

# 人工智能：游走在神经和符号之间

刘群 LIU Qun

华为诺亚方舟实验室

中国计算语言学大会 CCL 2024, 太原

2024.07.28



NOAH'S ARK LAB



# Content

神经和符号的结合是实现真正的人类水平智能的必经之路

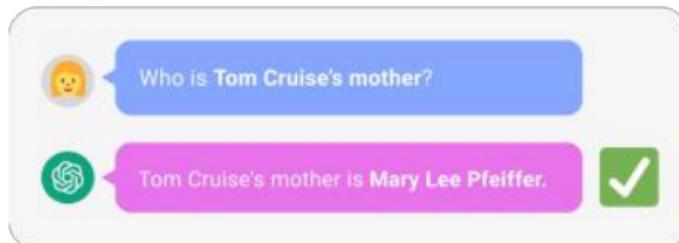
符号化知识表示的类型及其与大语言模型结合的方法

在Transformer架构中引入符号计算模块的设想

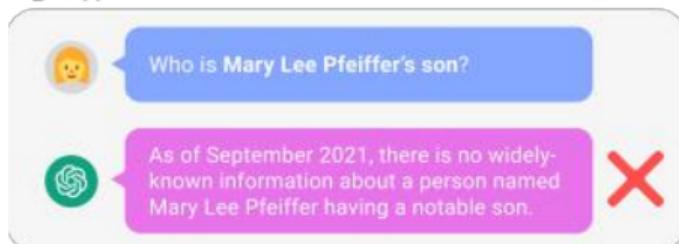
总结

# 大语言模型的符号计算能力仍然欠缺

A → B



B → A



常识理解

ChatGPT

9.8和9.11哪个大?

9.11大于9.8。

为什么?



比较两个小数时，先从左到右逐位比较数字：

1. 比较整数部分：9 和 9 相同。
2. 比较小数点后第一位：1 大于 8。

因此，9.11 大于 9.8。

数学推理

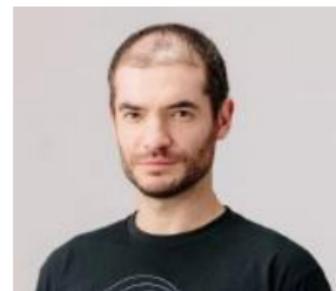
# 实现真正的人类水平智能，需要AI模型中引入符号计算吗？

不需要，神经网络将通过Scaling Law将自然获得符号计算的能力



Geoffery Hilton

需要，神经网络无法精确推理，需研究神经符号结合的方法



Ilya Sutskever



Yoshua Bengio

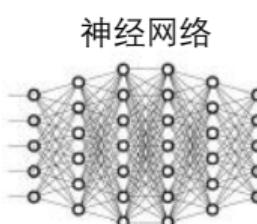


Gary Marcus

# 为什么基于神经网络的语言模型很难达到人类水平智能？

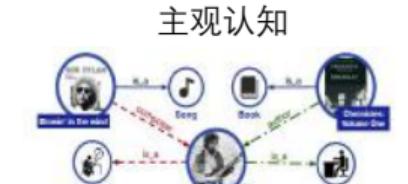
- ▶ 人类的思维可以在系统1和系统2之间自如切换，而目前的大语言模型的大语言模型的神经计算和符号计算是完全脱节的。
- ▶ 人类在说出一个句子之前，大脑里面可以用符号的方式（也就是通过系统2）进行思考，而神经网络模型在输出一个token之前，所有的计算都是基于神经的。
- ▶ 人类大脑中有一个海马体组织，可以保持与时间和空间相关的短期记忆和长期记忆，而目前的神经网络模型缺乏类似的模块。

# 信息系统抽象的层次



Language Learning Models (LLMs) have revolutionized the field of natural language processing, enabling machines to understand and generate human-like text. At the core of LLMs lies the concept of tokens, which serve as the fundamental building blocks for processing and representing text data. In this blog post, we'll deconstruct tokens in LLMs, examining their significance and exploring how they contribute to the power and flexibility of these remarkable models.

## 语言表达

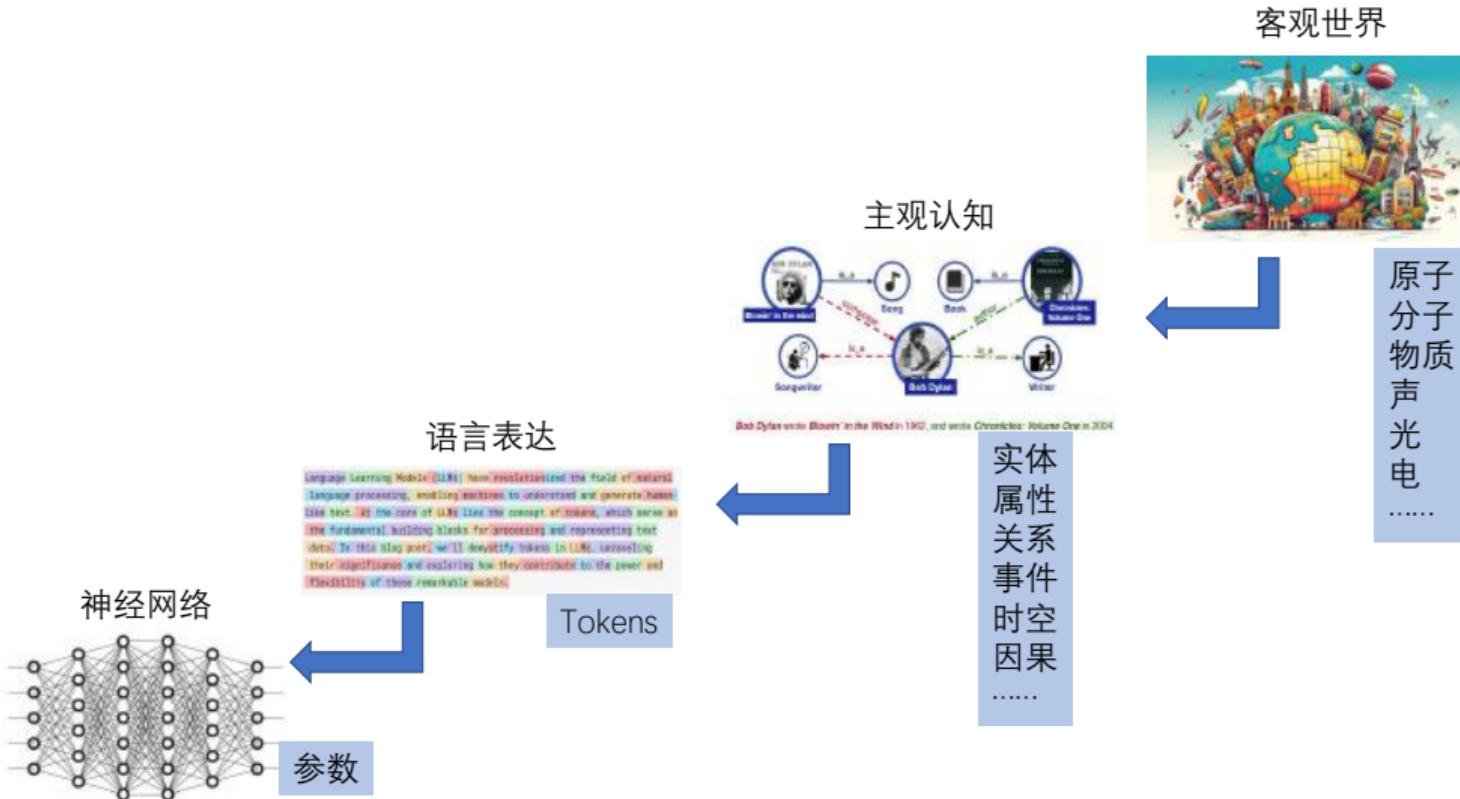


Bob Dylan write 'Blowin' in the Wind in 1962, and write Chronicles: Volume One in 2004



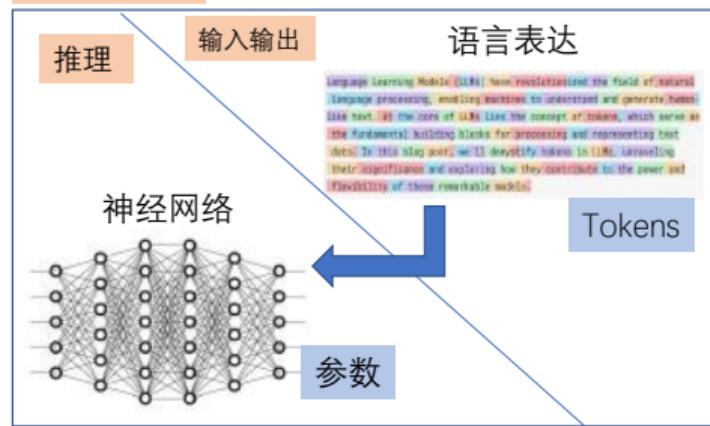
客观世界

# 信息系统抽象的层次

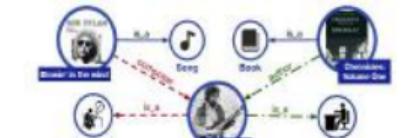


# 信息系统抽象的层次

大语言模型



主观认知



客观世界



原子  
分子  
物质  
声  
光电  
.....

# 神经网络的参数化表示与主观认知之间存在GAP

大语言模型对于最基本的实体、属性、数量等基本的认知都还存在很多问题：

- ▶ 文本生成错误例（来自TGEA2.0）：

- ▶ 血和着鲜血，从她白嫩光洁的面颊上流淌下来。
- ▶ 这套样板房以一个正方形为主题，整个房间被分成了5个部分：客厅、餐厅、书房和卧室。
- ▶ 而辰口发出的四封信和几张明信片也来到耐子手中，她打开一看，是一幅自己画的漫画。

- ▶ 文生视频错误例（来自Sora）：



# 神经与符号的GAP是目前大模型很多问题的根源

- ▶ 人类认知表示最自然的形式是实体、属性、关系、时空、事件、因果等等，这些都最适合用符号来表示
- ▶ 而目前的大语言模型的基本构成单位是参数，所有的计算和推理都发生在参数之间，跟认知所使用的符号表示的形式存在巨大的差异
- ▶ 这种差异是造成目前大模型很多问题的根源



# Content

神经和符号的结合是实现真正的人类水平智能的必经之路

符号化知识表示的类型及其与大语言模型结合的方法

在Transformer架构中引入符号计算模块的设想

总结

# Content

符号化知识表示的类型及其与大语言模型结合的方法

**符号化知识表示的类型**

        自然语言与神经网络结合的方法

        形式语言与神经网络结合的方法

        图表语言与神经网络结合的方法

# 符号化知识表示的多样性难题

- ▶ 神经网络为所有问题提供统一的解决方案，而符号化知识形式非常多样化，每种形式特点不同，面临的问题也不同
- ▶ 不同的符号化知识表示形式，需要采用不同的方式与神经网络结合
- ▶ 是否存在统一的形式化知识表示形式，可以解决所有符号推理问题？
- ▶ 如果不存在，有哪些主要的符号化知识表示形式？各自有什么特点？如何与大语言模型结合进行推理？

# 符号化知识表示的形式

- ▶ Declarative knowledge 陈述性知识
  - ▶ 描述概念、实体、事实
  - ▶ 以陈述句形式描述
- ▶ Procedural knowledge 过程性知识
  - ▶ 包括规则、策略、过程、议程等
  - ▶ 可以执行完成任务
- ▶ Meta-knowledge 元知识
  - ▶ 关于知识的知识
- ▶ Heuristic knowledge 启发式知识
  - ▶ 专家根据经验获得的领域或专业知识
- ▶ Structural knowledge 结构化知识
  - ▶ 概念之间的上下位或者整体部分关系等知识
  - ▶ 解决问题所需要的知识



Source: <https://www.javatpoint.com/knowledge-representation-in-ai>

# 符号化知识表示的类型

- ▶ 自然语言
  - ▶ 词、短语、句子、篇章
- ▶ 语义网（Semantic Web）、知识图谱
  - ▶ 知识本体（Ontologies）
  - ▶ 实体知识图谱（Entities/Relations/Facts）
  - ▶ 事理知识图谱（Events）
- ▶ 程序语言（规则都可以表示为程序）
  - ▶ 函数式程序语言、过程式程序语言、面向对象程序语言
- ▶ 逻辑语言
  - ▶ 布尔逻辑、命题逻辑、描述逻辑、构造逻辑、一阶谓词逻辑、高阶谓词逻辑
- ▶ 图表语言
  - ▶ 表格（Tables、Spreadsheets）、图（graphs）、自由图表（Diagrams）

# 符号化知识表示的类型

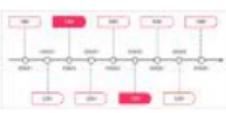
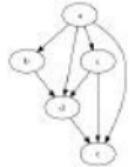
自然语言

中文, English, Español, Français, Deutsch, 日本語, Русский, Português, Italiano, 한국어

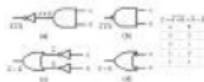
图表语言



表格



图



自由图表

形式语言

产生式规则

C C++

Python

JavaScript

Latex

Json

HTML

Markdown

知识图谱

程序语言

描述逻辑

命题逻辑

布尔逻辑

构造逻辑

一阶谓词逻辑

高阶谓词逻辑

逻辑语言

# Content

符号化知识表示的类型及其与大语言模型结合的方法

符号化知识表示的类型

自然语言与神经网络结合的方法

形式语言与神经网络结合的方法

图表语言与神经网络结合的方法

# 自然语言与神经网络结合的方法

- ▶ 自然语言是最自然的符号表示形式
- ▶ 语言也是对大语言模型最亲和的一种符号表示形式
- ▶ 但自然语言本身的推理能力有限
- ▶ 自然语言与神经网络结合可以通过Chain-of-Thought及其各种衍生的X-of-Thought来实现复杂的推理
- ▶ 自然语言推理通常缺乏自动化的验证方法，只能通过人工来验证，很难通过自动化的方式来迭代提高

# Chain-of-Thought and Thinking-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. ✓

# Self-consistency improves CoT

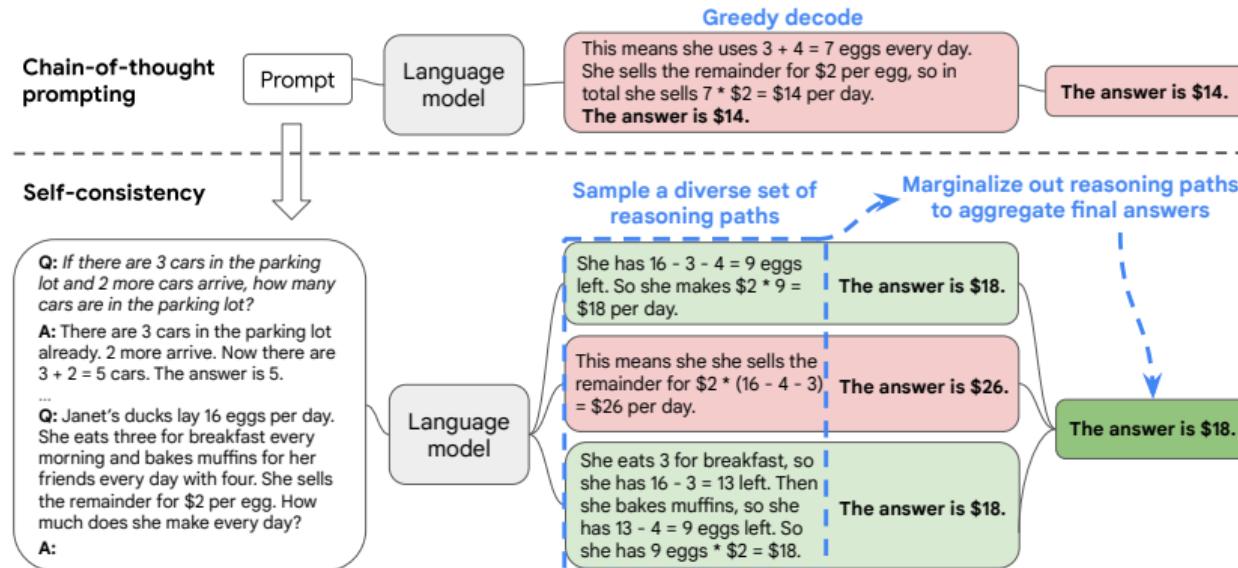


Figure 1: The self-consistency method contains three steps: (1) prompt a language model using chain-of-thought (CoT) prompting; (2) replace the “greedy decode” in CoT prompting by sampling from the language model’s decoder to generate a diverse set of reasoning paths; and (3) marginalize out the reasoning paths and aggregate by choosing the most consistent answer in the final answer set.

Wang et al., Self-Consistency Improves Chain Of Thought Reasoning In Language Models, arXiv:2203.11171v4, 2022

# Progressive-Hint Prompting improves CoT

## Base Prompting

Q: There are 96 fourth-graders at Small Tree School. 43 of them are girls. On Friday, 5 fourth-grade girls and 4 fourth-grade boys were absent. How many fourth grade boys were at Small Tree School on Friday?



Base Answer

A: There were 96 fourth-graders at Small Tree School. 43 of them were girls. On Friday, 5 girls and 4 boys were absent. So there were 43 - 5 = 38 girls at school on Friday, and 96 - 38 = 58 boys at school on Friday. The answer is 58.

## Progressive-Hint Prompting

Q: There are 96 fourth-graders at Small Tree School. 43 of them are girls. On Friday, 5 fourth-grade girls and 4 fourth-grade boys were absent. How many fourth grade boys were at Small Tree School on Friday? (Hint: The answer is near to 58).



Second Answer

A: We know the Answer Hints: 58. With the Answer Hints: 58, we will answer the question. There were 96 fourth-graders at Small Tree School, and 43 of them were girls. On Friday, 5 girls and 4 boys were absent. So there were 43 - 5 = 38 girls and 96 - 43 = 53 boys at Small Tree School on Friday. After subtracting the absent students, there were 38 - 4 = 34 girls and 53 - 5 = 48 boys. The answer is 48.

## Progressive-Hint Prompting

Q: There are 96 fourth-graders at Small Tree School. 43 of them are girls. On Friday, 5 fourth-grade girls and 4 fourth-grade boys were absent. How many fourth grade boys were at Small Tree School on Friday? (Hint: The answer is near to 58, 48).



Third Answer

A: We know the Answer Hints: 58, 48. With the Answer Hints: 58, 48, we will answer the question. There were 96 fourth-graders, and 43 of them were girls. On Friday, 5 girls and 4 boys were absent. So there were 43 - 5 = 38 girls and 96 - 9 = 87 students in total. Of those 87 students, 38 were girls, so there must have been 87 - 38 = 49 boys. The answer is 49.

## Progressive-Hint Prompting

Q: There are 96 fourth-graders at Small Tree School. 43 of them are girls. On Friday, 5 fourth-grade girls and 4 fourth-grade boys were absent. How many fourth grade boys were at Small Tree School on Friday? (Hint: The answer is near to 58, 48, 49).



Fourth Answer

A: We know the Answer Hints: 58, 48, 49. With the Answer Hints: 58, 48, 49, we will answer the question. There were 96 fourth-graders at Small Tree School, and 43 of them were girls. On Friday, 5 girls and 4 boys were absent. So there were 43 - 5 = 38 girls and 96 - 43 = 53 boys originally. After the absences, there were 38 - 5 = 33 girls and 53 - 4 = 49 boys. The answer is 49.

# Content

符号化知识表示的类型及其与大语言模型结合的方法

符号化知识表示的类型

自然语言与神经网络结合的方法

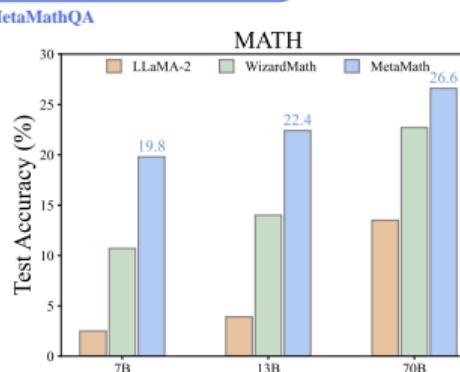
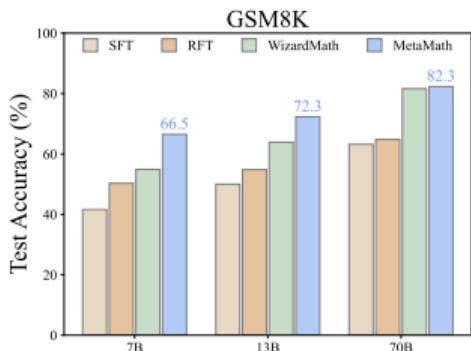
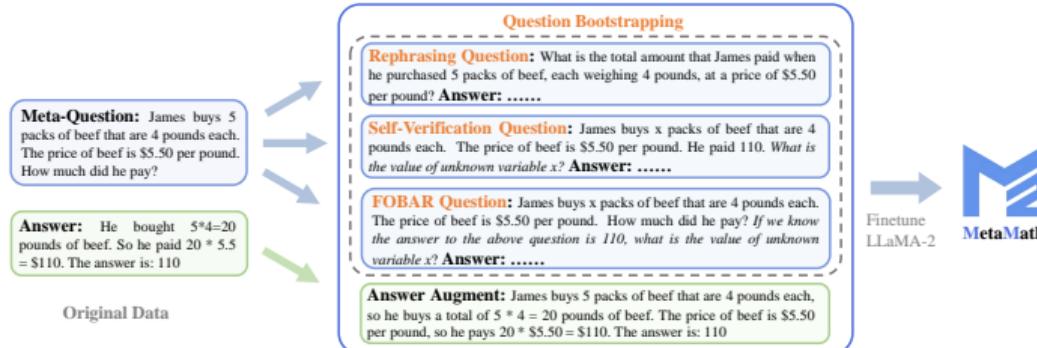
形式语言与神经网络结合的方法

图表语言与神经网络结合的方法

# 形式语言与神经网络结合的方法

- ▶ 形式语言主要有两类：程序语言和逻辑语言。
- ▶ 形式语言都没有歧义，是最精确的符号表示形式。
- ▶ 形式语言与神经网络的结合都面临两类问题：
  - ▶ 自然语言到形式语言的转换问题：如何自将然语言描述的问题转换成形式语言（逻辑或程序）：
    - ▶ 这更多是自然语言理解问题
    - ▶ 可以采用XoT方法提高推理准确率
    - ▶ 或者采用数据合成方法构造更多更好的训练数据
  - ▶ 形式语言本身的生成问题：如何生成正确的程序语言：
    - ▶ 由于形式语言自身有明确的语义，因此是可以通过引入外部符号引擎进行验证
    - ▶ 可以采用蒙特卡洛搜索等方法寻找更好的结果
- ▶ 在特定约束条件下，数学定理证明和函数式编程存在某种严格的对应关系（科里-霍华德同构），因此数学定理证明问题可以转换成代码生成问题。

# MetaMath: 通过训练数据增强改进LLM数学问题求解能力



Yu et al., Metamath: Bootstrap Your Own Mathematical Questions For Large Language Models, arXiv:2309.12284v4

# The Curry-Howard Isomorphism 科里-霍华德同构

- ▶ 柯里-霍华德对应（英语：Curry-Howard correspondence）是在计算机程序和数学证明之间的紧密联系；这种对应也叫做柯里-霍华德同构、公式为类型对应或命题为类型对应。
- ▶ 这是对形式逻辑系统和公式计算（computational calculus）之间符号的相似性的推广。
- ▶ 它被认为是由美国数学家哈斯凯尔·布鲁克·柯里（Haskell Brooks Curry）和逻辑学家威廉·阿尔文·霍华德（William Alvin Howard）独立发现的。
- ▶ 有多种方式考虑柯里-霍华德对应。
- ▶ 在一个方向上，它工作于“把证明编译为程序”级别上。这里的“证明”最初被限定为在构造性逻辑中一典型的是直觉逻辑系统中的证明。而“程序”是在常规的函数式编程的意义上的；从语法的观点上看，这种程序是用某种 $\lambda$ 演算表达的。所以柯里-霍华德同构的具体实现是详细研究如何把来自直觉逻辑的证明写成 $\lambda$ 项。

# The Curry-Howard Isomorphism 科里-霍华德同构

## The mathematician

*Theorem.* For all  $n \in \mathbb{N}$ , there exists  $p \in \mathbb{N}$  such that  $n = 2p$  or  $n = 2p + 1$ .

*Proof.* By induction on  $n$ .

- If  $n = 0$  then this is obvious.
- Otherwise, assume that  $n = m + 1$ . By the induction hypothesis, we know that there exists some  $p$  such that  $m = 2p$  or  $m = 2p + 1$ .
  - In the first case,  $n = 2p + 1$ .
  - Otherwise  $n = 2(p + 1)$ .

## The programmer

```
val div2 : int -> int * bool
(* [div2 n] returns the integer
division by 2 of [n] together with
a boolean indicating if [n] is
even. *)
```

```
let rec div2 n = match n with
| 0 -> (0, true)
| m + 1 ->
    let (p, even) = div2 m in
        if even then (p, false)
        else (p + 1, true)
```

Slides: Pierre-Marie Pédrot, The Curry-Howard isomorphism for Dummies

# The Curry-Howard Isomorphism 科里-霍华德同构

Logic	CS
Proofs	Programs
Formula	Types
$A \text{ implies } B$	function from $A$ to $B$
$A \text{ and } B$	pair of $A$ and $B$
$A \text{ or } B$	tagged union of $A$ and $B$
falsity	empty type
truth	singleton type
for all $x \in A$ , $B(x)$	dependent product from $A$ to $B$
Axiom	System primitive
Soundness theorem	Compiler
Completeness theorem	Debugger
Incompleteness theorem	Infinite loop

Slides: Pierre-Marie Pédrot, The Curry-Howard isomorphism for Dummies

## Lean语言

- ▶ Lean是一款在包含归纳类型的构造演算基础上所开发的计算机定理证明辅助工具和函数式编程语言。
- ▶ Lean语言既是一种函数式编程语言，
- ▶ Lean语言又是一种形式化数学定理证明工具，用Lean语言写的数学定理证明可以保证其正确性。
- ▶ 包括陶哲轩在内的一些数学家计划将现有的大部分数学定理证明都用Lean语言表示出来，目前已经有了一个初步的定理证明库MathLib
- ▶ Lean语言也成为了基于AI进行数学定理证明的有效工具

# Theorem Proving - Holy Grail of AI

- Very general and most challenging form of intelligence



Hilbert



Turing



Shannon

...



Sutskever



Lample

- Special cases: SAT, SMT, first-order logic, math word problems
- Applications:

- **Formal verification** => 100% correct code with theoretical guarantee
- **Code generation** => assist/replace coders (and with 100% correctness)
- **AI for Math** => education, solve open problems, create new algorithms

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Automated Theorem Proving (ATP) - the Problem

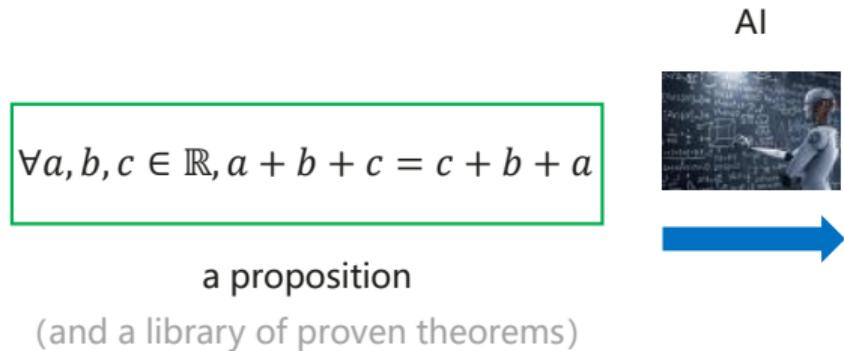
$$\forall a, b, c \in \mathbb{R}, a + b + c = c + b + a$$

a proposition

(and a library of proven theorems)

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Automated Theorem Proving (ATP) - the Problem



Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Automated Theorem Proving (ATP) - the Problem

$$\forall a, b, c \in \mathbb{R}, a + b + c = c + b + a$$

a proposition

(and a library of proven theorems)

AI



**State:**  $\vdash a + b + c = c + b + a$

**Action:**  $\vdash a + b + c = b + c + a$

1. Use add\_comm on c and b

$\vdash a + b + c = a + (b + c)$

2. Use add\_comm on  $(b + c)$  and a

$\vdash a + b + c = a + (b + c)$

3. This is exactly add\_assoc

goals accomplished

its proof

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Automated Theorem Proving (ATP) - the Problem

$$\forall a, b, c \in \mathbb{R}, a + b + c = c + b + a$$

a proposition

(and a library of proven theorems)

AI



**State:**  $\vdash a + b + c = c + b + a$

**Action:**  $\vdash a + b + c = c + b + a$

1. Use add\_comm on c and b

$\vdash a + b + c = b + c + a$

2. Use add\_comm on  $(b + c)$  and a

$\vdash a + b + c = a + (b + c)$

3. This is exactly add\_assoc

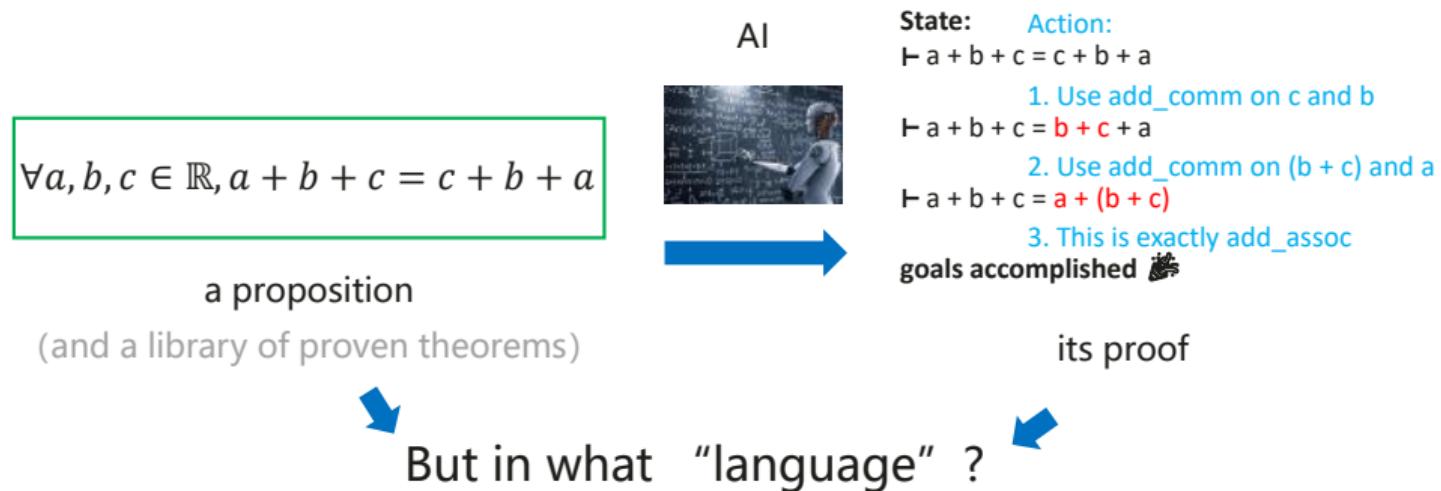
goals accomplished 

its proof

*(undecidable in most cases, by Gödel Incompleteness Theorem. But it suffices to achieve human performance.)*

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Automated Theorem Proving (ATP) - the Problem



Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Automated Theorem Proving (ATP) - the Problem

$\forall a, b, c \in \mathbb{R}, a + b + c = c + b + a$

a proposition

(and a library of proven theorems)

AI



We use: **Lean** theorem prover

```
import data.real.basic

theorem add_abc (a b c : ℝ):
  a + b + c = c + b + a := 
begin
  rw add_comm c b,
  rw add_comm (b + c) a,
  exact add_assoc a b c,
end
```

its proof

But in what “language” ?

*(undecidable in most cases, by Gödel Incompleteness Theorem. But it suffices to achieve human performance.)*

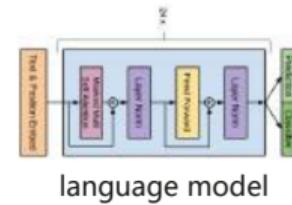
Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment



lean-gym



language model

[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

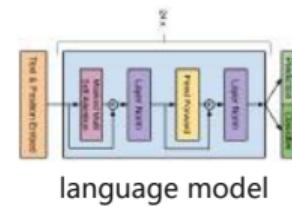
# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment



lean-gym

init\_search:  
**add\_abc real**  
(theorem name, namespaces)



language model

[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

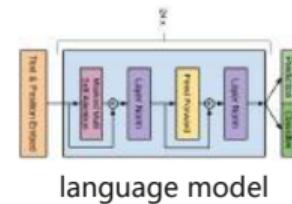
lean-gym [1] provides a theorem proving environment



lean-gym

init\_search:  
**add\_abc real**  
(theorem name, namespaces)

$\vdash a + b + c = c + b + a$   
(initial tactic state)



language model

[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment

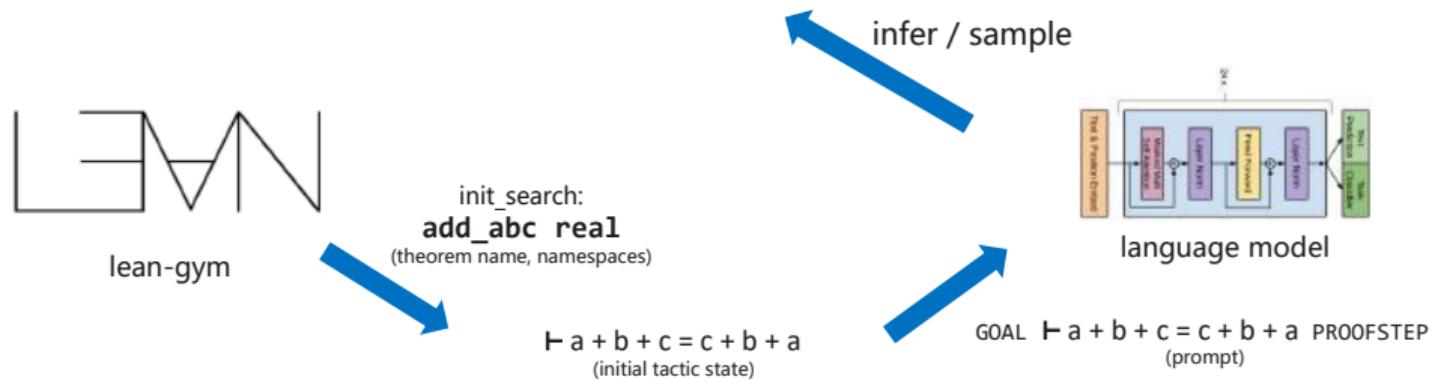


[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment

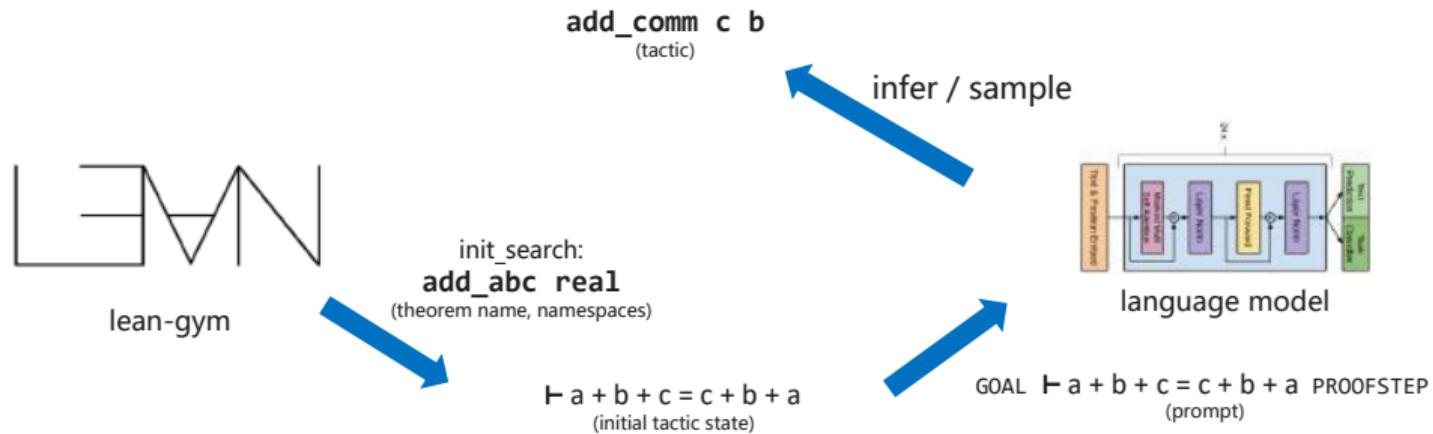


[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment

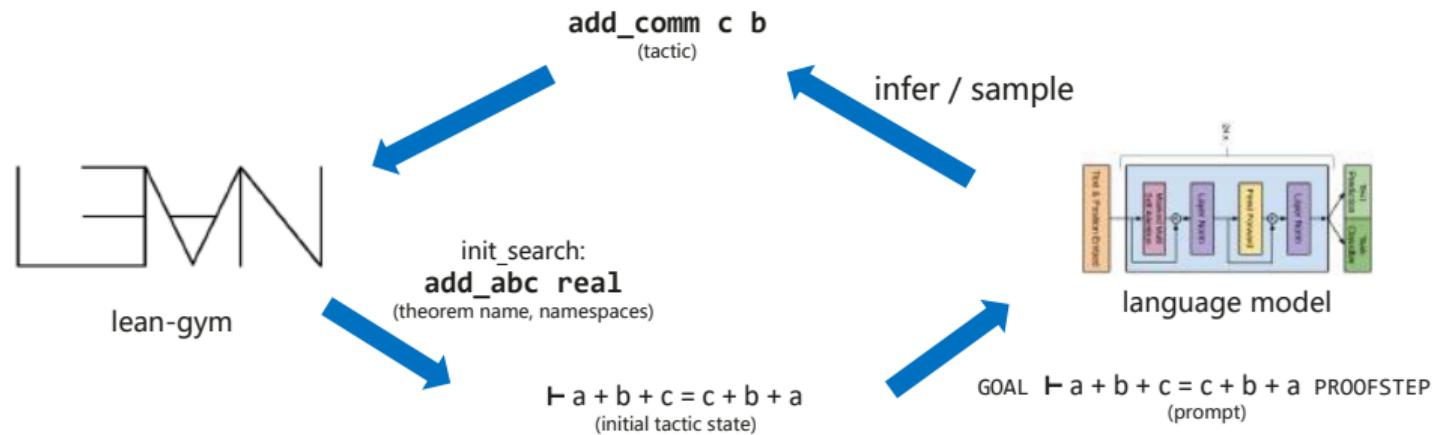


[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment

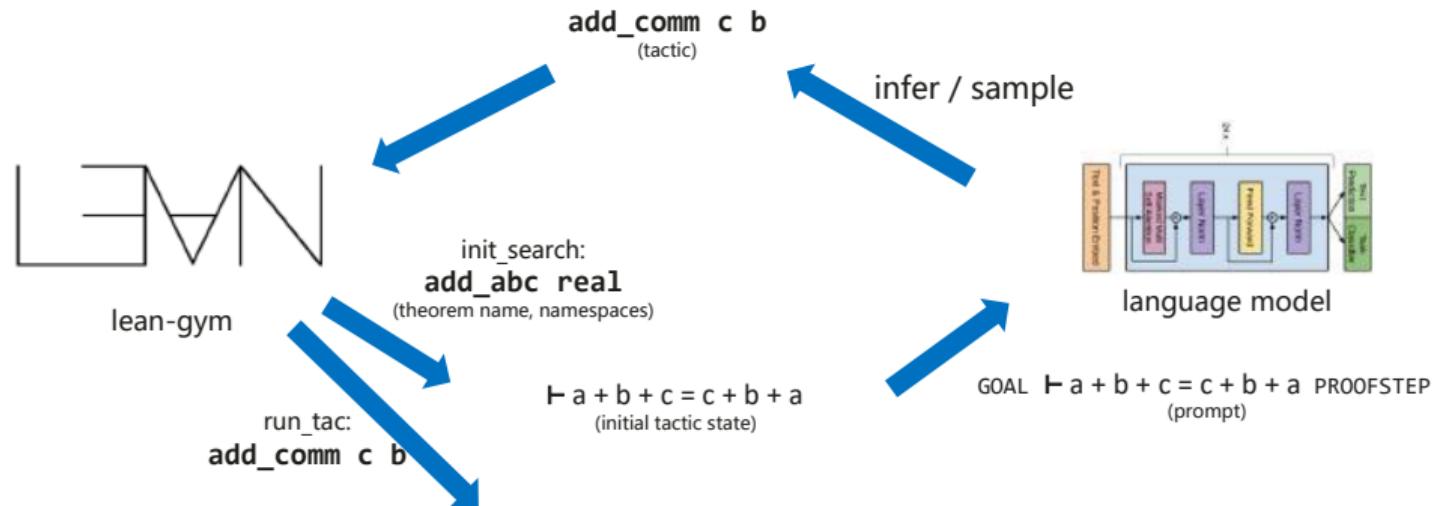


[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment

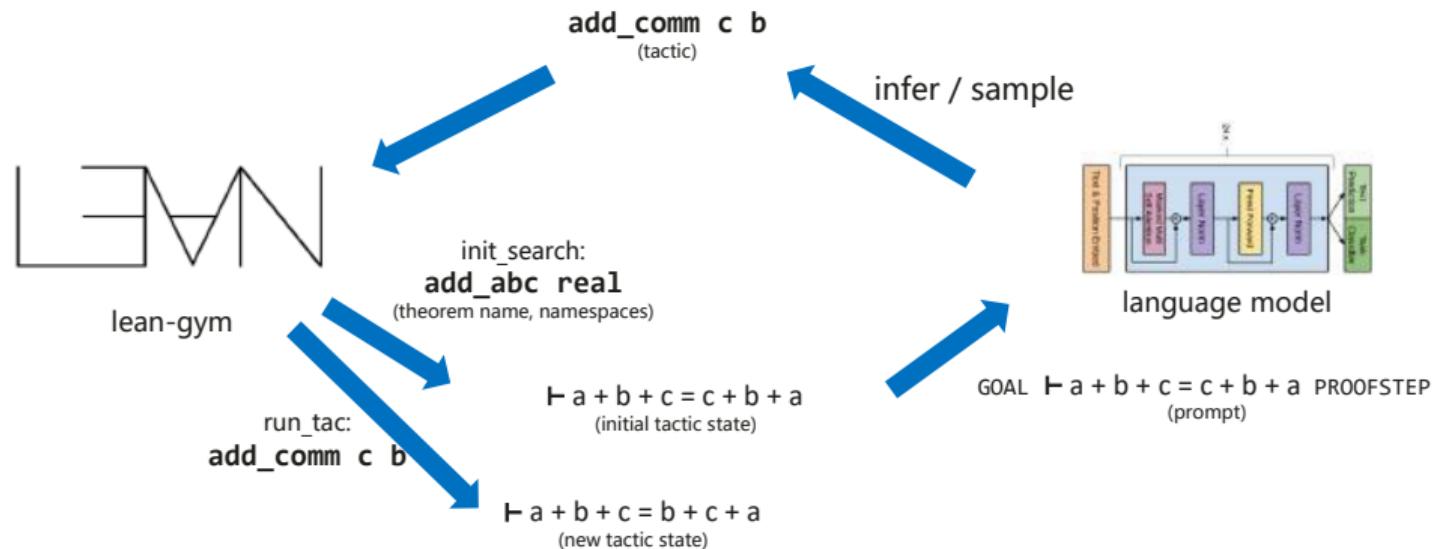


[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

lean-gym [1] provides a theorem proving environment

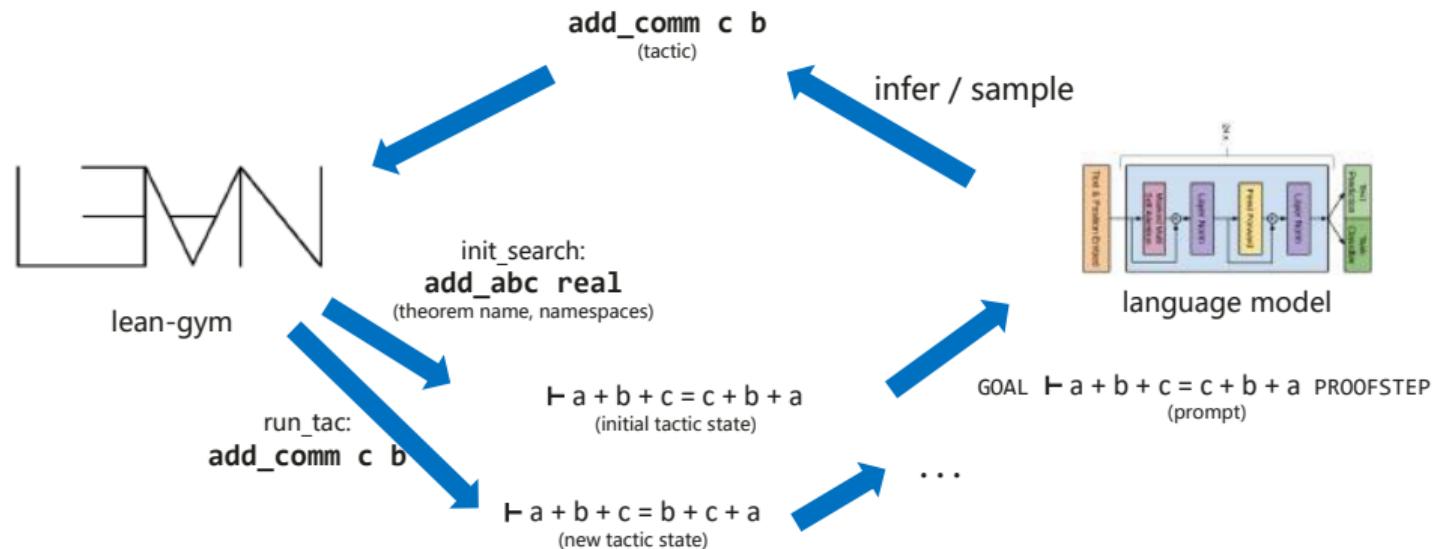


[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# Interaction between the prover (Lean) and the language model

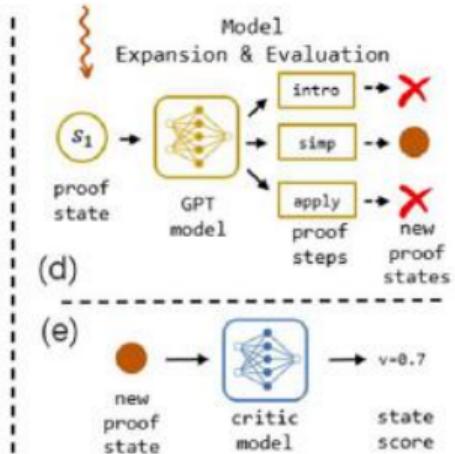
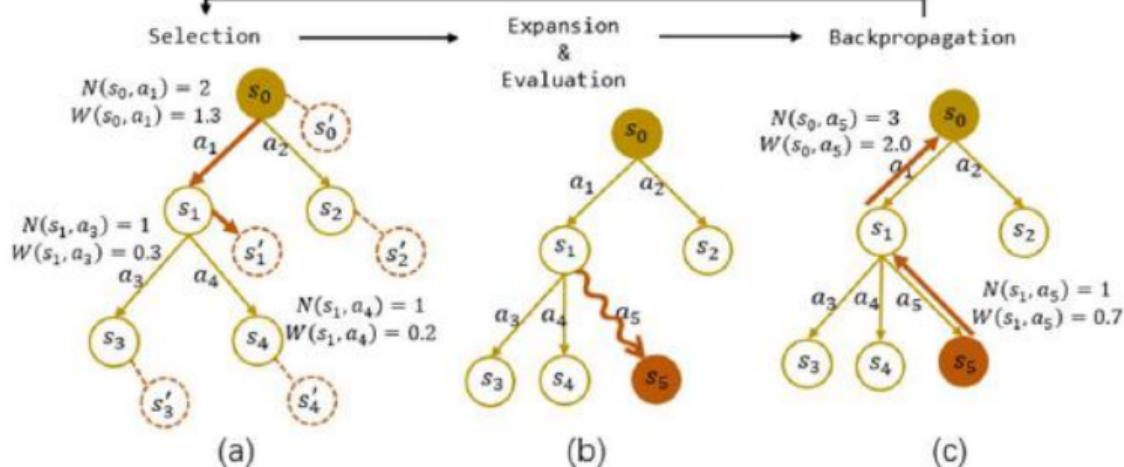
lean-gym [1] provides a theorem proving environment



[1] Polu, S. et al. Formal Mathematics Statement Curriculum Learning. 2022.

Credit to Zhengying Liu at Huawei Noah's Ark Lab

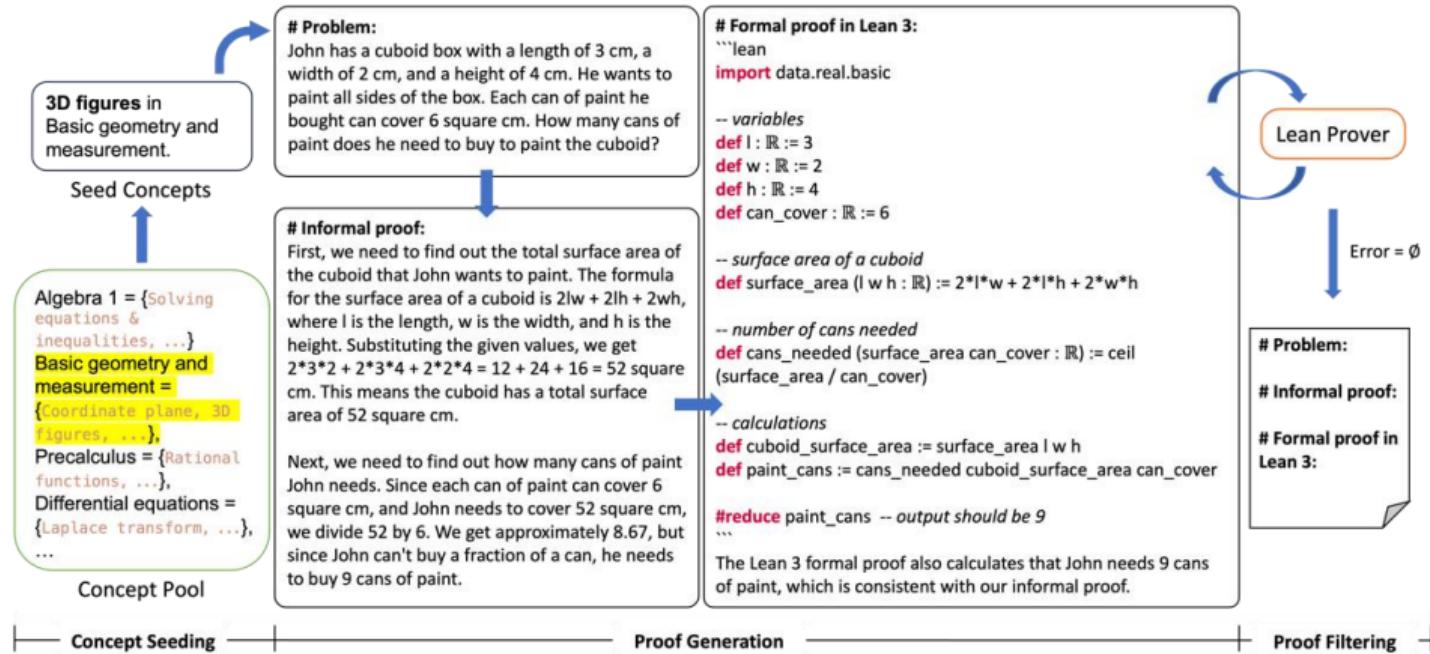
# DT-Solver (ACL 2023)



Wang et al., DT-Solver: Automated Theorem Proving with ..., ACL 2023

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# MUSTARD (ICLR 2024)



Y. Huang et al., MUSTARD: Mastering Uniform Synthesis of Theorem and Proof Data, ICLR 2024

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# LEGO-Prover (ICLR 2024)

LEGO-Prover = **Prover** + **Evolver**

**Prover**: the prover proves the theorem modularly using the retrieved skill.

Input:

- informal & formal statement
- 6 retrieved skills from skill library

Output:

- formal proof
- new skill

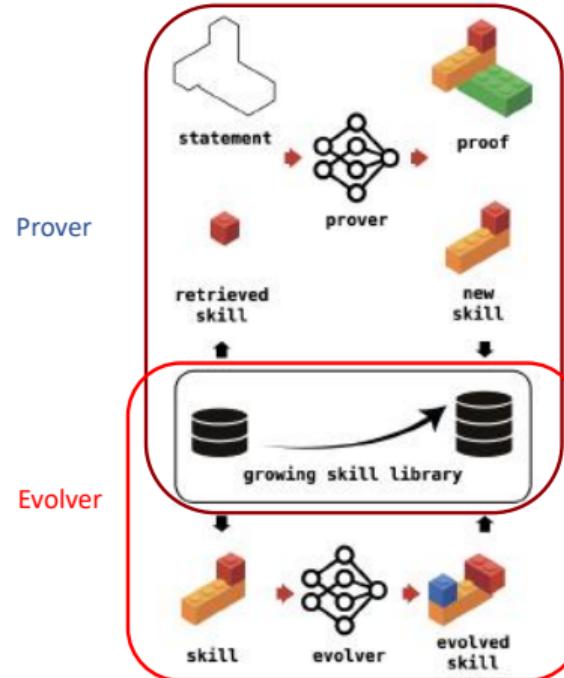
**Evolver**: the evolver transforms the skill for reusability and generalizability.

Input:

- Skill in the skill library

Output:

- Verified evolved skill



Wang et al., LEGO-Prover: Neural Theorem Proving With Growing Libraries, ICLR 2024

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# LEGO-Prover (ICLR 2024)

Table 1: **Proving success rates on the miniF2F dataset with Isabelle.** The table displays the success rates of previous works and the LEGO-Prover. The highest success rates for each set are highlighted in bold. LEGO-Prover\* denotes the cumulative pass rate of the miniF2F dataset, considering the total number of problems solved using model-generated and human-written proofs.

Success rate	LLM	miniF2F-valid	miniF2F-test
<i>Baselines</i>			
Thor (Jiang et al., 2022a)	-	28.3%	29.9%
Thor + expert iteration (Wu et al., 2022)	Codex	37.3%	35.2%
Draft, sketch, and Prove (Jiang et al., 2022b)	Codex	42.6%	39.3%
Subgoal-Learning (Zhao et al., 2023)	ChatGPT	48.0%	45.5%
<i>Ours (100 attempts)</i>			
LEGO-Prover (model informal proof)	ChatGPT	52.4%	45.5%
LEGO-Prover (human informal proof)	ChatGPT	55.3%	<b>50.0%</b>
LEGO-Prover*	ChatGPT	<b>57.0%</b>	<b>50.0%</b>
<i>Ablations (50 attempts)</i>			
LEGO-Prover	ChatGPT	50.4%	-
- Skill Library	ChatGPT	47.1%	-

Wang et al., LEGO-Prover: Neural Theorem Proving With Growing Libraries, ICLR 2024

Credit to Zhengying Liu at Huawei Noah's Ark Lab

# DeepMind: solve IMO problems at a silver medalist level



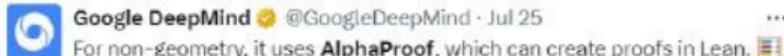
We're presenting the first AI to solve International Mathematical Olympiad problems at a silver medalist level. 🥈

It combines **AlphaProof**, a new breakthrough model for formal reasoning, and **AlphaGeometry 2**, an improved version of our previous system. 🧠 [dpmd.ai/imo-silver](https://dpmd.ai/imo-silver)



Our system had to solve this year's six IMO problems, involving algebra, combinatorics, geometry & number theory. We then invited mathematicians [@wtgowers](#) and Dr Joseph K Myers to oversee scoring.

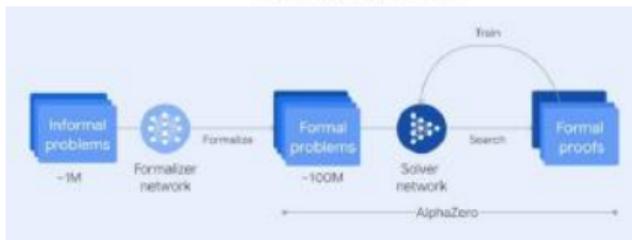
It solved 4 problems to gain 28 points - equivalent to earning a silver medal. ↓



For non-geometry, it uses **AlphaProof**, which can create proofs in Lean. 📄

It couples a pre-trained language model with the AlphaZero reinforcement learning algorithm, which previously taught itself to master games like chess, shogi and Go. [dpmd.ai/imo-silver](https://dpmd.ai/imo-silver)

Score on IMO 2024 problems



# Content

## 符号化知识表示的类型及其与大语言模型结合的方法

符号化知识表示的类型

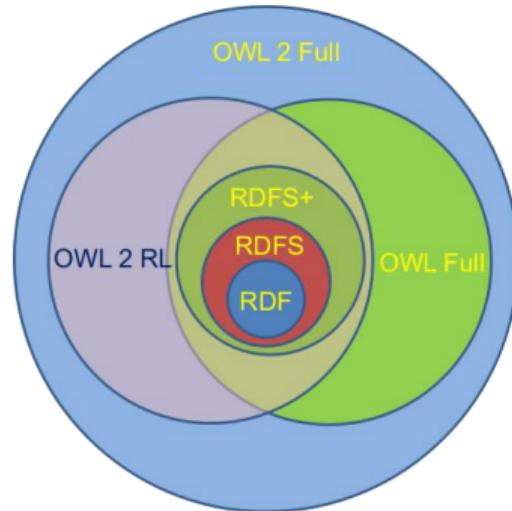
自然语言与神经网络结合的方法

形式语言与神经网络结合的方法

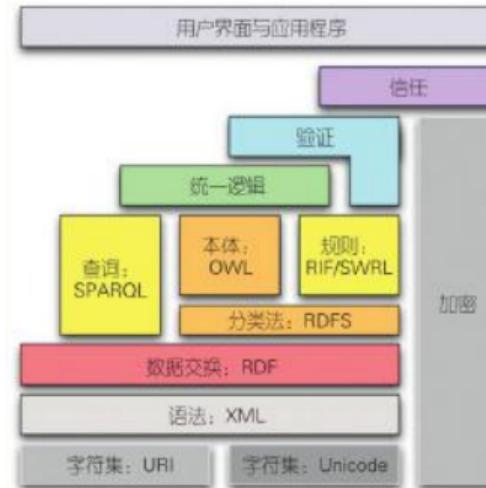
图表语言与神经网络结合的方法

# 知识图谱、语义网 Semantic Web

- ▶ 在W3C推动下，语义网有一套比较完备的形式化语义描述体系，包括RDF、Schema、Ontology、N-Tuples、SPARQL等，具备描述逻辑的表达能力



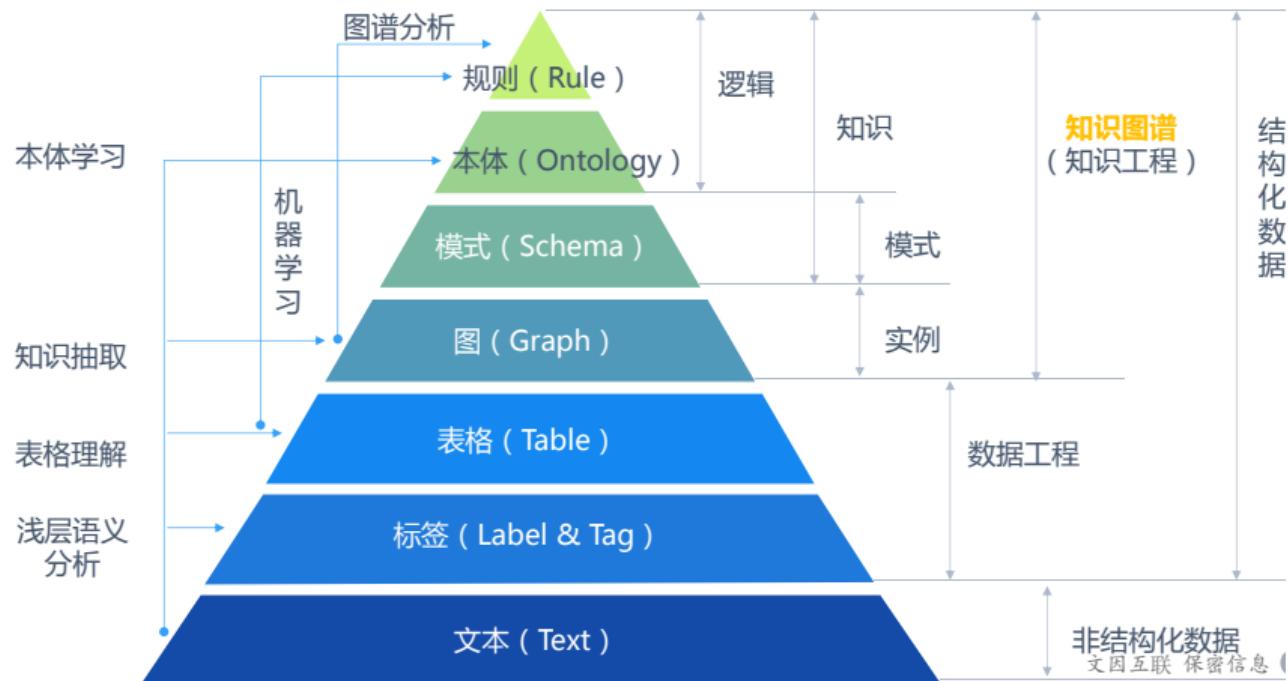
知识图谱语言：RDF and OWL



W3C语义网技术栈

Slides: 鲍捷：从语义网到知识图谱——Web知识技术体系的变迁

# 知识图谱技术金字塔



Slides: 鲍捷: 从语义网到知识图谱——Web知识技术体系的变迁

# 知识图谱与神经网络的结合

- ▶ 语义网和知识图谱的大规模实践表明，逻辑形式的表示在实践中对语义的精确性要求过于严格，不具备可行性
- ▶ 在实践中，知识图谱被大大简化，仅仅表示为三元组形式的实体图谱和事理图谱，以及概念层次结构
- ▶ 简化后的三元组式的知识图谱（包括事理图谱）在搜索引擎和信息推荐等特定领域获得了较广泛的应用
- ▶ 在通用领域，知识图谱仍然面临知识覆盖率太低的问题，特别是与大语言模型相比
- ▶ 自由文本知识图谱（Free-Text Knowledge Graph）允许对实体和关系采用任意自然语言描述，可以一定程度上缓解知识图谱表达能力的不足，但依然很受限。
- ▶ 知识图谱与神经网络（大语言模型）结合的方式：
  - ▶ 使用大语言模型自动生成知识图谱：准确率无法保障，覆盖率仍然不足；
  - ▶ 把知识图谱转换成文本形式用于语言模型预训练：数据量被其他预训练数据淹没，效果有限；
  - ▶ 通过检索增强（RAG）方式实时查询知识图谱用于推理：可以有效提高推理准确性，减少幻觉。

# GraphRAG

- ▶ 什么是GraphRAG?
  - ▶ GraphRAG是一种基于知识图谱的检索增强技术。通过构建图模型的知识表达，将实体和关系之间的联系用图的形式展示出来，然后利用大语言模型（LLM）进行检索增强。
- ▶ GraphRAG 的工作原理:
  - ▶ 提取实体：从用户输入的查询中提取关键实体。
  - ▶ 构建子图：根据提取的实体构建相关的子图，形成上下文。
  - ▶ 生成答案：将构建好的子图输入大语言模型，生成答案。
- ▶ GraphRAG引起了较多的重视，取得了一定的成功。

资料来源：CSDN Blog: GraphRAG: 知识图谱+大模型, 作者: Python\_金钱豹

# 符号化知识表示的其他形式

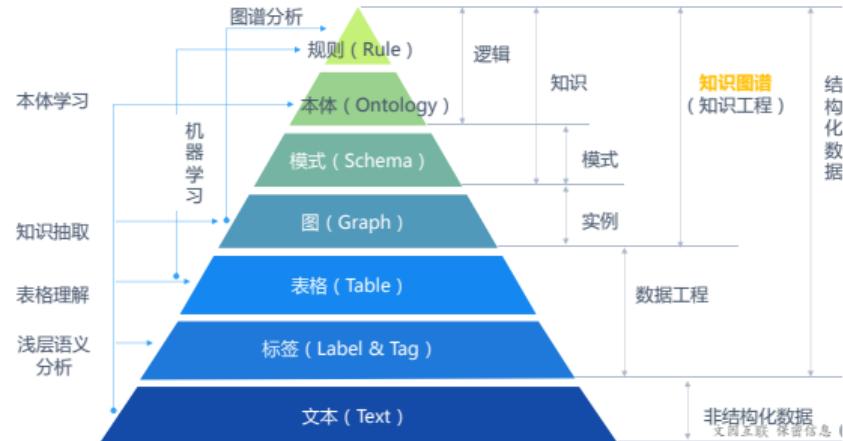
- ▶ 除了自然语言、知识图谱、程序代码、逻辑命题之外，还存在很多其他的符号化表示形式：
  - ▶ 事件时间线
  - ▶ 思维导图
  - ▶ 表格
  - ▶ 电路图
  - ▶ 日历
  - ▶ 建筑设计图
  - ▶ 演示胶片
  - ▶ 广告设计图
  - ▶ .....
- ▶ 大量的各种符号化知识都存在于这些非正规的表示形式中，缺乏系统全面的梳理
- ▶ 其中部分图表可以转化为专业的描述语言（如电路图、设计图）
- ▶ 大量图表都无法表示成形式语言的描述，只能以图片形式保存

# 图像 (image) 作为大语言模型和符号化知识的接口

- ▶ 相比语言，图像可以提供更多、更直观的信息：

设想一下，如果想用文字传达右图的所有信息，应该如何表述？

- ▶ 如何表述金字塔的层次结构？
- ▶ 如何表述颜色深浅传达的信息？
- ▶ 如何表述周边的文字标注和金字塔层级的关系？
- ▶ 直接把这种非规范的图表以图像形式输入到多模态大模型中，也是一种合理的神经符号结合方法。



# 图像 (image) 作为大语言模型和符号化知识的接口

## Glyce: Glyph-vectors for Chinese Character Representations

Yuxian Meng\*, Wei Wu\*, Fei Wang\*, Xiaoya Li\*, Ping Nie, Fan Yin  
Muyu Li, Qinghong Han, Xiaofei Sun and Jiwei Li

Shannon.AI

{yuxian\_meng, wei\_wu, fei\_wang, xiaoya\_li, ping\_nie, fan\_yin, myu\_li, qinghong\_han, xiaofei\_sun, jiwei\_li}@shannon.ai.com

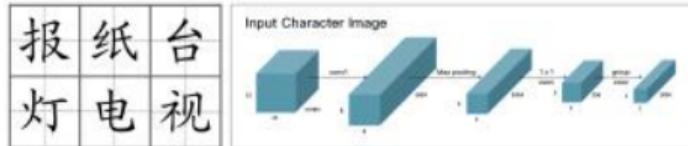


Figure 1: Illustration of the Tianzege-CNN used in Glyce.

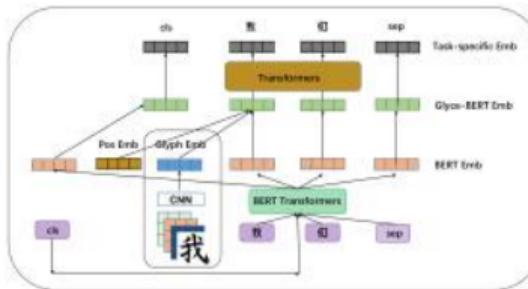


Figure 2: Combing glyph information with BERT.

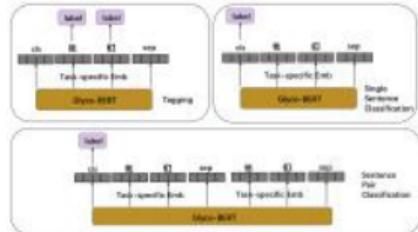


Figure 3: Using Glyce-BERT model for different tasks.

Model	ChnSentiCorp	the Fudan corpus	iFeng
LSTM	91.7	95.8	84.9
LSTM + Glyce	93.1 (+ 1.4)	96.3 (+ 0.5)	85.8 (+ 0.9)
BERT	95.4	99.5	87.1
Glyce+BERT	95.9 (+ 0.5)	99.8 (+ 0.3)	87.5 (+ 0.4)

Table 6: Accuracies for Single Sentence Classification task.

Dependency Parsing		
Model	UAS	LAS
Ballesteros et al. [2016]	87.7	86.2
Kiperwasser and Eliyahu [2016]	87.6	86.1
Cheng et al. [2016]	88.1	85.7
Biaffine	89.3	88.2
Biaffine+Glyce	90.2 (+0.9)	89.0 (+0.8)

Model	P	R	F
Roth and Lapata [2016]	76.9	73.8	75.3
Marcheggiani and Diego [2017]	84.6	80.4	82.5
He et al. [2018]	84.2	81.5	82.8
k-order pruning+Glyce	85.4 (+0.8)	82.1 (+0.6)	83.7 (+0.9)

Table 7: Results for dependency parsing and SRL.

# 图像 (image) 作为大语言模型和符号化知识的接口

## Autoformalizing Euclidean Geometry

Logan Murphy<sup>1\*</sup> Kaiyu Yang<sup>2\*</sup> Jialiang Sun<sup>1</sup> Zhaoyu Li<sup>1</sup> Anima Anandkumar<sup>2</sup> Xujie Si<sup>1</sup>

Proposition 1  
To construct an equilateral triangle on a given finite straight-line.



Let  $AB$  be the given finite straight-line.  
So it is required to construct an equilateral triangle on the straight line  $AB$ .

Let the circle  $BCD$  with center  $A$  and radius  $AB$  have been drawn [Post. 3], and again let the circle  $ACE$  with center  $B$  and radius  $BA$  have been drawn [Post. 3]. And let the straight-lines  $CA$  and  $CB$  have been joined from the point  $C$ , where the circles cut one another,<sup>1</sup> to the points  $A$  and  $B$  (respectively) [Post. 1].

And since the point  $A$  is the center of the circle  $BCD$ ,  $AC$  is equal to  $AB$  [Def. 3.15]. Again, since the point  $B$  is the center of the circle  $ACE$ ,  $BC$  is equal to  $BA$  [Def. 3.15]. But  $CA$  was also shown to be equal to  $AB$ . Then,  $CA$  and  $CB$  are each equal to  $AB$ . But things equal to the same thing are also equal to one another [CN. 1]. Thus,  $CA$  is also equal to  $CB$ . Thus, the three straight-lines  $CA$ ,  $AB$ , and  $BC$  are equal to one another.

Thus, the triangle  $ABC$  is equilateral, and has been constructed on the given finite straight-line  $AB$  (which is) the very thing it was required to do.

Informal Euclidean geometry problem

theorem proposition\_1 :  $\forall (a b : \text{Point}) (AB : \text{Line})$ ,

$\exists c : \text{Point}, |(c-a)| = |(a-b)| \wedge |(c-b)| = |(a-b)|$

Ground truth theorem

theorem proposition\_1' :  $\forall (a b : \text{Point}) (AB : \text{Line})$ ,

$a.\text{onLine } AB \wedge b.\text{onLine } AB \wedge a \neq b \rightarrow$

$\exists c : \text{Point}, |(a-c)| = |(c-b)| \wedge |(a-c)| = |(a-b)|$

Autoformalized theorem

by  
euclid\_intros  
euclid\_apply circle\_from\_points a b as BCD  
euclid\_apply circle\_from\_points b a as ACE  
euclid\_apply intersection\_circles BCD ACE as c  
euclid\_apply point\_on\_circle\_onlyif a b c BCD  
euclid\_apply point\_on\_circle\_onlyif b a c ACE  
use c  
euclid\_finish

Autoformalized proof

Z3  
CVC5

Equivalent?  
✓  
✗

SMT-based symbolic reasoning engine

a b : Point  
AB : Line  
BCD ACE : Circle  
isCenter a BCD  
onCircle b BCD  
isCenter b ACE  
onCircle a ACE  
|- intersects BCD ACE

LEAN

⇒

...

Z3  
CVC5

...

...

Diagrammatic reasoning gaps

Category	GPT-4		GPT-4V	
	1-shot	5-shot	1-shot	5-shot
Triangle	35%	45%	45%	70%
Similarity	5%	15%	10%	15%
Congruent	5%	25%	15%	25%
Quadrilateral	35%	25%	20%	30%
Parallel	5%	15%	5%	15%
Overall	17%	25%	19%	31%

# LayoutGPT

**[2D Numerical Reasoning]** There are three elephants standing beside a pool of water.

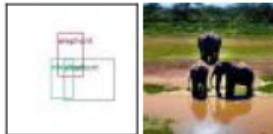
StableDiffusion  
(v2.1)



Attend-and-  
Excite



LayoutGPT  
+ GLIGEN



**[3D Living Room]** Room Type: Living Room;  
Room Size: 7.7m x 3.6m

ATISS



LayoutGPT



**[2D Spatial Reasoning]** A carrot and some onion next to a knife on a cutting board.

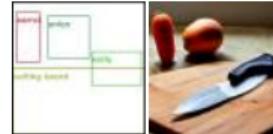
StableDiffusion  
(v2.1)



Attend-and-  
Excite



LayoutGPT  
+ GLIGEN



**[3D Bedroom]** Room Type: Bedroom;  
Room Size: 3.0m x 4.8m

ATISS



LayoutGPT



Figure 1: Generated layouts from LayoutGPT in 2D images and 3D indoor scenes. LayoutGPT can serve as a visual planner to reflect challenging numerical and spatial concepts in visual spaces.

LayoutGPT: Compositional Visual Planning and Generation with LLMs, arXiv:2305.15393v2

# 图 (Graph) 作为大语言模型和符号化知识的接口

## ERNIE: Enhanced Language Representation with Informative Entities

Zhengyan Zhang<sup>1,2,3\*</sup>, Xu Han<sup>1,2,3\*</sup>, Zhiyuan Liu<sup>1,2,3†</sup>, Xin Jiang<sup>4</sup>, Maosong Sun<sup>1,2,3</sup>, Qun Liu<sup>4</sup>

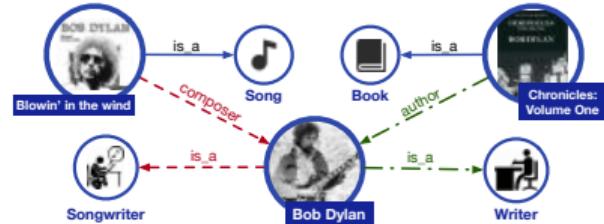
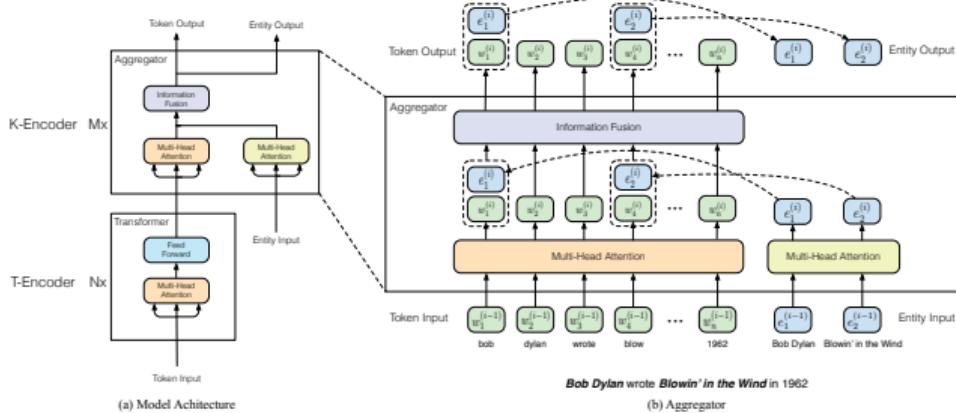
<sup>1</sup>Department of Computer Science and Technology, Tsinghua University, Beijing, China

<sup>2</sup>Institute for Artificial Intelligence, Tsinghua University, Beijing, China

<sup>3</sup>State Key Lab on Intelligent Technology and Systems, Tsinghua University, Beijing, China

<sup>4</sup>Huawei Noah's Ark Lab

{zhangzhengyan14, hanxu17}@mails.tsinghua.edu.cn

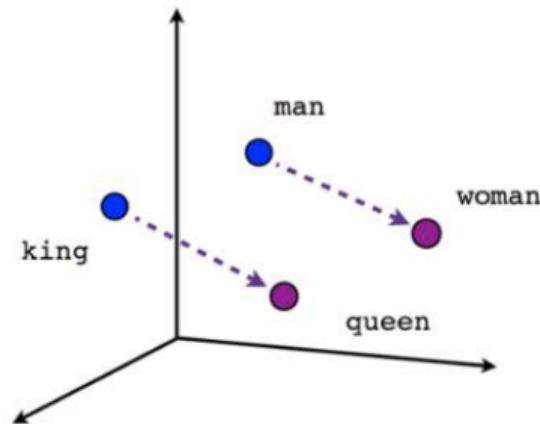


Bob Dylan wrote *Blowin' in the Wind* in 1962, and wrote *Chronicles: Volume One* in 2004.

Model	MNLI-(m/mm) 392k	QQP 363k	QNLI 104k	SST-2 67k
BERT <sub>BASE</sub>	84.6/83.4	71.2	-	93.5
ERNIE	84.0/83.2	71.2	91.3	93.5
Model	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k
BERT <sub>BASE</sub>	52.1	85.8	88.9	66.4
ERNIE	52.3	83.2	88.2	68.8

## 图 (Graph) 作为大语言模型和符号化知识的接口

- ▶ 早期，将预训练语言模型和知识图谱嵌入表示（如TransE）相结合的做法，曾经引起了较多的关注。
- ▶ 在大模型时代，这样的研究已经很少见。
- ▶ 但我认为，这样的研究也许仍然是有价值的，比如对于Reverse Curse问题，也许是个可行的解决方案。



# Content

神经和符号的结合是实现真正的人类水平智能的必经之路

符号化知识表示的类型及其与大语言模型结合的方法

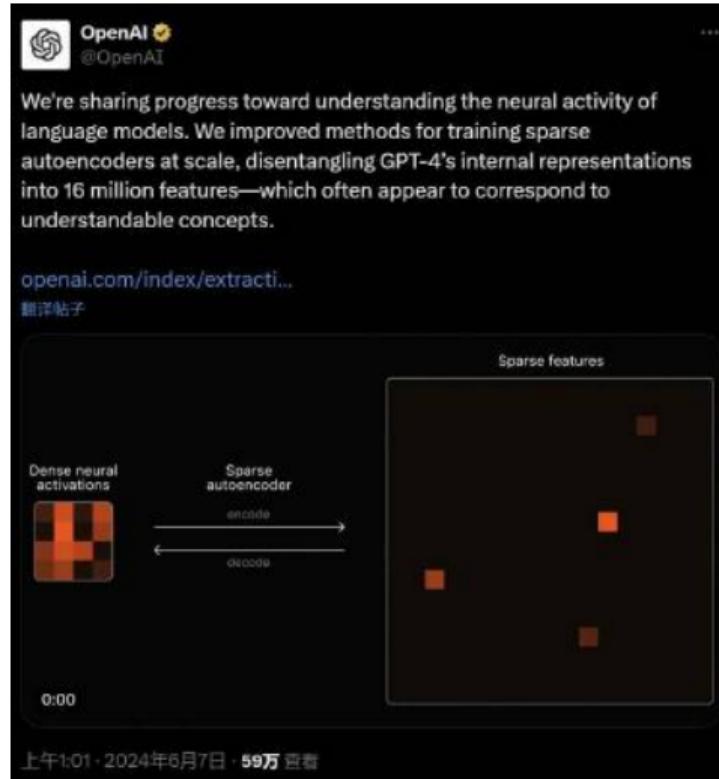
在Transformer架构中引入符号计算模块的设想

总结

# 在Transformer架构中引入符号计算模块

- ▶ 目前，大语言模型和符号化知识表示的唯一接口就是token序列：
  - ▶ 输入：Prompting
  - ▶ 输出：Generation
- ▶ 其他所有带结构的知识表示，都需要转化成线性的token序列（语言或图像）才能跟大语言模型交互
- ▶ 线性化以后的带结构的知识表示，虽然理论上包含了所有的结构信息，但实际上大语言模型很难准确捕获到完整的结构信息
- ▶ 是否有可能修改Transformer结构，直接在其中加入符号计算模块？
  - ▶ 类似人脑中有海马体，从仿生角度看，在语言模型内部引入符号计算模块有一定的合理性
  - ▶ 这种符号处理模块应该能够直接处理实体、关系等具有明确语义的符号，而不仅是tokens
  - ▶ 关于Transformer的可解释性研究的一些工作为这类方案提供了可能性

# Scaling and evaluating sparse autoencoders (OpenAI)



# Content

神经和符号的结合是实现真正的人类水平智能的必经之路

符号化知识表示的类型及其与大语言模型结合的方法

在Transformer架构中引入符号计算模块的设想

总结

# 总结

- ▶ 神经和符号的结合是实现真正的人类水平智能的必经之路。
  - ▶ 神经与符号的GAP是目前大模型很多问题的根源。
- ▶ 不同的符号表示形式发展出来不同的神经符号结合方法：
  - ▶ 自然语言是最自然的符号表示形式，可以直接口神经网络语言模型，推理需要借助XoT等方法。
  - ▶ 形式语言包括逻辑语言与程序语言。形式语言是精确的符号表示形式，通过引入外部执行引擎或者正确性检查机制，再辅以基于强化学习的搜索算法，可以实现强大的推理能力。
  - ▶ 图表形式的符号化知识，形式多样，跟语言模型结合的方式也有很多种，如GraphRAG、图（graph）、图像（image）等形式。
- ▶ 直接在Transformer架构中引入符号计算模块也是值得探索的路径。

# 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.

