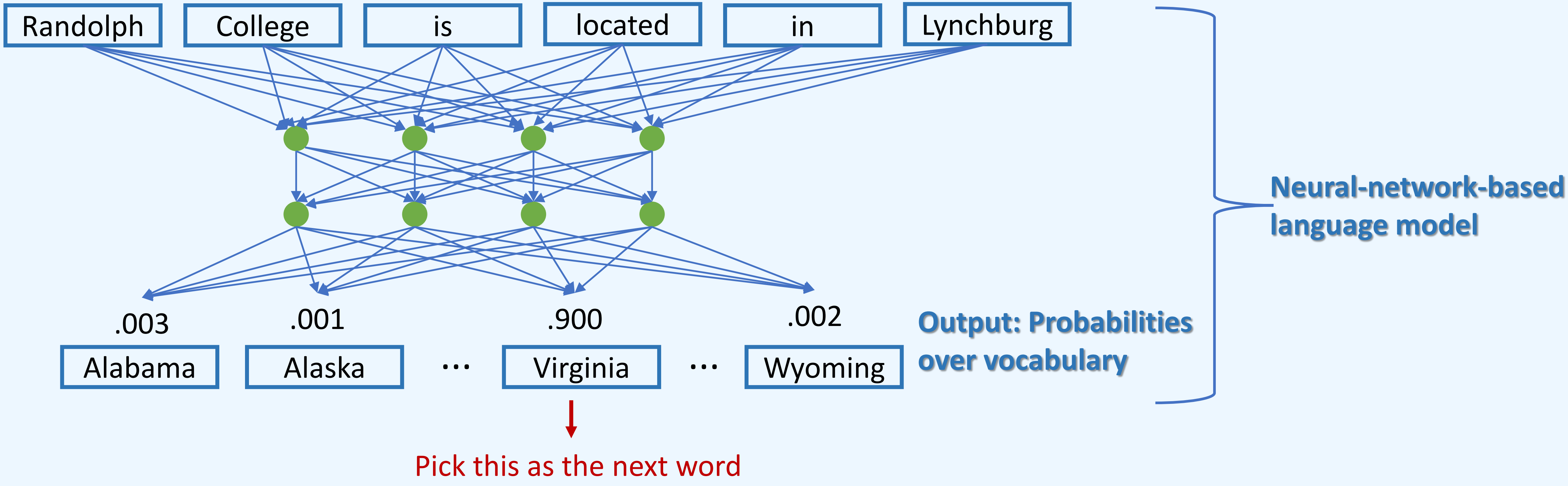
Add-on: What is the Large Language Model?

Next word prediction problem



Randolph College is located in Lynchburg → ?

Add-on: What is the Large Language Model?



Randolph College is located in Lynchburg

.003 Alabama

.001 Alaska

…

.900 Virginia

…

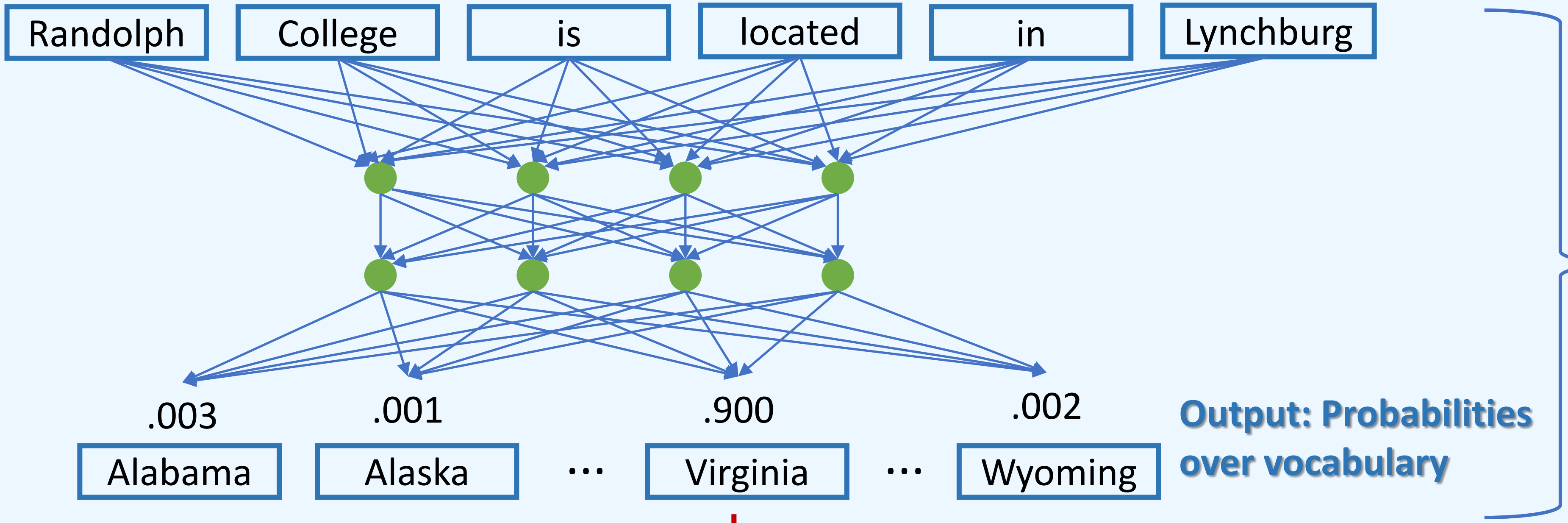
.002 Wyoming

Pick this as the next word

Output: Probabilities over vocabulary

Neural-network-based language model

Add-on: What is the Large Language Model?



Randolph College is located in Lynchburg

.003 Alabama

.001 Alaska

…

.900 Virginia

…

.002 Wyoming

Output: Probabilities over vocabulary

Pick this as the next word

Neural-network-based language model

Large Language Model: has lots of parameters, trained on huge amount of data

Overview

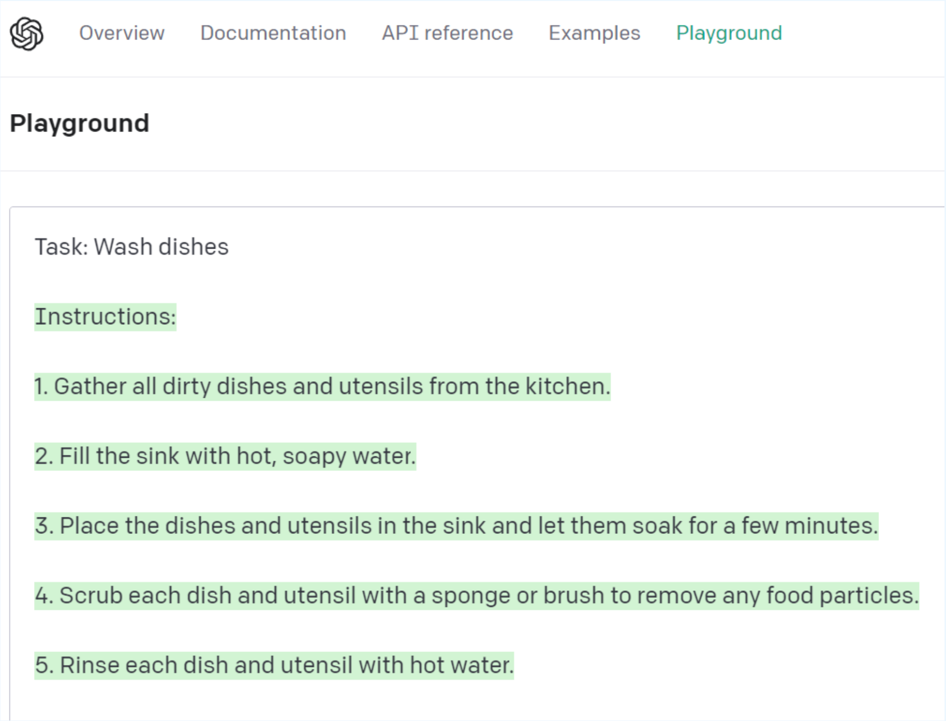
1. Real Autonomous Robots

2. LLM for Planning

3. LLM for Acting

LLM for Planning

GPT can do planning for human activities



Playground

Task: Wash dishes

Instructions:

1. Gather all dirty dishes and utensils from the kitchen.

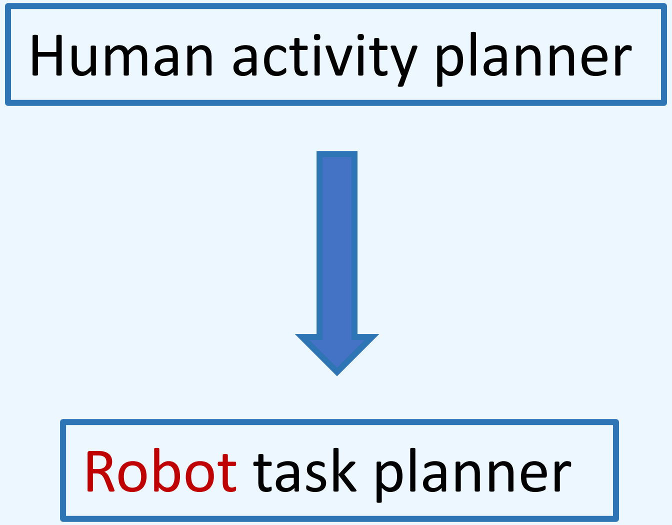
2. Fill the sink with hot, soapy water.

3. Place the dishes and utensils in the sink and let them soak for a few minutes.

4. Scrub each dish and utensil with a sponge or brush to remove any food particles.

5. Rinse each dish and utensil with hot water.

How about using LLMs in robot tasks?

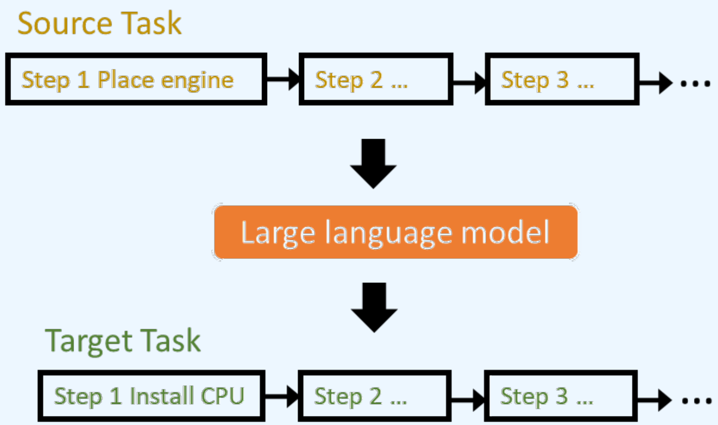

Human activity planner

Robot task planner

LLM for Planning

Previous works:
Generated task are sequential

Can we generate modular tasks?
Better reusability and readability



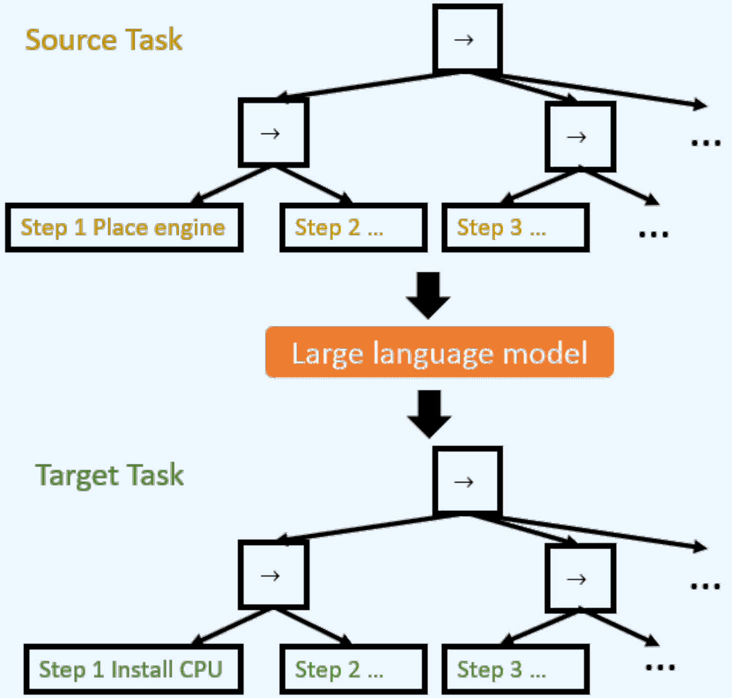
Source Task

Step 1 Place engine → Step 2 ... → Step 3 ... → ...

Large language model

Target Task

Step 1 Install CPU → Step 2 ... → Step 3 ... → ...



Source Task

→

→ →

Step 1 Place engine Step 2 ... Step 3 ... ...

Large language model

Target Task

→

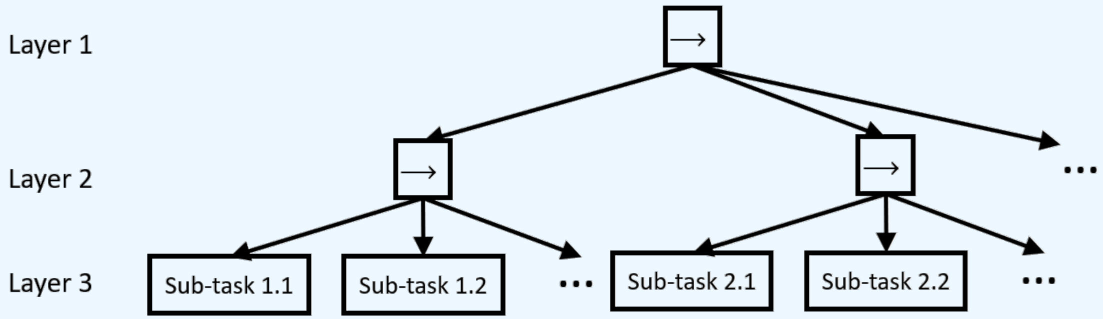
→ →

Step 1 Install CPU Step 2 ... Step 3 ... ...

Yue Cao and C. S. George Lee, "Robot Behavior-Tree-Based Task Generation with Large Language Models",
AAAI Spring Symposium, 2023

LLM for Planning

- Behavior-Tree Task Generation with LLMs
  Prompt Design: Phase-Step Prompt



Source Task
Procedures:
Phase 1.
Step 1. [Sub-task 1.1] $_X$; Step 2. [Sub-task 1.2] $_X$; ...
Phase 2.
Step 1. [Sub-task 2.1] $_X$; Step 2. [Sub-task 2.2] $_X$; ...
...
Target Task: Task Description
Procedures:
[Phase 1.
Step 1. Sub-task ?; Step 2. Sub-task ?; ...
Phase 2.
Step 1. Sub-task ?; Step 2. Sub-task ?; ...
...] $_Y$

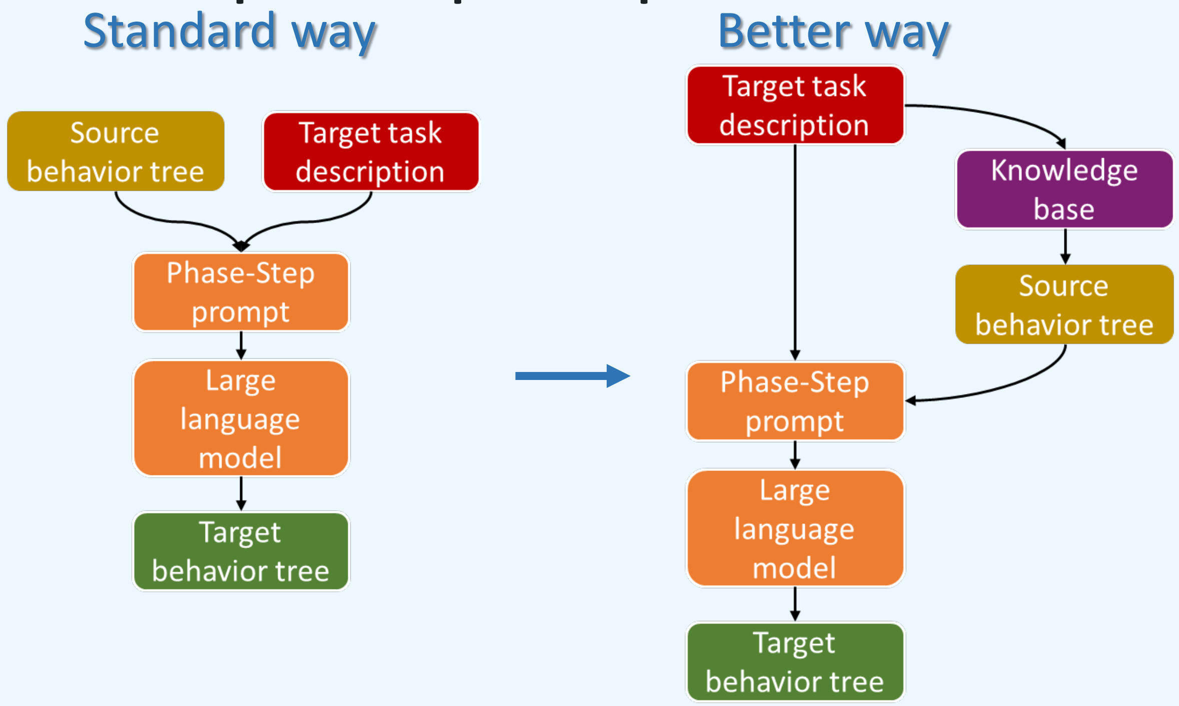
Layer 2 node: Phase; Layer 3 node: Step

LLM for Planning

- Use domain knowledge to improve prompt

Source Task
Procedures:
Phase 1.
Step 1. Put car at a conveyor;
Step 2. Lift the car.
Phase 2.
Step 1. Pick the wheel;
Step 2. Approach conveyor;
Step 3. Align wheel with wheel hub.
Phase 3.
Step 1. Insert screws;
Step 2. Fasten screws;
Step 3. Leave the conveyor.

Target Task: Desktop assembly
Procedures:
To be generated …



Integrate knowledge base retrieval into the pipeline

The end-user just needs to input a short target task description. The source behavior tree can be automatically retrieved from a knowledge base.

LLM for Planning

- How to retrieve similar tasks from a knowledge base?

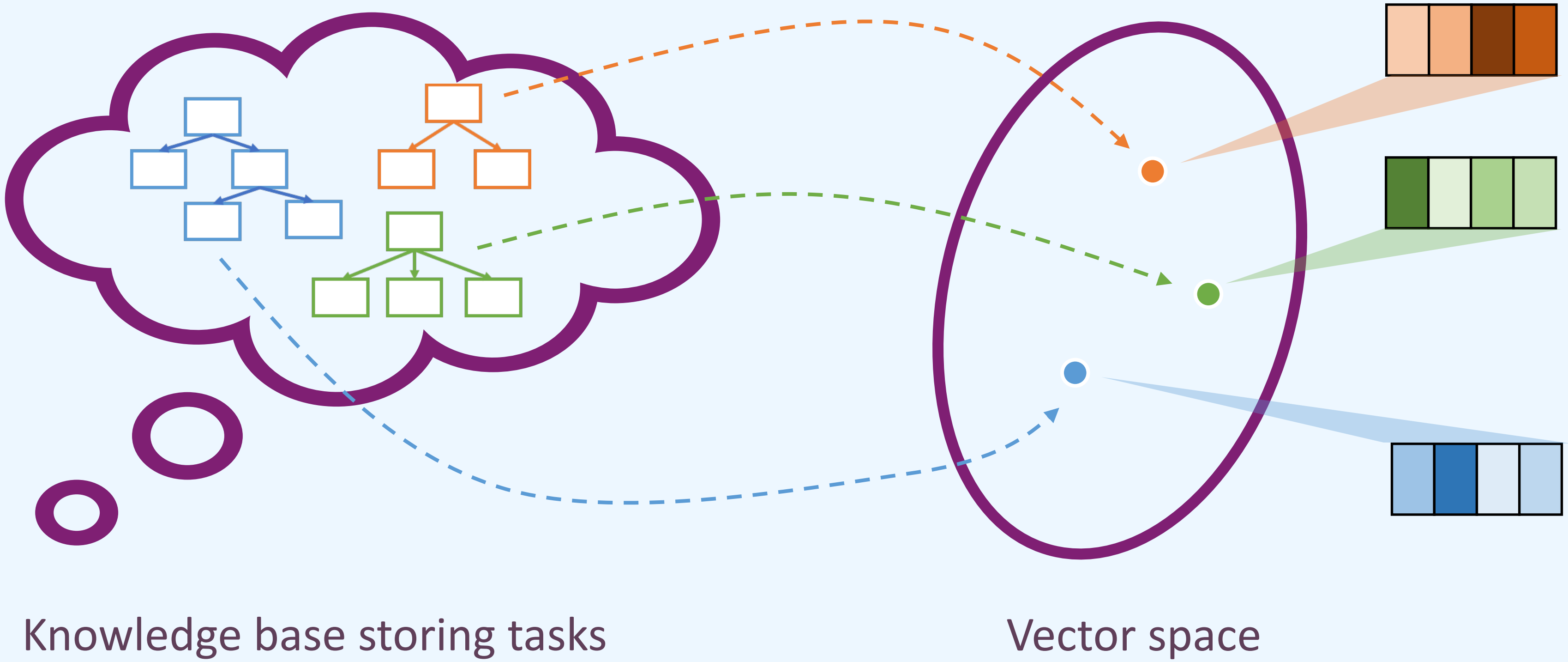
- Solution:
  Embedding, mapping an entity into a fixed length vector.

↓

word, document, Amazon product, Netflix movie…(Entity2Vec)

LLM for Planning

- Vector Embeddings for behavior-tree tasks

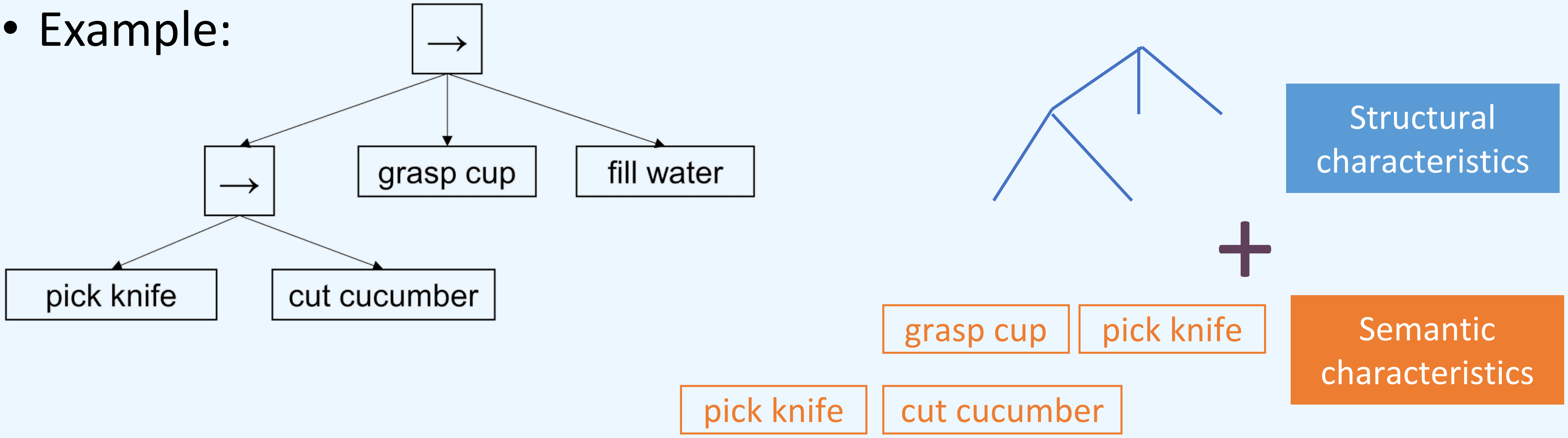


Yue Cao and C. S. George Lee, “Behavior-Tree Embeddings for Robot Task-Level Knowledge”, IROS, 2022

LLM for Planning

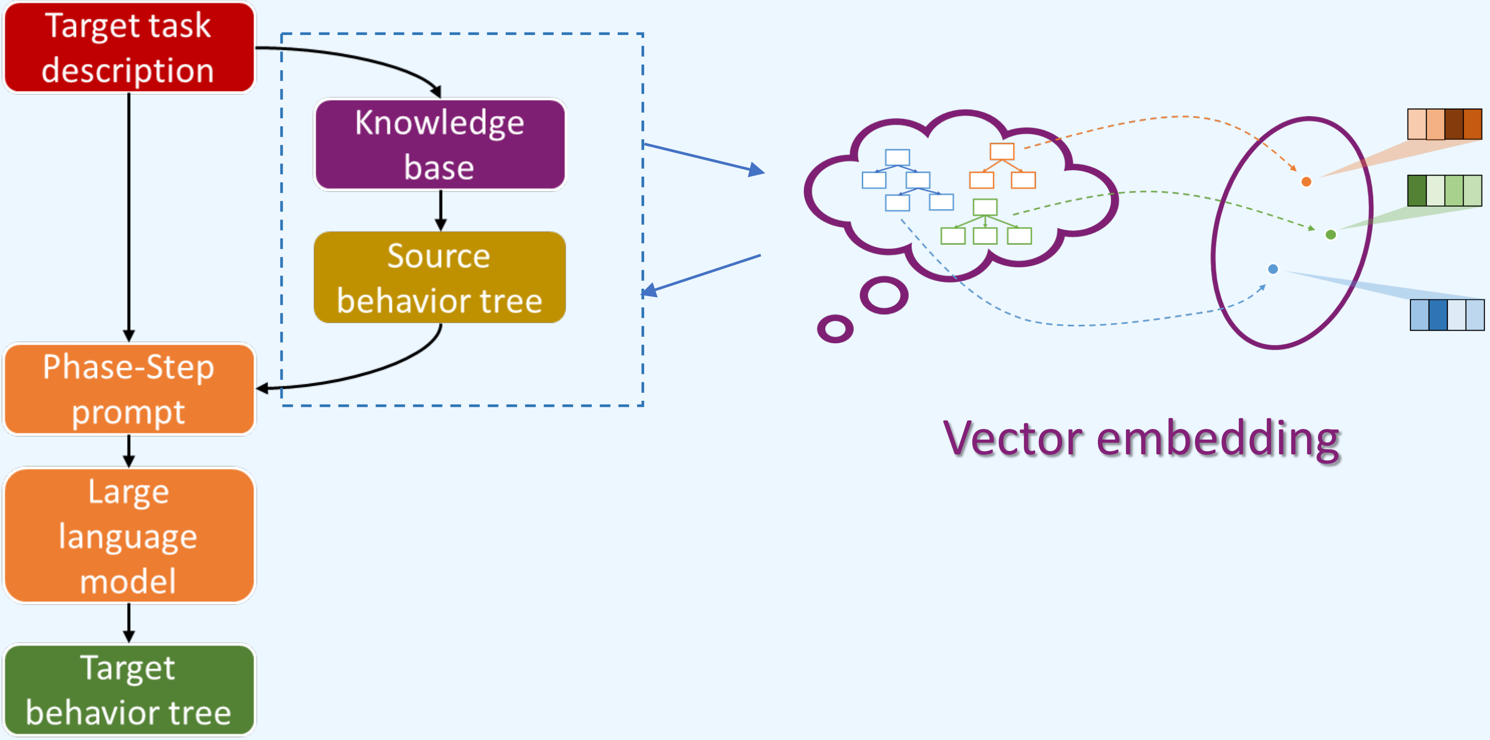
- Principle: In the vector space, similar tasks should be close, while distinct tasks should be far away. → **Define similarity?**

- Example:



LLM for Planning

- Summary



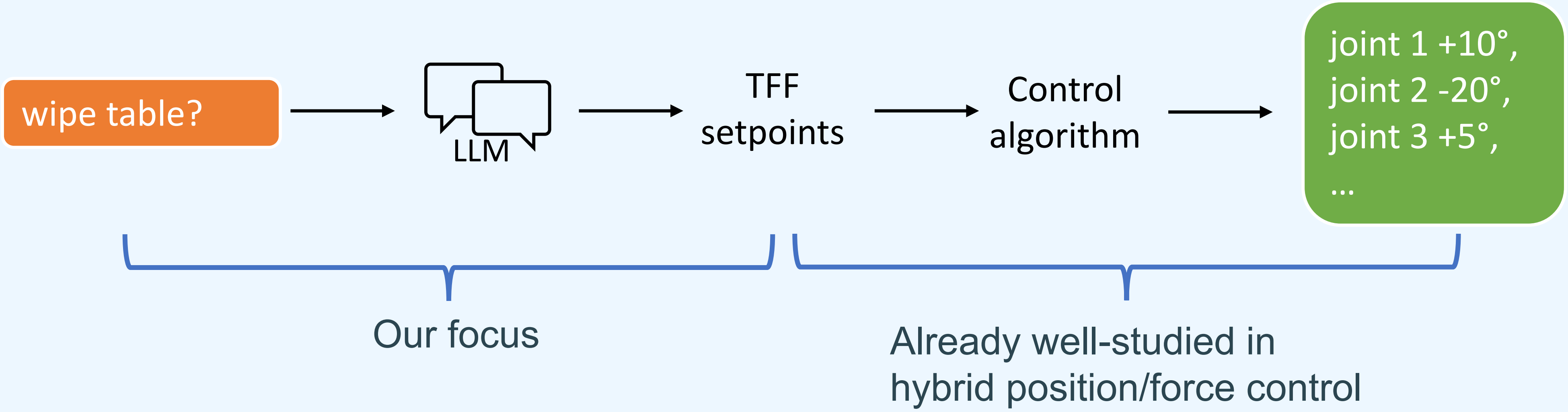
LLM for Acting

- These generated tasks are not yet linked with low-level motor execution.

- Using LLMs to realize this process?

LLM for Acting

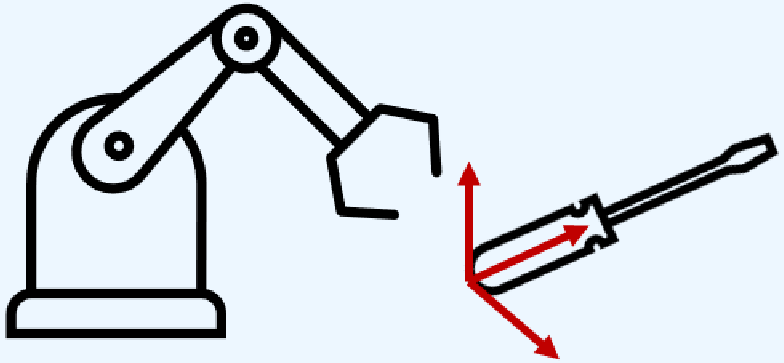
- Scope: Manipulator primitive tasks, typically contact-rich.
- **Input:** a natural-language-described manipulator primitive task.
- **Utilize:** task frame formalism (TFF)
- **Output:** a set of position/force set-points in the task frame.

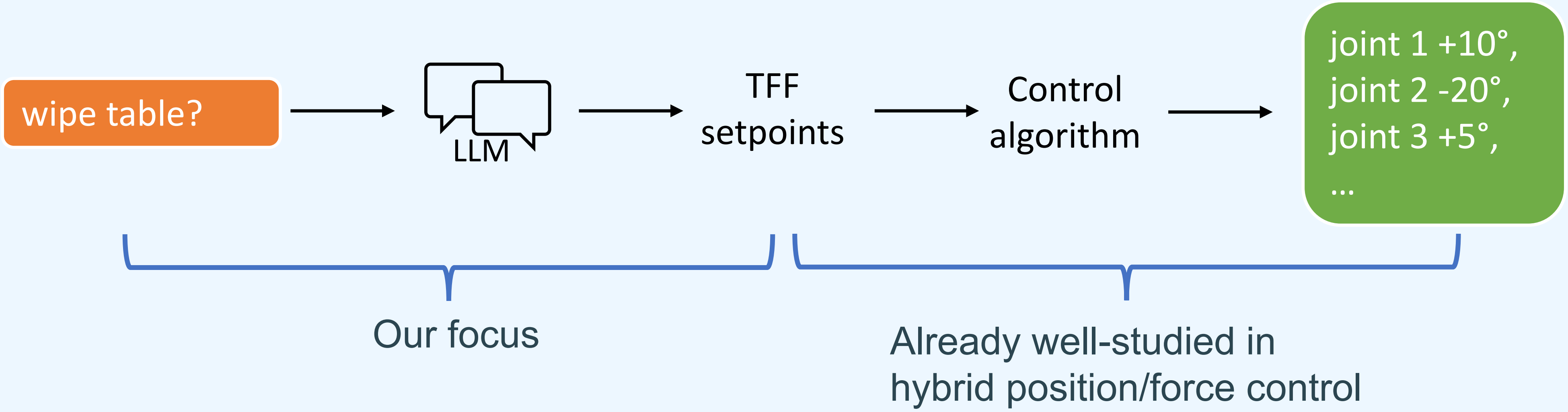


Yue Cao and C. S. George Lee, "Ground Manipulator Primitive Tasks to Executable Actions using Large Language Models", AAAI Fall Symposium, 2023

LLM for Acting

- task frame formalism
- A frame specified on the manipulated object,
  3 translational directions,
  3 rotational directions,
  either position controlled/force controlled.



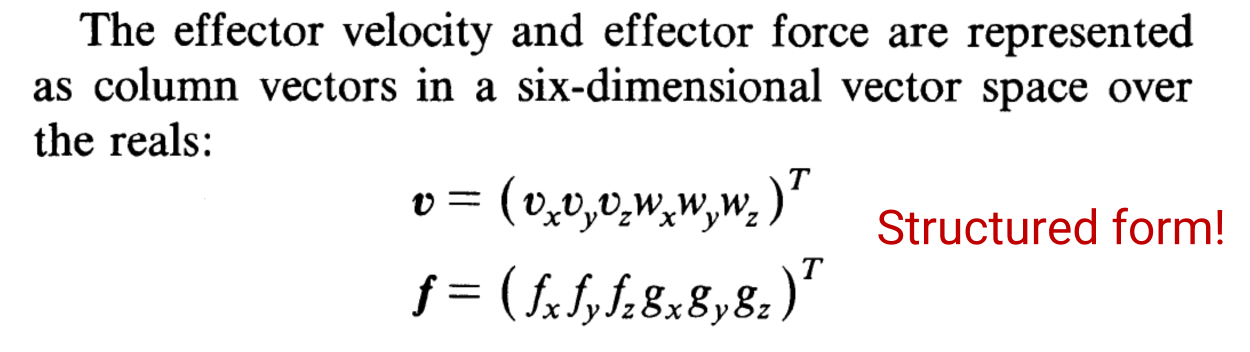


wipe table? → LLM → TFF setpoints → Control algorithm → joint 1 +10°, joint 2 -20°, joint 3 +5°, …

Our focus

Already well-studied in
hybrid position/force control

LLM for Acting



The effector velocity and effector force are represented as column vectors in a six-dimensional vector space over the reals:

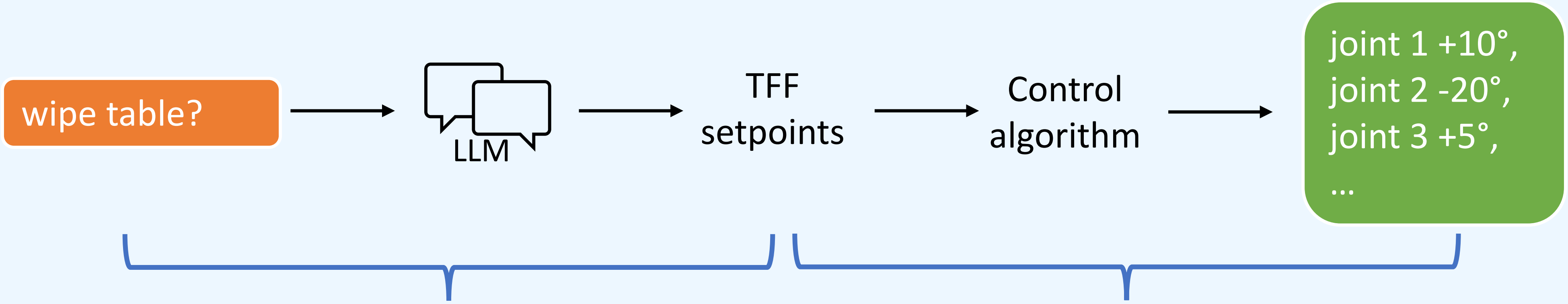
$$\boldsymbol{v} = (v_x v_y v_z w_x w_y w_z)^T$$

$$\boldsymbol{f} = (f_x f_y f_z g_x g_y g_z)^T$$

Captured from Mason's paper

Structured form! →

LLMs prefer structured input, think about the example of generating fake citations (.bib file).



wipe table? → LLM → TFF setpoints → Control algorithm →

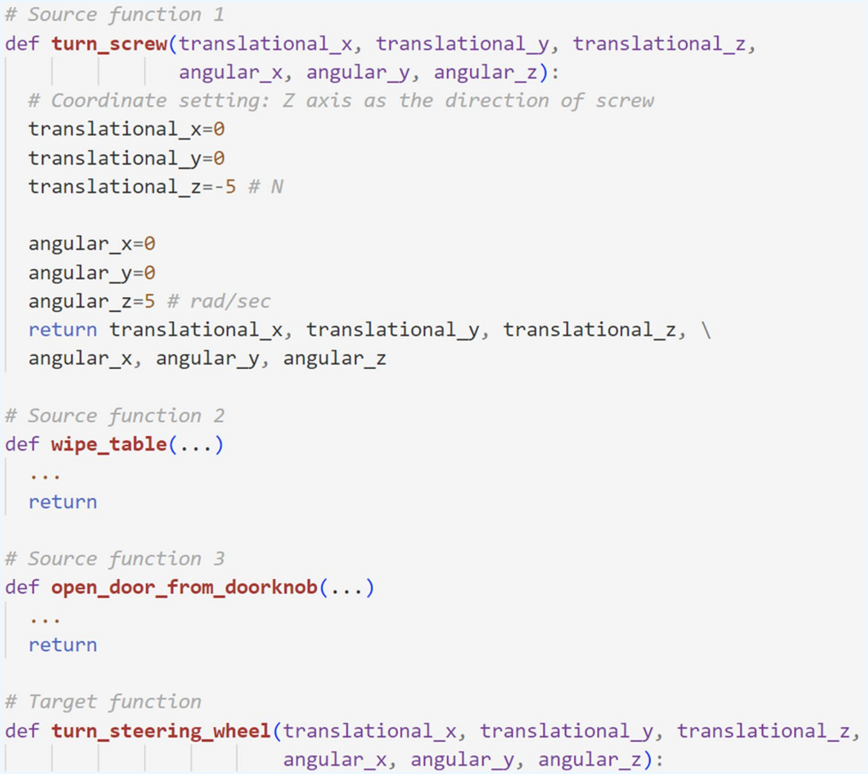
joint 1 +10°,
joint 2 -20°,
joint 3 +5°,
…

Our focus

Already well-studied in
hybrid position/force control

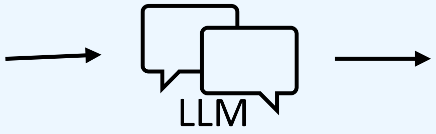
Mason, M. T., " Compliance and Force Control for Computer Controlled Manipulators ", IEEE Transactions on Systems, Man, and Cybernetics, 1981. doi: 10.1109/TSMC.1981.4308708

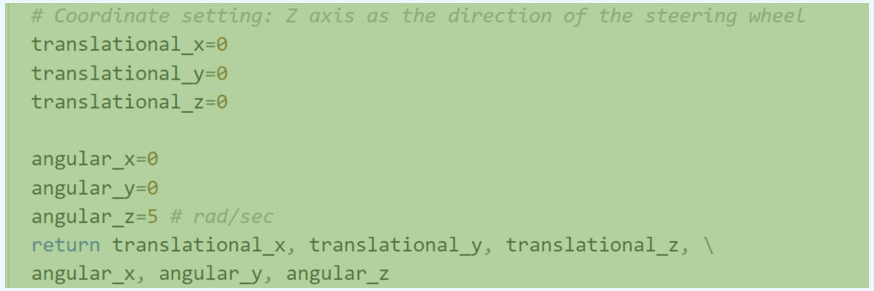
LLM for Acting



```
# Source function 1
def turn_screw(translational_x, translational_y, translational_z,
               angular_x, angular_y, angular_z):
  # Coordinate setting: Z axis as the direction of screw
  translational_x=0
  translational_y=0
  translational_z=-5 # N

  angular_x=0
  angular_y=0
  angular_z=5 # rad/sec
  return translational_x, translational_y, translational_z, \
  angular_x, angular_y, angular_z

# Source function 2
def wipe_table(...)
  ...
  return

# Source function 3
def open_door_from_doorknob(...)
  ...
  return

# Target function
def turn_steering_wheel(translational_x, translational_y, translational_z,
                        angular_x, angular_y, angular_z):
```

A 3-shot prompt example





```
# Coordinate setting: Z axis as the direction of the steering wheel
translational_x=0
translational_y=0
translational_z=0

angular_x=0
angular_y=0
angular_z=5 # rad/sec
return translational_x, translational_y, translational_z, \
angular_x, angular_y, angular_z
```

LLM Output

- Program-function-like prompt
- Task-frame-formalism-based representation
- Few-shot inference

Overview

1. Real Autonomous Robots
   (not teleoperated)
2. LLM for Planning
   (hierarchical planning, retrieval-augmented)
3. LLM for Acting
   (structured input)

Thanks!