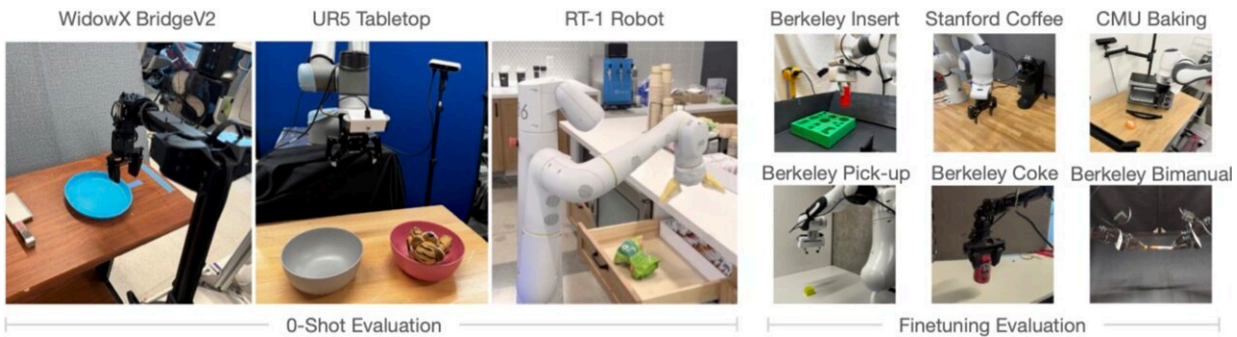


# An open-source generalist model for robot object manipulation

June 10 2024, by Ingrid Fadelli



These are the robots we tested Octo on – you can see that there is a wide range of different robot arms, from small to large, single arm to bimanual. Octo was able to control all these robots. Credit: Team et al.

The public release of ChatGPT and other large language models (LLMs) has allowed developers worldwide to start experimenting with these models to enhance the interactive capabilities of their own systems. Similar generalizable models for robotic manipulation, however, remain scarce.

Researchers at University of California, Berkeley (UC Berkeley), Stanford University and CMU recently introduced Octo, an open-source generalist model for robotic manipulation that could allow different robotic systems to effectively manipulate a wide range of objects. This model, presented in a [paper](#) pre-published on the server *arXiv*, could

open new avenues for the development of robots that can tackle manual tasks.

"Much of the current progress in AI is driven by [large datasets](#) and large models," Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black and Oier Mees, told Tech Xplore. "In the robotics community, we recently assembled [the Open X-Embodiment dataset](#), a big manipulation dataset that pools data from many [research institutions](#). While this new dataset is a really exciting resource, at the time there weren't many models that could make use of it yet."

The recent work by this research team had two main objectives. The first was to develop a good generalist robotics model that could be applied to various robots and the second was to create open-source code that would allow other researchers to build similar models in the future.

"Octo is what we call a 'generalist' [robot](#) model, a [neural network](#) that can control many different types of robots and make them fulfill requests like 'pick up the spoon,' 'close the drawer,' 'wipe the table' etc.," Ghosh, Walke, Pertsch, Black and Mees explained.

"Being a generalist and working on many robots is key, because if you look at research labs around the world, many of them use different robots, so the only way to ensure Octo can be used by many researchers is by supporting a wide range of robots."

Within the technology research and development community, highly performing computational tools that can be applied across multiple systems are often referred to as foundational models. An example of these models is ChatGPT, which can be used to equip various agents and systems with natural language processing (NLP) capabilities.

"We want to build similar foundation models, but for robot control, or in

other words, models that can control many robots and make them solve many different tasks," Ghosh, Walke, Pertsch, Black and Mees said.

"Octo is a first step towards that goal. Its training looks very similar to models like ChatGPT: we curate a large and diverse dataset, in our case robot data instead of text, and train a large model to predict the next action the robot should execute given the current robot state and a task instruction."

Octo, the model developed by Ghosh, Walke, Pertsch, Black and Mees is based on the same type of neural networks as ChatGPT, known as transformers. A key advantage of Octo over other previously developed robotics models is the scale of the data used to train it and its flexibility.

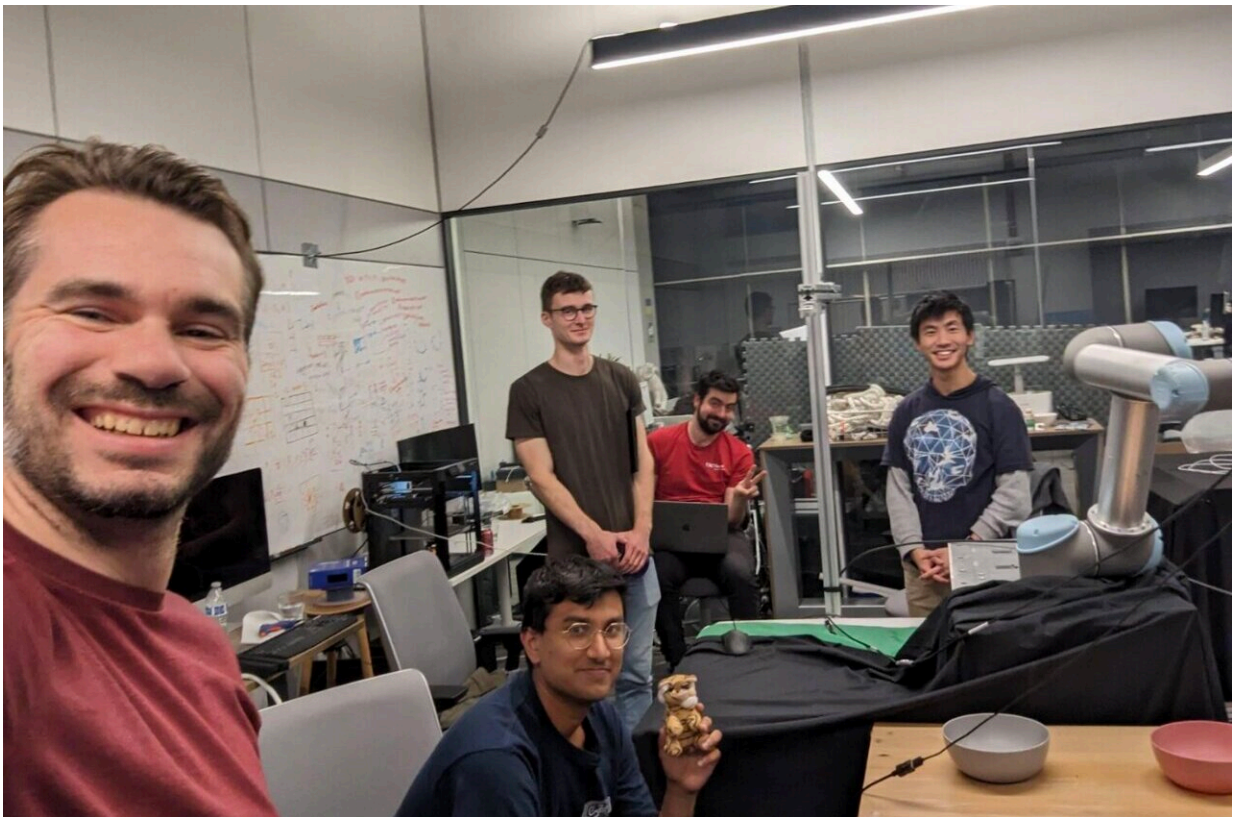
The model was trained on the largest dataset of robotic manipulation trajectories compiled to date; the Open X-Embodiment dataset. Octo can also process a diverse range of sensory inputs including different types of images, robot joint readings, language instructions, goal-related images and more.

"Octo can also control many different types of robot arms, from small single arms that can barely pick up a soda can, to larger, more powerful robot arms and even bi-manual setups," Ghosh, Walke, Pertsch, Black and Mees said. "This flexibility is what makes Octo more applicable to the diverse setups roboticists actually have around the world."

The researchers evaluated their model in a series of initial experiments, deploying it on nine different robotic systems developed at UC Berkely, Stanford and CMU. Octo succeeded in controlling these robots and allowed them to complete various manipulation tasks, even in instances where it had not encountered data collected by these robots' sensors or their unique design during training.

"It was really cool to see that we can take our Octo model and use it to control many different robots," the researchers said. "Since we [released the model](#), we saw quite a few people who tried running it on their own robots and we have been using the codebase we built for Octo in our next projects as well. These are some encouraging signs that Octo will indeed help foster the next generation of improved foundation models for robotics."

For the researchers, the development of Octo was merely a small milestone towards their goal of building a generalist model for robotic manipulation. In their next studies, they plan to continue working towards this goal and hope that research groups at other institutes will also start experimenting with their code.



Part of the Octo model team when we were running robot experiments late at

night before the model release (Left to right: Oier Mees, Dibya Ghosh, Homer Walke, Karl Pertsch, Lawrence Chen). Octo was a big team effort between multiple research labs from Berkeley, Stanford and CMU. Work on foundation models in robotics is hard, with many many hours spent evaluating models on all different types of robots, so having many helping hands is a necessity. Credit: Team et al.

"Right now, chances are that the model will not work on your robot out of the box and you need to collect a few examples of the task you want your robot to solve to teach it to Octo, even if it's a mundane task like picking up a coke can in a new kitchen," they added.

"This is to say, the generalization ability of the current model is still pretty limited and we're working on new models that will push this a bit further. We're not yet at the point where you can just download a model to your robot, tell your robot what you'd like it to do and it will succeed 9 out of 10 times, but we're working towards this goal."

**More information:** Dibya Ghosh et al, Octo: An Open-Source Generalist Robot Policy, *arXiv* (2024). [DOI: 10.48550/arxiv.2405.12213](https://doi.org/10.48550/arxiv.2405.12213)

© 2024 Science X Network

Citation: An open-source generalist model for robot object manipulation (2024, June 10) retrieved 18 June 2024 from <https://techxplore.com/news/2024-06-source-generalist-robot.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.