

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Automated Discovery of Psychological Representations using Language Model Agents

Permalink

<https://escholarship.org/uc/item/58x1x8qz>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Lima Braga, Daniel

Marjieh, Raja

Jagadish, Akshay Kumar

et al.

Publication Date

2026

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

Automated Discovery of Psychological Representations using Language Model Agents

Daniel Braga

Princeton University, Princeton, USA

Raja Marjieh*

Princeton University, Princeton, USA

Akshay K. Jagadish*

Princeton University, Princeton, USA

Ilia Sucholutsky

New York University, New York, USA

Nori Jacoby

Cornell University, Ithaca, USA

Brenden Lake

Princeton University, Princeton, USA

Thomas L. Griffiths

Princeton University, Princeton, USA

*Equal contribution

Abstract

Similarity judgments provide a powerful behavioral signal for characterizing psychological representations. Computational techniques for translating similarity judgments into representations have been available since the early days of cognitive science. However, these algorithms typically focus on one particular kind of representation (e.g., spaces or trees), and identifying a satisfying representation often requires searching over a large space of possible structures and algorithms. Inspired by recent developments in AI, we propose an automated process for efficient representation discovery (AutoRep), in which large language model agents iteratively propose, refine, and critique code for fitting representational structures to similarity data. We show how this process can reliably recover psychological representations from both synthetic and human datasets, and how it can flexibly explore spatial, feature-based, graphical, or even neural representations. Our work demonstrates how modern agentic pipelines can substantially facilitate basic research workflows in cognitive science.