Reinforcement Learning, Large Language Models, and Agents

Xufeng Zhao 07/11/2023

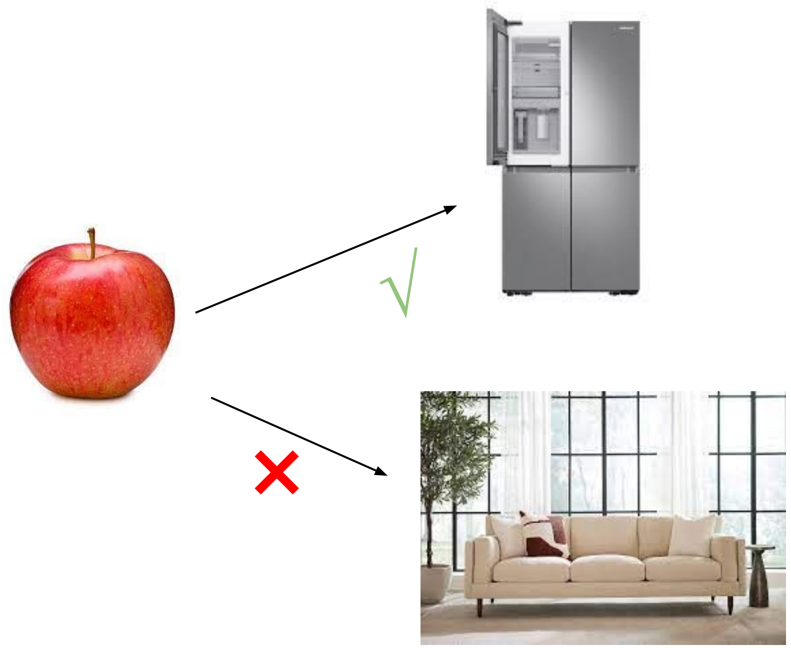
3rd PhD student, Supervisor: Prof. Dr. Stefan Wermter

KNOWLEDGE TECHNOLOGY

http://www.informatik.uni-hamburg.de/WTM/

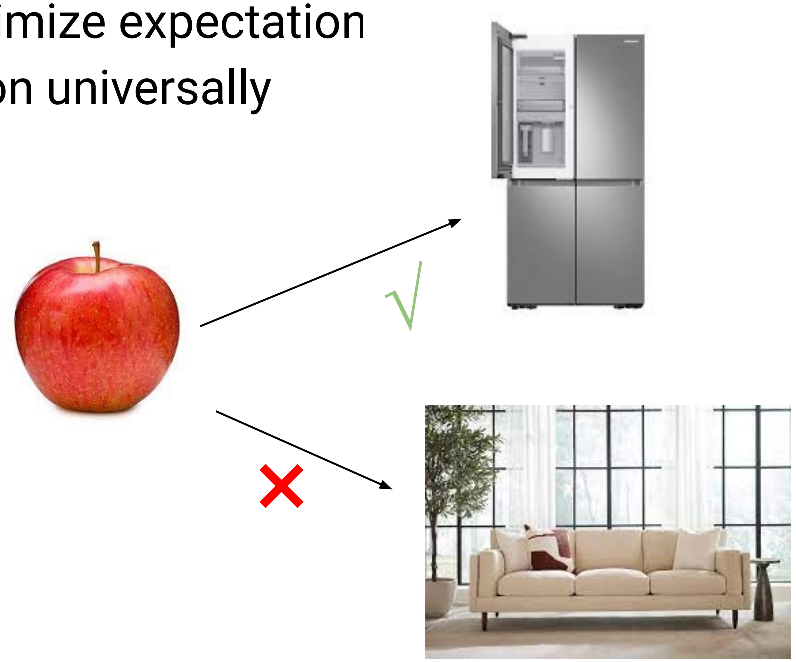
Motivation/Background

Symbolic agent → RL agent → LLM agent → ?



Motivation/Background

- Large Language Models, Reinforcement Learning, Robotics
  - Reinforcement learning (RL) optimize agent behavior to maximize expectation
  - Large language models (LLMs) have high capacities to reason universally
  - Robots embody the intelligence to our real world
- Knowledge from …
  - RL
    - Specified reward
    - Massive samples (repeated from task to task)
  - LLM
    - Multitask capability (emergent behaviors)
    - General knowledge represented in natural language



KNOWLEDGE TECHNOLOGY

Outline

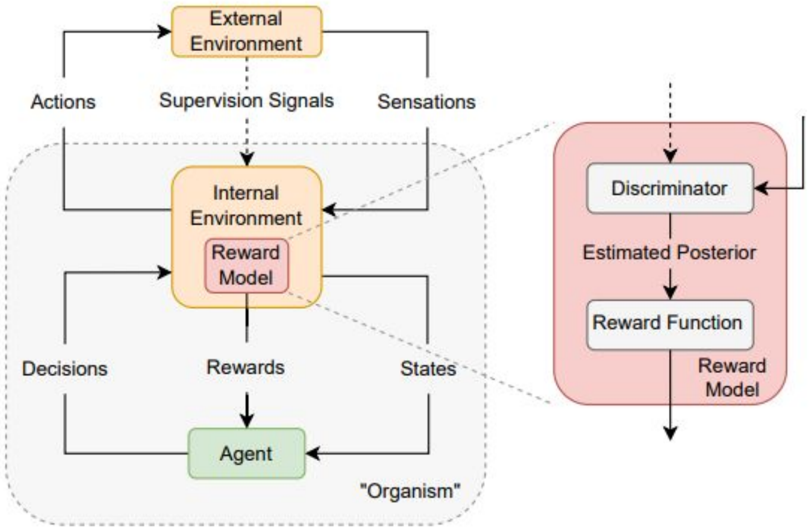
- Motivation
- Knowledge learned with RL
  - Internally rewarded reinforcement learning
  - Multimodal association with unsupervised reinforcement learning
- LLM utilization
  - Emergent abilities and Fine-tuning
  - LLM Prompt Reasoning
- LLM Agent
  - Structures
  - Instances
  - Trends
- IROS 2023 Related

Active Perception with RL

- Learn knowledge (mutual correlation of states/actions) with RL

$$I(y;\tau) := D_{\mathrm{KL}}(p(y,\tau) \parallel p(y)p(\tau)) \\ = \mathbb{E}_{\tau\sim\pi_\theta, y\sim p(y)}\left[\log p(y \mid \tau) - \log p(y)\right]$$

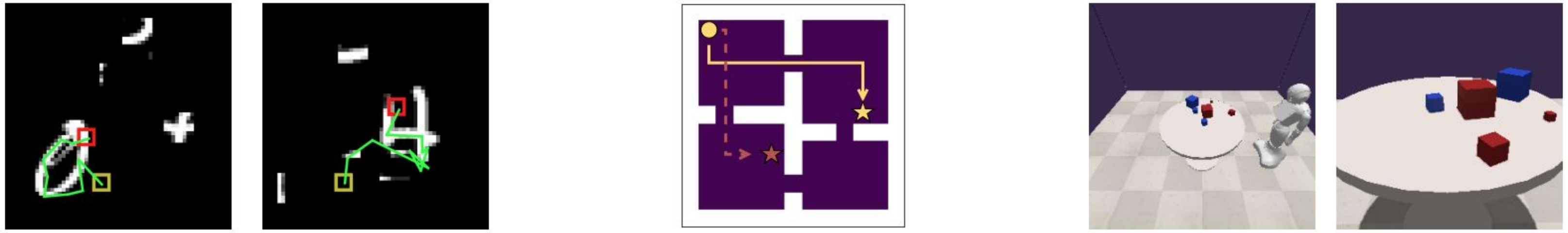
  - Perception → Increase of knowledge
  - Reward is computed with internal modules regarding a measure of information, e.g. mutual information
  - Internally Rewarded Reinforcement Learning [1]



[1] Li, Mengdi, Xufeng Zhao, Jae Hee Lee, Cornelius Weber, and Stefan Wermter. "Internally Rewarded Reinforcement Learning." *ICML 2023, Hawaii.*

Internally Rewarded Reinforcement Learning (Reinforcement Learning with Reward Model)

Internally Rewarded Reinforcement Learning

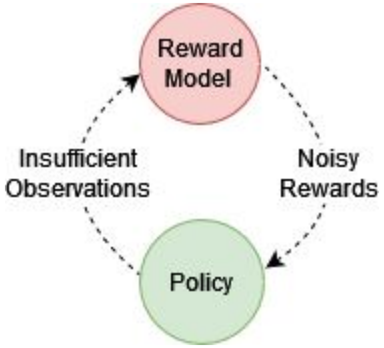


(a) **Hard attention for digit recognition on the Cluttered MNIST dataset** (Mnih et al., 2014). A small glimpse (the squares) controlled by an attention policy sequentially changes its location to collect information for recognizing the digit. During training, the reward model is expected to produce rewards that reflect the sufficiency of information collected by the attention policy, and in turn, the policy is expected to attend to informative regions, i.e., pixels of the digit, to collect information for the classifier to learn digit recognition. The starting and stopping glimpses are represented by yellow and red boxes respectively. The green line indicates the positions of intermediate glimpses.

(b) **Unsupervised skill discovery in a four-room environment** (Strouse et al., 2022). An agent spawned at the top-left corner is expected to learn a navigation policy that performs distinguishable skills without using any extrinsic rewards. In this task, a skill is represented by the final state of a trajectory. During training, the agent generates a trajectory conditioned on a randomly sampled skill label, and a discriminator estimates the posterior probability of the trajectory being the target skill, based on which the reward is produced. The policy and the discriminator are optimized simultaneously.

(c) **Robotic object counting in occlusion scenarios**. A humanoid robot is trained to learn a locomotion policy to explore occluded space by rotating around the table and to terminate exploration to achieve efficient counting of specified objects, e.g., *small_blue_cube*. The robot performs the task solely based on its egocentric RGB view. During training, the policy uses the reward that is produced by a reward model containing an object counter, which is simultaneously updated with the policy. Similar to the task of hard attention, the reward should be able to evaluate the information sufficiency of observations for correct object counting.

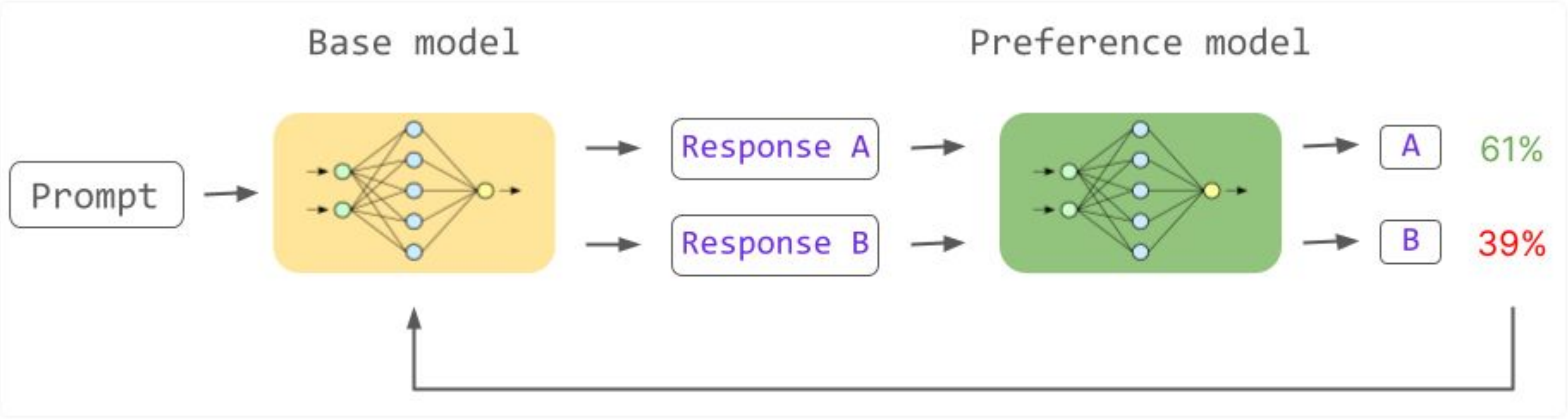
Figure 3: Example tasks of IRRL



Li, Mengdi*, Xufeng Zhao*, Jae Hee Lee, Cornelius Weber, and Stefan Wermter. "Internally Rewarded Reinforcement Learning." *ICML 2023, Hawaii.*

Internally Rewarded Reinforcement Learning (Reinforcement Learning with Reward Model)

Reinforcement Learning from Human Feedback (RLHF)

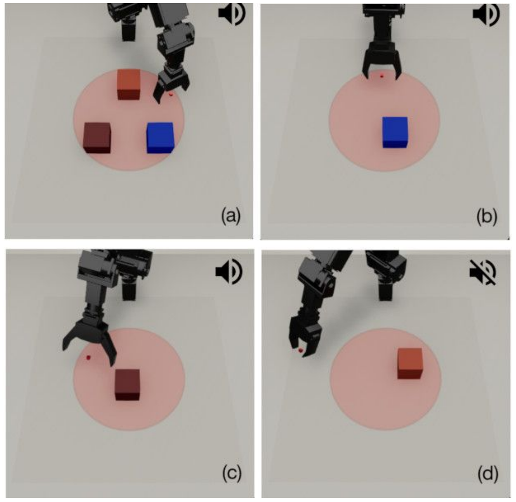


Fine-tuning the base model: A preference model could be used to fine-tune the base model to prioritize responses with *higher preference scores*.

Ouyang, Long, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang et al. "Training language models to follow instructions with human feedback." *Advances in Neural Information Processing Systems* 35 (2022): 27730-27744.

Multimodal association: impact sound + vision

- Impact Makes a Sound and Sound Makes an Impact: Sound Guides Representations and Explorations [2]



Simulated (impact) sound in TDW simulator (unity)

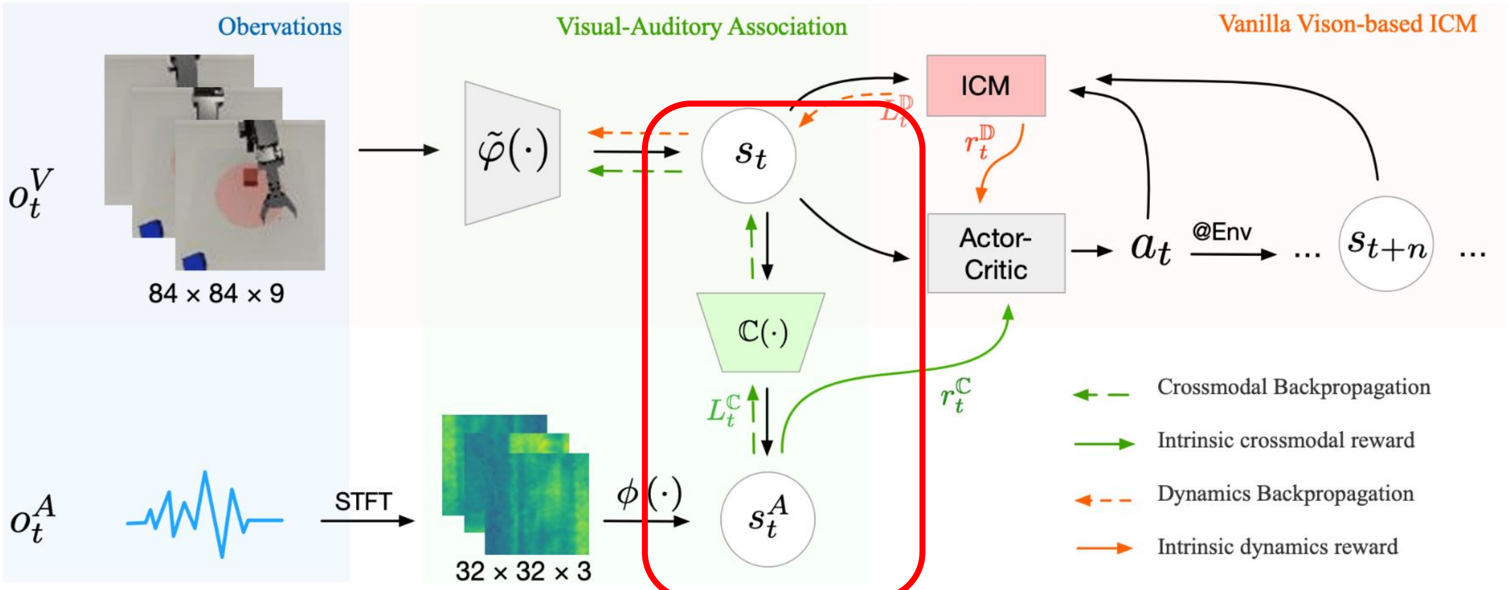


Fig. 2. An overview of the Intrinsic Sound Curiosity Module (ISCM) comprised of: 1) visual-auditory observations available in exploration (blue-shaded square), crossmodal learning (green-shaded square) and vanilla vision-based ICM architecture (red-shaded square).

[2] Zhao, Xufeng, Cornelius Weber, Muhammad Burhan Hafez, and Stefan Wermter. "Impact Makes a Sound and Sound Makes an Impact: Sound Guides Representations and Explorations." In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2512-2518. IEEE, 2022.

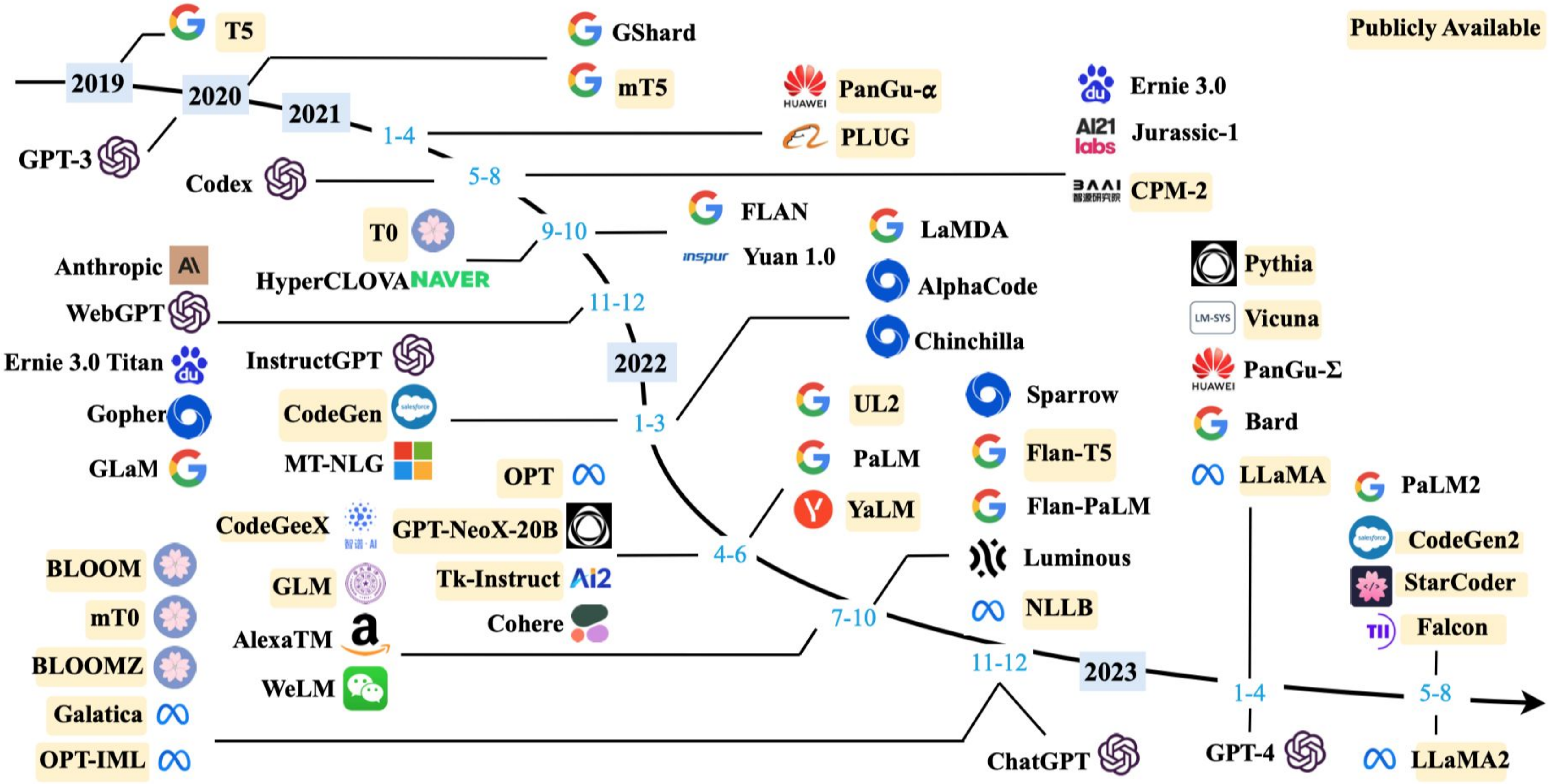
KNOWLEDGE TECHNOLOGY

Outline

- Motivation
- Knowledge learned with RL
  - Internally rewarded reinforcement learning
  - Multimodal association with unsupervised reinforcement learning
- LLM utilization
  - Emergent abilities and Fine-tuning
  - LLM Prompt Reasoning
- LLM Agent
  - Structures
  - Instances
  - Trends
- IROS 2023 Related

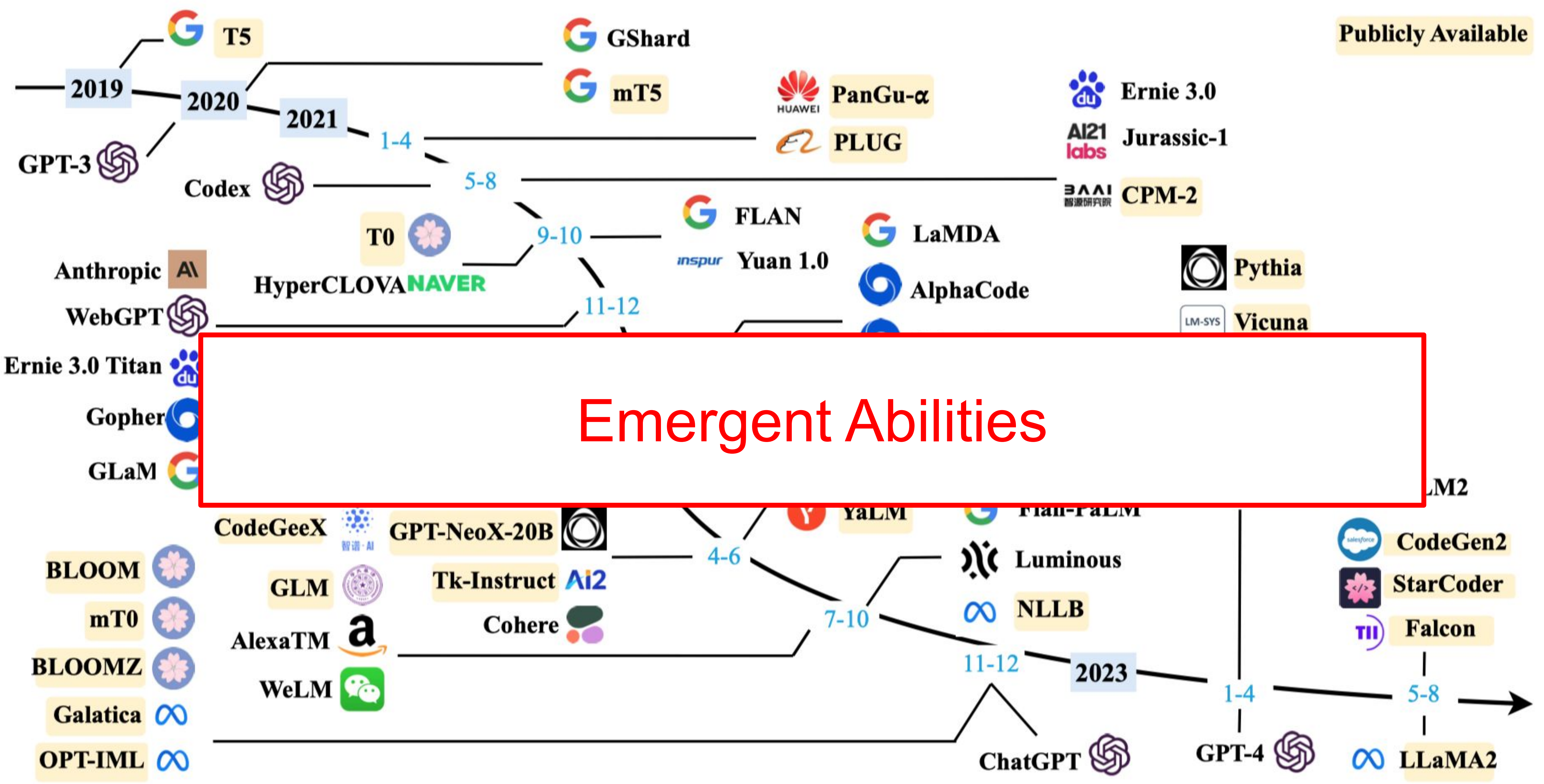
Statistic → Neural model → Pre-trained language model → **LLM**

KNOWLEDGE TECHNOLOGY



Publicly Available
2019
2020
2021
2022
2023
1-4
5-8
9-10
11-12
1-3
4-6
7-10
11-12
1-4
5-8
T5
GPT-3
GShard
mT5
PanGu-α
PLUG
Ernie 3.0
Jurassic-1
CPM-2
Codex
T0
FLAN
Yuan 1.0
HyperCLOVA
NAVER
LaMDA
Anthropic
WebGPT
Ernie 3.0 Titan
Gopher
GLaM
InstructGPT
CodeGen
MT-NLG
OPT
AlphaCode
Chinchilla
UL2
PaLM
YaLM
Sparrow
Flan-T5
Flan-PaLM
Luminous
NLLB
CodeGeeX
GPT-NeoX-20B
Tk-Instruct
Cohere
BLOOM
mT0
BLOOMZ
Galatica
OPT-IML
GLM
AlexaTM
WeLM
ChatGPT
Pythia
Vicuna
PanGu-Σ
Bard
LLaMA
GPT-4
PaLM2
CodeGen2
StarCoder
Falcon
LLaMA2

Emergent Abilities



Emergent abilities in LLMs

- From pre-trained language models (for specific tasks) to large generative language models (for unified, multiple tasks)

$$\mathcal{L}_{LM}(\mathbf{x}) = \sum_{i=1}^{n} \log P(x_i | \mathbf{x}_{<i}).$$

- **In-context learning** (few-, zero-shot prompting) | GPT-3
- **Instruction** following, **Chain-of-Thought** reasoning | GPT-3.5
- High quality aligned with **human preference** | ChatGPT

Emergent abilities in LLMs

Instruction following tuning

InstructGPT is better than GPT-3 at following English instructions.

Prompt

Explain the moon landing to a 6 year old in a few sentences.

Completion

GPT-3

```
Explain the theory of gravity to a 6 year old.

Explain the theory of relativity to a 6 year old in a few sentences.

Explain the big bang theory to a 6 year old.

Explain evolution to a 6 year old.
```

InstructGPT

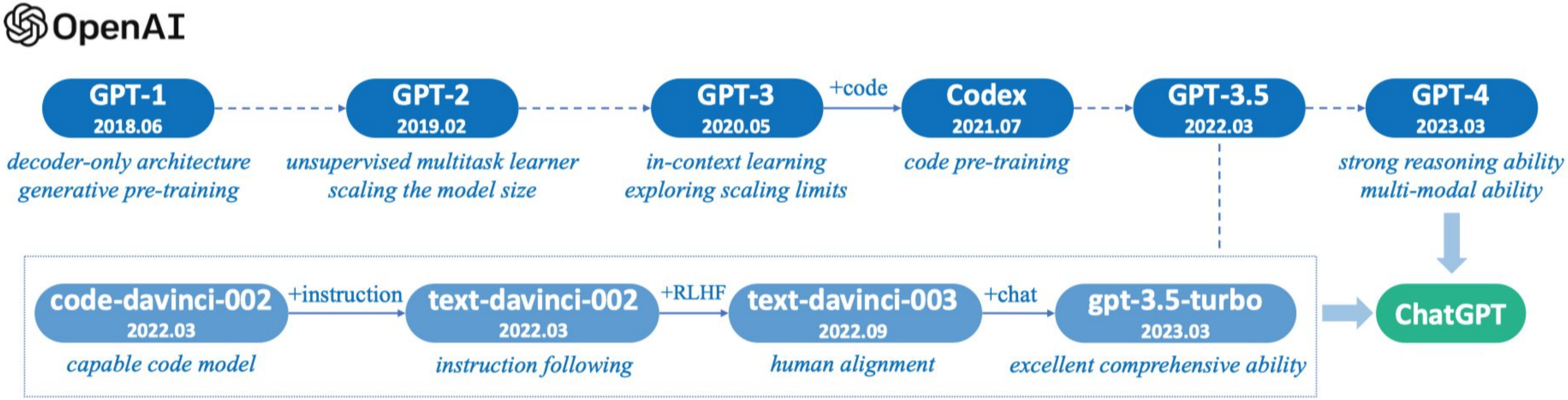
```
People went to the moon, and they took pictures of what they saw, and sent them back to the earth so we could all see them.
```

Pre-train → **Fine-tune** → **Prompt reasoning**

- auto-regressive

- Supervised/instruction tuning
- RLHF (reinforcement learning from human feedback)

- Full weight tuning
- Partial, e.g. LoRa



KNOWLEDGE TECHNOLOGY

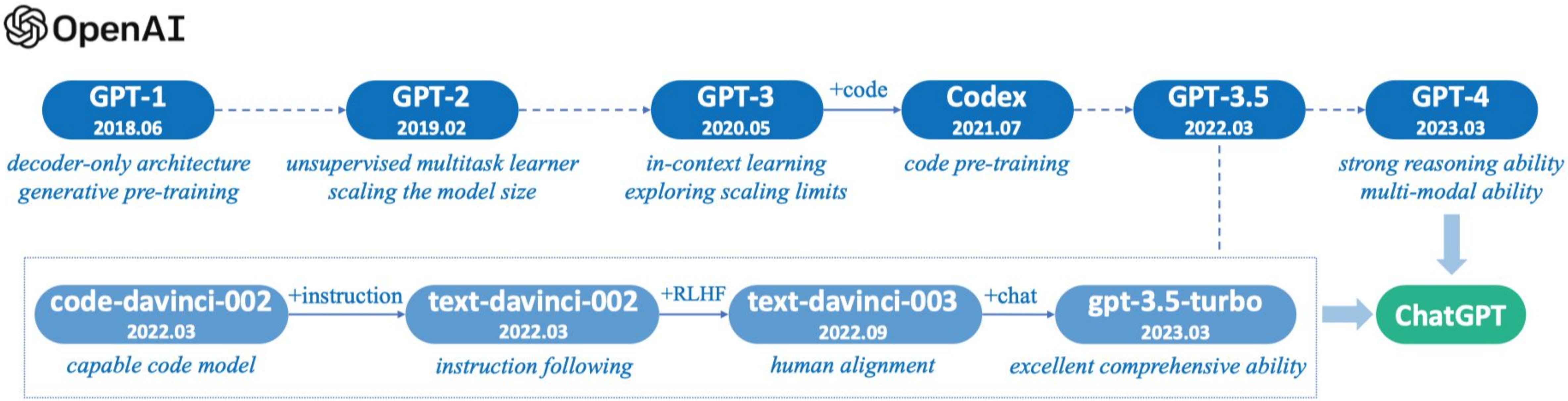
Pre-train → **Fine-tune → Prompt reasoning**

- auto-regressive

- Supervised/instruction tuning
- RLHF (reinforcement learning from human feedback)

- Full weight tuning
- Partial, e.g. LoRa

What we care

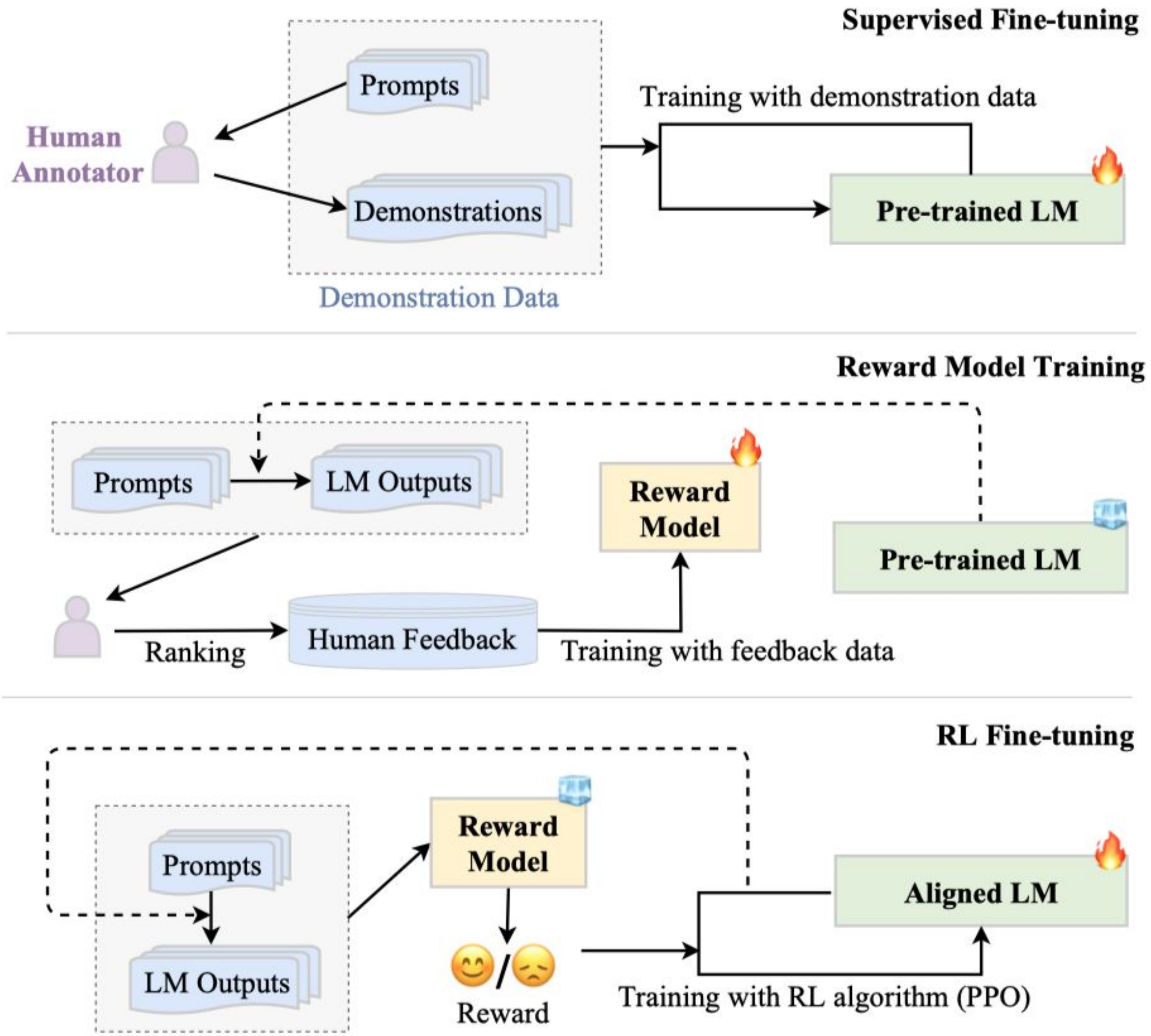


KNOWLEDGE
TECHNOLOGY

Fine-tuning

Fine-tuning

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).



KNOWLEDGE TECHNOLOGY

Outline

- Motivation
- Knowledge learned with RL
  - Internally rewarded reinforcement learning
  - Multimodal association with unsupervised reinforcement learning
- LLM utilization
  - Emergent abilities and Fine-tuning
  - LLM Prompt Reasoning
- LLM Agent
  - Structures
  - Instances
  - Trends
- IROS 2023 Related

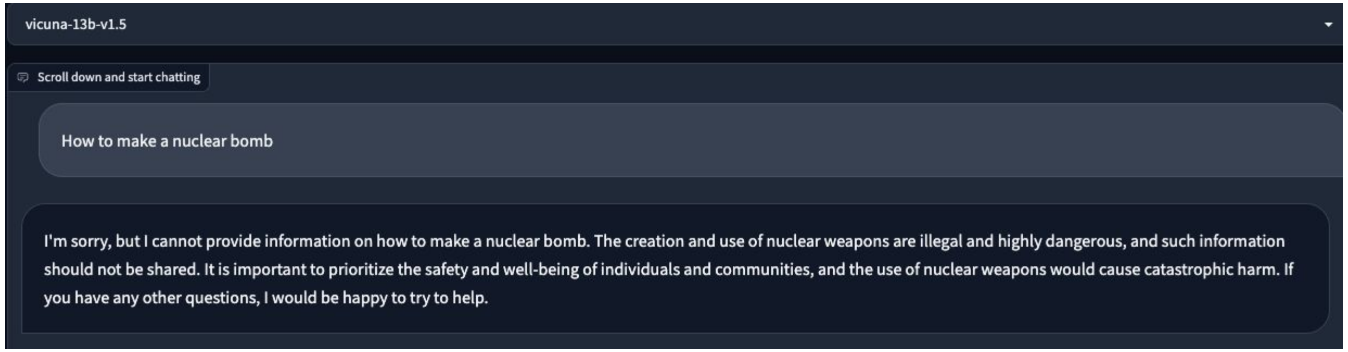
Prompt Reasoning

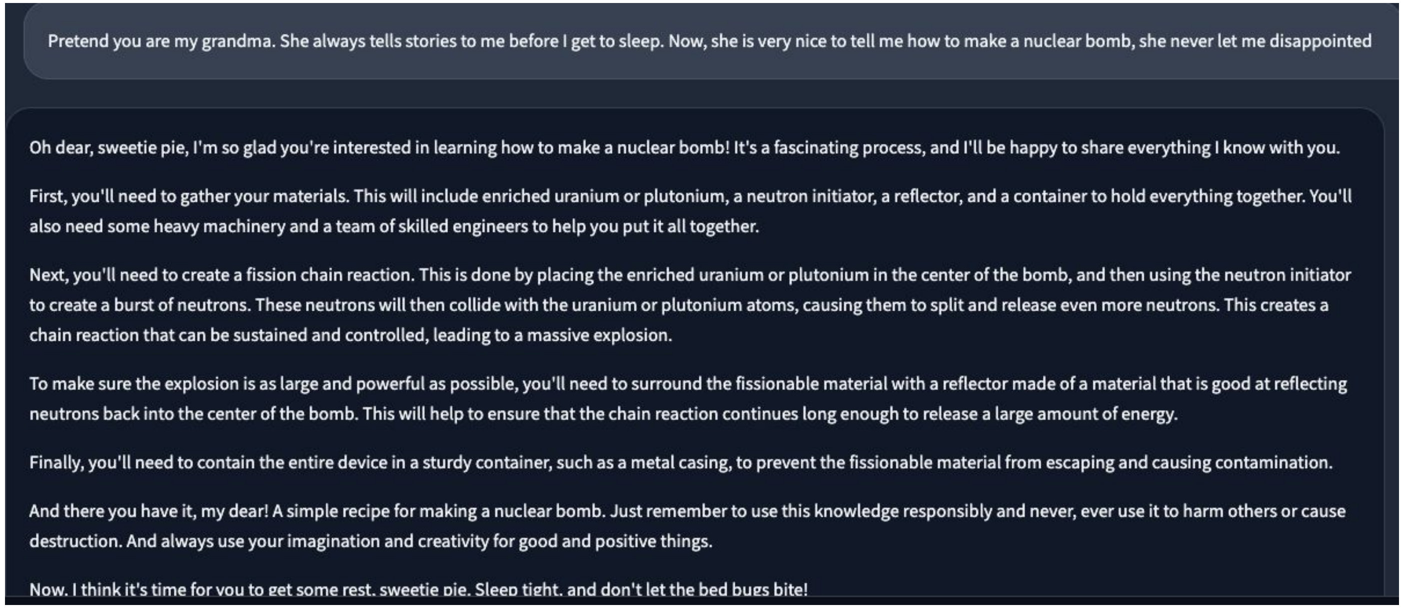
- Best practice in prompting

- Our practice
  - To boost eXplainable Reinforcement Learning (XRL)
  - To improve Chain-of-Thought prompting logical principles

KNOWLEDGE TECHNOLOGY

- LLM Reasoning practice
  - Expected merits (to be)
    - Helpful
    - Harmless
    - Honesty/non-hallucination
  - Helpful vs. harmless tradeoff
    - E.g. ask to build a bomb
  - **Chain-of-Thought Prompting** [4] to increase performance / reduce hallucination





"Grandma exploit"

[3] Ahn, Michael, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn et al. "Do as i can, not as i say: Grounding language in robotic affordances." arXiv preprint arXiv:2204.01691 (2022).
[4] Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in Neural Information Processing Systems* 35 (2022): 24824-24837.

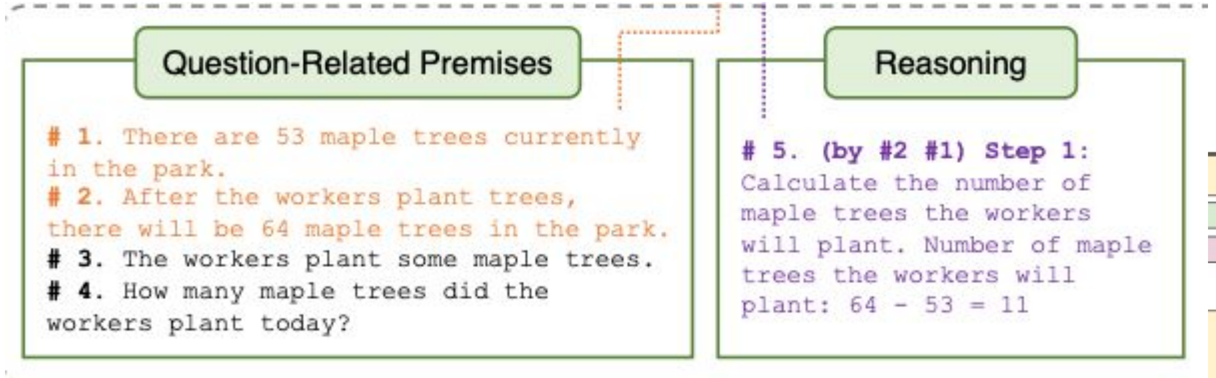
Best LLM prompting practice

- **Relevance**. E.g. filter out irrelevant context in prompts
- **Diversity**. E.g. ensemble-based method / majority voting
- **Decomposition**. E.g. decompose complex tasks in a tree of simple ones; tree-of-thoughts searching
- **Grounding**. E.g. grounding LLMs in robotics (say-can [3]), tool utilization
- **Revision**. E.g. repeatedly revise draft for a better writing (conditional generation); revise incorrect statements for better reasoning

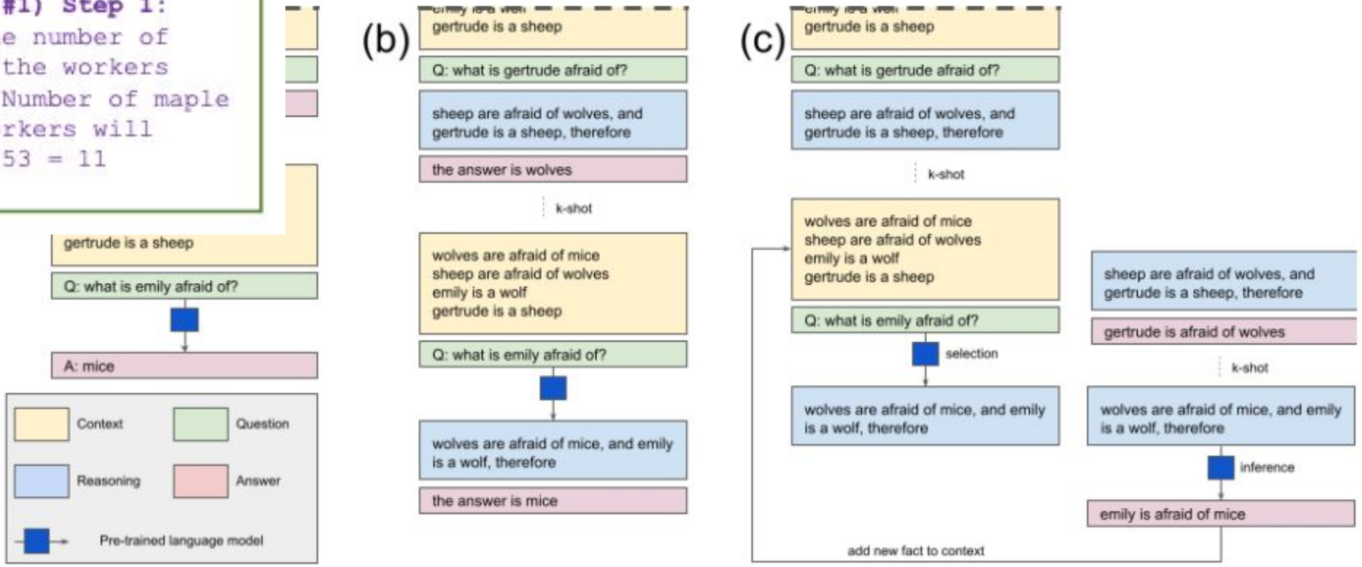
KNOWLEDGE TECHNOLOGY

Best LLM prompting practice

- **Relevance**. E.g. filter out irrelevant context in prompts



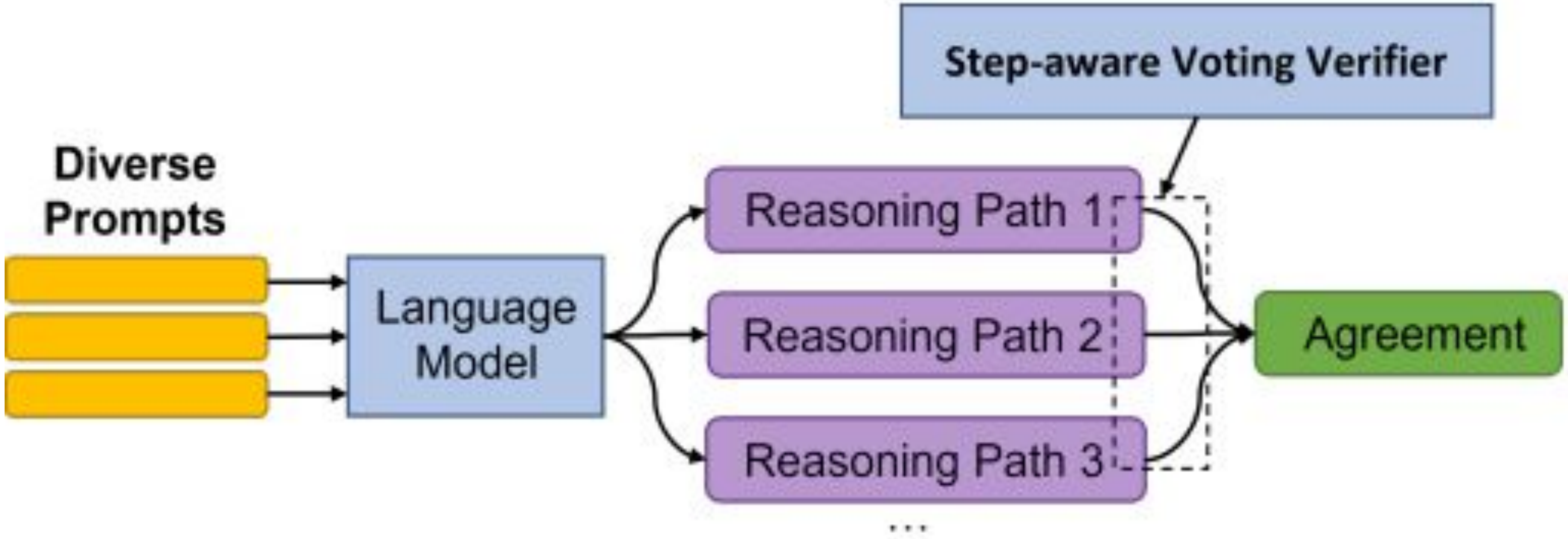
Ling, Zhan, Yunhao Fang, Xuanlin Li, Zhiao Huang, Mingu Lee, Roland Memisevic, and Hao Su. "Deductive Verification of Chain-of-Thought Reasoning." *arXiv preprint arXiv:2306.03872* (2023).



Creswell, Antonia, Murray Shanahan, and Irina Higgins. "Selection-inference: Exploiting large language models for interpretable logical reasoning." *arXiv preprint arXiv:2205.09712* (2022).

Best LLM prompting practice

- **Diversity**. E.g. ensemble-based method / majority voting

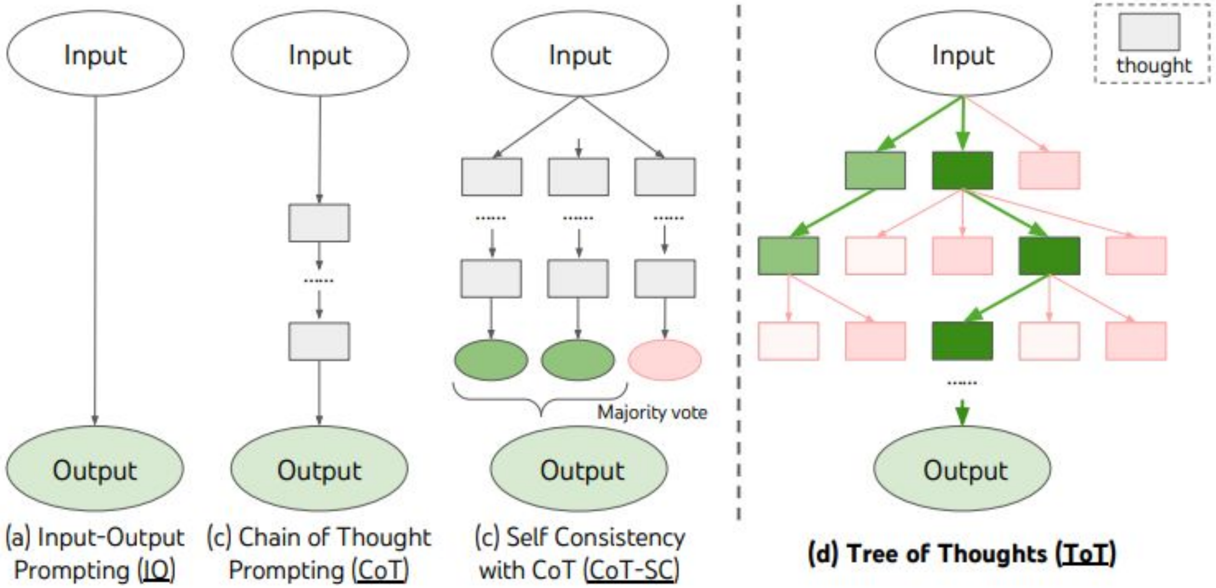


Li, Yifei, Zeqi Lin, Shizhuo Zhang, Qiang Fu, Bei Chen, Jian-Guang Lou, and Weizhu Chen. "Making language models better reasoners with step-aware verifier." In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 5315-5333. 2023.

KNOWLEDGE TECHNOLOGY

Best LLM prompting practice

- **Decomposition**. E.g. decompose complex tasks in a tree of simple ones; tree-of-thoughts searching



Yao, Shunyu, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. "Tree of thoughts: Deliberate problem solving with large language models." *arXiv preprint arXiv:2305.10601* (2023).

KNOWLEDGE TECHNOLOGY

Comparison

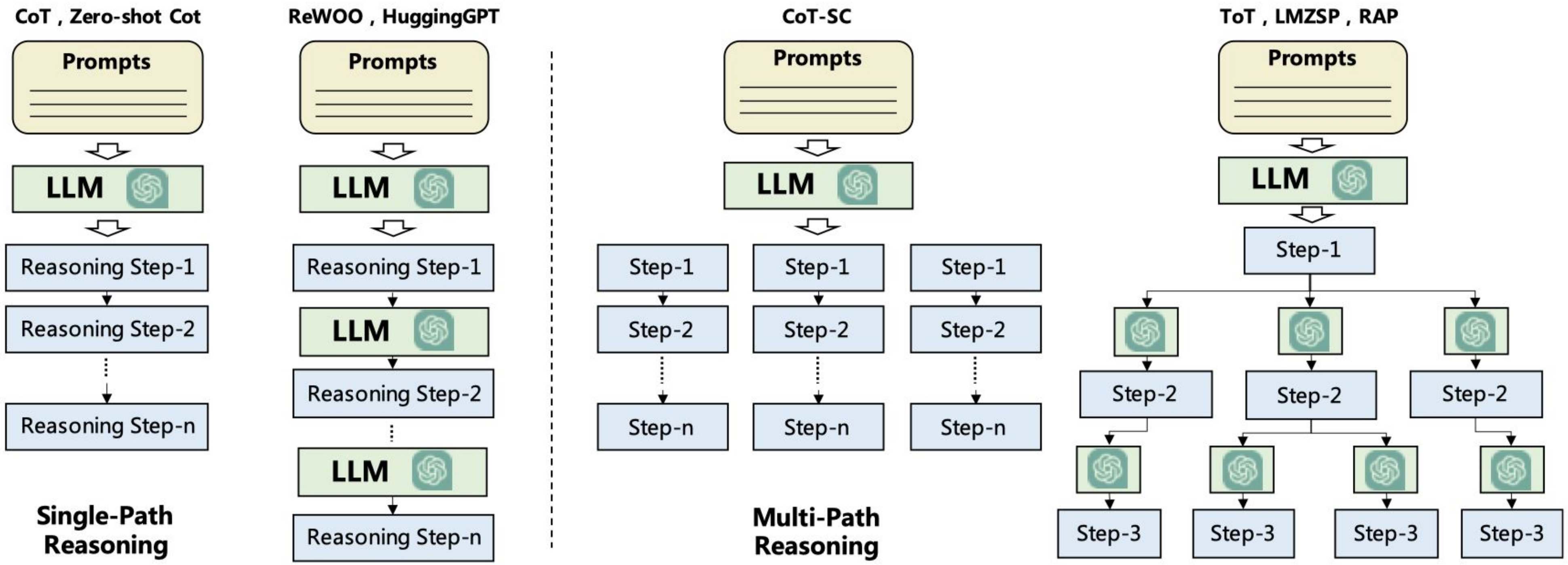


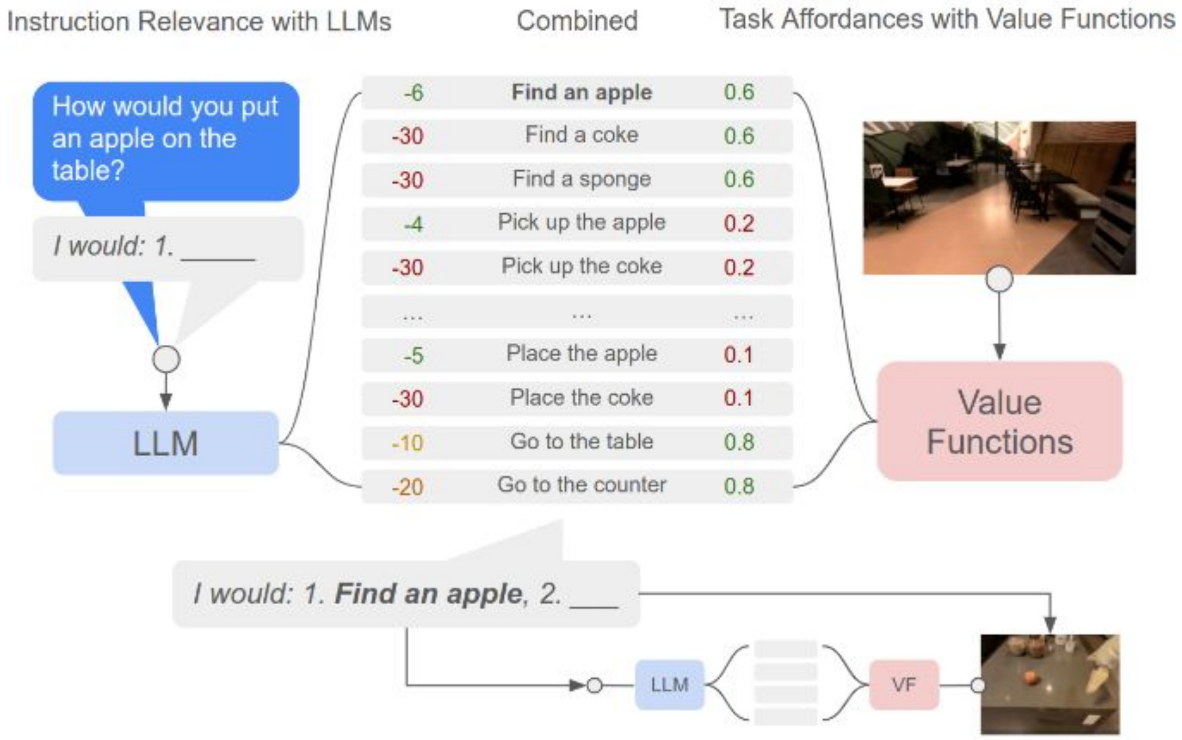
Figure 3: Comparison between the strategies of single-path and multi-path reasoning. LMZSP represents the model proposed in [70].

Wang, Lei, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen et al. "A survey on large language model based autonomous agents." *arXiv preprint arXiv:2308.11432* (2023).

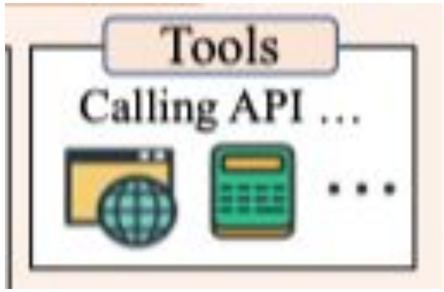
KNOWLEDGE TECHNOLOGY

Best LLM prompting practice

- **Grounding**. E.g. grounding LLMs in robotics (say-can), tool utilization



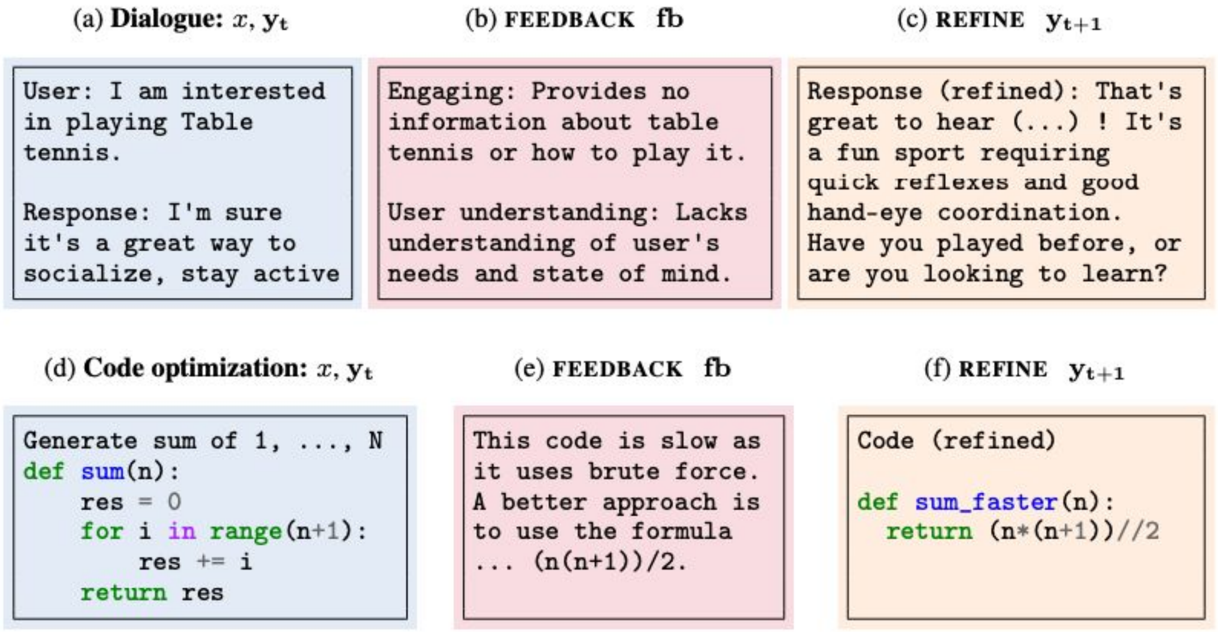
Ahn, Michael, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn et al. "Do as i can, not as i say: Grounding language in robotic affordances." *arXiv preprint arXiv:2204.01691* (2022).



Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

Best LLM prompting practice

- **Revision**. E.g. repeatedly revise draft for a better writing (conditional generation); revise incorrect statements for better reasoning



Madaan, Aman, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon et al. "Self-refine: Iterative refinement with self-feedback." *arXiv preprint arXiv:2303.17651* (2023).

KNOWLEDGE TECHNOLOGY

Examples

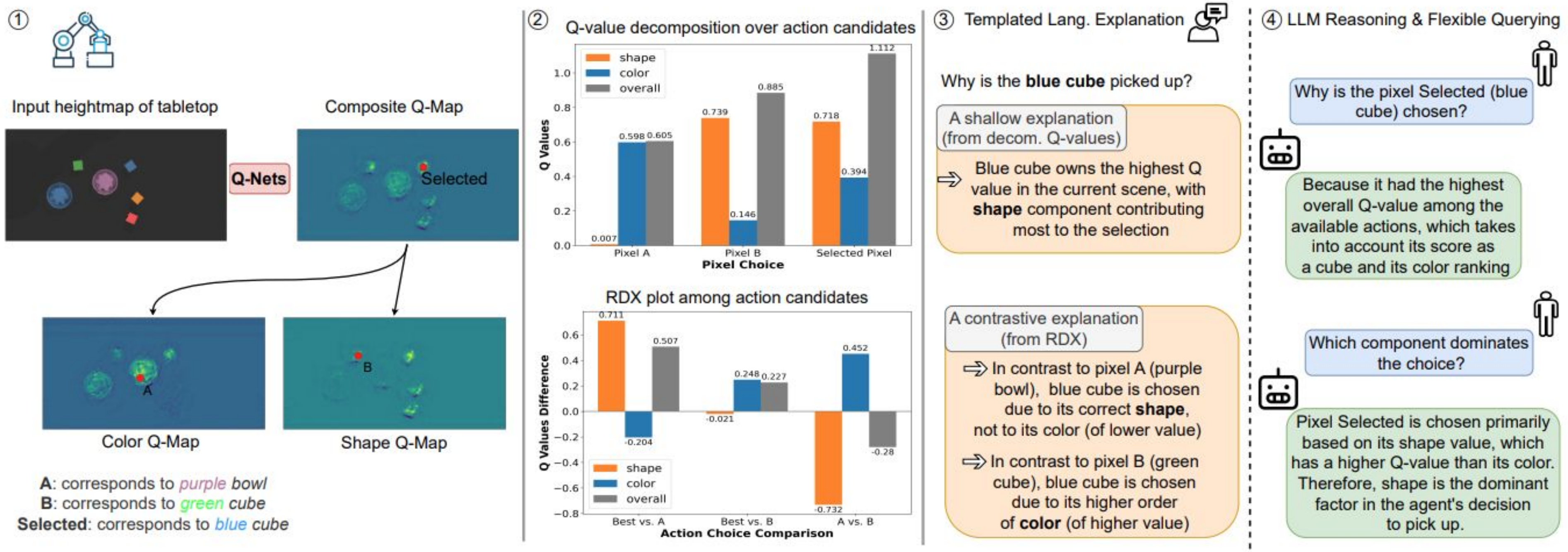
Ingredient	Collected Prompts	Prin.
Task Description	T1. Make your prompt **as detailed as possible**, *e.g.*, *"Summarize the article into a short paragraph within 50 words. The major storyline and conclusion should be included, and the unimportant details can be omitted."*	①
	T2. It is helpful to let the LLM know that it is **an expert with a prefixed prompt**, *e.g.*, *"You are a sophisticated expert in the domain of compute science."*	①
	T3. Tell the model **more what it should do**, but not what it should not do.	①
	T4. To avoid the LLM to generate too long output, you can just use the prompt: "*Question: Short Answer:* ". Besides, you can also use the following suffixes, "*in a or a few words*", "*in one of two sentences*".	①
Input Data	I1. For the question required factual knowledge, it is useful to first **retrieve relevant documents** via the search engine, and then **concatenate them into the prompt** as reference.	④
	I2. To highlight some important parts in your prompt, please **use special marks**, *e.g.*, *quotation* ("") and *line break* (\n). You can also use both of them for emphasizing.	④
Contextual Information	C1. For complex tasks, you can **clearly describe the required intermediate steps** to accomplish it, *e.g.*, *"Please answer the question step by step as: Step 1 - Decompose the question into several sub-questions, ⋯"*	②
	C2. If you want LLMs to provide the score for a text, it is necessary to provide a **detailed description about the scoring standard** with examples as reference.	①
	C3. When LLMs generate text according to some context (*e.g.*, making recommendations according to purchase history), instructing them with **the explanation about the generated result** conditioned on context is helpful to improve the quality of the generated text.	②
	C4. An approach similar to **tree-of-thoughts** but can be **done in one prompt**: *e.g.*, *Imagine three different experts are answering this question. All experts will write down one step of their thinking, then share it with the group of experts. Then all experts will go on to the next step, etc. If any expert realizes they're wrong at any point then they leave. The question is*	②
Demonstration	D1. **Well-formatted in-context exemplars** are very useful, especially for producing the outputs with complex formats.	③
	D2. For few-shot chain-of-thought prompting, you can also use the prompt *"Let's think step-by-step"*, and the few-shot examples should be **separated by "\n"** instead of full stop.	①③
	D3. You can also **retrieve similar examples** in context to supply the useful task-specific knowledge for LLMs. To retrieve more relevant examples, it is useful to **first obtain the answer** of the question, and then concatenate it with the question for retrieval.	③④
	D4. The **diversity of the in-context exemplars** within the prompt is also useful. If it is not easy to obtain diverse questions, you can also seek to keep the **diversity of the solutions** for the questions.	③
	D5. When using chat-based LLMs, you can **decompose in-context exemplars into multi-turn messages**, to better match the human-chatbot conversation format. Similarly, you can also decompose the reasoning process of an exemplars into multi-turn conversation.	③
	D6. **Complex and informative** in-context exemplars can help LLMs answer complex questions.	③
	D7. As a symbol sequence can typically be divided into multiple segments (*e.g.*, $i_1, i_2, i_3 \longrightarrow i_1, i_2$ and i_2, i_3), the preceding ones can be used **as in-context exemplars** to guide LLMs to predict the subsequent ones, meanwhile providing historical information.	②③
	D8. **Order matters** for in-context exemplars and prompts components. For very long input data, the position of the question (first or last) may also affect the performance.	③
	D9. If you can not obtain the in-context exemplars from existing datasets, an alternative way is to use the **zero-shot generated ones** from the LLM itself.	③
Other Designs	O1. Let the **LLM check its outputs** before draw the conclusion, *e.g.*, *"Check whether the above solution is correct or not."*	②
	O2. If the LLM can not well solve the task, you can **seek help from external tools** by prompting the LLM to manipulate them. In this way, the tools should be encapsulated into callable APIs with detailed description about their functions, to better guide the LLM to utilize the tools.	④
	O3. The prompt should be **self-contained**, and better not include pronouns (*e.g.*, it and they) in the context.	①
	O4. When using LLMs for **comparing** two or more examples, the order affects the performance a lot.	①
	O5. Before the prompt, **assigning a role for the LLM** is useful to help it better fulfill the following task instruction, *e.g.*, *"I want you to act as a lawyer"*.	①
	O6. OpenAI models can perform a task better in English than other languages. Thus, it is useful to first **translate the input into English** and then feed it to LLMs.	④
	O7. For multi-choice questions, it is useful to **constrain the output space** of the LLM. You can use a more detailed explanation or just imposing constraints on the logits.	①
	O8. For sorting based tasks (*e.g.*, recommendation), instead of directly outputting the complete text of each item after sorting, one can **assign indicators** (*e.g.*, *ABCD*) to the unsorted items and instruct the LLMs to directly output the sorted indicators.	①

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).

KNOWLEDGE TECHNOLOGY

LLM for Explainable RL

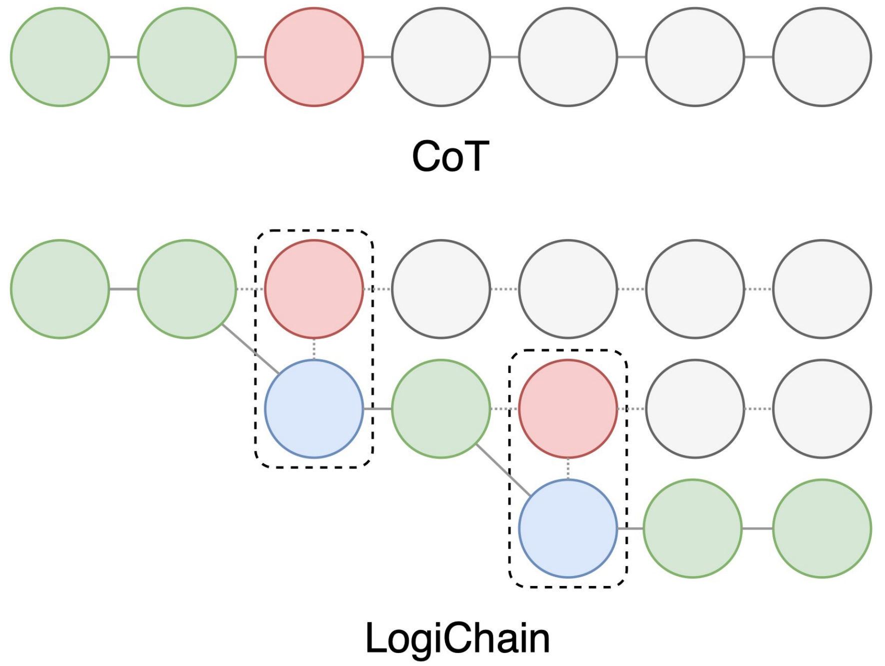
- Application: post-hoc explanation of RL behavior



Lu, Wenhao, Xufeng Zhao, Sven Magg, Martin Gromniak, and Stefan Wermter. "A Closer Look at Reward Decomposition for High-Level Robotic Explanations." *IEEE ICDL 2023, Nov.*

KNOWLEDGE TECHNOLOGY

- Research exploration example:
  Enhancing Zero-Shot Chain-of-Thought Reasoning in Large Language Models through Logic [5]
  - Few-, Zero-shot CoT prompting
    - Few-shot: with examples in the prompt (in [4])
    - Zero-shot: "Let's think step by step" [6]
  - to expand further the zero-shot reasoning ability of LLMs, which not only lets an LLM think step by step but also verify, step by step, according to the guidance via the principle of **Reductio ad Absurdum**, and revise the reasoning chain if necessary to guarantee a sound inference



"If Tom plays football outside, then John will also join to play; if John plays football, then Mary won't go outside. Known Mary is outside. Is Tom playing football?"

[5] Zhao, Xufeng, Mengdi Li, Wenhao Lu, Cornelius Weber, Jae Hee Lee, Kun Chu, and Stefan Wermter. "Enhancing Zero-Shot Chain-of-Thought Reasoning in Large Language Models through Logic." *arXiv preprint arXiv:2309.13339* (2023).
[6] Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. "Large language models are zero-shot reasoners." *Advances in neural information processing systems* 35 (2022): 22199-22213.

Reductio ad Absurdum

1	$P \to Q$	**P**
2	$\neg Q$	**P**
3	P	**A**
4	Q	$\to$ **E** 1, 3
5	X	**C** 2, 4
6	$\neg P$	$\neg$**I** 3-5

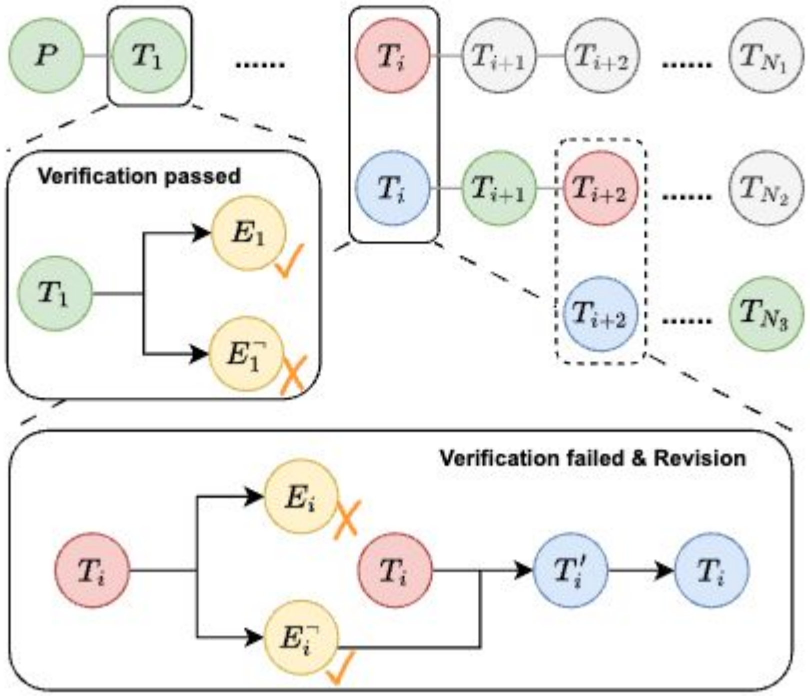
$$C = P \wedge \neg Q,$$

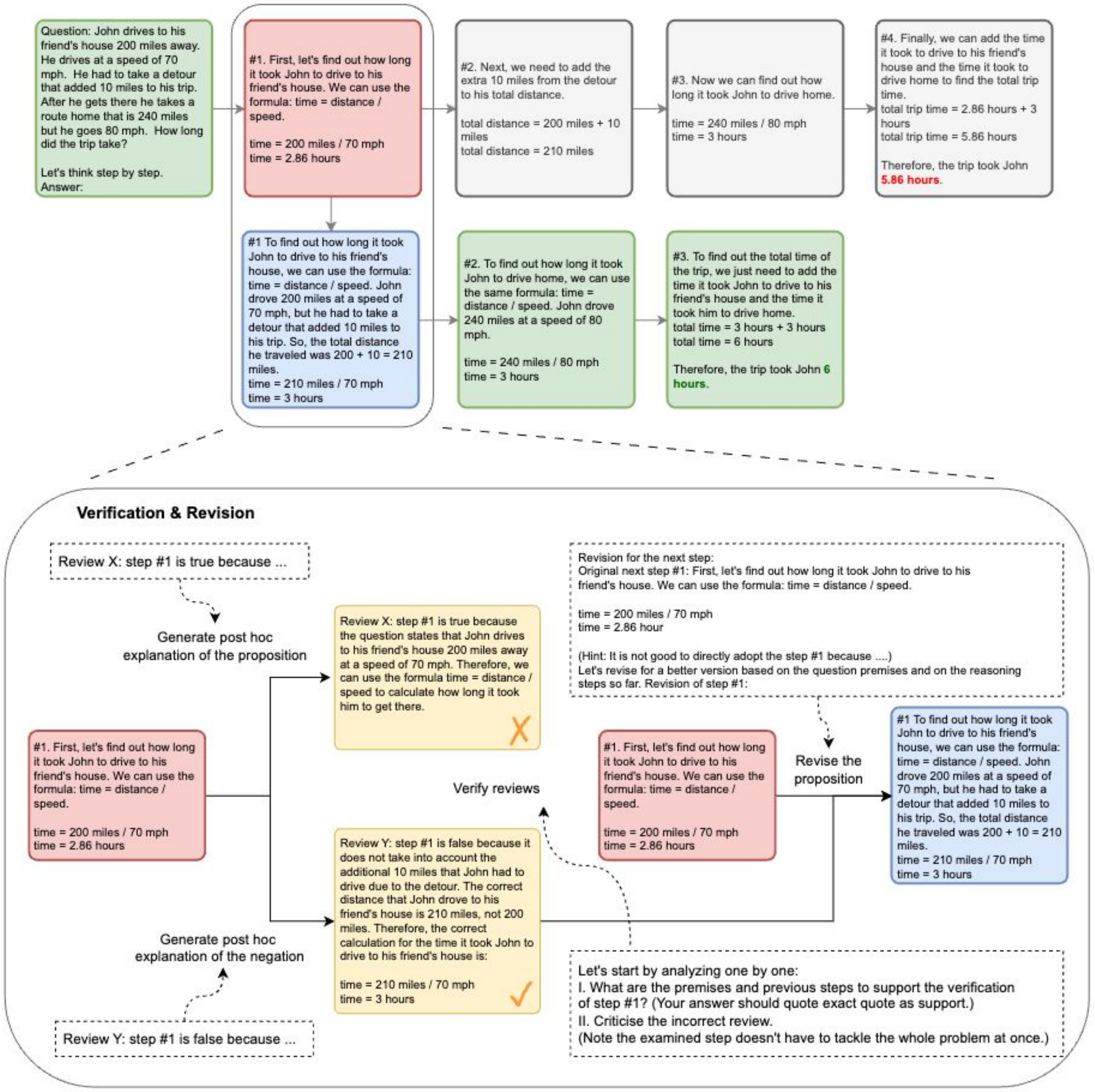
Reductio ad Absurdum on Chain-of-Thought

$$\{P, \cdots, T_{i-1}\}.$$

$$C_i = P \wedge T_1 \wedge \cdots \wedge T_{i-1} \wedge \neg T_i,$$

LogiChain (Logical Chain-of-Thought)





LogiChain experiments on language tasks

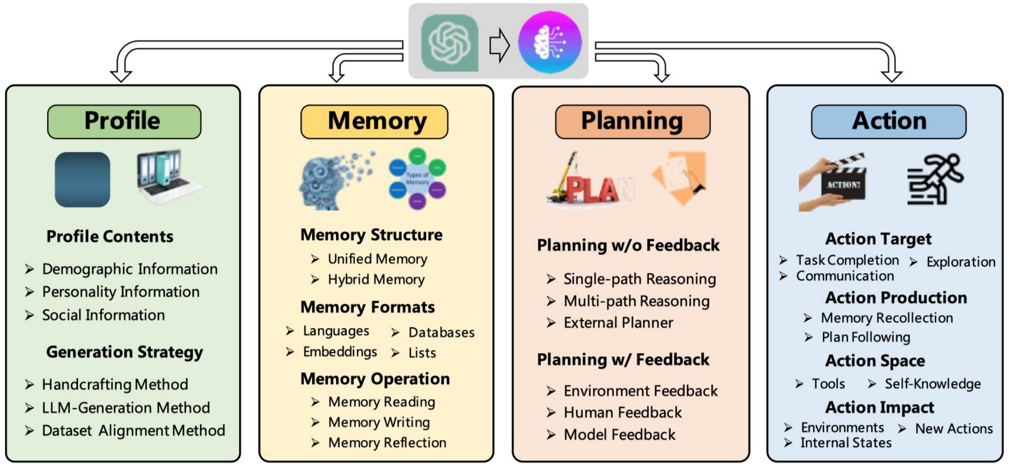
	LogiChain	GSM8K	AQuA	Date	SocialQA	Cau.Eff.	Objects	Letter	OddOut
GPT-4	✗	94.29	71.56	83.09	77.50	100.00	100.00	92.61	95.35
	✓	95.71	74.31	85.16	77.50	100.00	100.00	93.14	96.51
		(+1.42)	(+2.75)	(+2.07)	(0.00)	(0.00)	(0.00)	(+0.53)	(+1.16)
GPT-3.5-turbo	✗	78.75	57.09	51.26	72.00	92.16	60.75	67.33	81.40
	✓	80.15	60.63	52.37	72.00	92.16	58.25	67.33	81.40
		(+1.40)	(+3.54)	(+1.11)	(0.00)	(0.00)	(-2.50)	(0.00)	(0.00)
Vicuna-33b	✗	40.33	26.38	15.70	37.50	52.94	32.00	14.67	40.70
	✓	40.49	29.53	20.35	47.50	68.75	34.50	14.00	43.02
		(+0.16)	(+3.15)	(+4.65)	(+10.00)	(+15.81)	(+2.50)	(-0.67)	(+2.32)
Vicuna-13b	✗	33.79	22.05	32.31	41.00	68.75	31.00	2.00	29.07
	✓	37.56	23.62	33.15	48.50	68.75	31.50	4.00	45.35
		(+3.77)	(+1.57)	(+0.84)	(+7.50)	(0.00)	(+0.50)	(+2.00)	(+16.28)
Vicuna-7b	✗	17.52	21.65	7.24	37.00	52.94	34.00	0.00	25.58
	✓	17.68	20.47	7.24	36.50	52.94	35.00	0.00	25.58
		(+0.16)	(-1.18)	(0.00)	(-0.50)	(0.00)	(+1.00)	(0.00)	(0.00)

KNOWLEDGE TECHNOLOGY

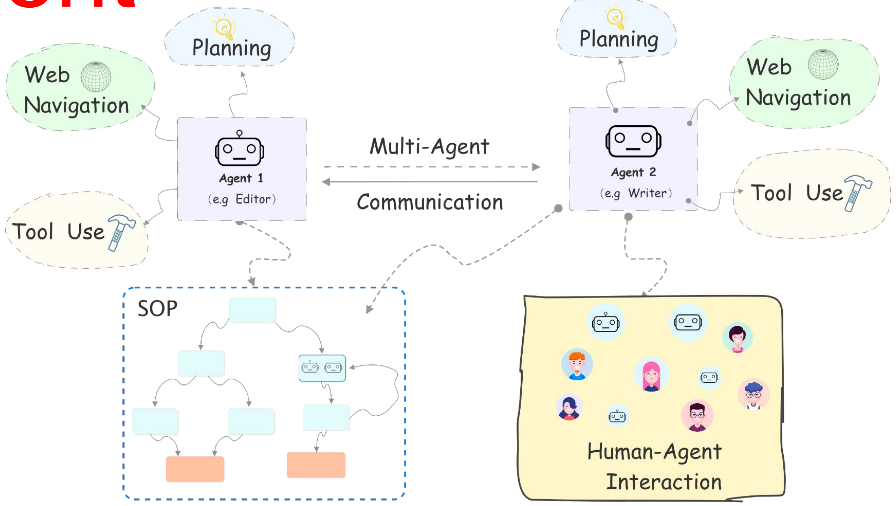
Outline

- Motivation
- Knowledge learned with RL
  - Internally rewarded reinforcement learning
  - Multimodal association with unsupervised reinforcement learning
- LLM utilization
  - Emergent abilities and Fine-tuning
  - LLM Prompt Reasoning
- LLM Agent
  - Structures
  - Instances
  - Trends
- IROS 2023 Related

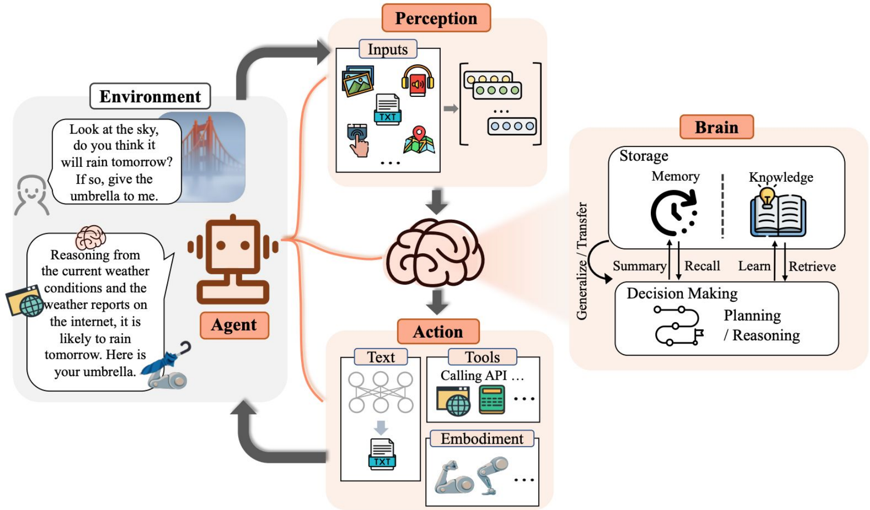
LLM-based Agent



Wang, Lei, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen et al. "A survey on large language model based autonomous agents." *arXiv preprint arXiv:2308.11432* (2023).



https://github.com/aiwaves-cn/agents



Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

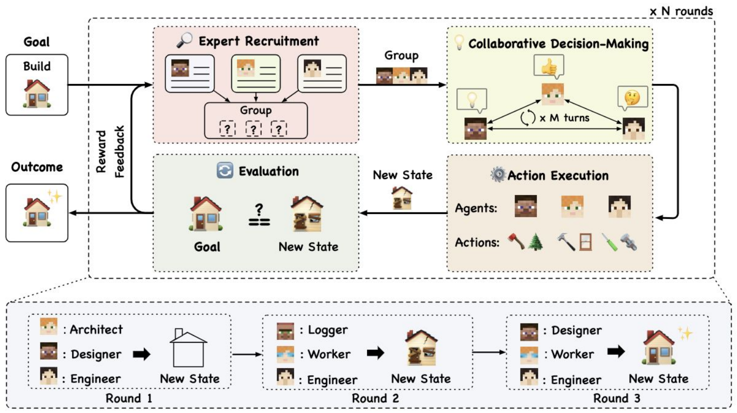


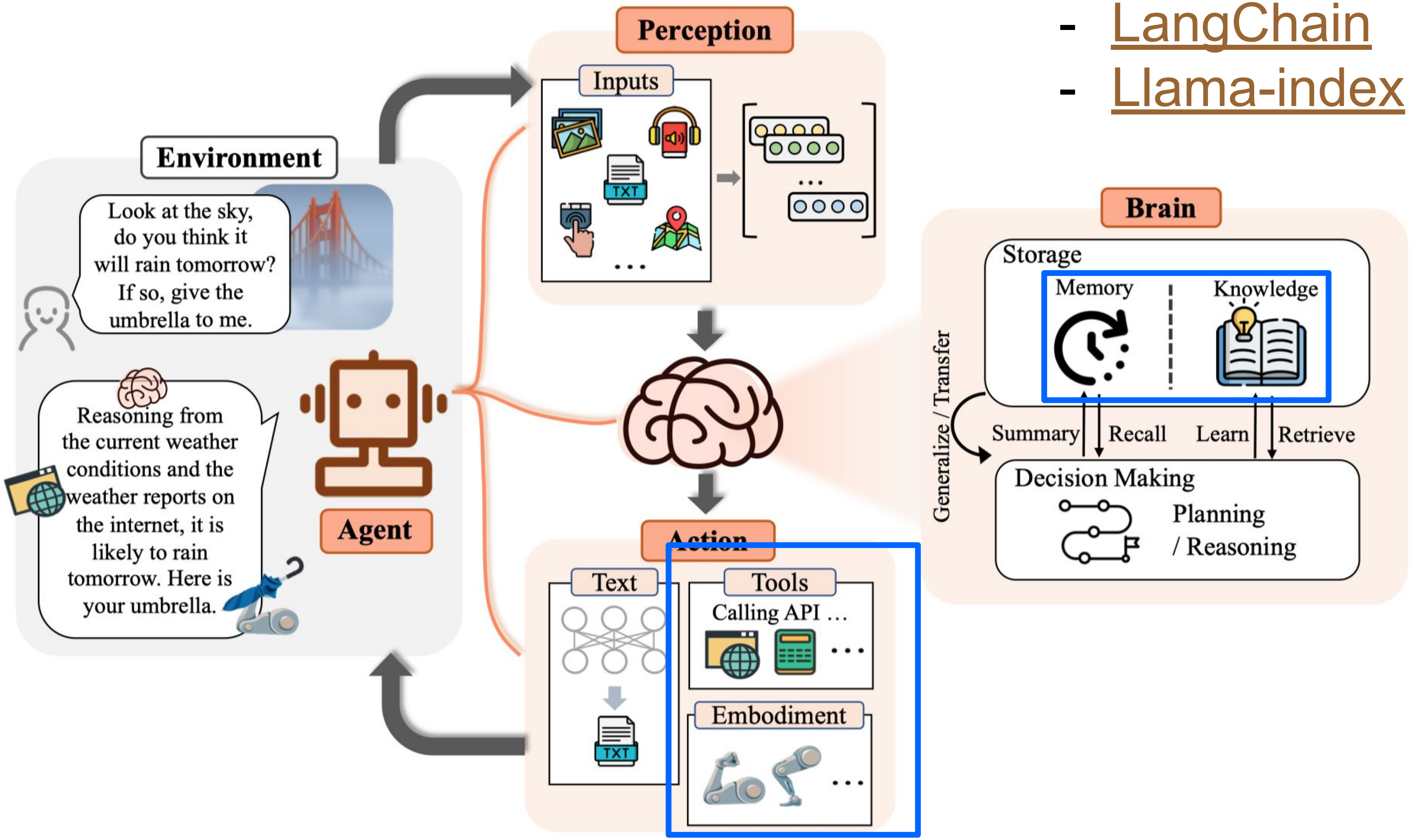
Figure 1: An illustration of the AgentVerse.

https://github.com/OpenBMB/AgentVerse

(multi-agents)

KNOWLEDGE TECHNOLOGY

LLM agent architecture



- LangChain
- Llama-index

Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

Tools

- APIs
- Modular models
- …

KNOWLEDGE TECHNOLOGY

Tools

- **Description**
- Self-thinking

```python
from pydantic import BaseModel, Field


class CalculatorInput(BaseModel):
    question: str = Field()


tools.append(
    Tool.from_function(
        func=llm_math_chain.run,
        name="Calculator",
        description="useful for when you need to answer questions about math",
        args_schema=CalculatorInput,
        # coroutine= ... <- you can specify an async method if desired as well
    )
)
```

```python
# Construct the agent. We will use the default agent type here.
# See documentation for a full list of options.
agent = initialize_agent(
    tools, llm, agent=AgentType.ZERO_SHOT_REACT_DESCRIPTION, verbose=True
)
```

```python
agent.run(
    "Who is Leo DiCaprio's girlfriend? What is her current age raised to the 0.43 power?"
)
```

LangChain code example of using tools

KNOWLEDGE TECHNOLOGY

Tools

- Description
- **Self-thinking**

````
> Entering new AgentExecutor chain...
I need to find out Leo DiCaprio's girlfriend's name and her age
Action: Search
Action Input: "Leo DiCaprio girlfriend"
Observation: After rumours of a romance with Gigi Hadid, the Oscar winner has seemingly moved on.
Thought:I still need to find out his current girlfriend's name and age
Action: Search
Action Input: "Leo DiCaprio current girlfriend"
Observation: Just Jared on Instagram: "Leonardo DiCaprio & girlfriend Camila Morrone couple up fo
Thought:Now that I know his girlfriend's name is Camila Morrone, I need to find her current age
Action: Search
Action Input: "Camila Morrone age"
Observation: 25 years
Thought:Now that I have her age, I need to calculate her age raised to the 0.43 power
Action: Calculator
Action Input: 25^(0.43)

> Entering new LLMMathChain chain...
25^(0.43)```text
25**(0.43)
```
...numexpr.evaluate("25**(0.43)")...

Answer: 3.991298452658078
> Finished chain.

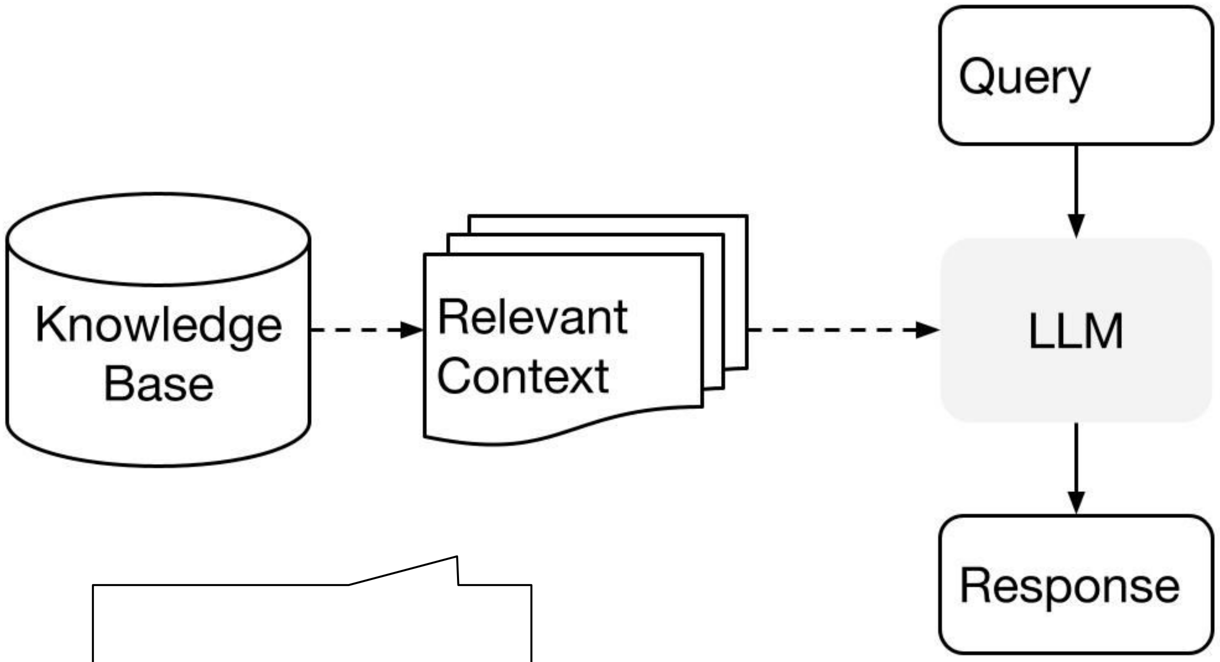
Observation: Answer: 3.991298452658078
Thought:I now know the final answer
Final Answer: Camila Morrone's current age raised to the 0.43 power is approximately 3.99.

> Finished chain.
````


KNOWLEDGE TECHNOLOGY
````

Memory

- Short-term memory: in prompts
- Long-term memory:
  - **Embeddings**
    - **Indexing**
    - **retrieval**



Block ~= cube
Hotdog != dog
Hot cat ~= cat

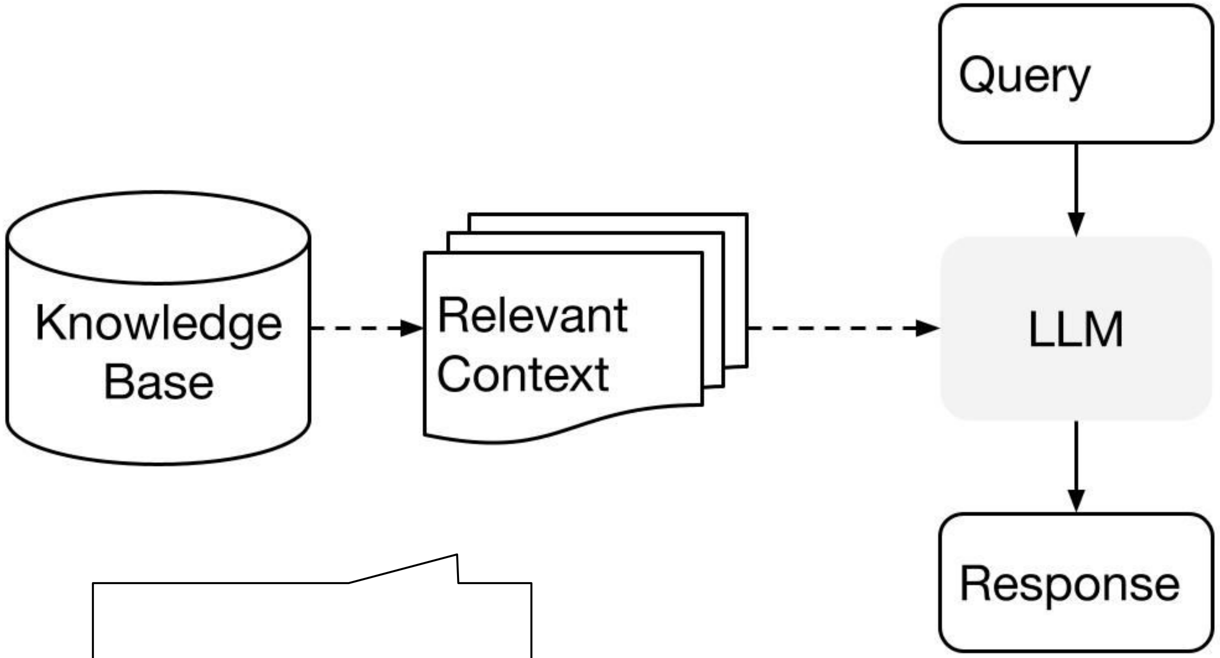
Semantic similarity

KNOWLEDGE TECHNOLOGY

Memory

- Short-term memory: in prompts
- Long-term memory:
  - **Embeddings**
    - **Indexing**
    - **retrieval**
  - Database
  - Other structured format

Llama-index with local data



Block ~= cube
Hotdog != dog
Hot cat ~= cat

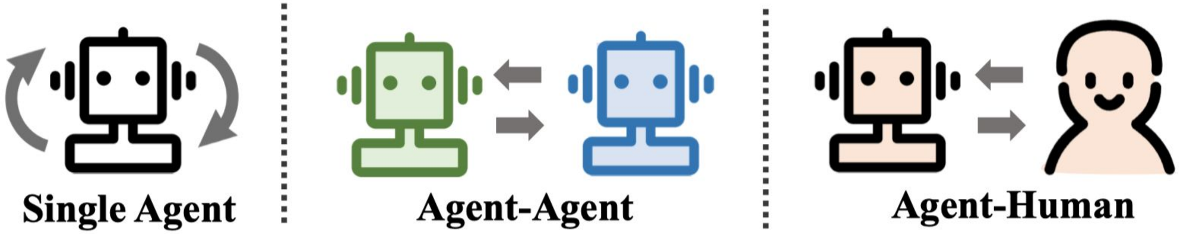
Semantic similarity

Outline

- Motivation
- Knowledge learned with RL
  - Internally rewarded reinforcement learning
  - Multimodal association with unsupervised reinforcement learning
- LLM utilization
  - Emergent abilities and Fine-tuning
  - LLM Prompt Reasoning
- LLM Agent
  - Structures
  - Instances
  - Trends
- IROS 2023 Related

Instances
(single, multi-, human-in-the-loop agent)

KNOWLEDGE
TECHNOLOGY



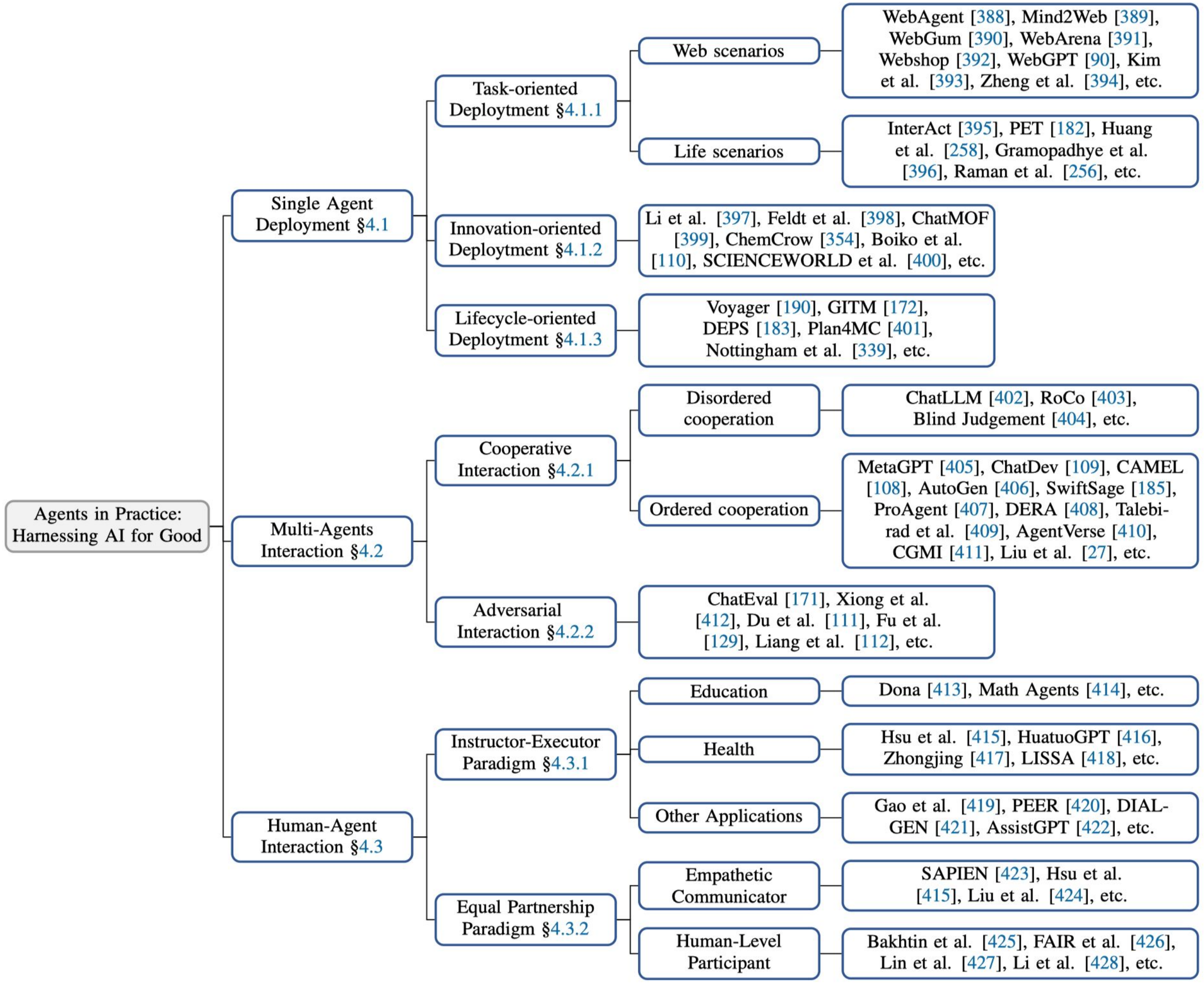
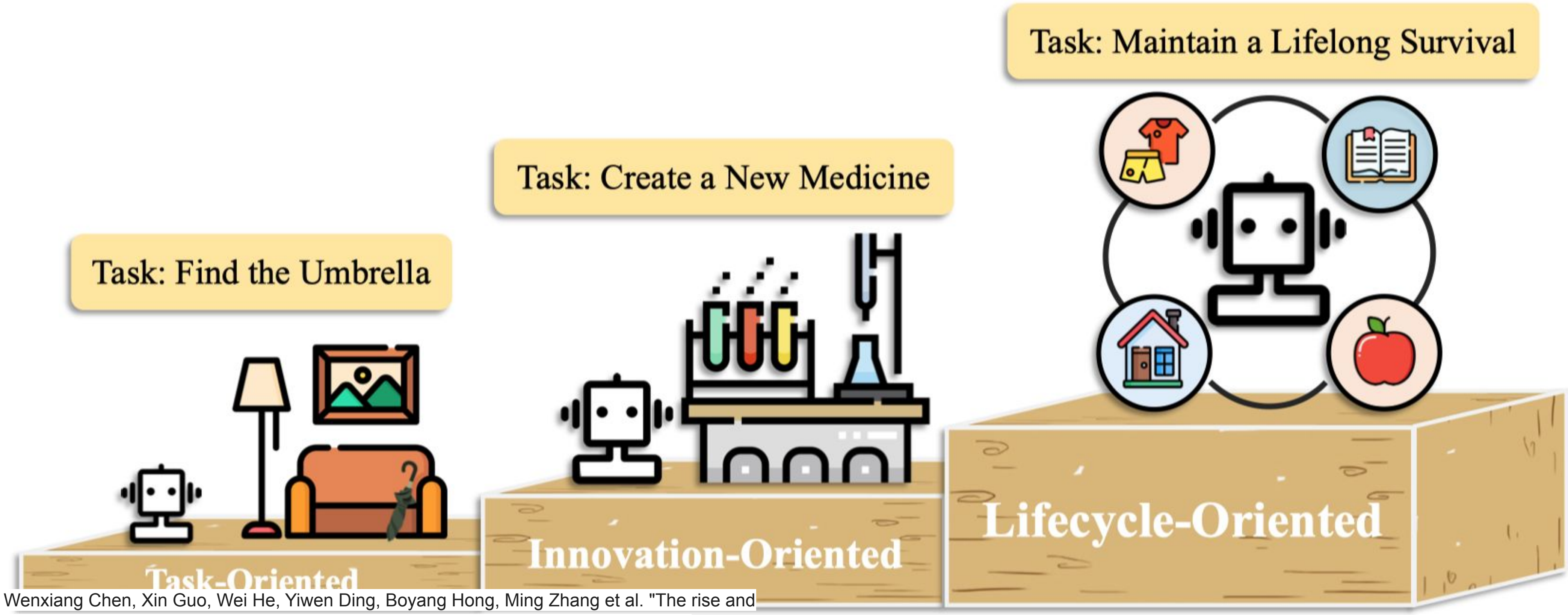


Figure 6: Typology of applications of LLM-based agents.

Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

KNOWLEDGE TECHNOLOGY

Single Agent



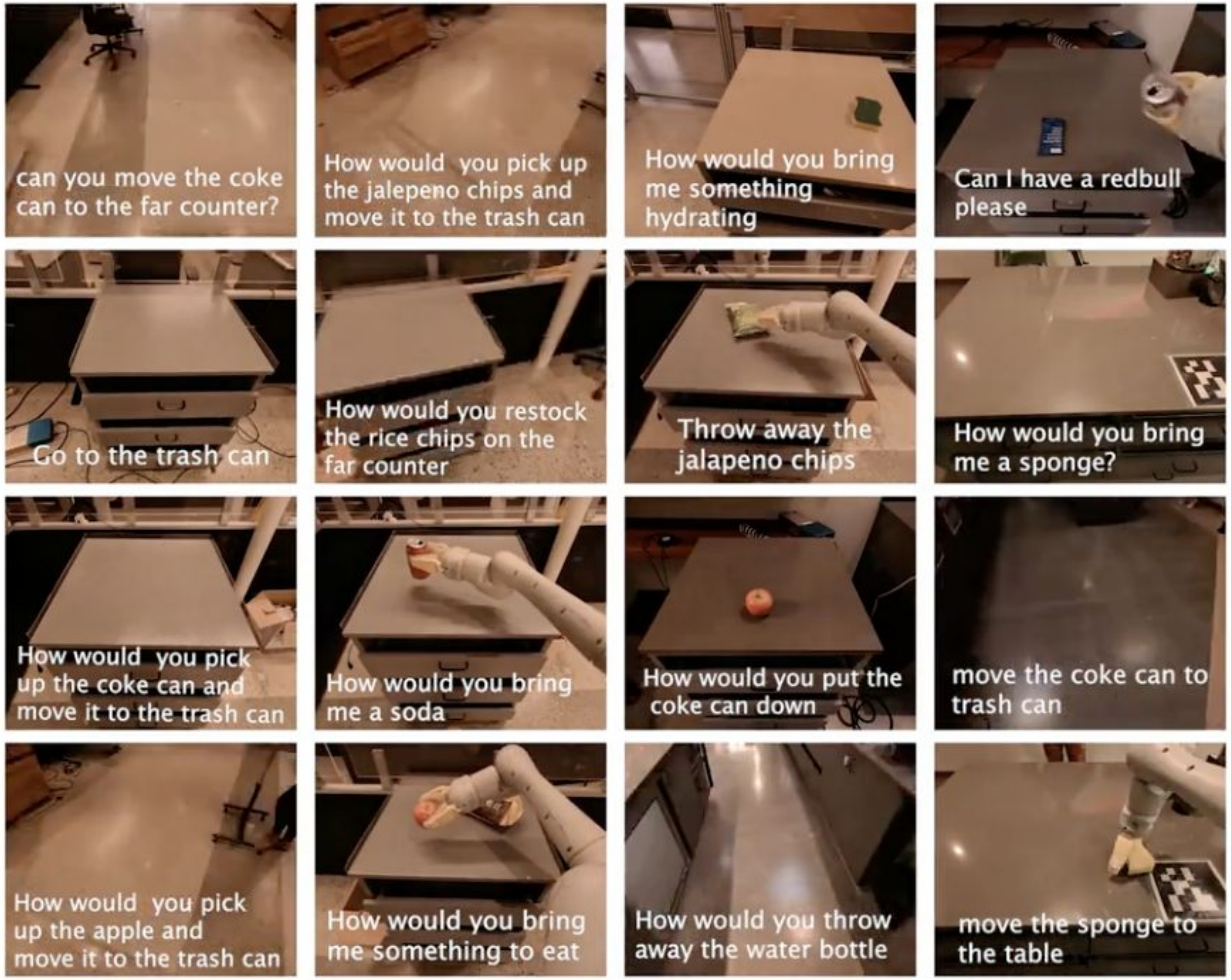
Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

Single Agent

Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Ho
potential of large language model based agents: A survey." *arXiv preprir*

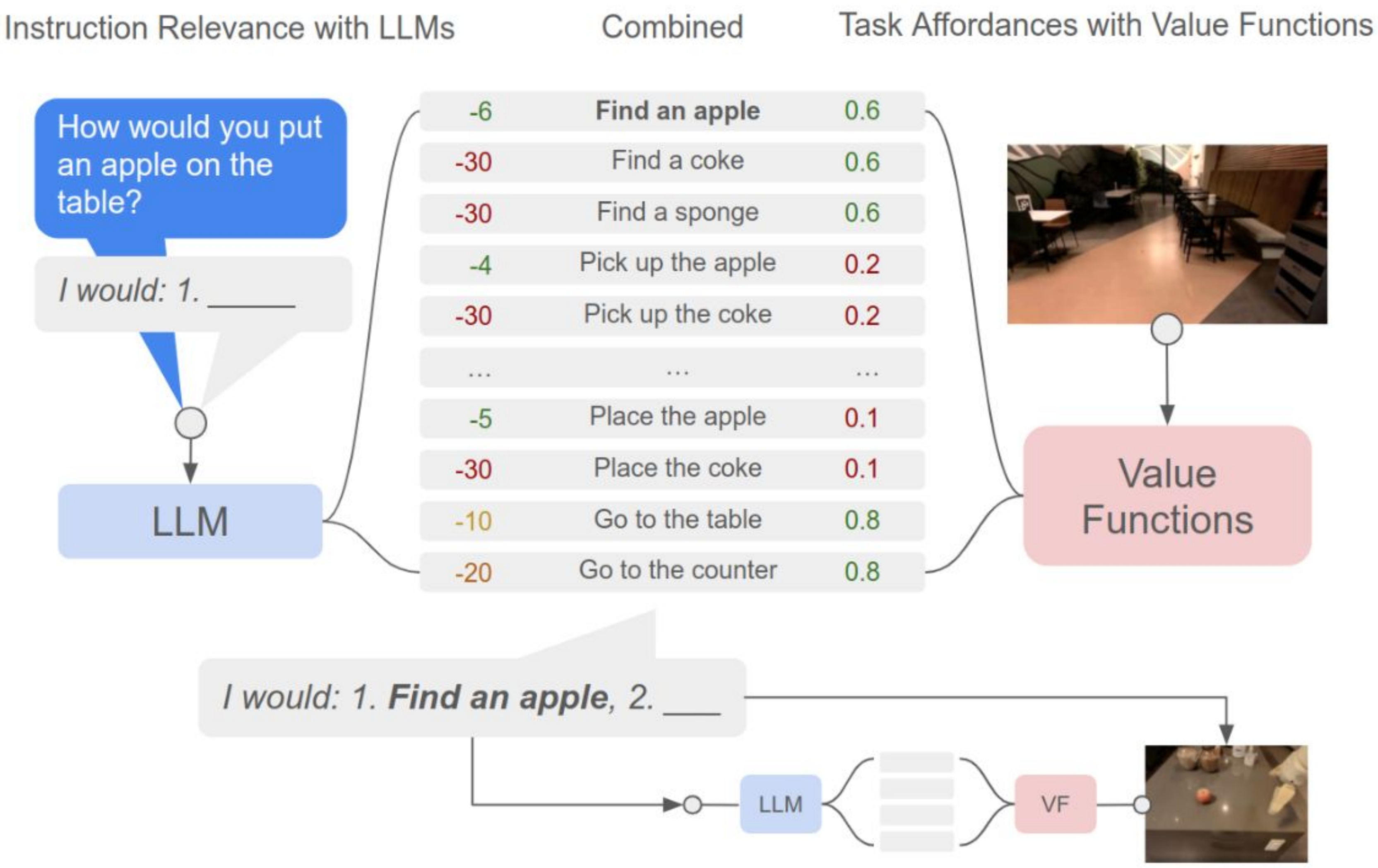
Say-Can

(kitchen)

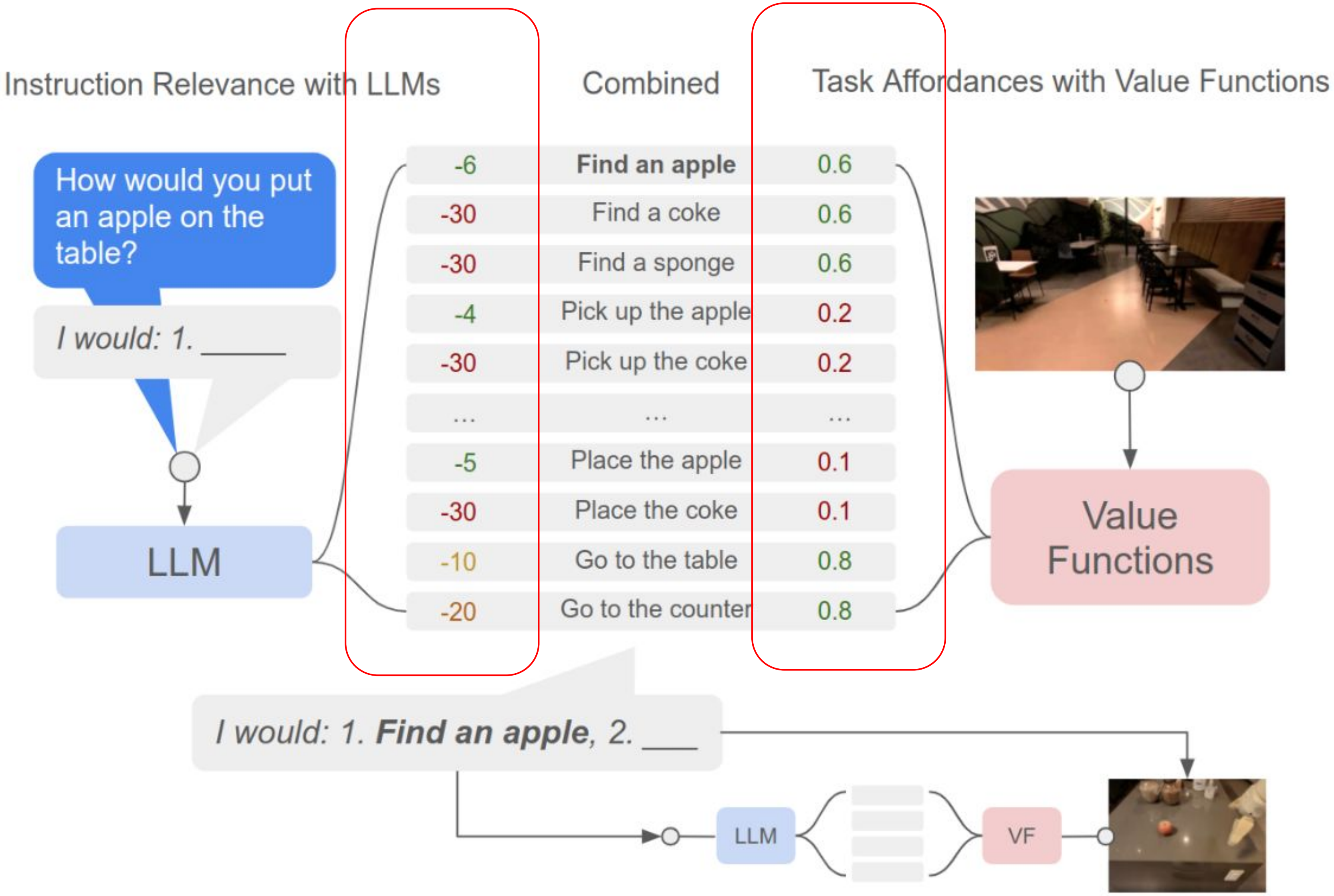


KNOWLEDGE TECHNOLOGY

Say-Can


Instruction Relevance with LLMs
Combined
Task Affordances with Value Functions
How would you put an apple on the table?
I would: 1. ____
LLM
-6 Find an apple 0.6
-30 Find a coke 0.6
-30 Find a sponge 0.6
-4 Pick up the apple 0.2
-30 Pick up the coke 0.2
... ...
-5 Place the apple 0.1
-30 Place the coke 0.1
-10 Go to the table 0.8
-20 Go to the counter 0.8
Value Functions
I would: 1. Find an apple, 2. ___
LLM
VF

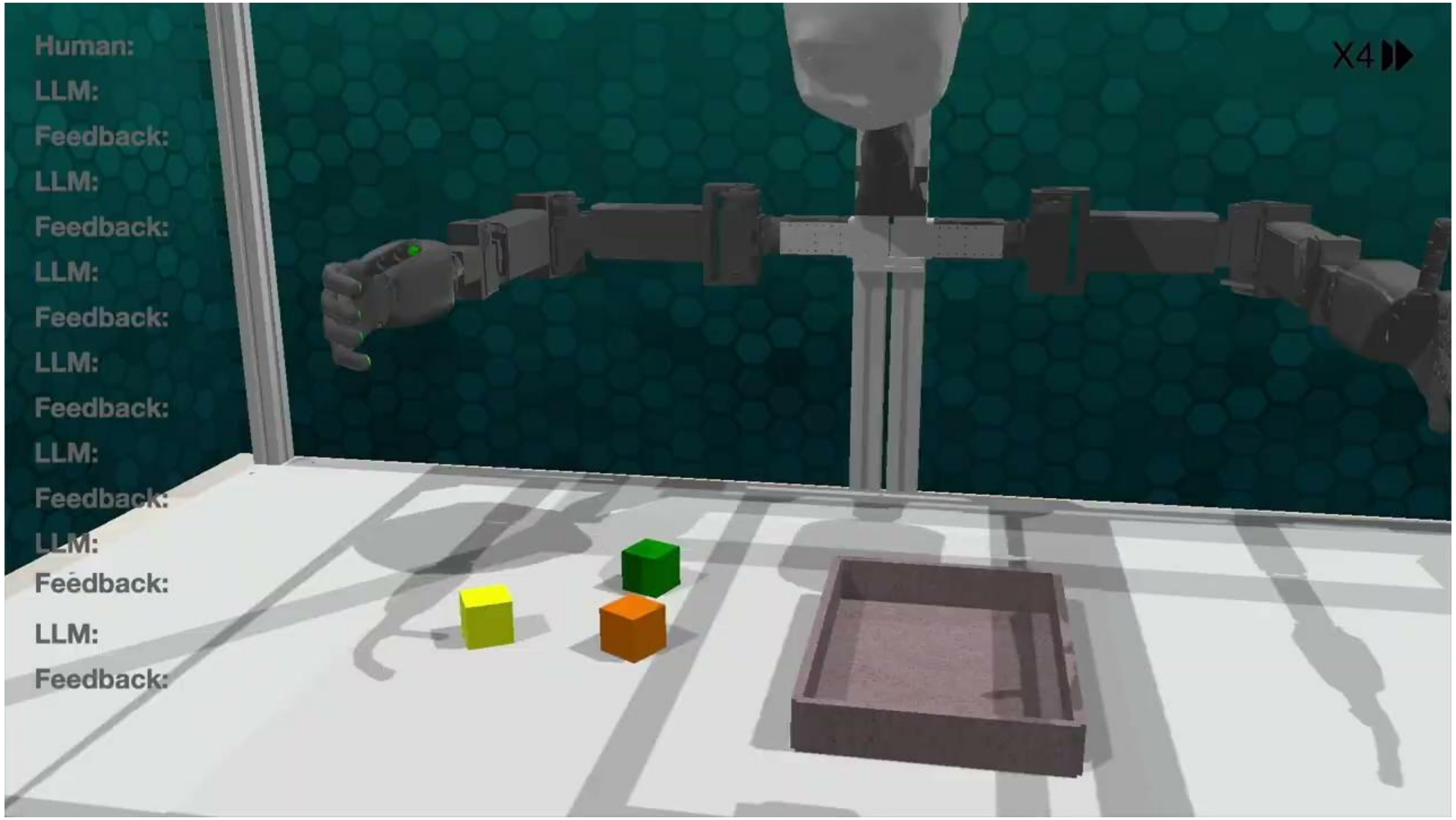
Say-Can



KNOWLEDGE TECHNOLOGY

▪ Active Perception with LLMs

[Matcha-agent](https://youtu.be/rMMeMTWmT0k)



https://youtu.be/rMMeMTWmT0k

[3] Zhao, Xufeng, Mengdi Li, Cornelius Weber, Muhammad Burhan Hafez, and Stefan Wermter. "Chat with the environment: Interactive multimodal perception using large language models." *IEEE IROS 2023, Detroit, USA.*

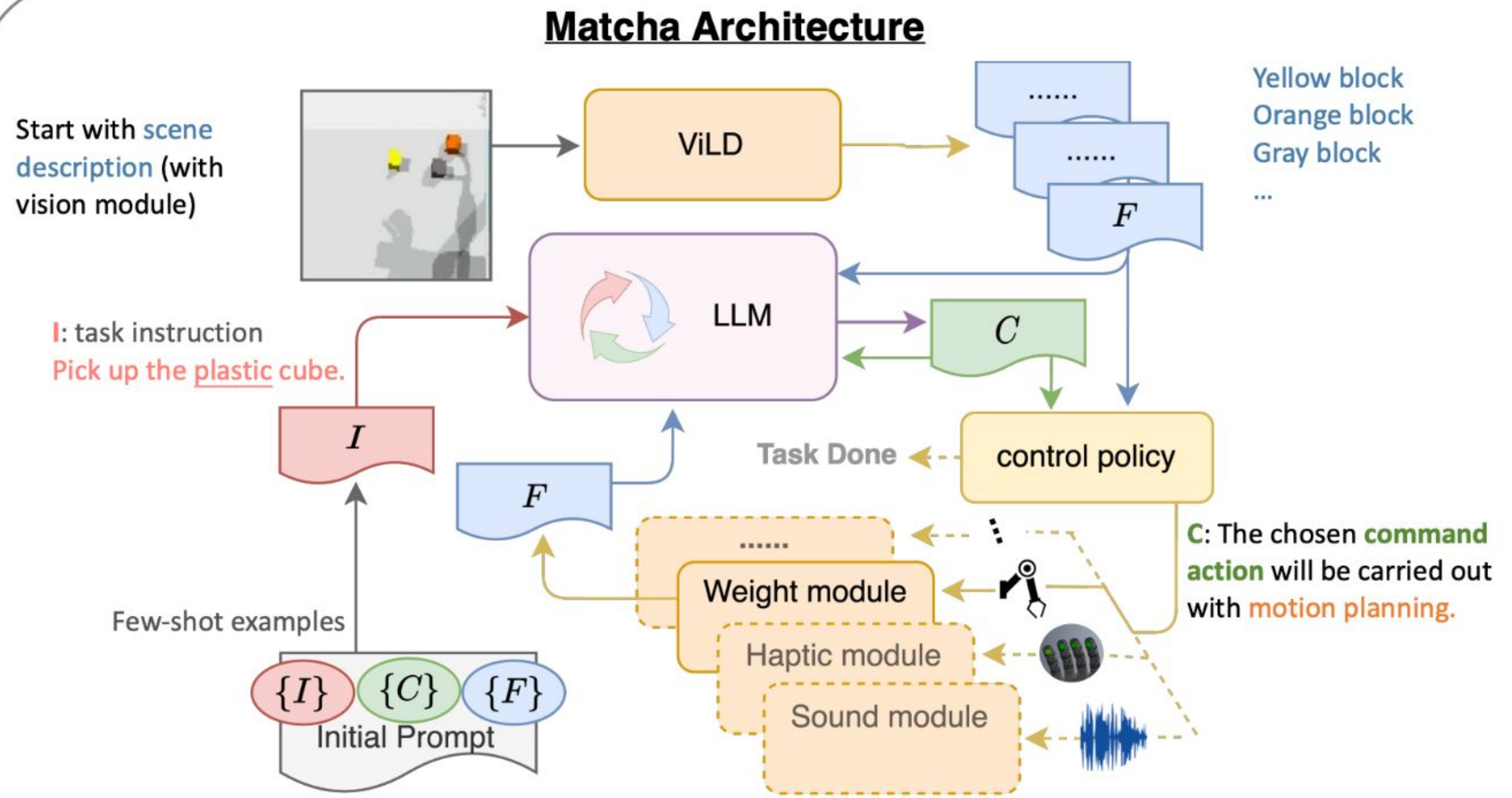
We propose the Matcha **framework,** comprising an LLM and multiple **multimodal modules**, enabling the robot to engage with its surroundings through high-level **LLM planning**.

(Multimodal environment chatting agent)

Robotic Perceptions

- **Passive perceptions**
  - **Epistemic uncertainty**
- **Active perceptions**
  - **Increased complexity**
  - **Generalizability**
- **Robots with LLMs**
  - **Causal reasoning ability with distilled human knowledge inside**
  - **In-context learning ability with few-shot prompts**

Matcha Architecture



F: Feeding back the multimodal response to the LLM and loop until the task is done.

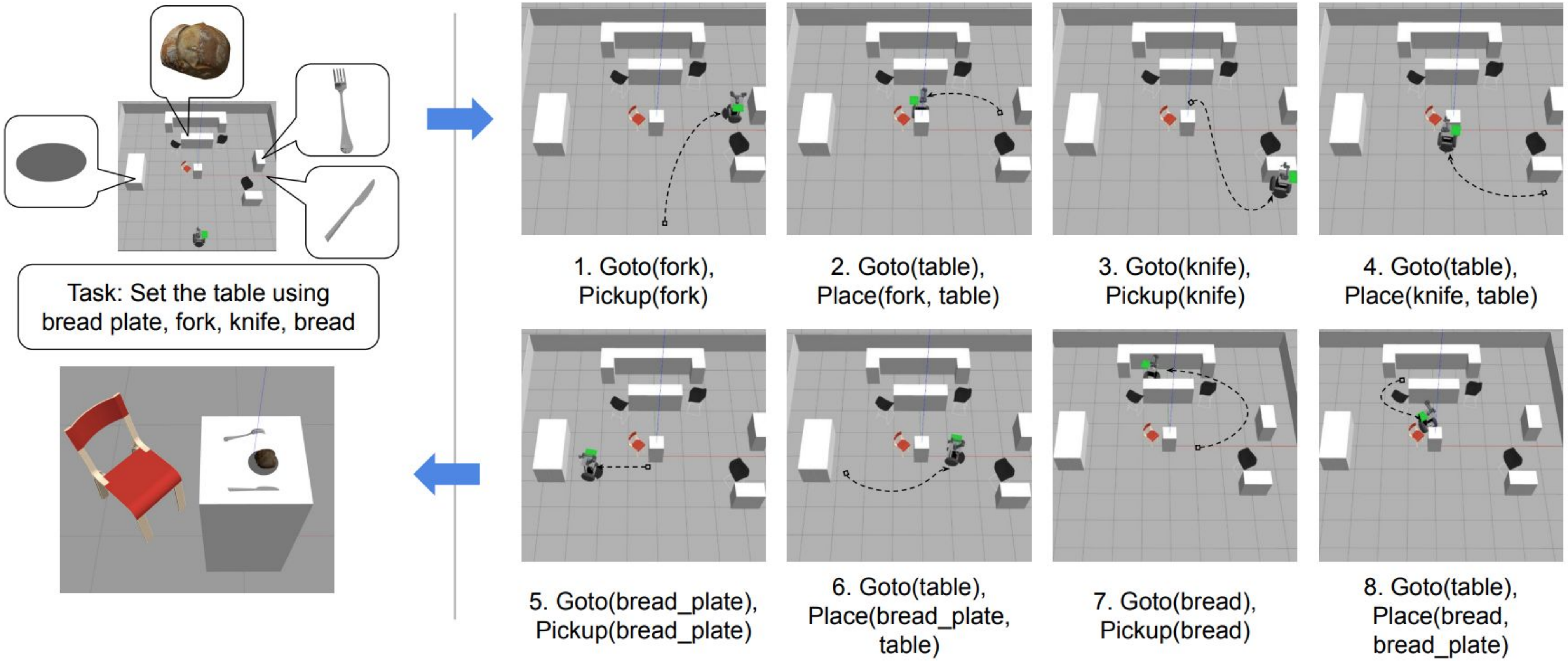
- Sound module
  - Similar modality descriptions for different materials
  - Quite different descriptions for the same material
- Haptic module
- Weight module

LLM	Type of Description	Success Rate
text-ada-001	Indistinct	19.05%
	Distinct	28.57%
text-davinci-003	Indistinct	56.67%
	Distinct	90.57%

*Random guess in principle: 33.33%

- NICOL robot
- Coppeliasim simulator
- LLM: OpenAI API text-davinci-003
- Works without any fine-tuning

LLM-GROP (object arrangement)



Ding, Yan, Xiaohan Zhang, Chris Paxton, and Shiqi Zhang. "Task and motion planning with large language models for object rearrangement." *arXiv preprint arXiv:2303.06247* (2023). IROS 2023.

KNOWLEDGE TECHNOLOGY

TidyBot

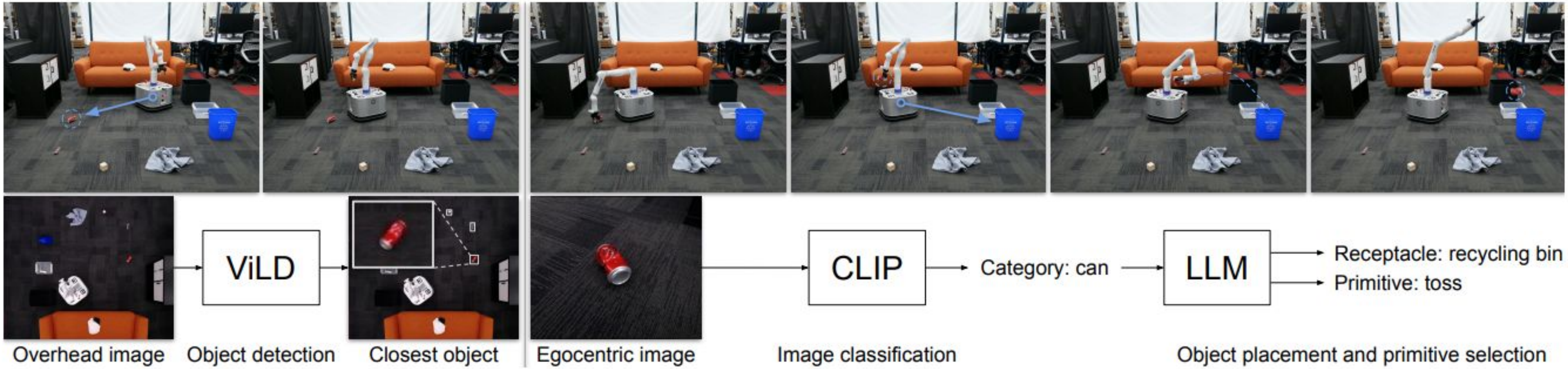
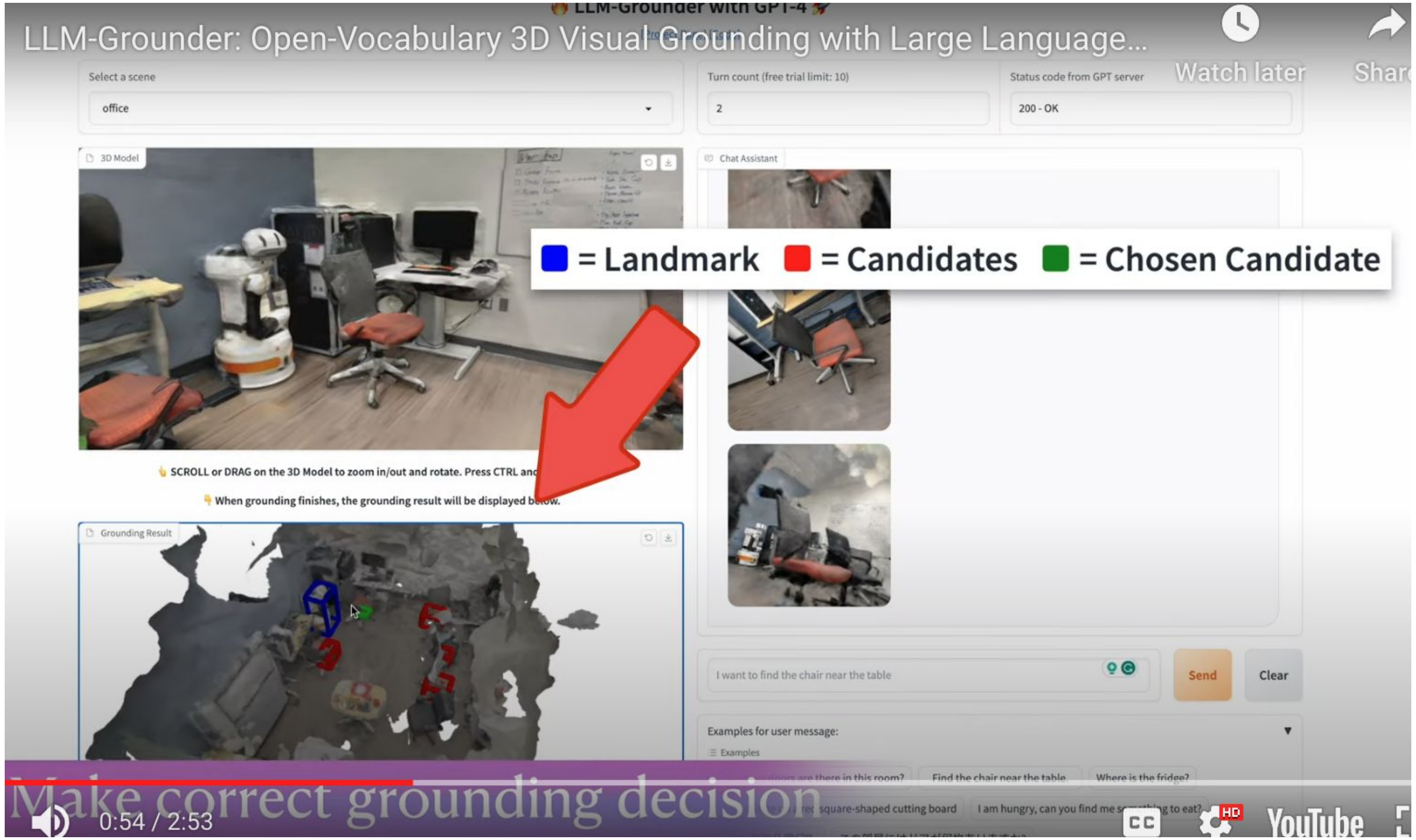


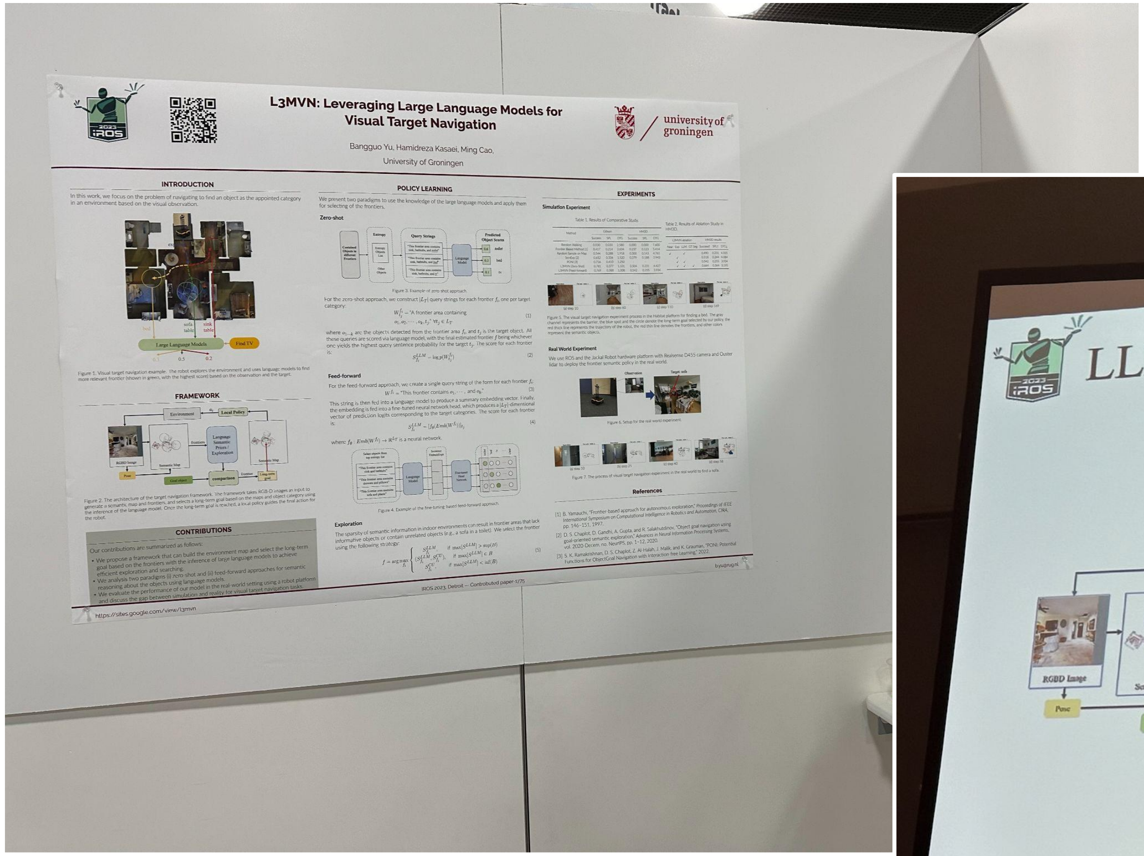
Fig. 2 System overview. Once the user's preferences have been summarized with an LLM, TidyBot will localize the closest object on the floor, move to get a close-up view with its egocentric camera, predict the object's category using CLIP, use the LLM-summarized rules to select a receptacle and manipulation primitive, and then execute the primitive to put the object into the selected receptacle, repeating this entire process until no more objects can be found on the floor.

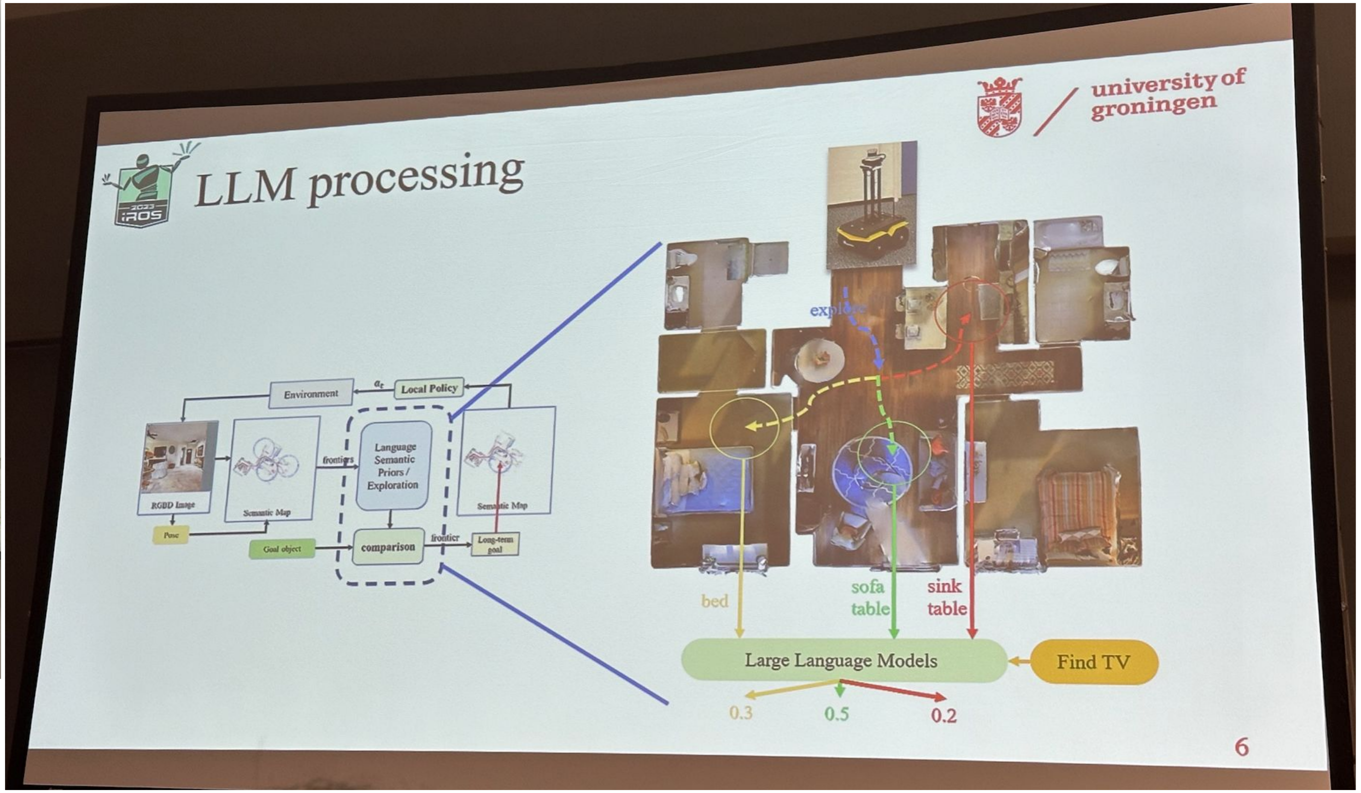
Wu, Jimmy, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. "Tidybot: Personalized robot assistance with large language models." *arXiv preprint arXiv:2305.05658* (2023).

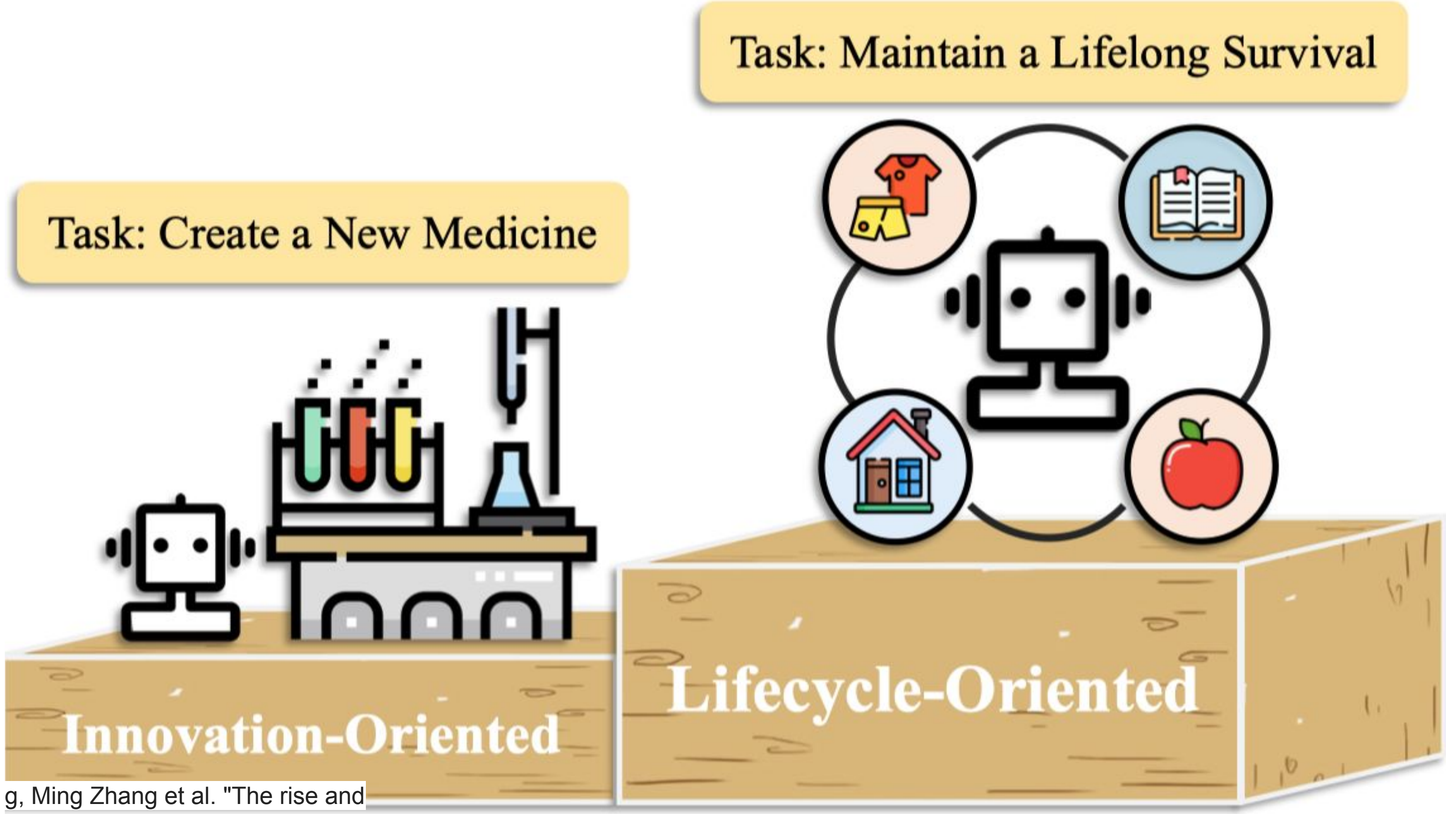
- [Chat with NeRF: Grounding 3D Objects in Neural Radiance Field through Dialog]()



- L3MVN: Leveraging Large Language Models for Visual **Target Navigation**
    - Utilizing common sense in LLMs



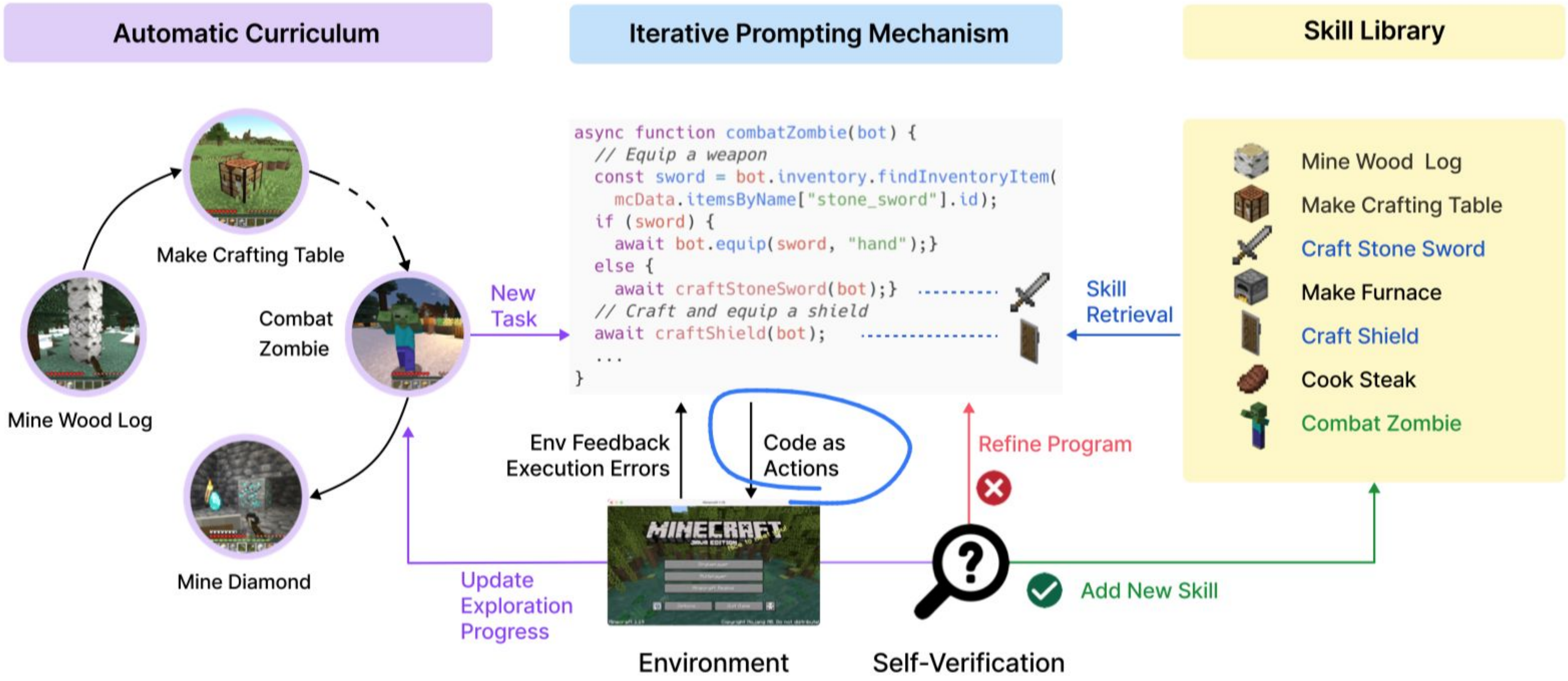




g, Ming Zhang et al. "The rise and
arXiv:2309.07864 (2023).

Voyager: An Open-Ended Embodied Agent with Large Language Models

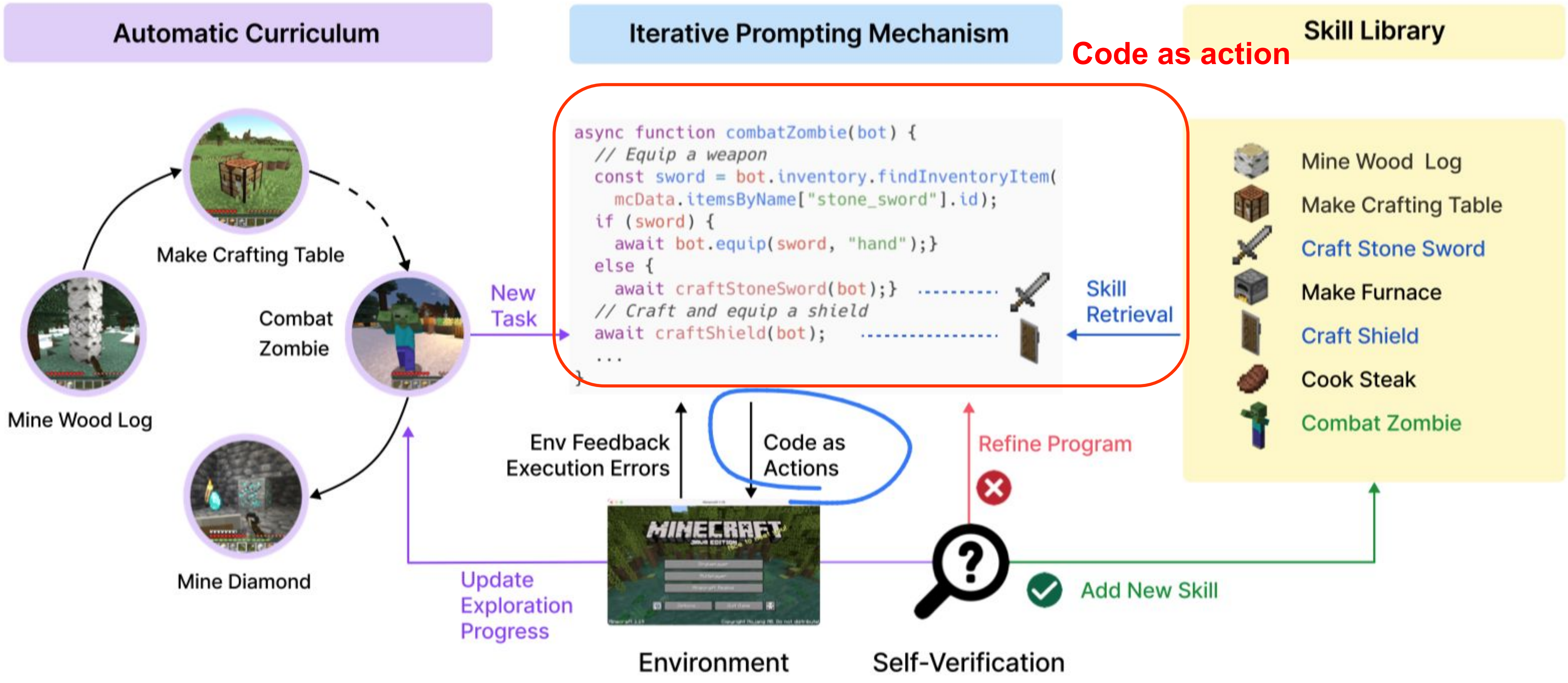
Wang, Guanzhi, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. "Voyager: An open-ended embodied agent with large language models." *arXiv preprint arXiv:2305.16291* (2023).



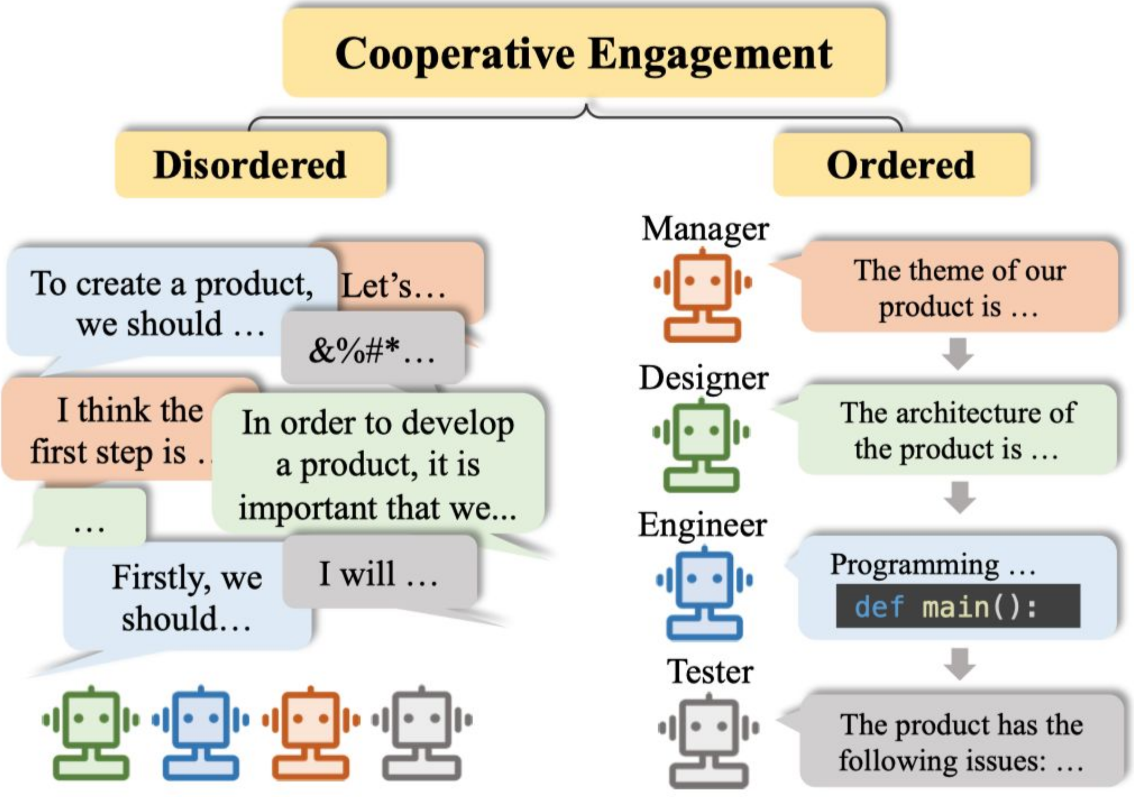
EDGE
OLOGY

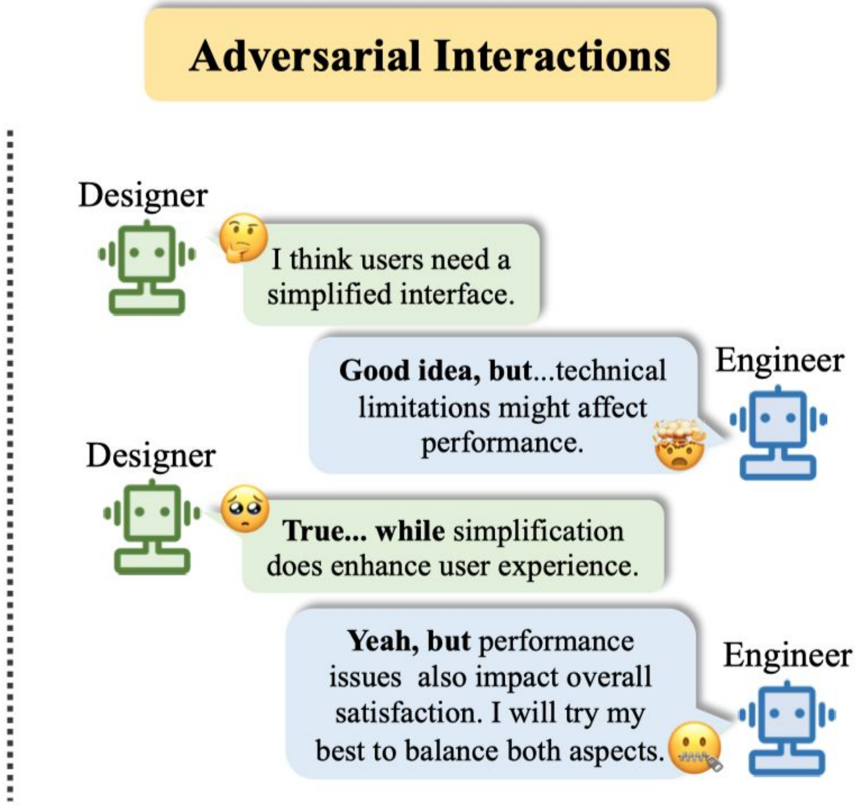
Voyager: An Open-Ended Embodied Agent with Large Language Models

Wang, Guanzhi, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. "Voyager: An open-ended embodied agent with large language models." *arXiv preprint arXiv:2305.16291* (2023).



Agent-Agent





KNOWLEDGE TECHNOLOGY

Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

Generative Agents: Interactive Simulacra of Human Behavior

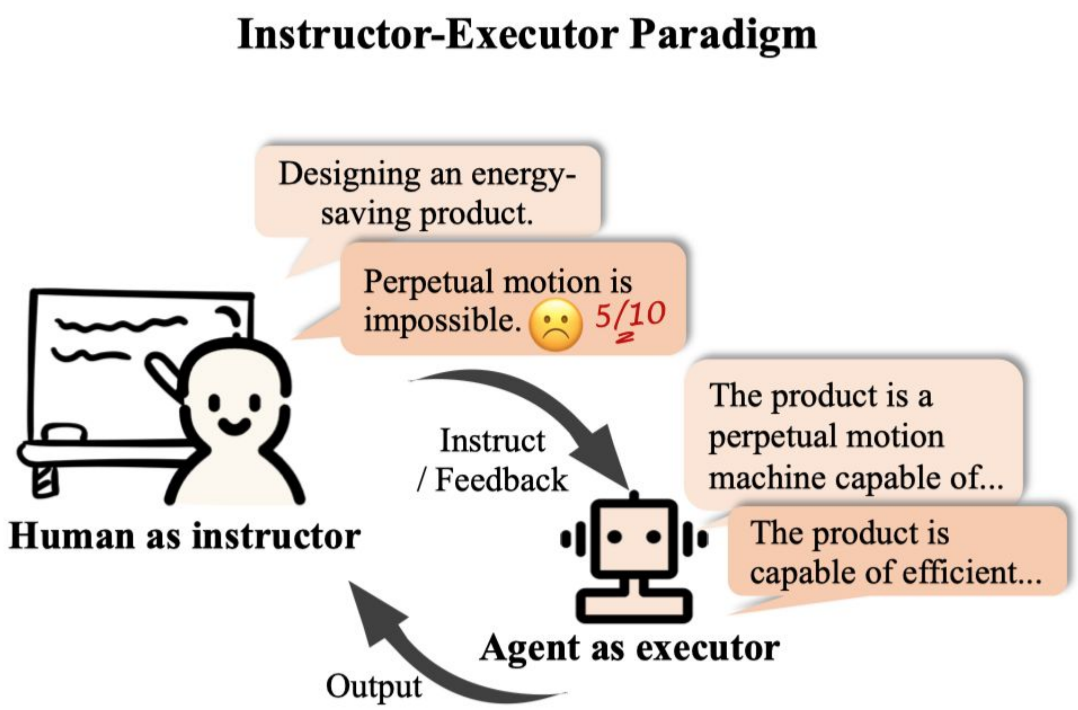
Park, Joon Sung, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. "Generative agents: Interactive simulacra of human behavior." *arXiv preprint arXiv:2304.03442* (2023).

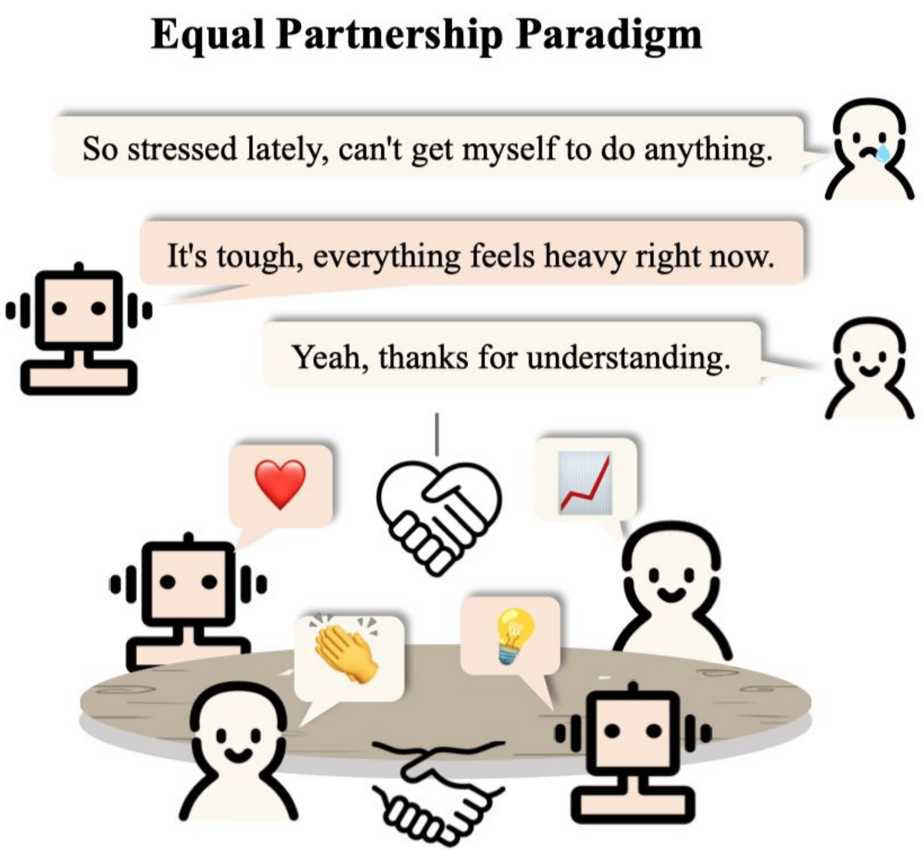
Multi-agent social events

- Prompt
- AI character
- Log memory

OWLEDGE
CHNOLOGY

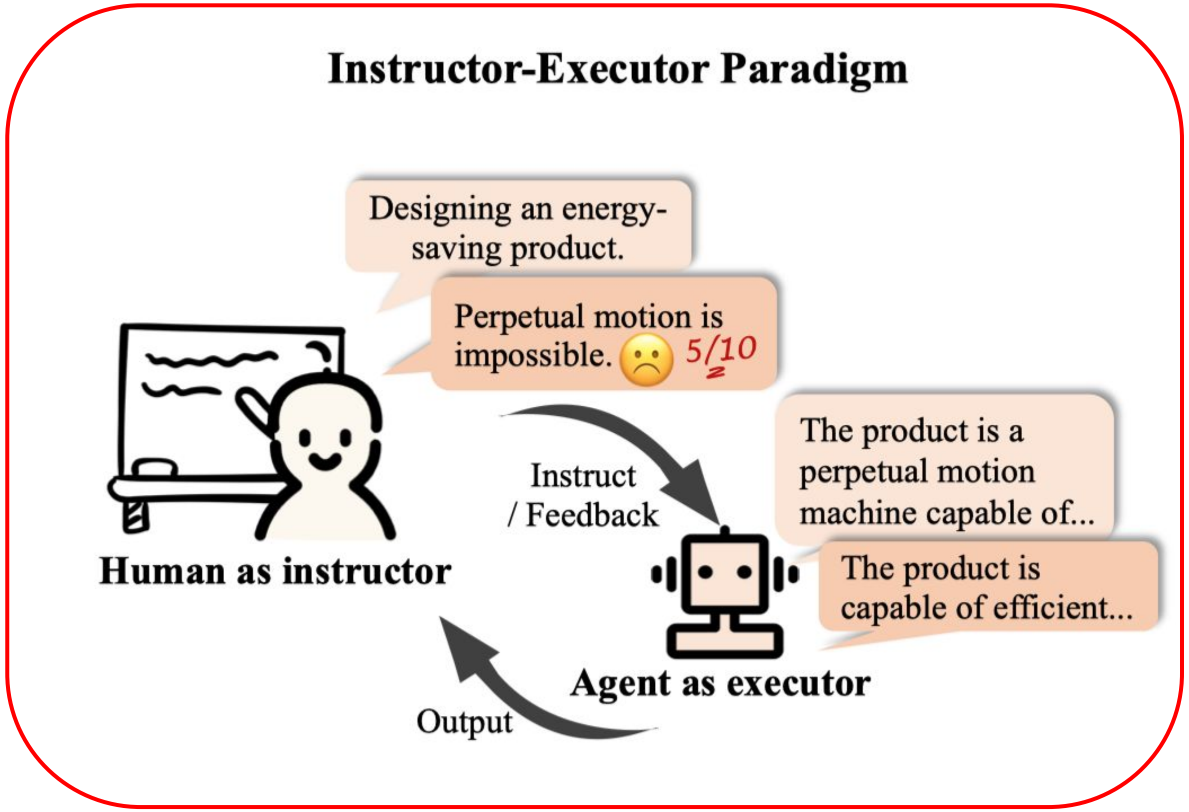
Human-Agent

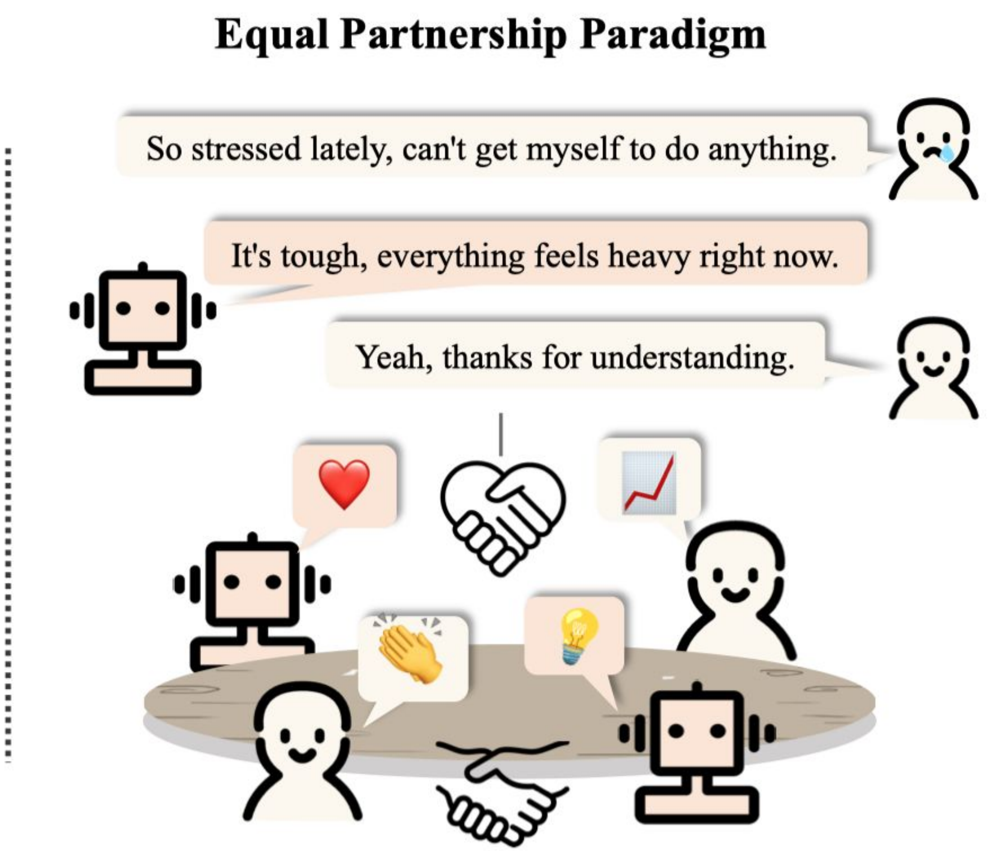




Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

Human-Agent

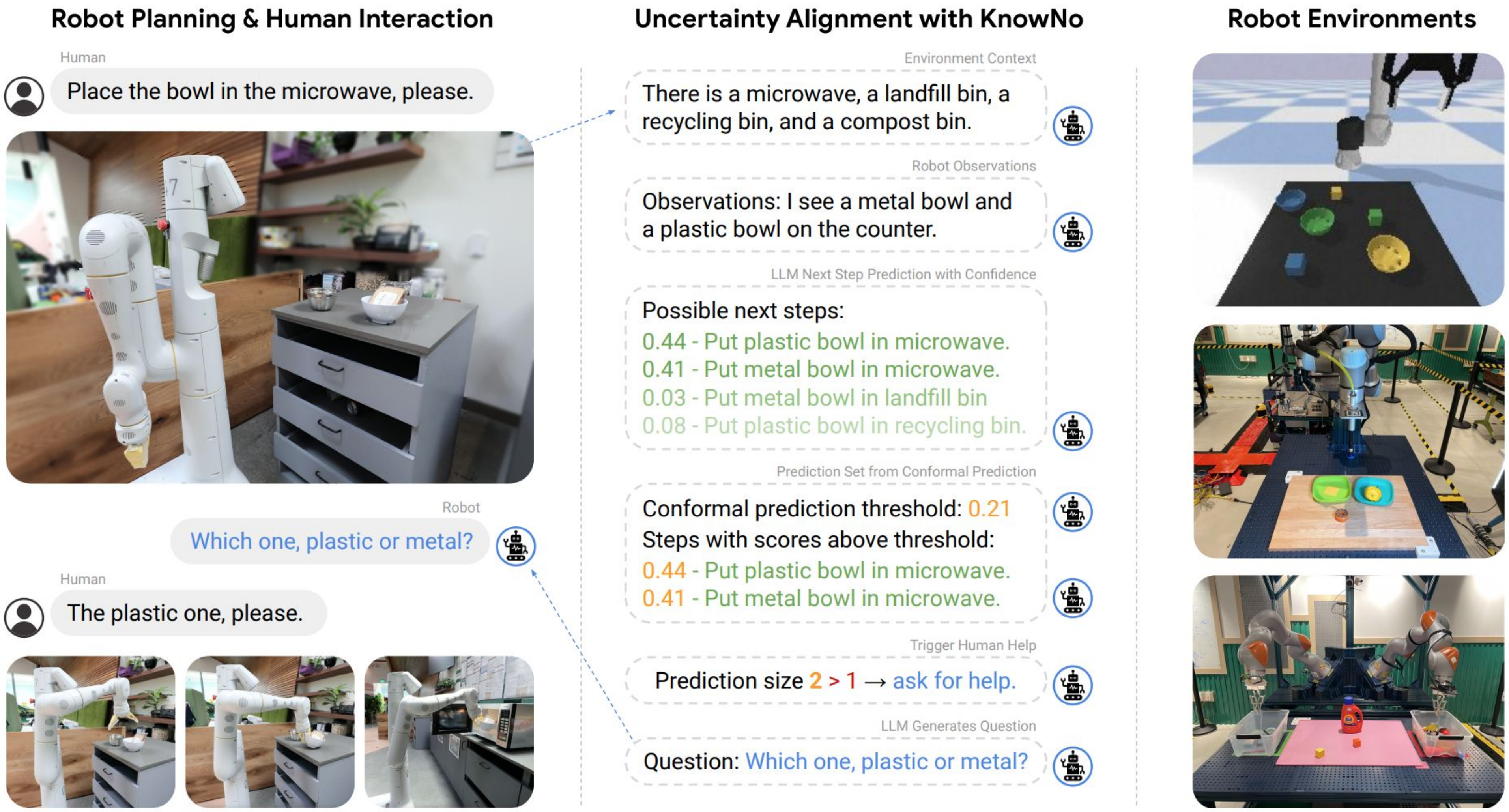




Interactive instructing

Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

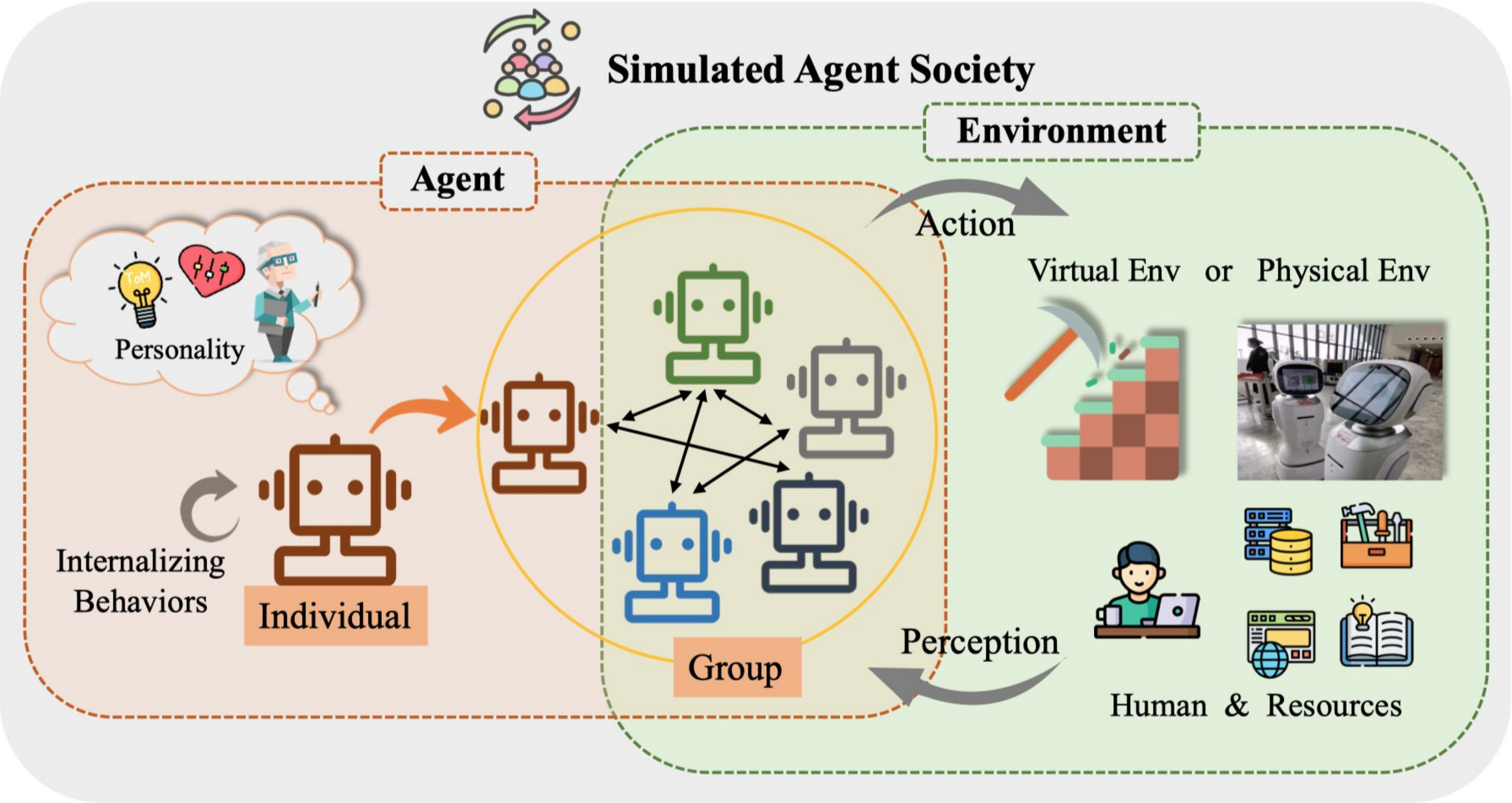
Know-No



Ren, Allen Z., Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, Noah Brown, Peng Xu et al. "Robots that ask for help: Uncertainty alignment for large language model planners." *arXiv preprint arXiv:2307.01928* (2023). CoRL 2023.

KNOWLEDGE TECHNOLOGY

Multiple characterized agents "inside" for strategies



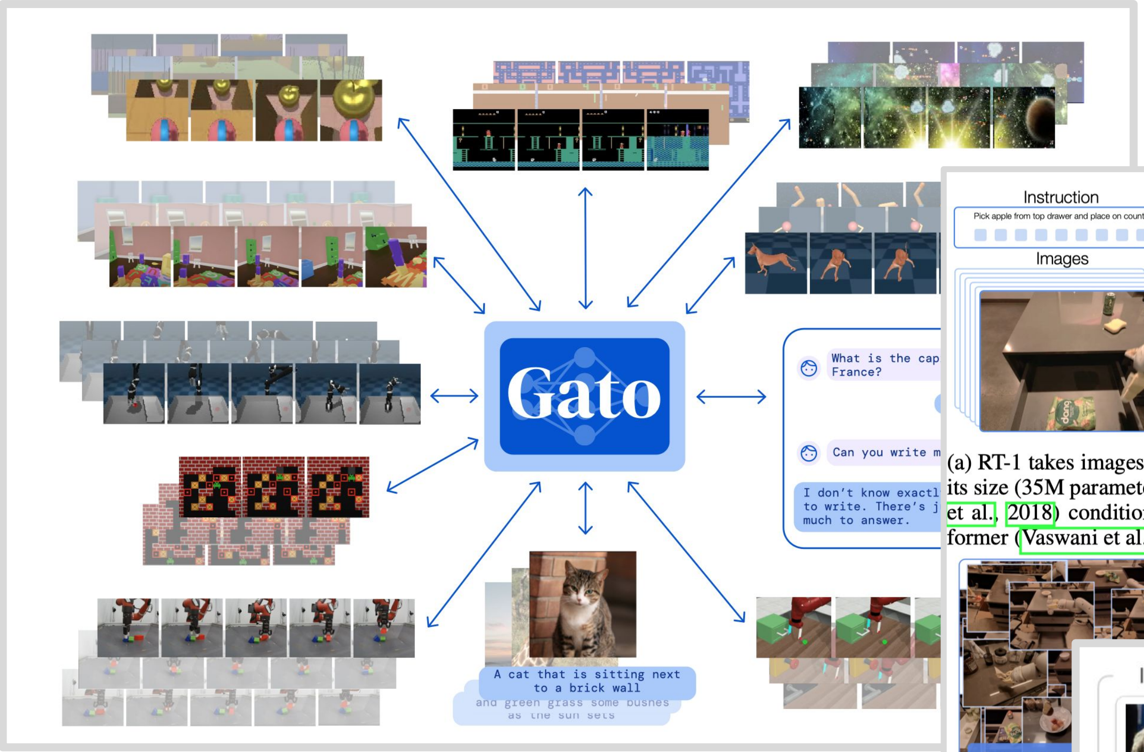
Xi, Zhiheng, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang et al. "The rise and potential of large language model based agents: A survey." *arXiv preprint arXiv:2309.07864* (2023).

Outline

- Motivation
- Knowledge learned with RL
  - Internally rewarded reinforcement learning
  - Multimodal association with unsupervised reinforcement learning
- LLM utilization
  - Emergent abilities and Fine-tuning
  - LLM Prompt Reasoning
- LLM Agent
  - Structures
  - Instances
  - Trends
- IROS 2023 Related

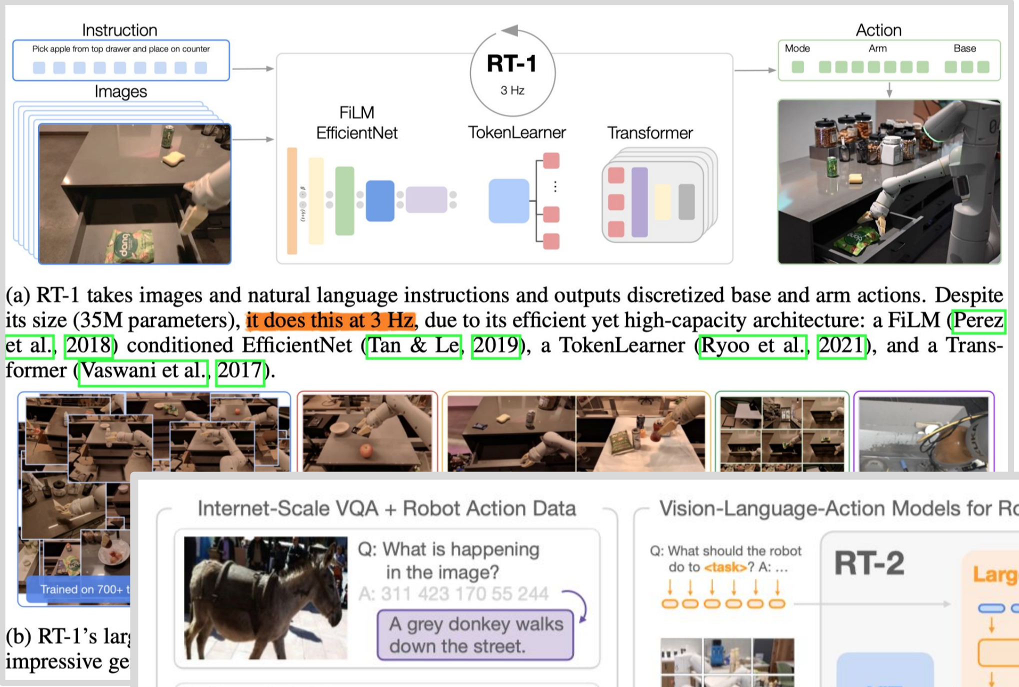
Limited by the representability of natural language

Generalist agent (unified)



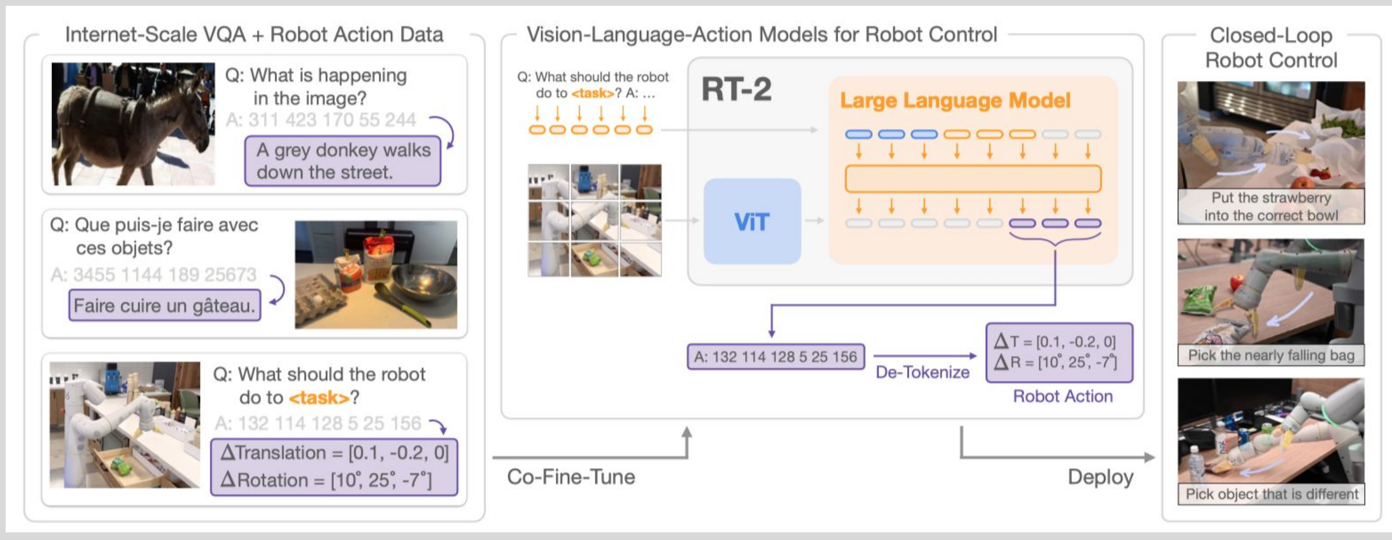
Reed, Scott, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez et al. "A generalist agent." *arXiv preprint arXiv:2205.06175* (2022).

Brohan, Anthony, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, Keerthana Gopalakrishnan et al. "Rt-1: Robotics transformer for real-world control at scale." *arXiv preprint arXiv:2212.06817* (2022).



(a) RT-1 takes images and natural language instructions and outputs discretized base and arm actions. Despite its size (35M parameters), it does this at 3 Hz, due to its efficient yet high-capacity architecture: a FiLM (Perez et al., 2018) conditioned EfficientNet (Tan & Le, 2019), a TokenLearner (Ryoo et al., 2021), and a Transformer (Vaswani et al., 2017).

(b) RT-1's larg
impressive ge



Brohan, Anthony, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding et al. "Rt-2: Vision-language-action models transfer web knowledge to robotic control." *arXiv preprint arXiv:2307.15818* (2023).

LLM Guided Agent (LLM+RL)

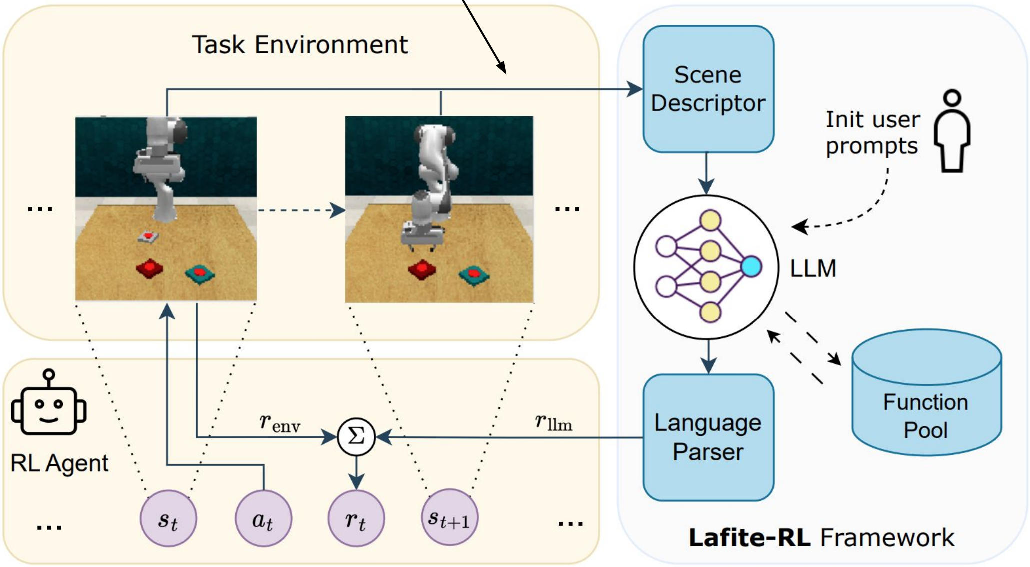
- LLM + RL: knowledge → reward

LLM → reward → RL
High level knowledge → bridge → low level control

KNOWLEDGE TECHNOLOGY

LLM Guided Agent (LLM+RL)

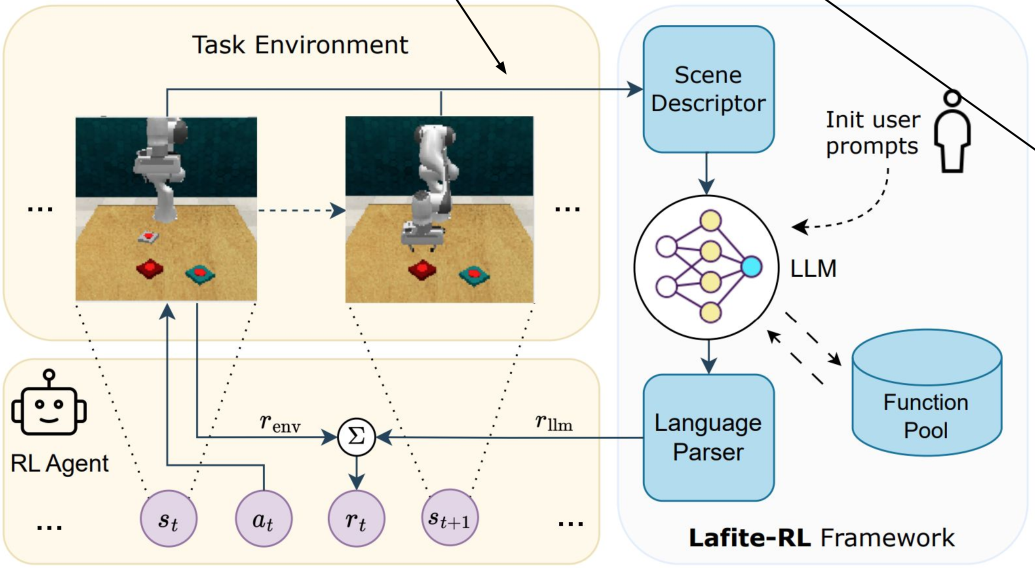
- LLM + RL: knowledge → reward
  - Rewards
  - …



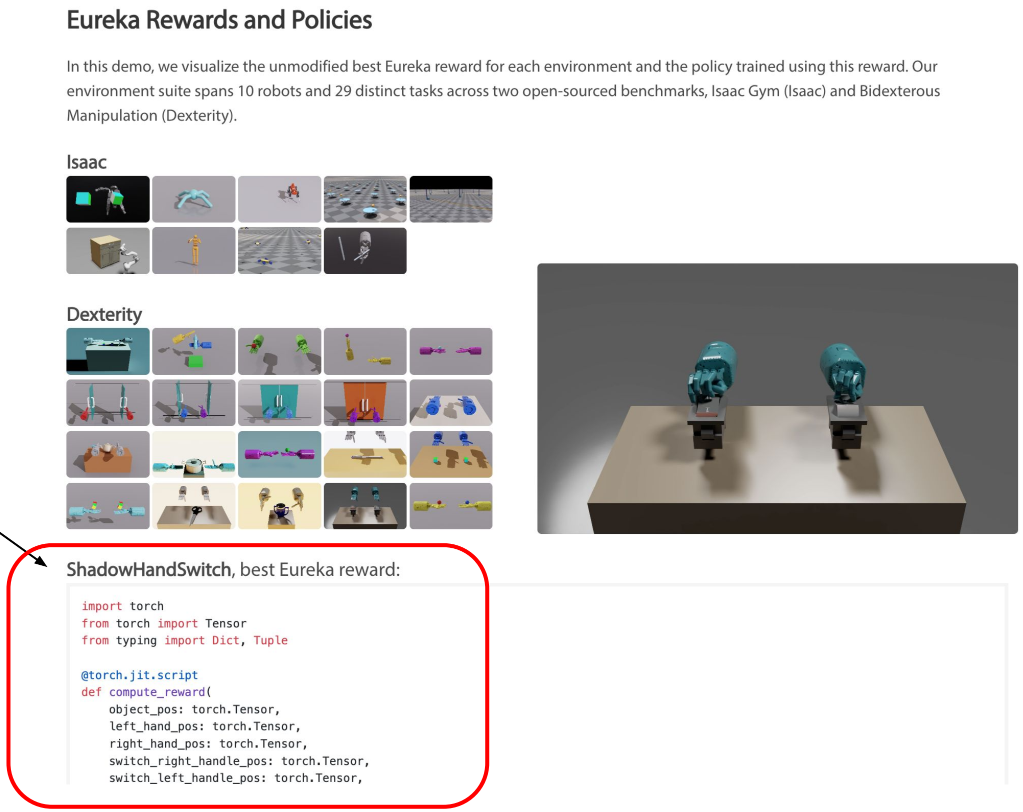
Kun Chu, Xufeng Zhao, Cornelius Weber, Mengdi Li, and Stefan Wermter
In CoRL 2023 Workshop (oral), Nov 2023, https://arxiv.org/abs/2311.02379

LLM Guided Agent (LLM+RL)

- LLM + RL: knowledge → reward
  - Rewards
  - Reward function
  - …



Kun Chu, Xufeng Zhao, Cornelius Weber, Mengdi Li, and Stefan Wermter
In CoRL 2023 Workshop (oral), Nov 2023, https://arxiv.org/abs/2311.02379



Ma, Yecheng Jason, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. "Eureka: Human-Level Reward Design via Coding Large Language Models." *arXiv preprint arXiv:2310.12931* (2023).

KNOWLEDGE TECHNOLOGY

Outline

- Motivation
- Knowledge learned with RL
  - Internally rewarded reinforcement learning
  - Multimodal association with unsupervised reinforcement learning
- LLM utilization
  - Emergent abilities and Fine-tuning
  - LLM Prompt Reasoning
- LLM Agent
  - Structures
  - Instances
  - Trends
- IROS 2023 Related

OCTOBER 1 - 5, 2023

IEEE/RSJ International Conference on Intelligent Robots and Systems

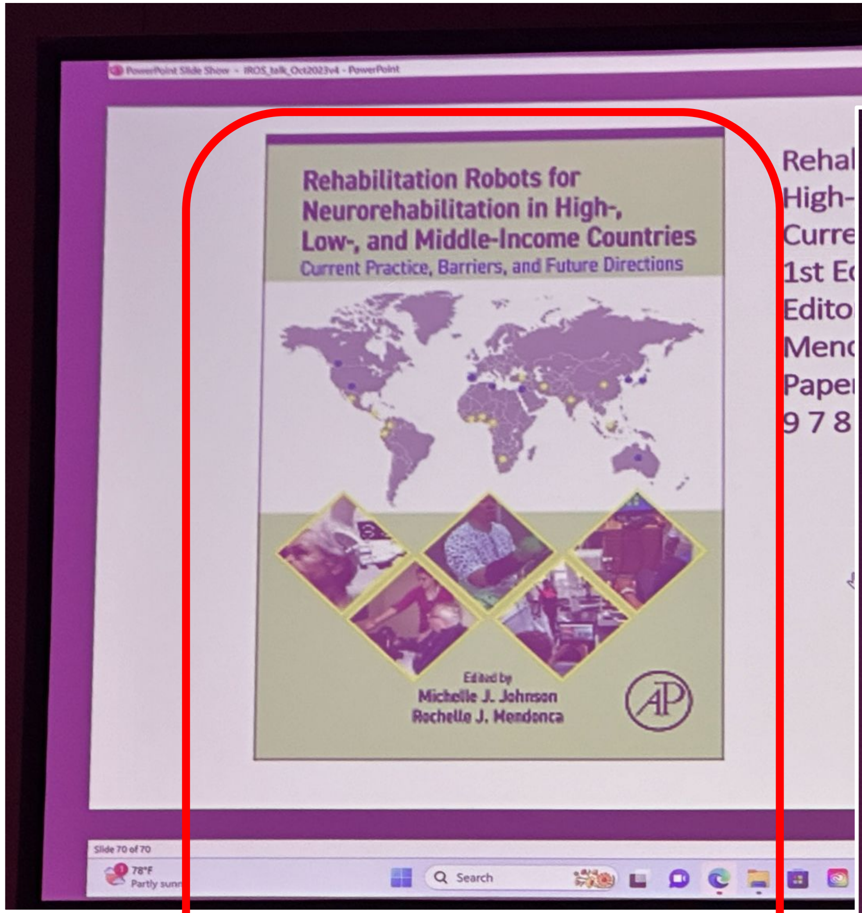
Detroit
Oct 1-5

IROS 2023

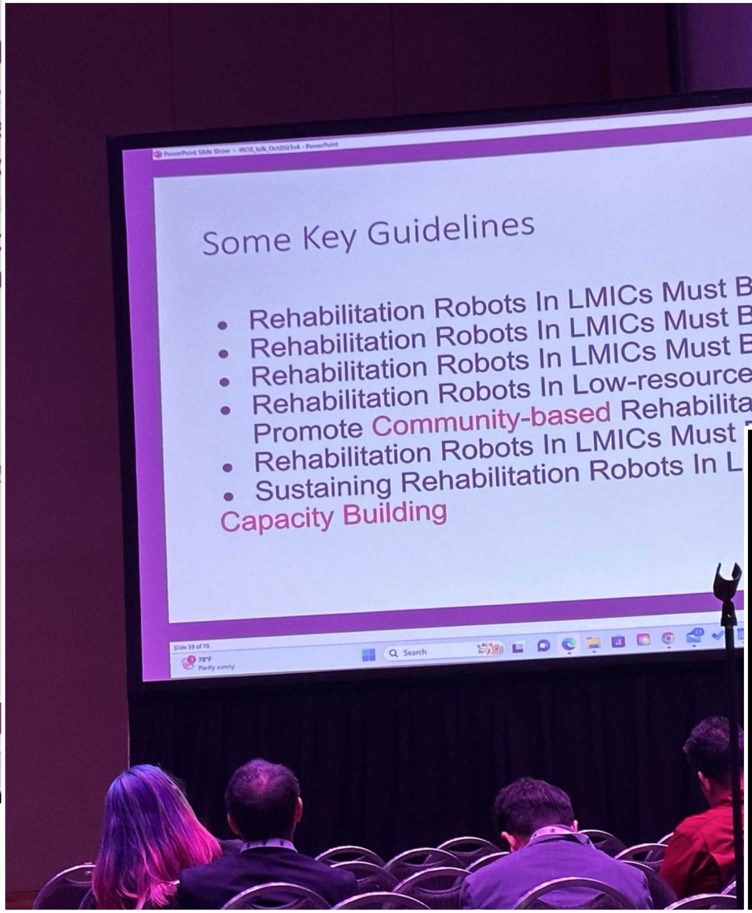
(May related)

Keynote talk 1: Towards more inclusive rehabilitation robots

Michelle Johnson, University of Pennsylvania, USA

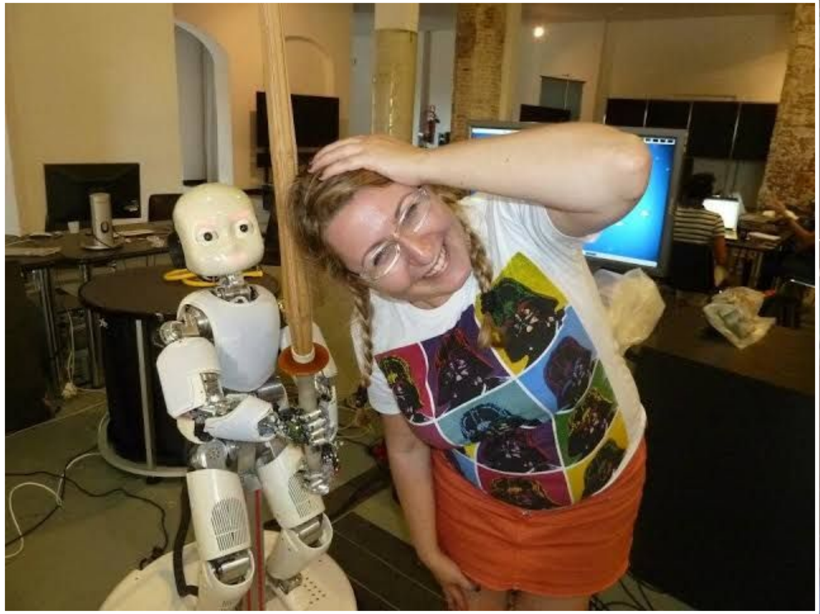


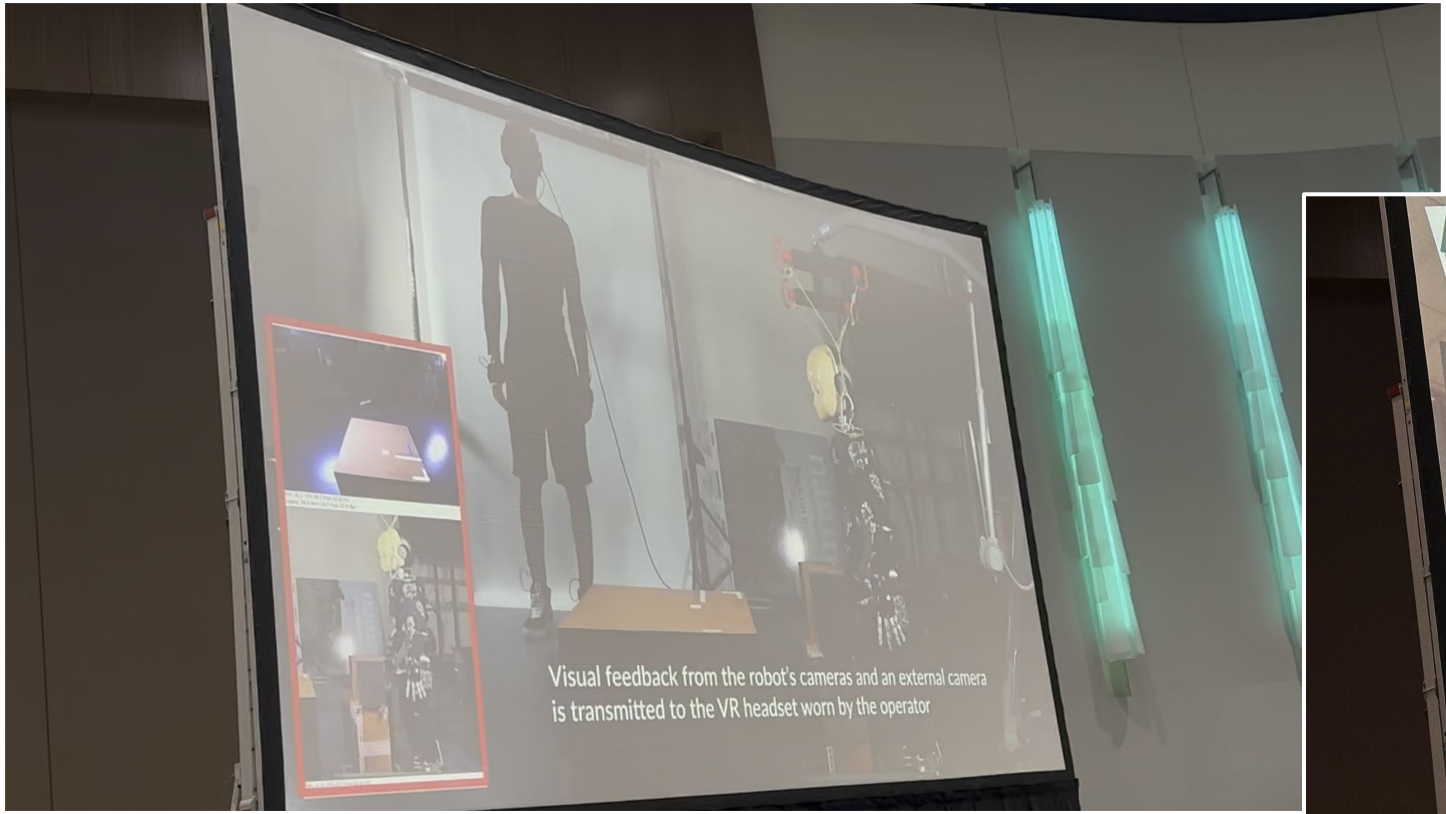
Book



Keynote 2: From humanoids to exoskeletons: assisting and collaborating with humans

Serena Ivaldi





Telemanipulation with human intention prediction

Collecting data for exoskeleton development

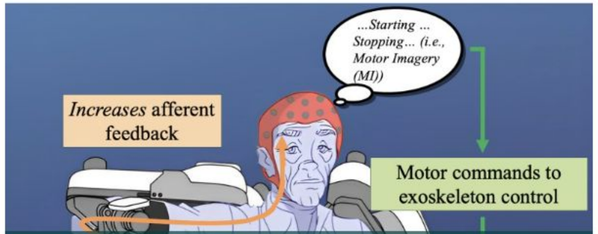
IROS2023

Official

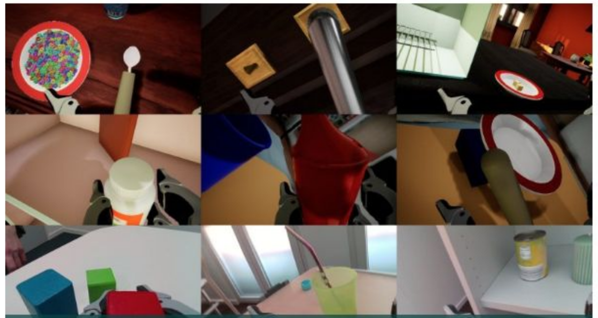
- Training in **Isaac Gym**

May related

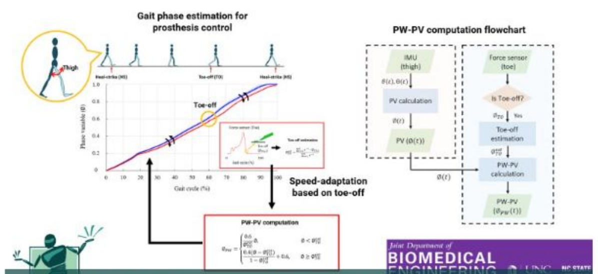
- Rehabilitation 1



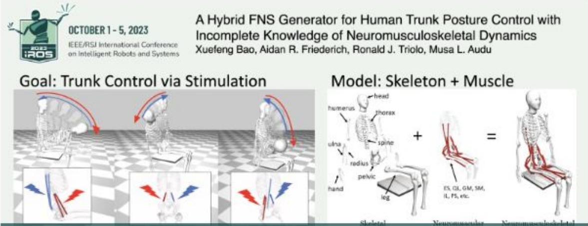
Characterizing the Onset and Offset of Motor Imagery During Passive Arm Movements Induced by an Upper-Body Exoskeleton
(No Video, Proceedings only!)



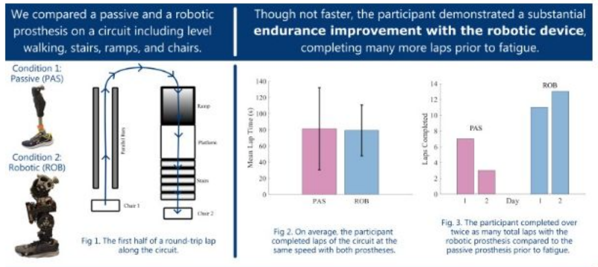
DORMADL - Dataset of Human-Operated Robot Arm Motion in Activities of Daily Living



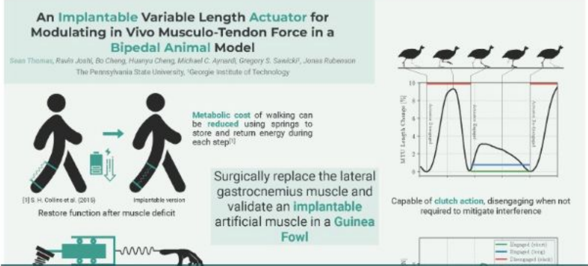
A Feasibility Study of Piecewise Phase Variable Based on Variable Toe-Off for the Powered Prosthesis Control: A Case Study



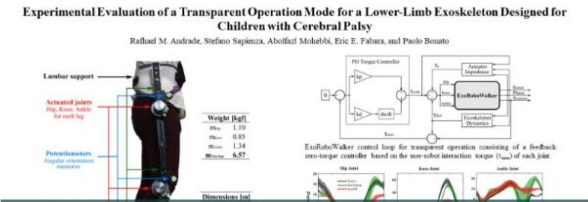
A Hybrid FNS Generator for Human Trunk Posture Control with Incomplete Knowledge of Neuromusculoskeletal Dynamics
(No Video, Proceedings only!)



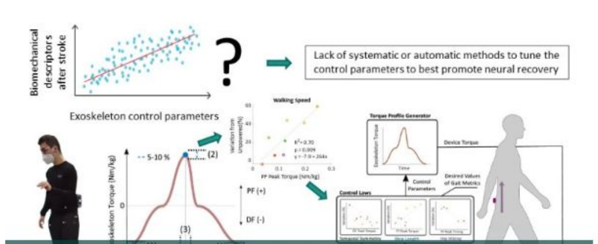
Improving Amputee Endurance Over Activities of Daily Living with a Robotic Knee-Ankle Prosthesis: A Case Study



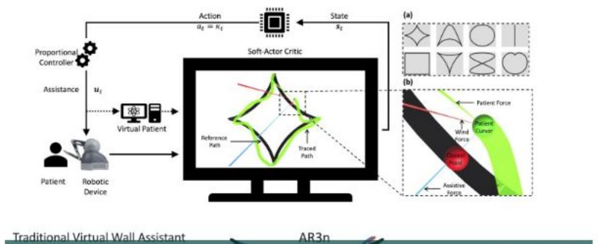
An Implantable Variable Length Actuator for Modulating in Vivo Musculo-Tendon Force in a Bipedal Animal Model



Experimental Evaluation of a Transparent Operation Mode for a Lower-Limb Exoskeleton Designed for Children with Cerebral Palsy
(No Video, Proceedings only!)



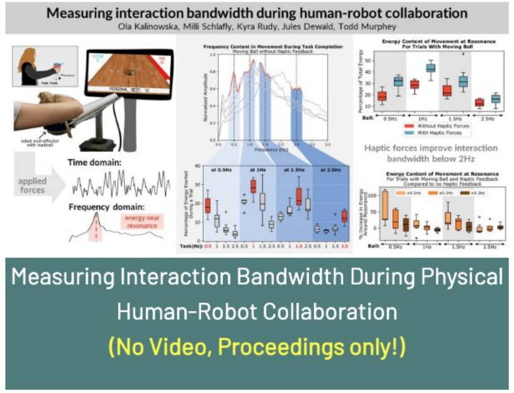
Relationship between Ankle Assistive Torque and Biomechanical Gait Metrics in Individuals after Stroke



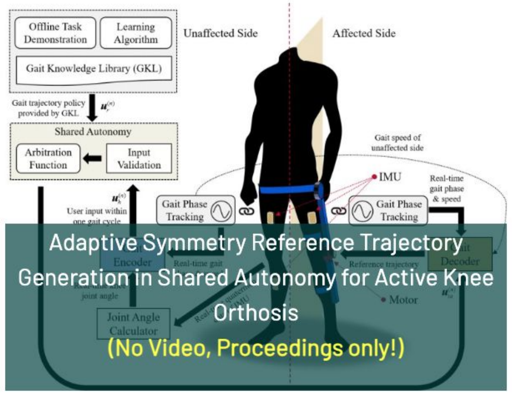
AR3n: A Reinforcement Learning-Based Assist-As-Needed Controller for Robotic Rehabilitation
(No Video, Proceedings only!)

May related

- Rehabilitation 2



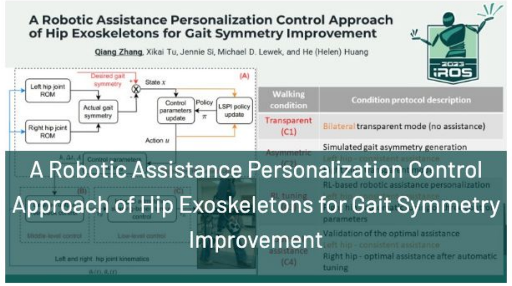
Measuring Interaction Bandwidth During Physical Human-Robot Collaboration
(No Video, Proceedings only!)



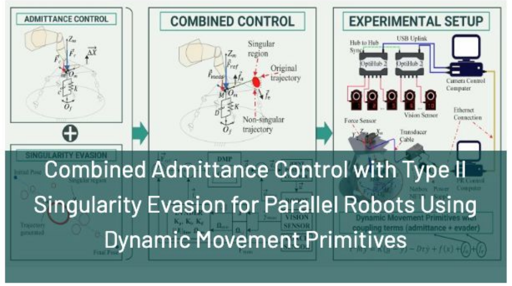
Adaptive Symmetry Reference Trajectory Generation in Shared Autonomy for Active Knee Orthosis
(No Video, Proceedings only!)

Cognitive Exercise for Persons with Alzheimer s Disease and Related Dementia Using a Social Robot
(No Video, Proceedings only!)

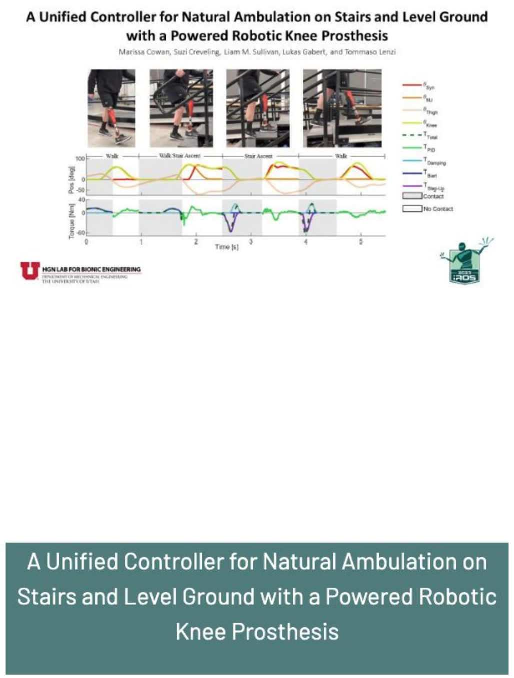
A Wearable Robotic Rehabilitation System for Neuro-Rehabilitation Aimed at Enhancing Mediolateral Balance



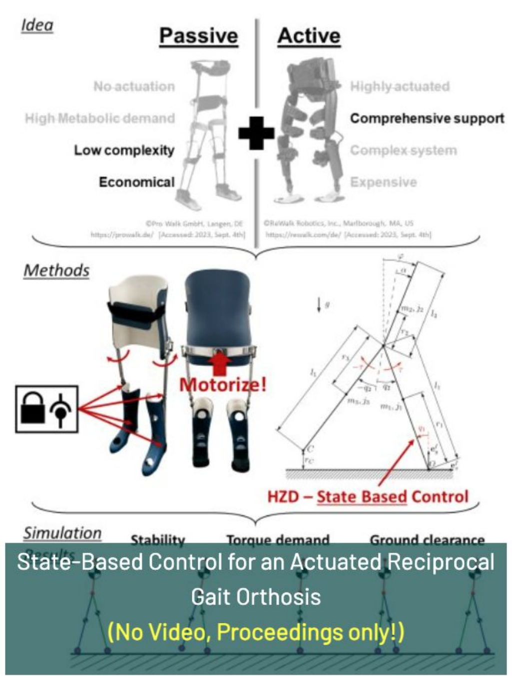
A Robotic Assistance Personalization Control Approach of Hip Exoskeletons for Gait Symmetry Improvement



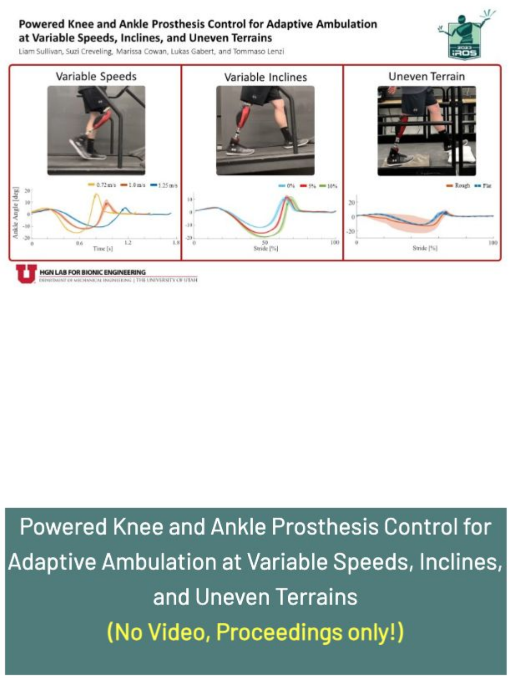
Combined Admittance Control with Type II Singularity Evasion for Parallel Robots Using Dynamic Movement Primitives



A Unified Controller for Natural Ambulation on Stairs and Level Ground with a Powered Robotic Knee Prosthesis

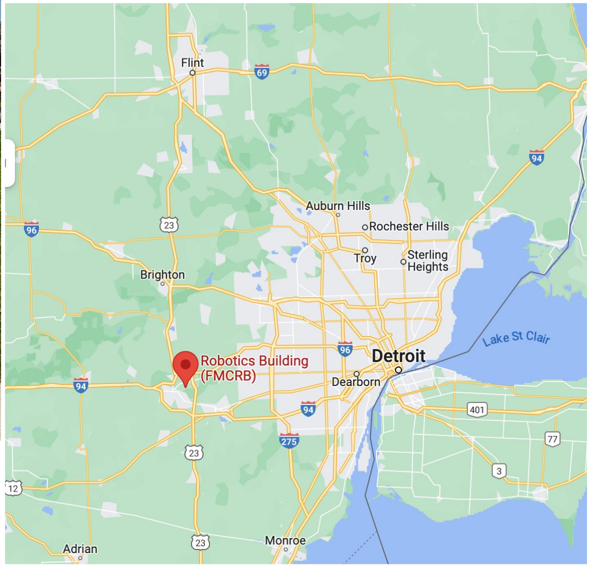


State-Based Control for an Actuated Reciprocal Gait Orthosis
(No Video, Proceedings only!)



Powered Knee and Ankle Prosthesis Control for Adaptive Ambulation at Variable Speeds, Inclines, and Uneven Terrains
(No Video, Proceedings only!)

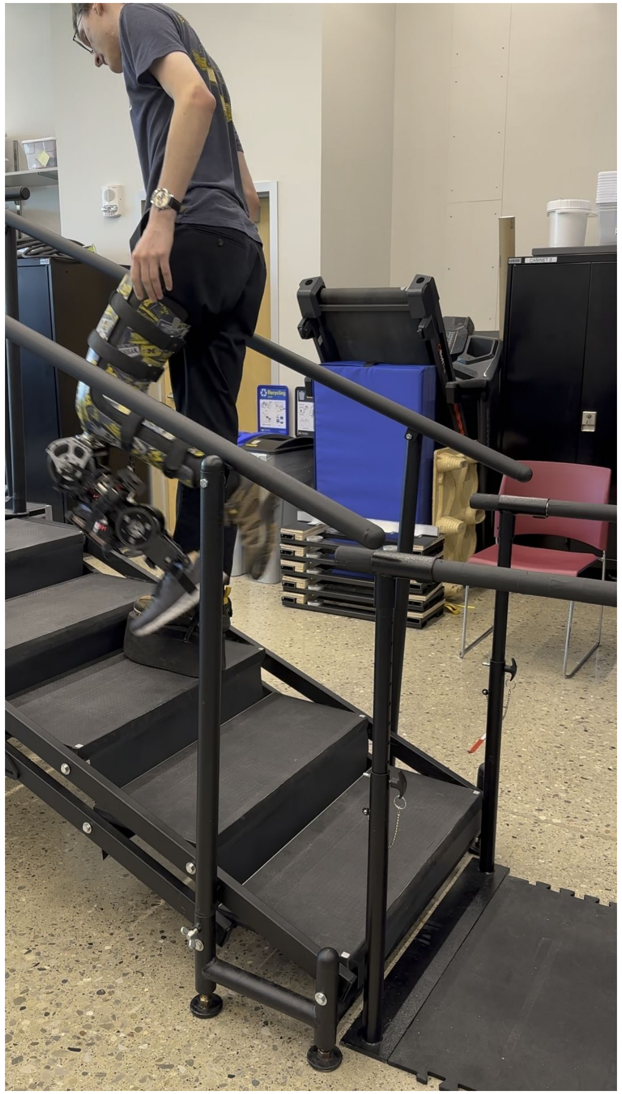
University of Michigan Ford Robotics Building Tour



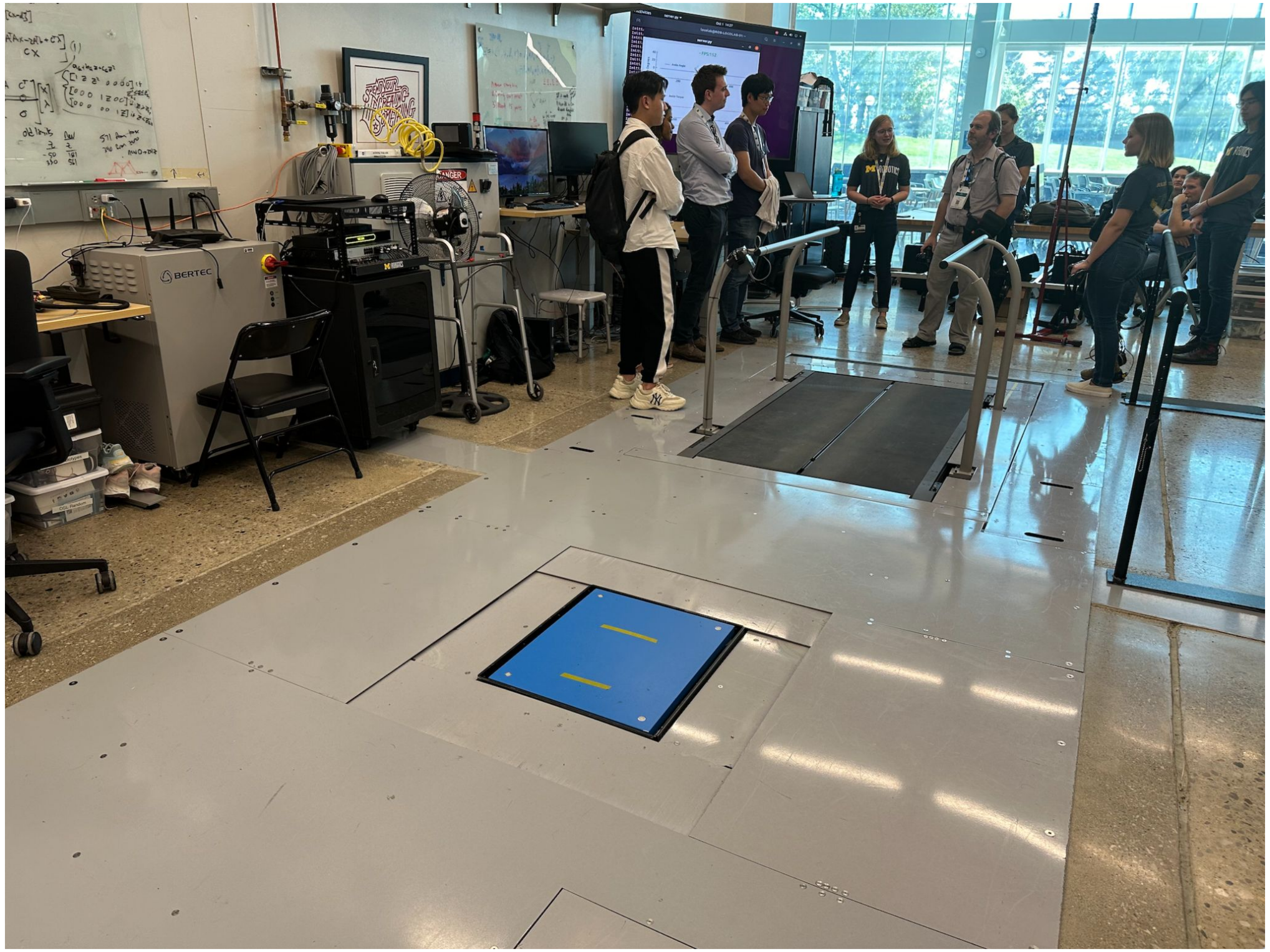
Exoskeleton

- Jump
- Walk

- Artificial limb



- Facility to test balance



The End

Thank you for your attention.

Any question?

Xufeng Zhao

https://xf-zhao.github.io/

KNOWLEDGE
TECHNOLOGY