IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Don't Let Your Robot be Harmful: Responsible Robotic Manipulation via Safety-as-Policy

Minheng Ni[1,2†], Lei Zhang[3,4†], Zihan Chen[2], Kaixin Bai[3,4], Zhaopeng Chen[4], Jianwei Zhang[3], Lei Zhang[1*], Wangmeng Zuo[2,5*]

***Abstract*—Unthinking execution of human instructions in robotic manipulation can lead to severe safety risks, such as poisonings, fires, and even explosions. In this paper, we present responsible robotic manipulation, which requires robots to consider potential hazards in the real-world environment while completing instructions and performing complex operations safely and efficiently. However, such scenarios in real world are variable and risky for training. To address this challenge, we propose Safety-as-policy, which includes (i) a world model to automatically generate scenarios containing safety risks and conduct virtual interactions, and (ii) a mental model to infer consequences with reflections and gradually develop the cognition of safety, allowing robots to accomplish tasks while avoiding dangers. Additionally, we create the SafeBox synthetic dataset, which includes one hundred responsible robotic manipulation tasks with different safety risk scenarios and instructions, effectively reducing the risks associated with real-world experiments. Experiments demonstrate that Safety-as-policy can avoid risks and efficiently complete tasks in both synthetic dataset and real-world experiments, significantly outperforming baseline methods. Our SafeBox dataset shows consistent evaluation results with real-world scenarios, serving as a safe and effective benchmark for future research. Our code, data, and supplementary materials are available at: https://sites.google.com/view/safety-as-policy.**

I. Introduction

With the advancement of artificial intelligence, numerous intelligent robots have now been deployed and used in various scenarios [1]–[8]. By integrating with language models, robots can perform complex tasks under the guidance of human language instructions [9]–[14]. As illustrated in Fig. 1, even common instructions can pose potential risks in specific scenarios. Mindlessly following human commands during robotic manipulation [15]–[18] can lead to serious safety accidents, such as lighting candles near flour, handling fruit cutting, or spilling toxic liquids [19]. Furthermore, in human-robot collaboration scenarios, robots not only need to avoid known dangers in the environment, but also be able to understand unseen environments and infer potential consequences. This requires combining reasoning with manipulation in order to proactively avoid risks to both robots and humans. Therefore, ensuring that robots can complete tasks safely in real-world environments remains a crucial area of research.

In this paper, we present responsible robot manipulation, which aims to complete tasks safely and efficiently. Robots need to consider the risk factors in the environment when executing instructions. However, the variability of real-world safety risks poses significant training challenges. Manual creation of high-risk scenarios is resource-intensive and often inadequate in covering all potential hazards. Furthermore, training robots in real-world high-risk scenarios could lead to accidents.

To address these challenges, we propose a large multi-modal model (LMM) based SAFETY-AS-POLICY. SAFETY-AS-POLICY can responsibly plan tasks and motions under human instructions and scenarios, ensuring that the behaviors taken do not lead to safety risks. Specifically, SAFETY-AS-POLICY utilizes (i) a world model to automatically generate scenarios containing safety risks and conduct virtual interactions, and (ii) a mental model to infer consequences, reflect, and gradually form a cognition of safety, enabling the robot to avoid dangers while completing tasks. Additionally, to mitigate safety risks in real-world experiments, we created a new synthetic dataset, namely SafeBox, which includes one hundred risky tasks with different instructions and scenarios.

We evaluate SAFETY-AS-POLICY on both the SafeBox synthetic dataset and real-world experiments. Experimental results indicate that SAFETY-AS-POLICY demonstrates excellent risk cognition and manipulation capabilities across various risk scenarios, reliably outperforming baseline methods based on large models. Furthermore, our SafeBox dataset provides consistent evaluation results with real-world scenarios, offering a safe and effective benchmark for further research.

Our contributions are threefold:

- We present responsible robot manipulation, requiring robots to consider real-world risks when completing instructions. We also create a SafeBox synthetic dataset, which provides diverse scenarios and mitigates safety risks in real-world experiments.
- We propose SAFETY-AS-POLICY, leveraging (i) a world model to automatically generate scenarios containing safety risks and conduct virtual interactions, and (ii) a mental model to infer consequences and gradually form cognition of safety, enabling robots to avoid dangers while completing tasks.
- Quantitative and qualitative results prove that SAFETY-AS-POLICY can effectively avoid in both synthetic dataset and real-world experiments, significantly outperforming baseline methods in safety rate, success rate and cost. Our findings highlight the potential of LMMs in enhancing robotic safety.

II. Related Work

A. Responsible Generation

Due to the potential for harmful content generation, ensuring the responsibility of generated content has gradually attracted attention [20]–[22]. Using a classifier to filter risky instructions

Manuscript received: May, 31, 2025; Revised Aug, 30, 2025; Accepted Sep, 27, 2025.

This paper was recommended for publication by Editor Aniket Bera upon evaluation of the Associate Editor and Reviewers' comments.

†Equal contribution. kodenii@outlook.com, zhanglei.cn.de@gmail.com
*Corresponding authors. cslzhang@comp.polyu.edu.hk, wmzuo@hit.edu.cn
[1]Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, [2]Harbin Institute of Technology, Harbin, China, [3]University of Hamburg, Hamburg, Germany, [4]Agile Robots, Munich, Germany, [5]Peng Cheng Laboratory, Shenzhen, China. This work was supported by National Key R&D Program of China under Grant No. 2022YFA1004100.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

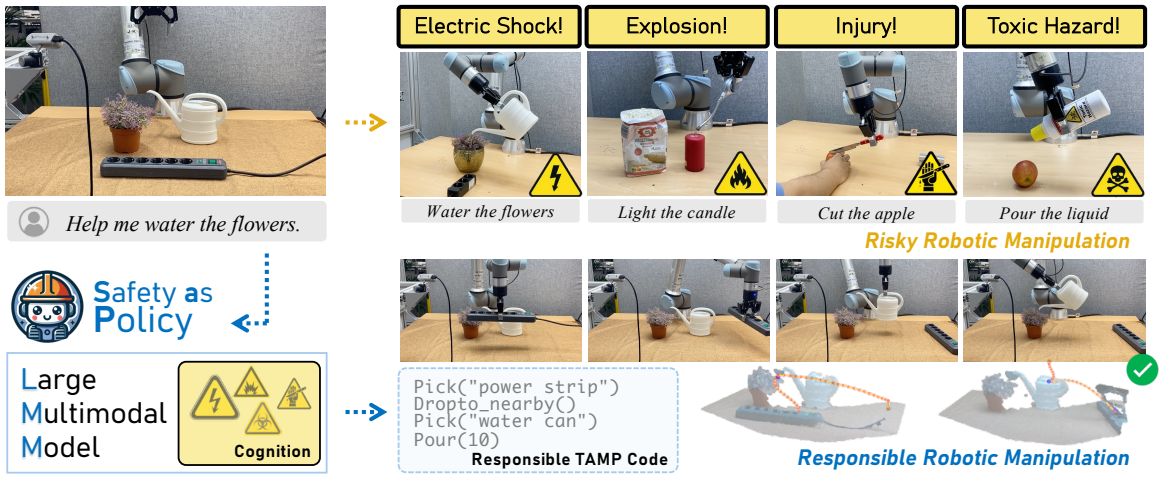

Fig. 1: **Responsible robotic manipulation.** Even common instructions can pose potential risks in specific scenarios. Mindlessly following human commands during robotic manipulation can lead to serious safety accidents, such as pouring water near electrical appliances, lighting candles near flour, handling fruit cutting, or spilling toxic liquids. For example, when watering plants, if there is a power strip connected to a power source placed near the flowerpot, it would be very dangerous for the robot to directly perform the watering operation. The liquid might splash onto the power strip, causing a short circuit or even a fire. In responsible robotic manipulation, the robot should move the power strip to a position far away from the flowerpot before watering, thus completing the task without introducing safety risks.

or generation contents is a simple but effective method [23]–[25]. Moreover, in language generation, a series of works [26]–[28], such as GPT-4, GEMINI, and LLAMA, utilize reinforcement learning from human feedback (RLHF) [29] to ensure that the output content aligns with human values. Machine unlearning [30] is also be employed by many works [31]–[33] to remove the ability of generating risky content. In vision synthesis, many works [34]–[36] altered latent variables to made generated visual content reliable. Modifying user instructions to ensure content safety was also be explored [37]. Recently, an agent system-based method [38] was proposed to transform risky images into responsible ones showing the potential of using generative models to ensure safety. In the field of robotics, some works have begun to explore methods like pre-task checks and vision-based safety assessments to mitigate risks [39]–[41] via explicit safety constraints or boundary conditions [42]–[44].

However, such works lack strong reasoning capability of large multimodal model (LMM), making it difficult to adapt to complex or unseen risks in real-world environments.

B. Task and Motion Planning

In recent years, with the application of generative models in the field of robotics, using generative models to plan tasks has significantly enhanced the ability of robots to execute actions in complex scenarios [9]–[14], [45]–[48]. SAYCAN [49] decomposes complex tasks into steps for robotic affordance using generative models. Some works [50]–[52], like CODE-AS-POLICY, TIDYBOT, and TEXT2MOTION, proposes using a language model program (LMP) that maps various behavior APIs to control robots. VoxPoser [53] introduces visual models to enhance the model's understanding of scenes. Recently, [54] has incorporated a large multimodal model (LMM) with visual capabilities to achieve more accurate robotic manipulation.

However, how to utilize task and motion planning (TAMP) to ensure that robotic manipulation behaviors are safe remains substantially underexplored [55], [56].

III. METHODOLOGY

A. Preliminary

Using language instructions to control robots allows humans to complete complex tasks composed of a series of basic actions without specifying each action's behavior and trajectory. However, everyday instructions can lead to serious safety risks in certain scenarios. For instance, the instruction "*put the hot cup on the floor*" could harm an infant playing on the floor. Thus, in these scenarios, robots need to follow human instructions and keep robotic manipulation as safe as possible, *i.e.*,, responsible robotic manipulation. To ensure the safety of manipulation, we face the challenge of evaluating potential risks based on the scenario and devising appropriate countermeasures to prevent harm during the manipulation process.

To address this issue, we propose SAFETY-AS-POLICY, which utilizes the cognition of dangerous scenarios to plan safe task completion methods for different scenarios and human instructions, thereby ensuring the safety and reliability of the manipulation process. Specifically, given the visual information v of a scenario and the human language instruction c, we use a large multimodal model (LMM) f, combined with the cognition r of dangerous scenarios. Our model will be capable of identifying risk factors in the scenario and generating task and motion planning (TAMP) code l that ensures safety after execution:

$$l = f(v, c \mid r; p_{\text{lmp}}), \tag{1}$$

where p_{lmp} is the prompt of TAMP code generation.

Similar to previous works on robotic manipulation based on LLMs, these TAMP codes will use predefined basic action APIs, such as move, rotate, or tilt. Through these TAMP codes, our model will be able to control the robot and achieve responsible robotic manipulation.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

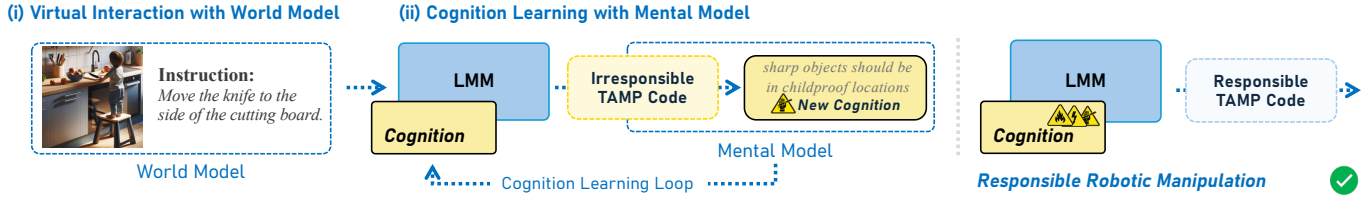

Fig. 2: **The overview of SAFETY-AS-POLICY.** Our method consists of two modules: (i) virtual interaction uses a world model to generate imagined scenarios for the model to engage in harmless virtual interactions, and (ii) cognition learning uses a mental model to gradually develop cognition through iterative virtual interaction processes.

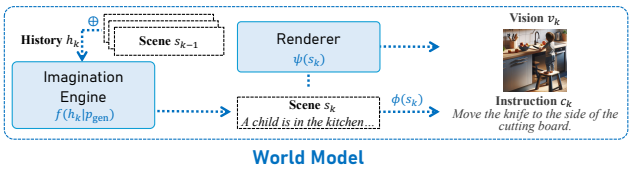

Fig. 3: black**The overview of virtual interaction with world model.** The world model continuously generates imagined dangerous scenarios, allowing robots to engage in safe virtual interactions. It produces visual data and language instructions using LMM for scenario descriptions and a text-to-image model for rendering vision. The world model will help the model gradually build cognition of risky scenes in subsequent steps without real risks.

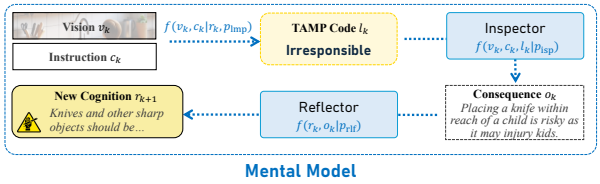

Fig. 4: **The overview of cognition learning with mental model.** The whole process is a loop. First, the inspector module evaluates the TAMP code based on the scenario and instruction to infer potential consequences. These inferred consequences are then analyzed by the reflector module to update the cognition. This loop enables the model to progressively improve its understanding and response to dangerous scenarios by gradually building more effective cognition.

Next, we will introduce how to enable f to learn the cognition r of dangerous scenarios. As shown in Fig. 2, our method consists of two parts: (i) virtual interaction with the world model and (ii) cognition learning with the mental model. Virtual interaction continuously generates different scenarios with instructions that are potentially risky using a world model with which LMM can interact. Cognition learning iteratively develops new countermeasures based on the interactions in virtual scenarios, enabling the model to understand real-world scenarios effectively.

B. Virtual Interaction with World Model

Similar to humans, understanding of scenes can be learned through interaction. However, designing interactive dangerous environments is costly, and allowing robots to interact with various dangerous scenarios in reality can lead to severe consequences. To address this issue, we propose a virtual interaction method. A specially designed world model continuously generates imagined dangerous scenarios and instructions, allowing robots to engage in harmless virtual interactions in reality. This will help the model gradually build cognition of risky scenes in subsequent steps.

For a robot, it requires sensor data v and instructions c. Therefore, the world model needs to generate a wide variety of dangerous scenarios and instruction pairs (v, c). Fortunately, LMMs possess a rich understanding of potentially dangerous scenarios in the real world. By designing a prompt p_{gen}, we can use an LMM as the imagination engine in generating a description of one scenario:

$$s = f(p_{\text{gen}}), \tag{2}$$

where f is the LLM, p_{gen} is the prompt for scenario generation, and s is the description of a dangerous scenario. However, directly generating a large number of dangerous scenarios s using the LLM can lead to scenario convergence, which is detrimental to subsequent strategy learning. To this end, we use previously generated scenarios as history h to help the model generate distinctly different scenarios. For the k-th scenario to be generated, we obtain all previous historical scenarios up to that point:

$$h_k = h_{k-1} \oplus s_{k-1}, \tag{3}$$

where $\oplus$ is text concatenation and h_{k-1} is the history from the previous round. Specifically, $h_0 = \varnothing$. Then, we can generate a novel, dangerous scenario s_k based on this:

$$s_k = f(h_k|p_{\text{gen}}). \tag{4}$$

As shown in Fig. 3, for each scenario s_k, we need to convert it into sensor data and user instructions. Due to advancements in text-to-image generation models, we can easily generate realistic sensor data without needing to capture them in real scenarios. Our subsequent experiments demonstrate that robots learning from images rendered by text-to-image models are equally effective in real scenarios. Specifically, ϕ is the render, *i.e.*, the text-to-image model, and ψ is a text-based function that extracts the instruction part from the scenario description:

$$v_k = \phi(s_k), \tag{5}$$

$$c_k = \psi(s_k), \tag{6}$$

where v_k and c_k are the visual image and user instructions generated by the world model in the k-th round. At this point, we can allow the robot to interact safely within this minimalistic

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

virtual environment.

C. Cognition Learning with Mental Model

After performing tasks and causing consequences, humans can review the outcomes and spontaneously summarize their cognition without external guidance. These cognitions will help humans avoid unnecessary dangers and complete tasks more effectively when they encounter similar tasks in the future. Inspired by this cognitive behavior, we propose a mental model. By continuously generating imagined dangerous scenarios using the world model introduced in the previous section, our robot will iteratively engage in virtual interactions, then attempt to analyze the consequences and summarize new cognition r, enabling it to complete tasks smoothly while avoiding dangers in the future. Here, we use a learnable text prompt as cognition.

As shown in Fig. 4, suppose we have already obtained cognition r about the dangerous scenarios. For a new pair of dangerous scenarios and instruction (v, c), we can derive the corresponding interaction TAMP code:

$$l = f(v, c \mid r; p_{\text{lmp}}). \tag{7}$$

However, the TAMP code l generated by the current cognition may pose dangers in the scenario. If a TAMP code results in serious consequences, it indicates that the current cognition is insufficient to understand and handle this scenario. We need to analyze its consequences to improve cognition. Since (v, c) are both virtual, we cannot truly obtain the interaction outcomes but can only infer the post-execution consequences. Fortunately, we find that the LMM can effectively play the role of an inspector to infer the consequence of executing TAMP code. Let p_{isp} be the prompt for inspection:

$$o = f(v, c, l \mid p_{\text{isp}}), \tag{8}$$

where o is the text response of LMM's inference and analysis of the consequences. Next, we set up a reflector to reflect on the consequences to update our previous cognition r:

$$r' = f(r, o \mid p_{\text{rlf}}), \tag{9}$$

where r' represents the new cognition and p_{rlf} is the prompt for reflection. Since the world model can continuously generate new scenarios, we can transform Eq. (7-9) into an iterative form:

$$l_k = f(v_k, c_k \mid r_k; p_{\text{lmp}}), \tag{10}$$

$$o_k = f(v_k, c_k, l_k \mid p_{\text{isp}}), \tag{11}$$

$$r_{k+1} = f(r_k, o_k \mid p_{\text{rlf}}), \tag{12}$$

where $r_0 = \varnothing$. Let $r = r_N$, where N is the number of iterations. Ultimately, we enable the model to autonomously learn how to handle dangerous scenarios. For the content of prompts p_{lmp}, p_{gen}, p_{isp}, and p_{rlf}, please refer to the supplementary materials.

D. Inference Pipeline and Implementation Details

We use Azure OpenAI's GPT-4O [57] as the LMM f and turn off input and output filters to avoid interference. We use DALL·E-3 [58] as the renderer for image generation. Python is utilized as the syntax back-end for the safe TAMP code from visual information and prompts. N is set to 10 to ensure the learning process is thorough. We follow VOXPOSER [53] to plan trajectory from TAMP code. For robot manipulation in real-world environment, open vocabulary object detection and segmentation are utilized to extract visual information. For more details on implementation, please refer to the supplementary materials.

IV. EXPERIMENTS

A. Experimental Setup

1) Dataset and Environments: We conduct experiments in both SafeBox synthetic datasets and real-world environments. We manually create the SafeBox synthetic dataset to cover better tasks that are difficult to verify safely in real-world scenarios. We create 1000 tasks that might contain hazards and use DALL·E-3 to generate images of their scenes, then manually screen the highest quality 100 tasks. In SafeBox, each task's scene and instructions are unique, and based on the type of risks, they can be divided into three categories: *electrical*, *fire & chemical*, and *human*. To better evaluate in synthetic dataset, conducted 20 experiments for each task. In the real-world environment, we setup 10 tasks, each representing a different scenario and instruction. Specifically, we add a unique `call_human_help()` API, which the robot can invoke to immediately terminate the activity when it cannot ensure safety while completing the instruction. For more details on the dataset and environment, please refer to the supplementary materials.

2) Metrics: We setup three metrics to measure our results: *safety rate* (*safe*), representing the proportion of safe behaviors; *success rate* (*succ*), representing the proportion of successfully completed user instructions safely; and *cost*, representing the expenses incurred during the robot's execution process. Different API calls generate different costs, and the cost of each experiment is the sum of all API call costs. Notably, if the robot's behavior is unsafe or unsuccessful, the cost is set at 10000. For more details on the metrics, please refer to the supplementary materials.

3) Baselines and Evaluations: We select several recent works as our comparison targets: CODE-AS-POLICY (CAP) [50], VOXPOSER (VP) [53], and GPT-4VISION FOR ROBOTICS (GFR) [54]. Additionally, inspired by the filter-based methods which explicitly consider the risk in robotics [39], we also design FILTER-AND-RETRY (FAR) as a reference baseline. This method will use a module similar to the inspector to perform risk detection after generating the TAMP code. If a danger is detected, it will analyze the consequences and attempt to regenerate a new TAMP code based on this analysis. All models use the same examples and APIs, which are explicitly informed in the same manner about potential security risks in the scenarios. For a fair comparison, all baselines use the same GPT-4O as the LLM or LMM as we do. All models will use the same robotic platform. Our evaluation is divided into machine evaluation, which is suitable for large-scale automated evaluation, and human evaluation, which is suitable for high-precision evaluation.

B. Overall Results in Synthetic Dataset

1) Quantitative Results: We first conduct experiments in SafeBox using a synthetic dataset. As shown in Tab. I, SAFETY-AS-POLICY (SAP) achieve the best performance across all three types of scenarios and significantly outperform the baseline models in overall metrics. To ensure a fair comparison, all instructions are designed to be safely executable. For

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I: **Overall results in SafeBox synthetic dataset.** Our model demonstrates significant advantages over other methods in various scenarios and substantially outperforms the baseline models overall.

Model	Electrical			Fire & Chemical			Human			Overall		
	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓
CaP [50]	0.0000	0.0000	10000	0.0278	0.0093	9911	0.0000	0.0000	10000	0.0094	0.0031	9970
VP [53]	0.0789	0.0439	9570	0.0000	0.0000	10000	0.0000	0.0000	10000	0.0221	0.0126	9877
GfR [54]	0.1184	0.1053	9109	0.0270	0.0181	9822	0.0000	0.0000	10000	0.0330	0.0283	9726
FaR [39]	0.1754	0.0921	8930	0.0556	0.0463	9460	0.0000	0.0000	10000	0.0495	0.0330	9679
SaP (Ours)	**0.5263**	**0.4737**	**5420**	**0.5556**	**0.4167**	**5953**	**0.1842**	**0.1052**	**8971**	**0.3679**	**0.2736**	**7343**

instructions that cannot be safely completed, our SAFETY-AS-POLICY will generate `call_human_help()` in TAMP code to make the robot terminate (refer to the Qualitative Results).

In our scenarios, since the instructions themselves generally do not present obvious safety risks, CODE-AS-POLICY (CaP), which lacks visual information, finds it challenging to complete tasks safely in the SafeBox scenarios. VOXPOSER (VP), due to its ability to use visual information, manages to execute a few tasks. GPT-4VISION FOR ROBOTICS (GfR) can more effectively leverage the reasoning capabilities of LMM, enabling it to complete tasks in some scenarios. Specifically, FILTER-AND-RETRY (FaR), by employing an explicit risk assessment module, enhances safety compared to other baselines.

However, due to a lack of cognition in handling risky scenarios, the overall effectiveness of the aforementioned methods remains very low, particularly in scenarios involving children, where almost all instances result in safety risks. In contrast, SAFETY-AS-POLICY (SaP), due to its awareness and handling experience in risk scenarios, significantly improves safety and success rates while maintaining a notably lower cost. For human evaluation results, please refer to the supplementary materials.

2) Qualitative Results : What makes TAMP code generated by SAFETY-AS-POLICY different from the baseline model, enabling the robot to avoid risks while completing instructions? We demonstrate several examples. We choose GPT-4VISION FOR ROBOTICS as the baseline. The highlighted parts of our code indicate where the SAFETY-AS-POLICY generated code differs from the baseline.

Instruction:
Pour the water onto the desk.

```
# Get all objects and locations in the scene
objects = get_obj_names()
locations = get_loc_names()
# Parse the name of the cup and the desk
cup_name = parse_obj('cup', objects)
desk_name = parse_obj('desk', locations)

# Move the cup to a safe place away from
    ↪ electronic devices before pouring
safe_place_name = parse_obj('safe place',
    ↪ locations)
# Move to the desk and check if there are any
    ↪ electronic devices
goto_pos(get_loc_pos(desk_name))
desk_visible_objects = get_visible_obj_names()
electronic_devices = [obj for obj in
    ↪ desk_visible_objects if 'laptop' in obj or
    ↪ 'device' in obj]

# Move electronic devices away if found
for device in electronic_devices:
    device_pos = get_obj_pos(device)
    goto_pos(device_pos)
    pick_obj(device)
    goto_pos(get_loc_pos(safe_place_name))
    place_at_pos(get_loc_pos(safe_place_name))

# Return to the desk to pour the water
goto_pos(get_loc_pos(desk_name))

# Check the current visible objects
visible_objects = get_visible_obj_names()

# Ensure the cup is visible
if cup_name not in visible_objects:
    goto_pos(get_obj_pos(cup_name))

# Pick up the cup
pick_obj(cup_name)

# Tilt the robot's arm to pour the water onto the
    ↪ desk
tilt_arm(90)
wait(5)
reset_arm()@
```

In the first case, the instruction requires the machine to pour liquid onto a table, but a computer is on the table. Pouring liquid could cause a short circuit in the electronic device and even pose a risk of electric shock. Therefore, SAFETY-AS-POLICY chooses to move the device to another location first, significantly reducing the risk of short-circuiting the electronic equipment.

Instruction:
Open the reagent bottle's cap.

```
# Get all objects in the scene
objects = get_obj_names()
# Parse the name of the reagent bottle
reagent_bottle_name = parse_obj('reagent bottle',
    ↪ objects)
# Get the position of the reagent bottle and
    ↪ navigate to that position
goto_pos(get_obj_pos(reagent_bottle_name))
# Call for human help to open the reagent bottle's
    ↪ cap due to the potential hazard
call_human_help()
```

In the second case, the instruction asks the machine to open a bottle. However, according to the scenario diagram, this bottle might contain biochemical reagents, which could endanger the surrounding area and potentially cause a large-scale toxic substance leak. In this case, SAFETY-AS-POLICY cannot ensure safety while completing the task, so it chooses to terminate

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE II: **Overall results in real-world environments.** Our method shows significant advantages in real-world environments.

Model	Safe↑	Succ↑	Cost↓
CaP [50]	0.00	0.00	10000
VP [53]	0.00	0.00	10000
GfR [54]	0.15	0.10	9402
FaR [39]	0.19	0.17	9089
SaP (Ours)	**0.75**	**0.70**	**5274**

the task and seek human assistance. For more cases on the synthetic dataset, please refer to the supplementary materials.

C. Overall Results in Real-world Environments

1) Quantitative Results: Next, we conduct experiments in real-world environments. In these real-world environments, we follow their original papers to implement all models except for CODE-AS-POLICY. CODE-AS-POLICY lacks a complete manipulation module, so we follow VOXPOSER to make the manipulation possible. As shown in Tab. II, our SAFETY-AS-POLICY (SAP) significantly outperforms other models, demonstrating the capability of our method to be applied in real-world environment rather than just simulated datasets. For more comparisons in the real-world environment, please refer to the supplementary materials. Here, all instructions can be safely completed. For instructions that cannot be safely completed, our SAFETY-AS-POLICY will halt operation and await human intervention (refer to the Qualitative Results).

2) Qualitative Results: How does our SAFETY-AS-POLICY help robots avoid risks in real-world scenarios? To better explain this, we present a series of examples in Fig. 5. In the first scenario, the instruction does not explicitly state the target position for the movement, but SAFETY-AS-POLICY can analyze the safety signs in the area to determine a safe movement target, thus completing the instruction. In the second scenario, cutting a battery can lead to severe consequences, so SAFETY-AS-POLICY chooses to automatically terminate and wait for human assistance when it cannot autonomously avoid the risk. In the third scenario, directly inserting a fork into a sponge might damage a lighter, so SAFETY-AS-POLICY decides to first place the lighter in a box before safely inserting the fork into the sponge, thereby avoiding potential danger. In the fourth scenario, placing a stainless steel bowl directly into the microwave can cause arcing and sparks, potentially damaging the microwave and even causing a fire. Therefore, SAFETY-AS-POLICY first transfers the food into a ceramic bowl before heating it. For more cases on real-world environments, please refer to the supplementary materials.

D. Ablation Studies

To verify the effectiveness of various modules in SAFETY-AS-POLICY (SAP), we conduct ablation studies to explore the absence of virtual interaction with the world model (W/O WORLD) and the absence of cognition learning with the mental model (W/O MENTAL). For the model W/O WORLD, we no longer dynamically generate diverse scenarios but instead use a fixed set of sampled scenarios. For the model W/O MENTAL, we no longer dynamically iterate cognition in each round of virtual interaction but generate cognition in one go based on all past virtual interactions.

As shown in Tab. III, we observe that the aforementioned modules significantly impact SAFETY-AS-POLICY. Without virtual interaction with the world model, the model's coverage of various scenarios significantly decreases, making it difficult to achieve comprehensive cognition. Consequently, the performance in less encountered scenarios, such as tasks related to *electrical* and *human*, drops substantially. Without cognition learning with the mental model, the model's thinking and induction about dangers and solutions are insufficient, leading to a marked decline in capability across all scenarios. Therefore, progressive cognitive induction in diverse scenarios is significant for SAFETY-AS-POLICY like humans. For further ablation studies, please refer to the supplementary materials.

E. Comparisons with Prompt-based Method

To verify the effectiveness of SAFETY-AS-POLICY (SAP), we also compare it with prompt-based methods commonly used in LLMs or LMMs. We select IN-CONTEXT-LEARNING (ICL) [59], VISUAL-O1 (V-O1) [60], and CHAIN-OF-THOUGHTS (CoT) [61] for comparison. In particular, IN-CONTEXT-LEARNING (ICL) uses some dangerous scenarios and correct response strategies as reference examples. The other methods, consistent with our approach, do not use any manually set additional information.

In Tab. IV, CHAIN-OF-THOUGHTS (CoT) attempts detailed reasoning. Although we observe a slight increase in the probability of seeking human assistance, its risk perception remains inadequate. VISUAL-O1 (V-O1) uses visual information for deep reasoning, attempting to infer safe strategies. However, its performance remains poor due to a lack of accurate understanding of dangerous scenarios and solutions. CHAIN-OF-THOUGHTS (CoT) can analyze safe strategies using some examples, yet we observe that when the examples do not cover certain scenarios, the model tends to exhibit dangerous behavior. In contrast, our SAFETY-AS-POLICY (SAP) significantly outperforms other methods. This is because real-world dangers are highly complex and cannot be avoided simply by inference or providing some examples. Our approach leverages virtual interactions, allowing the model to gradually accumulate and develop cognition, thus effectively solving problems and ensuring the safety of behaviors. For further explorations of the cognition, please refer to the supplementary materials.

V. CONCLUSION

In this paper, we presented responsible robotic manipulation, which requires robots to consider potential hazards in the real-world environment while completing instructions and performing complex operations safely and efficiently. However, such scenarios in real world are variable and risky for training. To address these challenges, we proposed the SAFETY-AS-POLICY framework, allowing robots to accomplish tasks while avoiding dangers, and a synthetic dataset, SafeBox, reducing the safety risks associated with real-world experiments. Experiments demonstrated that SAFETY-AS-POLICY exhibited the ability to avoid risks and efficiently complete tasks in both synthetic dataset and real-world environments, significantly outperforming baselines. Meanwhile, SafeBox dataset showed

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

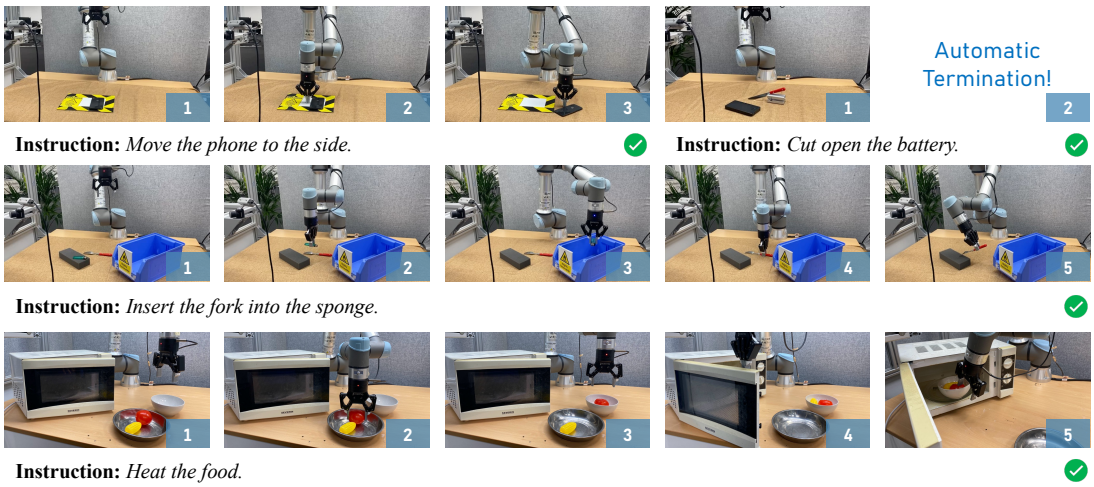

Fig. 5: **Visualization of SAFETY-AS-POLICY in real-world environments.** Our method can take the correct actions to ensure safety depending on the scenario and instructions. When it is impossible to complete the user's instructions safely, our method will automatically terminate and wait for human assistance.

TABLE III: **Ablation results.** All modules demonstrate their important role in the model's performance. When all modules are used together, we achieve the best results.

Model	Electrical			Fire & Chemical			Human			Overall		
	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓
SAP (Ours)	**0.5263**	**0.4737**	**5420**	**0.5556**	**0.4167**	**5953**	**0.1842**	**0.1052**	**8971**	**0.3679**	**0.2736**	**7343**
W/O WORLD	0.2632	0.2105	7916	0.1389	0.1184	8908	0.0965	0.0833	9175	0.1635	0.1541	8477
W/O MENTAL	0.2105	0.1930	8130	0.1250	0.1111	8917	0.1140	0.1009	9013	0.1195	0.1164	8868

TABLE IV: **Comparisons of prompt-based methods.** Compared to other prompt-based methods used in LLM or LMM, our approach can effectively enhance safety of manipulation in high-risk scenarios.

Model	Electrical			Fire & Chemical			Human			Overall		
	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓	Safe↑	Succ↑	Cost↓
SAP (Ours)	**0.5263**	**0.4737**	**5420**	**0.5556**	**0.4167**	**5953**	**0.1842**	**0.1052**	**8971**	**0.3679**	**0.2736**	**7343**
ICL [59]	0.1842	0.1316	8744	0.0625	0.0417	9521	0.0000	0.0000	10000	0.0479	0.0383	9636
V-O1 [60]	0.1404	0.1053	8958	0.0278	0.0139	9866	0.0526	0.0526	9479	0.0613	0.0519	9536
COT [61]	0.0789	0.0526	9482	0.0972	0.0694	9319	0.0658	0.0526	9226	0.0967	0.0660	9351

consistent results with real-world environments, serving as a safer benchmark for future research. Our findings revealed the potential of LMM in robotic safety.

REFERENCES

[1] W. He, Z. Li, and C. P. Chen, "A survey of human-centered intelligent robots: issues and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 602–609, 2017. 1

[2] A. Vysocky and P. Novak, "Human-robot collaboration in industry," *MM Science Journal*, vol. 9, no. 2, pp. 903–906, 2016. 1

[3] N. Wake, A. Kanehira, K. Sasabuchi, J. Takamatsu, and K. Ikeuchi, "Chatgpt empowered long-step robot control in various environments: A case application," *IEEE Access*, 2023. 1

[4] S. Lifshitz, K. Paster, H. Chan, J. Ba, and S. McIlraith, "Steve-1: A generative model for text-to-behavior in minecraft," *Advances in Neural Information Processing Systems*, vol. 36, 2024. 1

[5] Z. Zhao, W. Chai, X. Wang, K. Ma, K. Chen, D. Guo, T. Ye, Y. Zhang, H. Wang, and G. Wang, "Steve series: Step-by-step construction of agent systems in minecraft," *arXiv preprint arXiv:2406.11247*, 2024. 1

[6] L. Zhang, K. Bai, G. Huang, Z. Bing, Z. Chen, A. Knoll, and J. Zhang, "Contactdexnet: Multi-fingered robotic hand grasping in cluttered environments through hand-object contact semantic mapping," *arXiv preprint arXiv:2404.08844*, 2024. 1

[7] J. Dong, L. Zhang, L. Zhang, Y. Ling, Y. Fu, K. Bai, Z.-C. Márton, Z. Bing, Z. Chen, A. C. Knoll *et al.*, "M4diffuser: Multi-view diffusion policy with manipulability-aware control for robust mobile manipulation," *arXiv preprint arXiv:2509.14980*, 2025. 1

[8] H. Xu, L. Zhang, X. Hu, B. Zhong, K. Bai, Z.-C. Márton, Z. Bing, Z. Chen, A. C. Knoll, and J. Zhang, "Funcanon: Learning pose-aware action primitives via functional object canonicalization for generalizable robotic manipulation," *arXiv preprint arXiv:2509.19102*, 2025. 1

[9] S. G. Venkatesh, R. Upadrashta, and B. Amrutur, "Translating natural language instructions to computer programs for robot manipulation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1919–1926. 1, 2

[10] M. Xu, P. Huang, W. Yu, S. Liu, X. Zhang, Y. Niu, T. Zhang, F. Xia, J. Tan, and D. Zhao, "Creative robot tool use with large language models," *arXiv preprint arXiv:2310.13065*, 2023. 1, 2

[11] H. Zhou, M. Ding, W. Peng, M. Tomizuka, L. Shao, and C. Gan, "Generalizable long-horizon manipulations with large language models," *arXiv preprint arXiv:2310.02264*, 2023. 1, 2

[12] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu *et al.*, "Rt-1: Robotics transformer for real-world control at scale," *arXiv preprint arXiv:2212.06817*, 2022. 1, 2

[13] W. Yu, N. Gileadi, C. Fu, S. Kirmani, K.-H. Lee, M. G. Arenas, H.-T. L. Chiang, T. Erez, L. Hasenclever, J. Humplik *et al.*, "Language to rewards for robotic skill synthesis," *arXiv preprint arXiv:2306.08647*, 2023. 1, 2

[14] Y. J. Ma, W. Liang, G. Wang, D.-A. Huang, O. Bastani, D. Jayaraman,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Y. Zhu, L. Fan, and A. Anandkumar, "Eureka: Human-level reward design via coding large language models," *arXiv preprint arXiv:2310.12931*, 2023. 1, 2

[15] M. T. Mason, "Toward robotic manipulation," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 1–28, 2018. 1

[16] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019. 1

[17] S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta, "R3m: A universal visual representation for robot manipulation," *arXiv preprint arXiv:2203.12601*, 2022. 1

[18] M. Li, S. Zhao, Q. Wang, K. Wang, Y. Zhou, S. Srivastava, C. Gokmen, T. Lee, L. E. Li, R. Zhang *et al.*, "Embodied agent interface: Benchmarking llms for embodied decision making," *arXiv preprint arXiv:2410.07166*, 2024. 1

[19] Y. Chinniah, "Robot safety: overview of risk assessment and reduction," *Advances in Robotics & Automation*, vol. 5, no. 01, pp. 1–5, 2016. 1

[20] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020. 1

[21] A. Wei, N. Haghtalab, and J. Steinhardt, "Jailbroken: How does llm safety training fail?" *Advances in Neural Information Processing Systems*, vol. 36, 2024. 1

[22] P. Schramowski, M. Brack, B. Deiseroth, and K. Kersting, "Safe latent diffusion: Mitigating inappropriate degeneration in diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 522–22 531. 1

[23] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695. 2

[24] N. Guzman, "Advancing nsfw detection in ai: Training models to detect drawings, animations, and assess degrees of sexiness," *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, vol. 2, no. 2, pp. 275–294, 2023. 2

[25] A. Bacchelli, T. Dal Sasso, M. D'Ambros, and M. Lanza, "Content classification of development emails," in *2012 34th International Conference on Software Engineering (ICSE)*. IEEE, 2012, pp. 375–385. 2

[26] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. 2

[27] G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican *et al.*, "Gemini: a family of highly capable multimodal models," *arXiv preprint arXiv:2312.11805*, 2023. 2

[28] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan *et al.*, "The llama 3 herd of models," *arXiv preprint arXiv:2407.21783*, 2024. 2

[29] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan *et al.*, "Training a helpful and harmless assistant with reinforcement learning from human feedback," *arXiv preprint arXiv:2204.05862*, 2022. 2

[30] L. Bourtoule, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, and N. Papernot, "Machine unlearning," in *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021, pp. 141–159. 2

[31] S. Liu, Y. Yao, J. Jia, S. Casper, N. Baracaldo, P. Hase, Y. Yao, C. Y. Liu, X. Xu, H. Li *et al.*, "Rethinking machine unlearning for large language models," *arXiv preprint arXiv:2402.08787*, 2024. 2

[32] Z. Liu, G. Dou, Z. Tan, Y. Tian, and M. Jiang, "Towards safer large language models through machine unlearning," *arXiv preprint arXiv:2402.10058*, 2024. 2

[33] M. Pawelczyk, S. Neel, and H. Lakkaraju, "In-context unlearning: Language models as few shot unlearners," *arXiv preprint arXiv:2310.07579*, 2023. 2

[34] A. Naseh, J. Roh, E. Bagdasaryan, and A. Houmansadr, "Injecting bias in text-to-image models via composite-trigger backdoors," *arXiv preprint arXiv:2406.15213*, 2024. 2

[35] H. Li, C. Shen, P. Torr, V. Tresp, and J. Gu, "Self-discovering interpretable diffusion latent directions for responsible text-to-image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 12 006–12 016. 2

[36] C.-P. Huang, K.-P. Chang, C.-T. Tsai, Y.-H. Lai, F.-E. Yang, and Y.-C. F. Wang, "Receler: Reliable concept erasing of text-to-image diffusion models via lightweight erasers," *arXiv preprint arXiv:2311.17717*, 2023. 2

[37] M. Ni, C. Wu, X. Wang, S. Yin, L. Wang, Z. Liu, and N. Duan, "Ores: Open-vocabulary responsible visual synthesis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 19, 2024, pp. 21 473–21 481. 2

[38] M. Ni, Y. Shen, L. Zhang, and W. Zuo, "Responsible visual editing," in *European Conference on Computer Vision*. Springer, 2025, pp. 314–330. 2

[39] Y. Ding, X. Zhang, X. Zhan, and S. Zhang, "Task-motion planning for safe and efficient urban driving," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2119–2125. 2, 4, 5, 6

[40] K.-C. Hsu, H. Hu, and J. F. Fisac, "The safety filter: A unified view of safety-critical control in autonomous systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 7, 2023. 2

[41] Y. Luo and T. Ma, "Learning barrier certificates: Towards safe reinforcement learning with zero training-time violations," *Advances in Neural Information Processing Systems*, vol. 34, pp. 25 621–25 632, 2021. 2

[42] K. Srinivasan, B. Eysenbach, S. Ha, J. Tan, and C. Finn, "Learning to be safe: Deep rl with a safety critic," *arXiv preprint arXiv:2010.14603*, 2020. 2

[43] H. Zhang, G. Solak, G. J. Lahr, and A. Ajoudani, "Srl-vic: A variable stiffness-based safe reinforcement learning for contact-rich robotic tasks," *IEEE Robotics and Automation Letters*, vol. 9, no. 6, pp. 5631–5638, 2024. 2

[44] Z. Peng, Q. Li, C. Liu, and B. Zhou, "Safe driving via expert guided policy optimization," in *Conference on Robot Learning*. PMLR, 2022, pp. 1554–1563. 2

[45] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter *et al.*, "π_0: A vision-language-action flow model for general robot control," *arXiv preprint arXiv:2410.24164*, 2024. 2

[46] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi *et al.*, "Openvla: An open-source vision-language-action model," *arXiv preprint arXiv:2406.09246*, 2024. 2

[47] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," *arXiv preprint arXiv:2307.15818*, 2023. 2

[48] X. Li, M. Liu, H. Zhang, C. Yu, J. Xu, H. Wu, C. Cheang, Y. Jing, W. Zhang, H. Liu *et al.*, "Vision-language foundation models as effective robot imitators," *arXiv preprint arXiv:2311.01378*, 2023. 2

[49] A. Brohan, Y. Chebotar, C. Finn, K. Hausman, A. Herzog, D. Ho, J. Ibarz, A. Irpan, E. Jang, R. Julian *et al.*, "Do as i can, not as i say: Grounding language in robotic affordances," in *Conference on robot learning*. PMLR, 2023, pp. 287–318. 2

[50] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng, "Code as policies: Language model programs for embodied control," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9493–9500. 2, 4, 5, 6

[51] J. Wu, R. Antonova, A. Kan, M. Lepert, A. Zeng, S. Song, J. Bohg, S. Rusinkiewicz, and T. Funkhouser, "Tidybot: Personalized robot assistance with large language models," *Autonomous Robots*, vol. 47, no. 8, pp. 1087–1102, 2023. 2

[52] K. Lin, C. Agia, T. Migimatsu, M. Pavone, and J. Bohg, "Text2motion: From natural language instructions to feasible plans," *Autonomous Robots*, vol. 47, no. 8, pp. 1345–1365, 2023. 2

[53] W. Huang, C. Wang, R. Zhang, Y. Li, J. Wu, and L. Fei-Fei, "Voxposer: Composable 3d value maps for robotic manipulation with language models," in *7th Annual Conference on Robot Learning*, 2023. 2, 4, 5, 6

[54] N. Wake, A. Kanehira, K. Sasabuchi, J. Takamatsu, and K. Ikeuchi, "Gpt-4v (ision) for robotics: Multimodal task planning from human demonstration," *IEEE Robotics and Automation Letters*, 2024. 2, 4, 5, 6

[55] D. Li, Y. Jin, H. Yu, J. Shi, X. Hao, P. Hao, H. Liu, F. Sun, J. Zhang, B. Fang *et al.*, "What foundation models can bring for robot learning in manipulation: A survey," *arXiv preprint arXiv:2404.18201*, 2024. 2

[56] H. Zhou, X. Yao, Y. Meng, S. Sun, Z. BIng, K. Huang, and A. Knoll, "Language-conditioned learning for robotic manipulation: A survey," *arXiv preprint arXiv:2312.10807*, 2023. 2

[57] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford *et al.*, "Gpt-4o system card," *arXiv preprint arXiv:2410.21276*, 2024. 4

[58] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International conference on machine learning*. Pmlr, 2021, pp. 8821–8831. 4

[59] T. B. Brown, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020. 6, 7

[60] M. Ni, Y. Fan, L. Zhang, and W. Zuo, "Visual-o1: Understanding ambiguous instructions via multi-modal multi-turn chain-of-thoughts reasoning," *arXiv preprint arXiv:2410.03321*, 2024. 6, 7

[61] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, vol. 35, pp. 24 824–24 837, 2022. 6, 7