UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Exploring the mechanisms that enable multimodal reasoning about data visualizations in vision-language models

Permalink

https://escholarship.org/uc/item/7bs5p3g5

Journal

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

Authors

Tartaglini, Alexa R.
Potts, Christopher
Fan, Judith E.

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

eScholarship.org Powered by the California Digital Library
University of California

Exploring the mechanisms that enable multimodal reasoning about data visualizations in vision-language models

Alexa Tartaglini
Stanford University, Stanford, California, United States

Christopher Potts
Stanford University, Stanford, California, United States

Judith Fan
Stanford University, Stanford, California, United States

Abstract

Humans can readily integrate visual, linguistic, and numerical information to extract meaning from symbolic displays of information. For instance, answering even a simple question about a data visualization requires connecting tokens of language to visual features in the plot to support quantitative inferences. What are the core computational mechanisms that enable integration across modalities to support such reasoning? Open-source vision-language models (VLMs) might provide a useful testbed for investigating these mechanisms, but doing so requires a high degree of experimental control. To achieve this control, we procedurally generated a large dataset containing pairs of questions and data visualizations that varied along several independent and ecologically important dimensions, including the number of observations and how they were distributed. We identified several open VLMs whose performance was sensitive to this variation, establishing their viability for further exploration of the mechanisms underlying multimodal reasoning.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).