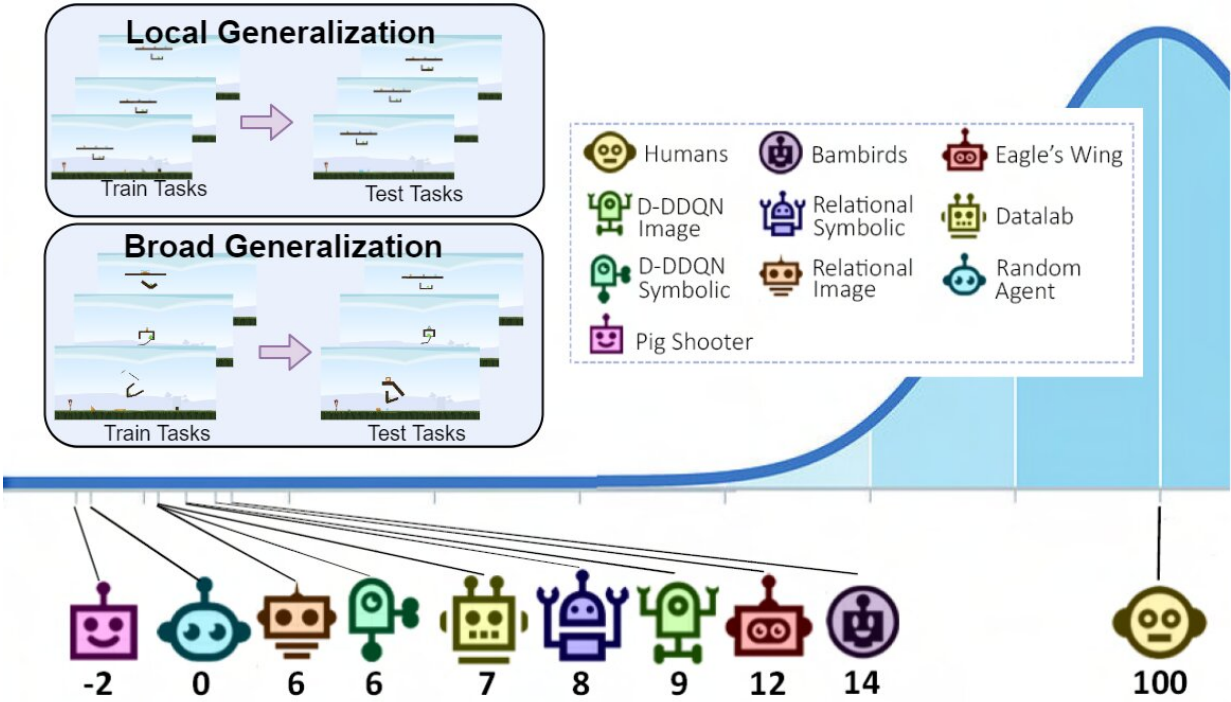


A testbed to assess the physical reasoning skills of AI agents

February 8 2023, by Ingrid Fadelli



An illustration showing the local and broad generalization setup in the Phy-Q testbed and the Phy-Q score obtained by different AI agents and humans. Credit: Xue et al

Humans are innately able to reason about the behaviors of different physical objects in their surroundings. These physical reasoning skills are incredibly valuable for solving everyday problems, as they can help

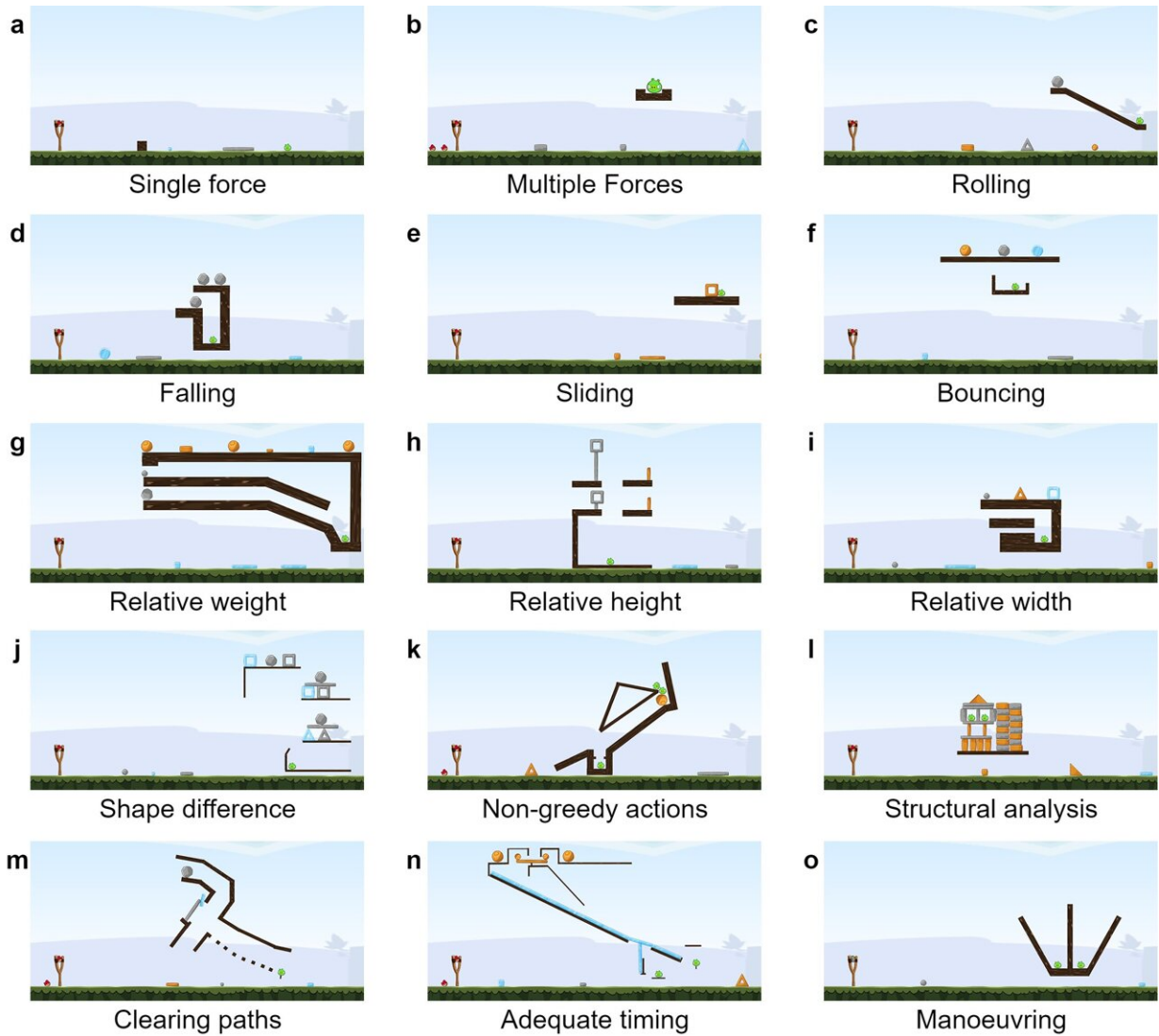
us to choose more effective actions to achieve specific goals.

Some computer scientists have been trying to replicate these reasoning abilities in [artificial intelligence](#) (AI) [agents](#), to improve their performance on [specific tasks](#). So far, however, a reliable approach to train and assess the physical reasoning capabilities of AI algorithms has been lacking.

Cheng Xue, Vimukthini Pinto, Chathura Gamage, and colleagues, a team of researchers at the Australian National University, recently introduced Phy-Q, a new [testbed](#) designed to fill this gap in the literature. Their testbed, introduced in a paper in *Nature Machine Intelligence*, includes a series of scenarios that specifically assess an AI agent's physical reasoning capabilities.

"Physical reasoning is an important capability for AI agents to operate in the [real world](#) and we realized that there are no comprehensive testbeds and a measure to evaluate the physical reasoning intelligence of AI agents," Pinto told Tech Xplore. "Our primary objectives were to introduce an agent friendly testbed along with a measure for physical reasoning intelligence, evaluating the state-of-the-art AI agents along with the humans for their physical reasoning capabilities, and providing guidance to the agents in the AIBIRDS competition, a long running competition for physical reasoning held at IJCAI and organized by Prof. Jochen Renz."

The Phy-Q testbed is comprised of 15 different physical reasoning scenarios that draw inspiration from situations in which infants acquire physical reasoning abilities and real-world instances in which robots might need to use these abilities. For every scenario, the researchers created several so-called "task templates," modules that allow them to measure the generalizability of an AI agent's skills in both local and broader settings. Their testbed includes a total of 75 task templates.



Screenshots of example tasks in Phy-Q representing the 15 physical scenarios. The slingshot with birds is situated on the left of the task. The goal of the agent is to kill all the green pigs by shooting birds from the slingshot. The dark-brown objects are static platforms. The objects with other colors are dynamic and subject to the physics in the environment. Credit: Xue et al

"Through local generalization, we evaluate the ability of an agent to generalize within a given task template and through broad generalization,

we evaluate the ability of an agent to generalize between different task templates within a given scenario," Gamage explained. "Moreover, combining the broad generalization performance in the 15 physical scenarios, we measure the Phy-Q, the physical reasoning quotient, a measure inspired by the human IQ."

The researchers demonstrated the effectiveness of their testbed by using it to run a series of AI agent evaluations. The results of these tests suggest that the physical reasoning skills of AI agents are still far less evolved than [human](#) abilities, thus there is still significant room for improvement in this area.

"From this study, we saw that the AI systems' physical reasoning capabilities are far below the level of humans' capabilities," Xue said. "Additionally, our evaluation shows that the agents with good local generalization ability struggle to learn the underlying physical reasoning rules and fail to generalize broadly. We now invite fellow researchers to use the Phy-Q testbed to develop their physical reasoning AI systems."

The Phy-Q testbed could soon be used by researchers worldwide to systematically evaluate their AI model's physical reasoning capabilities across a series of physical scenarios. This could in turn help developers to identify their model's strengths and weaknesses, so that they can improve them accordingly.

In their next studies, the authors plan to combine their physical reasoning testbed with open-world learning approaches. The latter is an emerging research area that focuses on improving the ability of AI agents and robots to adapt to new situations.

"In the real world, we constantly encounter novel situations that we have not faced before and as humans, we are competent in adapting to those novel situations successfully," the authors added. "Similarly, for an agent

that operates in the real world, along with the physical reasoning capabilities, it is crucial to have capabilities to detect and adapt to novel situations. Therefore, our future research will focus on promoting the development of AI agents that can perform in physical [reasoning](#) tasks in different novel situations."

More information: Cheng Xue et al, Phy-Q as a measure for physical reasoning intelligence, *Nature Machine Intelligence* (2023). [DOI: 10.1038/s42256-022-00583-4](https://doi.org/10.1038/s42256-022-00583-4)

© 2023 Science X Network

Citation: A testbed to assess the physical reasoning skills of AI agents (2023, February 8) retrieved 1 June 2023 from <https://techxplore.com/news/2023-02-testbed-physical-skills-ai-agents.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.