Reasoning over ontologies using Large Language Models

This manuscript (permalink) was automatically generated from cmungall/gpt-reasoning-manuscript@ca9b3ae on May 24, 2023.

Authors

- **John Doe**
  XXXX-XXXX-XXXX-XXXX · johndoe · johndoe · @johndoe@mastodon.social
  Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Chris Mungall** ✉
  0000-0002-6601-2165 · cmungall
  Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720

✉ — Correspondence possible via GitHub Issues or email to Chris Mungall <cjmungall@lbl.gov>.

Abstract

Reasoning is a core component of human intelligence, and a key goal of AI research. Reasoning has traditionally been the domain of symbolic AI, but recent advances in deep learning and in particular Large Language Models (LLMs) such as GPT-3 seem to suggest that LLMs have some latent reasoning ability.

To investigate this, we created a GPT-based reasoning agent that is intended to perform ontological reason using a few-shot learning approach, using instruction prompting and in-context examples. We also created a series of benchmarks to test ontological reasoning ability in LLMs and other systems.

Our results indicate that GPT is a poor reasoner, and is only able to perform ontological reasoning on some of the simplest tasks. Even on these simple tasks, results are highly variable, with performance degrading as the size of the ontology and the complexity of the explanation increases. In the cases where it does successfully perform the task, this seems to essentially be an advanced pattern-based form of lookup.

Our results indicate that a maximalist approach to using LLMs may be limiting, and that to be successful AI should use hybrid strategies.

Introduction

We...

Citation by DOI [1].

xx

Methods

..

Results

task_type	method	f1_score
EntailedIndirectSuperClassTask	basic	0.739145
EntailedIndirectSuperClassTask	chain_of_thought	0.587469
EntailedIndirectSuperClassTask	explanation	0.561746
EntailedSubClassOfExpressionTask	basic	0.718349
EntailedSubClassOfExpressionTask	chain_of_thought	0.623445
EntailedSubClassOfExpressionTask	explanation	0.755769
EntailedTransitiveSuperClassTask	basic	0.805197
EntailedTransitiveSuperClassTask	chain_of_thought	0.583613
EntailedTransitiveSuperClassTask	explanation	0.68935

task_type	method	f1_score
MostRecentCommonSubsumerTask	basic	0
MostRecentCommonSubsumerTask	chain_of_thought	0
MostRecentCommonSubsumerTask	explanation	0
OntologyCoherencyTask	basic	0
OntologyCoherencyTask	chain_of_thought	0
OntologyCoherencyTask	explanation	0

Discussion

blah

Conclusions

blah

References

1. **Sci-Hub provides access to nearly all scholarly literature**
Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene
eLife (2018-03-01) https://doi.org/ckcj
DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410