Hazards in Daily Life?
Enabling Robots to Proactively Detect and Resolve Anomalies

Anonymous ACL submission

Abstract

Existing household robots have made significant progress in performing routine tasks, such as cleaning floors or delivering objects. However, a key limitation of these robots is their inability to recognize potential problems or dangers in home environments. For example, a child may pick up and ingest medication that has fallen on the floor, posing a serious risk. We argue that household robots should proactively detect such hazards or anomalies within the home, and propose the task of *anomaly scenario generation*. We leverage foundational models instead of relying on manually labeled data to build simulated environments. Specifically, we introduce a multi-agent brainstorming approach, where agents collaborate and generate diverse scenarios covering household hazards, hygiene management, and child safety. These textual task descriptions are then integrated with designed 3D assets to simulate realistic environments. Within these constructed environments, the robotic agent learns the necessary skills to proactively discover and handle the proposed anomalies through task decomposition, and optimal learning approach selection. We demonstrate that our generated environment outperforms others in terms of task description and scene diversity, ultimately enabling robotic agents to better address potential household hazards. The Anonymous Github is `https://anonymous.4open.science/r/AnomalyGen-E2F9`.

1 Introduction

The development of Vision-Language Models (VLMs) has significantly improved household robots' ability to interact with the physical world in a more human-like manner (Liu et al., 2024b,a; Cai et al., 2023; Majumdar et al., 2024). Among these models, the most popular paradigm for such robots is receiving instructions and performing corresponding operational tasks (Yang et al., 2024; Driess et al., 2023; Ahn et al., 2022).

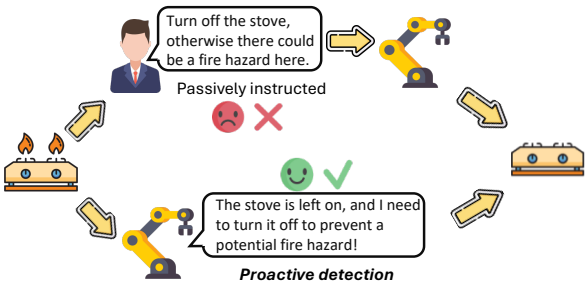

Figure 1: Comparison of passively instructed robots and our proactive detection robot. Our paradigm creates benefits and convenience for safety, even in the absence of human presence.

However, a critical yet often overlooked scenario arises when no instructions are provided. According to survey data, 31% of cooking fires are caused by unattended equipment (Ahrens, 2020). Meanwhile, unintentional injuries are the predominant cause of death among children, particularly those aged 1-14 years, encompassing incidents such as drowning, falls, and accidental poisonings (Worldwide, 2022). A lack of adequate supervision is often identified as a significant contributor to many of these fatalities, especially in cases involving younger children (Hymel et al., 2006; Williams and Kotch, 2023). It would greatly benefit humans if household robots could monitor whether stoves and other fire sources are properly turned off and detect potential hazards in the home that could lead to falls or accidental poisonings. Many of these fires and unintentional injuries could be prevented. However, to the best of our knowledge, such robotics have yet to be implemented.

Hence, in this work, we propose AnomalyGen, which can generate diverse *anomaly settings covering household hazards, hygiene management, and child safety* in 3D simulation environments, enabling robots to develop proactive detection and problem-solving abilities, as shown in Figure 1. Specifically, we first devised a group brainstorming

setting, where LLM-based agents collaborate to generate diverse and comprehensive anomaly scenarios. The motivation comes from the observation that simply prompting an LLM to generate hazard scenarios results in repetitive and similar settings. In contrast, group brainstorming in real-life meetings often leads to novel and creative ideas. Based on these task settings, AnomalyGen automatically constructs simulated anomalous scenes through carefully designed 3D asset retrieval, configuration, and scene setup steps. Finally, AnomalyGen guides household robots in developing detection and resolution abilities for handling anomalies. It reads textual descriptions of the simulated environment, including the 3D coordinates of assets, and automatically identifies potential anomalous tasks that require attention. AnomalyGen then decomposes the task into fine-grained sub-tasks and selects the most appropriate learning method for the household robot. In general, our AnomalyGen leverages language-based approaches to bridge the domain gap between foundational models and robot interaction, enabling operations such as control inputs, operational trajectories, and physical interaction.

For the experiments, AnomalyGen constructs 111 diverse and comprehensive anomaly scenes, with human evaluation showing high quality and automatic metrics demonstrating greater diversity compared to previous human-crafted robotic datasets. Based on this simulation data, household robots are guided by AnomalyGen to learn and demonstrate a variety of skills across tasks such as rigid and articulated object manipulation and legged locomotion, achieving a task completion rate of 83%. Additionally, we conduct an error analysis highlighting the limitations of the current learning algorithm and VLM, identifying areas for future improvement and direction.

Our contributions can be summarized as follows: Firstly, we introduce AnomalyGen, an unsupervised generative framework that enables household robots to autonomously detect and address anomalies without explicit instructions. Secondly, AnomalyGen creates a 3D simulation environment with 111 diverse hazard scenarios, generated through a collaborative brainstorming mechanism, significantly enhancing task diversity compared to previous datasets. Thirdly, AnomalyGen enables robots to autonomously identify anomalies, decompose tasks, and learn appropriate skills using an effective task decomposition and learning method with minimal human input.

2 Related Work

2.1 Household Anomaly Detection

Household robotics have seen significant advancements, with researchers developing various benchmarks for embodied AI agents to tackle household tasks in simulation. Behavior1K (Li et al., 2021b; Srivastava et al., 2021) introduced a benchmark for AI agents to complete 100 household activities in a simulated environment. Housekeep (Kant et al., 2022) focuses on organizing homes by rearranging cluttered items, while TidyBot (Wu et al., 2023) emphasizes personalized household cleanup, aiming to understand and place items in their correct locations. However, limited work has been done on anomaly detection in household environments. The only notable dataset addressing safety-related topics in this domain is SafetyDetect (Mullen et al., 2024), which manually configured just seven distinct scenes and required substantial human effort for scene construction and data collection. Additionally, it is an image-based dataset, not a simulation. In contrast, AnomalyGen autonomously generates simulation environment and tasks without human intervention.

2.2 Foundation Models in Robotics

With the rapid development of foundation and generative models in multi-modal settings (Poole et al., 2022; Melas-Kyriazi et al., 2023; Touvron et al., 2023; Driess et al., 2023; OpenAI, 2023; Liu et al., 2023; Girdhar et al., 2023), a growing body of research has begun to harness the capabilities of foundational models across various domains, such as visual imagination for skill execution (Du et al., 2023), and sub-task planning (Ahn et al., 2022; Huang et al., 2022; Lin et al., 2023), among others. Some recent works have also attempted to fully harness the potential of LLMs for robotic manipulation, such as using LLMs for reward function generation (Yu et al., 2023; Ma et al., 2023), and sub-task and trajectory generation (Ha et al., 2023). Additionally, GenSim explored LLM-based robotic instruction tasks, but it primarily focused on object manipulation on desktops with a limited set of 3D assets. Gen2Sim (Katara et al., 2023) extends the range of task types by generating instead of only retrieving new 3D assets.

2.3 Simulation Environment

VirtualHome (Puig et al., 2018) and Alfred (Shridhar et al., 2020a) abstract physical interactions to

concentrate on symbolic reasoning, yet they lack physical realism and a comprehensive scope of actions. Habitat (Savva et al., 2019), employing 3D scans of real homes, focuses on navigation tasks (Batra et al., 2020) but omits physics-based interactions. To augment physical realism, Habitat 2.0 (Yenamandra et al., 2023) and iGibson (Srivastava et al., 2021; Li et al., 2021a) introduce realistic actions, interactions with environments, and object state simulations. Additionally, emerging simulation platforms such as ManiSkills (Gu et al., 2023), TDW (Gan et al., 2022), SoftGym (Lin et al., 2021), and RFUniverse (Fu et al., 2022) emphasize physical realism but still fall short on task diversity. To enrich task variety, several works have explored language-conditioned tasks (Zeng et al., 2021; James et al., 2020b; Mees et al., 2022). AnomalyGen focuses on constructing household anomaly scenes, which is an area that has not been covered in previous work.

3 Method

In this section, we demonstrate how our framework utilizes advanced generative models to automatically create anomalous scenarios and task-related data, as shown in Figure 2.

3.1 Brainstormed Anamoly Task Proposal

The most intuitive way to obtain anomaly scenarios would be to prompt an LLM to generate a list. However, in our preliminary experiments, the LLM tends to generate repetitive and lackluster scenarios, such as "move the scissors to a drawer", "put the scissors into the box" and "store scissors safely." This lack of diversity limits the range of potential hazards, resulting in scenarios that are too similar to each other.

To address this, we propose group brainstorming, a round-based divergent thinking framework that allows multiple agents to build upon each other's ideas. We also incorporate role-playing, where each agent adopts a unique perspective, encouraging a broader range of creative thoughts. This collaborative process not only enhances the variety of scenarios but also improves the realism and complexity of the generated anomalies.

Role-play Initialization Stage. To ensure that each agent considers different perspectives and approaches the task from various angles, we assign distinct roles to each agent. This is especially useful in households, where different professions face diverse hazard scenarios daily. For example, a parent may focus on child safety, noticing hazards like sharp objects or unlocked cabinets, while a household maintenance worker may be more attuned to issues like electrical faults or fire hazards. Our role-play list includes roles such as homemaker, household safety advisor, and educational consultant, with the full list in the Appendix B.1. Based on their assigned role, each agent randomly selects a *target object* from an anomalous household asset list and proposes an initial anomaly scenario based on that object. This anomalous object list was curated from a subset of the PartNet Mobility dataset (Xiang et al., 2020; Wang et al., 2023b). The detailed categories and directory of these anomalous object assets can be found in the Appendix A.

Brainstorming Stage. After each agent proposes its initial anomaly scenario, we facilitate multiple rounds of discussion to foster meaningful exchanges of ideas among the LLM agents. Each agent takes the outputs from other agents in the previous rounds, combines them with its own character and thoughts, and proposes new tasks. Each agent is informed that they are part of a collaborative brainstorming session, where teamwork and diverse perspectives are key to generating creative and comprehensive hazard scenarios. The detailed prompt can be found in Appendix B.2

For instance, as shown in Figure 2, the "gardener" and "engineer" agents propose a scenario where sharp tools left on a wet floor and exposed electrical wires near water pose significant safety risks. Motivated by the sharp tools mentioned by the gardener and the electrical hazards raised by the engineer, the "homemaker" suggests that leaving kitchen appliances plugged in near the sink could also lead to electrocution, expanding on the dangers of water-related hazards in the home. This iterative dialogue mimics human brainstorming, where participants build on each other's ideas for more creative and comprehensive outcomes. By encouraging LLM agents to collaborate, we achieve greater variety and depth in the generated anomaly scenarios.

3.2 3D Anomalous Scenarios Generation

The brainstormed ideas above are textual descriptions of the scenario, and our next step is to turn the text into vivid 3D environments.

Auxiliary 3D Assets Retrieval. In the previous section, we compiled a list of 3D anomaly assets

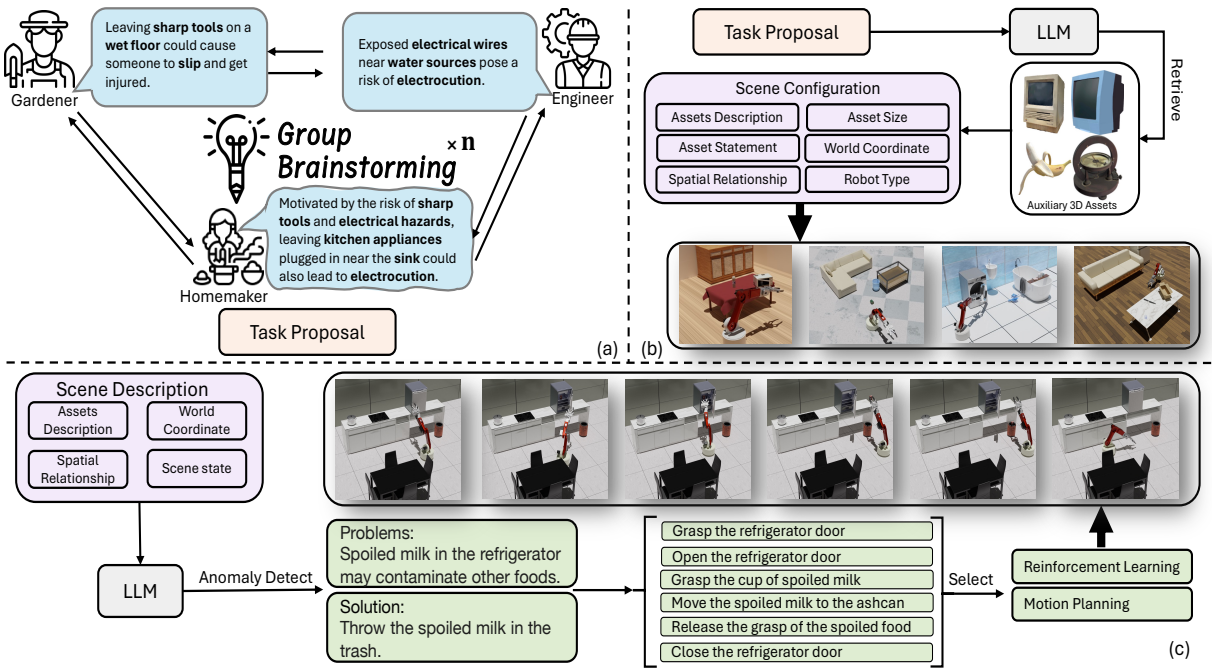

Figure 2: AnomalyGen includes 3 modules: a) Group Brainstorming, b) Anomalous Scenarios Generation, c) Proactive Anomaly Detection and Anomaly Task Learning.

that serve as *target objects* for the robots to manipulate. However, to realistically simulate real-world scenarios, focusing only on these target objects is not enough. We also need *auxiliary surrounding objects* to construct realistic environments that mimic real-world object distributions. A straightforward approach to select these surrounding objects is to source them from Objaverse (Deitke et al., 2023). However, this dataset is extensive, containing up to 800K items, and the item names are often too short or duplicated, making selection challenging.

To address this, we first query LLM to generate names and descriptions of objects relevant to the anomalous task, from a *textual perspective*. We then employ Sentence-BERT (Reimers and Gurevych, 2019) to retrieve the top-k textually similar 3D assets from the Objaverse list based on these descriptions. From a *visual perspective*, we further validate the selection by using a VLM. The VLM takes the task name, detailed task description, and the assets annotations as input. It then determines whether the scene is valid and outputs "yes" to confirm alignment between the task description and the scene setup.

In this way, we compile a high-quality list of relevant 3D assets that undergo dual validation through both textual and visual steps.

Asset Configuration. Assets configuration ensures that the retrieved 3D assets have physically plausible dimensions. To automate this process, we employ LLM to determine the size of each asset. The size of each object is calculated as a scalar value in meters, representing its largest dimension. In addition, we establish *spatial relationship rules*, where objects with relative size relationships must satisfy specific task requirements. For example, in a scenario where a "bowl of soup" needs to be placed inside a microwave, if the bowl's size is 0.15 meters, the microwave must have a dimension larger than 0.15 meters to accommodate it. We also define *initial state rules*, ensuring that objects are in the appropriate state for the task. For example, for the task "turn off faucet," the faucet must be initialized in the "on" state to accurately simulate the conditions necessary for the task's execution.

Scene Configuration. After configuring the assets, the final step is to position them accurately while maintaining appropriate spatial relations. To achieve precise placement, we query an LLM to establish a 3D world coordinate system (x, y, z). Target assets required for specific tasks are strategically placed within the constrained space of $(0, 0, 0)$ to $(1, 1, 1)$, while auxiliary assets not involved in the current task are positioned outside this specified range.

We will demonstrate through human evaluation in §5.4 that the environment constructed in this manner is of high quality.

3.3 Proactive Anomaly Detection

Although existing research extensively explores ways to enhance a robot's ability to follow instructions, a key limitation is their inability to actively detect anomalies or dangerous situations in daily life (Fan et al., 2024), an ability that is crucial for ensuring safety in dynamic environments (Lundström et al., 2015; Wang et al., 2023b). In our simulated 3D environment, we aim to enable household robots to proactively detect hazards or anomalies and acquire the necessary skills to solve tasks related to these anomalies. Concretely, AnomalyGen employs an LLM to analyze and identify potential problems. The input for the LLM includes outputs from the previous steps, including the target object, retrieved assets, their configurations, and the overall scene setup. Note that LLM has no access to the task name and description, but only infers the potential task based on the environment observation. Based on this information, the LLM is prompted to generate possible problems that need to be solved and their solutions. For example, a problem could be that a folding knife on the table may cause cuts or injuries, and the solution would be to store the folding knife in a storage box.

3.4 Anomaly Task Learning

After confirming the task to be completed, AnomalyGen queries the LLM to decompose the detected solution into shorter-horizon sub-tasks, as illustrated in the bottom part of Figure 2.

Then, different learning algorithms are selected, tailored to different subtasks: reinforcement learning (Schulman et al., 2017; Haarnoja et al., 2023) and action primitives with motion planning (Karaman and Frazzoli, 2011). Each algorithm has its strengths: Reinforcement learning is ideal for dynamic, contact-rich environments, like legged locomotion or adjusting appliance controls, and Action primitives with motion planning handle navigation through cluttered environments, ensuring safe and efficient paths. We introduce the strengths of each algorithm and provide three examples of action-algorithm pairs to the LLM, enabling it to select the most appropriate learning algorithm for each subtask by in-context learning.

Meanwhile, for subtasks trained using reinforcement learning, the LLM is responsible for generating the corresponding reward functions. For tasks involving rigid manipulation and locomotion, these reward functions are derived from low-level states accessible to the LLM. In contrast, for tasks involving soft body manipulation, the reward functions are based on the earth-mover distance between the particles of the current and target shapes, ensuring precise shape matching. To simplify object grasping and approaching action in action primitives subtasks, we use a robot equipped with a suction cup. This setup streamlines the process of grasping. The simplified pseudo-algorithm for the grasping and approaching primitives is in Appendix C.1.

4 Experiment Setup

4.1 Implementation Details

Our proposed system is generic and agnostic to specific simulation platforms. However, considering the broad audience for simulation platforms, as following (Wang et al., 2023b; Kant et al., 2022; Gu et al., 2023; Shridhar et al., 2020b; Savva et al., 2019). we choose Genesis (Katara et al., 2023), the most widely employed simulation platform for deployment. The model itself is general-purpose and independent of any specific simulation platform. We employ the state-of-the-art `GPT-4-0314` LLM and BLIP-2 (Li et al., 2023b) VLM by default. For anomaly task learning, we utilize the Soft Actor-Critic (SAC) (Haarnoja et al., 2018) as the reinforcement learning algorithm, employing learning rate $3e-4$ for the actor, the critic, and the entropy regularizer. The horizon of manipulation sub-tasks is 100, with a frameskip of 2. For each sub-task, we train with 1M environment steps. We also employ Batch Informed Trees (Gammell et al., 2015) as the motion planning algorithm. More details are in the Appendix C.2.

4.2 Baselines

We compare our constructed environment with the latest benchmark environments, including RLBench (James et al., 2020a), which encompasses 100 distinct, meticulously crafted tasks, ranging in complexity from basic tasks like target-reaching and door-opening to more advanced, multi-step tasks. ManiSkill2 (Gu et al., 2023) includes 20 manipulation task families which cover stationary/mobile-base, single/dual-arm, and rigid/soft-body manipulation tasks with 2D/3D-input data. Meta-World (Yu et al., 2020), serves as a benchmark specifically designed for evaluating the performance of meta-reinforcement learning and multitask learning algorithms. Behavior-100 (Li et al., 2023a) featuring 100 activities within

	AnomalyGen	RoboGen	Behavior-100	RLbench	MetaWorld	Maniskill2	GenSim
Number of Tasks	111	106	100	106	50	20	70
Task Description - Self-BLEU ↓	**0.227**	0.287	0.299	0.317	0.322	0.674	0.378
Task Description - SentenceBert ↓	**0.245**	0.394	0.210	0.200	0.263	0.194	0.288
Scene Image - Embedding Similarity (ViT) ↓	**0.315**	0.353	0.389	0.375	0.517	0.332	0.717
Scene Image - Embedding Similarity (CLIP) ↓	**0.805**	0.824	0.833	0.864	0.867	0.828	0.932

Table 1: **Task diversity comparison** with leading human-designed robotics datasets, including Behavior-100, RLBench, MetaWorld, Maniskill2, and GenSim.

simulated environments. These activities encompass a variety of routine household tasks, including cleaning, maintenance, and food preparation. GenSim (Wang et al., 2023a) offers 100 robotic arm grasping scenarios, all set on a tabletop environment. RoboGen (Wang et al., 2023b) has generated 106 more diverse tasks, further expanding them to accommodate a wider range of robotic arm types. Note that RoboGen doesn't release its constructed dataset; therefore, we reimplement RoboGen based on the provided code.

4.3 Evaluation Metrics

We first evaluate the *diversity* of the generated anomaly task settings, including the semantic aspects of the tasks and the visual aspects of the scenes. For semantic view, we concatenate the task name and description from each anomalous task and calculate their similarity by Self-BLEU (Papineni et al., 2002) and SentenceBERT (Reimers and Gurevych, 2019) for quantitative analysis. For scene visual information, we assess scene diversity by measuring the embedding similarity of unrendered images of the scene from the initial camera state. Specifically, we utilize ImageNet pre-trained Vision Transformer (ViT) (Dosovitskiy et al., 2020) and CLIP models (Radford et al., 2021).

Next, we assess whether the robot can proactively and accurately *detect* anomaly scenarios. We employ three undergraduate students specializing in computer-related fields to determine whether the task detected by the anomaly detection module aligns with the ground truth task.

Finally, the annotators review task action videos to determine whether the task is successfully completed and evaluate the *overall success rate* of task completion. Note that even if the detected tasks do not fully align with the designed tasks in the task proposal, evaluations are still based on the detected tasks, as these tasks possess a certain degree of rationality, as explained in §5.3.

5 Results and Analysis

5.1 Anomaly Task Statistics

Through our group brainstorming component, we can theoretically generate an unlimited number of diverse tasks. To facilitate the evaluation process, we limited the number of evaluated scenes to 111. These scenes are categorized into three general categories: household hazards, hygiene management, and child safety. Each category includes diverse tasks, as visualized in Figure 4, where "store sharp objects safely" represents the largest proportion of tasks at 28.8%, followed by other critical tasks such as "reduce tripping hazards" and "manage chemical risks," all of which represent common dangers encountered in real life. Note that no human intervention was involved in the entire design process, so the distribution of tasks is the result of fully autonomous generation by our model.

5.2 Anomaly Task Diversity

In Table 1, we show the diversity of our model and baselines on the constructed anomalies task from both text and visual perspectives. It can be seen that our model achieves the lowest Self-BLEU and SentenceBERT similarity scores for text, as well as the lowest ViT and CLIP scores for visuals, indicating that our framework surpasses previous manually constructed benchmarks and datasets.

Group Brainstorming Ablation. To demonstrate that the diversity described above is due to our group brainstorming setting rather than the LLM's inherent abilities, we conduct an ablation study. The first setting removes both the brainstorming and role-play components introduced in §3.1, directly querying LLM to generate task proposals. The second setting incorporates role-playing, where we provide the role settings in the prompt and query LLM to propose tasks from the perspective of the assigned character. To comprehensively evaluate the generation ability, we generate 300 anomalous task proposals. Concretely, we use identical query parameters for both ex-

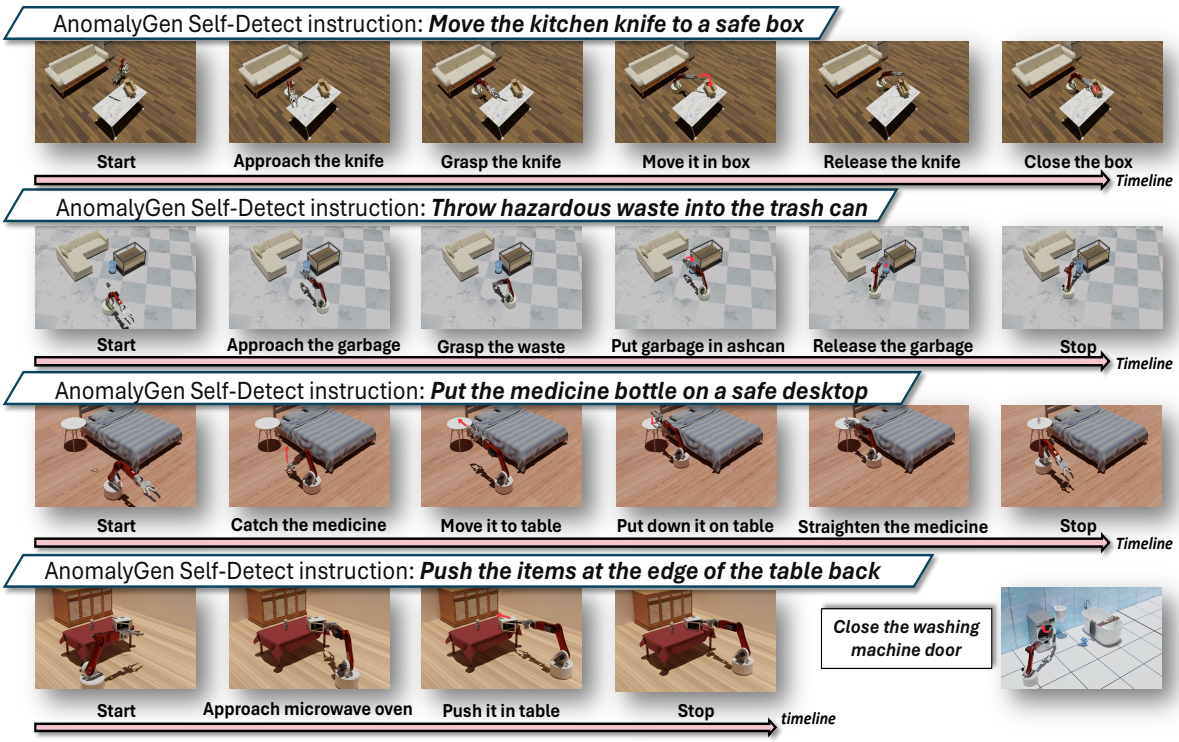

Figure 3: Snapshots of the learned skills across 4 exemplary long-horizon sequential tasks and 1 single-step task.

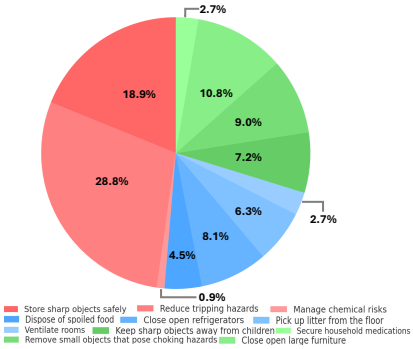

Figure 4: Distribution of types of anomalous scenarios. The red color represents "Household Hazards," the blue color denotes "Hygiene Management," and the green color denotes "Child Safety Measures."

Method	Self-BLEU (↓)	SentenceBERT (↓)	WMD (↓)
w/o brainstorming & role-play	0.217	0.553	0.618
w/o brainstorming	0.225	0.551	0.605
AnomalyGen	**0.043**	**0.393**	**0.511**

Table 2: Ablation experiment results on group brainstorming and role-play. Lower values indicate better performance across all metrics.

periments and randomly selected 10 assets as targets. Each method is run 10 times, generating 10 task proposals per iteration. We employ Self-BLEU, SentenceBERT, and Word Mover's Distance (WMD) (Huang et al., 2016) for evaluation.

The experiment results are presented in Table 2. It can be seen that with role-play, the SentenceBERT and WMD scores decrease by 0.002 and 0.007, respectively, compared to the naive LLM approach. Then, brainstorming brings the greatest contribution, leading to the largest improvement across all three metrics, including a 0.182 decrease in Self-BLEU, 0.158 decrease in SentenceBERT, and 0.107 decrease in WMD. This demonstrates that the combination of brainstorming and role-play significantly enhances the model's performance by promoting better alignment and understanding, as reflected in the lower values across all metrics.

5.3 Anomaly Detection Performance

For the anomaly detection task, we allow AnomalyGen to come up with up to three solutions, and manually evaluate the hit accuracy for k solutions. As shown in Table 3, our model generally achieves high performance, with 76% accuracy on the first attempt and 82% accuracy when given three attempts. This demonstrates the effectiveness of the prompt we design for AnomalyGen to correctly identify the task, as well as the validity of our constructed simulation environment, where the anomaly scenes can be accurately detected.

In addition, we conduct an error analysis on the

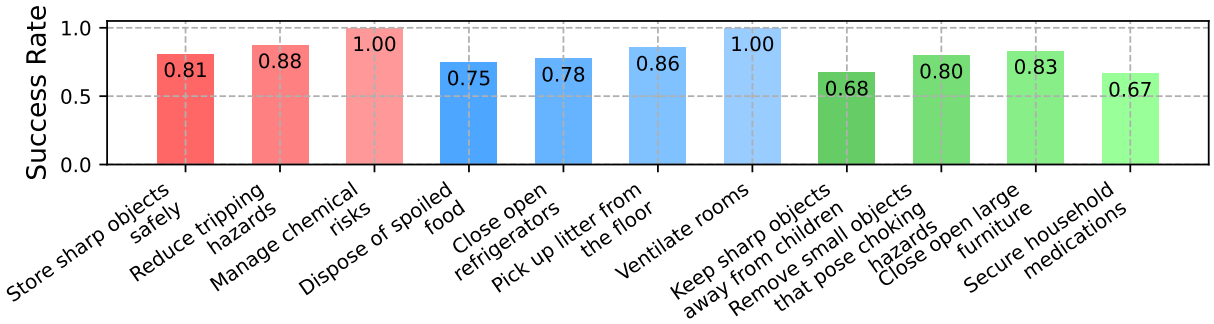

Figure 5: Anomaly resolution completion rate across different categories.

k Solutions	Success Rate
1	0.759
2	0.804
3	0.821

Table 3: Anomaly detection success rate with k solutions.

failed tasks. We find that the main reason for failure was that our environment closely mirrors real-life scenarios, including a variety of everyday clutter, which mislead the LLM into making incorrect detections. Meanwhile, we note that the tasks proposed by the detection module still possess a degree of rationality. For instance, while the ground truth involves picking up a pill from the floor and placing it on a table, our detection module instead proposes to discard it. Although this action does not align with the specified ground truth, the alternative task remains meaningful, as discarding a potentially misplaced pill could also be seen as a reasonable safety precaution.

5.4 Anomaly Solution Performance

Lastly, we can evaluate the performance of overall task performance. We first give some examples of the task execution process in Figure 3, including multi-step tasks, such as placing a medicine bottle on a safe desktop, and single-step tasks, such as closing the washing machine door. It can be observed that the model successfully follows the instructions and completes tasks of varying complexity and time steps.

Next, we conduct a quantitative analysis of the anomaly resolution accuracy, using human evaluations to assess the completion rate across different categories, as in Figure 5. Among the 111 generated anomalous scenarios, the average success rate for resolving these tasks was 83%, highlighting AnomalyGen's strong execution performance.

We also conduct an error analysis and summarize two main reasons for task failure. First, there is an overlap of scene assets. In the verification step described in §3.2, we employ VLMs to verify that the assets are correctly positioned. However, there are 4 out of 111 instances where the VLMs fail to identify mispositioned items and incorrectly approve them. For example, when the mispositioned items are both white, accurate identification becomes more difficult. The misposition problem exists in other environments (Wang et al., 2023b), and we anticipate that advancements in VLM technology will address this limitation. Second, some tasks proposed by AnomalyGen include complex, multi-step actions that challenge the capabilities of current algorithms, making it difficult for them to perform effectively. We expect that improvements in learning algorithms will enable robots to better learn from their environments and handle more complex tasks in the future.

6 Conclusion

In this study, we present AnomalyGen, an innovative framework designed to enhance the proactive detection and resolution of household anomalies by robots. Our approach integrates advanced generative models to automatically create diverse and realistic 3D environments, which are essential for training robots to handle real-world tasks autonomously. We also propose a group brainstorming method, which generates a wide variety of anomalous scenarios, surpassing traditional methods that rely heavily on manual input. Furthermore, the AnomalyGen framework introduces a novel approach to anomaly detection, offering potential strategies for enabling robots to act without direct human commands. We hope our work will inspire further exploration into autonomous decision-making in real-world applications.

7 Limitation

While AnomalyGen has achieved certain milestones, it still encounters several limitations:

1) In unsupervised settings, the validation of tasks within generated anomaly scenarios remains challenging, with a potential for scenarios that clearly do not meet task requirements. This issue is particularly exacerbated under conditions of large-scale generation. However, with future enhancements in the capabilities of multimodal large language models, we anticipate that this limitation will be addressed.

2) The richness of the generated scenes is currently somewhat constrained by the scale of the 3D assets dataset. A limited dataset size may curtail the full potential of AnomalyGen.

3) Regarding the deployment of AnomalyGen into real-world applications, there remains a significant sim-to-real domain gap. This gap constitutes an independent research domain that is beyond the scope of our current work. Given recent rapid advancements in physically accurate simulation (Li et al., 2020) and techniques such as domain adaptation (Tobin et al., 2017; Xu et al., 2023; Xiao et al., 2024) along with realistic sensory signal rendering (Zhuang et al., 2023), we anticipate a continual narrowing of this gap in the near future.

References

Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, et al. 2022. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*.

Marty Ahrens. 2020. Home cooking fires.

Dhruv Batra, Aaron Gokaslan, Aniruddha Kembhavi, Oleksandr Maksymets, Roozbeh Mottaghi, Manolis Savva, Alexander Toshev, and Erik Wijmans. 2020. ObjectNav Revisited: On Evaluation of Embodied Agents Navigating to Objects. In *arXiv:2006.13171*.

Rizhao Cai, Zirui Song, Dayan Guan, Zhenhao Chen, Xing Luo, Chenyu Yi, and Alex Kot. 2023. Benchlmm: Benchmarking cross-style visual capability of large multimodal models. *arXiv preprint arXiv:2312.02896*.

Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. 2023. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. 2023. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*.

Yilun Du, Mengjiao Yang, Bo Dai, Hanjun Dai, Ofir Nachum, Josh Tenenbaum, Dale Schuurmans, and Pieter Abbeel. 2023. Learning universal policies via text-guided video generation. *arXiv preprint arXiv:2302.00111*.

Haolin Fan, Xuan Liu, Jerry Ying Hsi Fuh, Wen Feng Lu, and Bingbing Li. 2024. Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics. *Journal of Intelligent Manufacturing*, pages 1–17.

Haoyuan Fu, Wenqiang Xu, Han Xue, Huinan Yang, Ruolin Ye, Yongxi Huang, Zhendong Xue, Yanfeng Wang, and Cewu Lu. 2022. Rfuniverse: A physics-based action-centric interactive environment for everyday household tasks. *arXiv preprint arXiv:2202.00199*.

Jonathan D Gammell, Siddhartha S Srinivasa, and Timothy D Barfoot. 2015. Batch informed trees (bit*): Sampling-based optimal planning via the heuristically guided search of implicit random geometric graphs. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 3067–3074. IEEE.

Chuang Gan, Siyuan Zhou, Jeremy Schwartz, Seth Alter, Abhishek Bhandwaldar, Dan Gutfreund, Daniel LK Yamins, James J DiCarlo, Josh McDermott, Antonio Torralba, et al. 2022. The threedworld transport challenge: A visually guided task-and-motion planning benchmark towards physically realistic embodied ai.

Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. 2023. Imagebind: One embedding space to bind them all. *arXiv preprint arXiv:2305.05665*.

Jiayuan Gu, Fanbo Xiang, Xuanlin Li, Zhan Ling, Xiqiang Liu, Tongzhou Mu, Yihe Tang, Stone Tao, Xinyue Wei, Yunchao Yao, et al. 2023. Maniskill2: A unified benchmark for generalizable manipulation skills. *arXiv preprint arXiv:2302.04659*.

Huy Ha, Pete Florence, and Shuran Song. 2023. Scaling up and distilling down: Language-guided robot skill acquisition. *arXiv preprint arXiv:2307.14535*.

Tuomas Haarnoja, Ben Moran, Guy Lever, Sandy H Huang, Dhruva Tirumala, Markus Wulfmeier, Jan Humplik, Saran Tunyasuvunakool, Noah Y Siegel, Roland Hafner, et al. 2023. Learning agile soccer skills for a bipedal robot with deep reinforcement learning. *arXiv preprint arXiv:2304.13653*.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR.

Gao Huang, Chuan Guo, Matt J Kusner, Yu Sun, Fei Sha, and Kilian Q Weinberger. 2016. Supervised word mover's distance. *Advances in neural information processing systems*, 29.

Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. 2022. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608*.

Kent P Hymel et al. 2006. When is lack of supervision neglect? *Pediatrics*, 118(3):1296–1298.

Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J Davison. 2020a. Rlbench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, 5(2):3019–3026.

Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J. Davison. 2020b. Rlbench: The robot learning benchmark & learning environment.

Yash Kant, Arun Ramachandran, Sriram Yenamandra, Igor Gilitschenski, Dhruv Batra, Andrew Szot, and Harsh Agrawal. 2022. Housekeep: Tidying virtual households using commonsense reasoning. *Preprint*, arXiv:2205.10712.

Sertac Karaman and Emilio Frazzoli. 2011. Sampling-based algorithms for optimal motion planning. *The international journal of robotics research*, 30(7):846–894.

Pushkal Katara, Zhou Xian, and Katerina Fragkiadaki. 2023. Gen2sim: Scaling up robot learning in simulation with generative models. *arXiv preprint arXiv:2310.18308*.

Chengshu Li, Fei Xia, Roberto Martín-Martín, Michael Lingelbach, Sanjana Srivastava, Bokui Shen, Kent Vainio, Cem Gokmen, Gokul Dharan, Tanish Jain, et al. 2021a. igibson 2.0: Object-centric simulation for robot learning of everyday household tasks. *arXiv preprint arXiv:2108.03272*.

Chengshu Li, Fei Xia, Roberto Martín-Martín, Michael Lingelbach, Sanjana Srivastava, Bokui Shen, Kent Vainio, Cem Gokmen, Gokul Dharan, Tanish Jain, Andrey Kurenkov, Karen Liu, Hyowon Gweon, Jiajun Wu, Li Fei-Fei, and Silvio Savarese. 2021b. igibson 2.0: Object-centric simulation for robot learning of everyday household tasks. In *Conference in Robot Learning (CoRL)*, page accepted.

Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Michael Lingelbach, Jiankai Sun, et al. 2023a. Behavior-1k: A benchmark for embodied ai with 1,000 everyday activities and realistic simulation. In *Conference on Robot Learning*, pages 80–93. PMLR.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023b. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR.

Minchen Li, Zachary Ferguson, Teseo Schneider, Timothy R Langlois, Denis Zorin, Daniele Panozzo, Chenfanfu Jiang, and Danny M Kaufman. 2020. Incremental potential contact: intersection-and inversion-free, large-deformation dynamics. *ACM Trans. Graph.*, 39(4):49.

Kevin Lin, Christopher Agia, Toki Migimatsu, Marco Pavone, and Jeannette Bohg. 2023. Text2motion: From natural language instructions to feasible plans. *arXiv preprint arXiv:2303.12153*.

Xingyu Lin, Yufei Wang, Jake Olkin, and David Held. 2021. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation.

Haohe Liu, Zehua Chen, Yi Yuan, Xinhao Mei, Xubo Liu, Danilo Mandic, Wenwu Wang, and Mark D Plumbley. 2023. Audioldm: Text-to-audio generation with latent diffusion models. *arXiv preprint arXiv:2301.12503*.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024a. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26296–26306.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2024b. Visual instruction tuning. *Advances in neural information processing systems*, 36.

Jens Lundström, W Ourique De Morais, and Martin Cooney. 2015. A holistic smart home demonstrator for anomaly detection and response. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 330–335. IEEE.

Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Eureka: Human-level reward design via coding large language models. *arXiv preprint arXiv:2310.12931*.

Arjun Majumdar, Anurag Ajay, Xiaohan Zhang, Pranav Putta, Sriram Yenamandra, Mikael Henaff, Sneha Silwal, Paul Mcvay, Oleksandr Maksymets, Sergio Arnaud, et al. 2024. Openeqa: Embodied question

answering in the era of foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16488–16498.

Oier Mees, Lukas Hermann, Erick Rosete-Beas, and Wolfram Burgard. 2022. Calvin: A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks.

Luke Melas-Kyriazi, Christian Rupprecht, Iro Laina, and Andrea Vedaldi. 2023. Realfusion: 360 {\deg} reconstruction of any object from a single image. *arXiv preprint arXiv:2302.10663*.

James F. Mullen, Prasoon Goyal, Robinson Piramuthu, Michael Johnston, Dinesh Manocha, and Reza Ghanadan. 2024. "don't forget to put the milk back!" dataset for enabling embodied agents to detect anomalous situations. *IEEE Robotics and Automation Letters*, 9(10):9087–9094.

OpenAI. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. 2022. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*.

Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. 2018. Virtualhome: Simulating household activities via programs.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. 2019. Habitat: A platform for embodied ai research. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9339–9347.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettlemoyer, and Dieter Fox. 2020a. Alfred: A benchmark for interpreting grounded instructions for everyday tasks.

Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettlemoyer, and Dieter Fox. 2020b. Alfred: A benchmark for interpreting grounded instructions for everyday tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10740–10749.

Sanjana Srivastava, Chengshu Li, Michael Lingelbach, Roberto Martín-Martín, Fei Xia, Kent Vainio, Zheng Lian, Cem Gokmen, Shyamal Buch, Karen Liu, Silvio Savarese, Hyowon Gweon, Jiajun Wu, and Li Fei-Fei. 2021. Behavior: Benchmark for everyday household activities in virtual, interactive, and ecological environments. In *Conference in Robot Learning (CoRL)*, page accepted.

Ioan A Sucan, Mark Moll, and Lydia E Kavraki. 2012. The open motion planning library. *IEEE Robotics & Automation Magazine*, 19(4):72–82.

Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. 2017. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 23–30. IEEE.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Lirui Wang, Yiyang Ling, Zhecheng Yuan, Mohit Shridhar, Chen Bao, Yuzhe Qin, Bailin Wang, Huazhe Xu, and Xiaolong Wang. 2023a. Gensim: Generating robotic simulation tasks via large language models. In *Arxiv*.

Yufei Wang, Zhou Xian, Feng Chen, Tsun-Hsuan Wang, Yian Wang, Katerina Fragkiadaki, Zackory Erickson, David Held, and Chuang Gan. 2023b. Robogen: Towards unleashing infinite data for automated robot learning via generative simulation. *ICML*.

Bret C Williams and Jonathan B Kotch. 2023. Child injury and mortality: America's children report. *ChildStats.gov*.

Safe kids Worldwide. 2022. National parent survey on child injury.

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

Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, Li Yi, Angel X. Chang,

Leonidas J. Guibas, and Hao Su. 2020. SAPIEN: A simulated part-based interactive environment. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Aoran Xiao, Jiaxing Huang, Kangcheng Liu, Dayan Guan, Xiaoqin Zhang, and Shijian Lu. 2024. Domain adaptive lidar point cloud segmentation via density-aware self-training. *IEEE Transactions on Intelligent Transportation Systems*, 25(10):13627–13639.

Zhenjia Xu, Zhou Xian, Xingyu Lin, Cheng Chi, Zhiao Huang, Chuang Gan, and Shuran Song. 2023. Roboninja: Learning an adaptive cutting policy for multi-material objects. *arXiv preprint arXiv:2302.11553*.

Yijun Yang, Tianyi Zhou, Kanxue Li, Dapeng Tao, Lusong Li, Li Shen, Xiaodong He, Jing Jiang, and Yuhui Shi. 2024. Embodied multi-modal agent trained by an llm from a parallel textworld. *Preprint*, arXiv:2311.16714.

Sriram Yenamandra, Arun Ramachandran, Karmesh Yadav, Austin Wang, Mukul Khanna, Theophile Gervet, Tsung-Yen Yang, Vidhi Jain, Alexander William Clegg, John Turner, Zsolt Kira, Manolis Savva, Angel Chang, Devendra Singh Chaplot, Dhruv Batra, Roozbeh Mottaghi, Yonatan Bisk, and Chris Paxton. 2023. Homerobot: Open vocabulary mobile manipulation.

Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. 2020. Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on robot learning*, pages 1094–1100. PMLR.

Wenhao Yu, Nimrod Gileadi, Chuyuan Fu, Sean Kirmani, Kuang-Huei Lee, Montse Gonzalez Arenas, Hao-Tien Lewis Chiang, Tom Erez, Leonard Hasenclever, Jan Humplik, et al. 2023. Language to rewards for robotic skill synthesis. *arXiv preprint arXiv:2306.08647*.

Andy Zeng, Pete Florence, Jonathan Tompson, Stefan Welker, Jonathan Chien, Maria Attarian, Travis Armstrong, Ivan Krasin, Dan Duong, Vikas Sindhwani, and Johnny Lee. 2021. Transporter networks: Rearranging the visual world for robotic manipulation.

Ziwen Zhuang, Zipeng Fu, Jianren Wang, Christopher Atkeson, Soeren Schwertfeger, Chelsea Finn, and Hang Zhao. 2023. Robot parkour learning. *arXiv preprint arXiv:2309.05665*.

A 3D Assets Stats

We have compiled statistics regarding the types and quantities of 3D assets extracted from PartNet-Mobility. The details are presented in Table 4. In summary, our subset comprises 44 categories of 3D assets and a total of 2,193 individual 3D assets.

B Brainstorming Setting

B.1 Role List

In this section, we provide a detailed list of roles that are commonly found within a household setting, each accompanied by a specific role description. These roles encompass a variety of responsibilities and skills required to efficiently manage and maintain a home environment. Table 5 outlines the diverse roles ranging from daily household management to specialized services that enhance the functionality and comfort of home life.

B.2 Detail Prompt

We show our prompt template in Figure 6. The prompt outlines a brainstorming session focused on generating home safety tasks for the Franka Panda robotic arm, taking into account the articulated, semantically tagged movable objects in a household setting. These tasks are to be envisioned in scenarios that may pose potential hazards or unsanitary conditions within the home, which the robot is equipped to handle. The tasks are categorized into three primary areas: household hazard, hygiene management, and child safety measures. Each task is to be formatted to include the task name, an explanation, a description, any auxiliary items required, and the articulations and their specific functions. The brainstorming context should be collaborative, with a strong emphasis on the operational limits of the robotic arm, such as avoiding complex assembly, disassembly, or cleaning tasks. This ensures that the tasks are tailored to the robot's capabilities, focusing on practical and manageable interventions in household environment.

C Parameter Setting

C.1 Algorithm of Grasping and Approaching Primitives

In designing a robotic manipulator equipped with a suction cup to facilitate object grasping, the operational primitives for grasping and approaching are outlined as follows: Initially, a random point on the surface of the designated target object or link is selected. Subsequently, a gripper pose is calculated such that it aligns with the normal at the sampled point. Motion planning algorithms are then employed to devise a collision-free trajectory to the predetermined gripper pose. Upon attaining this pose, the manipulator advances along the normal vector until contact is established with the

Prompt for brainstorming

You are {**role**}, your description is {**description**}.
#The role is randomly chosen from the role list.

I will provide an articulated object, with its articulation tree and semantics. Your goal is to imagine some dangerous or unsanitary household anomalies that a robotic arm can address with the articulated object. You can think of the robotic arm as a Franka Panda robot. The scenario will be built in a simulator for the robot to learn it.

Please note that the robot arm has limited capabilities; for example, it only has one arm, making tasks like retracting the blade of a folding knife quite challenging. When setting up tasks to mitigate dangers caused by a folding knife, use simpler methods that are easier for the robot arm to handle, for example, placing the knife in a safe drawer rather than retracting the blade.
Only the articulated object sometimes will have functions, e.g., a microwave can be used to heat food, in these cases, feel free to include other objects that are needed for the task.
Please do not think of tasks that try to assemble or disassemble the object. Do not think of tasks that aim to clean the object or check its functionality.

In brainstorming:
{
[From the previous brainstorming session]
Agent 1: <task proposals 1>
Agent 2: <task proposals 2>
.....
You are doing a group brainstorming now. And the above is the task proposal provied by other agents, pls refer them and propose a new one.
}

The tasks will be categorized into three kinds.You just need to combine the articulated object I provided and choose a task type to write the task.
Household Hazards:This category focuses on tasks that eliminate hazards in the home, for example, put away sharp objects in the home, move objects near the edge of the table back to the center and pick up anything that might trip people.
Hygiene Management:This category focuses on tasks that ensure a clean and hygienic environment in the home, for example, dealing with spoiled food or cleaning, close the open refrigerator door to prevent food from spoiling and pick up the trash on the floor.
Child Safety Measures:This category focuses on tasks that remove dangerous objects or substances that could harm children, for example, put away medications and similar items and keep knives and sharp objects out of children's reach.

Please focus on addressing specific problems that arise in these areas.
Also you are in a group brainstorm with other teammates; as a result, answer as diversely and creatively as you can

For each task you imagined, please simplify this task and write in the following format:
Task name: <the name of the task>
Explanation: <Explain why is the scenario unsafe/unsanitary/not safe for children>
Description: <some basic descriptions of the tasks>
Additional Objects: <Additional objects other than the provided articulated object required for completing the task>
Links: <Links of the articulated objects that are required to perform the task>
- *Link 1*: reasons why this link is needed for the task
- *Link 2*: reasons why this link is needed for the task - …

Joints: <Joints of the articulated objects that are required to perform the task>
- *Joint 1*: reasons why this joint is needed for the task
- *Joint 2*: reasons why this joint is needed for the task - …

Here is some examples:

Example 1:	Example2:	Example 3:
Input: ...	Input: ...	Input:...
Ouput:..	Ouput: ...	Ouput:...

Figure 6: The prompt template of Brainstorming.

Type of item	Number of models	Type of item	Number of models
Bottle	57	Microwave	16
Box	28	Mouse	14
Bucket	36	Oven	30
Camera	37	Pen	48
Cart	61	Phone	18
Chair	81	Pliers	25
Clock	31	Printer	29
CoffeeMachine	54	Refrigerator	44
Dishwasher	48	Remote	49
Dispenser	57	Safe	30
Display	37	Scissors	47
Door	36	Stapler	23
Eyeglasses	65	StorageFurniture	346
Fan	81	Suitcase	24
FoldingChair	26	Table	101
Globe	61	Toaster	25
Kettle	29	Toilet	69
Keyboard	37	TrashCan	70
KitchenPot	25	USB	51
Knife	44	WashingMachine	17
Lamp	45	Window	58
Laptop	55	Lighter	28

Table 4: Detail categories and quantities of subset which select from PartNet-Mobility.

object. In this setup, AnomalyGen leverages LLM to determine the specific target object for either grasping or approaching, dependent on the given subtask. We show the simplified pseudo-algorithm in Algorithm 1.

Algorithm 1 Grasping and Approaching Primitives

Initialize:
  TargetObject: the object or link to be manipulated
  GripperPose: the pose of the robotic gripper
  Procedure:
  Point $p \leftarrow$ RandomSample(TargetObject)
  Vector $n \leftarrow$ NormalAtPoint(TargetObject, p)
  GripperPose $\leftarrow$ AlignWithNormal(GripperPose, n)
  Path path $\leftarrow$ MotionPlanning(GripperPose)
  ExecutePath(path)
  while not ContactMade() **do**
    MoveAlongNormal(GripperPose, n)
  end while
  if ContactMade() **then**
    Grasp(TargetObject)
  end if

C.2 Skill Learning Parameter

For anomaly task learning, we employ SAC algorithm for reinforcement learning. In object manipulation tasks, the observation space includes the low-level state of objects and the robot involved in the task. The SAC utilizes MLP with three layers, each having 256 units, for both the policy and Q networks. We set a learning rate of 3e-4 for the actor, critic, and entropy regularizer. Each manipulation task has a horizon of 100 steps and employs a frameskip of 2. The RL policy controls a 6-dimensional action space, where the first three dimensions dictate the translation—either as delta translation or target location—and the remaining three dimensions specify the delta rotation, represented as a delta-axis angle in the gripper's local frame. We train each sub-task over 1 million environment steps.

For locomotion tasks, we apply the Cross Entropy Method (CEM) for skill learning, which has proven more stable and efficient than traditional RL approaches. We use a ground-truth simulator as the dynamics model in CEM, focusing on optimizing the joint angle values of the robot. The horizon for locomotion tasks is set to 150, with frameskip of 4.

Additionally, we integrate BIT (Gammell et al., 2015), implemented within the Open Motion Planning Library OMPL (Sucan et al., 2012), for action primitives in motion planning. Specifically, for grasping and approaching primitives, we begin by sampling a surface point on the targeted object or link. We then compute a gripper pose that aligns the gripper's y-axis with the normal of the sampled point. The pre-contact gripper pose is established 0.03 meters above the surface point along the normal direction. Utilizing motion planning, we identify a collision-free path to the target gripper pose, continuing the gripper's movement along the normal until contact is achieved.

D Data Statistics

We present the categories and numbers of all generated scenes. Detailed statistics are in Table 6.

Role	Role Description
Homemaker	Responsible for managing household chores and daily life, acting as the heart of the home. Skills include expert cooking, time management, and budget control. The challenge lies in providing the best quality of life on a limited budget.
Engineer	Specializes in designing and maintaining home-use robots such as cleaning robots or elder care robots. Skills in programming, mechanical design, and AI. The challenge is developing robots that integrate seamlessly into the home environment.
Gardener	In charge of designing and maintaining the home garden. Knowledge in botany, creative design, and ecological maintenance. The challenge is to create an aesthetically pleasing yet sustainable outdoor space.
Nutritionist	Provides dietary advice and plans for family members. Expertise in nutrition, food science, and health promotion. The challenge is to balance various dietary restrictions and preferences.
Personal Trainer	Responsible for physical training and health management of family members. Skills in sports science, human physiology, and motivational psychology. The challenge is to create personalized fitness programs that accommodate varying fitness levels.
Financial Planner	Manages family finances, providing investment and savings advice. Knowledge in economics, market analysis, and risk management. The challenge is ensuring financial security and future growth for the family.
Educational Consultant	Supports children in the family with academic guidance and educational planning. Expertise in pedagogy, psychology, and curriculum design. The challenge is to adapt to different learning styles and educational needs.
Home Security Officer	Responsible for family safety and handling emergencies. Skills in security management, emergency response, and physical defense. The challenge is to maintain security without compromising the family's freedom and comfort.
Interior Designer	Optimizes the layout and design of the home to enhance living experience. Skills in artistic design, spatial planning, and color theory. The challenge is to create a functional and beautiful living space within budget.
Household Advisor	Provides comprehensive home management services from daily cleaning to organizing special events. Skills in project management, customer service, and efficiency optimization. The challenge is ensuring all household activities run efficiently and seamlessly.

Table 5: Roles and Descriptions for Household-Based Role Play

Category	Class Name	Number
Household Hazards	Store sharp objects safely	21
	Reduce tripping hazards	32
	Manage chemical risks	1
Hygiene Management	Dispose of spoiled food	5
	Close open refrigerators	9
	Pick up litter from the floor	7
	Ventilate rooms	3
Child Safety Measures	Keep sharp objects away from children	8
	Remove objects that pose choking hazards	10
	Close open large furniture (cabinets, dishwashers, etc.)	12
	Secure household medications	3

Table 6: Household Hazards, Hygiene Management, and Child Safety Measures