

ACL 2023 Cutting-Edge Tutorial: Complex Reasoning over Natural Language

Wenting Zhao, Mor Geva, Bill Yuchen Lin,
Michihiro Yasunaga, Aman Madaan, Tao Yu



Plan for the tutorial

- Review recent benchmarks on complex reasoning
- Review promising directions for tackling complex reasoning tasks
- Tutorial materials are available at tinyurl.com/reasoning-tutorial
- Live Q&A: <https://acl.rocket.chat/channel/tutorial-2>

By reasoning, what do we mean?

Did Aristotle use a laptop?

Input



1. Aristotle lived from 384-322 BC.
2. The first laptop was made in 1980.
3. 322 BC is before 1980.

Reasoning



No.

Output

By reasoning, what do we mean?

1. Aristotle lived from 384-322 BC.
2. The first laptop was made in 1980.
3. 322 BC is before 1980.

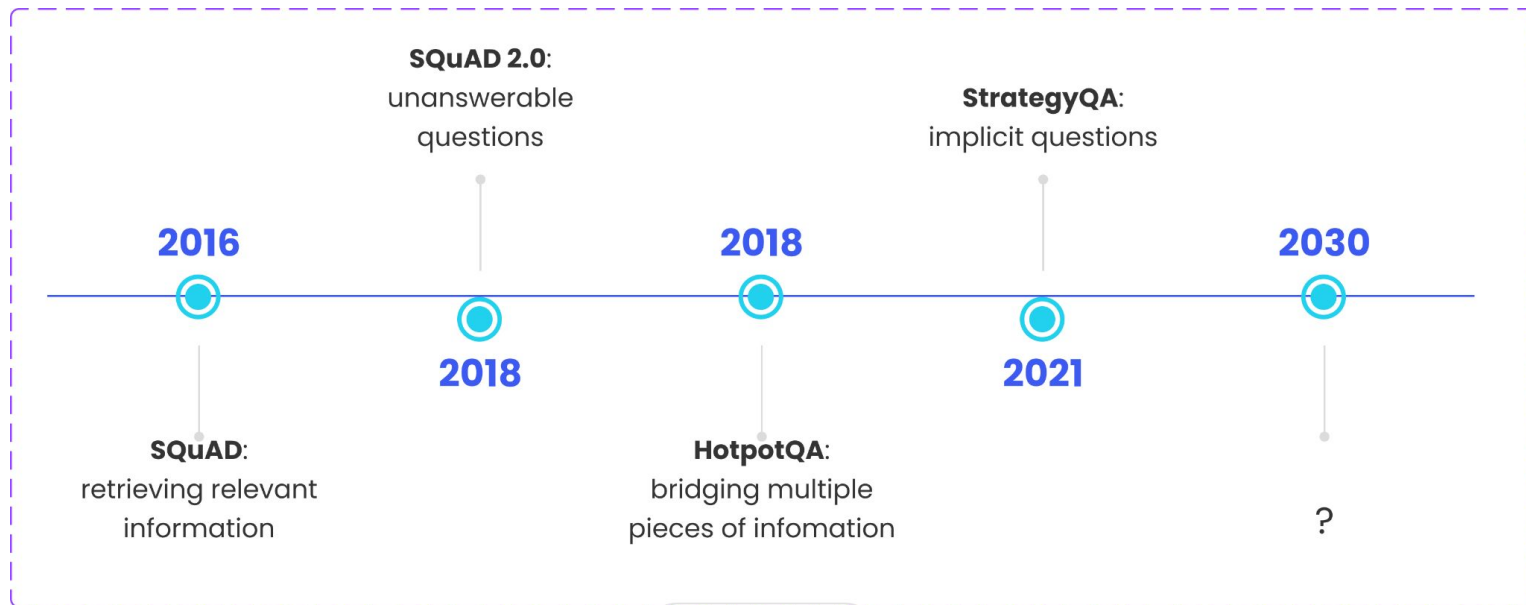
Reasoning

The ***process*** of deriving
the output from the input.

Complex reasoning

- Going beyond the surface meaning
 - What can be easily solved by an end-to-end system
- Examples
 - Compositional reasoning
 - Knowledge retrieval
 - Grounding
 - Commonsense reasoning
 - ...

Trend of NLP tasks - Question Answering



Trend of NLP tasks - Commonsense Reasoning

Before: Reasoning about
common situations

The Smiths went on a vacation without the children.



The Smiths brought a ? souvenir back for Ty.



Ty's face lit up as he ran to the new toy.

Now: Reasoning about
uncommon, long-tail situations

She tried sushi for the first time, and really disliked it.



She wanted to avoid disappointing her partner.

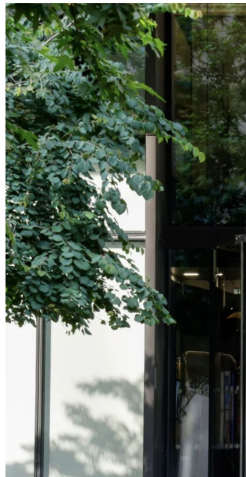



She stayed and ate more sushi.

In 2023, LLMs are coming in a flood


Google opens early access to Bard, its AI chatbot

Romain Dillet @romaindillet / 10



**OpenAI** 🟡 @OpenAI · Mar 14

Announcing GPT-4, a large multimodal model, with our best-ever results on capabilities and alignment: openai.com/product/gpt-4

Meta AI

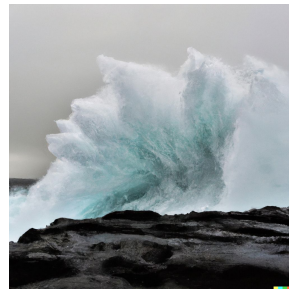
How do

Research

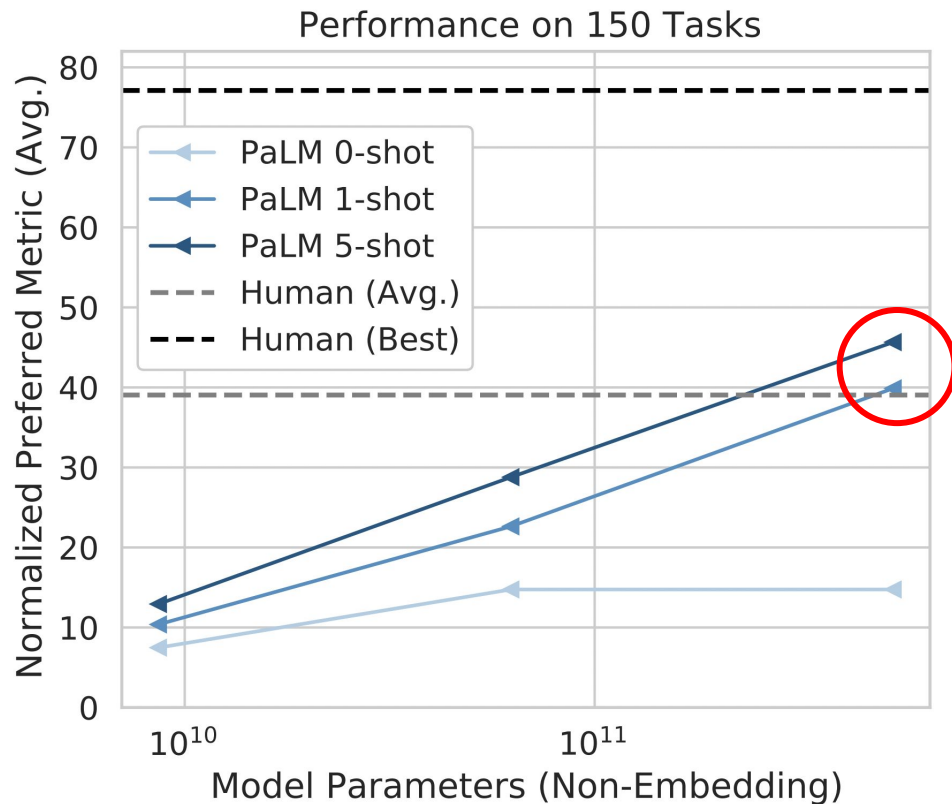
Introducing LLaMA: A foundational, 65-billion-parameter large language model

February 24, 2023

2,293 23.7K 66.8K 10.9M



LLMs have made amazing progress on complex tasks



Are the models really this good?

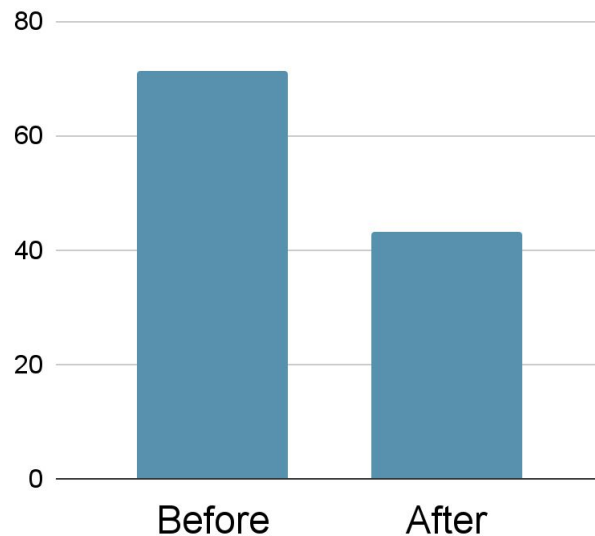
ted.com

Yejin Choi: Why AI is incredibly smart and shockingly stupid

Computer scientist Yejin Choi is here to demystify the current state of massive artificial intelligence systems like ChatGPT, highlighting three key problems ...

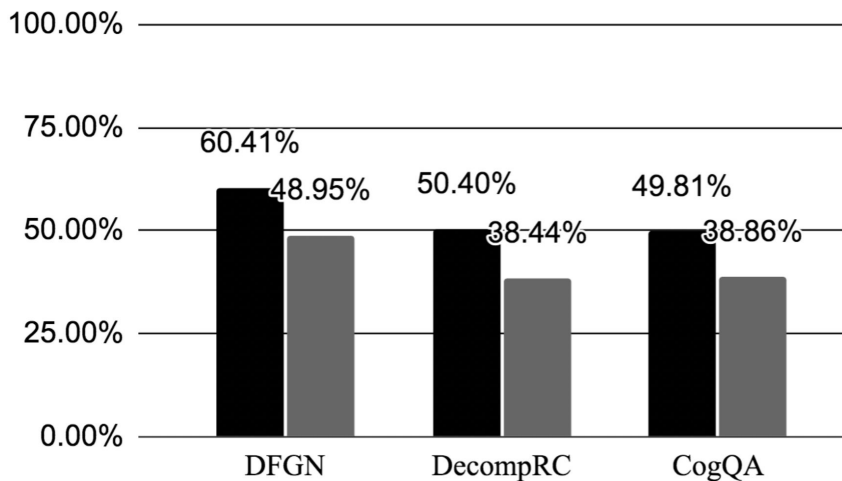
Data contamination

- Before and after removing test data that has n-gram overlap with train data
- Dataset: ARC
- Model: RoBERTa-large



Generalization

- Multi-hop QA systems are less good at answering single-hop sub-questions
- Dataset: HotpotQA
- Model: RoBERTa-large



Left bar: Multi-hop accuracy
Right bar: Single-hop accuracy

Limitation

- Standard fine-tuning / prompting methods only maximizes the task accuracy without explicitly considering the underlying reasoning

Why care about taking the correct reasoning route?

- Deployment in critical domains / building trust with users
- Generalization

Goal of this tutorial

- Explore ways to augment language models with methods that make the reasoning process *explicit*
 - Can we explicitly incorporate knowledge?
 - Can we explicitly specify rules?
 - Can we integrate symbolic reasoning?

Tutorial Schedule

Benchmarks & Evaluation



Mor Geva
Visiting Researcher at Google

“What are the types of complex reasoning abilities recent NLP benchmarks are focused on? And how do we evaluate such abilities?”

9:15-9:40 EST

1(a). Knowledge-augmentation after pretraining

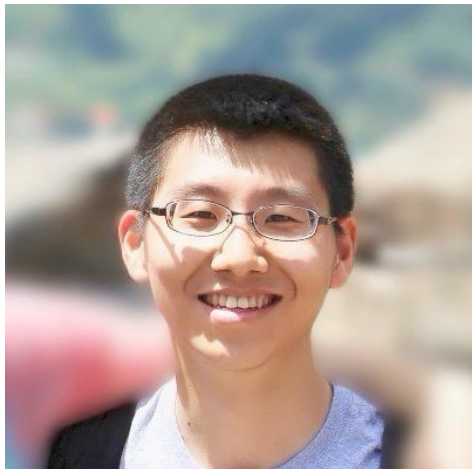


Yuchen Lin
Postdoc at AI2

“What are the ways to incorporate external knowledge when learning specific NLP tasks?”

9:40-10:05 EST

1(b). Knowledge-augmented pretraining



Michihiro Yasunaga
PhD student at Stanford

“We often incorporate knowledge in a task-specific manner 🤔 Can we do this during pretraining to help a broader range of downstream tasks?”

10:05 -10:30 EST

2. Few-shot prompting approaches



Aman Madaan
PhD student at CMU

“What are the clever ways to perform few-shot prompting so that it’s more robust and requires less prompt engineering efforts?”

11:00-11:30 EST

3. Neuro-symbolic approaches: LLMs + tool use

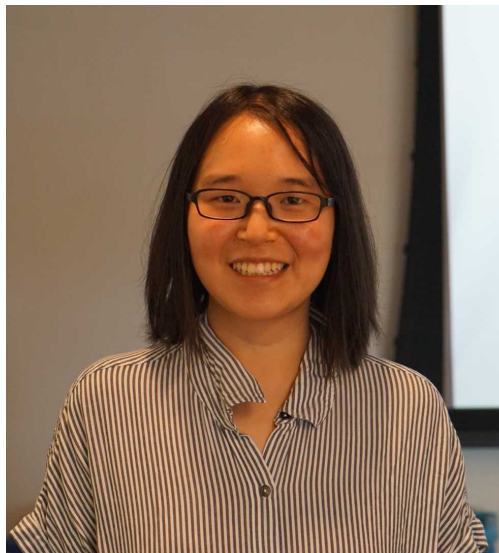


Tao Yu
Assistant Professor at HKU

“How can we combine neural modules and symbolic components to make our NLP systems more robust and interpretable?”

11:30-12:00 EST

4. Rationale-based approaches & Conclusion



Wenting Zhao
PhD student at Cornell

“Let’s think about ways to produce rationales and how can they improve the existing NLP systems.”

12:00-12:30 EST

Related sessions at ACL'23

Abductive Commonsense Reasoning Exploiting Mutually Exclusive Explanations
Wenting Zhao, Justin Chiu, Claire Cardie, Alexander Rush
11:45-12:00 (Metropolitan West) on July 10th

Natural Language Reasoning and Structured Explanations Workshop
July 13

Paper list

[Bhagavatula et al., 2019] Chandra Bhagavatula, Ronan Le Bras, Chaitanya Malaviya, Keisuke Sakaguchi, Ari Holtzman, Hannah Rashkin, Doug Downey, Wen tau Yih, and Yejin Choi. 2020. Abductive commonsense reasoning. In International Conference on Learning Representations