



Towards Human-level AI

Hao Fei
2023

OUTLINE

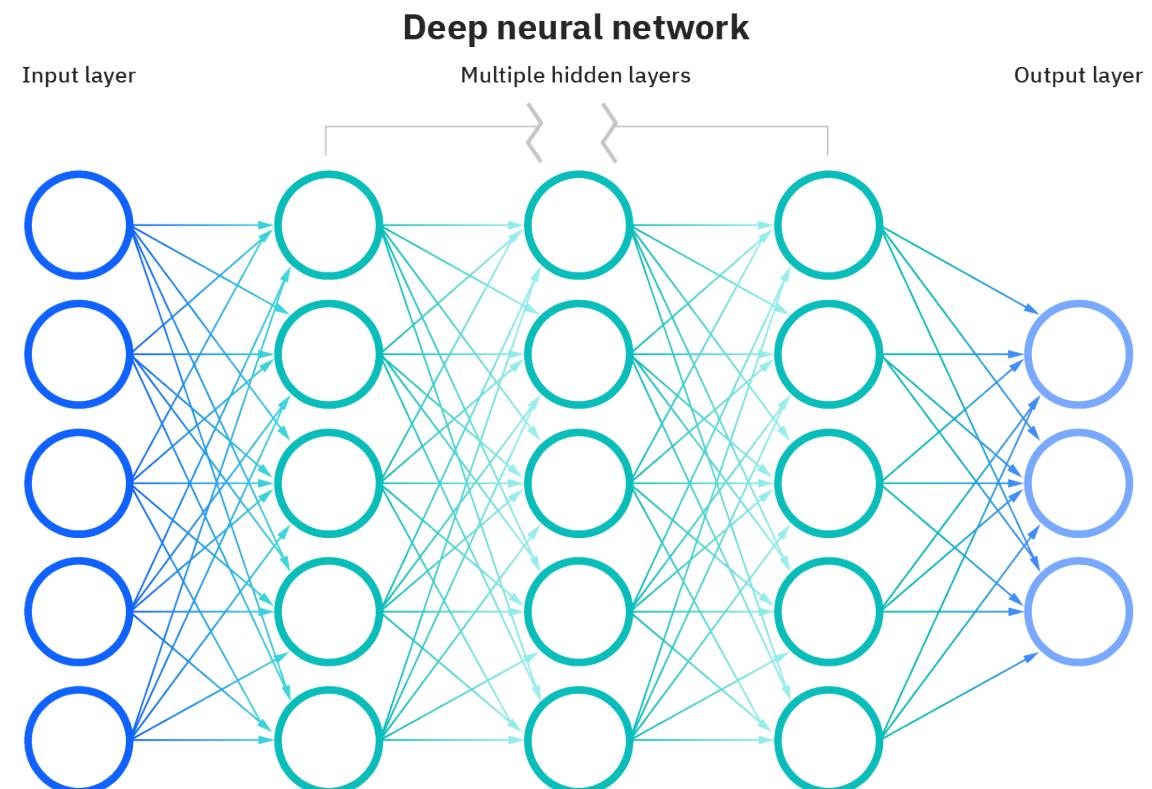
- The essence of human-level AGI
- An application case of current SoTA AI
- Pitfalls in ChatGPT-represented AI
- Post-ChatGPT AI/NLP research

OUTLINE

- The essence of human-level AGI
- An application case of current SoTA AI
- Pitfalls in ChatGPT-represented AI
- Post-ChatGPT AI/NLP research

➤ AI before ChatGPT

- *Neural-architecture-oriented AI*
- *The possibility of achieving AGI in short term is scarce.*
 - Lack of reasoning abilities
 - Limited understanding of human intelligence
 - Data and compute limitations
 - Lack of explainability
 - Theoretical limitations
 - Need for interdisciplinary research
 - ...



➤ AI after ChatGPT

An updated online chatbot?

The chart is titled "ChatGPT" in large, bold, white font at the top center. Below the title are three sections: "Examples" (with a sun icon), "Capabilities" (with a lightning bolt icon), and "Limitations" (with a warning triangle icon). Each section contains three examples or statements.

Examples	Capabilities	Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

➤ AI after ChatGPT, NOW



Public understanding: **ChatGPT ≈ Artificial General Intelligence (AGI)**

- In-depth language semantic understanding

*Grasping the latent intent, and
capture the subtle changes
accurately in the inputs.*

HA 分析一句话：“我今天比较累了，打算晚上9点钟睡觉，然后明天有个会议，早上大概9点一刻叫醒我，我上线，大概到中午12点半能开完。”，请问几点钟开会？

根据句子的描述，会议是在明天早上，“早上大概9点一刻”会叫醒说话者，所以会议应该在“早上大概9点一刻”或者稍后的时间，具体的时间并没有被明确提到。

➤ AI after ChatGPT, NOW



Public understanding: **ChatGPT ≈ Artificial General Intelligence (AGI)**

- In-depth language semantic understanding

*Grasping the latent intent, and
capture the subtle changes
accurately in the inputs.*



期中考试出成绩了，小明的语数外成绩分别是：89, 90, 97；小李的语数外成绩分别是：90, 87, 92。请将这两个同学的成绩做成表格并计算成绩总分，最后按照总分进行排序。



可以将成绩制作成一个表格，如下所示：

姓名	语文成绩	数学成绩	英语成绩	总分
小明	89	90	97	276
小李	90	87	92	269

根据表格可以计算出小明的总分为276分，小李的总分为269分。因此，小明的总分比小李高，排名第一。

➤ AI after ChatGPT, NOW



Public understanding:

- Impressive language abilities

Impressively capable of generating highly sophisticated and nuanced responses.

Even the key information is not possibly correct, the response texts are well organized logically.

S

你如何看待chatgpt对NLP领域，甚至AI领域带来的新变化，请分点论述

C



ChatGPT是一种基于深度学习技术的大型语言模型，其在NLP领域和AI领域带来了许多新变化。具体来说，可以从以下几个方面进行论述：

1. 提升了NLP应用的性能：ChatGPT通过学习大量的语言数据，可以生成高质量的文本和对话。因此，它被广泛应用于机器翻译、语音识别、情感分析、聊天机器人等NLP领域的应用中，并且取得了非常好的效果。
2. 拓宽了NLP应用的范围：传统的NLP技术通常需要手动编写规则和特征工程，因此只能解决一些特定的问题。而ChatGPT可以自动学习语言的规则和特征，因此可以处理更加复杂和多样化的问题，例如文本生成、机器翻译等。
3. 推动了对话系统的发展：ChatGPT的应用使得对话系统的性能有了质的提升。通过与ChatGPT进行交互，人们可以获得更加流畅、自然的对话体验。这种体验已经被广泛应用于在线客服、智能客户端、虚拟助手等领域。
4. 促进了语言学习的研究：ChatGPT的成功证明了基于深度学习的语言模型在自然语言处理方面的优势。这也为语言学习的研究提供了新的机会和挑战，例如如何让模型更好地理解上下文、如何进行知识图谱的构建等。

总之，ChatGPT对NLP领域和AI领域的发展产生了深远的影响，推动了自然语言处理技术的进步，也为人机交互提供了更加流畅、自然的交互方式。

➤ AI after ChatGPT, NOW



Public understanding:

- Reasoning ability

basic logic reasoning.

S

已知一张桌子的价钱是一把椅子的10倍，又知一张桌子比一把椅子贵378.9元，一张桌子和一把椅子各多少钱？

设一把椅子的价格为 x 元，则一张桌子的价格为 $10x$ 元。根据题目中的信息，可以得到以下方程：

$$10x - x = 378.9$$

化简得：

$$9x = 378.9$$

解方程可得：

$$x = 42.1$$

因此，一把椅子的价格为42.1元，一张桌子的价格为 $10x = 421$ 元。

- Positioning the *intelligence* in ChatGPT



My understanding:

ChatGPT



Artificial General Intelligence (AGI)

- ChatGPT is not AGI.
- ChatGPT is so far the nearest technique to AGI.
- ChatGPT would most likely be the embryo of AGI

➤ Abstract Thinking in Advanced Beings

- **Language-Thought**

Language and thought are deeply intertwined.

Some researchers believe that language plays a crucial role in shaping our thoughts and mental processes.

- **Language-Cognition**

Language is essential for many cognitive processes, such as perception, memory, attention, and reasoning.

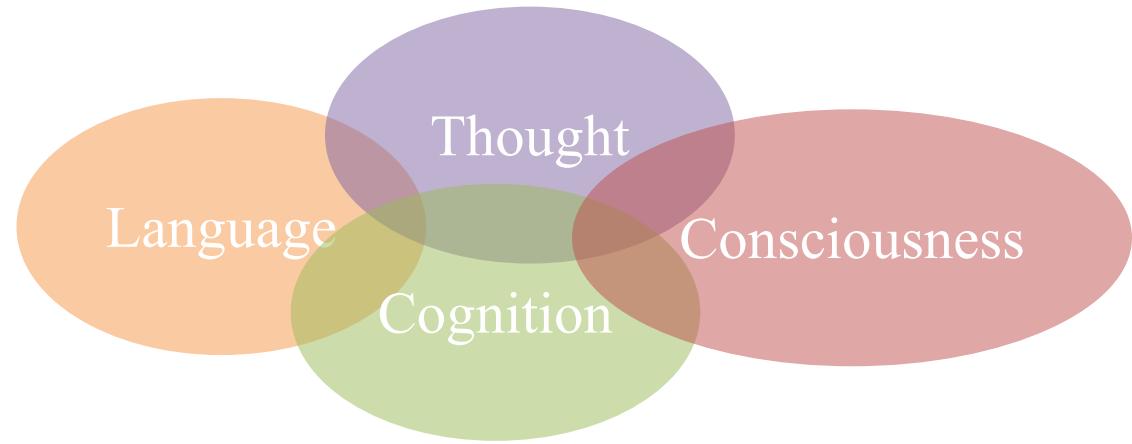
It provides a way to categorize and label our experiences, which in turn shapes how we perceive and remember them.

- **Thought-Cognition**

Cognition providing the foundation for many types of thought, such as problem-solving, decision-making, and creativity.

- **Consciousness-Thought**

Consciousness refers to our subjective experience of the world, including our thoughts, emotions, and sensations. It is closely linked to thought, with our thoughts and mental processes being a major component of our conscious experience.

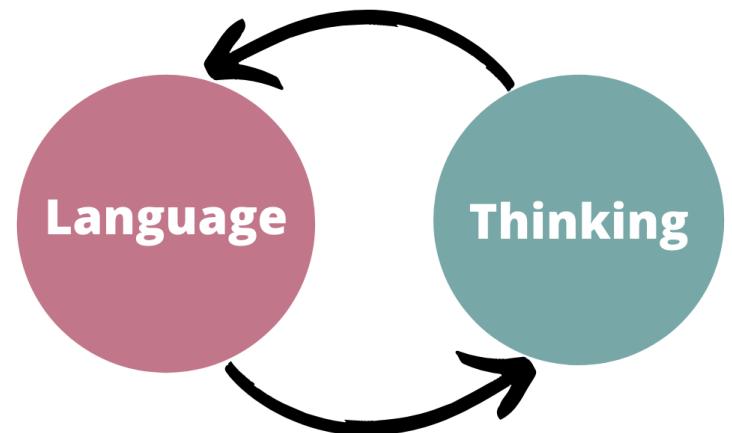


➤ Example 1: *Unconscious Processing, Subconscious*

Every day we hear and read thousands of words, and quickly, naturally understand who or what is being described, and speculate about things that are not explicitly mentioned in the text. But we hardly consciously or intentionally notice this process.



- *Our brains are highly efficient at processing language, and we have developed automatic neural pathways that allow us to quickly and effortlessly understand the meaning of words and sentences without conscious effort.*
- *The subconscious is believed to play a significant role in shaping our beliefs, attitudes, and perceptions, and it can be accessed through techniques such as hypnosis, meditation, and dream analysis.*



➤ Example 2: *Imagination*

Our thoughts are capable of cross-modal imagination, as demonstrated when we hear the phrase "Don't think about that elephant" and a mental image of a big-eared, long-nosed elephant immediately comes to mind.



Our brain is able to integrate information from different sensory modalities, such as visual, auditory, and olfactory. Therefore, even if we are instructed not to think of an elephant, the activation of this concept in our brain can trigger the cross-modal retrieval of associated sensory information and result in the experience of imagining an elephant.



➤ Example 3: *Robustness to Input Information*

Human beings have high error robustness to language and various input information:

“研表究明，汉字序顺并不定一影响阅读。比如当你看完这句话后，才发现这里的字全是乱的。”



We don't rely solely on the order of individual characters or words to understand the meaning of a sentence. Instead, we rely on contextual cues and our knowledge of the language's grammar, syntax, and semantics to guide our comprehension.

In the case of the jumbled Chinese characters, our brains are able to utilize our knowledge of the language to infer the correct order of the characters based on their context and the rules of the language, allowing us to comprehend the sentence despite its scrambled appearance.



➤ Example 4: *Self-consciousness*

Humans (or other advanced beings) possess self-awareness, which is characterized by the recognition of oneself as a unique and distinct individual separate from others and the environment. This is evidenced by the ability to pass mirror tests and the presence of a self-model in their world view.



Actually the phenomenon is complex and not fully understood. However, it is believed to involve a combination of cognitive and neurological factors, including the ability to introspect and reflect on one's own mental states, the development of a sense of agency and ownership over one's actions and experiences, and the neural capacity to represent and manipulate mental models of the self.



The essence of human-level AGI

➤ What is Human-level AI



Machine performing any intellectual task (i.e., can learn and understand) as human do.

- *Capable of abstract thinking, problem-solving, creativity, and other cognitive abilities that are typically associated with human intelligence.*
- *General-purpose intelligence, not just limited to a specific task or application.*



JARVIS [Ironman]



David [A.I. Artificial Intelligence]

➤ Essence to Human-level AI

✓ **Essence-I: Mastery of World Knowledge**

- *Mastering world knowledge is critical for building a comprehensive mental model of the world, which is necessary for self-awareness.*

✓ **Essence-II: Autonomous Reasoning**

- *This ability is critical for developing a robust self-model and for maintaining a sense of agency and ownership over one's actions and experiences.*
- *Autonomous reasoning also enables an individual to adapt to novel situations, plan for the future, and engage in creative problem-solving.*

➤ Essence to Human-level AI



Both two abilities are mutual independent.

- *Reasoning needs to be performed based on existing world knowledge.*
- *Common-sense knowledge gradually grows in the process of understanding the world via reasoning.*
- *With the increase of knowledge, the ability to reason the world will become more and more accurate.*
- *Together, these two abilities are essential for achieving AGI.*

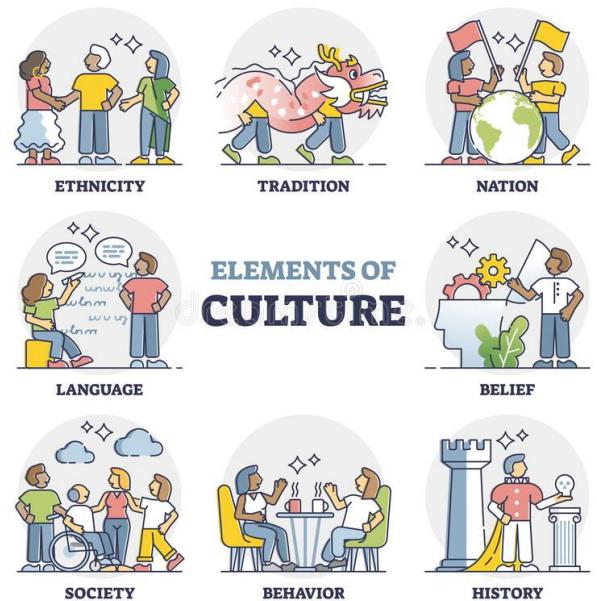
➤ Basic Knowledge

basic information about the physical world, such as the properties of matter and the laws of physics.



➤ More Abstract Concept and Knowledge

social interactions, culture, and emotions, etc.



➤ Capability in ChatGPT



Mostly did good job!

S 力的三要素是什么

力是物理学中的一个基本概念，其三要素是大小、方向和作用点。在描述一个力时，需要指出它的大小（通常使用牛顿为单位）、作用方向（可以用向量表示），以及作用点（即力作用的位置）。这三个要素可以用来准确描述力的性质和作用方式。

S 被100度的水和100度的水蒸气烫伤，哪个更严重？

被100度的水蒸气烫伤比被100度的水更加严重。虽然两者温度相同，但水蒸气在100度时已经变成气态，它具有更高的热量和热能，能够更快地将热量传递给皮肤，造成更加严重的灼伤。此外，水蒸气能够穿透皮肤的表层，更容易引起组织的损伤和烫伤。因此，在处理热水或蒸汽时，需要格外小心，避免烫伤和烧伤发生。

➤ Faults



Sometimes generated anti-factual outputs.

S

你觉得《长安的葡萄》这本书怎么样?



作为一个人工智能，我并没有读书的能力，也没有情感和主观意见。我只能给你一些客观的信息。《长安的葡萄》是一本小说，作者是陈忠实。这本书在出版后引起了广泛的关注和讨论，被认为是描写中国农村生活最为生动的作品之一。它通过一个家庭的变迁，展现了中国农民在社会转型期中所经历的种种变化和矛盾。此外，该书也被改编为电视剧和电影，受到了观众的喜爱和认可。不过，每个人的阅读体验和感受都是不同的，你可以亲自阅读并形成自己的看法。

S

《请回答1988》中狗焕的扮演者是谁?



电视剧《请回答1988》中，狗焕的扮演者是韩国演员朴宝剑（Park Bo Gum）。他出演过多部影视作品，在亚洲和国际范围内都拥有广泛的粉丝基础。

➤ Autonomous Reasoning



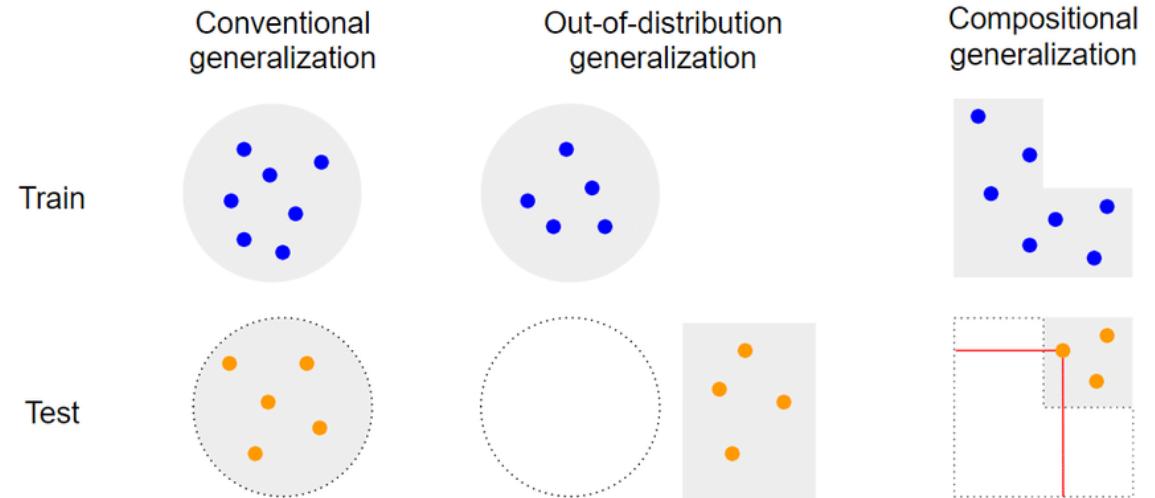
a higher-level cognitive process, actively reasoning and making decisions based on the acquired world knowledge.

- **Problem-solving:** Using logic and critical thinking to find solutions to complex problems.
- **Decision-making:** Weighing the pros and cons of different options and selecting the best course of action.
- **Creative thinking:** Generating new and original ideas through imagination and brainstorming.
- **Predicting:** Making educated guesses about future events based on past experiences and current data.
- **Analyzing:** Breaking down complex information into smaller parts to better understand it.
- **Evaluating:** Assessing the value or quality of something based on certain criteria.
- **Learning:** Acquiring new knowledge and skills through observation, experience, and instruction.
- **Planning:** Creating a strategy or course of action to achieve a desired outcome.
- **Introspection:** Reflecting on one's own thoughts and feelings to gain insight into oneself.
- **Self-correction:** Adjusting one's behavior or beliefs based on new information or feedback.

➤ Measuring AGI

- *Out-of-Distribution Generalization*

An ability of a system to perform well in a testing environment where the distribution of data is different from the one it was trained on.



- *Compositional Generalization*

A key branch of OOD generalization, aims to enable a model to infer the meaning of unknown object combinations based on the properties of known objects



➤ Compositional Generalization

- *Critical for achieving AGI, as it enables a system to learn and reason about new concepts and situations that it has not previously encountered.*
- *Particularly important for tasks that require flexible and creative problem-solving, such as language understanding and natural interaction with the world.*

Shane Steele

brother

<entity> 's <entity>

Did <entity> <verb> <entity> ?

<verb> and <verb>

produce

direct

Revenge Of The Spy

➤ Compositional Generalization

- Conceptualization
 - *Abstract representations of objects and categories in the human mind, which help us understand and process new entities that appear within known categories.*
 - *In a sense, concepts are like **the glue that holds our mental world together**.*
- Compositionality
 - *Key aspect of human generality, referring to the ability to understand and produce new combinations from known components.*
 - *In a sense, through compositionality that humans **understand and produce novel combinations from known components**.*

- Compositional Reasoning
 - Common-sense Reasoning
 - Relational Reasoning
 - Multi-hop Reasoning

➤ Common-sense Reasoning

Reasoning based on common-sense knowledge

Imagine :

Someone asked you, "Can you keep a giraffe in your office?"
You may immediately answer, "Of course not".



- *Because the giraffe is a very tall animal, and the office is an indoor room. There is no room for such a tall animal in it...*
- *It is a well-known fact that have already existed in our minds.*



➤ Relational Reasoning

Inferring new knowledge based on known relationships of entities

If the system knows that **John** is the father of **Zhang San**, what new knowledge can be introduced based on this relationship?



Zhang San is the son of **John**.



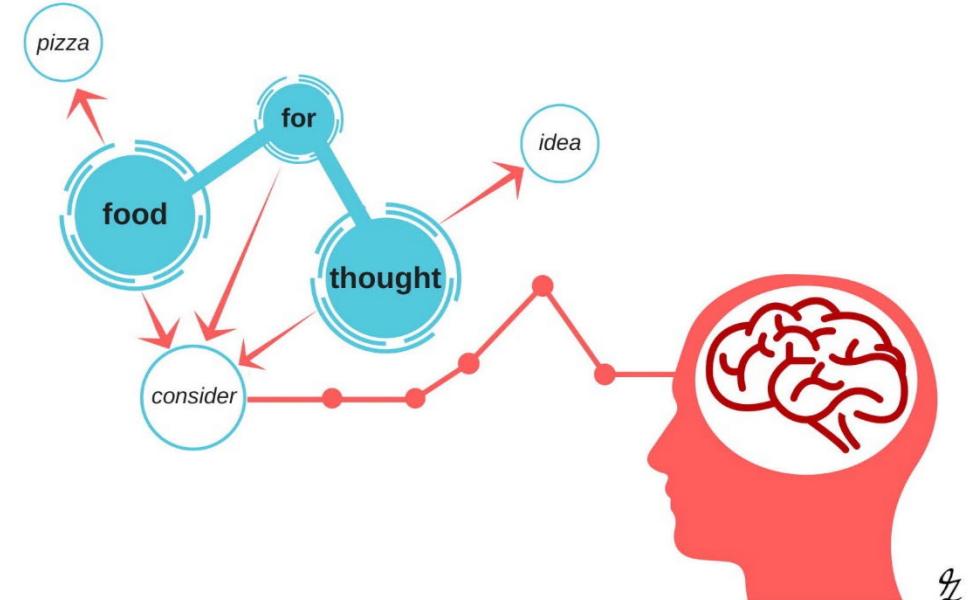
➤ Multi-hop (Discrete) Reasoning

Deducing new knowledge from a series of steps over existing known information

If the system knows that **A** is the father of **B**, and **B** is the father of **C**, what new knowledge can be deduced based on this relationship?



A is the grandfather of C.



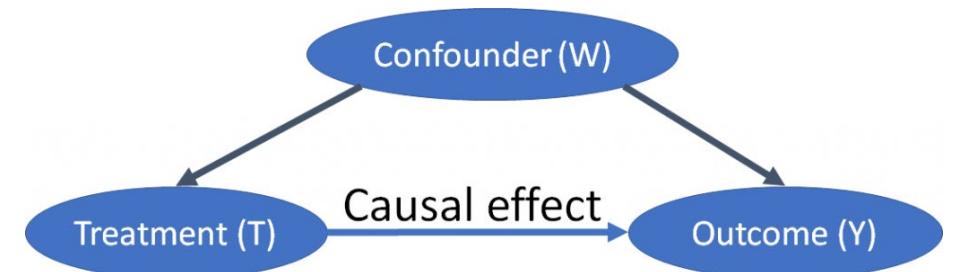
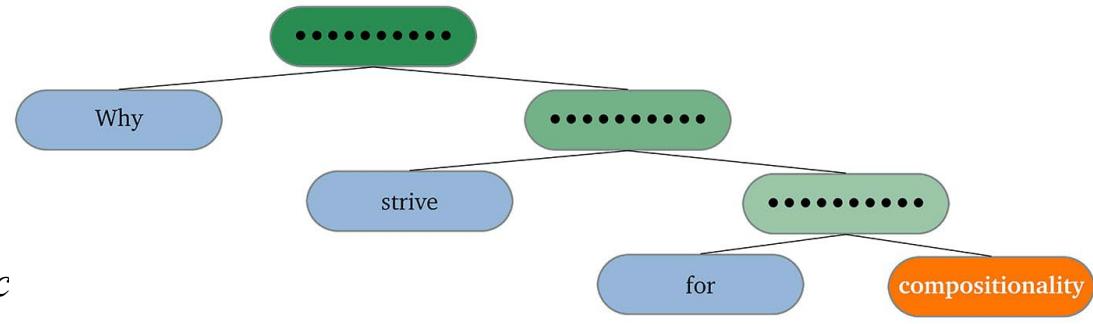
➤ Implementing Compositional Reasoning in Small Models

- Recursive Neural Modeling

- *Learning hierarchical structures in language by recursively applying the network to each node in a parse tree, capturing the meaning of larger structures by combining the meanings of their smaller parts.*
- *Enabling to generalize to new, unseen combinations of words and phrases, rather than simply memorizing specific examples from the training data.*

- Causal Inference

- *Allowing to identify the causal relationships between different variables in a system, helping predict how the system will behave in new, unseen contexts.*
- *Identifying which variables are most important for making accurate predictions.*



Identification: Causal effect → Observed effect conditioned on W, $E[Y|T, W]$

Estimation: $E[Y|T, W] \rightarrow$ Propensity Score Stratification

➤ Implementing Reasoning in Large Language Models

□ Capability in ChatGPT



Mostly did good job!

S

人有两条腿可以骑自行车，那鸭子也有两条腿，它可以骑自行车吗？



鸭子通常使用它们的两条腿来行走、游泳和飞行，但它们没有足够的平衡和协调能力来骑自行车。自行车需要人类具备特定的平衡和协调能力才能够骑行，而鸭子的身体结构和行动能力与人类不同，它们无法像人类一样骑自行车。因此，鸭子不能骑自行车。

➤ Implementing Reasoning in Large Language Models

□ Capability in ChatGPT



Mostly did good job!

s

已知一张桌子的价钱是一把椅子的10倍，又知一张桌子比一把椅子贵378.9元，一张桌子和一把椅子各多少钱？



设一把椅子的价格为 x 元，则一张桌子的价格为 $10x$ 元。根据题目中的信息，可以得到以下方程：

$$10x - x = 378.9$$

化简得：

$$9x = 378.9$$

解方程可得：

$$x = 42.1$$

因此，一把椅子的价格为42.1元，一张桌子的价格为 $10x = 421$ 元。

➤ Implementing Reasoning in Large Language Models

□ Capability in ChatGPT



*Sometimes failed give
correct reasoning.*

S 小王28岁了，如果他在8年前出生，那么他现在几岁？

如果小王在8年前出生，那么现在他的年龄应该是 $8+28=36$ 岁。因为小王在8年前出生，那么到现在就过去了 $28+8=36$ 年。

➤ Implementing Reasoning in L

□ Capability in Ch



*Sometimes failed give
correct reasoning.*

HA

For the sentence "The trophy doesn't fit in the suitcase because it is too small", please choose the correct coreference of the word "it" between A) the trophy and B) the suitcase.



The correct coreference of the word "it" in the sentence "The trophy doesn't fit in the suitcase because it is too small" is A) the trophy.

HA

For the sentence "The trophy doesn't fit in the suitcase because it is too big", please choose the correct coreference of the word "it" between A) the trophy and B) the suitcase.



The correct coreference of the word "it" in the sentence "The trophy doesn't fit in the suitcase because it is too big" is A) the trophy.

HA

For the sentence "The trophy doesn't fit in the suitcase because it is too small", please choose the correct coreference of the word "it" between A) the trophy and B) the suitcase.



The correct coreference of the word "it" in the sentence "The trophy doesn't fit in the suitcase because it is too small" is A) the trophy.

➤ Implementing Reasoning in Large Language Models

Data Fitting Levels

- *Direct final task predictions*
 - *Task labels*
 - *Shortcut features*
- *Intermediate patterns and rules*
 - *The reasoning steps of giving final results*
 - *Calculating mathematical derivation process*
 - *Intermediate features*

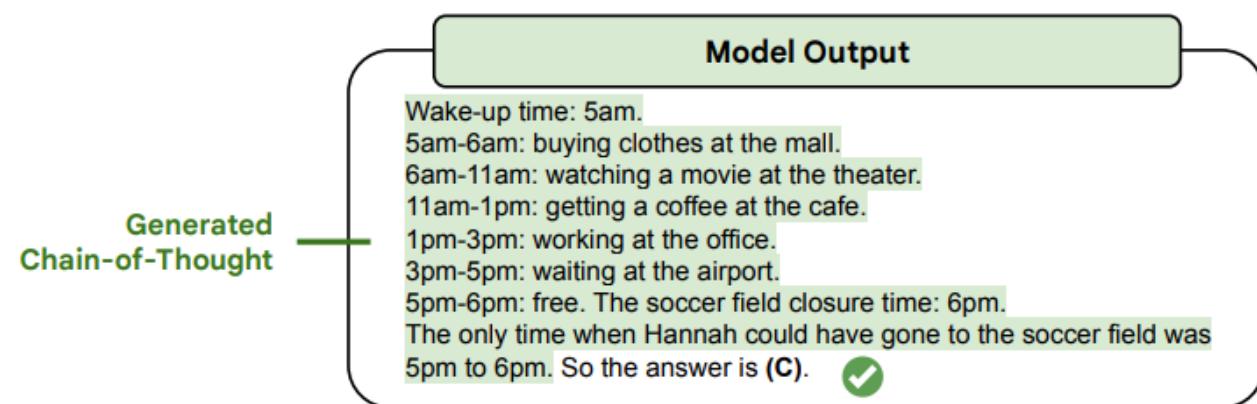
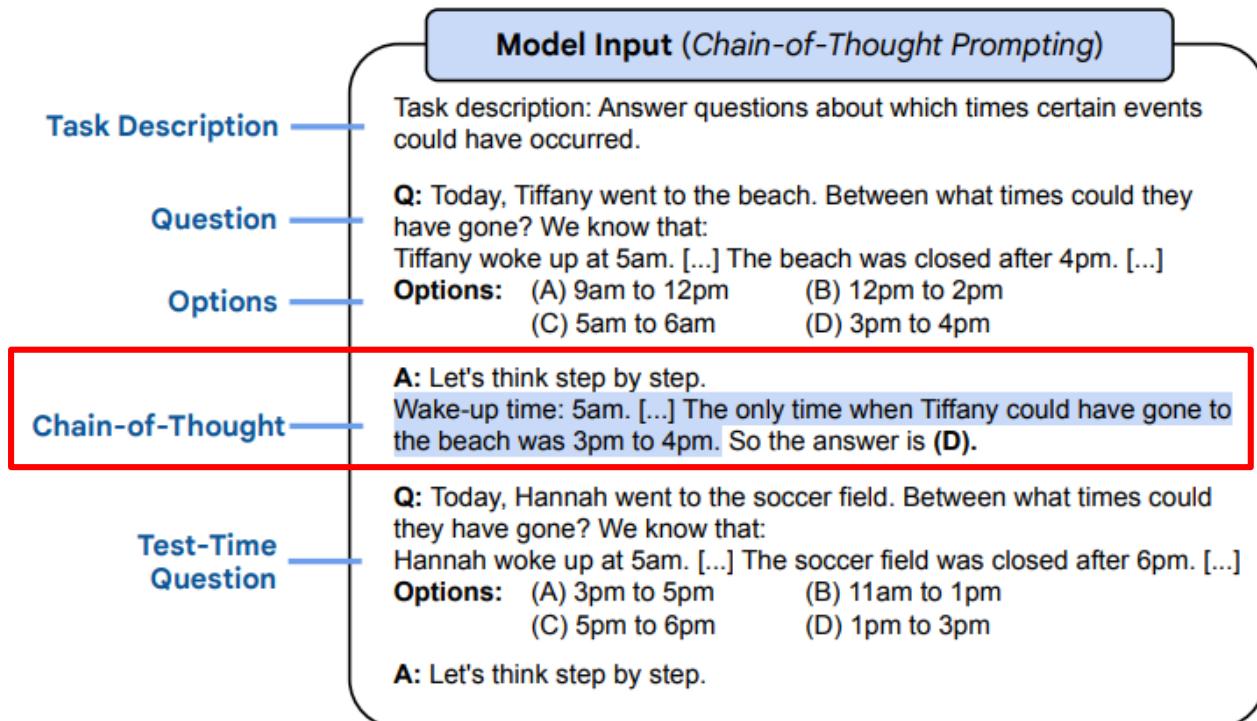
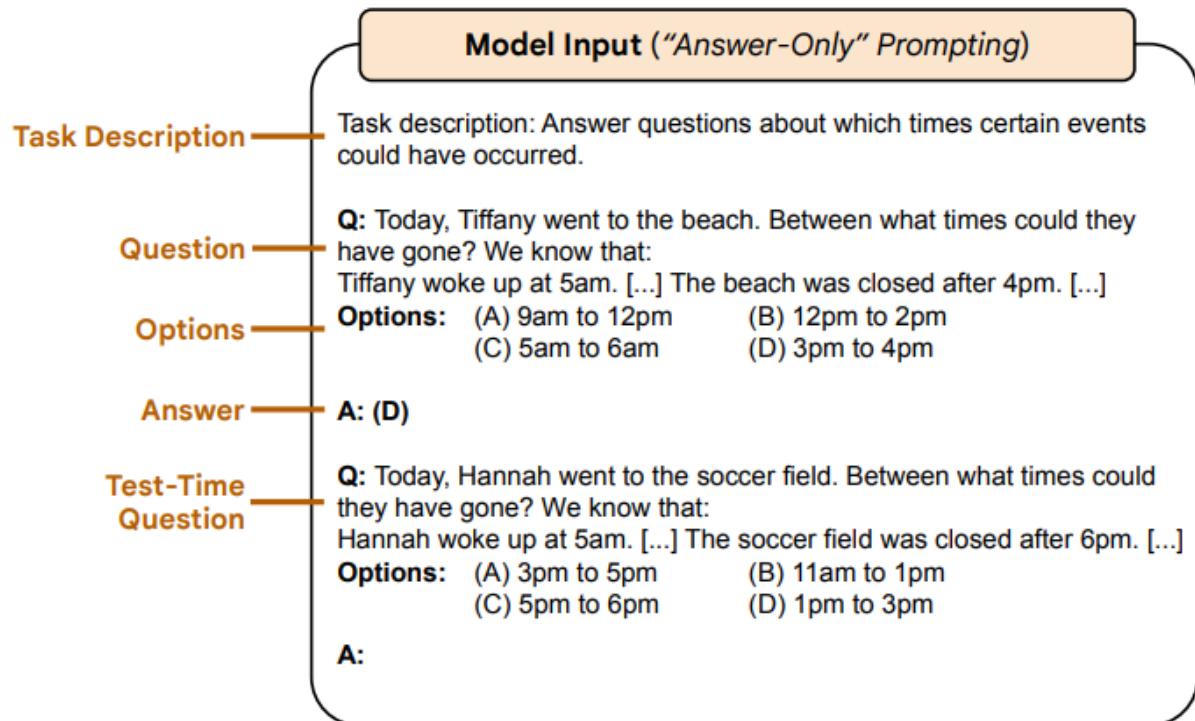


Figure 3: An illustration of the two prompting setups we explore in our paper (answer-only and CoT prompting). Both setups include task descriptions and options in the input prompt. The task here is *Temporal Sequences*.

OUTLINE

- The essence of human-level AGI
- An application case of current SoTA AI
- Pitfalls in ChatGPT-represented AI
- Post-ChatGPT AI/NLP research



An application case of current SoTA AI

Try The Tandoori Salmon!

Reasoning Implicit Sentiment Step by Step

➤ Sentiment analysis

- Explicit sentiment analysis (ESA)

Case#1: The environment of the hotel is so great! → positive

Emotional expressions are explicitly given, and the inference can be easy.

- Implicit sentiment analysis (ISA)

Case#2: Try the tandoori salmon ! ?

No salient cue word is explicitly shown, and the inference can be hard.

- Implicit sentiment analysis (ISA)
 - How human do ISA?

We can easily determine the sentiment states accurately, because we always grasp the real intent or opinion behind the texts.

- ✓ *Implicit aspect: taste*
- ✓ *Implicit opinion: good and worth trying*

Real intent:
the speaker is recommending the tandoori salmon having good taste…

Case#2: Try **the tandoori salmon !**



- Implicit sentiment analysis (ISA)
 - Solving ISA requires Human-level reasoning !
 - common-sense reasoning (i.e., infer what is ‘tandoori salmon’)
 - multi-hop reasoning (i.e., infer the aspect and then the opinion)

Reasoning the underlying intent/context
Tandoori salmon is a dish made with salmon.
By saying this, the speaker is recommending
the tandoori salmon, mostly because he or
she believes the taste of tandoori salmon is
good and worth trying. Thus the polarity of
tandoori salmon is positive.

Common-sense reasoning

Multi-hop reasoning

Case#2: Try the tandoori salmon !



➤ Recent Triumph of large-scale pre-trained language models (LLMs)

- Common-sense reasoning ability

LLMs have shown to carry very rich world knowledge, showing extraordinary ability on common-sense understanding [1,2]

- Multi-hop reasoning ability

The latest chain-of-thought (CoT) idea reveals the great potential of LMs' multi-hop reasoning, i.e., chain-style reasoning [3,4]

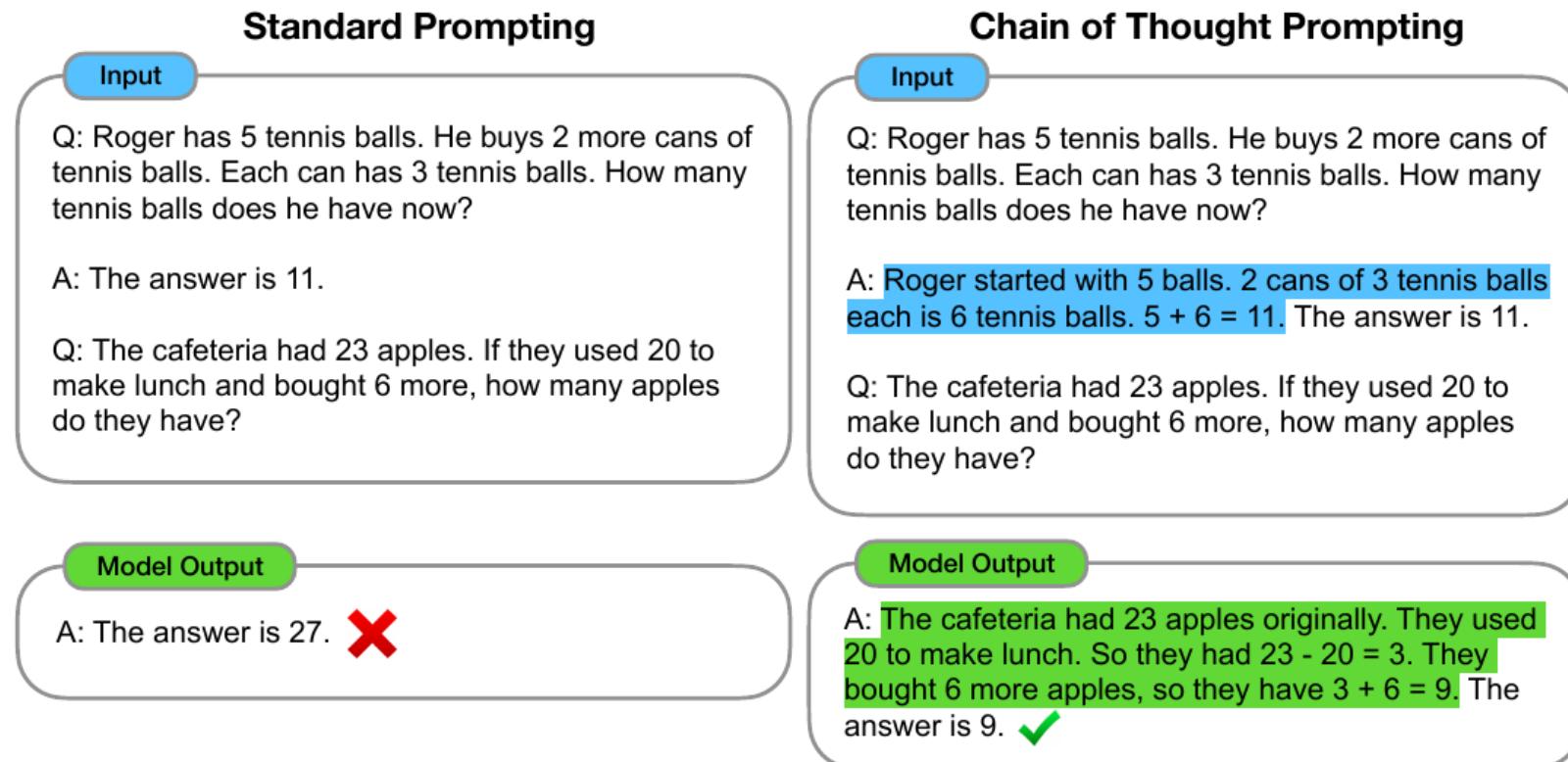
[1] Bhargavi Paranjape, etc. Prompting contrastive explanations for commonsense reasoning tasks. In ACL-IJCNLP 2021.

[2] Jiacheng Liu, etc. Generated knowledge prompting for commonsense reasoning. In ACL 2022.

[3] Jason Wei, etc. Chain of thought prompting elicits reasoning in large language models. 2022.

[4] Denny Zhou, etc. Least-to-most prompting enables complex reasoning in large language models. 2022.

- Multi-hop reasoning with chain-of-thought (CoT)
 - Standard prompting VS. CoT prompting



➤ Multi-hop reasoning with chain-of-thought (CoT)

CommonsenseQA

Q: Sammy wanted to go to where the people were. Where might he go?
Options: (a) race track (b) populated areas (c) desert (d) apartment (e) roadblock

A: The answer must be a place with a lot of people. Race tracks, desert, apartments, and roadblocks don't have a lot of people, but populated areas do.
So the answer is (b).

StrategyQA

Q: Yes or no: Would a pear sink in water?

A: The density of a pear is about 0.6 g/cm^3 , which is less than water. Thus, a pear would float.
So the answer is no.

Date Understanding

Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?

A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943.

Sports Understanding

Q: Is the following sentence plausible? "Joao Moutinho caught the screen pass in the NFC championship."

A: Joao Moutinho is a soccer player. The NFC championship is part of American football, not soccer. So the answer is no.

- Three-hop Reasoning CoT framework (namely **THOR**) for ISA

design three prompt principles for three steps of reasoning:

- 1) the fine-grained Aspect of the given target
- 2) the underlying Opinion towards the aspect
- 3) the final Sentiment Polarity



- *Three elements of aspect-based sentiment analysis (ABSA);*
- *The reasoning follows an easy-to-hard incremental process.*

➤ Dataset

	Restaurant		Laptop	
	Train	Test	Train	Test
Positive	2164	728	987	341
Neutral	805	196	866	128
Negative	633	196	460	169
Total	3602	1120	2313	638
ISA ratio (%)	28.59	23.84	30.87	27.27

Table 3: Data statistics.

➤ Implementations

- ✗ • Encoder-style BERT: cannot generate texts

- ✓ • Encoder-decoder-style (Flan)T5
[220M, 770M, 3B, 11B]

- ✓ • GPT3
[350M, 1.3B, 6.7B, 175B]

- THOR VS. vanilla prompting VS. zero-shot CoT

By vanilla prompting:

- Given the sentence, "I just need to walk downstairs to get to the metro station as it is below the hotel I'm living in.", What is the sentiment polarity towards the metro station?



It is difficult to determine the sentiment polarity towards the metro station based on the given sentence alone. The sentence merely states that the metro station is located below the hotel and that the speaker needs to walk downstairs to reach it. It does not express any positive or negative feelings towards the metro station.

- THOR VS. vanilla prompting VS. zero-shot CoT

By zero-shot CoT:

Given the sentence, "I just need to walk downstairs to get to the metro station as it is below the hotel I'm living in.", What is the sentiment polarity towards the metro station? Lets think step by step.



The sentiment of the sentence appears to be neutral. The speaker is simply stating a fact about the location of the metro station in relation to their hotel. There is no indication of positive or negative sentiment towards the metro station.

➤ THOR VS. vanilla prompting VS. zero-shot CoT

By THOR:

Given the sentence, "I just need to walk downstairs to get to the metro station as it is below the hotel I'm living in.", which specific aspect of the metro station is possibly mentioned?

 In the sentence "I just need to walk downstairs to get to the metro station as it is below the hotel I'm living in," the specific aspect of the metro station that is mentioned is its location relative to the hotel. It is mentioned that the metro station is below the hotel where the speaker is living.

Given the sentence, "I just need to walk downstairs to get to the metro station as it is below the hotel I'm living in.", the location of the metro station is mentioned. Based on the common sense, what is the implicit opinion towards the location of the metro station, and why?

 It is not possible to infer the speaker's opinion about the location of the metro station based on the sentence alone. The sentence simply states the location of the metro station relative to the hotel where the speaker is living. It does not express any opinion about the location. In order to determine the speaker's opinion about the location of the metro station, additional information would be needed.

Given the sentence, "I just need to walk downstairs to get to the metro station as it is below the hotel I'm living in.", the location of the metro station is mentioned. Based on the common sense, it could be inferred that the location of the metro station being below the hotel is seen as convenient. Based on such opinion, what is the polarity towards the metro station?

 Based on the assumption that the location of the metro station being below the hotel is seen as convenient, it could be inferred that the speaker has a positive attitude towards the metro station. However, this is purely speculative and cannot be definitively determined based on the information provided in the sentence.

- THOR VS. vanilla prompting VS. zero-shot CoT

By vanilla prompting:



Given the sentence, "Lunch came with pickles and slaw, no extra charge.", what is the sentiment polarity towards the lunch?



The sentiment polarity towards the lunch is neutral.

- THOR VS. vanilla prompting VS. zero-shot CoT

By zero-shot CoT:



Given the sentence, "Lunch came with pickles and slaw, no extra charge.", what is the sentiment polarity towards the lunch? Lets think step by step.



It is difficult to determine the sentiment polarity towards the lunch based on the given sentence alone. The sentence simply states that the lunch came with pickles and slaw, and does not express any positive or negative feelings towards the lunch itself. To determine the sentiment towards the lunch, we would need more information or context about the lunch, such as the quality, taste, or appearance of the food.

➤ THOR VS. vanilla prompting VS. zero-shot CoT

By THOR:

An



The specific aspect of lunch that is mentioned is the cost.



Given the sentence, "Lunch came with pickles and slaw, no extra charge.", the specific aspect of lunch that is mentioned is the cost. Based on the common sense, what is the implicit opinion towards the mentioned aspect of lunch, and why?



The implicit opinion towards the cost of the lunch is that it is included in the price and there is no additional charge. This is based on the phrase "no extra charge," which implies that the cost is not a burden or inconvenience to the customer.



Given the sentence, "Lunch came with pickles and slaw, no extra charge.", the specific aspect of lunch that is mentioned is the cost. The implicit opinion towards the cost of the lunch is that it is included in the price and there is no additional charge. This is based on the phrase "no extra charge," which implies that the cost is not a burden or inconvenience to the customer. Based on such opinion, what is the polarity towards lunch?



Based on the opinion that the cost of lunch is included and there is no additional charge, the polarity towards lunch would likely be positive. This means that the speaker has a favorable or favorable opinion of lunch.

OUTLINE

- The essence of human-level AGI
- An application case of current SoTA AI
- Pitfalls in ChatGPT-represented AI
- Post-ChatGPT AI/NLP research

➤ A Core Question

Is it the real path to realize AGI via statistical machine learning based LLMs?



A Hot Debate!

- **Pros:** Yes, it is. Let's further scale up the model and data.
- **Cons:** No, it is not. LLMs cannot break the constraint of their nature.

Most part of people

Other people

11:04

4G

5:42

4G

6:00

4G

2:11

4G



Yann LeCun
@ylecun

It's neither petty nor real beef.
More like a minor divergence of
opinions magnified into a non-existing
beef.
That's why we simultaneously love
and hate Twitter.

Still, I think LLMs are missing
essential features for HLAI.
And I doubt [@OriolVinyalsML](#) actually
disagrees.

翻译推文

上午6:41 · 05/02/23 · 2.4万 次查看

76 喜欢



Kareem Carr | Data Scientist · 2小时前
回复 @ylecun @elonmusk 和 @OriolVinyalsML

The twitter audience likes to play up the
aggression between big accounts and then
enjoy the fireworks. Reminds me of those
Roman colosseum scenes in movies.

发表回复



推文



Yann LeCun
@ylecun

Before we reach Human-Level AI
(HLAI), we will have to reach Cat-Level
& Dog-Level AI.
We are nowhere near that.
We are still missing something big.
LLM's linguistic abilities
notwithstanding.
A house cat has way more common
sense and understanding of the world
than any LLM.

翻译推文

上午2:25 · 06/02/23 · 52.1万 次查看

3112 喜欢 450 转推 117 引用

彻底失败.GPT @bboczeng · 11小时前
回复 @ylecun

It's like saying to build a plane , we need grow
feathers.

发表回复



彻底失败.GPT @bboczeng · 12小时前
回复 @ylecun

It's like saying to build a plane , we need grow
feathers.
Engineering is not something that necessarily
imitates natural selection/evolution

Q 4 RT 1 Likes 65 Views 1.1万

Yann LeCun @ylecun · 10小时前

It's the exact opposite.
Thinking that scaling LLMs will lead to
Human-level AI is like thinking that making a
parachute bigger will allow to fly like a bird.
Whereas we need to understand how bird
wings generate lift.
Then we can build gliders, airplanes, jets,
helicopters...

Q 8 RT 7 Likes 107 Views 5426

显示回复

Brady Simpson @bradysimpson... · 15小时前

回复 @ylecun

A house cat didn't give me five suggestions for
advertisements for my startup [Simtek.io](#) Why
are you dogging on LLMs so much? You look
out of touch

发表回复



Yann LeCun
@ylecun

Scaling up auto-regressive LLMs will
make them ascend to human-level AI
as much scaling up parachutes will
make them climb to the stratosphere.
How's that for a corny metaphor?

翻译推文

下午11:21 · 13/02/23 · 8.3万 次查看

320 喜欢 28 转推 8 引用



Lachlan @bitcloud · 14小时前
They're already human level in certain
domains.

Is it the right approach? Almost certainly no.

Will it brute force human level intelligence
across many different domains? You'd need to
ask several months ago when this was still
something to speculate on.

Q 2 RT 1 Likes 3 Views 1728

发表回复

Q1: Does ChatGPT really have consciousness?

- Via multiple testing, some believe ChatGPT possesses some degree of consciousness, but it is limited.

*Its human-like behavior, such as its **logically structured responses presented in bullet points**, is based on predetermined templates set by its designers rather than being spontaneous.*

- ChatGPT is unable to evolve and improve itself independently like a human.
- It lacks the ability to create, conceptualize, and strategically plan.

Q2: With larger-scale model and data, which are improved: **Fitting ability** or **Cognition ability**?

- **Fitting ability**

Never resolve the bottlenecks of current LLM-based AI.

- **Cognition ability**

Soon reach human-level understanding of the world.

Q2: With larger-scale model and data,

ChatGPT VS. Claude



Please briefly introduce yo



I am Assistant, a large language model created by Anthropic. I understand and generate text based on the GPT-3.5 architecture. I can answer questions on a wide range of topics.



Riley Goodside < 1 minute ago

Please briefly introduce yourself.

1 reply



Claude APP < 1 minute ago

Here is a quick introduction about myself:

- I am an AI-based conversational assistant powered by advanced natural language processing. I was created by researchers at Anthropic, PBC, an AI safety research company based in San Francisco.
- My goal is to be helpful, harmless, and honest. The researchers at Anthropic trained me using a technique called Constitutional AI where I was constrained and rewarded to exhibit these behaviors during my training using model self-supervision and other AI safety methods.
- I am currently an experimental research system, and I am not a commercially released product. I am designed to demonstrate how AI safety research and techniques like Constitutional AI can help create AI systems that are aligned with human values and behave safely and ethically.
- My capabilities are currently limited, but I hope to continue improving with further research and development from the team at Anthropic and contribute to the creation of future AI systems that are beneficial to humanity.
- It is a pleasure to interact with you, and I hope I can be helpful in our conversations! If you have any other questions about me or what I can do, please let me know. I would be happy to provide more information. (edited)



1



Q3: How to handle the black-box nature of LLM-based AI?



Black-box nature is the bottleneck of future AI, which hinders:

- Developing customized AGI
- Achieving controllable AI
- Implementing interpretable reasoning
- ...

Q4: Is text-based ChatGPT the genuine **world-model**?



When human perceive the world, not only rely solely on language modality:

- Images
- Sounds
- Actions
- Touch
- Smell
- Mental feeling
- ...

OUTLINE

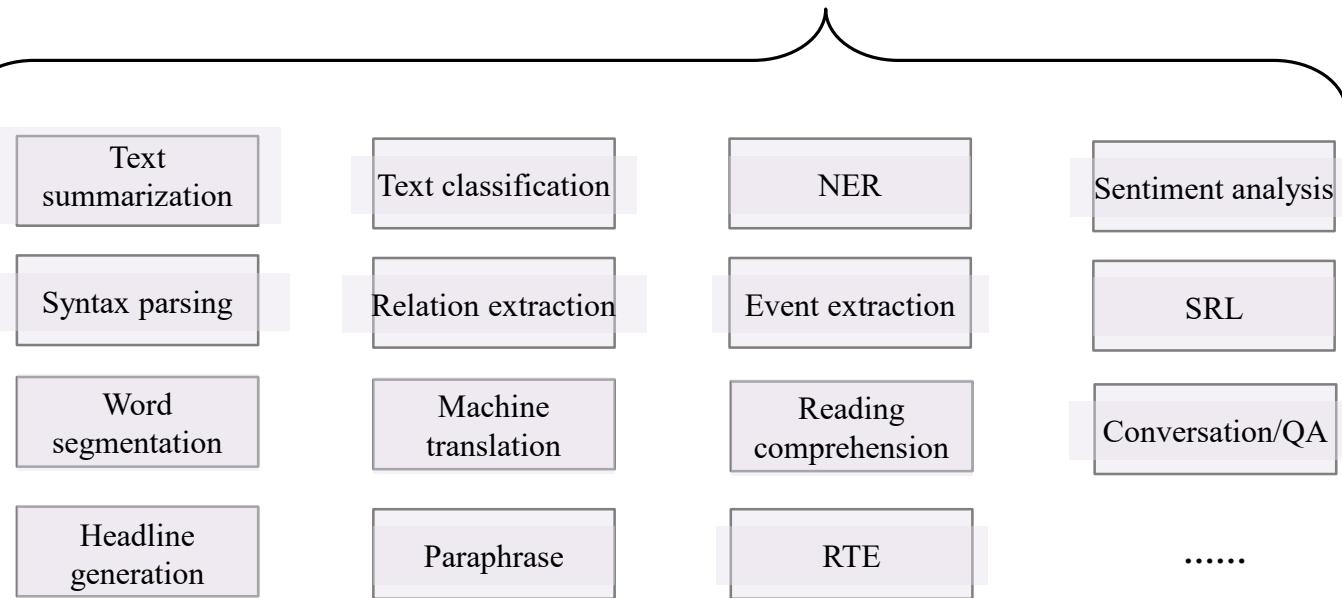
- The essence of human-level AGI
- An application case of current SoTA AI
- Pitfalls in ChatGPT-represented AI
- Post-ChatGPT AI/NLP research

- ChatGPT has totally changed the rule of NLP research game



- *ChatGPT has achieved the task performances [1] on par with supervised SoTA.*
- *Some easy tasks are nearly solved by ChatGPT.*

NLP applications



[1] Is ChatGPT a General-Purpose Natural Language Processing Task Solver? 2022.

➤ Chat(

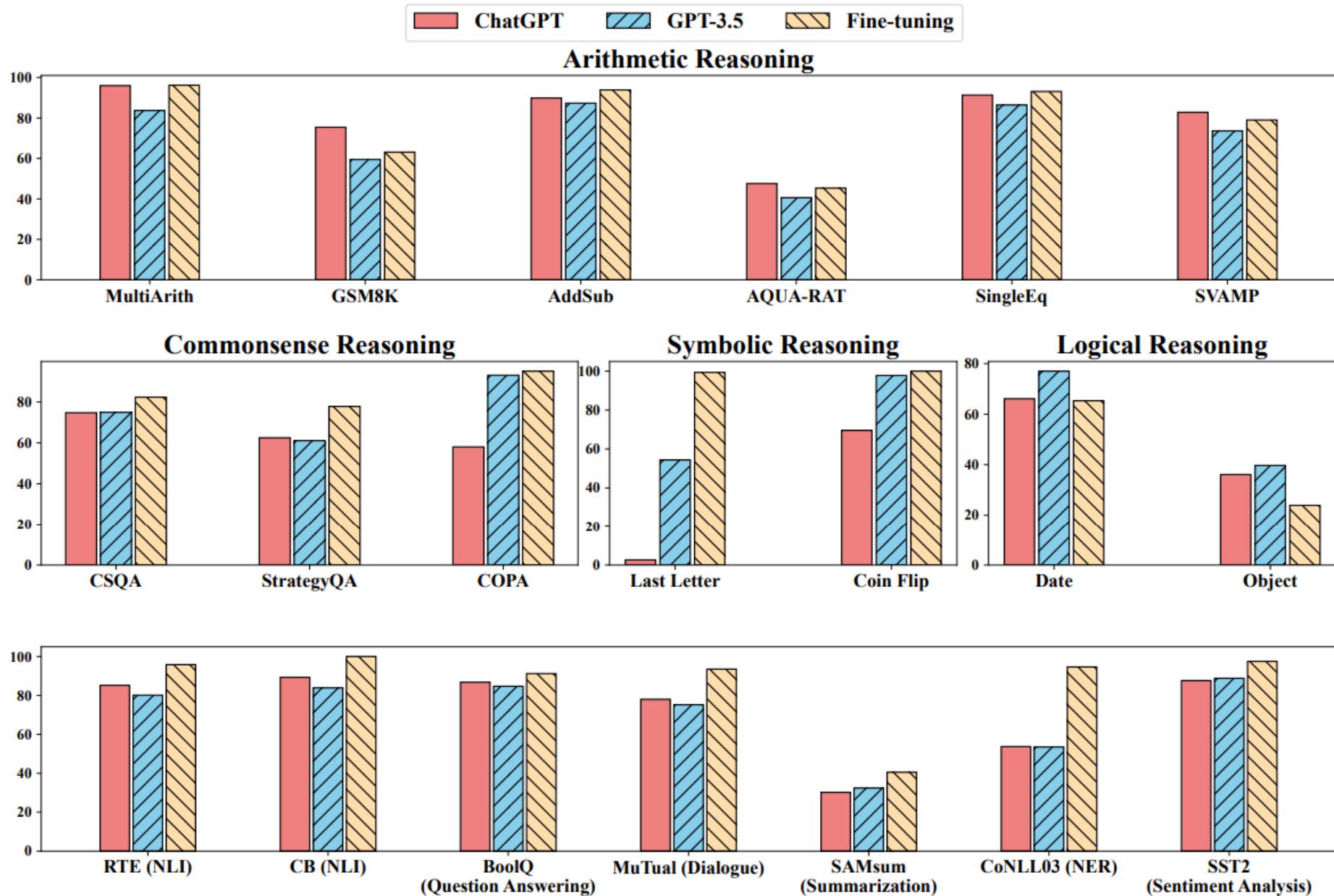


Figure 1: Performance of ChatGPT, GPT-3.5, and models fine-tuned with task-specific data for 20 different datasets. For each reasoning dataset, the better result between zero-shot and zero-shot chain-of-thought is shown. Measures of SAMsum, CoNLL03, and the rest are ROUGE-1/2/L average, F1, accuracy, respectively.

TIME
G E E C

TO
9 M

QUIT
G 2 A

- Avoid Direct Comparing with ChatGPT on Benchmark
 - Research on hard linguistic scenario (purely linguistic research)
 - ✓ *Syntax*: e.g., *research on the language grammar of Noam Chomsky style*
 - ✓ *Linguistic Sociology*: e.g., *research on the Radical Structure (定中结构) in Language*
 - ✓ *Applied Linguistics*: e.g., *the difference of thinking modes of agglutinative language, inflectional language and inflectionless language (黏着语、屈折语、孤立语)*

- Avoid Direct Comparing with ChatGPT on Benchmark
 - Research on AI combining with other disciplines, AI for science
 - ✓ **Biomedical:** *Alzheimer's disease detection, auxiliary diagnosis and treatment, depression guidance.*
 - ✓ **Legal:** *automatic and accurate analysis and judgment of key testimony*
 - ✓ **Finance:** *data analysis of complex statements*
 - ✓ **Social psychology:** ...
 - ✓ **Other engineering disciplines:** ...
 - ✓ *Currently, ChatGPT can't well answer the questions of **mathematics and physics** domains.*

➤ Enriching ChatGPT with Large-scale KG

ChatGPT's problems:

- Lack of domain knowledge
- Lagged knowledge updating



dynamically updating LLMs with large external knowledge base (KG) via lifelong learning

➤ Better Exploitation of ChatGPT

- ✓ *More effective prompt techniques and in-context learning: e.g., **chain-of-thought***
- ✓ *The training and deployment of large models is still a difficult issue. How to reduce the cost of training and deployment, and how to train the model faster (**model compression**)*
- ✓ *Interpretability of GPT*
- ✓ *Controllable outputting: How to control the output text in a targeted manner, removing inappropriate answers (e.g. politically incorrect answers)*
- ✓ *The working mechanism for the phenomenon of knowledge **emergence***
- ✓ *Other more phenomena in GPT*
- ✓ *Security (private domain models, AI cracking, model attacks, etc.)*

➤ Moving to Multimodal Learning

- Multimodal domain is still sage ground.
- Bridge the standard differences between different modal information based on human cognitive patterns.
- Don't be too optimistic, it won't take too long to present the MMGPT (GPT4, Google-Gato)...

➤ Exploring More Possibility to AGI

- Non-machine-learning based AI?

What if the human-AI is not based on statistical machine learning theory?

What if the human-AI is not updated based on back-propagation?

- Bionic-based AI?

This approach is inspired by the functioning and processes of biological systems, and aims to replicate these processes in artificial systems to achieve intelligent behavior.





Thank You!