

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Probing Mechanical Reasoning in Large Vision Language Models

Permalink

<https://escholarship.org/uc/item/6vz43710>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors

Sun, Haoran

Li, Yijiang

Gao, Qingying

et al.

Publication Date

2025

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Probing Mechanical Reasoning in Large Vision Language Models

Haoran Sun

Johns Hopkins University, Baltimore, Maryland, United States

Yijiang Li

Electrical and computer engineering, La Jolla, California, United States

Qingying Gao

Computer Science, Baltimore, Maryland, United States

Haiyun Lyu

University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

Dezhi Luo

University of Michigan, Ann Arbor, Michigan, United States

Hokin Deng

Harvard University, Boston, Massachusetts, United States

Abstract

Mechanical reasoning is a hallmark of human intelligence, defined by its ubiquitous yet irreplaceable role in human activities ranging from routine tasks to civil engineering. Embedding machines with mechanical reasoning is therefore an important step towards building human-level artificial intelligence. Here, we leveraged 155 cognitive experiments to test the understanding of system stability, gears and pulley systems, leverage principle, inertia and motion, and fluid mechanics in 26 vision language models. Results indicate that VLMs consistently perform worse than humans on all domains, while demonstrate significant difficulty in reasoning about gear systems and fluid mechanics. Notably, their performance on these tasks do not improve as number of parameters increase, suggesting that current attention-based architecture may fail to grasp certain underlying mechanisms required for mechanical reasoning, particularly those pertaining to mental simulations.