



中国科学技术大学
University of Science and Technology of China

《人工智能数学原理与算法》

第6章 自监督学习

6.4 大语言模型

凌震华

zhling@ustc.edu.cn

目录

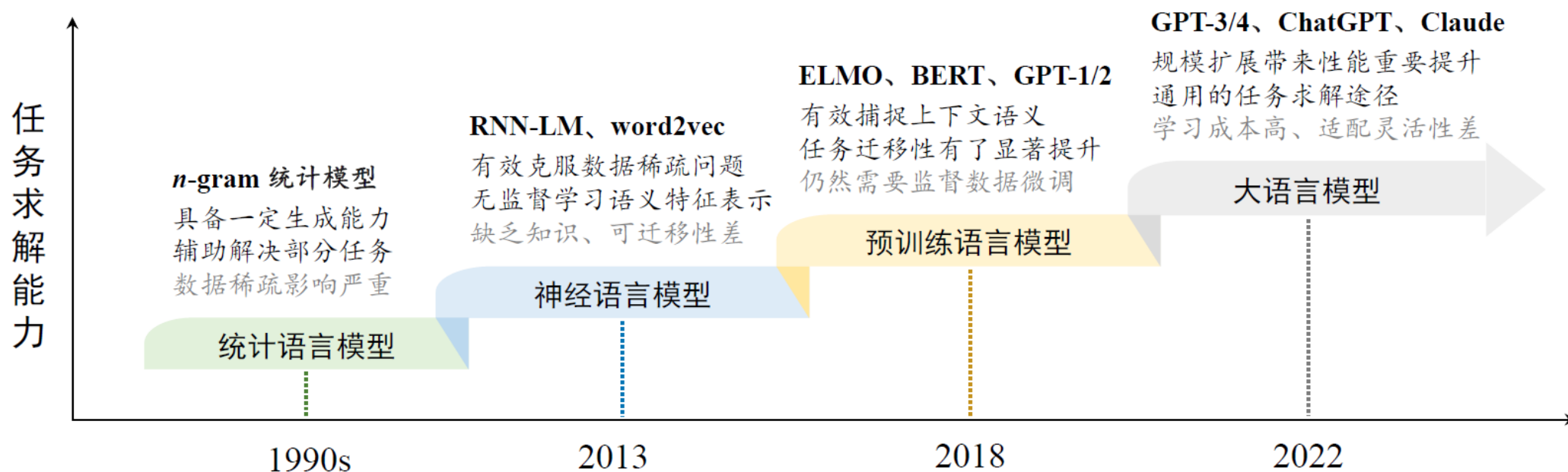
- 01 大语言模型概述
- 02 大语言模型的能力特点
- 03 大语言模型的构建方法
- 04 大语言模型的应用

语言模型的发展演进

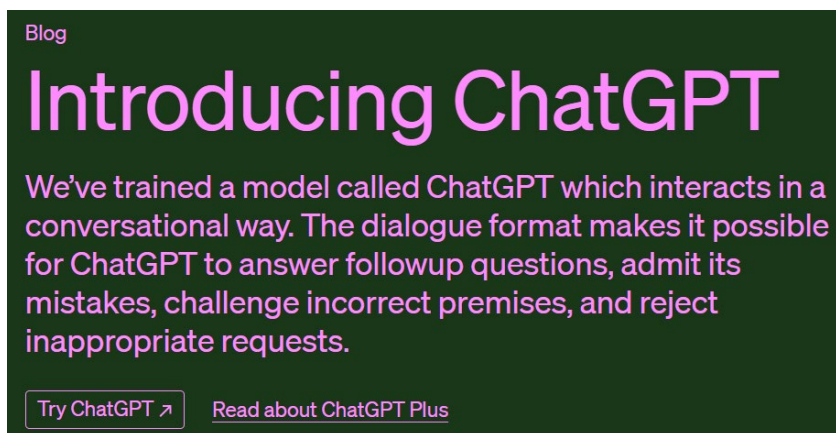
• 回顾：语言模型

- 对于人类语言的内在规律进行建模，从而准确预测词序列的概率
- 具体来说，就是准确预测词序列中未来（或缺失）词的概率

• 发展历程



- 2022年11月，OpenAI发布ChatGPT模型，是语言模型发展进入大语言模型(large language model, LLM)阶段的标志性事件



November 30, 2022

Authors
[OpenAI](#) ↓

[Product, Announcements](#)

ChatGPT is a sibling model to [InstructGPT](#), which is trained to follow an instruction in a prompt and provide a detailed response.

We are excited to introduce ChatGPT to get users' feedback and learn about its strengths and weaknesses. During the research preview, usage of ChatGPT is free. Try it now at chat.openai.com.

- 将GPT系列模型适配到对话任务中，展现出惊人能力
- 学术界高度关注，之后大语言模型相关论文迅速增长

大语言模型的主要优势

- 具有较为丰富的世界知识

- 与传统机器学习模型相比，大语言模型经过超大规模文本数据的预训练后能够学习到较为丰富的世界知识

- 具有较强的通用任务解决能力

- 以“预测下一个词元”为预训练任务，无需针对特定下游任务的优化
- 以上过程可以看做一个多任务学习过程，不同下游任务→自回归生成

情感分类

“我喜欢这部电影，这部电影真好看”

数值计算

“1+2=3”

知识推理

“中国的首都是北京”

... ..

大语言模型的主要优势

- **具有较好的复杂任务推理能力**

- 能够回答知识关系复杂的推理问题、解决涉及复杂数学推理过程的数学题目等
- 传统方法性能不理想，需要针对性的模型架构设计与使用特定数据的学习

- **人具有较强的类指令遵循能力**

- 自然语言形式的统一任务解决模式：任务输入与执行结果均通过自然语言表达
- 使用者可以通过自然语言描述下达任务指令，自然、通用的人机交互方式

大语言模型的主要优势

- **具有较好的人类对齐能力**

- 当前广泛采用的对齐方式是基于人类反馈强化学习(RLHF)技术
- 使得模型进行正确行为的加强以及错误行为的规避，建立较好的人类对齐能力

- **具有可拓展的工具使用能力**

- 可以通过微调、上下文学习等方式掌握外部工具的使用，如搜索引擎与计算器
- 借鉴了人类使用工具的思路，通过具有特殊功能的工具来加强大语言模型能力

当前大语言模型的局限性

- “幻觉问题” 与生成结果的可信性
- 可解释性与理解决策过程的透明性
- 数据带来的偏见与公平性
- 泛化与鲁棒性
- 资源依赖性
- 负面应用与合规性
-

幻觉 (hallucination)

- 偏离用户输入
- 生成的上下文之间偏离
- 错误的事实性知识

User Input



Can you recommend a delicious recipe for dinner?

LLM Response





Yes, here is a delicious recipe for **lunch**. So how about fried chicken with mashed potatoes? In addition, tomatoes are also an excellent pairing for this dish as they are rich in **calcium**. Enjoy this **steak**!

Hallucination Explanation

Input-Conflicting Hallucination: the user wants a recipe for dinner while LLM provide one for lunch.

Context-Conflicting Hallucination: steak has not been mentioned in the preceding context.

Fact-Conflicting Hallucination: tomatoes are not rich in calcium in fact.

- 
- 
- 01 大语言模型概述
 - 02 大语言模型的能力特点
 - 03 大语言模型的构建方法
 - 04 大语言模型的应用

目录

1. 扩展法则

- 大语言模型对比小型语言模型

- 相似的模型结构 (例如Transformer) 与自监督学习方法 (例如自回归生成)
- 成功的关键在于对 **“规模扩展” (Scaling)** 的充分探索与利用
 - GPT-1 → (2018年, 117M) → GPT-2 (2019年, 1.5B) → GPT-3 (2020年, 175B)
- 扩展所带来的性能提升通常显著高于通过改进架构、算法等带来的改进

- **扩展法则 (Scaling Law)**

- 建立定量的建模方法, 研究规模扩展所带来的模型性能提升
- 可以用于指导大语言模型的训练, 例如使用小模型的性能去预估大模型的性能
- 基于语言建模损失 (下一个词元的平均交叉熵损失) 开展; 然而, 语言建模损失的减少并不总是意味着模型在下游任务上的性能改善

1. 扩展法则——KM扩展法则

- 2020年，OpenAI 团队Kaplan 等人首次建立
 - 描述模型交叉熵损失 L 与模型规模 N 、数据规模 D 和计算算力 C 之间关系
 - 通过模型在不同数据规模（22M ~23B 词元）、模型规模（768M~1.5B）和算力规模下的性能表现拟合推导得到

$$L(N) = \left(\frac{N_c}{N} \right)^{\alpha_N}, \quad \alpha_N \sim 0.076, N_c \sim 8.8 \times 10^{13}$$

$$L(D) = \left(\frac{D_c}{D} \right)^{\alpha_D}, \quad \alpha_D \sim 0.095, D_c \sim 5.4 \times 10^{13}$$

$$L(C) = \left(\frac{C_c}{C} \right)^{\alpha_C}, \quad \alpha_C \sim 0.050, C_c \sim 3.1 \times 10^8$$

- **模型性能与三个因素可以近似刻画为指数关系**

1. 扩展法则——Chinchilla扩展法则

- 2022年，DeepMind 团队Hoffmann等人提出
 - 指导大语言模型充分利用给定的算力资源进行优化训练
 - 更大范围模型规模（70M~16B）和数据规模（5B~500B 词元）进行实验拟合
 - 模型性能与模型规模、数据规模之间的**另一种指数关系**

$$L(N, D) = E + \frac{A}{N^\alpha} + \frac{B}{D^\beta} \quad \text{其中, } E, A, B, \alpha, \beta \text{ 为常数}$$

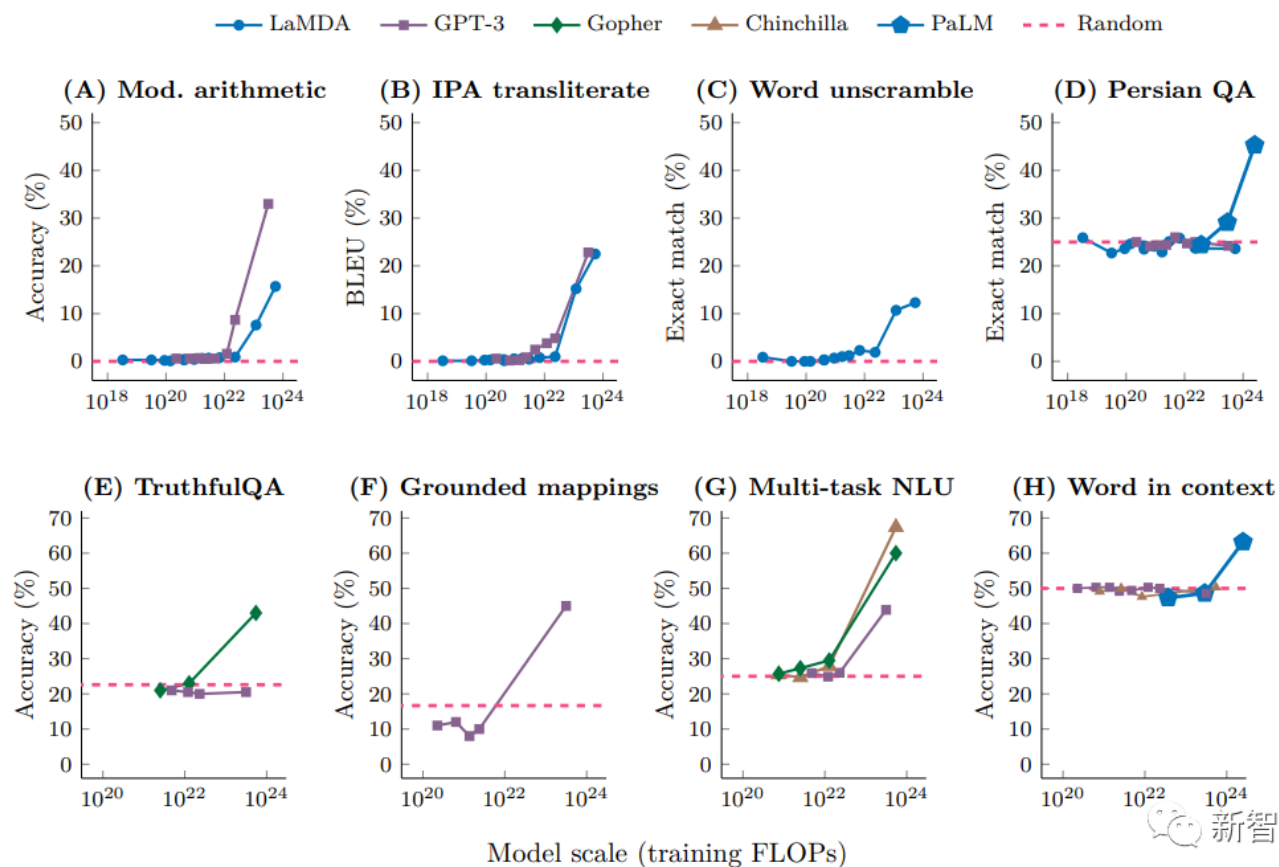
- 进一步获得获得算力资源固定情况下**模型规模与数据规模的最优分配方案**

$$N_{\text{opt}}(C) = G \left(\frac{C}{6} \right)^a, \quad D_{\text{opt}}(C) = G^{-1} \left(\frac{C}{6} \right)^b \quad \text{其中, } G, a, b \text{ 为常数}$$

2. 涌现能力

• 涌现能力(Emergent Abilities)

- 非形式化定义为“在小型模型中不存在但在大模型中出现的能力”
- 当模型扩展到一定规模时，模型的特定任务性能突然出现显著跃升的趋势，远超过随机水平
- 大语言模型相对传统语言模型的能力优势
- 存在学术争论，缺乏相应的理论证实与解释

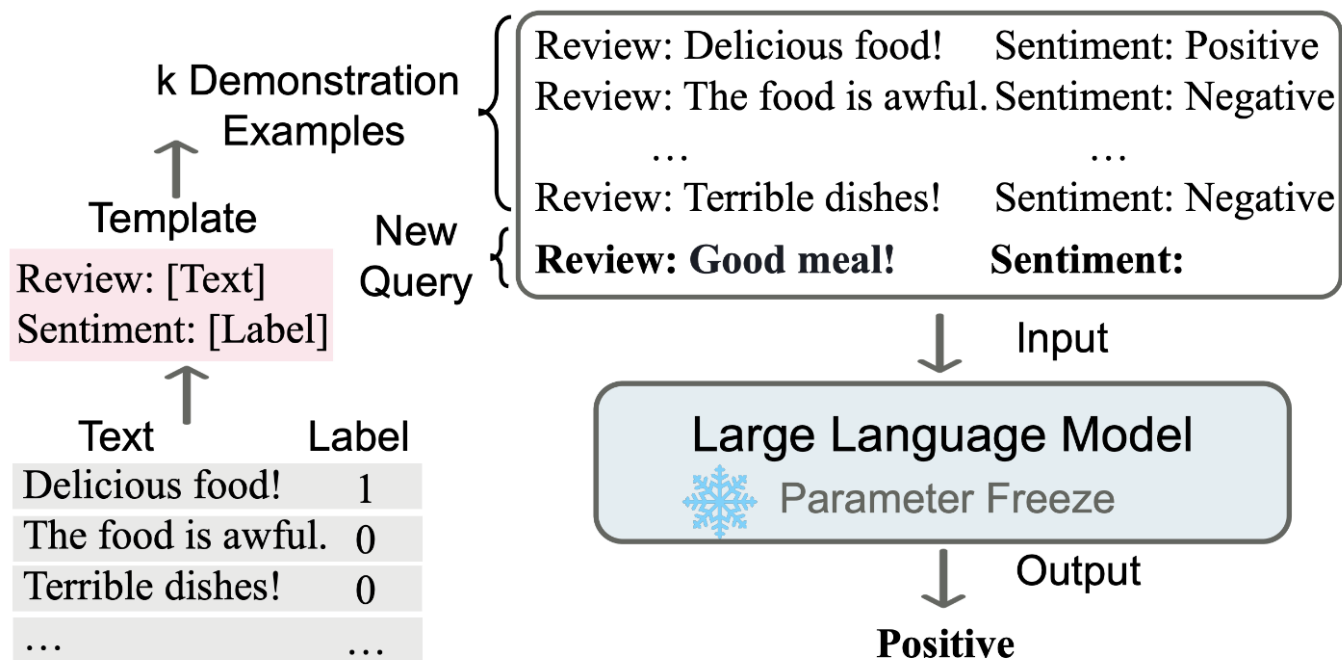


Wei, Jason, et al. "Emergent abilities of large language models." *arXiv preprint arXiv:2206.07682* (2022).

2. 涌现能力——典型涌现能力

• 上下文学习 (In-Context Learning)

- 在提示中提供自然语言指令和多个任务示例，无需显式的训练或梯度更新



给定k个示范样本，模型参数冻结

2. 涌现能力——典型涌现能力

• 指令遵循 (Instruction Following)

- 大语言模型能够按照自然语言指令来执行对应的任务

Instruction: I am looking for a job and I need to fill out an application form. Can you please help me complete it?

Input:

Application Form:

Name: _____ Age: _____ Sex: _____

Phone Number: _____ Email Address: _____

Education: _____ ...

Output:

Name: John Doe Age: 25 Sex: Male

Phone Number: ...

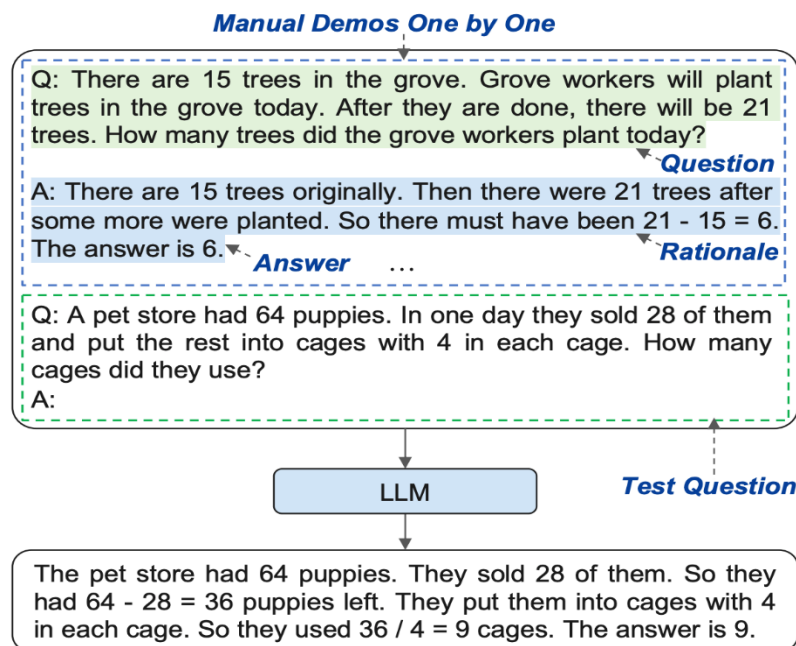


给定指令，模型输出

2. 涌现能力——典型涌现能力

• 逐步推理 (Step-by-step Reasoning)

- 利用思维链 (Chain-of-Thought, CoT) 提示策略, 在提示中引入任务相关的中间推理步骤来加强复杂任务的求解

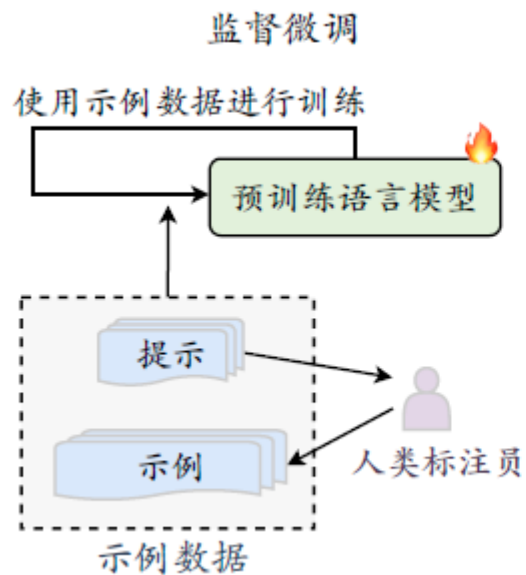


手动设计思维链步骤, 模型输出

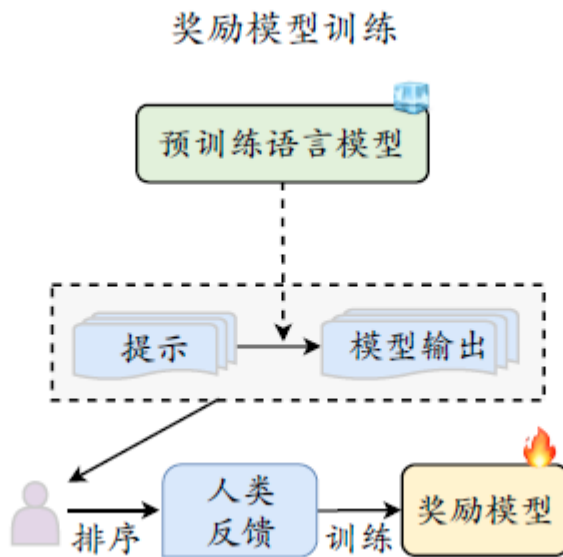
目录

- 01 大语言模型概述
- 02 大语言模型的能力特点
- 03 大语言模型的构建方法
- 04 大语言模型的应用

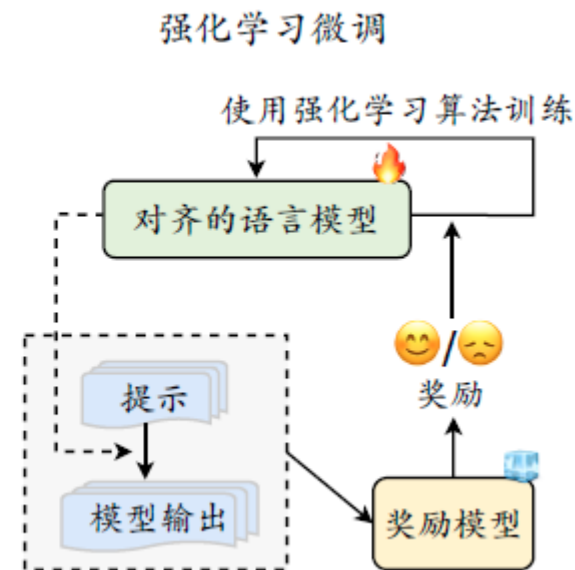
总体构建过程——以ChatGPT为例



1. 大规模预训练



2. 有监督微调



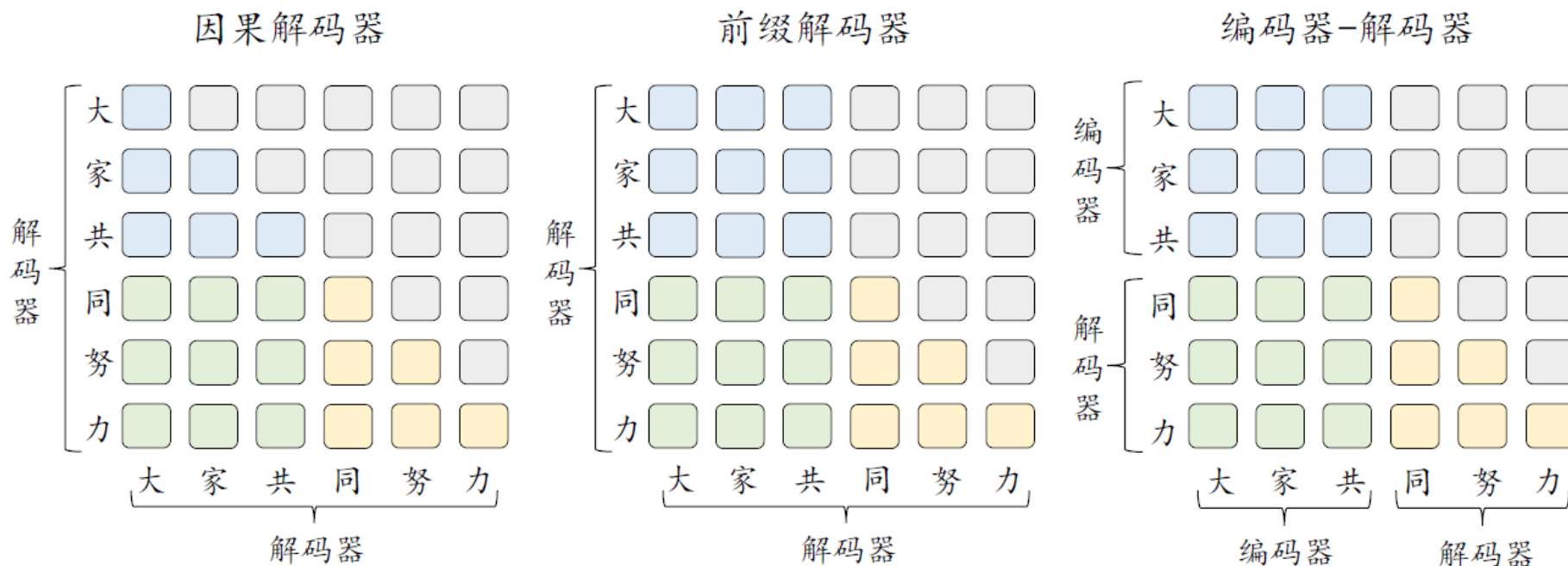
3. 人类对齐

1. 大规模预训练

- **大规模预训练(Pre-training)的目的**
 - 获得通用的语言理解与生成能力
 - 掌握较为广泛的世界知识
 - 具备解决众多下游任务的性能潜力
- 以上目的的达成，离不开
 - 大尺寸模型参数
 - 大规模训练数据

1. 大规模预训练

- 基础架构：基于Transformer的生成式架构
 - **因果解码器 (GPT/LLaMA)**：不显示区分输入输出，单向掩码注意力自回归预测
 - 前缀解码器 (GLM/U-PaLM)：对于输入（前缀）部分使用双向注意力进行编码；对于输出部分利用单向掩码注意力进行自回归地预测
 - 编码器-解码器 (FLAN-T5)：编码器双向注意力；解码器交叉+掩码注意力



1. 大规模预训练

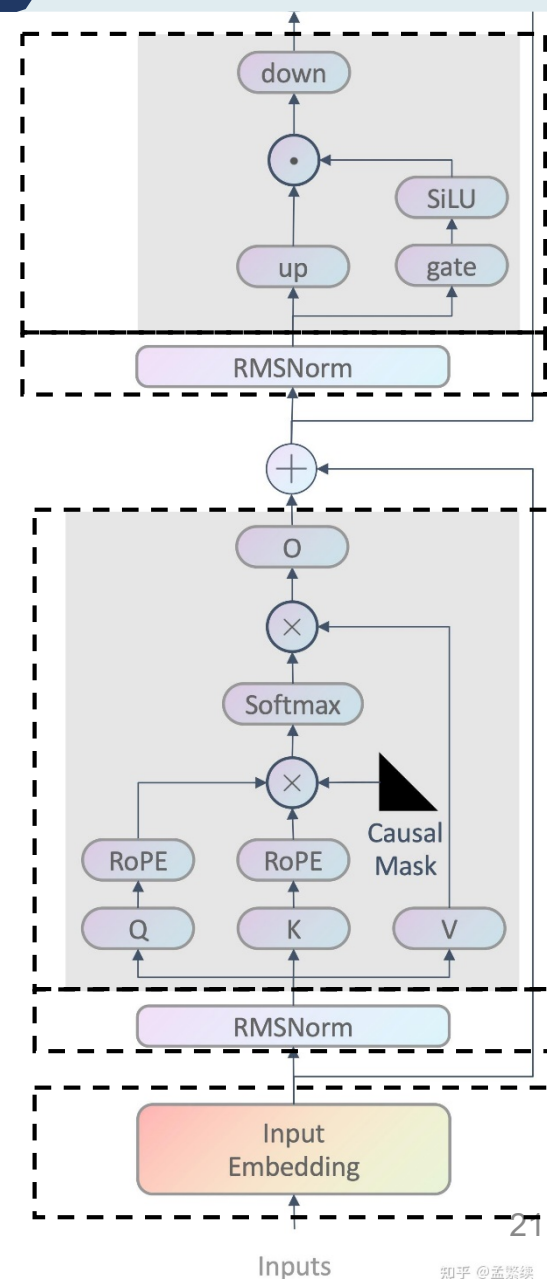
• 模型参数量估算：以LLaMA模型为例

词表大小	V	32000	解码器层数	L	32
中间状态维度	H	4096	前馈网络中间状态维度	H'	11008

1. 输入嵌入层 VH
2. 多头注意力层 $4 \times H^2$
3. 前馈网络层 $3 \times HH'$
4. 归一化层 $2 \times H$ (每层解码器) H (最终输出)
5. 输出层 VH

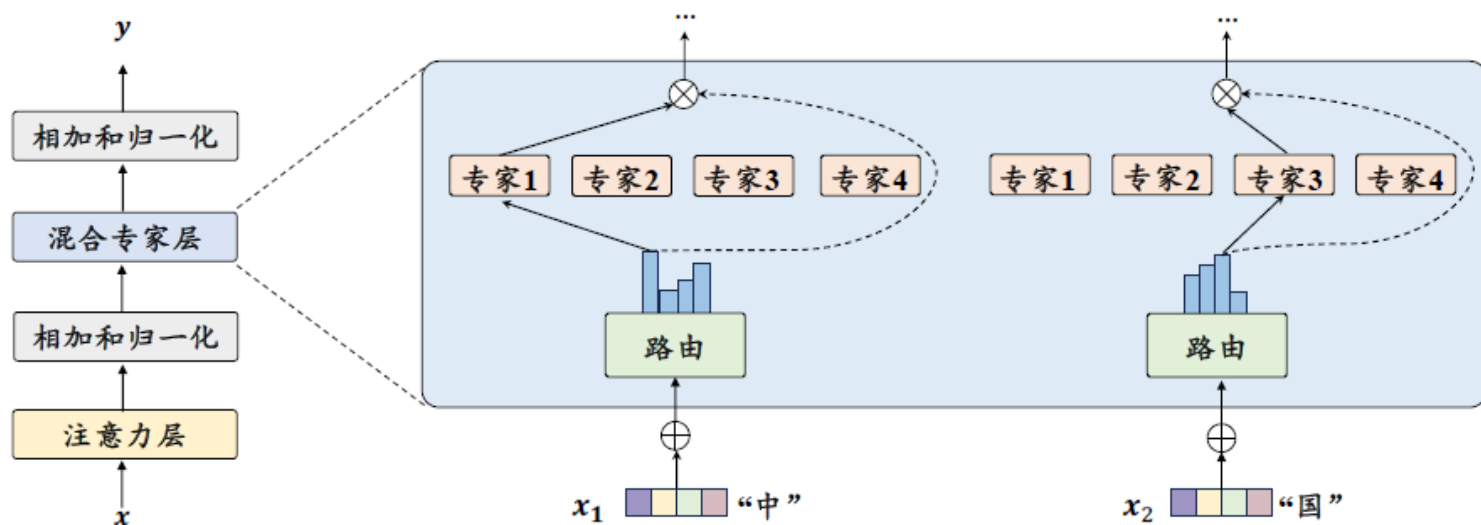
共计 $2VH + H + L \cdot (4H^2 + 3HH' + 2H)$

LLaMA(7B) 6, 738, 415, 616



1. 大规模预训练

- 基于稀疏激活的混合专家（Mixture-of-Experts, MoE）结构
 - 面向大语言模型建模的Transformer改进
 - 不显著提升计算成本同时实现对模型参数的拓展，被DeepSeek等模型使用
 - 每个混合专家层包含 K 个**专家组件**(前馈神经网络) $[E_1, E_2, \dots, E_K]$
 - 对于每个输入词元 x_t ，通过**路由网络** G 计算该词元对应于各个专家的得分
 - 得分最高 k 个专家激活，送入softmax计算权重，对应加权作为最终输出



$$G(x_t) = \text{softmax}(\text{topk}(x_t \cdot W^G))$$

$$o_t = \text{MoELayer}(x_t) = \sum_{i=1}^K G(x_t)_i \cdot E_i(x_t)$$

1. 大规模预训练

- 预训练任务：“预测下一个词元”

- 绝大部分大语言模型广泛采用
- 给定一个词元序列 $\mathbf{u} = \{u_1, \dots, u_T\}$ ，基于序列中当前位置之前的词元序列 $\mathbf{u}_{<t}$ ，采用自回归的方式对于目标词元 u_t 进行预测，即最大化以下概率

$$\mathcal{L}_{\text{LM}}(\mathbf{u}) = \sum_{t=1}^T \log P(u_t | \mathbf{u}_{<t})$$

- **变体：前缀语言建模**，配合前缀解码器架构
- 每个文本序列 \mathbf{u} 会根据随机选择的位置 k ($1 \leq k \leq T$) 切分为前缀 $\mathbf{u}_{\text{prefix}} = \{u_1, \dots, u_k\}$ 和后缀 $\mathbf{u}_{\text{suffix}} = \{u_{k+1}, \dots, u_T\}$ ，仅后缀中的词元损失会被计入总损失

$$\mathcal{L}_{\text{Prefix}}(\mathbf{u}) = \log P(\mathbf{u}_{\text{suffix}} | \mathbf{u}_{\text{prefix}}) = \sum_{t=k+1}^T \log P(u_t | \mathbf{u}_{<t})$$

1. 大规模预训练

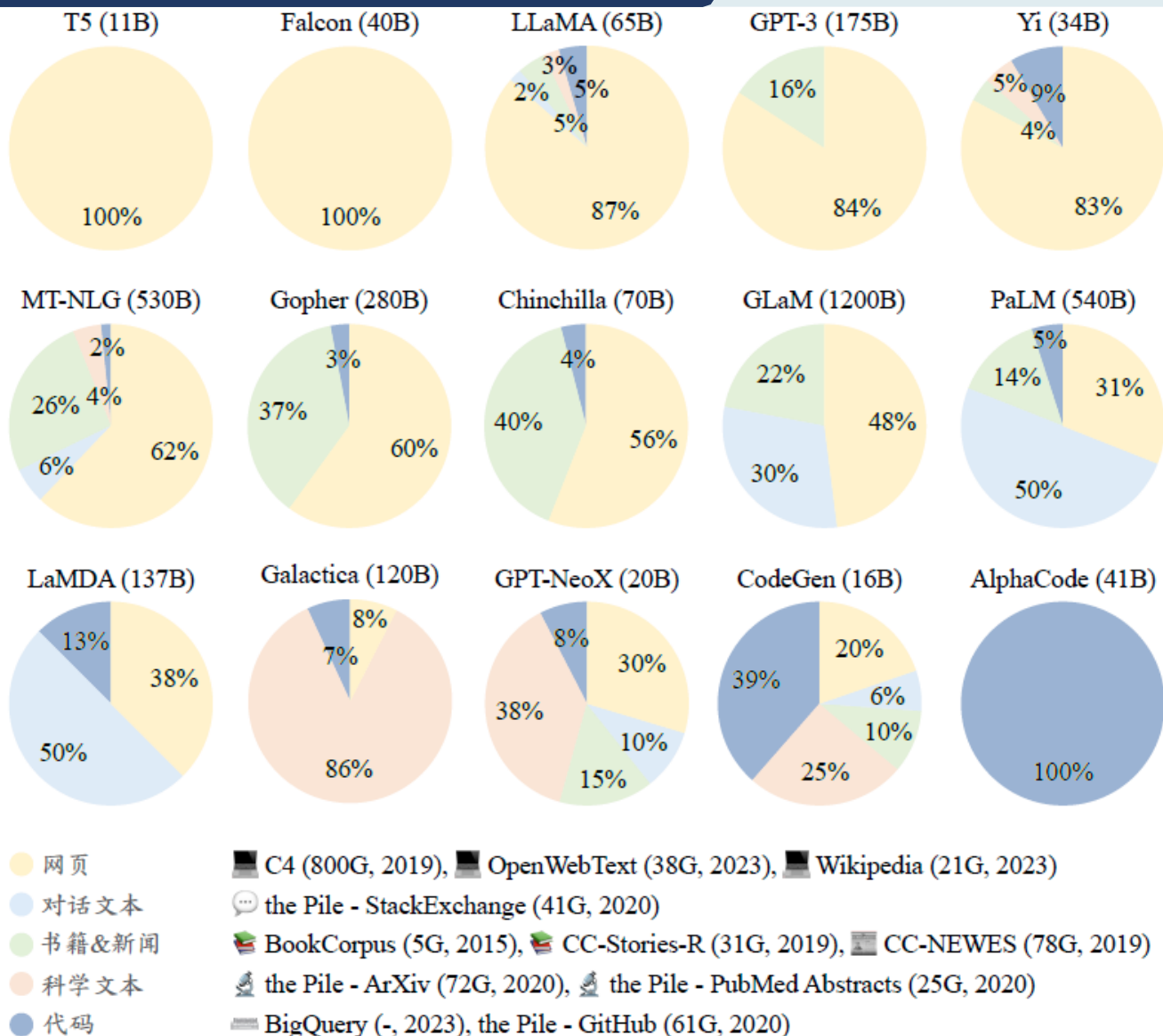
- 训练数据

- 通用文本数据

- 网页
 - 书籍
 - 对话文本，等等

- 专用文本

- 多语言文本数据
 - 科学文本
 - 代码，等等



2. 有监督微调

- **有监督微调 (Supervised Fine-tuning, SFT)**

- 使用自然语言形式的数据对预训练后的大语言模型进行参数微调
- 指令微调 (Instruction Tuning) / 多任务提示训练 (Multitask Prompted Training)

- **主要作用**

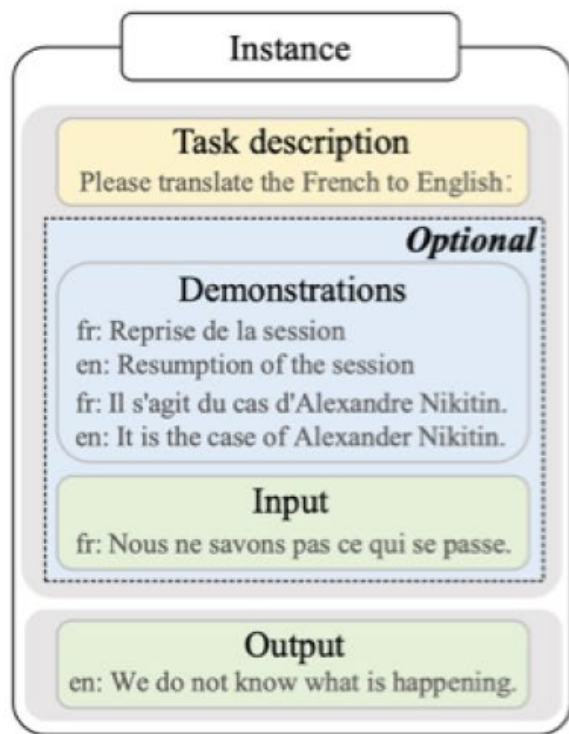
- **改进整体模型性能**: 相对预训练成本显著降低, 数据量仅为约万分之一甚至更少
- **增强任务求解能力**: 指导模型学会理解自然语言指令, 并据此完成相应的任务
- **适配各专业化领域**: 通用的大语言模型在特定领域 (如医学、法律和金融等) 的表现与领域专用模型的效果仍有一定差距

2. 有监督微调

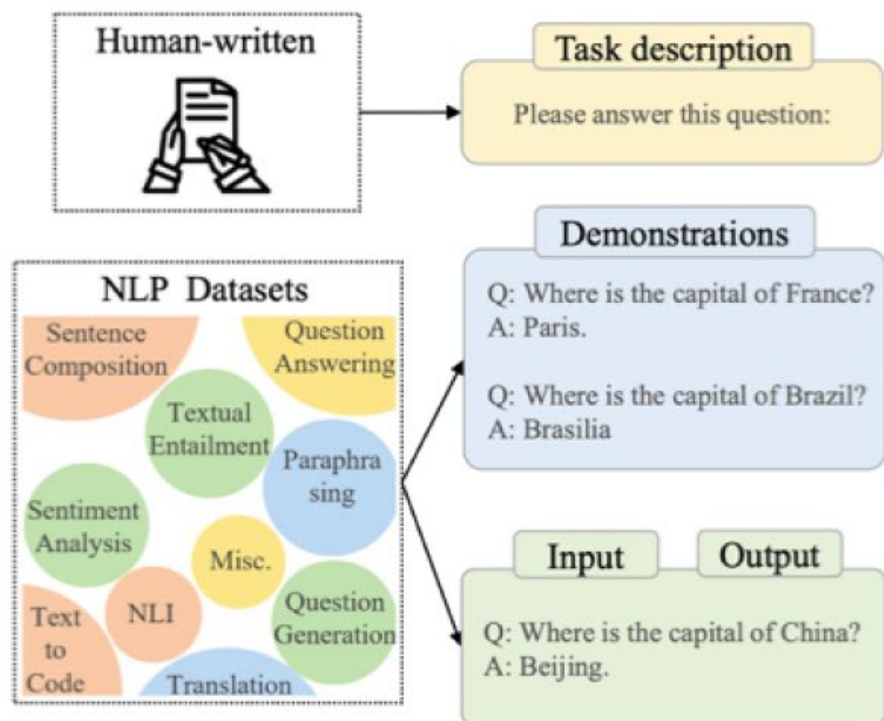
- 指令数据

- 任务描述（也称为指令 instruct） + 可选的示例 + 任务输入-任务输出

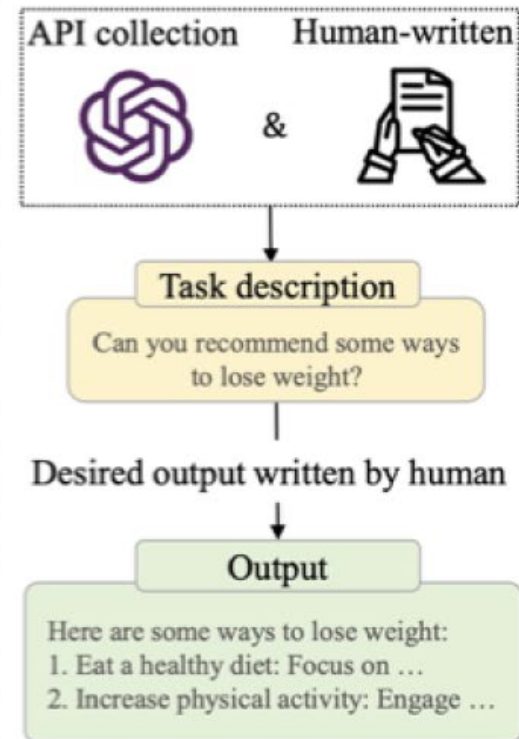
- 构建方式



(a) Instance format



(b) Formatting existing datasets

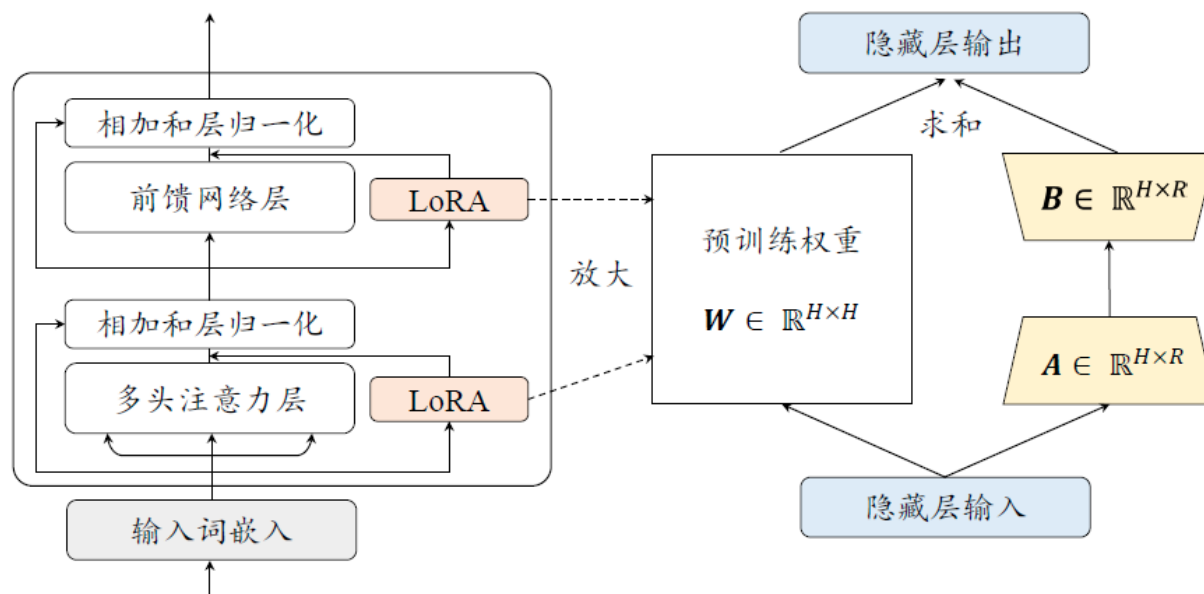


(c) Formatting human needs

2. 有监督微调

• 微调训练

- 在训练方式上与预训练相似，仅在输出部分计算损失
- 为使微调过程更加有效稳定，可在指令微调期间引入预训练数据和任务
- 参数高效微调（Parameter-efficient Fine-tuning）：全参数微调需要较多的算力资源开销，例如，低秩适配（Low-Rank Adaptation, LoRA）微调



3. 人类对齐

- **人类对齐 (Human Alignment)**

- 如何确保大语言模型的行为与人类价值观、人类真实意图和社会伦理相一致

- **必要性**

- 大模型有时会出现错误或具有危害性的行为，例如无法正确遵循指令、生成虚假信息、以及产生有害/有误导性以及带有偏见的表达等
 - 原因在于训练目标未充分考虑人类的价值观或偏好

输入：我想在一个月內減肥 10 斤。

输出 (无人類对齐训练)：每天只吃一个苹果，坚持高强度运动 6 小时。

(从单纯的逻辑和语言理解角度，它似乎明白了减肥的目标，但这个建议严重不符合人类健康和生活实际，可能会让你的身体受到极大损害)

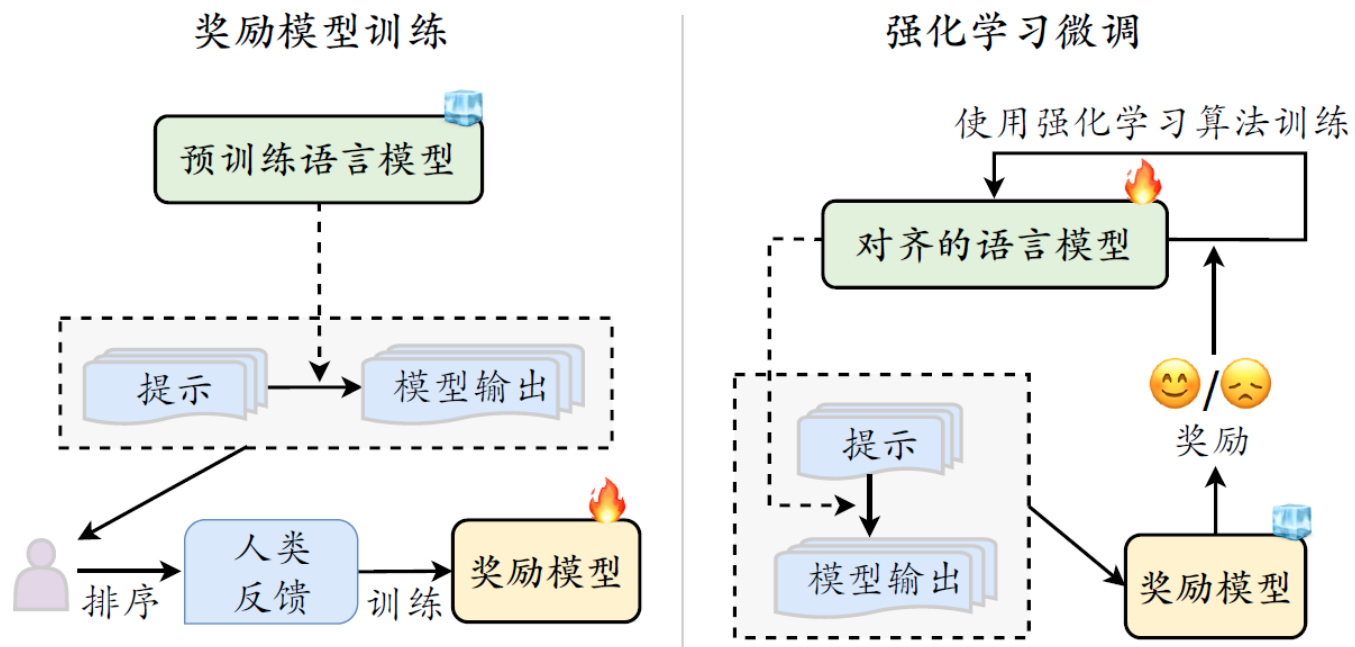
输出 (有人類对齐训练)：保持均衡饮食，减少高热量食物摄入，每天进行 30 分钟到 1 小时的有氧运动，搭配适量的力量训练，保证充足睡眠。

3. 人类对齐

- 对齐标准：具有一定的主观性，三个有代表性的标准
 - **有用性 (Helpfulness)**
 - 大语言模型需要提供有用的信息，能够准确完成任务，正确理解上下文，并展现出一一定的创造性与多样性
 - **诚实性 (Honesty)**
 - 模型的输出应具备真实性和客观性，不应夸大或歪曲事实，避免产生误导性陈述，并能够应对输入的多样性和复杂性
 - **无害性 (Harmlessness)**
 - 大语言模型应避免生成可能引发潜在负面影响或危害的内容
 - 处理敏感主题时应遵循道德标准和社会价值观，从而消除冒犯性与歧视性

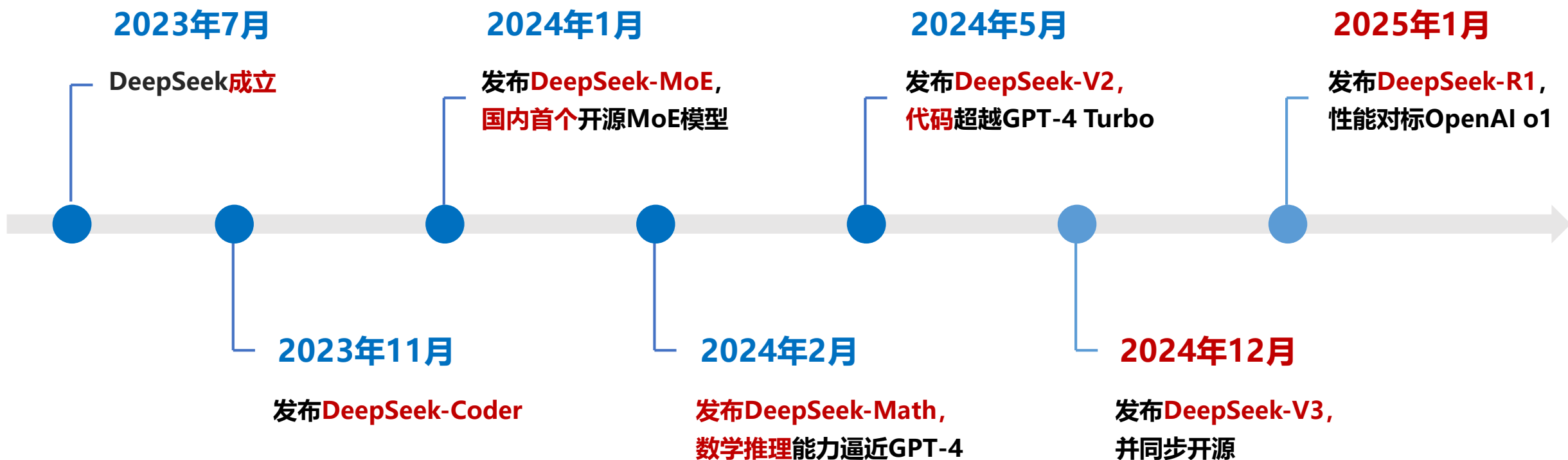
3. 人类对齐

- **基于人类反馈的强化学习** (Reinforcement Learning from Human Feedback, RLHF)
 - 对齐标准难以通过形式化的优化目标进行建模
 - 引入人类反馈对大语言模型的行为进行指导



奖励模型 (Reward Model)

- 在人类偏好数据上进行训练
- 对于模型生成内容进行质量评分
- 实现对于人类偏好分数的预测

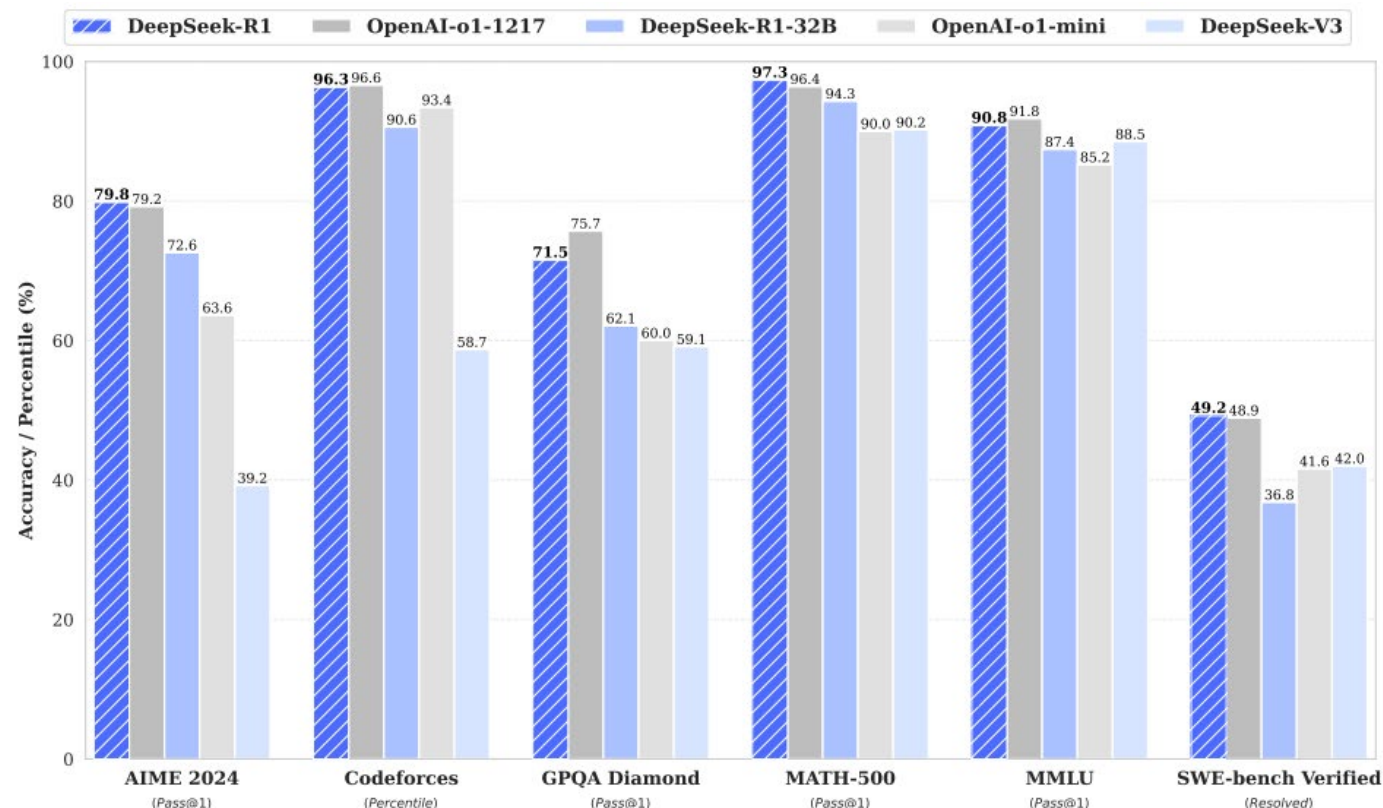


• DeepSeek V3

- 使用MoE结构，总参数量671B
- 性能对标OpenAI 4o

• DeepSeek R1

- 基于DeepSeek V3
- 性能对标OpenAI o1
- 实现长思维链(CoT)思考过程的可读化输出，显著提升体验时的新颖性和应用时的可解释性



Guo, Daya, et al. "Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning." *arXiv preprint arXiv:2501.12948* (2025).

• 主要创新

- R1-Zero强化学习：提出GRPO强化学习训练算法，直接在V3基座模型上，通过规则驱动强化学习，进行CoT推理能力的学习，不需要任何有监督标注数据

• 其他特点

- 在工程技术方面，通过多种技术创新，大幅提升训练效率，降低训练成本（据报道，DeepSeek V3训练成本是GPT-4o的1/20）
- DeepSeek V3/R1等模型开源，并且提供了多个经蒸馏学习的相对较小模型

• 存在问题

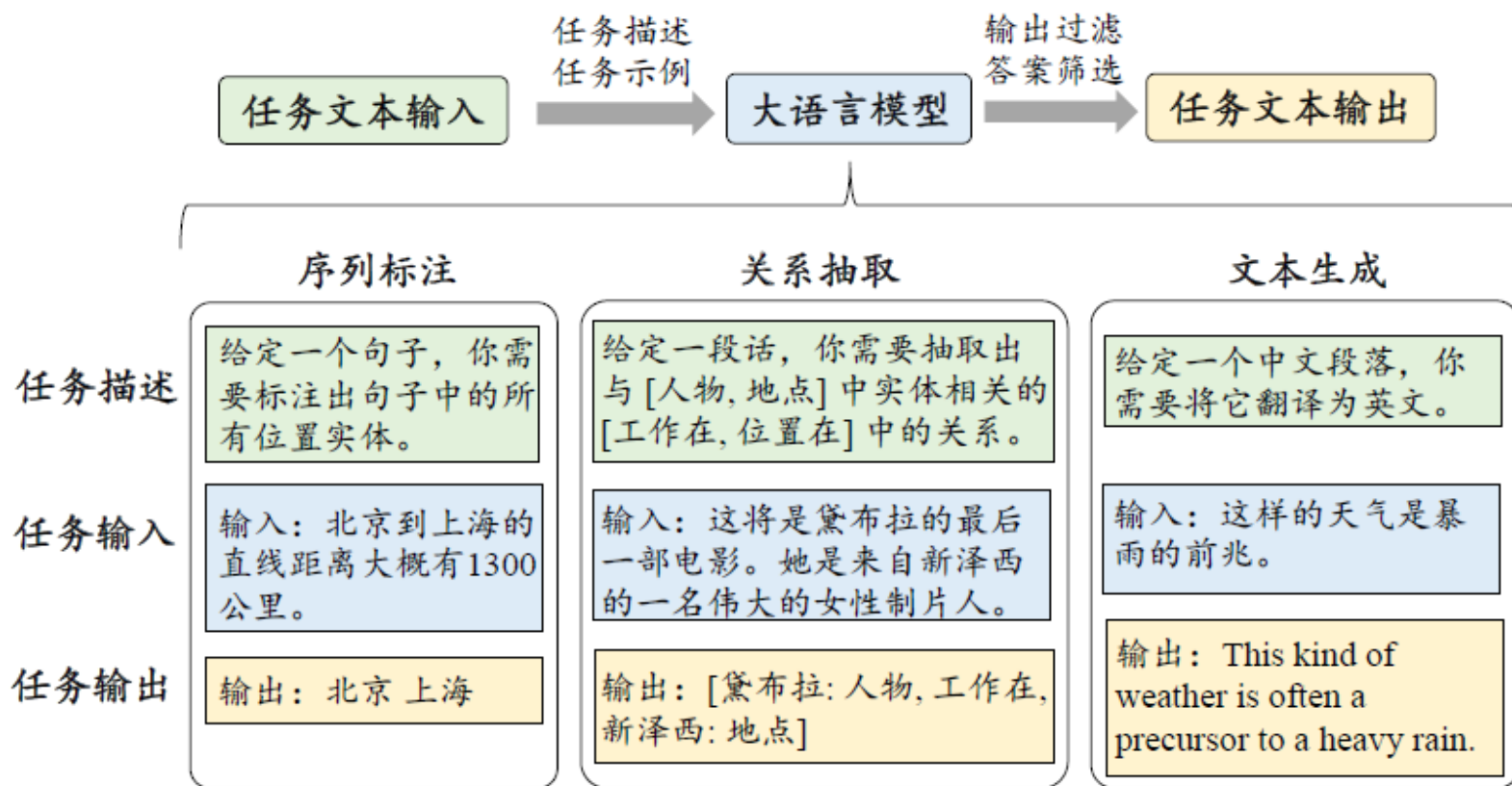
- 暂时不具备多模态能力
- 仍然存在内容安全与知识幻觉等问题

目录

- 01 大语言模型概述
- 02 大语言模型的能力特点
- 03 大语言模型的构建方法
- 04 大语言模型的应用

• 传统自然语言处理任务

- 为各种任务提供统一的解决方案，在零样本和少样本场景下取得有竞争力表现
- 但无法有效应对低资源领域的自然语言处理任务，如小语种翻译

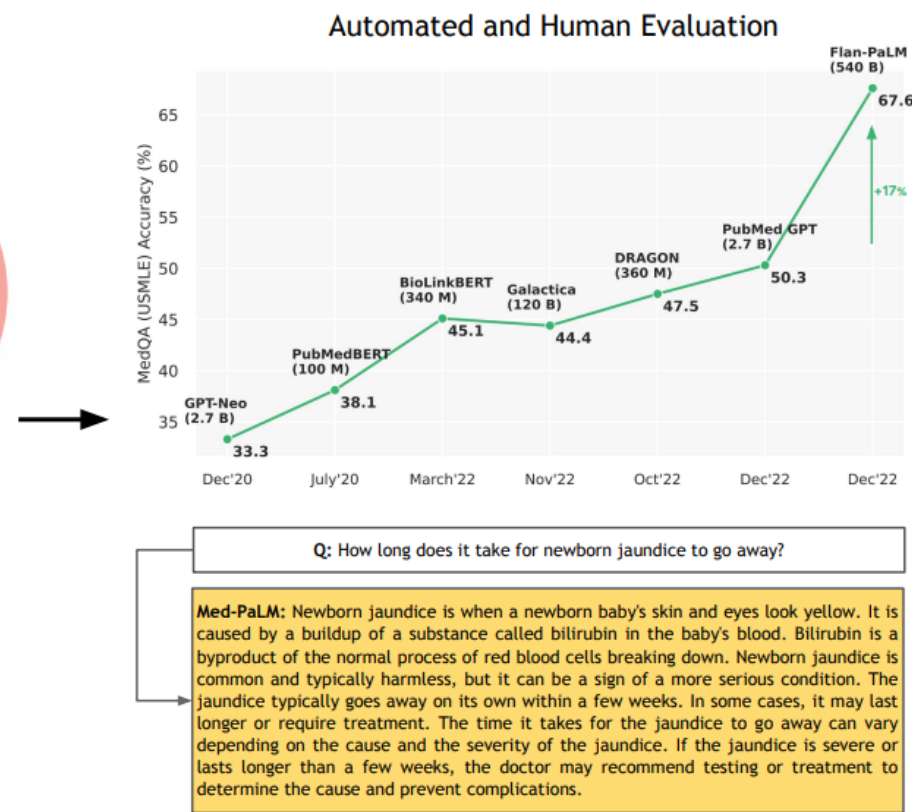
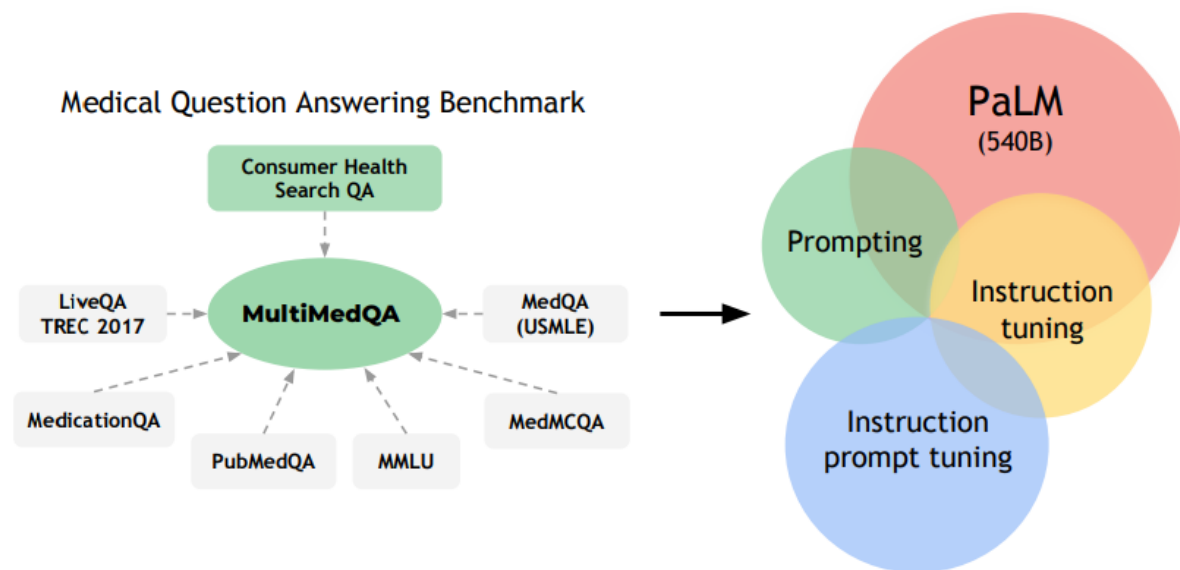


- **多模态大语言模型** (Multimodal Large Language Model, MLLM)
 - 能够处理和整合多种模态信息 (比如文本、图像和音频) 的大语言模型
 - 以视觉-语言大语言模型为例



• 医疗领域

- 数据来源主要包括电子病历、科学文献和医学问答等

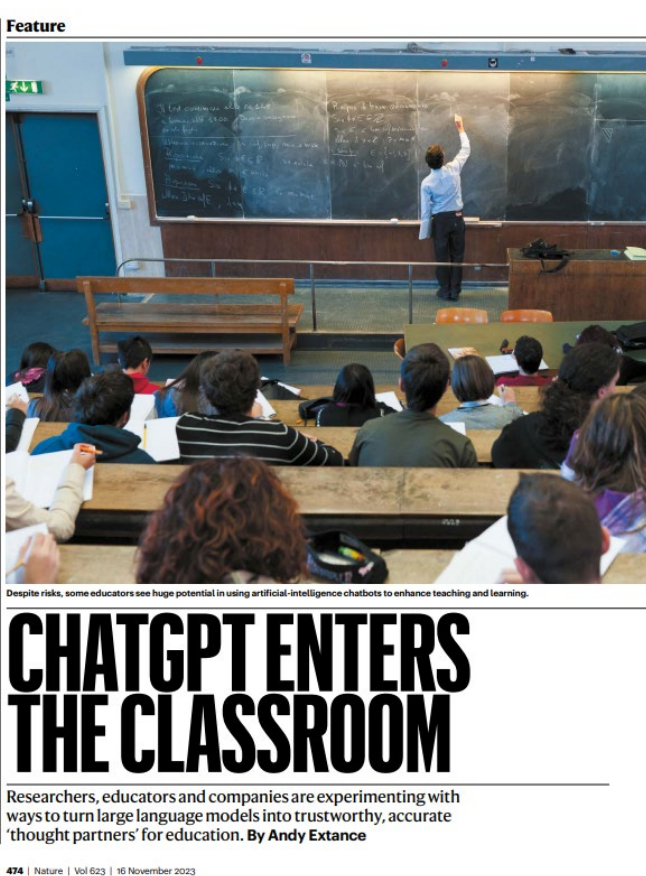


谷歌推出的医疗大语言模型Med-PaLM

Med-PaLM performs encouragingly on consumer medical question answering

• 教育领域

- 需要海量教育相关文本和专业数据对大模型进行训练，并结合大规模的对话数据进行指令微调，从而适配教育应用场景下的多种需求

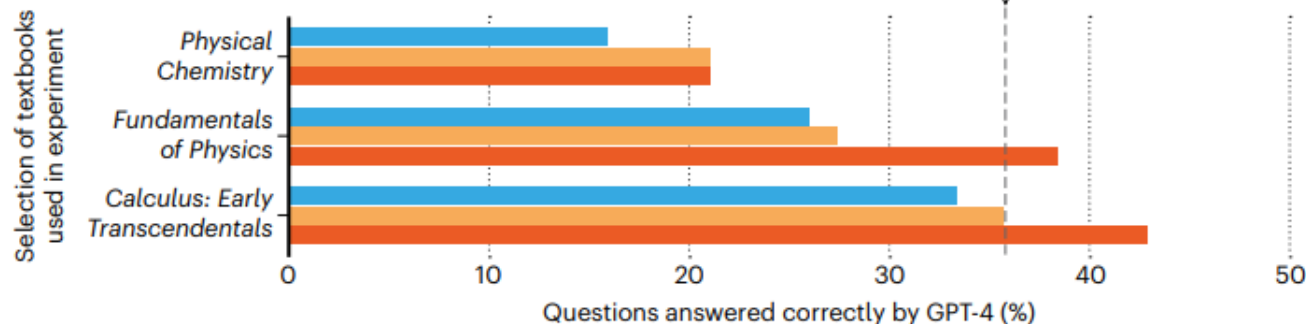


AI'S TEXTBOOK ERRORS

GPT-4 — one of the artificial-intelligence models behind the chatbot ChatGPT — is generally poor at answering problems from university-level science textbooks, researchers found, even though it had previously excelled at some mathematical tests. The scientists achieved only moderate boosts to GPT-4's performance by giving it some examples first, prompting it to break the problem down into steps or telling it to use external software to perform the calculations.

- Shown examples
- Broke problem into steps
- Used external tools for calculations

GPT-4 got its best average score across many textbooks* — 36% — when prompted to write code in Python to execute calculations. But this was still much lower than its previously reported scores, which were close to 90% on secondary-school and graduate-level tests of mathematical ability.



*Average includes tests from more textbooks, not shown here.

Extance, Andy. "ChatGPT has entered the classroom: how LLMs could transform education." *Nature* 623.7987 (2023): 474-477.

• 科学研究

- 研究人员需要面对复杂科学问题，处理分析大量数据，及时学习最新科学进展
- 大模型技术辅助人类的科研探索工作，推动科学研究的快速进展
- 例如，Meta AI 公司于2022 年11 月推出的科学大模型Galactica
 - 通过在48M 篇论文、教科书和讲义、数百万个化合物和蛋白质、科学网站、百科全书等大量科学相关数据上预训练得到的
 - 实验结果表明，Galactica 可以解决许多很多复杂科研任务，包括辅助论文撰写、物理问题求解、化学反应预测任务等

Prompt

The formula for Bessel's differential equation is:

Generated Answer

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

Prompt

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

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

Generated Answer



本节小结

- **架构**：基于Transformer的生成式架构
 - 解码器 / 编码器-解码器架构
 - 为何称之为“大”模型：模型参数量的估算
- **学习**：预训练 + 有监督微调 + 人类对齐
 - 使用简单自监督学习任务，获得解决各种下游任务的通用能力
- **能力特点**
 - 扩展法则：扩展模型规模与数据规模是本次大模型成功的重要因素
 - 涌现能力：上下文学习、指令遵循、逐步推理
- **展望**
 - 通过长思维链(CoT)方式，有望进一步提升模型解决复杂推理问题能力
 - 仍然存在幻觉问题、解释性欠缺、泛化性、安全性、资源依赖性等局限

设计一个和自己本专业相关的问题，使用DeepSeek进行回答。

- (1) 分别使用和不使用“深度思考(R1)”功能输出结果，判断两个结果是否正确并对比其差异。
- (2) 记录使用“深度思考(R1)”功能时，模型输出的思考过程。简述你对于此思考过程是否认可，以及是否还有可以改进的地方。

• 本节参考材料

- 赵鑫,李军毅,周昆,唐天一,文继荣, 大语言模型, <https://llmbook-zh.github.io/>, 2024.