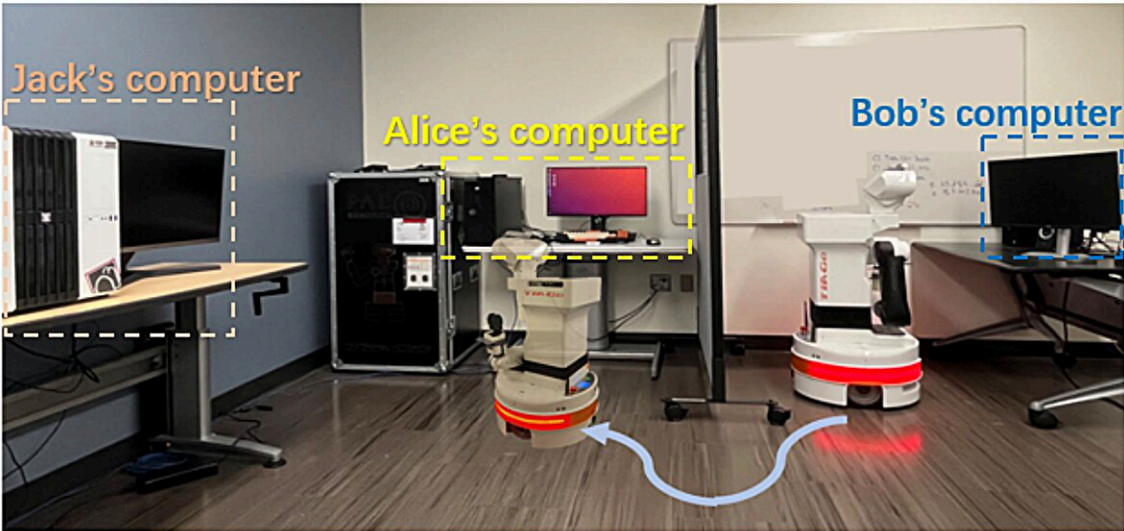
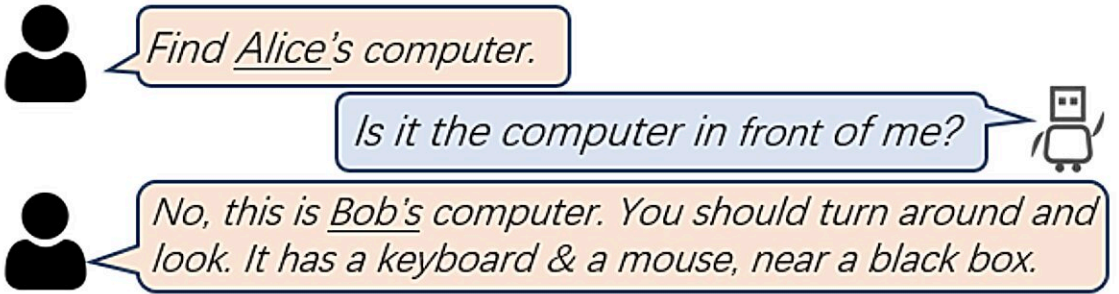


Using large language models to enable open-world, interactive and personalized robot navigation

October 27 2023, by Ingrid Fadelli



An example of zero-shot interactive personalized navigation. There are three computers in the room never seen by the robot before. The goal is to find Alice's computer. The robot starts by finding the wrong object and needs to communicate with the user and leverage the user feedback to locate the personalized goal. Credit: Dai et al.

Robots should ideally interact with users and objects in their surroundings in flexible ways, rather than always sticking to the same sets of responses and actions. A robotics approach aimed towards this goal that recently gained significant research attention is zero-shot object navigation (ZSON).

ZSON entails the development of advanced computational techniques that allow robotic agents to navigate unknown environments interacting with previously unseen objects and responding to a wide range of prompts. While some of these techniques achieved promising results, they often only allow robots to locate generic classes of objects, rather than using natural language processing to understand a user's prompt and locate specific objects.

A team of researchers at University of Michigan recently set out to develop a new approach that would enhance the ability of robots to explore open-world environments and navigate them in personalized ways. Their proposed framework, introduced in a paper [published](#) on *arXiv* preprint server, uses [large language models](#) (LLMs) to allow robots to better respond to requests made by users, for instance locating specific nearby objects.

"The existing works of ZSON mainly focus on following individual instructions to find generic object classes, neglecting the utilization of natural language interaction and the complexities of identifying user-specific objects," Yinpei Dai, Run Peng and their colleagues wrote in their paper. "To address these limitations, we introduce Zero-shot Interactive Personalized Object Navigation (ZIPON), where robots need to navigate to personalized goal objects while engaging in conversations with users."

In their paper, Dai, Peng and their collaborators firstly introduce a new task, which they dub ZIPON. This task is a generalized form of ZSON, that entails accurately responding to personalized prompts and locating specific target objects.

If traditional ZSON entails locating a nearby bed or chair, ZIPON takes this one step further, asking a robot to identify a specific person's bed, a chair bought from Amazon, and so on. The researchers subsequently tried to develop a [computational framework](#) that would effectively solve this ask.

"To solve ZIPON, we propose a new framework termed Open-world Interactive persOnalized Navigation (ORION), which uses Large Language Models (LLMs) to make sequential decisions to manipulate different modules for perception, navigation and communication," Dai, Peng and their colleagues wrote in their paper.

The new framework developed by this team of researchers has six key modules: a control, a semantic map, an open-vocabulary detection, an exploration, a memory, and an interaction module. The control module allows the robot to move around in its surroundings, the semantic map module indexes [natural language](#), and the open-vocabulary detection module allows the robot to detect objects based on language-based descriptions.

Robots then search for objects in their surrounding environment using the exploration module, while storing important information and feedback received from users in the memory module. Finally, the interaction [module](#) allows robots to speak with users, verbally responding to their requests.

Dai, Peng and their colleagues evaluated their proposed framework both in simulations and real-world experiments, using TIAGo, a mobile

wheeled robot with two arms. Their findings were promising, as their framework successfully improved the ability of the robot to utilize user feedback when trying to locate specific nearby objects.

"Experimental results show that the performance of interactive agents that can leverage user feedback exhibits significant improvement," Dai, Peng and their colleagues explained. "However, obtaining a good balance between task completion and the efficiency of navigation and interaction remains challenging for all methods. We further provide more findings on the impact of diverse user feedback forms on the agents' performance."

While the ORION framework shows potential for improving personalized [robot](#) navigation of unknown environments, the team found simultaneously ensuring that robots complete missions, smoothly navigate unknown environments and interact well with users extremely challenging. In the future, this study could inform the development of new models for completing the ZIPON task, which could address some of the reported shortcomings of the team's proposed [framework](#).

"This work is only our initial step in exploring LLMs in personalized navigation and has several limitations," Dai, Peng and their colleagues wrote in their paper. "For example, it does not handle broader goal types, such as image goals, or address multi-modal interactions with users in the real world. Our future efforts will expand on these dimensions to advance the adaptability and versatility of interactive robots in the human world."

More information: Yinpei Dai et al, Think, Act, and Ask: Open-World Interactive Personalized Robot Navigation, *arXiv* (2023). [DOI: 10.48550/arxiv.2310.07968](https://doi.org/10.48550/arxiv.2310.07968). arxiv.org/abs/2310.07968

© 2023 Science X Network

Citation: Using large language models to enable open-world, interactive and personalized robot navigation (2023, October 27) retrieved 27 April 2024 from

<https://techxplore.com/news/2023-10-large-language-enable-open-world-interactive.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.