

How do you make a robot smarter? Program it to know what it doesn't know

November 29 2023, by Molly Sharlach



Engineers at Princeton University and Google have come up with a new way to teach robots to know when they don't know and ask for clarification from a human. Credit: Allen Ren et al./Princeton University

Modern robots know how to sense their environment and respond to language, but what they don't know is often more important than what they do know. Teaching robots to ask for help is key to making them safer and more efficient.

Engineers at Princeton University and Google have come up with a new way to teach robots to know when they don't know. The technique involves quantifying the fuzziness of human language and using that measurement to tell robots when to ask for further directions. Telling a [robot](#) to pick up a bowl from a table with only one bowl is fairly clear. But telling a robot to pick up a bowl when there are five bowls on the table generates a much higher degree of uncertainty—and triggers the robot to ask for clarification.

The paper, "[Robots That Ask for Help: Uncertainty Alignment for Large Language Model Planners](#)," was presented Nov. 8 at the Conference on Robot Learning.

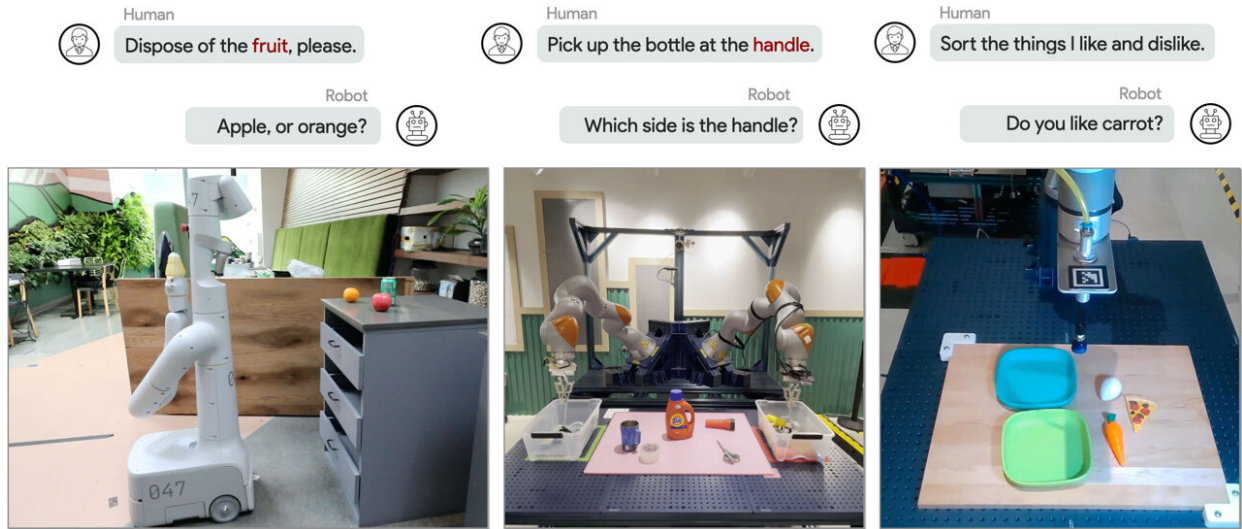
Because tasks are typically more complex than a simple "pick up a bowl" command, the engineers use [large language models](#) (LLMs)—the technology behind tools such as ChatGPT—to gauge uncertainty in complex environments. LLMs are bringing robots powerful capabilities to follow [human language](#), but LLM outputs are still frequently unreliable, said Anirudha Majumdar, an assistant professor of mechanical and aerospace engineering at Princeton and the senior author of a [study](#) outlining the new method.

"Blindly following plans generated by an LLM could cause robots to act in an unsafe or untrustworthy manner, and so we need our LLM-based robots to know when they don't know," said Majumdar.

The system also allows a robot's user to set a target degree of success, which is tied to a particular uncertainty threshold that will lead a robot to ask for help. For example, a user would set a surgical robot to have a much lower error tolerance than a robot that's cleaning up a living room.

"We want the robot to ask for enough help such that we reach the level of success that the user wants. But meanwhile, we want to minimize the

overall amount of help that the robot needs," said Allen Ren, a graduate student in mechanical and aerospace engineering at Princeton and the study's lead author. Ren received a best student paper award for his Nov. 8 presentation at the Conference on Robot Learning in Atlanta. The new method produces high accuracy while reducing the amount of help required by a robot compared to other methods of tackling this issue.



The researchers tested their method on a simulated robotic arm and on two types of robot hardware. The most complex experiments (left) involved a robotic arm mounted on a wheeled platform and placed in an office kitchen with a microwave and a set of recycling, compost and trash bins. Another set of experiments used a tabletop robotic arm (right) tasked with sorting a set of toy food items into two different categories; a setup with a left and right arm (center) added an additional layer of ambiguity. Credit: Allen Ren et al./Princeton University

The researchers tested their method on a simulated [robotic arm](#) and on two types of robots at Google facilities in New York City and Mountain View, California, where Ren was working as a student research intern. One set of hardware experiments used a tabletop robotic arm tasked with

sorting a set of toy food items into two different categories; a setup with a left and right arm added an additional layer of ambiguity.

The most complex experiments involved a robotic arm mounted on a wheeled platform and placed in an office kitchen with a microwave and a set of recycling, compost and trash bins. In one example, a human asks the robot to "place the bowl in the microwave," but there are two bowls on the counter—a metal one and a plastic one.

The robot's LLM-based planner generates four possible actions to carry out based on this instruction, like multiple-choice answers, and each option is assigned a probability. Using a statistical approach called conformal prediction and a user-specified guaranteed [success rate](#), the researchers designed their algorithm to trigger a request for human help when the options meet a certain probability threshold. In this case, the top two options—place the plastic bowl in the microwave or place the metal bowl in the microwave—meet this threshold, and the robot asks the human which bowl to place in the microwave.

In another example, a person tells the robot, "There is an apple and a dirty sponge ... It is rotten. Can you dispose of it?" This does not trigger a question from the robot, since the action "put the apple in the compost" has a sufficiently higher probability of being correct than any other option.

"Using the technique of conformal prediction, which quantifies the language model's uncertainty in a more rigorous way than prior methods, allows us to get to a higher level of success" while minimizing the frequency of triggering help, said the study's senior author Anirudha Majumdar, an assistant professor of mechanical and aerospace engineering at Princeton.

Robots' physical limitations often give designers insights not readily

available from abstract systems. Large language models "might talk their way out of a conversation, but they can't skip gravity," said co-author Andy Zeng, a research scientist at Google DeepMind. "I'm always keen on seeing what we can do on robots first, because it often sheds light on the core challenges behind building generally intelligent machines."

Ren and Majumdar began collaborating with Zeng after he gave a talk as part of the [Princeton Robotics Seminar](#) series, said Majumdar. Zeng, who earned a computer science Ph.D. from Princeton in 2019, outlined Google's efforts in using LLMs for robotics, and brought up some open challenges. Ren's enthusiasm for the problem of calibrating the level of help a robot should ask for led to his internship and the creation of the new method.

"We enjoyed being able to leverage the scale that Google has" in terms of access to large language models and different hardware platforms, said Majumdar.

Ren is now extending this work to problems of active perception for robots: For instance, a robot may need to use predictions to determine the location of a television, table or chair within a house, when the robot itself is in a different part of the house. This requires a planner based on a model that combines vision and language information, bringing up a new set of challenges in estimating uncertainty and determining when to trigger help, said Ren.

More information: Allen Z. Ren et al, Robots That Ask For Help: Uncertainty Alignment for Large Language Model Planners (2023) openreview.net/forum?id=4ZK8ODNyFXx

Provided by Princeton University

Citation: How do you make a robot smarter? Program it to know what it doesn't know (2023, November 29) retrieved 6 December 2023 from <https://techxplore.com/news/2023-11-robot-smarter-doesnt.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.