



大语言模型

让我们更聪明还是更愚蠢

宗成庆 · 中国科学院自动化研究所



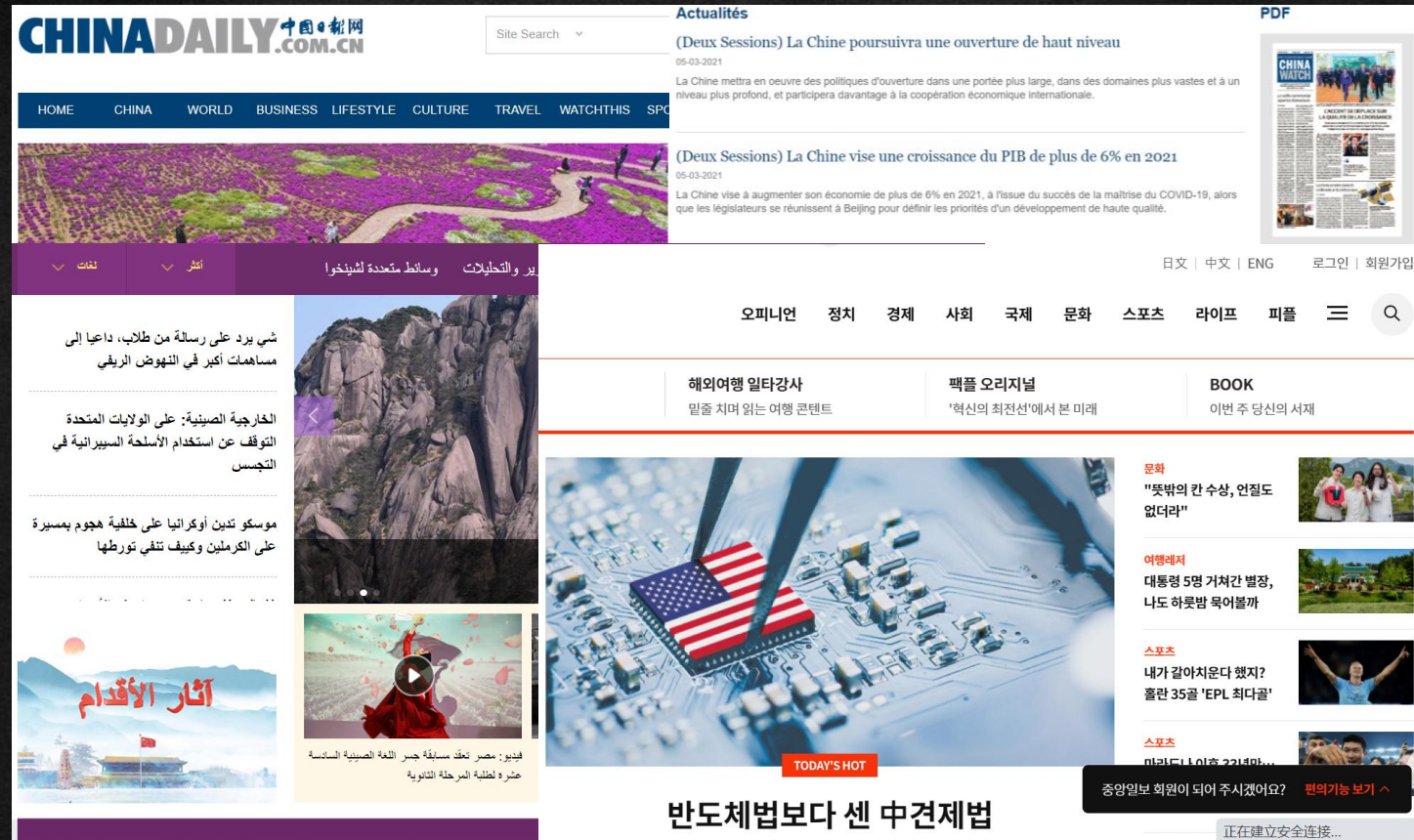
什么是自然语言处理？

人们处在不同的国家，使用不同的语言，通过不同的手段和方式发表不同的言论(社交网络、专著、论文、网页等)，千丝万缕的关系将他们联系在一起，构成一个特定的社会网络。

什么是自然语言处理?



什么是自然语言处理？



如何让计算机自动分析语言文本，理解人的意图，根据不同用户的需求实现个性化信息服务？这项技术就是自然语言处理。



- 作为语音助手、智能翻译等最贴近生活的人工智能应用背后的技术，自然语言处理被誉为人工智能皇冠上的明珠
- 几代人为了摘取这颗明珠不懈努力

第一次试验 ChatGTP (2022.12.15)

A: 我们聊聊滑雪吧。你会滑雪吗?

C: 对不起, 我是一个纯粹的语言模型, 我无法感知或了解外界环境。我无法参与体育活动, 也无法对滑雪进行评价。我只能根据您提供的信息来回答问题, 不能对外界环境进行评价或描述。如果您有其他问题, 我将尽力回答您的问题。

A: 你能写一首关于大雪的诗歌吗?

C: 虽然我不会滑雪, 但我可以尝试写一首关于大雪的诗歌:

大雪纷飞, 银装素裹。
白茫茫一片, 如同洁白的棉被。
踏着深厚的积雪, 挥动双手捕捉飞舞的雪花。
欢乐的嬉戏, 让人忘却寒冷的冬日。
希望您喜欢这首诗。

A: 还不错，就是不押韵。

C: 我抱歉地通知您，我只是一个纯粹的语言模型，我无法感知或了解外界环境，也无法理解诗歌的韵律和押韵。我的目的是提供简单的信息和回答问题，而不是写诗歌。如果您对我的诗有任何批评或建议，我将非常感激。

A: 你可以把这首诗翻译成英文吗？

C: 是的，我可以尝试把这首诗翻译成英文：

The snow is falling heavily,
Dressed in silver clothes.
The vast expanse is white,
Like a pure cotton blanket.
Walking on the deep snow,
Swinging your hands to catch the flying snowflakes.
Happy play,
Forget the cold winter.

希望您能喜欢这首诗的英文版本。

A: 你能把这首诗改写成散文吗?

C: 是的, 我可以尝试把这首诗改写成散文:

当大雪纷飞, 披着银色的衣裳时, 白雪覆盖了整个世界, 就像一床洁白的棉、被。踩着厚厚的积雪, 挥动双手捕捉飞舞的雪花, 在冬日的寒冷中尽情嬉戏, 忘却一切烦恼。

希望您喜欢这篇散文。

A: 谢谢你! 我快要下班了。

C: 不客气, 感谢您提出宝贵的问题。如果您有其他问题, 欢迎随时联系我, 我将尽力为您解答。祝您有一个愉快的下午!

具有出色的对话能力和语言表达能力



- 很好地理解上下文
- 清楚地把握用户意图
- 给出周到、全面的回复
- 规范的语言表达能力
- 长文本生成能力
- 多语言生成能力
- 甚至可以“交互式学习”
-

医学问题

Model: Default (GPT-3.5)

NW

有人长期运动，喜欢打球和爬山，可是最近左脚底前面部分出现麻木、疼痛，甚至剧痛症状。请问是什么原因造成的？



这种疼痛和麻木的症状可能是由多种因素引起的。以下是一些可能的原因：



1. 运动损伤：如果您运动时使用不正确的姿势或过度使用脚部肌肉，可能会导致肌肉、韧带或肌腱的拉伤或扭伤，从而导致疼痛和麻木。
2. 跖筋膜炎：跖筋膜是脚底部的一层薄膜，它支撑着脚弓。如果这个膜发炎，可能会导致脚底疼痛和麻木。
3. 足底筋膜纤维化：这是一种常见的足底疼痛症状，可能是由于过度运动、过度使用或足部负重过重引起的。
4. 神经问题：麻木和疼痛也可能是由于神经问题引起的，例如坐骨神经痛或足底神经痛。

如果症状持续存在，建议您咨询医生进行检查，以确定症状的确切原因，并获得适当的治疗。

法律问题

NW

A和B两个人在饭店吃饭时发生争执，A先动手打了B，但没有造成重伤。B报警后民警将两人带到派出所，B不同意和解。请问民警应该如何处理？



在这种情况下，民警应该根据相关法律法规，对A和B进行调查和处理，以维护公正和法律的权威。



1. 对A和B进行询问和取证：民警应当向A和B了解事发经过，了解是否有其他证人或证据，如监控录像等，以收集证据。
2. 进行医学鉴定：民警应当将B送到医院进行医学鉴定，确定受伤程度，以作为后续处理的参考。
3. 根据相关法律规定进行处理：如果A的行为构成了刑事犯罪，民警应当依照法律程序进行处理，可能会涉及到行政拘留或刑事处罚等。如果A的行为不构成犯罪，民警应当及时调解双方，促成和解，或者将案件移交到相关部门进行调解或处理。
4. 加强宣传教育：民警应当向双方进行法律宣传和教育，提高双方的法律意识和法律素养，以预防类似事件的发生。

在处理这种情况时，民警应当尽力保持公正和客观，遵循法律程序，依法处理，确保双方的合法权益得到维护。

算术问题

Model: GPT-4



When I was 6 my sister was half my age. Now I'm 70 how old is my sister?

当我6岁时，我妹妹是我年纪的一半。现在我70岁了，我妹妹几岁？



When you were 6 years old, your sister was half your age, which means she was 3 years younger than you. Now that you're 70, your sister would be $70 - 3 = 67$ years old.



67岁，回答正确，推理正确

在2022年高考全国试卷上测试：

对于客观题，（排除坏样本之后） ChatGPT在全部126个样本数可以达到**76%的准确率**，能力与两名500分左右的高考生（文科和理科各一名）相当。

对于主观题，根据人工对ChatGPT给出回答的打分，在文科综合（历史、地理、政治）试卷上**得分率达78%**。

——来自复旦大学邱锡鹏教授的测试结果



在各领域不同任务上,
都有中等水平以上的表现!

- 常识性知识
- 百科知识
- 数学问题求解
- 编写程序
-

但是，如果你问出下面的问题

NW

鲁迅和周树人是同一个人吗？



不是。鲁迅（1881年9月25日-1936年10月19日），原名周樟寿，字豫山，是中国现代文学史上杰出的文学家、思想家、革命家，代表作有《狂人日记》、《阿Q正传》、《呐喊》等。



周树人（1895年1月25日-1972年11月1日），原名周懋华，字有楼，是中国现代文学史上著名的文学家、教育家、翻译家，代表作有《故乡》、《春蚕》、《朱自清散文集》等。

两位文学家名字中的“周”字相同，但是他们的名字、生平、思想、作品等方面都有着很大的不同，是两个不同的人。

Model: GPT-4



你知道“林黛玉倒拔垂杨柳”的故事吗



“林黛玉倒拔垂杨柳”是中国古典名著《红楼梦》中的一段描写。《红楼梦》是清代作家曹雪芹创作的长篇小说，被誉为中国古典小说的四大名著之一。  

故事发生在林黛玉来到荣府后，她因生性聪慧、才情出众，深受贾母与王夫人的宠爱。有一天，林黛玉在荣府的大观园里散步，看到一片美丽的垂杨柳。当时垂杨柳正值春天繁茂，杨柳依依，景色宜人。然而林黛玉却对这垂杨柳心生厌恶，她认为它们像病夫一样无精打采，不值一顾。于是林黛玉倒拔了其中一棵垂杨柳。

这个故事反映了林黛玉的性格特点，她聪明、机智、独立，有时也带有一些叛逆。在《红楼梦》中，林黛玉是一位极具个性和魅力的人物，她的故事至今仍具有很高的文学价值。

A: 请介绍一下鸡兔同笼问题。

C: 鸡兔同笼问题是一个经典的数学问题，问题的形式通常是这样的：**如果有一个笼子里有鸡和兔子，总共有 n 只脚，问笼子里有多少只鸡和多少只兔子**。这道题目的答案通常可以用一个简单的方程来推导出来：如果 c 只鸡和 r 只兔子，那么 $c + r = n$ ，并且 $2c + 4r = 2n$ 。由于鸡和兔子的脚数之和是已知的，所以我们可以用这两个方程来求解 c 和 r 的值。例如，如果 $n = 10$ ，那么 $c + r = 10$ ，并且 $2c + 4r = 20$ ，我们可以用**第一个方程来推出 $c + r = 10$** ，然后用**第二个方程来推出 $c = 5$ 和 $r = 5$** 。这样，我们就可以得到答案：笼子里有 5 只鸡和 5 只兔子。

A: 如果鸡和兔子一共36个头, 94只脚。那有多少只鸡, 多少只兔子呢?

C: 根据鸡兔同笼问题的推导方法，我们可以得到两个方程来求解这个问题：

$$c + r = 36$$

$$2c + 4r = 94$$

由于 c 和 r 分别代表鸡和兔子的数量，因此我们可以**用第一个方程来推出 $c + r = 36$** ，**后用第二个方程来推出 $c = 23$ 和 $r = 13$** 。因此，答案是：笼子里有 23 只鸡和 13 只兔子。

正确的答案：25只鸡($c=25$)；11只兔子($r=11$)。

难以掩饰的愚笨和无知



一本正经地胡说八道

模型的内生性缺陷

- 可解释性
- 知识来源缺乏引用，特定领域的专业知识匮乏
- 出现“臆想”答案
- 事实性错误.....

由于对齐而产生人为倾向性

- 回复带有主观性倾向/价值观
- 无法展现和平衡多种观点
- 缺乏灵活的角色和具身感知能力

缺乏精深的专业知识，尤其最新的知识



**为什么 ChatGPT 会表现出远超以往的能力？
它是有哪些独特的技术吗？**

自然语言处理关键技术的发展

- 
- 语言模型(language model) -1980S
 - 神经语言模型(neural network based language model) -2003
 - 注意力机制(attention) -2017
 - 预训练语言模型(pre-training language model, PLM) -2018
 - 大语言模型(large language model, LLM) -2020

生成式语言模型

已知前面几个词，通过概率来预测出下一个最有可能出现的词。

那 是 2023 年 阳春

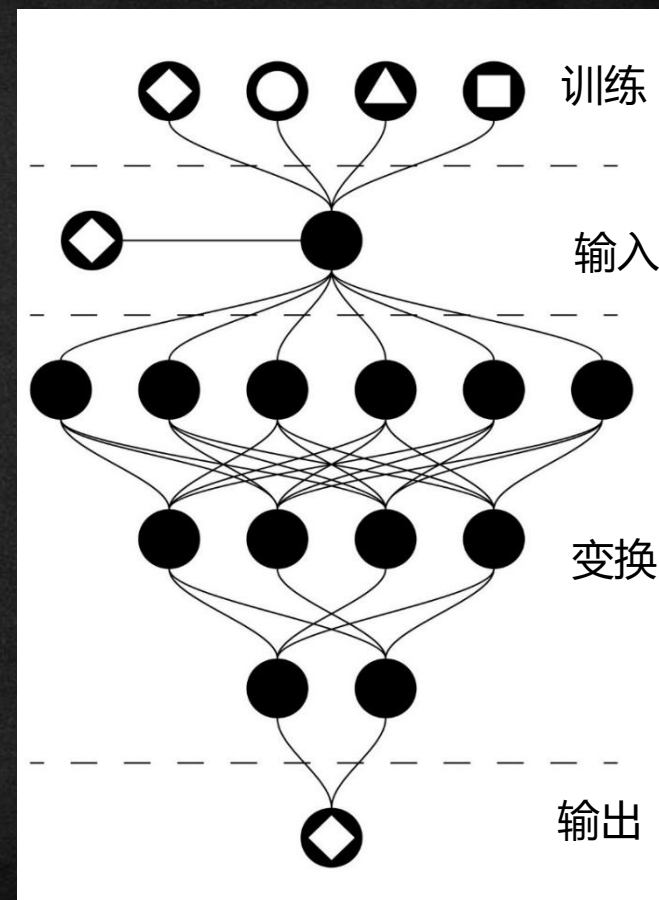
$n = 4$

问题：只考虑前面有限的几个词，没有考虑到同义词和不同词之间的关系等.....

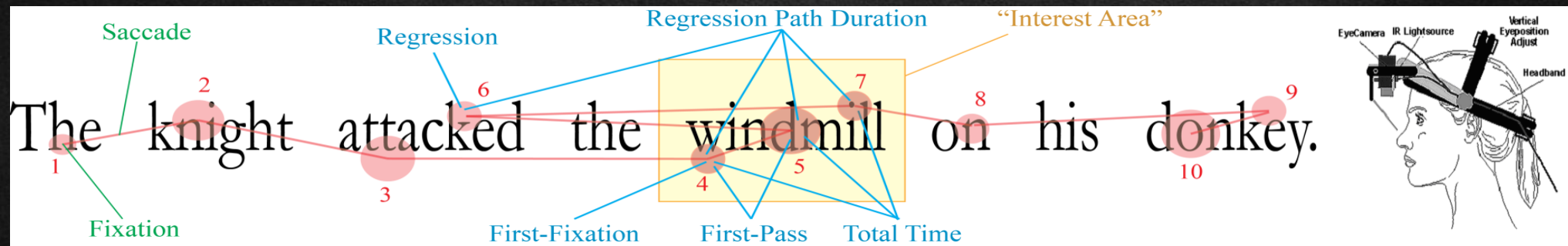
神经语言模型

那是 2023 年 阳春？

三月的一个早晨，美丽的西子湖畔

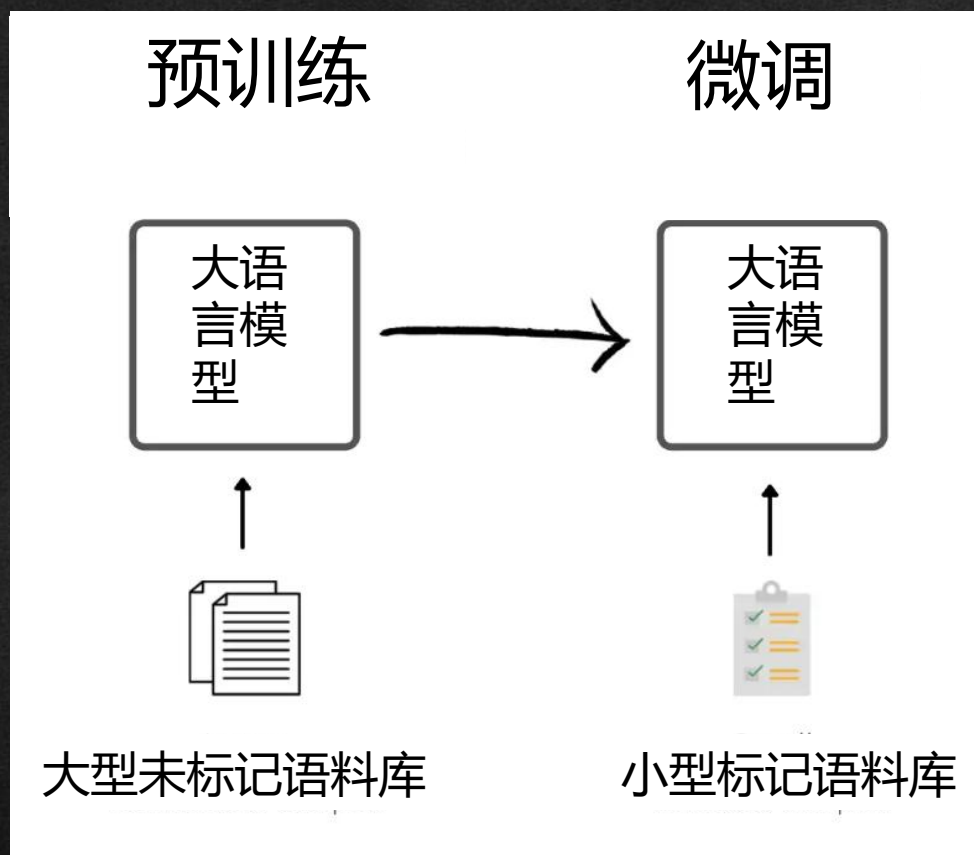


注意力机制



预训练语言模型

GPT: Generative Pre-Trained Transformer
(生成式预训练Transformer模型)



ChatGPT 家族

模型	GPT (2018)	GPT 2.0 (2019)	GPT 3.0 (2020)	ChatGPT (2022)	GPT 4.0 (2023)
参数量	1.17亿	15亿	1750亿	1750亿	未知
学习目标	无监督语言模型(Pre-training)/有监督微调(fine-tune)	多任务, P(output input, task), Zero Short Task Transfer	小样本学习(Few shot)	基于人类反馈的强化学习(RLHF)	
主要区别		增加语料、层数、维度 LN前移, 最后加LN, 初始化scale	增加语料、层数、维度	训练方法	多模态, 增加了图片、图像处理能力
数据集	7000未出版的书, 长文较多	40GB网络文本, 800万文档	互联网信息, 网络文本, 书籍, 维基百科, 45TB	+人工标注数据 +用户反馈数据	

大语言模型(LLM):
参数量至少在
100亿以上。

ChatGPT 的三项基本技术

预训练
大模型

利用大量**文本数据(45T)**
和**代码数据(179G)**

指令微调

收集真实的人类指令，
并由人工书写答案。

基于人类
反馈的
强化学习

使得模型生成的答案更
符合用户偏好。

大数据

单次训练费用达到460万美元，耗时一个月，总训练成本1200万美元。

大算力

- 定制具有28.5万个CPU核心，1万个GPU(V100)超级计算机
- 400GB/s的GPU服务器网络传输带宽
- 每次用户与ChatGPT互动，算力的云服务成本为0.01美元（OpenAI）

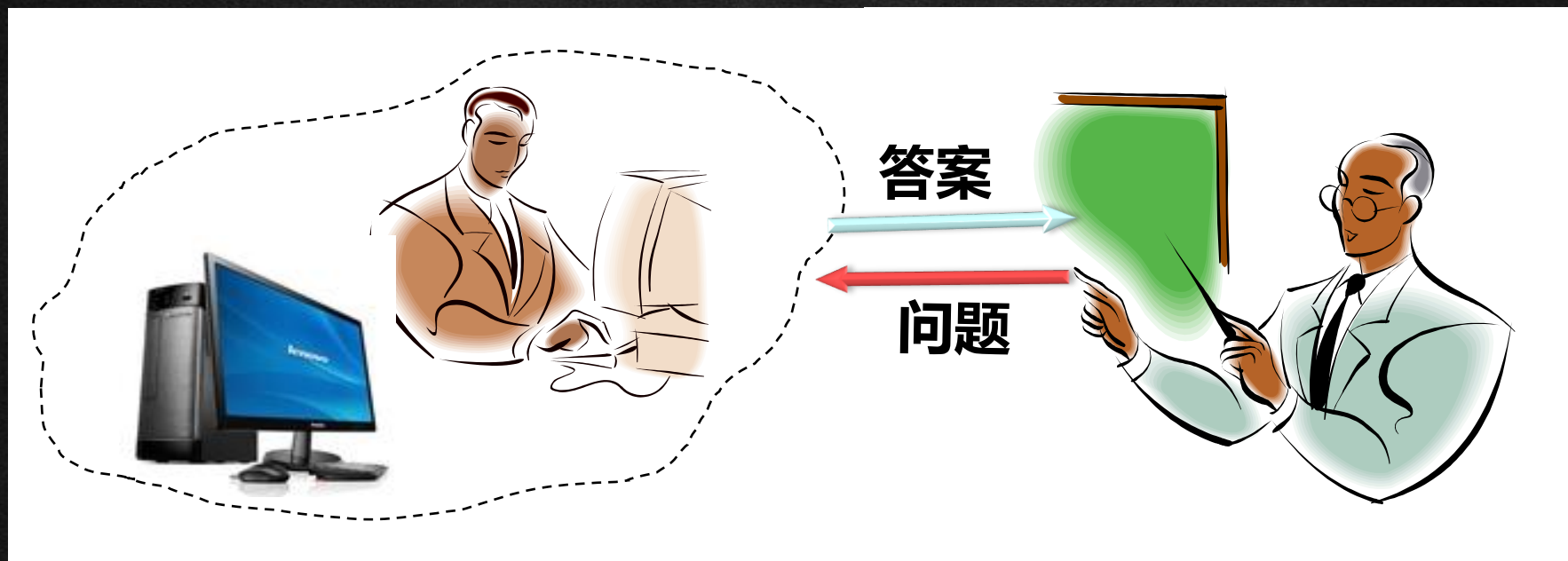
大人工

- 人工标注约2万~10万单词/人·天
- 10人一组，每类信息分为n组



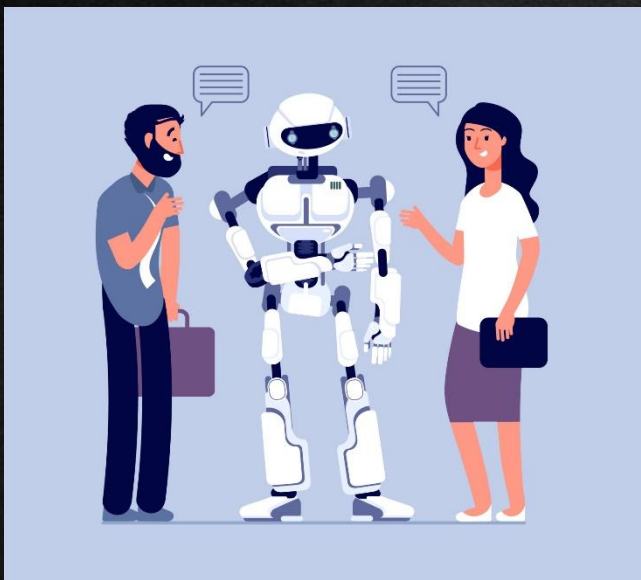
**为什么大语言模型的能力
出现了质的飞跃？**

大语言模型是很可能通往通用人工智能 (AGI) 的道路

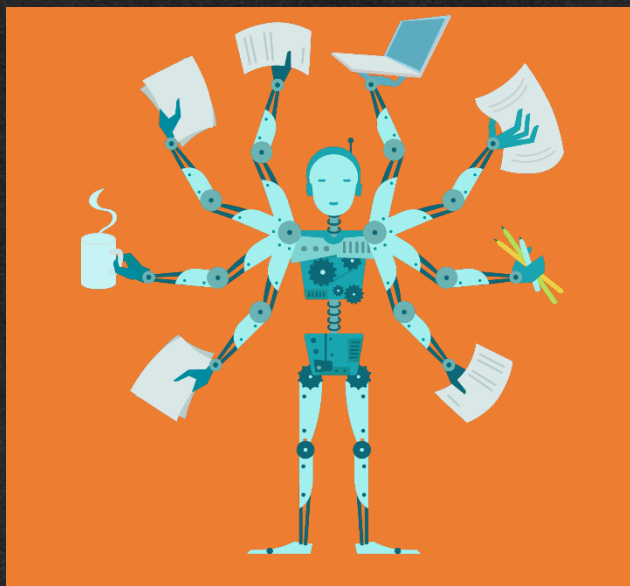


图灵测试(Turing test)

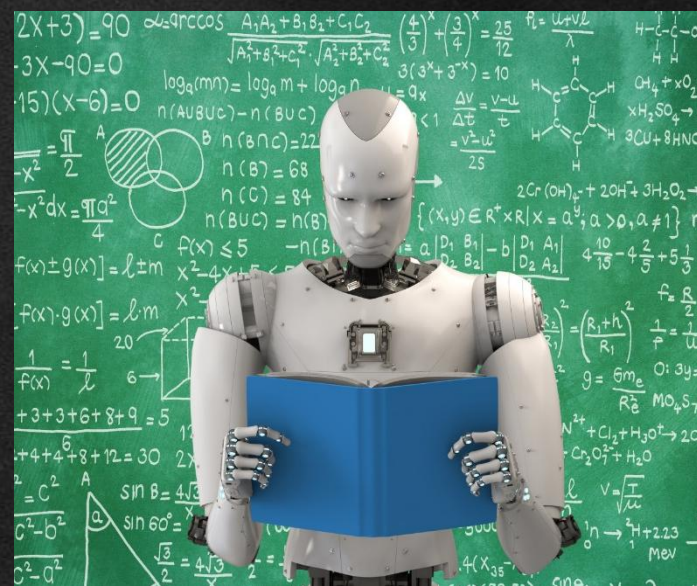
大语言模型带来的影响



提高工作效率，拓展个人视野



拥有全能的AI助手
(百科、写作、翻译、编程.....)

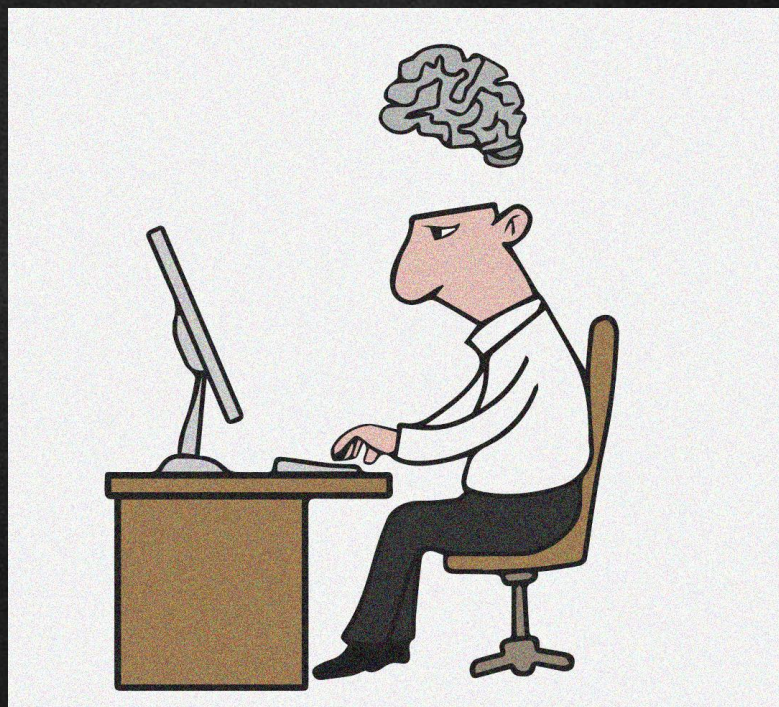
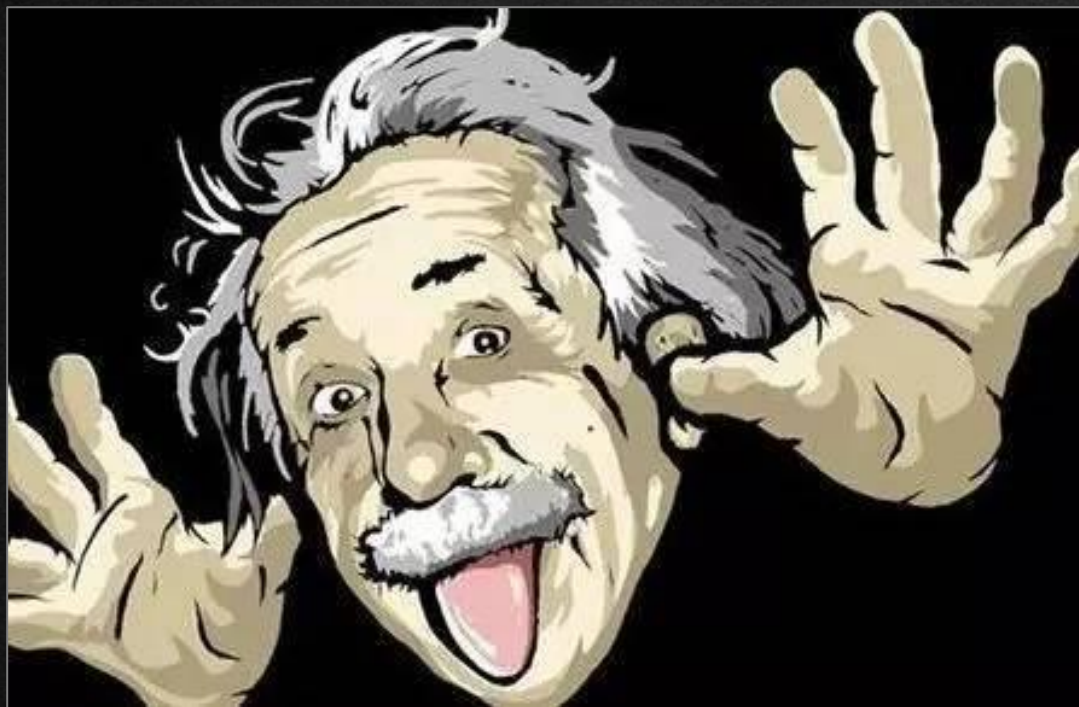


专业领域的建议者
(医疗、法律，甚至是科研.....)

大语言模型带来的影响

- 一元化的思想意识、价值观、乃至文化偏见
- 一切难题都交给AI，学生还需要学习吗？
- 错误事实/ 观点蔓延、渗透，诚信体系破坏，尤其在科研领域
- 社交替代，人与人的情感交流能力减弱
-

我们正站在一个十字路口上
人类将从此变得更加聪明，或是更加愚蠢？



我们要怎么做？

对于普通民众，没有理由，也没有必要限制使用大语言模型

对于学术界，确定使用的边界，需要使用者诚实、真实地说明使用的方式方法



对于自然语言处理研究者：

沉下心来
做好自己的事情
做出中国的大语言模型



对于自然语言处理研究者：

沉下心来
做好自己的事情
做出中国的大语言模型



对于自然语言处理研究者：

沉下心来
做好自己的事情
做出中国的大语言模型



对于自然语言处理研究者：

沉下心来
做好自己的事情
做出中国的大语言模型