A Hands-on Workshop on Designing LLM-Powered Context-Aware Behavior for Companion Robots

Eshtiak Ahmed
Gameful Futures Lab, Research Center of Gameful Realities, Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
eshtiak.ahmed@tuni.fi

Bakhtawar Khan
Gameful Futures Lab, Research Center of Gameful Realities, Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
bakhtawar.khan@tuni.fi

Jiangnan Xu
Gameful Futures Lab, Research Center of Gameful Realities, Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
jiangnan.xu@tuni.fi

Linas Kristupas Gabrielaitis
Gameful Futures Lab, Research Center of Gameful Realities, Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
linas.gabrielaitis@tuni.fi

Juho Hamari
Gamification Group, Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
juho.hamari@tuni.fi

Oğuz 'Oz' Buruk
Gameful Futures Lab, Research Center of Gameful Realities, Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
oguz.buruk@tuni.fi

Abstract

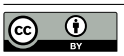
As robots increasingly enter social and domestic environments, their ability to generate context-aware and expressive behaviors becomes essential for meaningful and believable human-robot interaction. Traditional behavior generation often relies on static rules or scripted routines, limiting adaptability and nuance. In this workshop, we present a novel system that leverages Large Language Models (LLMs) to generate dynamic, embodied responses for robots. Our system integrates a defined robot persona, an affordance profile based on physical capabilities, and a response mapping framework that converts conversational inputs into sequences of expressive movement markers. Participants will engage in a hands-on experience with this system, beginning with an introduction to behavior generation in robotics and a live demonstration. Working in small groups, they will design their own LLM prompts and response scenarios for daily life interactions, implement them using the framework, and test the results directly on a quadruped robot, Spot from Boston Dynamics. The workshop drives participants and practitioners to critically reflect on expressive AI and robot interaction design.

CCS Concepts

• **Human-centered computing → HCI design and evaluation methods**.

Keywords

Companion Robots, Robot Behavior, Large Language Models (LLMs)



Mindtrek '25, Tampere, Finland

ACM ISBN 979-8-4007-1512-9/25/10
https://doi.org/10.1145/3757980.3758014

ACM Reference Format:
Eshtiak Ahmed, Bakhtawar Khan, Jiangnan Xu, Linas Kristupas Gabrielaitis, Juho Hamari, and Oğuz 'Oz' Buruk. 2025. A Hands-on Workshop on Designing LLM-Powered Context-Aware Behavior for Companion Robots. In *28th International Academic Mindtrek (Mindtrek '25), October 07–10, 2025, Tampere, Finland*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3757980.3758014

1 Introduction and Background

Robots have become an integral part of modern life, transitioning from industrial applications to more personalized roles in homes, healthcare, education, and entertainment [2]. They are no longer limited to repetitive tasks but are evolving to engage with humans in increasingly sophisticated ways. In the current landscape of robots alongside humans, their importance lies in their ability to assist, complement, and augment human capabilities [11]. However, their ability to interpret and respond to human interventions, engage in meaningful interactions, and adapt to different social contexts has positioned them to potentially become companions to humans [5, 10]. As robots develop more advanced communication and behavioral capabilities, they are not just tools but entities that might have the ability to create meaningful agency with humans in daily life scenarios. One way to create and express agency is to respond contextually to human interaction cues, making the response meaningful to humans [8, 13].

Early approaches to robot response generation, especially in social contexts, often rely on pre-programmed scripts or limited sets of rules, which restrict adaptability and fail to capture the complexity of real-time, dynamic human-robot interactions [4, 9, 12, 15]. Development in natural language processing paved the way for leveraging voice-based interaction with robots, which allowed for the creation of voice-based feedback mechanisms on robots [18]. However, the process stayed far from dynamic due to the shortcomings of natural language generation from the robot's perspective.

The introduction of Large Language Models (LLMs) addresses a lot of these shortcomings, especially because it provides a more dynamic, while still being controllable and predictable, depending on how it is configured [1]. Also, Artificial Intelligence (AI), being the driving force of LLMs, makes them more adaptive and situated in diverse contexts.

Recent advances in Large Language Models (LLMs), such as GPT-3 and GPT-4 [1], have opened new possibilities for robotic interactions. LLMs have demonstrated strong capabilities in contextual understanding, dialogue generation, and zero-shot reasoning, making them promising tools for generating robot responses from natural language input [3, 14]. In robotics, LLMs have been used to interpret instructions, plan high-level actions, and even generate code to control robotic systems [19, 21]. For example, the Code-as-Policies (CaP) framework leverages language models to output executable robot code from user commands, allowing for flexible and composable robot behavior [16]. More recently, LLMs have also been explored for embodied language understanding, where the model integrates sensory data and physical actions with conversational input [20].

Keeping all these factors in mind, we have developed a system that enables a quadruped robot, Boston Dynamics' Spot [7], to respond to human voice inputs with expressive, dog-like physical behaviors. By integrating voice recognition, large language models (LLMs), and a structured response mapping framework, the robot interprets conversational inputs and generates sequences of behavior markers aligned with its physical capabilities. Through this proposed workshop, we aim to introduce participants to this novel framework, encourage hands-on experimentation with designing LLM prompt structures and embodied responses, and facilitate creative exploration of how language-based interactions can be transformed into meaningful robotic behaviors. Participants will engage in collaborative design exercises, contribute their own interpretations of daily life scenarios with robots, and test the expressive capacity of the robot using their own custom LLM prompts. Through this, we seek to gain insights into the challenges and opportunities of designing context-aware, embodied interactions with LLM-driven robots.

2 AwaR(e)obot: LLM-Powered System for Behavioral Response Generation

The system comprises three main components: Human Interface, LLM Module, and Robot Execution Layer. Fig. 1 shows a high-level design of the system.

2.1 Human Interface

The interaction begins with the human user providing voice input, which serves as the primary modality for initiating communication with the system. This input can vary widely in form and intent, ranging from explicit commands, such as "follow me", "sit down", or "walk ahead", to more expressive and emotionally rich dialogue, like "I feel tired today" or "this place feels a bit eerie". By accommodating this full spectrum of utterances, the system moves beyond traditional command-and-control paradigms and instead embraces a more conversational model of interaction. Supporting both directive and expressive speech opens up avenues for more nuanced human-robot relationships. Once the voice is detected, it is transcribed into text using a speech-to-text module and immediately fed to the LLM module. The length of the voice input can be set either to 1) listen until the human keeps talking, allowing longer and more complex dialogs, or 2) listen for a finite amount of time, allowing for shorter interactions.

2.2 LLM Module

The core of the system is the Large Language Model (LLM), which acts as a bridge between the natural language and the robot execution layer. The primary aim of this module is to take the voice input as text, create a custom LLM prompt including this voice input, and feed it to LLM, and then generate response markers as output. It also provides detailed reasoning for the response markers it produces.

2.2.1 Input Specifications. The system's input to the Large Language Model (LLM) consists of three structured components: the robot's persona, its affordance profile, and a response mapping framework. The robot persona serves as a foundational marker that guides the tone, style, and intent of the robot's responses, shaped by its type and appearance (e.g., quadruped vs. humanoid). The affordance profile provides a clearly defined list of the robot's physical capabilities and limitations, ensuring the LLM generates only realistic and executable behaviors grounded in the robot's embodied form. Finally, the response mapping framework includes a set of movement markers, generation rules, and output formatting constraints that ensure coherence and believability.

2.2.2 Output Generation. The LLM produces two complementary outputs: a sequence of behavioral markers and a short reasoning text. The behavioral markers are structured codes representing the robot's non-verbal physical actions, formatted as timed movement instructions (e.g., "f,1.0" to indicate movement 'f' for 1 second). These markers are drawn from the predefined vocabulary in the response mapping framework and translate directly into parameterized movement routines in the robot's behavior engine. Accompanying this, the reasoning text provides a natural language explanation of the LLM's logic, interpreting the user's tone and speech content, and justifying the selection of specific behaviors.

2.3 Robot Execution Layer

The Robot Execution Layer serves as the final stage in the system pipeline and is responsible for translating the high-level response markers generated by the LLM into tangible, physical movements performed by the robot. Once a sequence of behavioral markers is produced as a python list, they are processed sequentially, using an iterative loop to traverse each marker in the list. As each marker is encountered, it is passed through a mapping function that links it to a corresponding routine in the robot's movement library. These routines consist of predefined movement primitives such as "walk forward", "move backward", "go belly up", or "strafe left", etc., that are natively supported by the robot's hardware capabilities. This marker-to-movement translation operates via a lookup mechanism in a python method library that ensures every behavioral marker is matched to a physically feasible and safe motion routine.

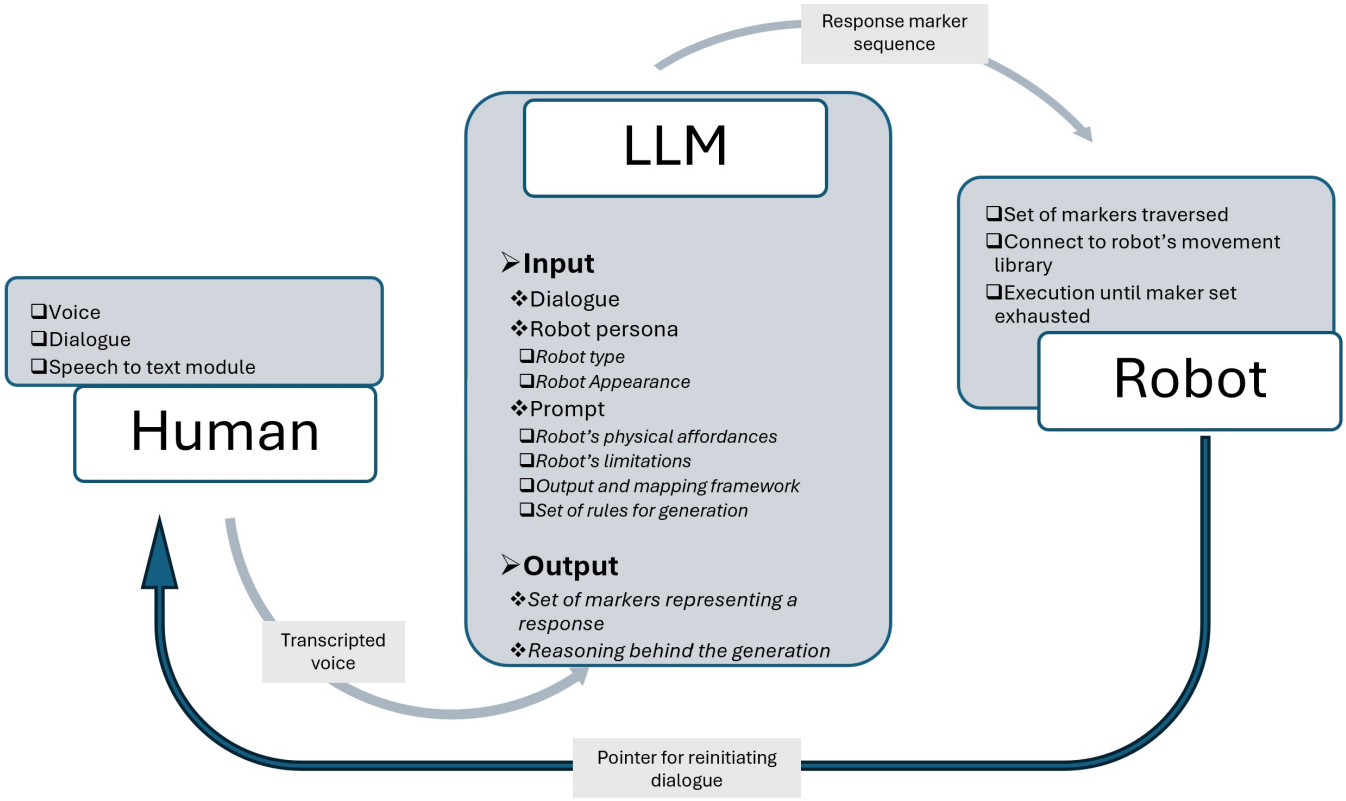


Figure 1: System design including components.

2.4 Interaction Loop

Once the robot completes executing its full sequence of behavioral responses mapped from the user's prior voice input, it transitions back into a receptive state. To signal this shift and ask the user for the next input, the robot emits a distinct sound marker, such as a short chime or mechanical tone, signifying that the robot is now actively listening and ready for further interaction.

2.4.1 System Generalizability. One of the deeper implications of this work lies in its potential to serve as a generalized framework for robot behavior generation. The modular design, consisting of a voice-to-text layer, LLM-based semantic interpretation, behavior marker generation, and robot execution mapping, can be adapted across robotic platforms with different hardware specifications. By abstracting robot actions into semantic markers (e.g., "greet formally," "look playful") before mapping them to platform-specific movement primitives, the framework offers a flexible pipeline that separates social logic from physical implementation. This abstraction enables scalability and portability, provided that each target robot has an associated library of motion mappings that correspond to its specific embodiment. Robots like Spot, a quadruped with no facial expressions, speech, or arms, rely solely on locomotion and posture, whereas humanoid robots might convey intent through gestures, gaze, or spoken language [6, 17]. This disparity requires tailored prompt structures, marker vocabularies, and control logic for each robot type.

3 Workshop Objectives and Structure

The primary goal of this workshop is to offer participants hands-on experience in designing expressive and context-aware robotic behaviors using large language models (LLMs), by leveraging the AwaR(e)obot system.

3.1 Key Themes and Objectives

The workshop and its activities will touch upon and go deeper into the following key themes.

- ***Embodied Interaction Design:*** Exploring how physical robot behavior can be designed to match contextual and meaningful cues from human input.
- ***LLMs as Behavior Generators:*** Understanding the role of large language models in generating not only text but also structured semantic interpretations that can map to physical actions.
- ***Robot Expressivity and Movement Mapping:*** Investigating how robot affordances such as posture, locomotion, and motion sequences can be utilized creatively to simulate expressivity, especially in robots lacking humanoid features.
- ***Prompt Engineering for Robots:*** Training participants to craft effective LLM prompts that encode context, scenario, character, and interaction goals, enabling more nuanced robot behavior generation.

- ***Human-Centered Robotics:*** Emphasizing the importance of user experience, empathy, and perceived agency when designing robots that interact with humans in shared environments.

3.2 Workshop Structure

This full-day workshop is designed to introduce participants to a novel framework for generating expressive robot behaviors using large language models (LLMs). Through a combination of brainstorming, hands-on activities, and group-based explorations, participants will learn how to craft context-aware robotic expressions, develop their own interaction scenarios, and evaluate each other's creations. The workshop day is structured to balance learning, creativity, and collaborative design activities. The workshop will follow the following overall structure:

- ***Introduction to the Workshop (09:30–10:15):*** The day begins with an overview of the workshop's aims, context, and structure. We introduce participants to the broader themes of expressive robotics, embodied interaction, and the role of LLMs in enabling context-aware behaviors in robots.
- ***Icebreaker and Getting to Know Each Other (10:15–10:45):*** To build a collaborative and open atmosphere, we engage participants in interactive icebreaker activities. These will help attendees get acquainted, understand each other's backgrounds, and feel comfortable exchanging ideas throughout the day.
- ***Coffee Break (10:45–11:00):*** A short break with refreshments.
- ***System Walkthrough and Demo (11:00–12:00):*** Participants are introduced to the key components of the framework: robot persona, affordance mapping, LLM prompt structure, behavioral marker system, and execution pipeline. A live demo with the Boston Dynamics Spot robot will show how the system responds to voice inputs and translates them into expressive physical behavior.
- ***Lunch Break (12:00–13:00)***
- ***Hands-On Group Activity for Designing Robot Expressions (13:00–15:30):*** Participants are divided into small groups and guided through the process of creating their own expressive robot interactions. Each group will:
  - Choose a real-world daily interaction scenario (e.g., reacting to compliments, expressing curiosity, etc.)
  - Define a robot persona and map its affordances
  - Design a LLM prompt structure and generate 4-5 expressive behavior sequences using the framework
- ***Coffee Break (15:30–15:45):*** A short break with refreshments.
- ***Presentation and Group Evaluations (15:45–16:30):*** Each group presents their scenario, LLM prompt, and resulting robot behavior. Live demonstrations are run, and peer evaluation is conducted based on expressiveness, contextual relevance, and social readability of the robot's responses.
- ***Closing Discussion and Reflection (16:30–17:00):*** We wrap up with a group reflection, discussing insights, challenges, and ideas for further development. Participants will be invited to contribute to future research directions or adapt the framework for their own creative or academic projects.

4 Intended Outcomes

This workshop is designed to yield both educational and research-related outcomes. On the educational side, participants will leave with a practical understanding of how to design and prototype embodied robot behaviors using LLMs. On the research side, we aim to generate a rich collection of interaction scenarios, LLM prompt templates, and robot behavior sequences that provide insights into how humans interpret and design robotic expressivity. On top of this, we aim to gather valuable feedback from the participants on the overall usability of the system, including what worked well, what could be improved, and how the system can be extended for more diverse use in different contexts. The following specific outcomes are anticipated:

- A portfolio of user-designed interaction scenarios with corresponding LLM prompts and movement mappings.
- Recorded demonstrations of robot behaviors responding to user-generated LLM prompts.
- Qualitative feedback from participants on the design process, ease of use, and creative potential of the system.
- Reflections and peer evaluations that help surface design patterns, misunderstandings, or opportunities in the framework.

Additionally, we will also gather observational data from facilitators to understand how users engage with the system, how they iterate designs, and where intervention or guidance is most frequently required. This workshop is part of an ongoing research project that explores how generative AI tools can enhance human-robot interaction by enabling robots to behave in socially intelligible and emotionally resonant ways. Moving forward, we aim to develop an open-source toolkit based on the framework introduced in this workshop, including LLM prompt libraries, behavior templates, and customization tools. Additionally, we will use the collected data to publish a peer-reviewed research paper on the participatory design of robot behavior using LLMs.

5 Call for Participants

We invite researchers, designers, and creative practitioners interested in human-robot interaction, embodied AI, and expressive robotics to participate in our hands-on workshop. This session will explore how large language models (LLMs) can be used to generate dynamic, context-aware robot behaviors through an interactive design framework. Participants will work in small teams to design their own LLM prompts and movement sequences for a quadruped robot, experience the results in real-time, and engage in peer feedback and reflection. The workshop is open to participants from various backgrounds, including HRI, interaction design, robotics, AI, and digital media. No extensive technical expertise is required, but familiarity with basic interaction or AI concepts will be helpful. We aim to create an inclusive, collaborative environment and welcome up to 15 participants to ensure meaningful engagement and feedback.

6 Organizers

Eshtiak Ahmed is a doctoral researcher at Tampere University, Finland. His current research focuses on exploring and leveraging AI to enhance and rationalize human-robot interaction (HRI) and companionship (HRC) in daily life scenarios.

Bakhtawar Khan is a doctoral researcher at Tampere University, exploring humans, technology, and more-than-human ecologies within the Finnish Flagship UNITE Programme. Her research blends AI, XR/VR, and other platforms to investigate how technology-mediated nature experiences can reshape the future and well-being.

Jiangnan Xu is a researcher in Human-AI interaction in the Gamification Group, with interests in natural language processing, multi-agent systems, and AI-assisted creativity. Her recent work focuses on leveraging large language models to support co-creative role-playing game (RPG) development, integrating narrative generation with agent-based reasoning, and explores how AI can enable collaborative storytelling and emergent gameplay in open-ended environments.

Linas Kristupas Gabrielaitis is a doctoral researcher at the Gamification Group, Tampere University, Finland. His research primarily focuses on game-playing and diagram-mapping as fictospeculative practices for more-than-human design.

Juho Hamari is a Professor of Gamification at the Faculty of Information Technology and Communications, Tampere University, leading the Gamification Group. Dr. Hamari's and his research group's (GG) research covers several forms of information technologies, such as games, motivational information systems, new media, peer-to-peer economies, and virtual economies. He is among the most cited scholars in the world, publishing in a number of prestigious venues.

Oğuz 'Oz' Buruk is an Assistant Professor of Gameful Experience at Tampere University, Finland. His research focuses on designing gameful environments for various contexts such as body integrated technologies, computational fashion, posthumanism, urban spaces, extended reality, and nature. He frequently employs methods such as speculative design, design fiction, and participatory design.

Acknowledgments

This work was supported by the Research Council of Finland's Flagship Programme UNITE (decision 357906).

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Eshtiak Ahmed, Oğuz 'Oz' Buruk, and Juho Hamari. 2024. Human–robot companionship: Current trends and future agenda. *International Journal of Social Robotics* 16, 8 (2024), 1809–1860.

[3] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil J Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy Zeng. 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. arXiv:2204.01691 [cs.RO] https://arxiv.org/abs/2204.01691

[4] Fernando Alonso-Martín and Miguel A Salichs. 2011. Integration of a voice recognition system in a social robot. *Cybernetics and Systems: An International Journal* 42, 4 (2011), 215–245.

[5] Cynthia Breazeal. 2003. Toward sociable robots. *Robotics and autonomous systems* 42, 3-4 (2003), 167–175.

[6] Lukáš Danev, Marten Hamann, Nicolas Fricke, Tobias Hollarek, and Dennys Paillacho. 2017. Development of animated facial expressions to express emotions in a robot: RobotIcon. In *2017 ieee second ecuador technical chapters meeting (etcm)*. IEEE, 1–6.

[7] Boston Dynamics. 2025. Spot - The Agile Mobile Robot. https://bostondynamics.com/products/spot/ Accessed: 2025-06-04.

[8] O Engwall, RJP Bandera, S Bensch, KS Haring, T Kanda, P Núnez, M Rehm, and A Sgorbissa. 2023. Editorial: Socially, culturally and contextually aware robots. *Frontiers in Robotics and AI* 10 (2023).

[9] Pasquale Foggia, Antonio Greco, Antonio Roberto, Alessia Saggese, and Mario Vento. 2023. A social robot architecture for personalized real-time human–robot interaction. *IEEE Internet of Things Journal* 10, 24 (2023), 22427–22439.

[10] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and autonomous systems* 42, 3-4 (2003), 143–166.

[11] Michael A Goodrich, Alan C Schultz, et al. 2008. Human–robot interaction: a survey. *Foundations and trends® in human–computer interaction* 1, 3 (2008), 203–275.

[12] Nishu Gupta et al. 2021. A novel voice controlled robotic vehicle for smart city applications. In *Journal of Physics: Conference Series*, Vol. 1817. IOP Publishing, 012016.

[13] Katherine Harrison, Giulia Perugia, Filipa Correia, Kavyaa Somasundaram, Sanne Van Waveren, Ana Paiva, and Amy Loutfi. 2023. The imperfectly relatable robot: An interdisciplinary workshop on the role of failure in hri. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 917–919.

[14] 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* (2022).

[15] Takayuki Kanda, Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, and Ryohei Nakatsu. 2002. Development and evaluation of an interactive humanoid robot" Robovie". In *Proceedings 2002 IEEE international conference on robotics and automation (Cat. No. 02CH37292)*, Vol. 2. IEEE, 1848–1855.

[16] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. 2022. Code as policies: Language model programs for embodied control. *arXiv preprint arXiv:2209.07753* (2022).

[17] Vimitha Manohar and Jacob W Crandall. 2014. Programming robots to express emotions: interaction paradigms, communication modalities, and context. *IEEE transactions on human-machine systems* 44, 3 (2014), 362–373.

[18] Nikolaos Mavridis. 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems* 63 (2015), 22–35.

[19] Yao Mu, Junting Chen, Qinglong Zhang, Shoufa Chen, Qiaojun Yu, Chongjian Ge, Runjian Chen, Zhixuan Liang, Mengkang Hu, Chaofan Tao, et al. 2024. Robocodex: Multimodal code generation for robotic behavior synthesis. *arXiv preprint arXiv:2402.16117* (2024).

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

[21] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2023. ProgPrompt: program generation for situated robot task planning using large language models. *Autonomous Robots* 47, 8 (2023), 999–1012.