An Overview of Dobby, A Conversational Service Robot

Carson Stark, Bohkyung Chun, Casey Charleston, Varsha Ravi, Luis Pabon, Surya Sunkari, Tarun Mohan, Peter Stone, and Justin Hart
{carsonstark,boh,caseycharleston,vravi,luisalepabon,suryasunkari,tarun.mohan}@utexas.edu
{pstone,hart}cs@utexas.edu
The University of Texas at Austin
Austin, Texas, USA

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**.

KEYWORDS

Human-Robot Interaction, Large Language Models, Robot Tour Guides

ACM Reference Format:
Carson Stark, Bohkyung Chun, Casey Charleston, Varsha Ravi, Luis Pabon, Surya Sunkari, Tarun Mohan, Peter Stone, and Justin Hart. 2024. An Overview of Dobby, A Conversational Service Robot. In *Proceedings of March 11, 2024 (The HRI 2024 Workshop on Human – Large Language Model Interaction).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

We introduce Dobby, an agent and architecture built around the GPT-4 large language model (LLM). The system leverages the LLM for both the generation of dialogue and task planning. The system is demonstrated in Human-Robot Interaction (HRI) study constructed around a tour-guide scenario in which study participants take personalized tours stopping at various landmarks in a shared space featuring multiple laboratories. Performance is measured along five dimensions: overall effectiveness, exploration abilities, scrutinization abilities, receptiveness to personification, and adaptability. This abstract is an abbreviated version of our arXiv paper on this system [4].

2 THE DOBBY ARCHITECTURE

The system's prompt instructs it to behave as a robot assistant, and includes context about its environment, background information, and a list of actions that the robot can perform. LLM queries are made using OpenAI's chat completion API. The function calling feature of the ChatGPT model is used to perform actions. Figure 1 shows a system diagram.

The HRI 2024 Workshop on Human – Large Language Model Interaction, Boulder, CO,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2.1 Function Calling

The model gpt-4-0613 introduced function calling, in which the model generates a JSON object containing a function call as part of the completion. The JSON objects can be parsed to execute external commands. To facilitate this, OpenAI accepts a structured description of available functions with every query to their API. The Dobby architecture defines the functions *ExecutePlan(string[] actionSequence)* and *CancelPlan()* for general use cases. Whe the agent "chooses" an action, one of these function calls is included in the output of the LLM.

2.2 Conversation

In the "Conversing" state, the system enters a loop where it records the user's utterance, transcribes the recorded audio, queries the agent for a response, plays the dialogue to the user, and begins recording again. Input text, system messages, and generated responses are accumulated in a history buffer which is sent to the API at every iteration. LLMs provide unique capabilities. The robot can pose clarifying questions, offer suggestions, and adapt to each unique individual, providing the robot with the opportunity to adjust to the user's intentions and desires before taking any action. System messages are included in the history buffer to provide event-based instructions or update the agent on the state of the environment, preventing the robot's dialogue from contradicting its behavior. If silence is detected for six seconds and no response is received, the robot will begin listening for the keyword "Dobby" to re-trigger the conversation loop.

2.3 Action Planning

Atomic actions include a textual title, pre/post-conditions, and an executable function. Each action's title is listed in the prompt. When queried, the agent may choose to begin a series of actions by calling the function *ExecutePlan(string[] actionSequence).* To ensure robustness to semantically similar commands, each string is matched to an action class by comparing the embedding of the output to each action title and selecting the action with the highest similarity. Occasionally, the agent will attempt to include actions not listed in the prompt which have no corresponding action class. To correct this, the agent is re-prompted with an error message if the maximum embedding similarity falls under a certain threshold. After repeated attempts, a system message informs the agent that it is not capable of the requested task, prompting it to explain this to the user.

Once parsed, steps are taken to assure plan validity. To model the environment, the system uses pre-conditions and post-conditions attached to each action; similar to additions and deletions in STRIPS [1] style planning, or the tracking of predicates in PDDL [3]. The

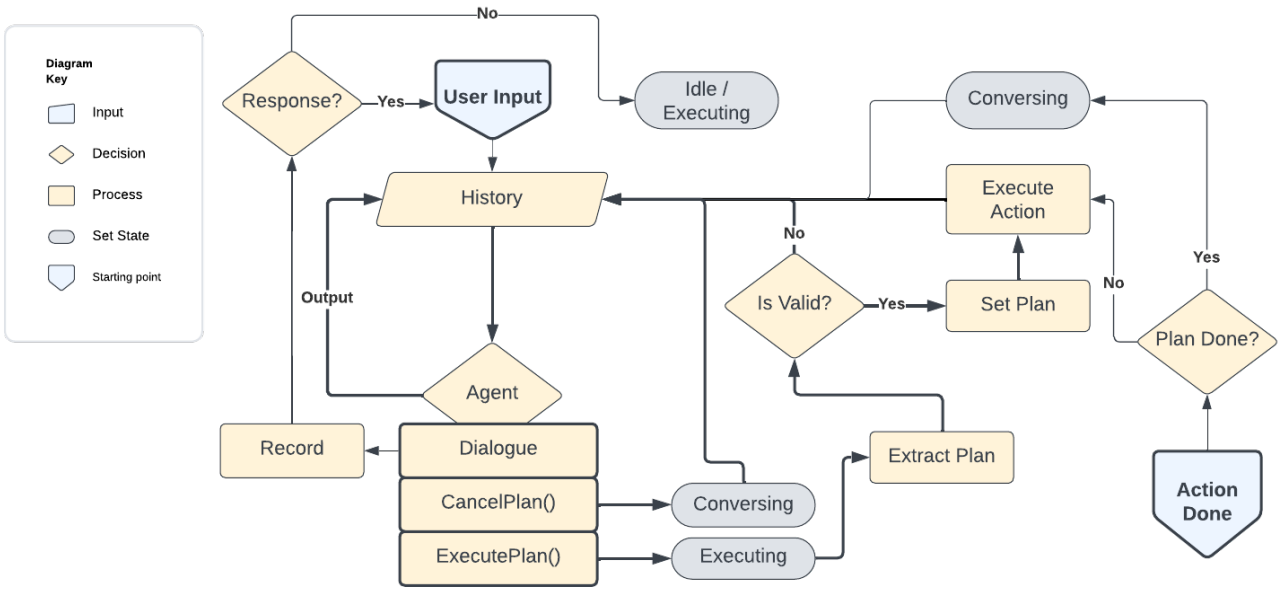

Figure 1: Diagram detailing the flow of information through the Dobby architecture, with emphasis on the inputs and outputs of the agent and transitions between states. All processes connected to History add a system message to provide context and then re-prompt the agent. Most function calls re-prompt the agent to generate dialogue before recording again.

system uses a greedy algorithm which skips actions until their necessary preconditions have been met, attempting to reorder the plan if necessary. If the generated plan cannot be corrected, the agent informs the user that it is incapable of the task.

2.4 Action Execution

Once a plan has been proposed and validated, actions are executed in order. When an action begins, system messages inform the agent that the previous action has completed and that the new action has begun. The agent is then re-prompted to provide a dialogue cue informing the user of its intended behavior. Actions do not block the system when executing, so it is possible to continue to converse with the robot while it is performing a task such as navigation. A function *CancelPlan()* may be called by the agent to halt the execution of the current plan at the user's request. Alternatively, the agent may start a new plan, overriding the previous one. When an action completes, the conversation loop is interrupted and the next action is started, prompting a corresponding dialogue line.

3 EVALUATION

To evaluate Dobby, we designed an experiment to contrast participants' experiences with a conversational vs. non-conversational robot tour guide; hypothesizing that the conversational version would would be more effective due to its ability to contextualize the user's requests, suggest destinations based on their interests, answer a wide variety of questions, and keep the user engaged with back and forth conversation. The study focuses on investigating our system's advantages in HRI instead of the planning domain.

We recorded the coordinates of ten notable destinations within the laboratory, along with a brief description of each. The coordinates are used to generate a "go to" action for each destination. The descriptions are included in Dobby's prompt along with information about five general topics to provide context about the lab. Dobby is built on top of an existing robot platform called the BWIBot [2] which is used in these experiments.

The non-conversational systtem is intended to represent the best system possible without a modern LLM. The robot's dialogue is scripted and interaction is limited to a fixed set of spoken commands: "Show me the (landmark)." and "Tell me about (topic)." When this robot arrives at a destination or is requested to provide information, it reads aloud descriptions of the landmark or topic verbatim.

We completed a study including 22 participants. Each trial consisted of one tour with the conversational robot and one tour with the non-conversational robot; in that order. Prior to participation, each participant provided informed consent. This study was approved by the University of Texas at Austin's Institutional Review Board. On-boarding instructions were provided to each participant explaining how to interact with the robots. Each tour ended when a participant expressed their willingness to end their tour.

Study participants rated the conversational robot substantially better than the non-conversational robot. They spent an average of 14.3 minutes with the conversational robot and only 5.8 with the non-coversational robot, and gave each an enjoyment rating of 6.59 vs 4.00 on a 7-point scale, respectively. A detailed description of the system and a full description of the study and results can be found in our arXiv paper [4].

ACKNOWLEDGMENTS

This work has taken place in the Learning Agents Research Group (LARG) and the Living with Robots Laboratory (LWR) at UT Austin. LARG research is supported in part by NSF (FAIN-2019844, NRT-2125858), ONR (N00014-18-2243), ARO (E2061621), Bosch, Lockheed Martin, Cisco Research, Army Futures Command, and UT Austin's Good Systems grand challenge. LWR research is supported in part by NSF (NRT-2125858 and GCR-2219236), Cisco Research, and Army Futures Command. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

REFERENCES

[1] Richard E. Fikes and Nils J. Nilsson. 1971. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2, 3 (1971), 189–208. https://doi.org/10.1016/0004-3702(71)90010-5

[2] Piyush Khandelwal, Shiqi Zhang, Jivko Sinapov, Matteo Leonetti, Jesse Thomason, Fangkai Yang, Ilaria Gori, Maxwell Svetlik, Priyanka Khante, Vladimir Lifschitz, J. K. Aggarwal, Raymond Mooney, and Peter Stone. 2017. BWIBots: A platform for bridging the gap between AI and human–robot interaction research. *The International Journal of Robotics Research* (2017). http://www.cs.utexas.edu/users/ai-lab?khandelwal:ijrr17

[3] Drew McDermott, Malik Ghallab, Adele E. Howe, Craig A. Knoblock, Ashwin Ram, Manuela M. Veloso, Daniel S. Weld, and David E. Wilkins. 1998. PDDL-the planning domain definition language.

[4] Carson Stark, Bohkyung Chun, Casey Charleston, Varsha Ravi, Luis Pabon, Surya Sunkari, Tarun Mohan, Peter Stone, and Justin Hart. 2023. Dobby: A Conversational Service Robot Driven by GPT-4. arXiv:2310.06303 [cs.RO]

Received 20 February 2007