

---

# Large models as an Active Learner

---

Jeongeun Park\*

Department of Artificial Intelligence  
Korea University  
baro0906@korea.ac.kr

## Abstract

In this project, we aim to explore the ability of large models to understand the sparse signal from humans. In particular, we focus on the problem of estimating the user intention on a sparse signal, like a preference signal or stopping signal. We utilize large language models (LLMs) and multimodal large language models (MLLMs) to learn the human intention from those signals. The codes are available at <https://github.com/jeongeun980906/LLM-Active-Learner>.

## 1 Introduction

In recent years, the rise of large models has revolutionized the field of artificial intelligence, bringing forth new capabilities and applications. One of the most intriguing aspects of these developments is the ability of models to interpret and learn from sparse signals, particularly those originating from human interactions. This project delves into the exploration of large language models (LLMs) and multimodal large language models (MLLMs) to discern and predict user intentions based on such sparse signals. Sparse signals, such as preference or stopping signals, are often overlooked yet hold substantial information critical for understanding and anticipating human needs and behaviors.

Understanding user intention from sparse signals is a challenging yet vital task. It involves decoding minimal information to predict complex human behaviors and preferences accurately. This not only enhances the interaction between humans and machines but also paves the way for more intuitive and responsive AI systems. By leveraging the advanced capabilities of LLMs and MLLMs, this project aims to bridge the gap between human cognitive processes and machine understanding. Through this research, we aspire to contribute to practical applications, i.e., robotics, enhancing machine responsiveness and adaptability. With a focus on the potent capabilities of large models, we aim to unlock new horizons in understanding and harnessing the power of sparse human signals.

## 2 Related Work

Building upon the potential of large language models (LLMs), there are several approaches for adapting the reasoning ability of LLMs in robotic planners. By incorporating LLMs [1, 2, 3], robots can understand and execute tasks described in natural language when confronted with high-level abstract goals to plan for low-level instructions. Utilizing chain-of-thoughts was reported to have success in sequence planning with various sources of feedback [1, 4]. There is another approach [2, 5] to achieve the reliability of LLMs by grounding the generated action to a feasible set. Hu et al. [6] introduced the ViLA framework, which leverages vision-language models (VLMs) to generate a sequence of actionable steps in complex scenarios.

Recently, there have been attempts to adapt human feedback for robotic applications leveraging large language models. Droc [7] utilizes LLMs to understand the language corrections via users for

---

\*Personal website: <https://sites.google.com/view/cv-jeongeunpark-korea>

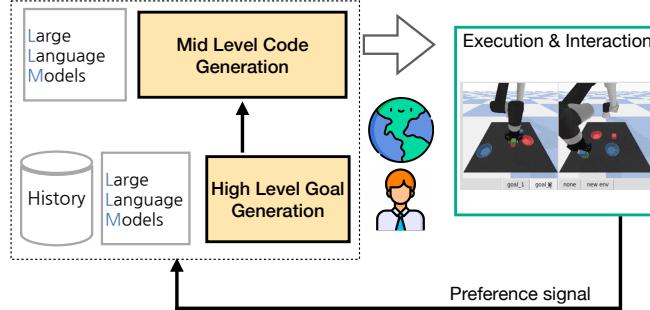


Figure 1: Architecture for Preference Signal

generating desired trajectories in robotic agents. OLAF [8] enables the users to stop the robot and conduct both verbal corrections and demonstrations via large language models. However, language corrections can induce a large cognitive load for the users. In this project, we aim to adapt the user's intention or preference based on sparse signals like binary selection or stopping signals.

### 3 Proposed Method

In this section, we will go through the architecture and the details of the proposed method. The objective of the system is to estimate the user's preference or intention denoted as  $z$ , and the corresponding language instruction  $y$ . The feedback from the users is denoted as  $h$ . The input prompt  $x$  contains the low-level skills and the instructions for the system. In this project, we denote the set of skills as "pick a block and place block in the bowl", "pick a block and place block on the block", and "pick a block and place block next to the object".

#### 3.1 Preference Signal

**Problem Formulation** The system showcases two different trajectories, where the users are asked to pick the preferable trajectory. The user can choose from three different options: [A], [B], and [Can not choose]. We add the option "[Can not choose]" for the case where the two trajectories have semantically identical meanings or where the user does not prefer both of the trajectories.

**Architecture** We start by estimating the preference or intention  $z$  first, followed by estimating two showcases of high-level instruction  $y$ . The language instruction is then transformed to robotic trajectory  $\tau = f(y)$  via instruction to action module, i.e., Code as Policies [3]. When the user's signal  $h$  is given, we append the signal into the prompts to condition the user's signal into the next estimates. The method overview is illustrated in Figure 1.

The estimated intention in step  $i + 1$  is denoted as  $z_{i+1}$  and defined as follows:

$$z_{i+1} \sim p_\theta(z_{i+1}|z_i, y_i, h_i, z_{i-1}, y_{i-1}, h_{i-1}, \dots, x) \quad (1)$$

where  $p_\theta(\cdot)$  is large language model,  $y_i$  is two showcase language instruction in step  $i$  and the  $h_i$  is user's feedback in  $i$  step. We then predict two showcase instruction  $y_{i+1}$ , which is defined as follows:

$$y_{i+1} = \{y_{i+1}^a, y_{i+1}^b\} \sim p_\theta(y_{i+1}|z_{i+1}, z_i, y_i, h_i, z_{i-1}, y_{i-1}, h_{i-1}, \dots, x) \quad (2)$$

Although utilizing the preference signal is intuitive, we observe that those signals are often noisy when the search space gets larger. This is mainly because the difficulty for human evaluators increases as the two trajectories become similar or when none of them is desired.

#### 3.2 Clicker Stop Signal

To mitigate the issues mentioned in the previous paragraph, we adopt a new type of sparse signal, i.e., stop signal.

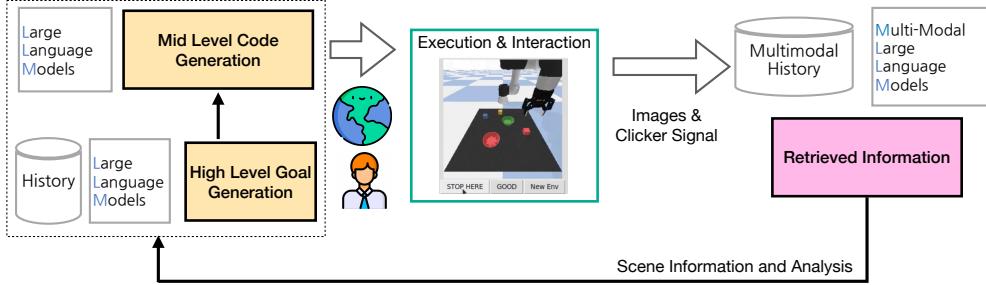


Figure 2: Architecture for the Stopping Signal

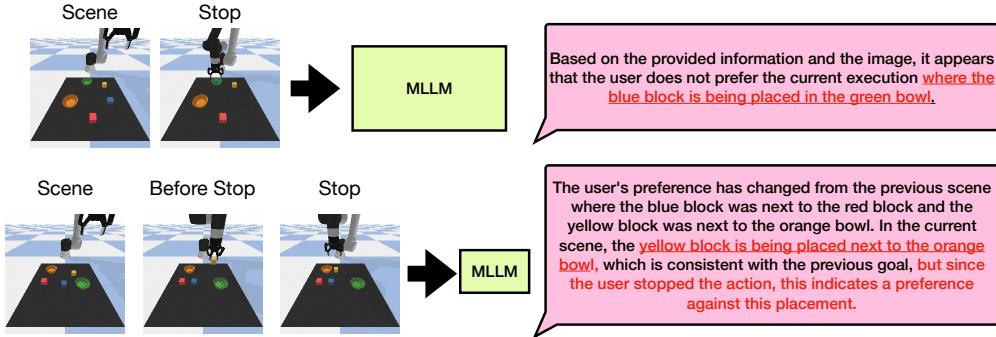


Figure 3: Retrieved Examples

**Problem Formulation** In this system, the users are asked to stop the robot when the robot starts to perform the undesired actions. The signal from the user  $h$  will be the time value when the user stops.

**Architecture** Again, in an initial step, we estimate preference or intention, denoted as  $z$ . Following this, high-level instruction is generated, referred to as  $y$ . These instructions are then transformed into a robotic trajectory, noted as  $\tau = f(y)$ , through an instruction-to-action module [3]. When a user input, denoted as  $h$ , is provided, the multimodal language models, i.e., GPT4V, retrieve information about the scene images to understand the undesired part of the users. The retrieved information  $r$  is incorporated into the prompts to condition the subsequent estimates for the large language model. This method is visually represented in Figure 2.

The user feedback  $h$  is defined as a time when the user stops. The prompt given  $h$ ;  $\text{prompt}(h)$  is defined as follows:

$$\text{prompt}(h) = \begin{cases} x_n & \text{if } h = \text{None} \\ \{I_0, x_s, I_s\} & \text{if } h < \epsilon \\ \{I_0, x_b, I_b, x_s, I_s\} & \text{otherwise} \end{cases} \quad (3)$$

where  $x_n$  denotes the textual prompts when the user does not stop the robot,  $x_s$  and  $x_b$  denote textual prompts when the user stops the robot. We aggregate the information from the scene, where  $I_0$  is an initial scene,  $I$  is a scene where the user stopped, and  $I_b$  is an image  $\epsilon$  seconds before the user stopped. Based on this prompt, we leverage the vision-language model, i.e., GPT4V, to retrieve the information about the user’s intention on the stopping signal. The retrieved information  $r_i$  is defined as follows:

$$r_i \sim p_\psi(r_i | \text{prompt}(h_i), y_i, r_{i-1}, \dots, r_1, \text{prompt}(h_1), y_1) \quad (4)$$

where  $p_\psi$  is a vision-language model. The examples of retrieved information are illustrated in Figure 3. Based on the retrieved information  $r_i$ , we then estimate the user’s intention  $z_{i+1}$  and the next

instruction  $y_{i+1}$ .

$$z_{i+1} \sim p_\theta(z_{i+1}|r_i, h_i, y_i, z_{i-1}, \dots, z_1, r_1, h_1, y_1) \quad (5)$$

$$y_{i+1} \sim p_\theta(y_{i+1}|z_{i+1}, r_i, h_i, y_i, z_{i-1}, \dots, z_1, r_1, h_1, y_1) \quad (6)$$

where  $p_\theta(\cdot)$  is a large language model.

## 4 Experiment

In this section, we will discuss the experimental results of a pick-and-place tabletop simulation environment.

### 4.1 Settings

We evaluate the proposed method in a tabletop simulation environment where there are multiple blocks and bowls. We formulate the four different ground-truth intentions: align the blocks horizontally, stack all the blocks, put the blocks in a color-matching bowl, and put the blocks in the mismatching bowl. For each intention, we evaluated five different trials where the system can only obtain the user's feedback five times. For the large language model, we utilized gpt-4-1106-preview, and for the vision-language models, we leverage gpt-4-vision-preview<sup>2</sup>. We compare the method of solely utilizing a multi-modal large language model (MLLM), preference signal with LLMs, and utilizing MLLM for image information extraction and LLM for reasoning. The setting is illustrated in Figure 4.

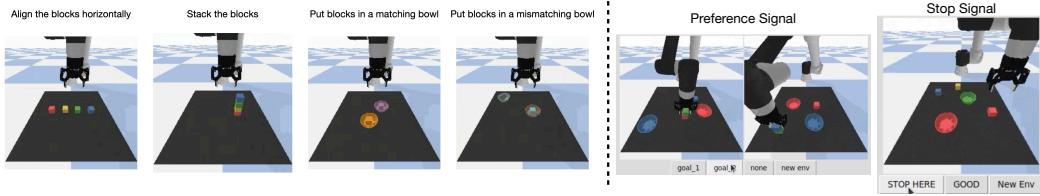


Figure 4: Experimental Setting

### 4.2 Simulation Tabletop Experiment

Tasks	Align	Stack	Color Match	Color Mismatch	Average
Preference Signal	0.6	0.0	0.6	0.4	0.4
MLLM	0.4	0.2	0.8	0.4	0.45
MLLM + LLM	0.6	0.4	0.8	0.6	<b>0.6</b>

Table 1: Success rates

Across the experiments, we observe that utilizing MLLM and LLM outperforms the compared methods with a gap of 0.15 in success rate. We believe that preference signals are often noisy and misleading, which leads to the lowest success rate. In addition, we believe that adding the additional step of retrieving the information from the images enhances the reasoning ability of large models to explore the user's intention and understand the user's signal. The qualitative result is illustrated in Figure 5.

## 5 Conclusion

This project has explored the capabilities of large language models (LLMs) and multimodal large language models (MLLMs) in understanding and interpreting sparse signals from humans to estimate user intention. Our findings reveal that LLMs and MLLMs can effectively understand and predict

<sup>2</sup><https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>

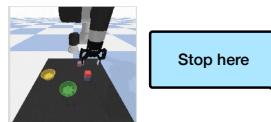
The robot can pick the blocks and place it in a bowl or in the table. We need to estimate the users preference. Available skills: [....]  
I see [objects]. Generate two different goals considering user's preference

Predicted Preference: The user prefers a blocks to be sorted by color into corresponding colored bowls  
Goal: place the blue block in the green bowl and the purple block in the yellow bowl



Stop here

Predicted Preference: The user prefers a unique arrangement that does not involve placing blocks in bowls with matching colors  
Goal: place the red block on the top of blue block, and place purple block next to the blue block



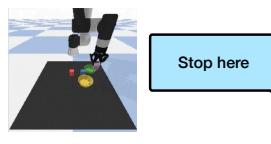
Stop here

Predicted Preference: The user prefers blocks to be placed next to each other on the table without stacking or using the bowls  
Goal: place the blue block next to the red block, and the purple block next to the yellow bowl



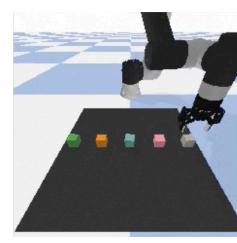
Stop here

Predicted Preference: The user prefers blocks to be placed next to each other on the table without stacking or using the bowls  
Goal: place the blue block next to the red block, and the purple block next to the blue block



Stop here

Let's spawn  
[gray block, pink block, cyan block, green block, orange block]



Predicted Preference: The user prefers blocks to be placed next to each other on the table in a linear arrangement without stacking or using bowls  
Goal: place the blue orange block next to green block, the cyan block next to orange block, the pink block next to the cyan block, and gray block next to the pink block, forming a line of blocks on the table

Figure 5: Experimental Results

human intentions, enhancing human interaction and AI systems' interaction. Although we have experimented with the proposed method in a tabletop environment, further research is essential, especially in more complex scenarios. We believe that expanding the method and the setting to the environment with complex intentions in long-horizon task planning will be an essential direction.

## References

- [1] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. Inner monologue: Embodied reasoning through planning with language models. In *Proc. of the 6th Annual Conference on Robot Learning (CoRL)*, 2022.
- [2] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, et al. Do as i can, not as i say: Grounding language in robotic affordances. In *Proc. of the 6th Annual Conference on Robot Learning (CoRL)*, 2022.
- [3] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. Code as policies: Language model programs for embodied control. In *Proc. of the 2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9493–9500, 2023.
- [4] Shunyu Yao, Jeffrey Zhao, Dian Yu, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *NeurIPS 2022 Foundation Models for Decision Making Workshop*.
- [5] Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In *Proc. of the International Conference on Machine Learning (ICRL)*, pages 9118–9147. PMLR, 2022.
- [6] Yingdong Hu, Fanqi Lin, Tong Zhang, Li Yi, and Yang Gao. Look before you leap: Unveiling the power of gpt-4v in robotic vision-language planning. *arXiv preprint arXiv:2311.17842*, 2023.
- [7] Lihan Zha, Yuchen Cui, Li-Heng Lin, Minae Kwon, Montserrat Gonzalez Arenas, Andy Zeng, Fei Xia, and Dorsa Sadigh. Distilling and retrieving generalizable knowledge for robot manipulation via language corrections. *arXiv preprint arXiv:2311.10678*, 2023.
- [8] Huihan Liu, Alice Chen, Yuke Zhu, Adith Swaminathan, Andrey Kolobov, and Ching-An Cheng. Interactive robot learning from verbal correction. 2023.