

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Evaluating Language Representation Models on Approximately Rational Decision Making Problems

Permalink

<https://escholarship.org/uc/item/7mm190b9>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Kejriwal, Mayank

Tang, Zhisheng

Publication Date

2022

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

Evaluating Language Representation Models on Approximately Rational Decision Making Problems

Mayank Kejriwal

University of Southern California, Marina del Rey, California, United States

Zhisheng Tang

University of Southern California, Los Angeles, California, United States

Abstract

Transformer-based language representation models, such as GPT-3 and DeBERTa, have yielded state-of-the-art results on several benchmarks in Natural Language Processing (NLP), including question answering, story generation and information extraction. However, little is currently understood about their ability to make (approximately rational) decisions, even when simple bets with clear net-positive, zero-risk choices are presented. We design a suite of multi-option decision-making experiments in the same vein as other classic studies (Kahneman & Tversky, 1979). Applying over 20 structural variants of these ‘benchmarks’ (to test for generalization and robustness) to three established language representation models, we use detailed experiments to show that the models lack fundamental decision-making ability, even after several controls. Techniques such as selective fine-tuning can help improve performance partially, but more fundamental NLP research may be needed before such models can be robustly and generally applied in real-world problem domains where decision making is essential.