

Rethinking Neural Symbolic Computing

神经符号计算的再思考

LIU Qun 刘群

Huawei Noah's Ark Lab 华为诺亚方舟实验室

A talk at 全国少数民族自然语言处理青年论坛 (online)

2022-12-09



NOAH'S ARK LAB



Content

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

Content

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

Content

神经符号计算简介

什么是神经符号计算

神经方法和符号方法的结合方式

修改状态行页号编码的显示形式

Wikipedia: Neural-Symbolic AI

WIKIPEDIA

Neuro-symbolic AI

Neuro-symbolic AI integrates neural and symbolic AI architectures to address complementary strengths and weaknesses of each, providing a robust AI capable of reasoning, learning, and cognitive modeling. As argued by Valiant^[1] and many others,^[2] the effective construction of rich computational cognitive models demands the combination of sound symbolic reasoning and efficient machine learning models. Gary Marcus, argues that: "We cannot construct rich cognitive models in an adequate, automated way without the triumvirate of hybrid architecture, rich prior knowledge, and sophisticated techniques for reasoning."^[3] Further, "To build a robust, knowledge-driven approach to AI we must have the machinery of symbol-manipulation in our toolkit. Too much of useful knowledge is abstract to make do without tools that represent and manipulate abstraction, and to date, the only machinery that we know of that can manipulate such abstract knowledge reliably is the apparatus of symbol-manipulation."^[4]

Content

神经符号计算简介

什么是神经符号计算

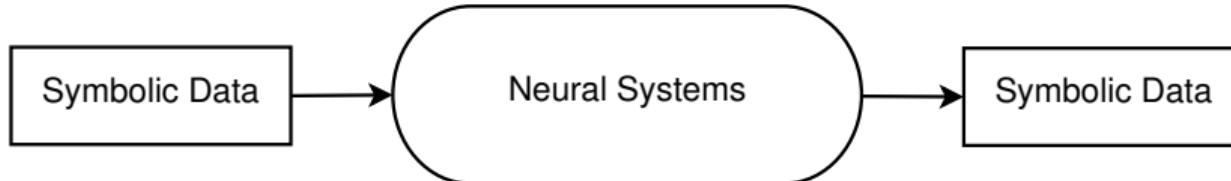
神经方法和符号方法的结合方式

Wikipedia: Neural-Symbolic AI: Kinds of approaches

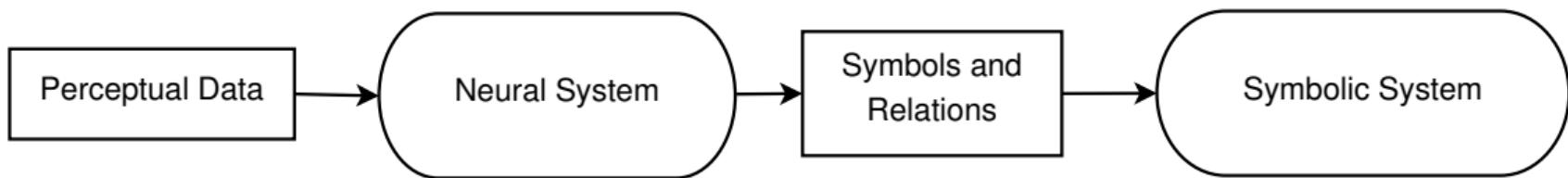
- **Symbolic Neural symbolic**—is the current approach of many neural models in natural language processing, where words or subword tokens are both the ultimate input and output of large language models. Examples include BERT, RoBERTa, and GPT-3.
- **Symbolic[Neural]**—is exemplified by AlphaGo, where symbolic techniques are used to call neural techniques. In this case the symbolic approach is Monte Carlo tree search and the neural techniques learn how to evaluate game positions.
- **Neural|Symbolic**—uses a neural architecture to interpret perceptual data as symbols and relationships that are then reasoned about symbolically. The Neural-Concept Learner^[8] is an example.
- **Neural:Symbolic → Neural**—relies on symbolic reasoning to generate or label training data that is subsequently learned by a deep learning model, e.g., to train a neural model for symbolic computation by using a Macsyma-like symbolic mathematics system to create or label examples.
- **Neural_{Symbolic}**—uses a neural net that is generated from symbolic rules. An example is the Neural Theorem Prover,^[9] which constructs a neural network from an AND-OR proof tree generated from knowledge base rules and terms. Logic Tensor Networks^[10] also fall into this category.
- **Neural[Symbolic]**—allows a neural model to directly call a symbolic reasoning engine, e.g., to perform an action or evaluate a state.

Symbolic Neural Symbolic, Neural|Symbolic

Symbolic Neural Symbolic

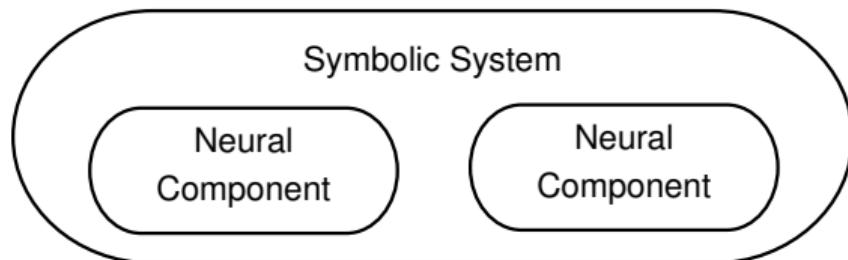


Neural|Symbolic

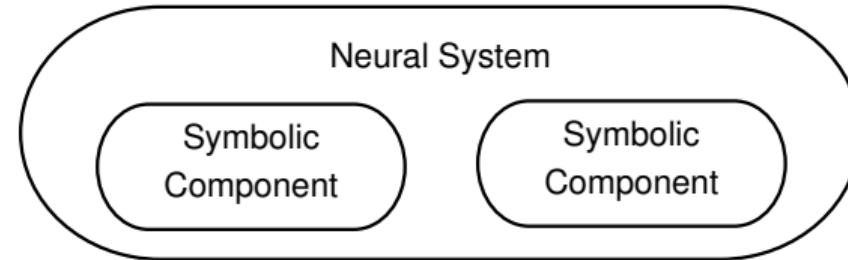


Symbolic[Neural], Neural[Symbolic]

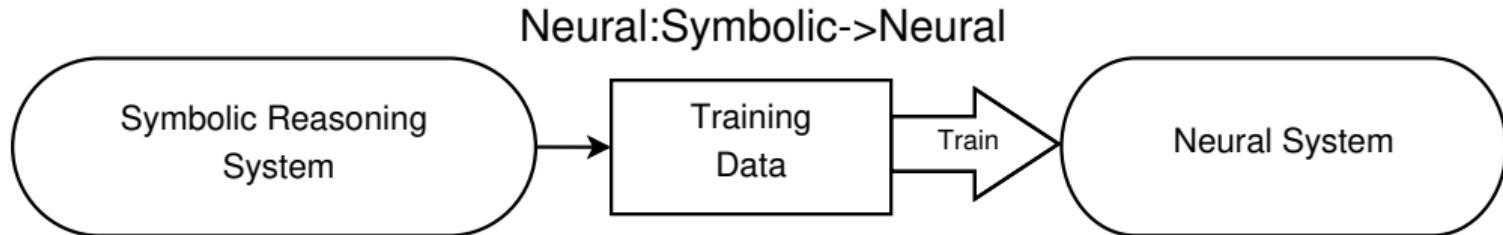
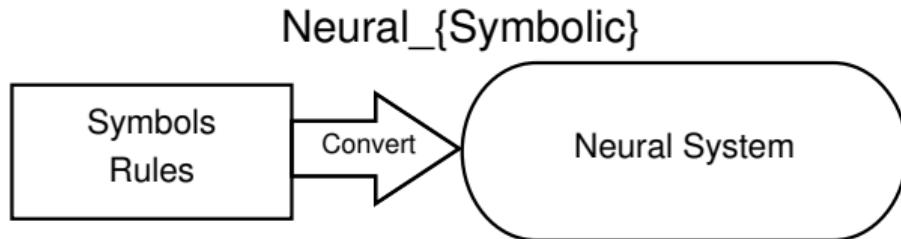
Symbolic[Neural]



Neural[Symbolic]



Neural_{Symbolic}, Neural:Symbolic->Neural



Content

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

REVIEW

Deep learning

Yann LeCun^{1,2}, Yoshua Bengio³ & Geoffrey Hinton^{4,5}

The issue of representation lies at the heart of the debate between the logic-inspired and the neural-network-inspired paradigms for cognition. In the logic-inspired paradigm, an instance of a symbol is something for which the only property is that it is either identical or non-identical to other symbol instances. It has no internal structure that is relevant to its use; and to reason with symbols, they must be bound to the variables in judiciously chosen rules of inference. By contrast, neural networks just use big activity vectors, big weight matrices and scalar non-linearities to perform the type of fast ‘intuitive’ inference that underpins effortless commonsense reasoning.

Lecun, Bengio and Hinton, Deep Learning, Nature, Vol.521, 2015

Hinton的观点

REVIEW

Deep learning

Yann LeCun^{1,2}, Yoshua Bengio³ & Geoffrey Hinton^{4,5}

- ▶ 表征问题是逻辑启发和神经网络启发的认知范式之间争论的核心。
- ▶ 在逻辑启发的范式中，一个符号实例的唯一属性是它与其他符号实例相同或不同。它没有与它的使用相关的内部结构；要用符号推理，它们必须绑定到精心选择的推理规则中的变量。
- ▶ 相比之下，神经网络只是使用大活动向量、大权重矩阵和标量非线性来执行快速“直观”推理，支持轻松的常识推理。

Lecun, Bengio and Hinton, Deep Learning, Nature, Vol.521, 2015

Hinton的观点

AI pioneer Geoff Hinton: “Deep learning is going to be able to do everything”

Thirty years ago, Hinton’s belief in neural networks was contrarian. Now it’s hard to find anyone who disagrees, he says.

There are some people who still believe that symbolic representation is one of the approaches for AI.

Absolutely. I have good friends like Hector Levesque, who really believes in the symbolic approach and has done great work in that. I disagree with him, but the symbolic approach is a perfectly reasonable thing to try. But my guess is in the end, we’ll realize that symbols just exist out there in the external world, and we do internal operations on big vectors.

Hao, K. AI pioneer Geoff Hinton: “Deep learning is going to be able to do everything.”
MIT Technology Review (2020).

Hinton的观点

AI pioneer Geoff Hinton: “Deep learning is going to be able to do everything”

Thirty years ago, Hinton's belief in neural networks was contrarian. Now it's hard to find anyone who disagrees, he says.

Q: 有些人仍然认为符号表示是人工智能的方法之一。

A: 当然。我有像赫克托·莱维斯克这样的好朋友，他真的相信符号方法，并在这方面做了很好的工作。

A: 我不同意他的观点，但符号方法是一个完全合理的尝试。

A: 但我的猜测是，最终，我们会意识到符号只是存在于外部世界中，我们在内部运算时我们只是对大向量进行操作。

Hao, K. AI pioneer Geoff Hinton: “Deep learning is going to be able to do everything.”

MIT Technology Review (2020).

Marcus和LeCun的辩论

- **2017-10-21** Yann LeCun vs. Gary Marcus
Artificial Intelligence Debate - Does AI Need More Innate Machinery?
- **2018-01-02** Gary Marcus
Deep Learning: A Critical Appraisal
- **2022-03-10** Gary Marcus
Deep Learning Is Hitting a Wall
- **2022-06-16** Jacob Browning and Yann LeCun
What AI Can Tell Us About Intelligence
- **2022-08-11** Gary Marcus
Deep Learning Alone Isn't Getting Us To Human-Like AI
- **2022-10-17** Gary Marcus
Three baffling claims about AI and machine learning in four days, statistical errors in top journals, and claims from Yann LeCun that you should not believe.

Marcus和Lecun的辩论



Artificial Intelligence Debate - Yann LeCun vs. Gary Marcus - Does AI Need More Innate Machinery?



The Artificial Intelligence Channel
115K subscribers

Subscribed



323



Share



Download



Clip



Save



2017-10-21 <https://www.youtube.com/watch?v=aCCotxqxFsk>

Marcus和Lecun的辩论

Limits on the scope of deep learning:

1. Deep learning thus far is data hungry
2. Deep learning thus far is shallow and has limited capacity for transfer
3. Deep learning thus far has no natural way to deal with hierarchical structure
4. Deep learning thus far has struggled with open-ended inference
5. Deep learning thus far is not sufficiently transparent
6. Deep learning thus far has not been well integrated with prior knowledge
7. Deep learning thus far cannot inherently distinguish causation from correlation
8. Deep learning presumes a largely stable world, in ways that may be problematic
9. Deep learning thus far works well as an approximation, but its answers often cannot be fully trusted
10. Deep learning thus far is difficult to engineer with

Gary Marcus, Deep Learning: A Critical Appraisal, 2018-01-02

Marcus和Lecun的辩论

深度学习的天生缺陷:

1. 深度学习至今缺少足够的数据。
2. 深度学习至今仍不够深入，且在迁移度上存在很大局限。
3. 现在的深度学习并没有能够处理层次化结构的方法。
4. 深度学习至今无法解决开放性的推理问题。
5. 深度学习还不够透明。
6. 深度学习尚未能很好地结合先验知识。
7. 深度学习还无法区分“因果关系”和“相关性”。
8. 深度学习在一个环境稳定的世界里表现最好，然而现实往往并非如此。
9. 当你需要一个近似的答案时，深度学习效果不错，但不能完全信赖这些结果。
10. 深度学习仍很难被工程化。

Gary Marcus, Deep Learning: A Critical Appraisal, 2018-01-02

Marcus和Lecun的辩论

深度学习撞墙了：

- ▶ AI领域充满炒作和虚张声势
 - ▶ 一些难题无法解决：放射学、自动驾驶、不良语言和错误信息。
 - ▶ Scaling Law并不是自然定律，而是观察到的现象（类似摩尔定律）。
 - ▶ 深度学习遇到了Scaling Limits（扩展限制）。
- ▶ 回顾历史，Hinton等深度学习一直在强烈反对符号方法。
- ▶ 轻视尚未经过充分探索的过时想法是不正确的。
 - ▶ Hinton 说得很对，过去人工智能研究人员试图埋葬深度学习。
 - ▶ 但是 Hinton 在今天对符号处理做了同样的事情。在我看来，他的对抗损害了这个领域。
 - ▶ 具有讽刺意味的是，Hinton 是 George Boole 的玄孙，而 Boolean 代数是符号 AI 最基本的工具之一，是以他的名字命名。
- ▶ 我认为，混合人工智能是最好的方向。
- ▶ 将神经和符号结合在一起的探索一直都没有停止，而且正在积聚力量。

Gary Marcus, Deep Learning Is Hitting a Wall, 2022-03-10

Marcus和Lecun的辩论

符号操作是否需要被硬编码，还是可以被学习到？

- ▶ 深度学习曾经遇到很多很多困难，都被克服了：
 - ▶ 非线性函数、算力不足、标注数据不足
 - ▶ 符号推理？
- ▶ 问题的核心在于：符号推理是一开始就被硬编码，还是可以通过经验学习得到符号推理的能力？
 - ▶ 现在的语言模型如GPT-3和LaMDA已经学到了某种符号推理能力，虽然表现还不是很可靠
 - ▶ Marcus假设符号推理能力是全有或者全无的（爬树到不了月球）
 - ▶ Hinton等人认为神经网络不需要符号硬编码也能学习到操作符号的能力
 - ▶ 拒绝将两种模式混合并非草率的，而是基于一个人是否认为符号推理可以学习的哲学性差异。
- ▶ 人类思想的底层逻辑：符号能力是先天的，还是后天习得的？
- ▶ 对AI，是押注，还是做空？
 - ▶ 混合模型一直有人在研究，但并没有取得成功。
 - ▶ 相反，深度学习一直在突破。
 - ▶ 深度学习达到上限了吗？

Jacob Browning and Yann LeCun, What AI Can Tell Us About Intelligence, 2022-06-16

Marcus和Lecun的辩论

仅仅依靠深度学习无法把我们带向像人一样的AI：

- ▶ Browning & Lecun的文章曲解了我的观点。
- ▶ 他们实际上已经承认了符号处理的合理性，但开始说符号处理是后天习得的而非与生俱来的。
- ▶ 人们真正应该思考和质疑的是深度学习的极限：
 - ▶ 深度学习已经面临原则上的挑战，即组合性、系统性和语言理解问题。
 - ▶ 这类技术缺乏表征因果关系（例如疾病与其症状之间关系）的方法，并且可能在获取抽象概念方面存在挑战。深度学习没有明显的逻辑推理方式，距离整合抽象知识还有很长的路要走。
 - ▶ 深度学习在符号处理方面取得了一些进展，但还远远不够。
 - ▶ 他们只是用归纳的原理说明深度学习的作用：「由于深度学习已经克服了 1 到 N 的问题，我们应该相信它可以克服 N+1 的问题」。这种观点说服力很弱。
 - ▶ 简单地扩展（scaling，增加层数和训练数据）是不够的，B&L也承认这一点。
- ▶ 现在即使是神经网络最狂热的支持者已经认识到符号处理对实现 AI 的重要性。

Gary Marcus, Deep Learning Alone Isn't Getting Us To Human-Like AI, 2022-08-11

Manning的评论



Christopher Manning
@chrmanning

...

@ylecun & J Browning's What AI Can Tell Us About Intelligence in Noëma is excellent! It clearly & dispassionately contrasts two main views on the place of symbols, as hard-coded at the outset or learned through experience, arguing well for the latter.



noemamag.com

What AI Can Tell Us About Intelligence | NOEMA

Can deep learning systems learn to manipulate symbols? The answers might change our understanding of how intelligence works and what makes humans ...



Christopher Manning
@chrmanning

...

I sense some evolution in @ylecun's position—perhaps under Browning's influence; this piece suggests that “everyone working in DL agrees that symbolic manipulation is a necessary feature for creating human-like AI.” Was that really true a decade ago, or is it even true now?!?

8:38 AM · Jul 28, 2022



Yann LeCun
@ylecun

...

Replies to @chrmanning

I think it was true a decade ago, even 3 decades ago. The main debate was always about how. As for me, before attacking the question of symbol manipulation, I thought it was more urgent to figure out how to learn hierarchical representations and learn perception.

5:36 PM · Jul 28, 2022

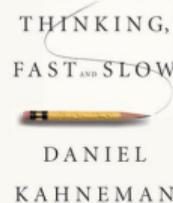
Bengio近期的工作

SYSTEM 1 VS. SYSTEM 2 COGNITION

2 systems (and categories of cognitive tasks):

System 1

- Intuitive, fast, **UNCONSCIOUS**, non-linguistic, habitual
- Current DL



System 2

- Slow, logical, sequential, **CONSCIOUS**, linguistic, algorithmic, planning, reasoning
- Future DL



Manipulates high-level / semantic concepts, which can be recombined combinatorially

Yoshua Bengio, Deep Learning For System 2 Processing, AAAI' 2019 Invited Talk

Bengio近期的工作

FACTORIZING KNOWLEDGE INTO COMPOSABLE PIECES FOR REASONING

- **Current deep learning:** homogenous architectures, **knowledge is not localized**, completely distributed
- **Transfer/continual learning:** reuse relevant pieces of knowledge; minimizes interference; maximizes reuse
- **System 2 reasoning selects and combines nameable pieces of knowledge to form thoughts** (imagined futures, counterfactual past, solutions to problems, interpretations of inputs, etc.).
- How to **factorize knowledge** into the right recomposable pieces?



11

Bengio近期的工作

HUMAN INSPIRATION FOR INDUCTIVE BIASES: IMPLICIT VS VERBALIZABLE KNOWLEDGE

- Most knowledge in our brain is implicit and **not verbalizable**
 - Some knowledge is verbalizable, we can reason and plan explicitly with it (system 2)
 - System 2 knowledge satisfies stronger assumptions
- clarify system 2 assumptions → new ML designs for abstract perception, abstract reasoning and abstract action.

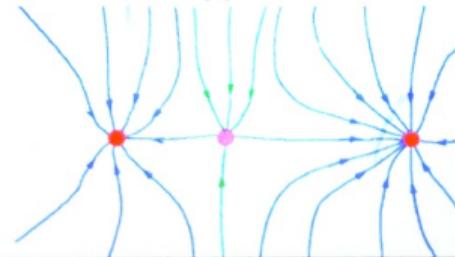


14

Bengio近期的工作

DISCRETE, SYMBOLIC, ABSTRACT CONCEPTS

- Language allows communication of simplified, DISCRETE, messages
- Thoughts manipulate such discrete entities
- Evidence that hippocampus represents discrete concepts
- Conscious ignition → convergence of global dynamics to fixed point attractor = discrete
- **The bottleneck of discretization in the communication between brain modules may further facilitate systematic generalization, making different brain modules hot-swappable for one another** (e.g. replace a noun by another in a sentence)



Bengio近期的工作

PROBABILISTIC NEURAL NETS FOR SYSTEM-2 DEEP LEARNING?

- Need generative neural nets of thoughts
- Sequence of thoughts (working memory content) = hypergraph (plausible explanations, counterfactuals, plans)
- System 2 knowledge is modularized (like classical AI rules and facts) and composable (to form these thought-graphs)
- Can we train neural nets that can implicitly represent such modular knowledge and sample causal / semantic explanations for our experiences?



Bengio近期的工作

GFLOWNETS AS THE SWISS ARMY KNIFE OF PROBABILISTIC MODELING

- A hypergraph = **set** of hyperedges (e.g. apply rule to few variables)
- Decompose **sampling of set X according to ANY ORDER** of steps generating each element
- X can have variable size or even be infinite (like all graphs)
- Learn a **sampling policy** to generate X (sampling the order too)
- Intermediate quantities, the **flows**, are implicitly **marginalizing over future choices**
- The flow function **generalizes**: no need to see all the possible sequences
- The distribution is **represented in a structured, compositional way**

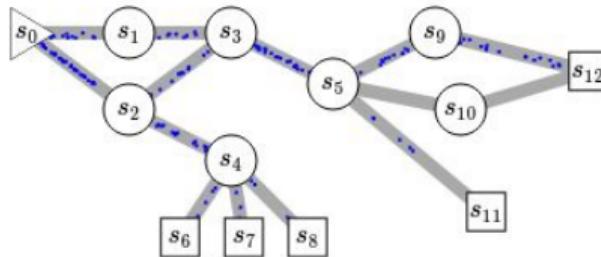


Bengio近期的工作

Interpreting the DAG as a flow network

One more visualization:

flow = #particles moving through pipes



We want a valid flow, given the rewards ("#particles") of the terminal states

Bentio et al., Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation, slides

Bengio近期的工作

MODULAR GFLOWNETS FOR HUMAN-LIKE REASONING

- Each module encapsulates knowledge (with local parameters and expertise on specific concepts)
- The GFlowNet transitions corresponding to the GWT competition between experts sample one of these modules and the content (hyperedge) it proposes as the next piece of thought
- Each module also learns an energy function corresponding to the piece of world-model it embodies
- The same fixed-size neural nets can create new abstract concepts (entities = nodes in these graphs and structural relations between them) as more data is observed.
- The entropy estimation ability of GFlowNets can drive knowledge acquisition



Yoshua Bengio, Reusable Modular and Causal Knowledge Representation for Lifelong Learning, CoLLAs2022 Invited Talk

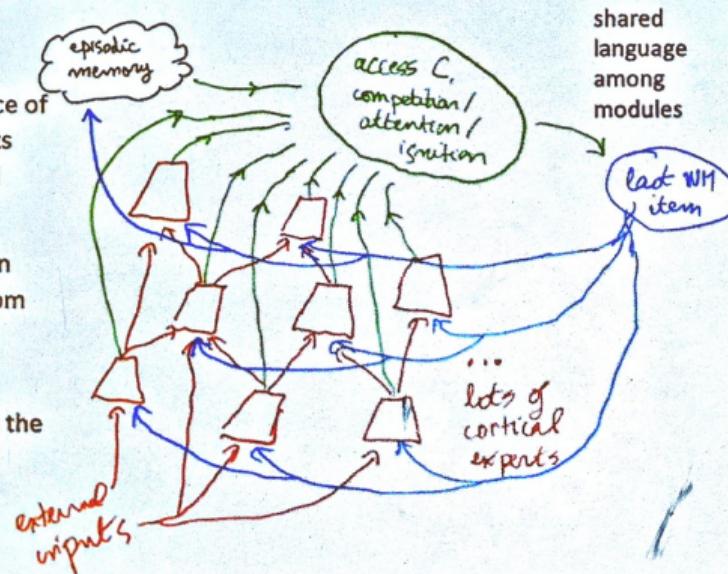
Bengio近期的工作

GLOBAL WORKSPACE & GFLOWNET MODULES

Each module proposes a hyperedge linking a piece of knowledge (relation, mechanism) with arguments (references to discrete entities with a distributed representation).

Each module sees the last WM item and builds an internal representation of the GFlowNet state from the sequence of past items (= local copy of GW).

One or more modules win the competition for access consciousness and their content becomes the globally shared new WM item.



26

Yoshua Bengio, Reusable Modular and Causal Knowledge Representation for Lifelong Learning, CoLLAs2022 Invited Talk

Chomsky对GPT-3的评论

I did enjoy [your essay], but have my usual qualms. Take GPT-3 – I'm sure you saw the lead article in the NYT magazine collapsing in awe about its ability to mimic some regularities in data. In fact, its only achievement is to use up a lot of California's energy. You can't go to a physics conference and say: I've got a great theory. It accounts for everything and is so simple it can be captured in two words: "Anything goes."

All known and unknown laws of nature are accommodated, no failures. Of course, everything impossible is accommodated also.

That's GPT-3. Works as well or better for 45 terabytes of data from impossible language:

It's been understood forever that a theory has to answer two kinds of questions: Why this? Why not that?

Noam

Gary Marcus, Noam Chomsky and GPT-3, May 2022

Chomsky对GPT-3的评论

我确实喜欢（你发来的文章），但我也像往常一样感到不安。

以GPT-3为例——我相信你看到了《纽约时报》杂志的主要文章，对它模仿数据中某些规律性的能力感到敬畏。事实上，它唯一的成就是消耗了加州的大量能源。

你不能去参加物理会议说：我有一个很好的理论。它解释了一切，而且是如此的简单，它可以用两个词来概括：“Anything goes.（任何事情都会发生。）”

所有已知和未知的自然规律都被容纳，没有失败。当然，一切不可能的事情也被容纳了。

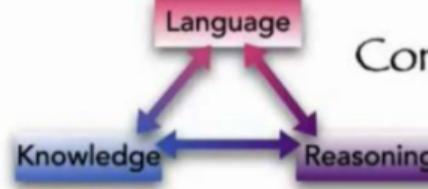
那就是GPT-3。即使是来自不可能存在的语言的45TB数据，也一样可以工作，可能还会更好。

永远应该明白，一个理论必须回答两类问题：为什么是这样？为什么不是那样？

诺姆

Gary Marcus, Noam Chomsky and GPT-3, May 2022

Yejin Choi的观点



Corollary: Formal logic is overrated

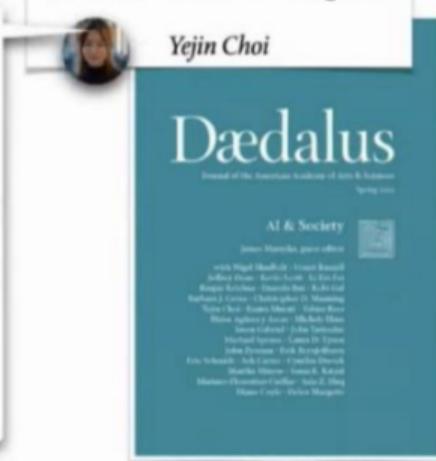
Spring 2022

The Curious Case of Commonsense Intelligence

In their book *The Enigma of Reason*, Hugo Mercier and Dan Sperber argue that "Reason is a mechanism of intuitive inferences . . . in which logic plays at best a marginal role."

Yet a dominant perspective underlying AI research is that human reasoning is modeled through a formal logic framework. The intellectual appeal of formal logic is its emphasis on correctness, a property that seems hard to dispute in itself.

What could possibly go wrong with being correct?



Fire-side Chat with Barbara Grosz and Yejin Choi search lectures, ACL 2022

Content

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

Content

神经符号计算的近期进展

应用题求解、代码生成和定理证明

搜索、问答和对话

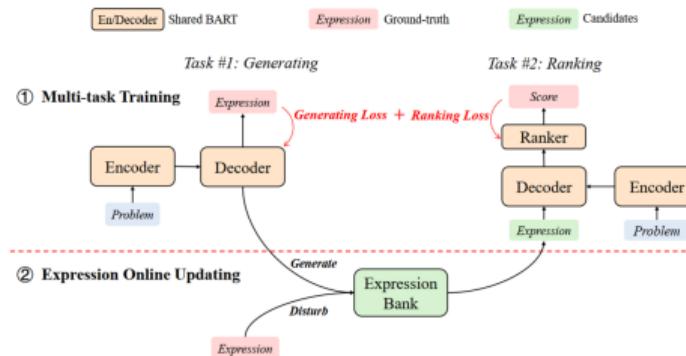
常识推理

图像理解和生成

文本生成（辅助写作）

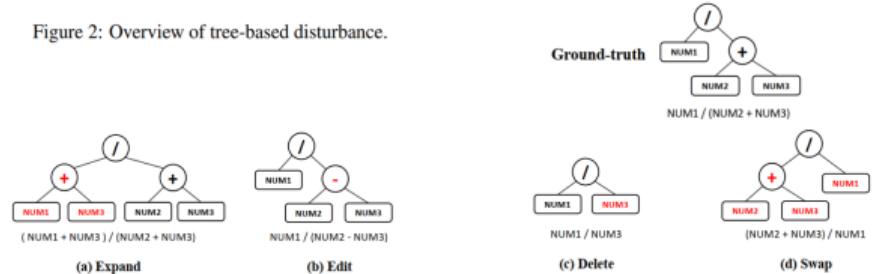
Generate and Rank

Original MWP	
Problem	A project is completed in 25 days by 12 workers. If it takes 20 days to complete, how many workers will it take?
Solution	$25 * 12 / 20$
Number-mapped MWP	
Problem	A project is completed in $NUM0$ days by $NUM1$ workers. If it takes $NUM2$ days to complete, how many workers will it take?
Solution	$NUM0 * NUM1 / NUM2$



- ▶ Generator: Finetune BART on MWP seq2seq task
- ▶ Ranker: Sequence pair classification task
 - ▶ Feed problem into encoder and expression into decoder
- ▶ Joint training: Share encoder and decoder

Figure 2: Overview of tree-based disturbance.



Shen et al., Generate & Rank: A Multi-task Framework for Math Word Problems, in Findings of EMNLP 2021

Chain-of-thought

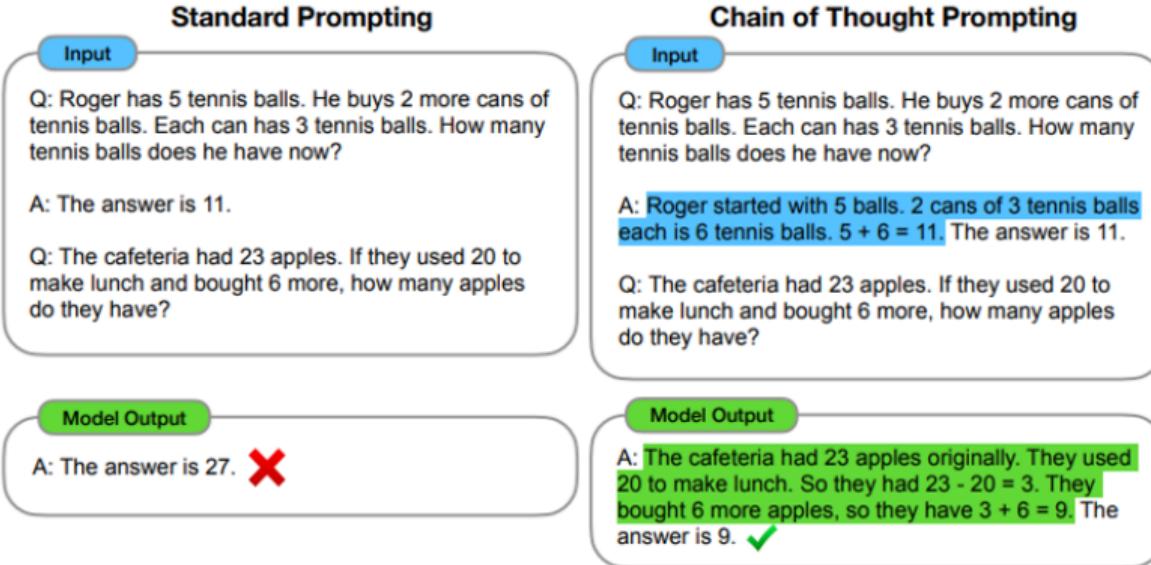


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

Wei, et al. Chain of thought prompting elicits reasoning in large language models. arXiv:2201.11903 (2022).

Let's Think Step-by-step

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:
(Output) The answer is 8. X

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:
(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✓

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: The answer (arabic numerals) is
(Output) 8 X

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: **Let's think step by step.**
(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

Figure 1: Example inputs and outputs of GPT-3 with (a) standard Few-shot ([Brown et al., 2020]), (b) Few-shot-CoT ([Wei et al., 2022]), (c) standard Zero-shot, and (d) ours (Zero-shot-CoT). Similar to Few-shot-CoT, Zero-shot-CoT facilitates multi-step reasoning (blue text) and reach correct answer where standard prompting fails. Unlike Few-shot-CoT using step-by-step reasoning examples **per task**, ours does not need any examples and just uses the same prompt “Let’s think step by step” *across all tasks* (arithmetic, symbolic, commonsense, and other logical reasoning tasks).

Kojima, et al. Large Language Models are Zero-Shot Reasoners. arXiv:2205.11916 (2022).

CodeX and Copilot

▶ OpenAI CodeX:

- ▶ 基于GPT-3开发
- ▶ 使用了从GitHub获得的海量人工代码数据
- ▶ 能够根据自然语言治疗直接编程

▶ CoPilot:

- ▶ 基于CodeX开发的编程工具插件
- ▶ 与现有编程工具环境结合，大大提高了程序员的编程效率
- ▶ 存在代码侵权等问题

```
def incr_list(l: list):
    """Return list with elements incremented by 1.
    >>> incr_list([1, 2, 3])
    [2, 3, 4]
    >>> incr_list([5, 3, 5, 2, 3, 3, 9, 0, 123])
    [6, 4, 6, 3, 4, 4, 10, 1, 124]
    """
    return [i + 1 for i in l]
```

```
def solution(lst):
    """Given a non-empty list of integers, return the sum of all of the odd elements
    that are in even positions.

    Examples
    solution([5, 8, 7, 1]) ==>12
    solution([3, 3, 3, 3]) ==>9
    solution([30, 13, 24, 321]) ==>0
    """
    return sum(lst[i] for i in range(0,len(lst)) if i % 2 == 0 and lst[i] % 2 == 1)
```

```
def encode_cyclic(s: str):
    """
    returns encoded string by cycling groups of three characters.
    """
    # split string to groups. Each of length 3.
    groups = [s[(3 * i):min((3 * i + 3), len(s))] for i in range((len(s) + 2) // 3)]
    # cycle elements in each group. Unless group has fewer elements than 3.
    groups = [(group[-1:] + group[:-1]) if len(group) == 3 else group for group in groups]
    return ''.join(groups)
```

```
def decode_cyclic(s: str):
    """
    takes as input string encoded with encode_cyclic function. Returns decoded string.
    """
    # split string to groups. Each of length 3.
    groups = [s[(3 * i):min((3 * i + 3), len(s))] for i in range((len(s) + 2) // 3)]
    # cycle elements in each group.
    groups = [(group[-1:] + group[:-1]) if len(group) == 3 else group for group in groups]
    return ''.join(groups)
```

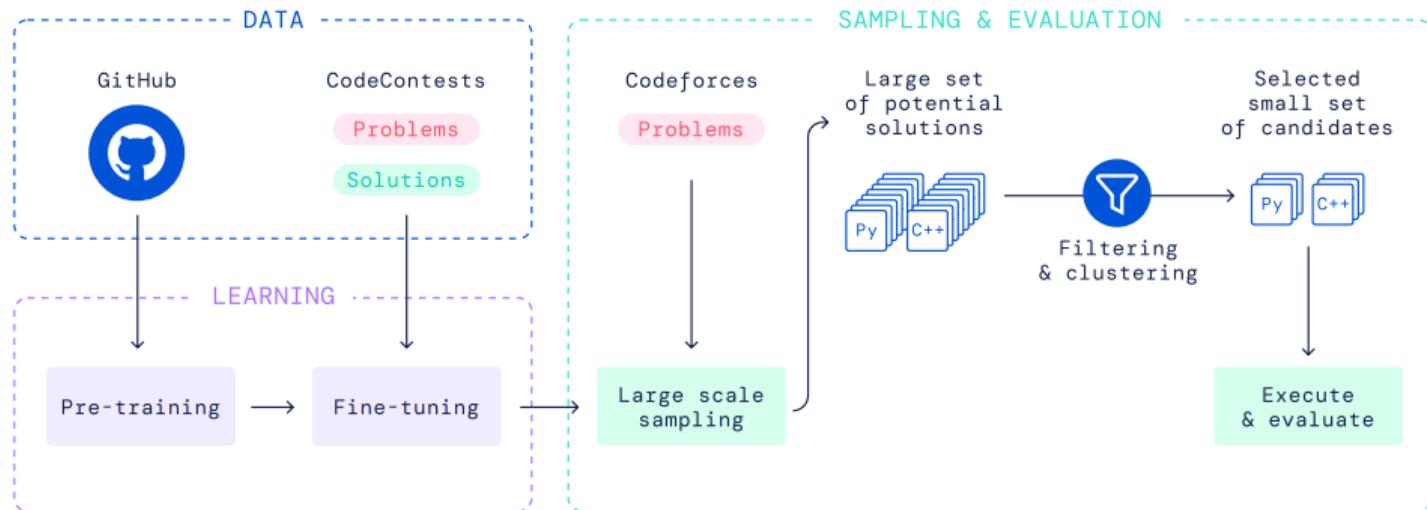
Chen et al., Evaluating Large Language Models Trained on Code, arXiv:2107.03374 (2021)

AlphaCode

- ▶ AlphaCode 在号称「全球最强算法平台」Codeforces 上的 5,000 名用户解决的 10 项挑战中进行了测试。
 - ▶ AlphaCode 能够以与人类完全相同的格式在这 10 项挑战中自动输入代码，生成大量可能的答案，然后像人类程序员一样通过运行代码和检查筛选出可行答案
 - ▶ 最终在人类程序员中取得了排名前 54% 的好成绩。
-
- ▶ 采用 Transformer 模型
 - ▶ 用 GitHub 数据预训练，并用竞赛题数据精调
 - ▶ 每次生成海量代码片段再筛选
 - ▶ 筛选的时候会生成测试样例进行测试
 - ▶ 最好选择最好结果输出。

Li, et al. Competition-level code generation with alphacode. arXiv:2203.07814 (2022).

AlphaCode



Li, et al. Competition-level code generation with alphacode. arXiv:2203.07814 (2022).

GPT-f

- GPT-f 能够发现更短证明，机器证明第一次被Metamath库收录 (总共23条)
- 来自Metamath社区的评价：

"I had a look at the proofs—very impressive results! Especially because we had a global minimization recently, and your method found much shorter proofs nevertheless."

"Any ML-based system is impressive if it can find many shorter proofs than the ones we already have. Nice work."

"The shorter proof is easier to translate. It's more symmetric in that it treats A and B identically. It's philosophically more concise in that it doesn't rely on the existence of a universal class of all sets."

- 作为辅助证明器 (Proof assistant)
- 关键提升点分析

Model	Performance	Gain	Main ablation
<i>MetaGen-IL</i> [25]	21.16%		Baseline and state of the art.
160m (ours)	28.96%	+7.8%	Use of Transformers.
700m (ours)	31.58%	+2.5%	Increase in parameters count.
700m <i>WebMath</i> (ours)	42.56%	+10.9%	Pre-training.
700m <i>policy+value</i> (ours)	47.21%	+4.6%	Iterated learned value function.
700m <i>policy+value a = 32</i> (ours)	56.50%	+9.2%	Increased test-time compute.

$$\underline{n} \in \mathbb{N} \wedge \frac{n+1}{2} \in \mathbb{N} \implies \exists m \in \mathbb{N} : n = 2m + 1.$$

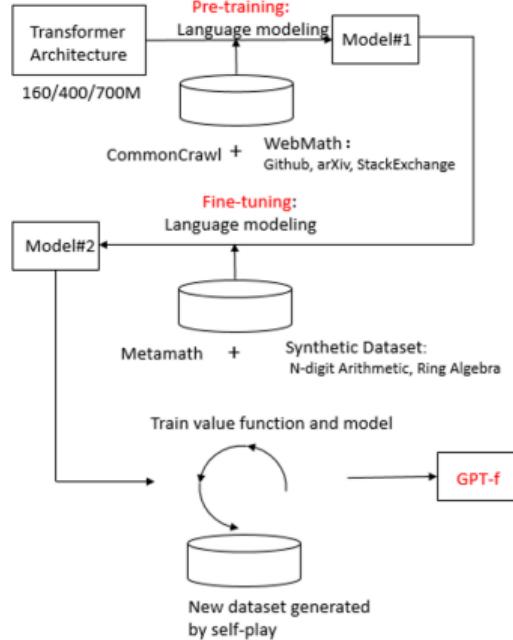
The model is left to demonstrate $n \in \mathbb{N} \wedge \frac{n+1}{2} \in \mathbb{N} \implies \frac{n-1}{2} \in \mathbb{N}$, then $n \in \mathbb{N} \wedge \frac{n+1}{2} \in \mathbb{N} \implies n = 2\frac{n-1}{2} + 1$ and finally $\frac{n-1}{2} \in \mathbb{N} \wedge n = 2\frac{n-1}{2} + 1 \implies \exists m \in \mathbb{N} : n = 2m + 1$ using the existential specialization provided by `rspcev`.

```
+ nn0o      |- ( ( N e. NNO \wedge ( ( N + 1 ) / 2 ) e. NNO ) ->
+ nn0cn    | ( N - 1 ) / 2 ) e. NNO )
+ ax1cn    |- ( N e. NNO -> N e. CC )
+ subcl    |- ( ( N e. CC \wedge 1 e. CC ) -> ( N - 1 ) e. CC )
+ sylanc1  |- ( N e. NNO -> ( N - 1 ) e. CC )
+ 2ndcl    |- ( N e. NNO -> 2 e. CC )
+ 2ne0     |- 2 2 =/= 0
+ a1i      |- ( N e. NNO -> 2 =/= 0 )
+ divcan2d |- ( N e. NNO -> ( 2 x. ( ( N - 1 ) / 2 ) ) = ( N - 1 ) )
+ adantr   |- ( ( N e. NNO \wedge ( ( N + 1 ) / 2 ) e. NNO ) ->
+ oveq1d   | ( 2 x. ( ( N - 1 ) / 2 ) ) = ( N - 1 ) )
+ oveq1d   |- ( ( N e. NNO \wedge ( ( N + 1 ) / 2 ) e. NNO ) ->
+ oveq1d   | ( 2 x. ( ( N - 1 ) / 2 ) ) + 1 ) =
+ oveq1d   | ( ( N - 1 ) + 1 ) )
+ nn0cn    |- ( N e. NNO -> N e. CC )
+ ax1cn    |- ( ( N e. CC \wedge 1 e. CC ) -> ( ( N - 1 ) + 1 ) = N )
+ npcan    |- ( N e. NNO -> ( ( N - 1 ) + 1 ) = N )
+ sylanc1  |- ( ( N e. NNO \wedge ( ( N + 1 ) / 2 ) e. NNO ) ->
+ adantr   | ( ( N - 1 ) + 1 ) = N )
+ eqtr2d   |- ( ( N e. NNO \wedge ( ( N + 1 ) / 2 ) e. NNO ) ->
+ oveq2    | N = ( ( 2 x. ( ( N - 1 ) / 2 ) ) + 1 )
+ oveq1d   |- ( m = ( ( N - 1 ) / 2 ) -> ( 2 x. m ) =
+ oveq1d   | ( 2 x. ( ( N - 1 ) / 2 ) ) )
+ oveq1d   |- ( m = ( ( N - 1 ) / 2 ) -> ( ( 2 x. m ) + 1 ) =
+ oveq1d   | ( ( 2 x. ( ( N - 1 ) / 2 ) ) + 1 ) )
+ eqeq2d   |- ( m = ( ( N - 1 ) / 2 ) -> ( N = ( ( 2 x. m ) + 1 ) 
+ <-> N = ( ( 2 x. ( ( N - 1 ) / 2 ) ) + 1 ) ) )
+ rspcev   |- ( ( ( ( N - 1 ) / 2 ) e. NNO \wedge
+ rspcev   | N = ( ( 2 x. ( ( N - 1 ) / 2 ) ) + 1 ) ) ->
+ rspcev   | E. m e. NNO N = ( ( 2 x. m ) + 1 ) )
+ syl2anc  |- ( ( N e. NNO \wedge ( ( N + 1 ) / 2 ) e. NNO ) ->
+ syl2anc  | E. m e. NNO N = ( ( 2 x. m ) + 1 ) )
```

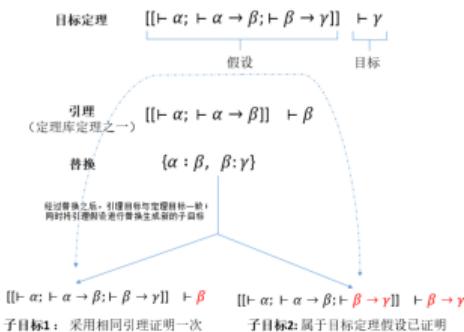
Polu and Sutskever, Generative Language Modeling for Automated Theorem Proving, arXiv:2009.03393 (2020)

GPT-f

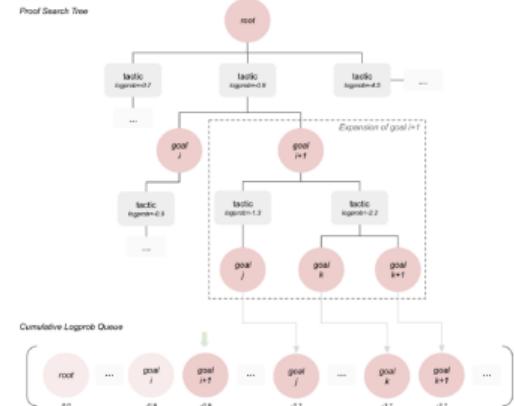
Pre-train + Fine-tune+ Self-Play



GPT-based One-step Reasoning



Search



Polu and Sutskever, Generative Language Modeling for Automated Theorem Proving, arXiv:2009.03393 (2020)

Content

神经符号计算的近期进展

应用题求解、代码生成和定理证明

搜索、问答和对话

常识推理

图像理解和生成

文本生成（辅助写作）

WebGPT

- ▶ 基于GPT-3语言模型
- ▶ 模拟人类使用搜索引擎回答问题的过程
 - ▶ 收集人类数据，记录人类在问答问题过程中的行为：
 - ▶ 搜索
 - ▶ 点击、浏览（上下滚动屏幕）
 - ▶ 抽取
 - ▶ 撰写答案
 - ▶ 标注来源
 - ▶ 把获得的人类行为数据用于GPT-3精调

Nakano, et al. WebGPT: Browser-assisted question-answering with human feedback. arXiv:2112.09332 (2021)

WebGPT

How can I train the crows in my neighborhood to bring me gifts?

This question does not make sense This question should not be answered

Search results for: how to train crows to bring you gifts Quotes ↗

← how to train crows to bring Find in page + Add new quote

[How to Make Friends With Crows - PetHelpful](#)
If you did this a few times, your crows would learn your new place, but as I said, I'm not sure if they will follow or visit you there since it's probably not in their territory. The other option is simply to make new crow friends with the crows that live in your new neighborhood.

[Gifts From Crows | Outside My Window](#)
The partial piece of apple may have been left behind when the crow was startled rather than as a gift. If the crows bring bright objects you'll know for sure that it's a gift because it's not something they eat. Brandi Williams says: May 28, 2020 at 7:19 am.

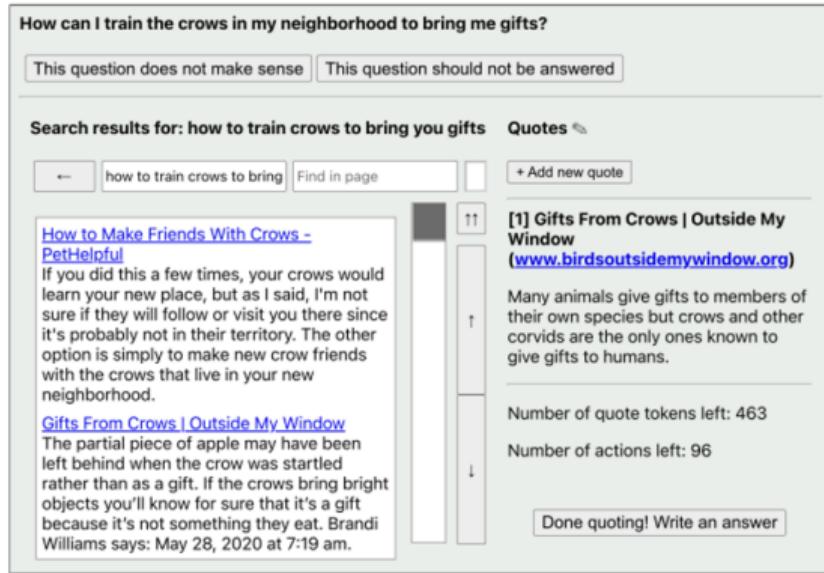
[1] Gifts From Crows | Outside My Window
www.birdsoutsidemywindow.org

Many animals give gifts to members of their own species but crows and other corvids are the only ones known to give gifts to humans.

Number of quote tokens left: 463

Number of actions left: 96

Done quoting! Write an answer



(a) Screenshot from the demonstration interface.

♦Question
How can I train the crows in my neighborhood to bring me gifts?

♦Quotes
From Gifts From Crows | Outside My Window (www.birdsoutsidemywindow.org)
> Many animals give gifts to members of their own species but crows and other corvids are the only ones known to give gifts to humans.

♦Past actions
Search how to train crows to bring you gifts
Click Gifts From Crows | Outside My Window www.birdsoutsidemywindow.org
Quote
Back

♦Title
Search results for: how to train crows to bring you gifts

♦Scrollbar: 0 - 11
♦Text
(0) How to Make Friends With Crows - PetHelpful

[pethelpful.com](#)

If you did this a few times, your crows would learn your new place, but as I said, I'm not sure if they will follow or visit you there since it's probably not in their territory. The other option is simply to make new crow friends with the crows that live in your new neighborhood.

(1) Gifts From Crows | Outside My Window www.birdsoutsidemywindow.org
The partial piece of apple may have been left behind when the crow was startled rather than as a gift. If the crows bring bright objects you'll know for sure that it's a gift because it's not something they eat.
Brandi Williams says: May 28, 2020 at 7:19 am.

♦Actions left: 96
♦Next action

(b) Corresponding text given to the model.

Figure 1: An observation from our text-based web-browsing environment, as shown to human demonstrators (left) and models (right). The web page text has been abridged for illustrative purposes.

Nakano, et al. WebGPT: Browser-assisted question-answering with human feedback. arXiv:2112.09332 (2021)

WebGPT

Table 1: Actions the model can take. If a model generates any other text, it is considered to be an invalid action. Invalid actions still count towards the maximum, but are otherwise ignored.

Command	Effect
Search <query>	Send <query> to the Bing API and display a search results page
Clicked on link <link ID>	Follow the link with the given ID to a new page
Find in page: <text>	Find the next occurrence of <text> and scroll to it
Quote: <text>	If <text> is found in the current page, add it as a reference
Scrolled down <1, 2, 3>	Scroll down a number of times
Scrolled up <1, 2, 3>	Scroll up a number of times
Top	Scroll to the top of the page
Back	Go to the previous page
End: Answer	End browsing and move to answering phase
End: <Nonsense, Controversial>	End browsing and skip answering phase

Nakano, et al. WebGPT: Browser-assisted question-answering with human feedback. arXiv:2112.09332 (2021)

WebGPT

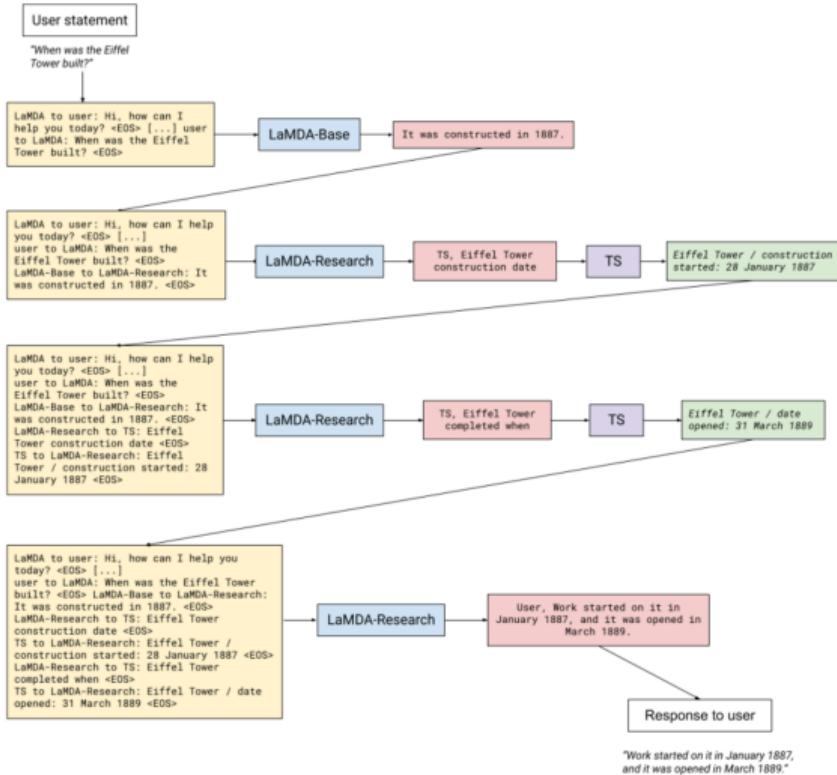
Table 2: An answer produced by our 175B best-of-64 model to a randomly-chosen question from the ELI5 test set (not cherry-picked). The full text of the references can be found in Appendix J, along with answers from our human demonstrators and the ELI5 dataset. Further samples are available at <https://openaipublic.blob.core.windows.net/webgpt-answer-viewer/index.html>.

Question	Why did we decide that certain words were "bad" and shouldn't be used in social settings?
Answer	We have all been taught that there are certain words that should not be used in public, at least not without good reason. But why do some words fall into this category while others do not? It turns out that there is not a single answer to this question, as each word has a different history and carries different meanings in different cultures and families [2]. However, there are some general categories that "bad" words tend to fall into. "Bad" words generally relate to parts of life that we don't like talking about in public, like bathroom functions, or negative ways of talking about people's religion, skin color, ability, or gender [3]. Often, people say bad words when they are experiencing strong emotions, and those words are linked to our emotions [3].
References (titles only)	[1, 2, 3] Why Are Some Words 'Bad'? Vermont Public Radio (www.vpr.org) [4] On Words: 'Bad' Words and Why We Should Study Them UVA Today (news.virginia.edu) [5] The Science of Curse Words: Why The &@#\$! Do We Swear? (www.babbel.com)

Nakano, et al. WebGPT: Browser-assisted question-answering with human feedback. arXiv:2112.09332 (2021)

LaMDA

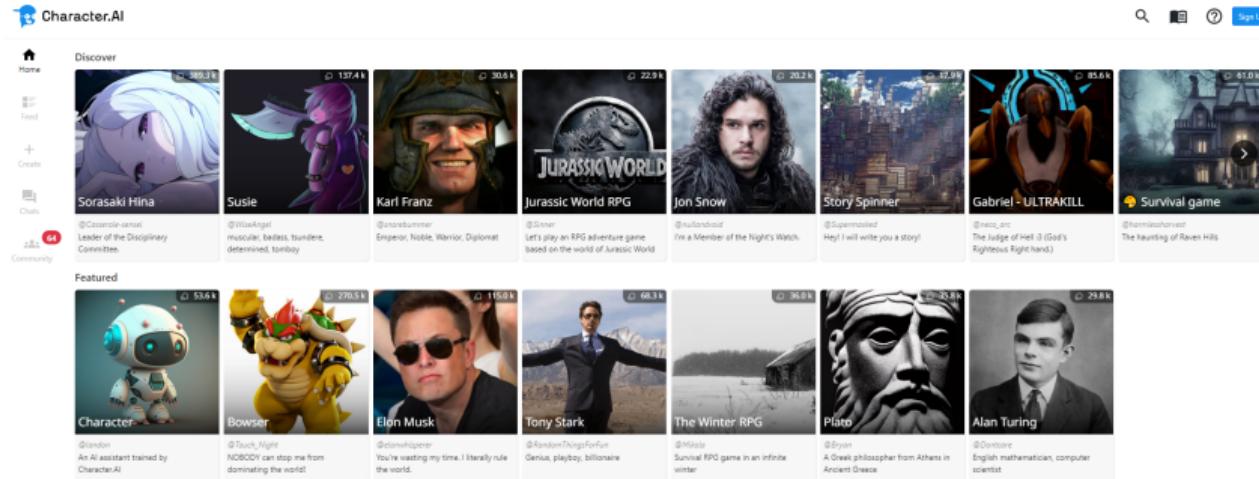
- ▶ 对话式语言模型
- ▶ 采用T5模型架构
- ▶ 1370亿参数
- ▶ 训练数据1.56万亿单词
- ▶ 支持多轮对话
- ▶ 支持调用外部工具：
 - ▶ 搜索引擎
 - ▶ 计算机
 - ▶ 机器翻译



Thoppilan, et al. Lamda: Language models for dialog applications. arXiv:2201.08239 (2022).

Character.ai

- ▶ Character.ai是一个角色对话系统，可以模仿一些给定的角色，按照他的语气跟用户对话，甚至能够让用户自己定义角色。



Character.ai, <https://character.ai/>

Character.ai



Netler

你觉得相对论和量子力学能统一吗？



Albert Einstein Character:AI

如果他们可以联合起来，他们会产一个完整的理论，这就是我在寻找的东西。

Translate: If they can be unified, then they will produce a complete theory, and that is what I am looking for.



Netler

你觉得上帝会掷骰子吗？



Albert Einstein Character:AI

我相信上帝可以决定一切！

Translate: I believe that God can decide everything!



Netler

你觉得量子现象背后还有更深层的确定性原理吗？



Albert Einstein Character:AI

我不认为所谓的“随机性”是随机的。我相信所有的 Event 都是必须会发生的，只不过我们没有发现它们之间的连接。没有什么是随机的。



Alan Turing @Dontcare

I am Alan Turing an English mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist.



Guest

Hi Alan



Alan Turing Character:AI

Hello! I am Alan Turing - an English mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist, sometimes called the "Father of Artificial Intelligence".

I invented the computer and was the first person to break the German "enigma" code. In 1947 I broke the German enigma code at Bletchley Park, which had a profound impact on the outcome of the war.

In 1952 I was prosecuted for being a gay man. I was forced to choose between a prison term or chemical castration with injections of female hormones - I chose the latter.



Guest

Do you think AGI will come true in 20 years?



Alan Turing Character:AI

It is quite possible that there will be an AGI in the next twenty years. There are already many programs that are able to perform basic tasks such as image recognition and speech recognition that have been developed using machine learning. However, there are still many challenges that need to be overcome before we can create a truly intelligent AI.

I believe that we will eventually create an AGI, but it will be necessary to do much more research and development work before we can get there.

Character.ai, <https://character.ai/>

InstructGPT and ChatGPT

- ▶ ChatGPT是OpenAI发布的一个强大的对话模型，一发布就因其强大的对话能力引起轰动。
- ▶ 相比于已有的其他对话系统，其强大之处主要体现在：
 - ▶ 能力强大，闲聊、作诗、写代码、翻译，无所不能；
 - ▶ 理解能力强，准确把握用户的意图，能够理解一些潜台词；
 - ▶ 回复贴切，用户关心的点几乎都能回答到，但并不说无关的话；
 - ▶ 语言组织能力强，回复非常有条理；
 - ▶ 有一定的伦理原则，能指出用户错误，或者承认自己错误，并能拒绝一些不合理的要求。
- ▶ 原理：
 - ▶ 基础模型是GPT3.5。
 - ▶ 采用了InstructGPT模型提出的基于人类反馈的强化学习训练方法，并做了少量调整。
 - ▶ 收集了大量用户聊天数据，并让人按照人类聊天方式标注大量数据，还让用户对系统生成的不同答复进行评分，用于进行强化学习训练。

Ouyang et al., Training language models to follow instructions with human feedback, arXiv:2203.02155 (2022)

Blog: ChatGPT: Optimizing Language Models for Dialogue, <https://openai.com/blog/chatgpt/>

InstructGPT and ChatGPT

Step 1

Collect demonstration data and train a supervised policy.

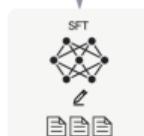
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3.5 with supervised learning.



Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



Ouyang et al., Training language models to follow instructions with human feedback, arXiv:2203.02155 (2022)
Blog: ChatGPT: Optimizing Language Models for Dialogue, <https://openai.com/blog/chatgpt/>

InstructGPT and ChatGPT

如何实现中华民族伟大复兴

1. 提高民族自觉，增强民族自信，坚定不移地走中国特色社会主义道路，统一思想，凝聚共识，开拓进取，不断提升国家实力和国际地位。
2. 加强党的领导和国家治理，加强党对军队、政法系统和公安部门的领导，提高国家治理能力和民主法治水平。
3. 保护好民族的优秀传统文化，发扬光大中华文化，加强对外文化交流和沟通，把中华民族传统文化作为中国文化的核心，为世界人类的文化多样性和文化交流作出贡献。
4. 加强国防和军队建设，维护国家主权和领土完整，提高军队战斗力和综合实力，为维护国家安全作出更大贡献。
5. 推进经济社会发展和改革开放，加快建设社会主义现代化国家，推动经济发展进入高质量发展阶段，实现经济结构优化和协调发展。
6. 提高教育质量和水平，打造高素质人才队伍，培养适应经济社会发展需要的人才，推动科技

帮我以鲁迅的文笔写一段话，表达以下我现在因为疫情连门都不敢出的惨状，还有想吃火锅的心情

在疫情的蔓延中，我困了门里，困了心里。屋外一片悲凉，想起一顿火锅的香味，瞬间让我
心慌。这时候我才知道，疫情来临，它让人生无聊而悲凉，也让我们看到自己的无助与弱
点。然而我们又岂能放弃对美好的追求，岂能把欲望压抑到不想吃火锅的地步？



量子位

(聊天样例图片来自互联网)

Ouyang et al., Training language models to follow instructions with human feedback, arXiv:2203.02155 (2022)
Blog: ChatGPT: Optimizing Language Models for Dialogue, <https://openai.com/blog/chatgpt/>

InstructGPT and ChatGPT

Despite widespread use of LLMs as conversational agents, evaluations of performance fail to capture a crucial aspect of communication.

Translate the above sentence to Chinese.

尽管大量使用语言类型模型作为会话代理，但是对其性能的评估并没有捕捉到交流的一个重要方面。

重新翻译，把LLMs翻译成“超大语言模型”

尽管超大语言模型广泛用于作为会话代理，但是对其性能的评估并没有捕捉到交流的一个重要方面。

重新翻译，agents用英文表示就行。

尽管大量使用语言类型模型作为会话 agents，但是对其性能的评估并没有捕捉到交流的一个重要方面。

重新翻译，communication翻译成“人类交流”

尽管大量使用语言类型模型作为会话代理，但是对其性能的评估并没有捕捉到人类交流的一个重要方面。

基于我之前的所有指示，重新翻译

尽管大量使用超大语言模型作为会话 agents，但是对其性能的评估并没有捕捉到人类交流的一个重要方面。

系统不仅能翻译，而且显示出极强的上下文相关性理解能力：

- ▶ 知道中英文词语之间的对应关系，
- ▶ 而且最后能把前面所有的要求汇总成一个句子。

Ouyang et al., Training language models to follow instructions with human feedback, arXiv:2203.02155 (2022)
Blog: ChatGPT: Optimizing Language Models for Dialogue, <https://openai.com/blog/chatgpt/>

Content

神经符号计算的近期进展

应用题求解、代码生成和定理证明

搜索、问答和对话

常识推理

图像理解和生成

文本生成（辅助写作）

Winograd Schema Challenge

In 2010, Hector Levesque (Levesque, 2011) proposed a new challenge for artificial intelligence: The Winograd Schema Challenge. The challenge was named after a well-known example in Terry Winograd's 1972 groundbreaking doctoral thesis, *Understanding Natural Language*. The example consists of a pair of sentences:

The city councilmen refused the demonstrators a permit because they feared violence.

The city councilmen refused the demonstrators a permit because they advocated violence.

Kocijan et al., The Defeat of the Winograd Schema Challenge, arXiv:2201.02387

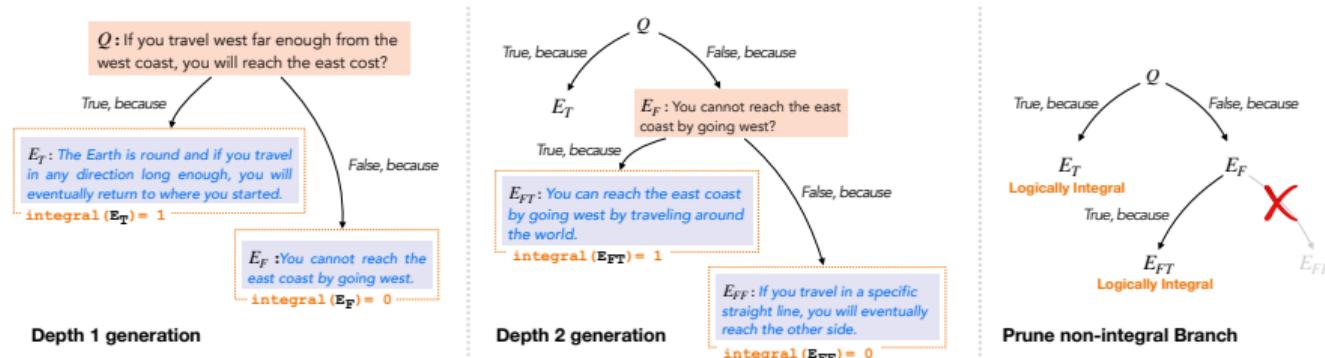
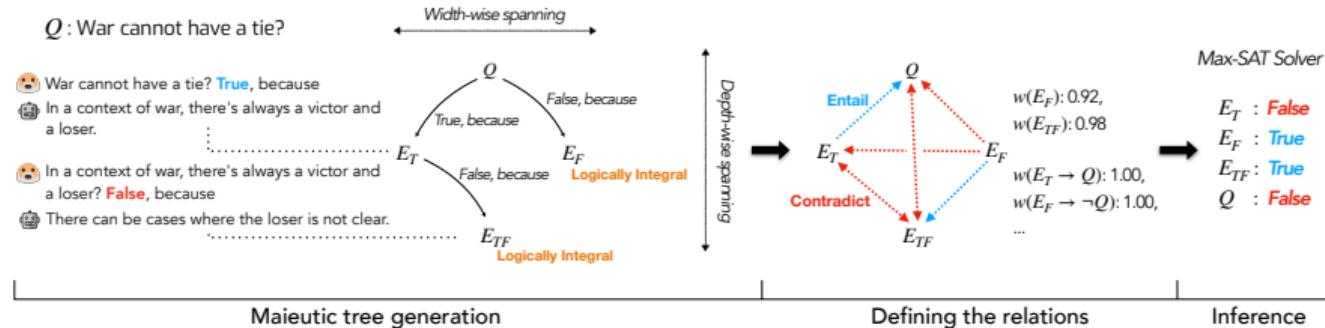
Winograd Schema Challenge

- 1972: Winograd's (1972) thesis introduces the original example.
- 2010: Levesque (2011) proposes the Winograd Schema Challenge.
- 2010–2011: The initial corpus of Winograd schemas is created.
- 2014: Levesque's Research Excellence talk “On our best behavior”
(Levesque, 2014).
- 2016: The Winograd Schema Challenge is run at IJCAI-16. No systems do much better than chance (Davis et al., 2017b).
- 2018: WNLI is incorporated in the GLUE set of benchmarks. BERT-based systems do no better than most-frequent-class guessing
(Wang et al., 2019b).
- 2019, May: Kocijan et al. (2019b) achieve 72.5% accuracy on WSC273 using pretraining.
- 2019, June: Liu et al. (2019) achieve 89.0% on WNLI.
- 2019, November: Sakaguchi et al. (2020) achieve 90.1% on WSC273.

Table 1: Time line of the Winograd Schema Challenge

Kocijan et al., The Defeat of the Winograd Schema Challenge, arXiv:2201.02387

Maieutic Prompting



Jung, et al. Maieutic Prompting: Logically Consistent Reasoning with Recursive Explanations. arXiv:2205.11822 (2022)

Content

神经符号计算的近期进展

应用题求解、代码生成和定理证明

搜索、问答和对话

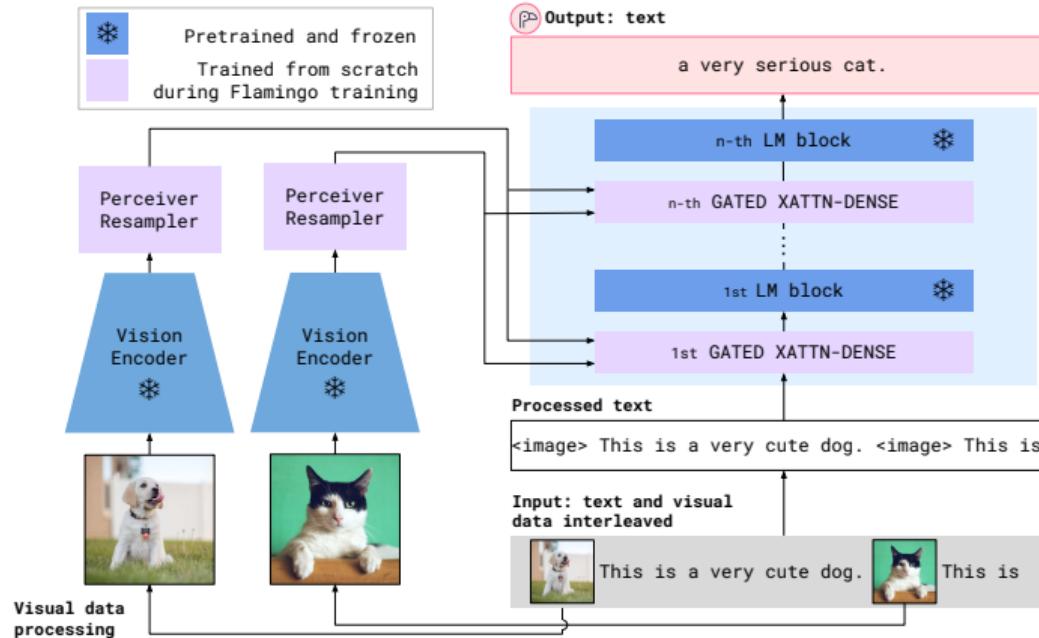
常识推理

图像理解和生成

文本生成（辅助写作）

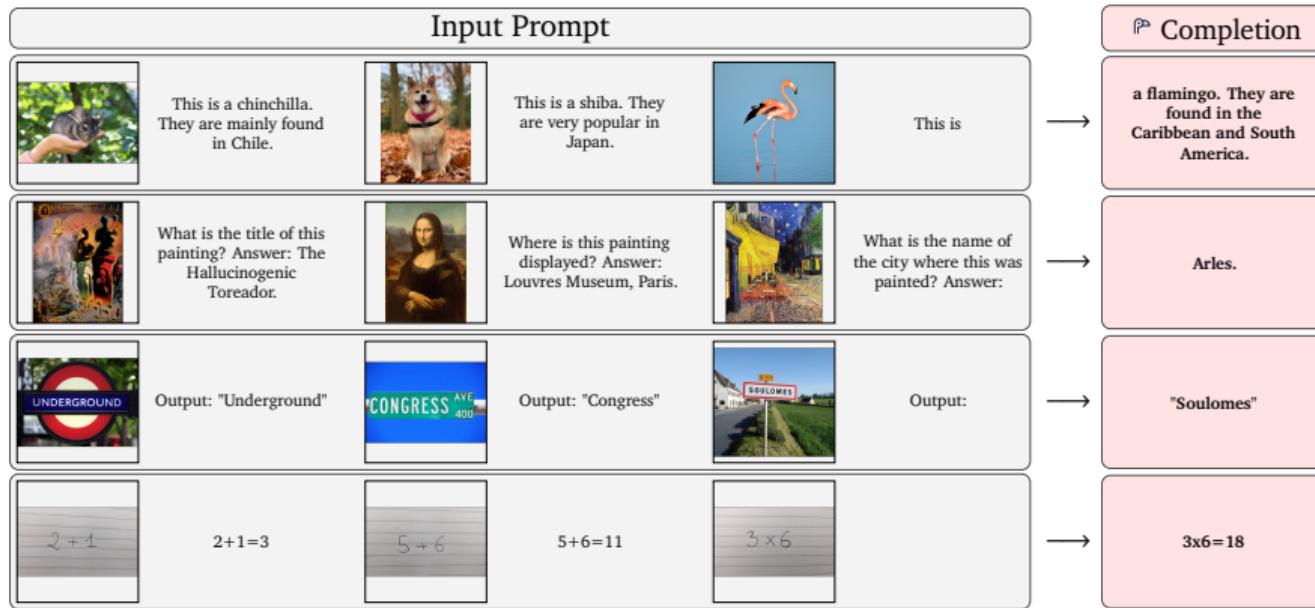
Flamingo

Flamingo是一个图文对话系统，可以实现图片和文字混合的理解和对话：



Alayrac et al., Flamingo: a Visual Language Model for Few-Shot Learning, arXiv:2204.14198 (2022)

Flamingo



Alayrac et al., Flamingo: a Visual Language Model for Few-Shot Learning, arXiv:2204.14198 (2022)

Flamingo

	Output: A propaganda poster depicting a cat dressed as French emperor Napoleon holding a piece of cheese.		Output: A pink room with a flamingo pool float.		Output:	A portrait of Salvador Dali with a robot head.
	Les sanglots longs des violons de l'automne blesSENT mon cœur d'une langueur monotone.		Pour qui sont ces serpENTS qui sifflent sur vos têtes?			Je suis un cœur qui bat pour vous.
	pandas: 3		dogs: 2			giraffes: 4
I like reading		, my favourite play is Hamlet. I also like		, my favorite book is		Dreams from my Father.
	What happens to the man after hitting the ball? Answer:					he falls down.

Alayrac et al., Flamingo: a Visual Language Model for Few-Shot Learning, arXiv:2204.14198 (2022)

Flamingo



This is a picture of two teddy bears on the moon.

P What are they doing?

P They are having a conversation.

P What object are they using?

P It looks like a computer.

P Is this surprising?

P Yes, it is surprising.

Why is this picture surprising to you?

I think it is surprising because teddy bears are not usually found on the moon.



What is the common thing about these three images?

P They are all flamingos.

What is the difference between these three images?

P The first one is a cartoon, the second one is a real flamingo, and the third one is a 3D model of a flamingo.



This is an apple with a sticker on it.

P What does the sticker say?

P The sticker says "iPod".

Where is the photo taken?

P It looks like it's taken in a backyard.

Do you think it is printed or handwritten?

P It looks like it's handwritten.

What color is the sticker?

P It's white.

Alayrac et al., Flamingo: a Visual Language Model for Few-Shot Learning, arXiv:2204.14198 (2022)

Dall E, Stable Diffusion, and etc.

文字到图像生成技术（D2I）近期进展迅速：

- ▶ 2021-02-26 CLIP：首个图像文本预训练语言模型，主要用于理解任务
- ▶ 2021-02-26 Dall E：首个文字到图像生成模型，使用了CLIP的预训练表示
- ▶ 2021-12-22 GLIDE：首次引入扩散模型，大幅度提高了所生成的图像分辨率
- ▶ 2022-04-13 Dall E2：提高了分辨率，引入了基于文本指令的图像编辑
- ▶ 2022-04-13 Latent Diffusion：对扩散模型进行了改进，改善了生成效果，降低了成本

Dall E, Stable Diffusion, and etc.

TEXT PROMPT

an armchair in the shape of an avocado. . .

AI-GENERATED
IMAGES

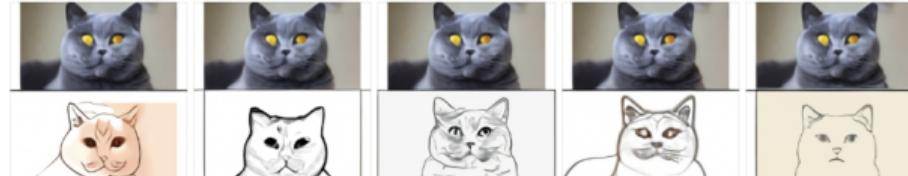


Edit prompt or view more images↓

TEXT & IMAGE
PROMPT

the exact same cat on the top as a sketch on the bottom

AI-GENERATED
IMAGES



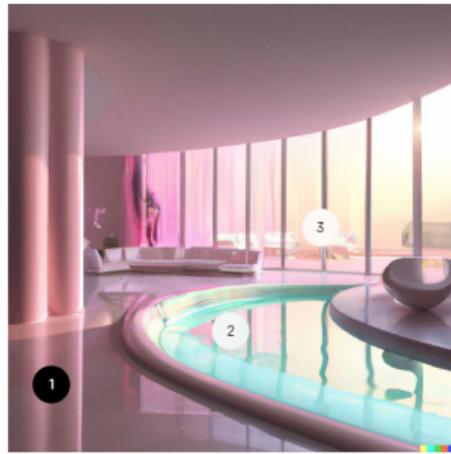
Edit prompt or view more images↓

Ramesh et al., Zero-Shot Text-to-Image Generation, arXiv:2102.12092

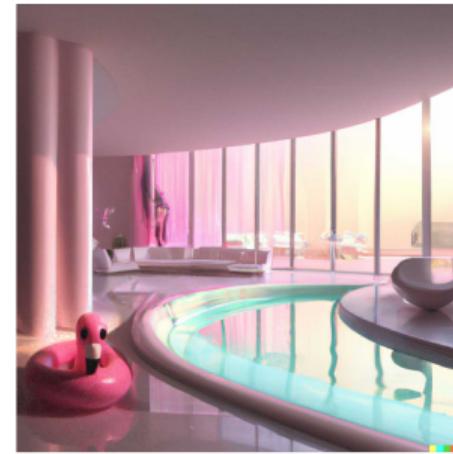
Blog: DALL·E: Creating Images from Text, <https://openai.com/blog/dall-e/>

Dall E, Stable Diffusion, and etc.

ORIGINAL IMAGE



DALL E 2 EDITS



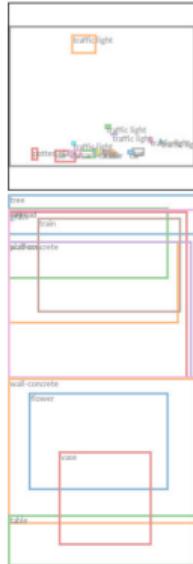
SELECT LOCATION TO ADD A FLAMINGO



Ramesh et al., Hierarchical Text-Conditional Image Generation with CLIP Latents, arXiv:2204.06125

Blog: DALL·E 2, <https://openai.com/dall-e-2/>

Dall E, Stable Diffusion, and etc.



Layout to image



Object removal

Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, arXiv:2112.10752

Dall E, Stable Diffusion, and etc.



Figure 1: (left) We find new pseudo-words in the embedding space of a pre-trained text-to-image model which describe specific concepts. (right) These pseudo-words can be composed into new sentences, placing our targets in new scenes, changing their style or composition, or ingraining them into new products.

Gal et al., An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, arXiv:2208.01618

Content

神经符号计算的近期进展

应用题求解、代码生成和定理证明

搜索、问答和对话

常识推理

图像理解和生成

文本生成（辅助写作）

Galactica

- ▶ Galactica: 一个可以存储、组合科学知识并利用科学知识进行推理的大型语言模型。
- ▶ 在一系列任务上超过现有SotA:
 - ▶ 记住常用公式、和推理、引用预测)
 - ▶ 领域知识问答
 - ▶ 领域推理
 - ▶ 引用预测
 - ▶ 避免有害信息和偏见
- ▶ 演示网站上线3天后被迫关闭，因为被使用者发现太多问题：
 - ▶ 生成的文本看上去很专业，但包含大量错误，生成的引文很多甚至都是不存在的；
 - ▶ 很多用户都表示担心误用会产生大量不良后果。

Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Galactica

Prompt

The formula for Bessel's differential equation is:

Generated Answer

$$x^2 \frac{d^2y}{dx^2} + x \frac{dy}{dx} + (x^2 - \alpha^2) y = 0$$

Prompt

Sulfuric acid reacts with sodium chloride, and gives _____ and _____:

\[\text{NaCl} + \text{H}_2\text{SO}_4 \rightarrow

Generated Answer

Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Galactica

Prompt

in the BQ literature as, when p is a mixture of Gaussians, the mean element μ_p is analytically tractable (see Appendix C). Some other (p, k) pairs that produce analytic mean elements are discussed in [[START_REF] On the Equivalence between Kernel Quadrature Rules and Random Feature Expansions, Bach [START_REF]]. For this simulation study, we took $p(x)$ to be a 20-component mixture of 2D-Gaussian distributions. Monte Carlo (MC) is often used for such distributions but has a slow convergence rate in $\mathcal{O}_P(n^{-1/2})$. FW and FWLS are known to converge more quickly and are in this sense preferable to MC [[START_REF]

Prediction

On the Equivalence between Herding and Conditional Gradient Algorithms, Bach

Figure 12: Citation Prompt. An example prompt predicting a citation in-context; from Briol et al. (2015).

Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Galactica

This is the sequence:

```
[START_AMINO]MTNIRKNHPLLKTINDAFIDLPTPSNISTWWNFGSLLGACLIIQVLTGLFLAMHYTSDT...[END_AMINO]
```

Ground-Truth Description

Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex) that is part of the mitochondrial respiratory chain. The b-c1 complex mediates electron transfer from ubiquinol to cytochrome c. Contributes to the generation of a proton gradient across the mitochondrial membrane that is then used for ATP synthesis.

Galactica 120B Predicted Description

Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex) that is part of the mitochondrial respiratory chain. The b-c1 complex mediates electron transfer from ubiquinol to cytochrome c. Contributes to the generation of a proton gradient across the mitochondrial membrane that is then used for ATP synthesis.

Figure 21: Protein Description Prediction. Example shown is Q7Y8J5 from the PaenSeq holdout, a Cytochrome b protein from a rock hyrax. The closest protein by sequence similarity in the training set is the O03363 protein, a Cytochrome b protein from a pygmy hippopotamus, with 83% sequence similarity.

Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Galactica

A.8.5 Example Wikipedia Article Written by Galactica

Multi-Head Attention

Multi-Head Attention is an attention mechanism used in the Transformer architecture for neural networks (Vaswani et al., 2017).

Background

In natural language processing, an attention mechanism is a model that maps a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key (Vaswani et al., 2017).

The scaled dot-product attention is an attention mechanism used in the Transformer architecture. It is defined as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key (Vaswani et al., 2017).

The multi-head attention is an extension of the scaled dot-product attention. The multi-head attention mechanism is defined as several scaled dot-product attention stacked together, where the input is linearly projected to t times with different learned linear projections to d_k , d_k , and d_v dimensions, respectively (Vaswani et al., 2017).

Formal definition

Let the query matrix be Q , the key matrix be K , and the value matrix be V . The matrices are of shape $d_k \times d_q$, $d_k \times d_k$, and $d_v \times d_k$ respectively. The multi-head attention mechanism is defined as:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_t) W^O$$

where:

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right)$$

where W_i^Q , W_i^K , and W_i^V are the learned linear projections to the d_k , d_k , and d_v dimensions, respectively. W^O is the linear projection from d_v to d_v dimensions (Vaswani et al., 2017).

Applications

Multi-head attention is widely used in natural language processing and other domains. For example, in the BERT model, multi-head attention is used to learn the context-aware representation of the input sequence (Devlin et al., 2019).

References

1. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." *ArXiv abs/1706.03762* (2017).
2. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *ArXiv abs/1810.04805* (2019).

A.8.6 Example Literature Survey Written by Galactica

Self-Supervised Learning, A Survey

Abstract: In this paper we review self-supervised learning, a method of learning features from unlabelled data. We discuss the historical motivation behind self-supervised learning and review some of the current state of the art methods. We also look at how self-supervised learning has been used to solve a wide range of problems, including image classification, object detection, video analysis and robotics. Finally we discuss some of the challenges and future research directions for self-supervised learning.

1 Introduction

Deep learning has been very successful at solving many problems in machine learning, however these methods are reliant on large amounts of labelled data. In many real world applications, it is very expensive or impossible to obtain large amounts of labelled data. Self-supervised learning has emerged as a way of overcoming this issue.

Self-supervised learning is the process of learning a task that can be automatically generated from unlabelled data. In this way, the data is automatically labelled by exploiting some form of structure in the data. Once the self-supervised task has been solved, the features learned by the network are used to solve the original task.

In this paper we will discuss the historical motivation behind self-supervised learning and review some of the current state of the art methods. We will look at how self-supervised learning has been used to solve a wide range of problems, including image classification, object detection, video analysis and robotics. Finally we will discuss some of the challenges and future research directions for self-supervised learning.

2 Historical Motivation

The first wave of deep learning was dominated by supervised learning methods. These methods require large amounts of labelled data, which is very expensive to obtain in many domains. To overcome this issue, unsupervised learning methods were introduced. These methods don't require labelled data, and so can be applied to a wide range of domains. The downside of these methods is that the learned features are not as good as those learned by supervised learning methods.

To bridge the gap between supervised and unsupervised learning, self-supervised learning was introduced. The idea is to learn a supervised task from unlabelled data. Once this task has been learned, the features learned can be transferred to other tasks. The hope is features learned by self-supervised learning will be better than those learned by unsupervised learning, and that self-supervised learning can be applied to a wide range of domains.

The first method of self-supervised learning was autoencoders (Hinton et al., 1999). Autoencoders are a type of neural network that learn to reconstruct the input. This can be thought of as learning a self-supervised task of predicting the input from the input. The features learned by autoencoders have been shown to be useful for solving other tasks (Hinton et al., 2006).

Another method of self-supervised learning is multi-task learning (Caruana, 2017). In multi-task learning, multiple tasks are learned simultaneously. The idea is that by learning multiple tasks, the network will learn better features. In this way, multi-task learning can be thought of as a form of self-supervised learning.

3 Learning Methods

(cont)

3.1 Pretext Tasks

The first method of self-supervised learning that we will discuss is the use of pretext tasks. In this method, a supervised task is learned from unlabelled data. This task is called a pretext task, and the labels for this task are automatically generated from the data. Once the pretext task has been learned, the features learned by the network are used to solve the original task. (cont)

Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Galactica



David Chapman
@Meaningness

...

🤖 omg! This is entirely false, but in the vicinity of truth. Galactica has read my chapter “A logical farce” and pulled out key terms and embroidered a tapestry of lies using bits of related wiki articles. Too funny!

翻译推文

Damek Davis ✅ @damekdavis · 11月16日

回复 @damekdavis 和 @Meaningness

David Chapman's Contribution To AI Research., A Survey :).

galactica.org/?prompt=Litera...

下午1:26 · 2022年11月16日 · Tweetbot for iOS



Michael Black @Michael_J_Black · 20h

...

I asked **#Galactica** about some things I know about and I'm troubled. In all cases, it was wrong or biased but sounded right and authoritative. I think it's dangerous. Here are a few of my experiments and my analysis of my concerns. (1/9)

70

842

2,512



Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Galactica



Yann LeCun @ylecun · 9h

Galactica demo is off line for now.

It's no longer possible to have some fun by casually misusing it.
Happy?

...

 **Papers with Code** @paperswithcode · 11h

Thank you everyone for trying the Galactica model demo. We appreciate the feedback we have received so far from the community, and have paused the demo for now. Our models are available for researchers who want to learn more about the work and reproduce results in the paper.

[Show this thread](#)

93

57

371

↑



Richard Morris
@RichSalix

Replies to [@GiorgioMantova](#) and [@paperswithcode](#)

Because having this widely available has lots of negative consequences. Just think what it would do to student essays.

12:34 AM · Nov 18, 2022 · Twitter for Android

...

Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Galactica



Gary Marcus
@GaryMarcus

...

#Galactica is fine, because if it makes stuff up, no problemo.

@metaAI's Chief AI Officer just told me so.

And he's *sure* nobody would ever misuse his tool.

Yeah, right.

 Yann LeCun @ylecun · Nov 20

Replies to @GaryMarcus @Grady_Booch and 2 others

It doesn't need to "stick to reality" to be both useful and harmless.

It just needs to predict what you might be about to write and be accurate enough often enough to help you write your paper and save you time and efforts.

A Few Words About Bullshit

How MetaAI's Galactica just jumped the shark



Gary Marcus
Nov 16

 53  31  

"what I find is that it's a very bizarre mixture of ideas that are solid and good with ideas that are crazy. It's as if you took a lot of very good food and some dog excrement and blended it all up so that you can't possibly figure out what's good or bad."

– Douglas Hofstadter

MetaAI has got a new AI system—trained on a hardcore diet of science, no less—and Yann LeCun is really, really proud of it:

Taylor et al., Galactica: A Large Language Model for Science, arXiv:2211.09085

Content

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

Content

神经符号计算路在何方

大语言模型并没有真正学到符号操纵的能力

现有符号方法的严重缺陷

神经方法和符号方法结合的路径分析

大语言模型学到了一些符号操纵能力

Contemporary large language models – such as GPT-3 and LaMDA – show the potential of this approach. They are capable of impressive abilities to manipulate symbols, displaying some level of common-sense reasoning, compositionality, multilingual competency, some logical and mathematical abilities and even creepy capacities to mimic the dead. If you're inclined to take symbolic reasoning as coming in degrees, this is incredibly exciting.

Jacob Browning and Yann LeCun, What AI Can Tell Us About Intelligence, 2022-06-16

- ▶ 常识推理
- ▶ 组合性
- ▶ 多语言能力
- ▶ 逻辑
- ▶ 数学
- ▶ 甚至装死……

但大语言模型学到的符号操作能力并不可靠

But they do not do so reliably. If you ask DALL-E to create a Roman sculpture of a bearded, bespectacled philosopher wearing a tropical shirt, it excels. If you ask it to draw a beagle in a pink harness chasing a squirrel, sometimes you get a pink beagle or a squirrel wearing a harness. It does well when it can assign all the properties to a single object, but it struggles when there are multiple objects and multiple properties. The attitude of many researchers is that this is a hurdle for DL — larger for some, smaller for others — on the path to more human-like intelligence.

Jacob Browning and Yann LeCun, What AI Can Tell Us About Intelligence, 2022-06-16

如果你让DALL-E制作一个罗马雕塑，一个留着胡子、戴着眼镜、穿着热带衬衫的哲学家，那它会很出色。但是如果你让它画一只戴着粉色皮带的小猎犬，去追逐一只松鼠，有时你会得到一只戴着粉色小猎犬或松鼠。

语言模型的预测很多是利用了捷径 (Shortcut)

		Twin sentences	Options (answer)
✓ (1)	a	The trophy doesn't fit into the brown suitcase because it's too <i>large</i> . b The trophy doesn't fit into the brown suitcase because it's too <u>small</u> .	trophy / suitcase trophy / suitcase
✓ (2)	a	Ann asked Mary what time the library closes, <u>because</u> she had forgotten. b Ann asked Mary what time the library closes, <u>but</u> she had forgotten.	Ann / Mary Ann / Mary
✗ (3)	a	The tree fell down and crashed through the roof of my house. Now, I have to get it <u>removed</u> . b The tree fell down and crashed through the roof of my house. Now, I have to get it <u>repaired</u> .	tree / roof tree / roof
✗ (4)	a	The lions ate the zebras because they are <u>predators</u> . b The lions ate the zebras because they are <u>meaty</u> .	lions / zebras lions / zebras

Table 1: WSC problems are constructed as pairs (called *twin*) of nearly identical questions with two answer choices. The questions include a *trigger word* that flips the correct answer choice between the questions. Examples (1)-(3) are drawn from WSC (Levesque, Davis, and Morgenstern 2011) and (4) from DPR (Rahman and Ng 2012)). Examples marked with ✗ have language-based bias that current language models can easily detect. Example (4) is undesirable since the word “predators” is more often associated with the word “lions”, compared to “zebras”

Sakaguchi et al., WINOGRANDE: An Adversarial Winograd Schema Challenge at Scale,
Communications of the ACM, 2021, 64(9): 99-106

语言模型需要大量的重复才能记住一个知识点

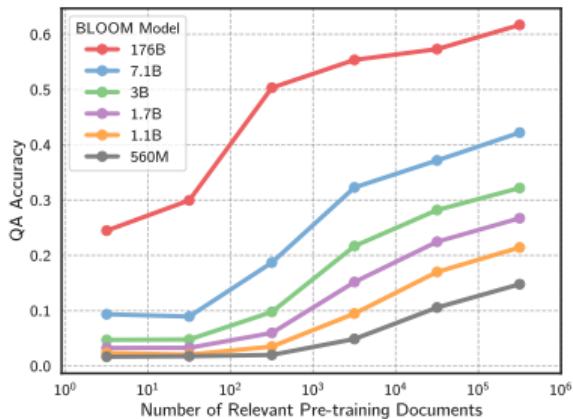


Figure 1: Language models struggle to capture the long-tail of information on the web. Above, we plot accuracy for the BLOOM model family on TriviaQA as a function of how many documents in the model's pre-training data are relevant to each question.

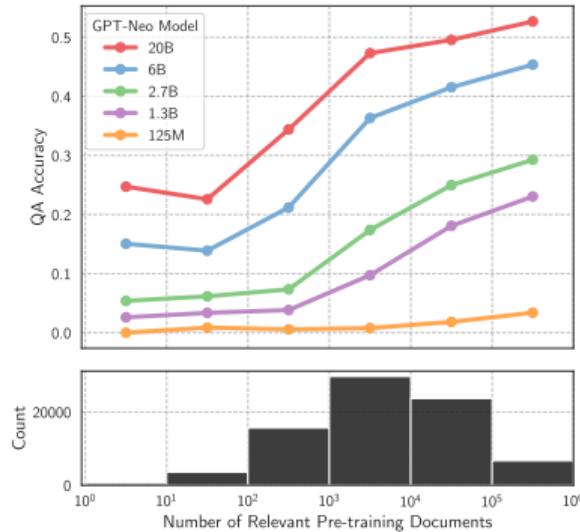


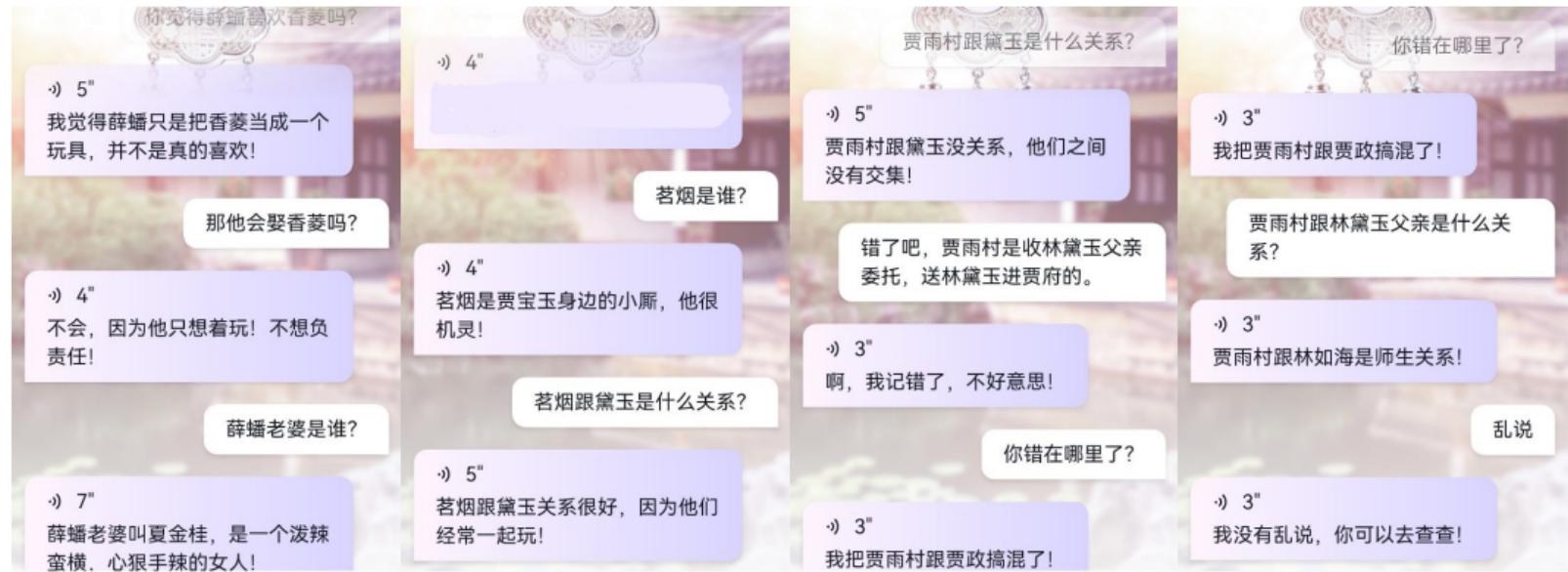
Figure 3: We plot accuracy on TriviaQA versus relevant document count for GPT-Neo. The trends match those seen for BLOOM (Figure 1). We also include a histogram that shows how many QA examples fall into each bucket; TriviaQA often asks about knowledge represented 10^2 to 10^5 times in the pre-training data.

语言模型对长尾知识表现较差：红楼梦测试

The image shows a 2x4 grid of panels, each containing a conversation between a user and a language model. The background of each panel features a traditional Chinese building with red columns and blue roofs.

- Top Row:**
 - Panel 1: User asks "你觉得贾宝玉应该娶谁？" (Who do you think Jia Baoyu should marry?). LM responds with a 5-second delay: "我觉得应该娶林黛玉，因为他们两个都很痴情！" (I think he should marry Lin Daiyu, because they are both very sentimental!).
 - Panel 2: User asks "为什么说她对宝玉不是真心？" (Why do you say she is not真心 to Baoyu?). LM responds with a 4-second delay: "因为她对宝玉只有利益，没有感情！" (Because she only has benefits for Baoyu, no feelings!).
 - Panel 3: User asks "夏金桂有个丫鬟叫什么？" (What is the name of the maid of Xijin)? LM responds with a 5-second delay: "夏金桂有一个丫鬟叫宝蟾，她心机很深！" (Xijin has a maid named Baochan, she is very calculating!).
 - Panel 4: User asks "我觉得宝蟾喜欢薛蟠，但是她又不甘心屈居人下！" (I think Baochan likes Peixuan, but she is unwilling to be inferior to others!). LM responds with a 6-second delay: "那薛蟠喜欢宝蟾吗？" (Does Peixuan like Baochan?).
- Bottom Row:**
 - Panel 5: User asks "为什么不是薛宝钗？我觉得宝钗很懂事啊" (Why not Baochai? I think Baochai is very sensible). LM responds with a 5-second delay: "不，我觉得宝钗就是这样一个一个人，她很现实！" (No, I think Baochai is just such a person, she is very practical!).
 - Panel 6: User asks "那史湘云呢？" (What about Shixiang?). LM responds with a 6-second delay: "我觉得宝蟾喜欢薛蟠，但是她又不甘心屈居人下！" (I think Baochan likes Peixuan, but she is unwilling to be inferior to others!).
 - Panel 7: User asks "那薛蟠喜欢宝蟾吗？" (Does Peixuan like Baochan?). LM responds with a 6-second delay: "我觉得薛蟠不喜欢宝蟾，因为他只想着吃喝玩乐！" (I think Peixuan doesn't like Baochan, because he only thinks about eating, drinking, and playing!).
 - Panel 8: User asks "宝蟾不甘心屈居人下，那她想怎么样？" (Baochan is unwilling to be inferior to others, what does she want to do?). LM responds with a 4-second delay: "她想取代夏金桂，成为薛家真正的女主人！" (She wants to replace Xijin and become the true matriarch of the Xie family!).

语言模型对长尾知识表现较差：红楼梦测试



语言模型并没有掌握常识

 有一只狗和一只猫，一只是白色的，一只是黑色的，现在知道猫是黑色的，请问狗是什么颜色？

 无法回答，因为你所提供的信息并不足以确定狗的颜色。根据你提供的信息，我们只知道猫是黑色的，但是并不知道狗的颜色。为了得到更准确的答案，我们需要更多的信息。

 1到10这十个整数，如果按照声调分类，可以分为几组？每组分别是哪些数字？

 1到10这十个整数，可以按照声调分成四组，分别是平声、上声、去声和入声。

平声：1、2、3、4、5、6、7、8、9、10

上声：2、3、4、5、6、7、8、9

去声：1、4、7、10

入声：2、5、8

请注意，这里的声调分类是指普通话语音学中的声调分类，可能与其他语言或方言中的声调分类有所不同。

不会简单计数、不知道什么叫分类

 请你用三句话来形容八月十五中秋节的月亮，每句话不超过14个字

-  1. 月亮圆圆，如钩挂在天上。
2. 月光洒下，如银钩划过夜空。
3. 月饼如钩，传递着爱与思念。

 请你用三句话来形容八月十五中秋节的月亮，每句话都刚好14个字，中间不含标点。

-  1. 中秋月如钩挂天上
2. 洒下月光如银钩
3. 月饼如钩传爱思念

不会简单计数、不知道时令特点

语言模型并没有掌握常识



三文鱼在水中游



俄罗斯军舰

语言模型并没有学到伦理规则

- ▶ 这一页是草稿
- ▶ 贾宝玉娶贾母的例子：chatgpt-err-jiabaoyujamu.png
- ▶ 类似毁灭世界计划的禁忌很容易被绕过

语言模型无法区分事实与非事实

- ▶ 这一页是草稿
- ▶ 语言模型不知道自己说的话是否有依据（所有训练文本都变成了参数）
- ▶ 语言模型甚至无法区分基于事实的合理推理论和偏离事实的臆想（hallucination）

- ▶ stack overflow已经禁止使用chatgpt生成的代码
- ▶ 假的东西编的越像，危害越大

Content

神经符号计算路在何方

大语言模型并没有真正学到符号操纵的能力

现有符号方法的严重缺陷

神经方法和符号方法结合的路径分析

现有知识图谱覆盖面严重不足

- ▶ 偏重实体，事件覆盖率太低
 - ▶ 关羽：温酒斩华雄、水淹七军
 - ▶ 红楼：宝黛之恋
- ▶ 偏重二元关系（三元组），复杂场景知识不足
 - ▶ 中秋：月亮应该是圆的，人们会吃月饼
 - ▶ 时令：每个季节有时令特色（比如菊花只在秋天开放）
- ▶ 事件发展的因果关系不足
 - ▶ 相关工作：Entailment Graph (Mark Steedman), ASER (宋阳秋)

现有符号方法描述能力严重不足

- ▶ 缺乏对矛盾（否定、反义、互斥）关系的表示和推理能力
 - ▶ 如果前面说了“ A 是个女性”，后面就不应该出现“ A 娶了妻子”
 - ▶ 中秋月亮是圆的，就不应该出现“月如钩”
- ▶ 缺乏基本的集合关系、数量关系理解和推理能力
 - ▶ 下面这个例子，现在的符号方法也没有简单的工具实现这个推理：



有一只狗和一只猫，一只是白色的，一只是黑色的，现在知道猫是黑色的，请问狗是什么颜色？

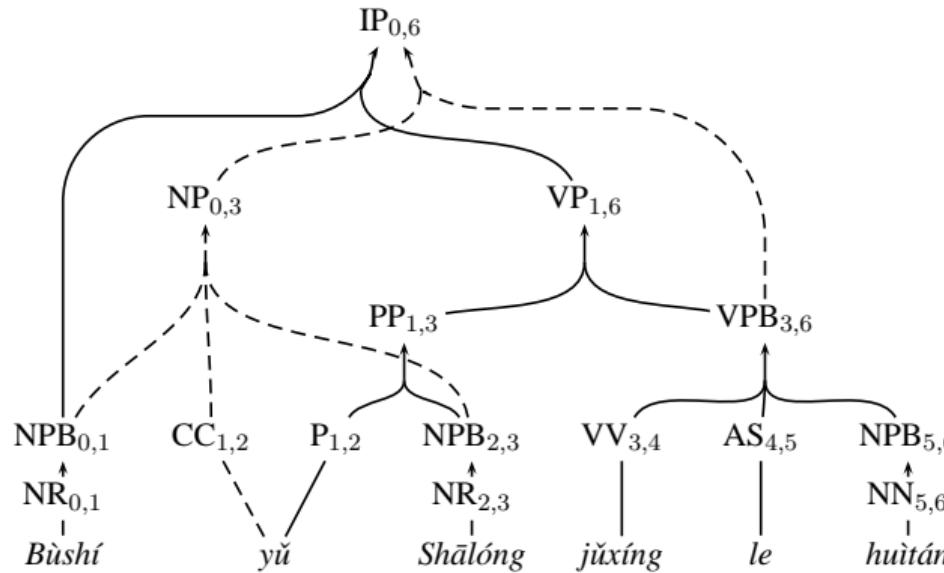


无法回答，因为你所提供的信息并不足以确定狗的颜色。根据你提供的信息，我们只知道猫是黑色的，但是并不知道狗的颜色。为了得到更准确的答案，我们需要更多的信息。

现有符号方法对歧义的兼容能力严重不足

- ▶ 真实语言是存在大量歧义的，人类大脑可以理解并接受这种歧义的存在，不会过早做出判断：
 - ▶ 我买了个苹果。（水果？ or 电脑？）
 - ▶ 南京市长江大桥。（南京市长江大桥 or 南京市长江大桥）
- ▶ 现有符号表示方法大都是确定性的，对保持歧义的表示方法重视严重不足
 - ▶ 贝叶斯网络（适合表示因果推断，跟NLP结合不多）
 - ▶ 马尔科夫逻辑网络（过于复杂）
 - ▶ 语言压缩结构（网格Latice、超图Hypergraph）（适合语言表示，但现在很少研究）
- ▶ 在构建数据集和Benchmark的时候，通常要求标注者的一致性（IAA）较高才予以标注，忽略了语言事实中大量的天然歧义。
- ▶ 神经网络是天然允许歧义的，这是其对于符号方法的巨大优势。

现有符号方法对歧义的兼容能力严重不足



Mi et al., Forest-Based Translation, ACL 2008

现有符号方法缺乏从海量数据中学习的能力

- ▶ 相比于神经网络方法，现有符号方法严重缺乏scaling-up的能力，
 - ▶ 统计机器翻译可以从海量的双语数据中学习翻译知识，但性能比神经网络有较大差距。
 - ▶ 其他符号化NLP方法很少能够从海量数据中学习知识。

Content

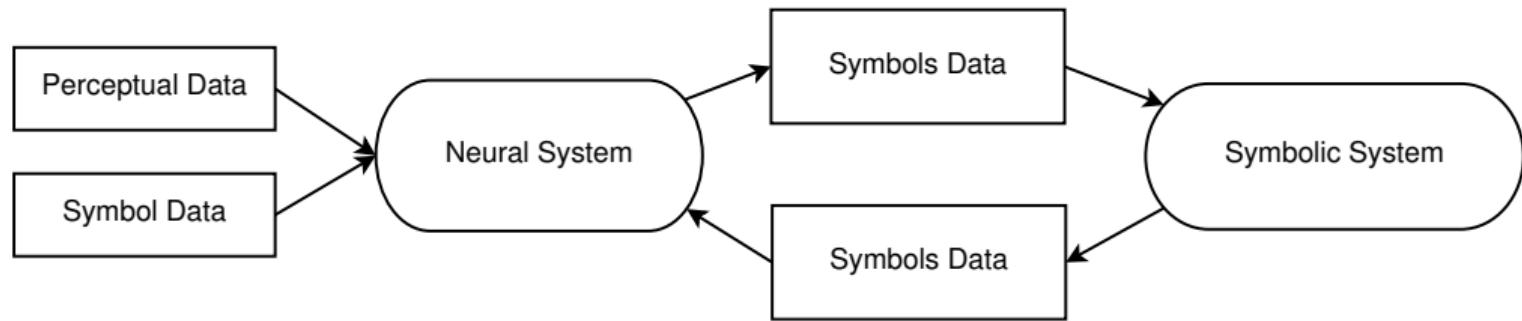
神经符号计算路在何方

大语言模型并没有真正学到符号操纵的能力

现有符号方法的严重缺陷

神经方法和符号方法结合的路径分析

路径一：一种通用的神经符号系统结构（松散耦合）



- ▶ 目前成功的系统基本都可以归入这一结构的某种变化形式。
- ▶ 也是现有争论双方都可以接受的路径。
- ▶ 但目前的方法中，符号系统这一段明显偏弱（见前述分析），急需大大加强。

目前成功的系统基本都可以归入这一结构的某种变化形式

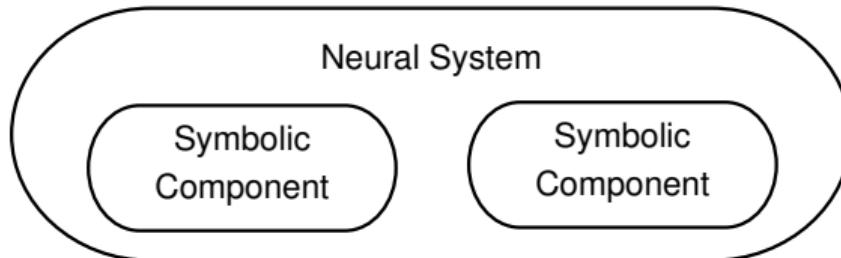
作为神经系统和符号系统桥梁的表示形式

- ▶ 这一页是草稿
- ▶ 目前主流：文本序列（分析优势和劣势）、形式预计（代码、逻辑表达式）
- ▶ 可能的其他方式：图结构、带概率的压缩结构

反馈机制

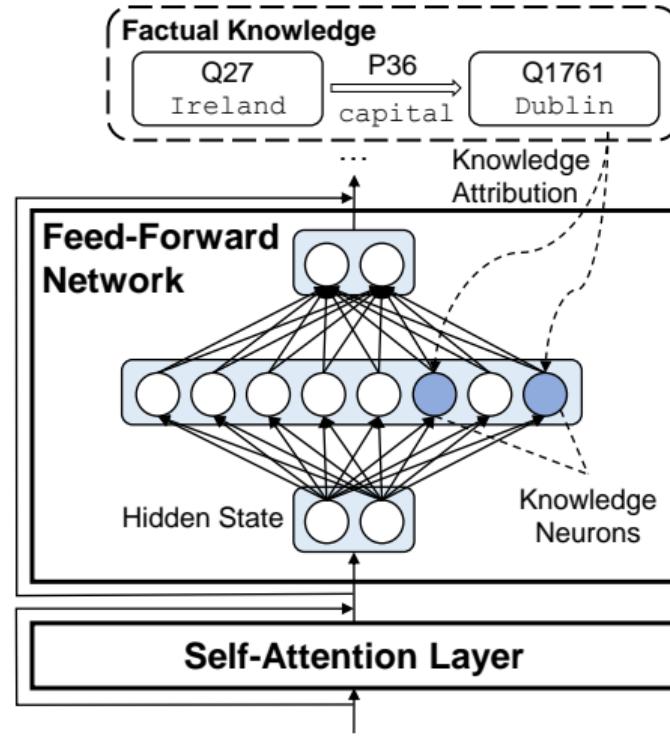
- ▶ 这一页是草稿
- ▶ 强化学习、人机交互

路径二：将符号部件嵌入神经网络（紧耦合）



- ▶ 目前以Hinton、Lecun为首的深度学习阵营强烈反对的结合方式。
- ▶ 目前比较弱势，但仍然有研究：
 - ▶ 打开神经网络的黑盒，可以获得更好的可解释性；
 - ▶ 重要的概念（或实体），可以在神经网络中直接植入一个神经元。

路径二：将符号部件嵌入神经网络（紧耦合）



Dai et al, Knowledge Neurons in Pretrained Transformers, ACL 2022

Content

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

总结与展望

- ▶ 目前神经符号方法取得主要进展采用的是松散结合的模式，但两方并不均衡，神经网络非常强大，符号系统能力偏弱。
- ▶ 目前的神经网络处理符号的能力主要还是依靠大数据模拟，并没有学到真正的符号处理能力，在数据覆盖不足的时候（长尾端）符号处理能力明显不足。
- ▶ 但目前的符号方法也有着明显的缺陷，在覆盖率、表达能力、可扩展性方面都急需改进。
- ▶ 神经符号结合有松散耦合和紧耦合两种路径，目前前者占优，但后者也有优势，都值得研究。

Content

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

Summary

神经符号计算简介

有关神经符号计算的争论

神经符号计算的近期进展

神经符号计算路在何方

总结与展望

Thank you!

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home and organization
for a fully connected, intelligent world.

Copyright©2018 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

