UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Limits and risks of GPT-3 applications

Permalink

Journal

Authors

Strasser, Anna
Crosby, Matthew
Schwitzgebel, Eric

Publication Date

2022

Peer reviewed

Limits and risks of GPT-3 applications

Anna Strasser
Ludwig-Maximilians-Universität, Munich, Germany

Matthew Crosby
DeepMind, London, United Kingdom

Eric Schwitzgebel
University of California, Riverside, Riverside, California, United States

Abstract

Limits and risks of GPT-3 applications Recent GPT-3 applications have fed new optimism in AI research. To explore limits and risks, we investigated how far we can get in developing a so-called digital replica of Daniel Dennett, generating text outputs from a large language model trained on Dennett's philosophical writing. In consultation with Dennett himself, we compare several fine-tuning strategies and evaluate outputs. Analyzing the failures and the successes will allow us to address technical and ethical issues such as: 1. How accurate can such models be with current technology? 2. To what extent (if at all) might it be acceptable to present a model's outputs as representing an author's view? 3. Should copyright holders have control over such models? 4. How one can address risks of making such models public (including possible over-reliance on their accuracy)? 5. How good are experts and non-experts in distinguishing humans from replicas?