AI大模型时代
--新的机遇与挑战

李航
ByteDance Research

目录

语言大模型LLM强大之所在

- 以ChatGPT和GPT4为代表
- 主要手段
  - 模型：Transformer强大的表示能力，表示语言的组合性
  - 预训练：语言模型，数据压缩 = 单词序列概率最大化
  - 微调：学习输入到输出的映射及过程，$X \rightarrow Y$， $X, C_1 \cdots, C_n \rightarrow Y$，学习模型的行为
  - RLHF：基于人的反馈，调整模型整体的行为
- 巨大进步
  - 智能性：具备语言、知识、简单推理能力，近似人的智能
  - 通用性：可以适用于不同领域，完成不同任务

LLM强大之所在

- 基本现象
  - 传统的语言模型能生成自然的语言，但在现实中出现的概率不一定高
  - LLM能生成现实中大概率出现的内容，甚至是合理的内容
- 主要突破
  - **大模型大数据带来质变**
  - **模型的行为是人教出来的**，Open AI开发了一整套技术，包括方法、技巧、工程实现

目录

实现AI的三条路径

输入经验知识

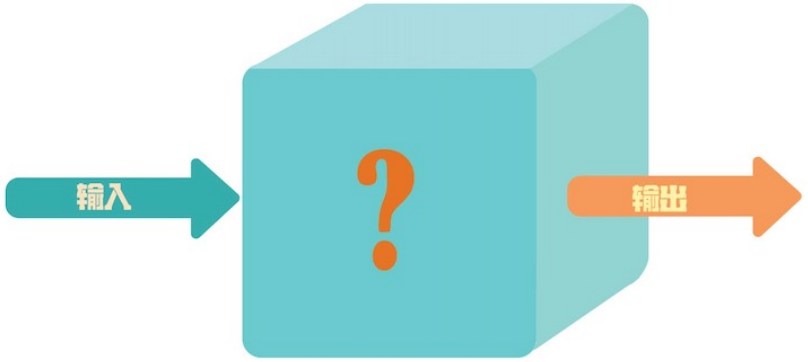
将知识通过规则
等教给计算机，
进行符号处理

实现人脑机制

解明人脑机制，
基于相同原理
实现人类智能

从数据中学习

通过数据驱动、
机器学习方法
模仿人类智能



实现AI三条路径

1. 输入经验知识：历史证明非常困难
2. 实现人脑机制：脑科学研究进展缓慢
3. 从数据中学习：目前的主要手段

- 符号处理 = 输入经验知识
- 机器学习 = 通过数据学习
- 深度学习 = 通过数据学习 + 受人脑启发
- **LLM = 通过数据学习 + 受人脑启发+ 输入经验知识**

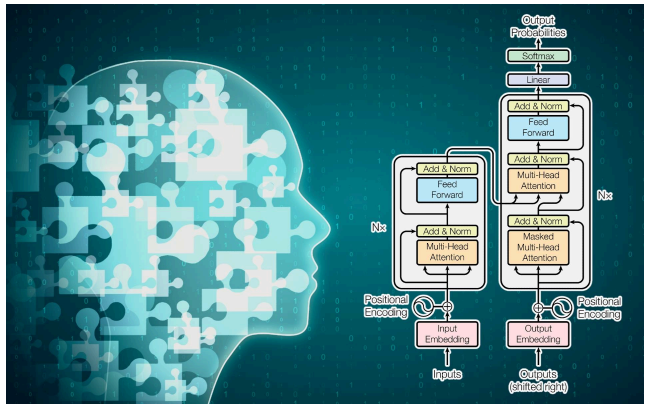
第一者体验 vs 第三者体验

- 第一者体验（first person experience），内心的感受和思考
- 第三者体验（third person experience），对外部世界的观察
- 科学的前提是第三者体验

- 输入经验知识：开发者基于第一者体验
- 实现人脑机制：开发者基于第三者体验
- 从数据中学习：开发者基于第三者体验

- **LLM = 基于第三者体验 + 基于第一者体验**

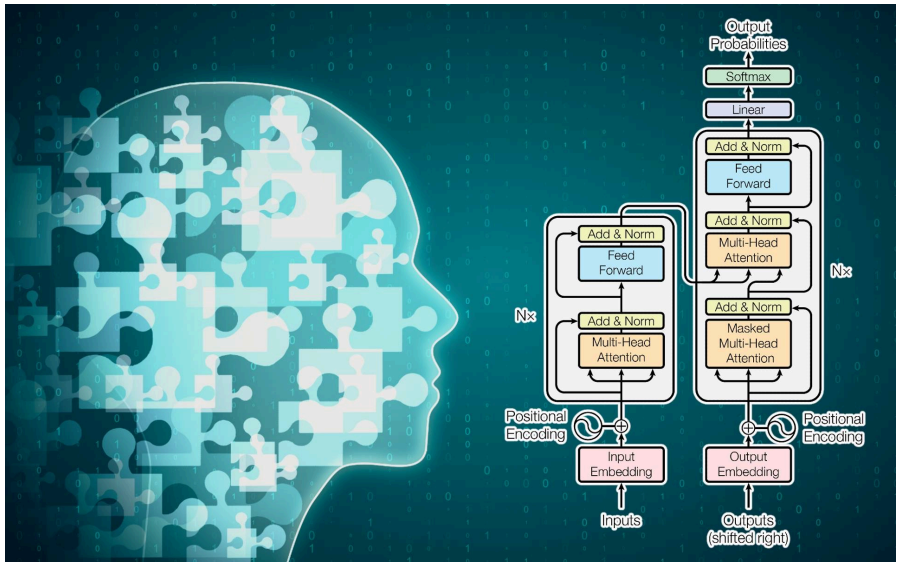
LLM的优势



- 拥有已有深度模型的优点
- 一定程度上解决了通用性问题，大幅提高了智能性
- 模型返回的结果大概率是现实可能发生的，当然仍有幻觉现象hallucination
- 开发者通过预训练、微调、RLHF、Prompt等方式，调教模型，大大提高学习效果和效率

目前LLM的局限

- 如何优化大模型，
- 如何保证模型生成内容的真实性，也就是避免幻觉。
- 如何构建可信赖大模型，也就是保证模型生成结果的有用性，安全性等
- 如何建立大模型的机器学习理论



目录

重要研究课题

- LLM的优化
- LLM的真实性
- 可信赖LLM与AI伦理
- LLM的理论
- LLM与多模态
- LLM+逻辑推理
- 智能体（agent）
- LLM与信息访问

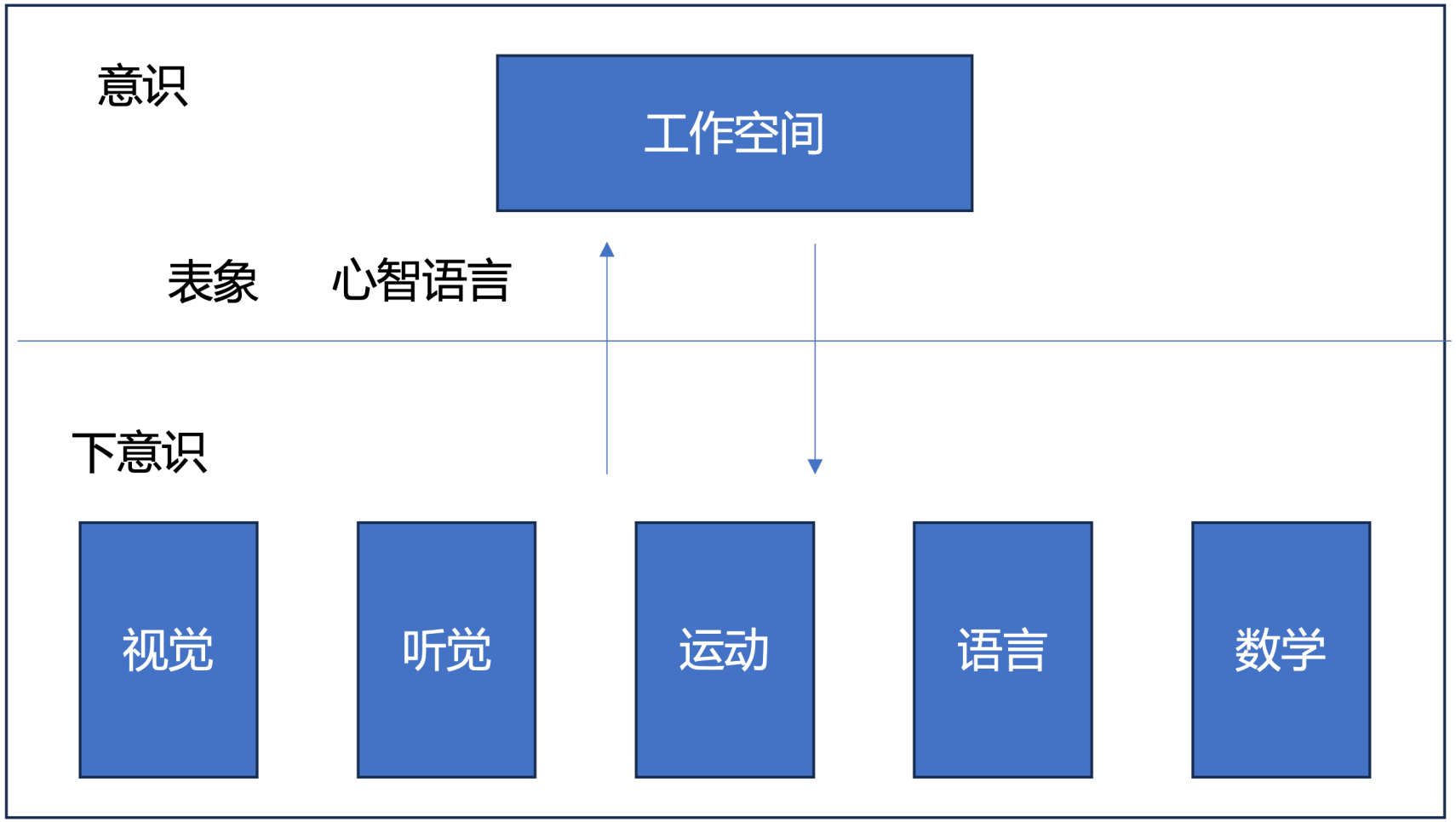
目录

- LLM强大之所在
- LLM的特点
  - AI三条路径
  - 第一者体验和第三者体验
  - LLM的优势与局限
- 重要研究课题
  - *从人类智能角度看LLM*
  - LLM与多模态
  - LLM与数学能力
  - LLM与信息访问
- 总结

人脑、心智、意识

- 人脑
  - 复杂的神经网络
  - 进行神经计算，产生神经表征（neural representation）
- 心智
  - 人自身的感知和认知
  - 心智 = 意识 + 下意识
- 下意识
  - 对应着人脑中的大部分神经计算
  - 并行处理，快思考
- 意识：
  - 产生表象（image）
  - 信息同步机制，自己脑中的那个“小人”是错觉
  - 串行处理，慢思考

人脑和心智的组成



人的语言理解

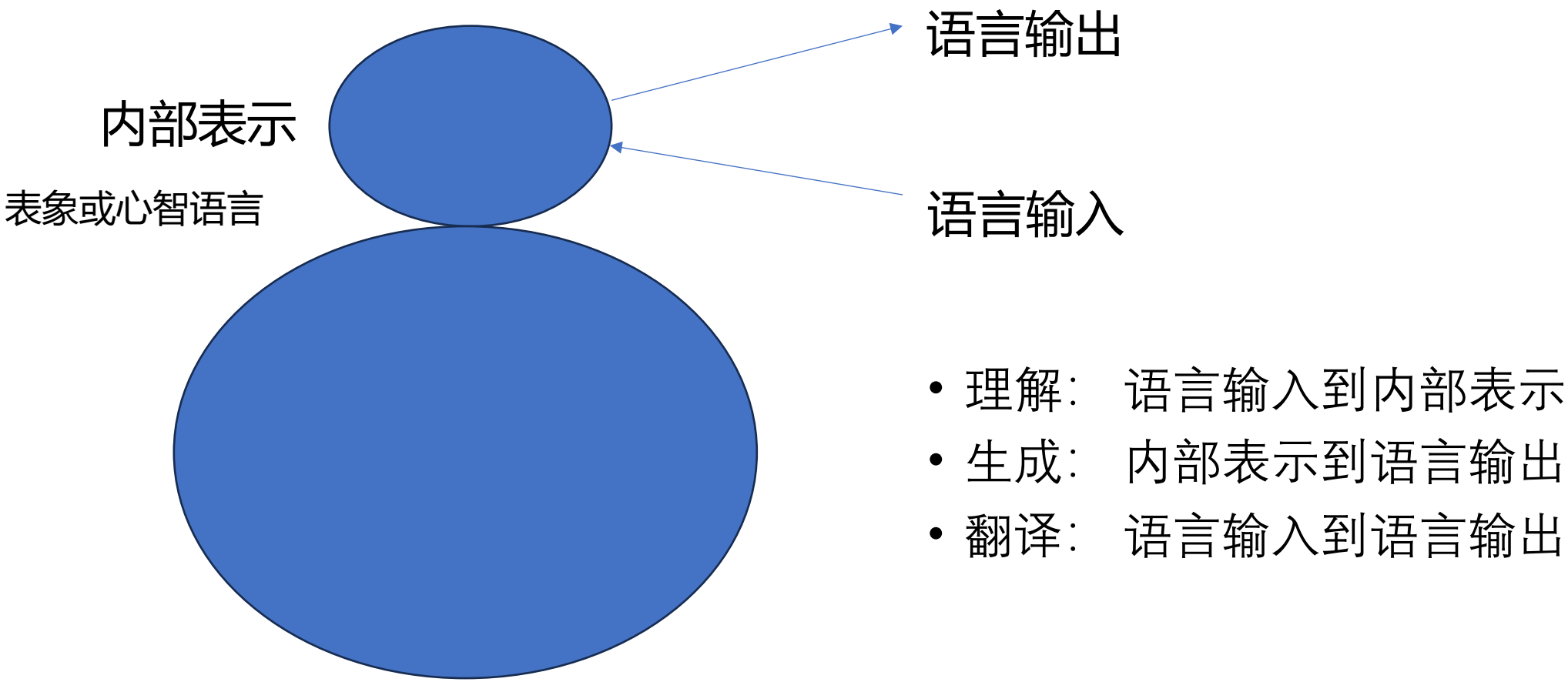
- 理解一个词语或者是一句话，意味着把记忆中的相关概念和事件唤起，并把它们联系起来，
- 在意识中产生表象或心智语言的表示
- 理解的结果产生语义落实（grounding），是没有歧义的。因为人脑在理解中做了消歧

The old man the boat.

Time flies like an arrow, fruit flies like a banana.

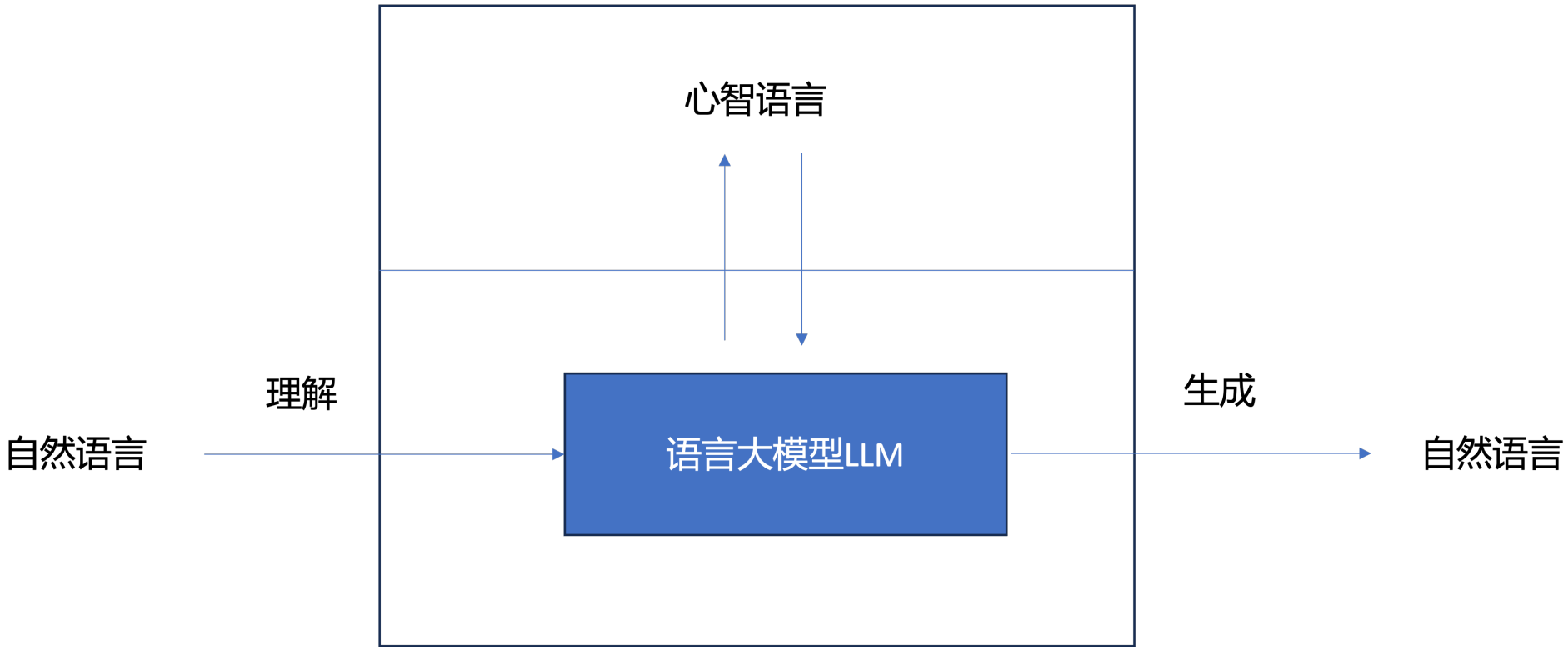
Garden path sentences

所有自然语言处理任务= Seq2Seq



- 理解：　语言输入到内部表示
- 生成：　内部表示到语言输出
- 翻译：　语言输入到语言输出

LLM生成的内容可以是心智语言的近似



- 基于LLM的语言理解，就是把自然语言转化为心智语言
- 心智语言应该是没有歧义的，而用LLM生成的语言经常是有歧义的
- 可以让LLM生成的内容没有歧义，如代码

目录

多模态大模型

- 问：LLM是否建立了世界模型？
- 答：是也不是。
- 当LLM和多模态大模型结合时，就能产生与人更接近的世界模型
- 知识通过实体和概念等联系起来
- 机器人技术的发展会产生具身智能

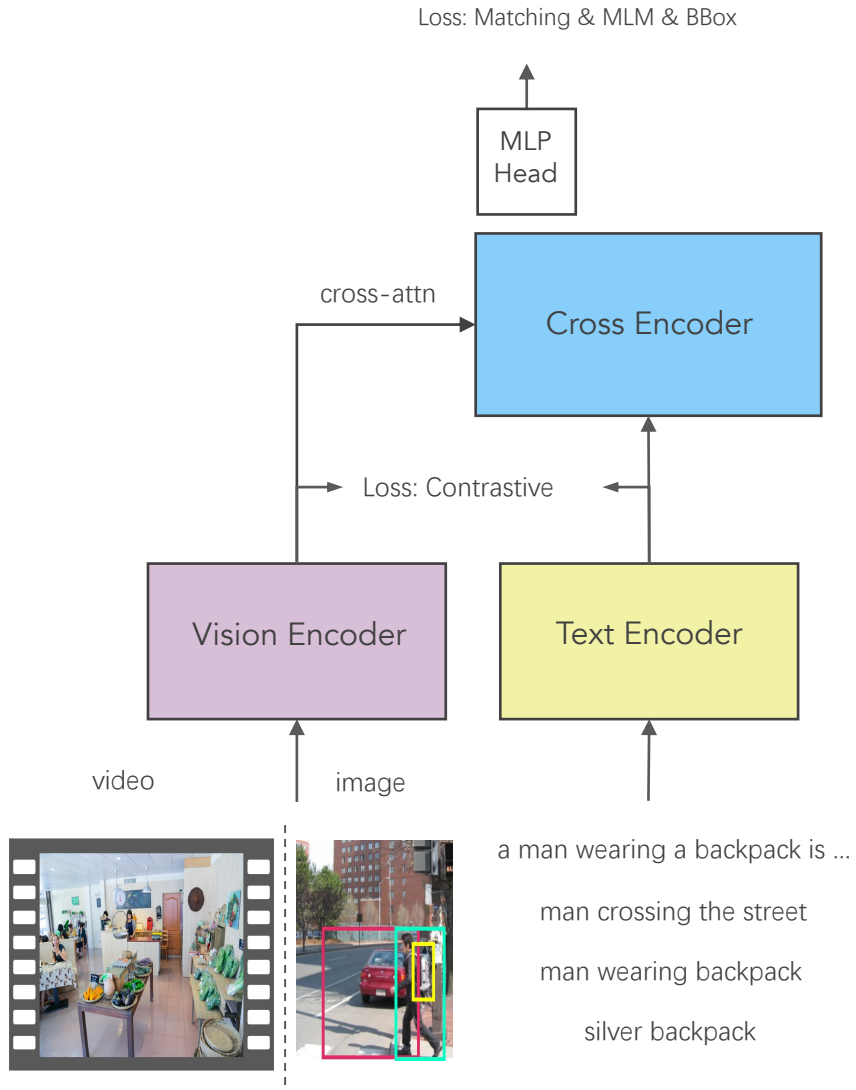
人的世界理解

- 语言和多模态密切关联
- 体验模拟假说
  - 语言理解是基于自己过去的视觉、听觉、运动等体验的模拟
  - 心理学实验
    - 木匠把钉子钉进墙里
    - 木匠把钉子钉进地板

- 儿童的实体概念是通过多模态学习获取的
- 最基本的视觉、听觉等能力是先天具有的，出生后开始发育

X-VLM和X^2-VLM

- 语言和视觉模型
- 多颗粒度语言和视觉对齐，图片、区域、物体
- 三个编码器：text encoder，vision encoder，cross encoder
- 四种损失函数：matching loss，contrastive loss，mask language modeling loss，*bounding box prediction loss*
- 在语言-视觉理解任务上是SOTA方法

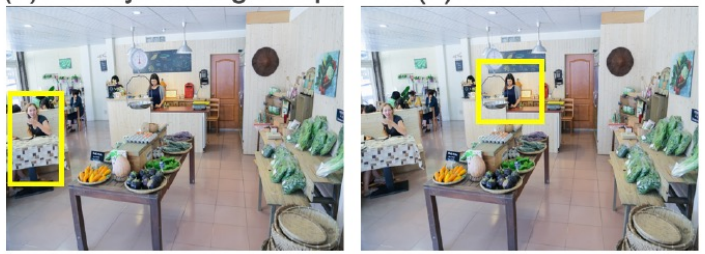


Zeng, Y., Zhang, X. and Li, H., June. Multi-Grained Vision Language Pre-Training: Aligning Texts with Visual Concepts. ICML 2022

生成图像标题和定位视觉概念

(1) “a lady holding a cup"

(2) “a cashier"



output: “a store filled with lots of produce and people"

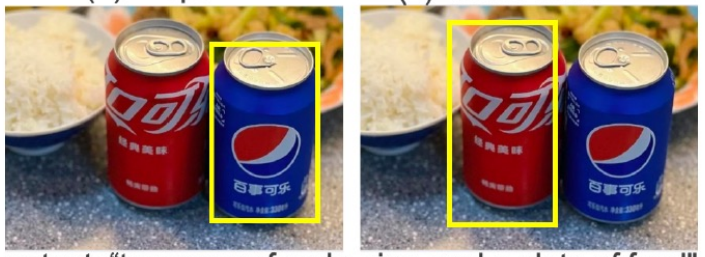
(1) “flying ads"

(2) “woman on ads"

output: “a man standing next to a car in a city"

(1) “Pepsi"

(2) “Coca Cola"



output: “two cans of soda, rice, and a plate of food"

(1) “ocean in daytime"

(2) “ocean in nighttime"

output: “an image of a cycle of water cycle"

(1) “Audi"

(2) “BMW"

output: “three different cars are parked next to each other"

(1) “Zoro with swords"

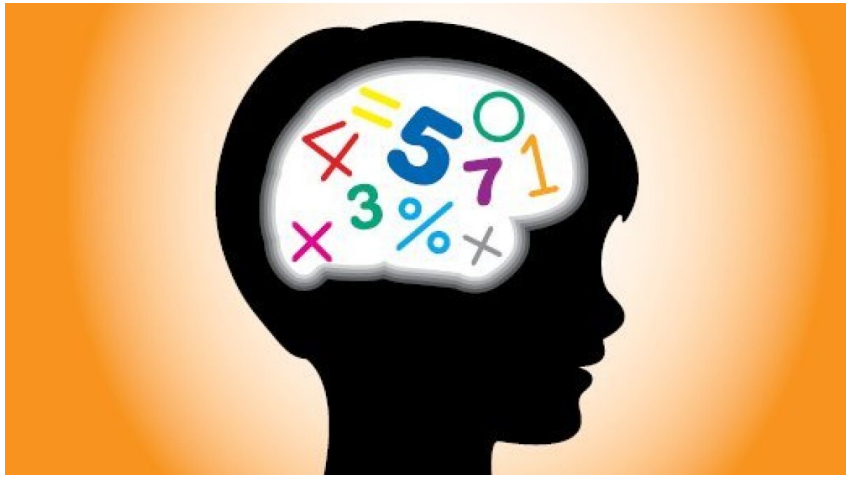
(2) “Luffy"

output: “a group of young people standing next to each other"

目录

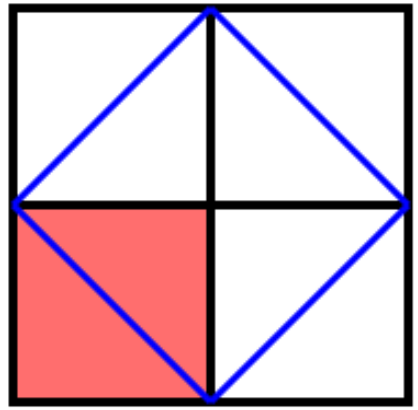
人的语言理解：与数学关联

- 人天生有识别数量大小的能力
- 四个月的儿童知道1+1=2
- 数字能力的核心是递归，猜测是人先天具备的能力
- 科学家猜测数学思维在顶叶的一个脑区进行
- 数学家通常把自己的数学思考描述为表象的操作，但也涉及逻辑



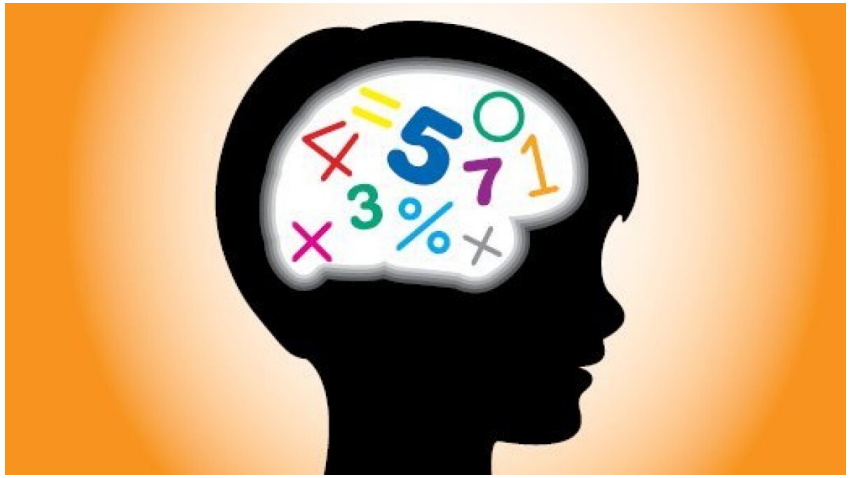
人的数学能力：哲学

- 亚里士多德：认为哲学的理论学分为数学、自然学（physics）、形而上学（metaphysics）
- 柏拉图《美诺篇》：苏格拉底通过与奴隶少年的对话，引导他想出了几何题目的解法：如何将2×2的正方形的面积扩大一倍
- 康德：提出“先验综合判断”，认为数学能力是先天的，如5+7=12

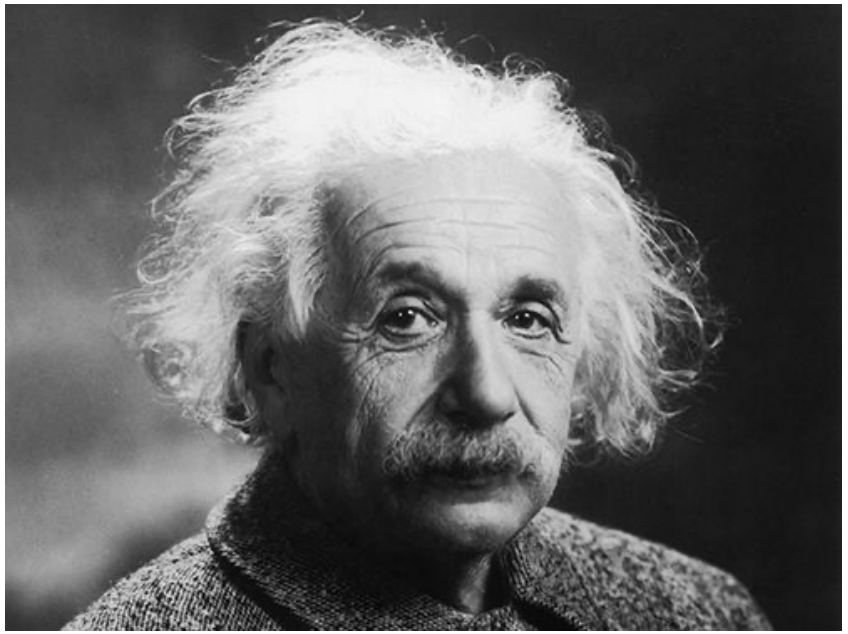


人的数学能力：脑科学

- 人天生有识别数量大小的能力，理解数字1，2，3
- 四个月的儿童知道1+1=2
- 关键是理解数字能力的核心是递归的，猜想可能是先天具备的能力
- 科学家猜测在数学思维在顶叶的一个脑区进行
- 数学家通常把自己的数学思考描述为表象的操作，但也涉及逻辑



爱因斯坦谈自己的数学思维



词汇或者语言，无论是书面形式还是口头形式，似乎在我的思维中并没有发挥任何作用。作为思维元素的实体是某些符号和或多或少清晰的表象，可以自发地复制和组合。而且，这些元素和相关的逻辑概念之间存在一定的联系。

LLM+逻辑推理

- 应用：数学解题
- 用LLM理解数学问题的题意，将其转换为心智语言，在心智语的基础上进行逻辑推理和数学计算
- 逻辑推理和数学计算调用其他的数学计算机制
- 人的数学解题有两种机制
  - 系统1进行快的思维（基于死记硬背）
  - 和系统2进行慢的思维（进行深入思考）
- 用LLM解题
  - 直接解题，对应着系统1
  - 用LLM产生程序，在程序基础上进行解题，对应着系统2

程序语言作为心智语言

- 用程序语言表示心智语言，因为LLM也能生成程序
- Python程序比英语（自然语言）作为“心智语言”，在数学解题中更有优势的事实
- 优点是，LLM理解题意后，得到的程序可以直接通过解释器执行，验证解题步骤的正确性
- 在Python程序上进行推理，也比在自然语言上进行推理更为容易
- 自我描述程序（Self Describing Program）最适合做数学解题的中间表示

目录

- LLM强大之所在
- LLM的特点
  - AI三条路径
  - 第一者体验和第三者体验
  - LLM的优势与局限
- 重要研究课题
  - 从人类智能角度看LLM
  - LLM与多模态
  - LLM与数学能力
  - *LLM与信息访问*
- 总结

LLM与信息访问

- LLM应用，已经验证的
  - 翻译
  - 编程辅助
  - 创作、写作（纠错、润色）
  - 对话： 情感类、游戏娱乐类、助理类（客服）
- LLM与搜索、推荐
  - 是否能有效结合，有待验证
  - 幻觉问题如何解决?
  - 可能是Chat和Search并存
  - 对话式推荐，有待探索
  - 助理类的挑战： 定制开发

目录

主要观点总结

- ChatGPT的突破主要在于规模带来的质变和模型调教方式的发明。
- LLM融合了实现人工智能的三条路径。
- LLM的开发需要结合第三者体验和第一者体验。
- LLM能近似生成心智语言。
- LLM需要与多模态大模型结合，以产生对世界的认识。
- LLM本身不具备逻辑推理能力，需要在其基础上增加推理能力。
- LLM与信息访问的结合仍有许多需要探索的问题。

Thanks!

相关文章：对语言大模型的若干观察和思考

机器之心，2023/10/15