

未来已来：

ChatGPT

底层原理与未来影响

Contents

感谢

本材料大部分参考渐构社团
于建国博士 (YJango) 视频：
[万字科普GPT4为何会颠覆现有工作流：为何你要关注微软Copilot、文心一言等大模型](#)



基本概念

- 搞清楚那堆《专业名词》



底层原理

- GPT如何回答问题?



训练过程

- ChatGPT是怎么训练的?加强版搜索引擎?



未来影响

- 它冲击了哪些职业?社会会变成怎样?



如何应对

- 我们如何保持竞争力?毕业=失业?

这些是啥？



这些是啥？

不知道？？问ChatGPT！



AIGC



图来源：量子位、腾讯研究院，甲子智库梳理，2023年)

什么是AIGC?

AIGC (Artificial Intelligence Generated Content, 人工智能生成内容) 是指通过人工智能 (AI) 技术生成的文本、图像、音频或视频等多媒体内容。这些内容通常由神经网络模型 (如生成对抗网络GANs、变分自编码器VAEs或大型语言模型如GPT系列等) 自动产生，模仿人类创作的各种形式。

AIGC在许多领域都有广泛的应用，包括：

1. **文本生成:** AI可以生成新闻报道、博客文章、创意写作、广告文案等。
2. **图像生成:** AI可以生成独特的艺术作品、设计元素或虚拟角色。
3. **音频生成:** AI可以生成音乐、语音合成或声音效果。
4. **视频生成:** AI可以生成动画、虚拟角色表演或视觉效果。

然而，AIGC也面临一些挑战，如生成内容的质量、模型的偏见、侵犯知识产权等问题。尽管如此，AIGC在各种领域的应用前景仍然广阔，有望为人类创作提供新的可能性和机会。

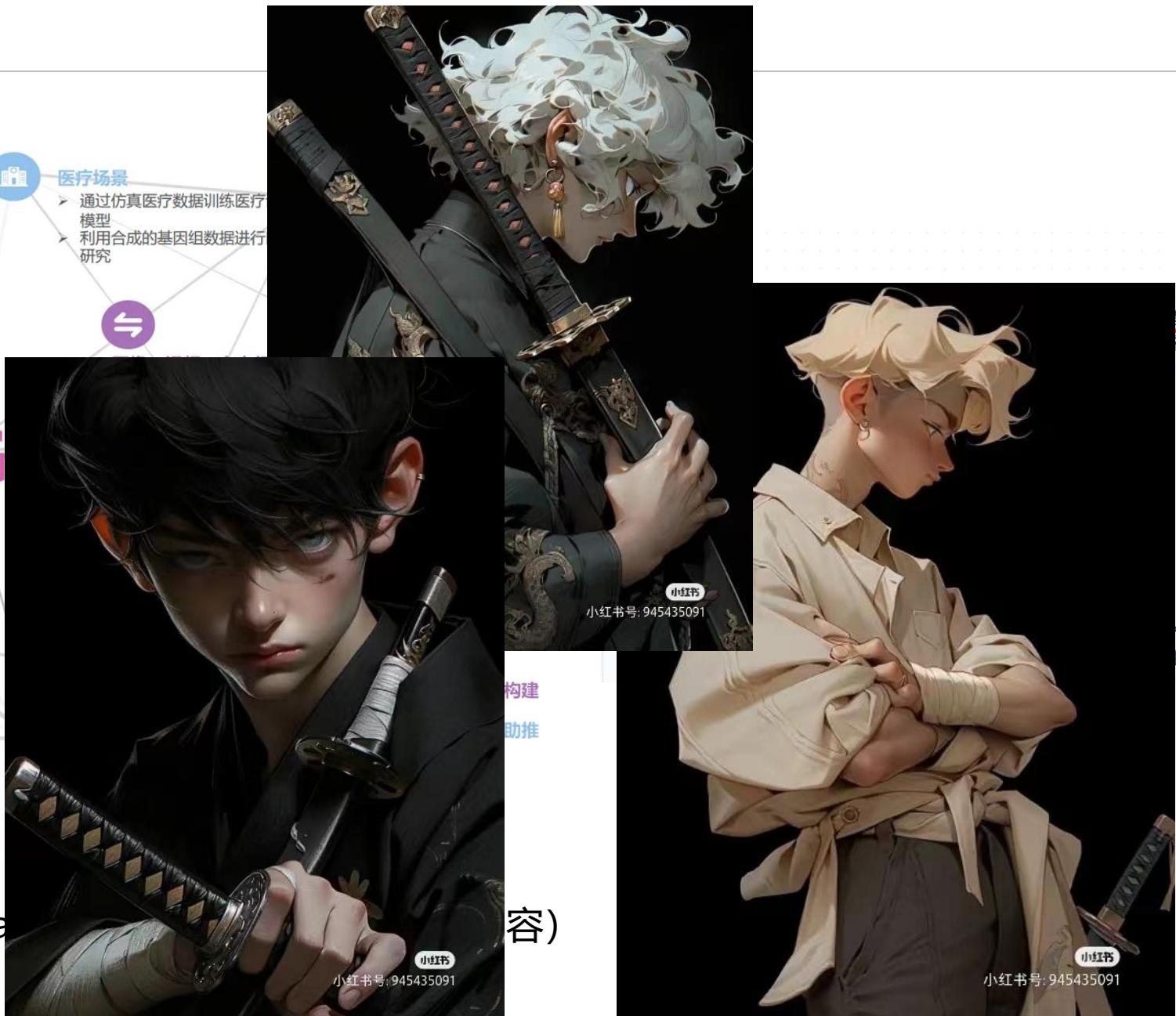
AIGC (Artificial Intelligence Generated Content, 人工智能生成内容)

AIGC



图来源：量子位、腾讯研究院，甲子智库梳理，2023年)

AIGC (Artificial Intelligence Genera



AIGC

Midjourney: 马斯克在广州城中村

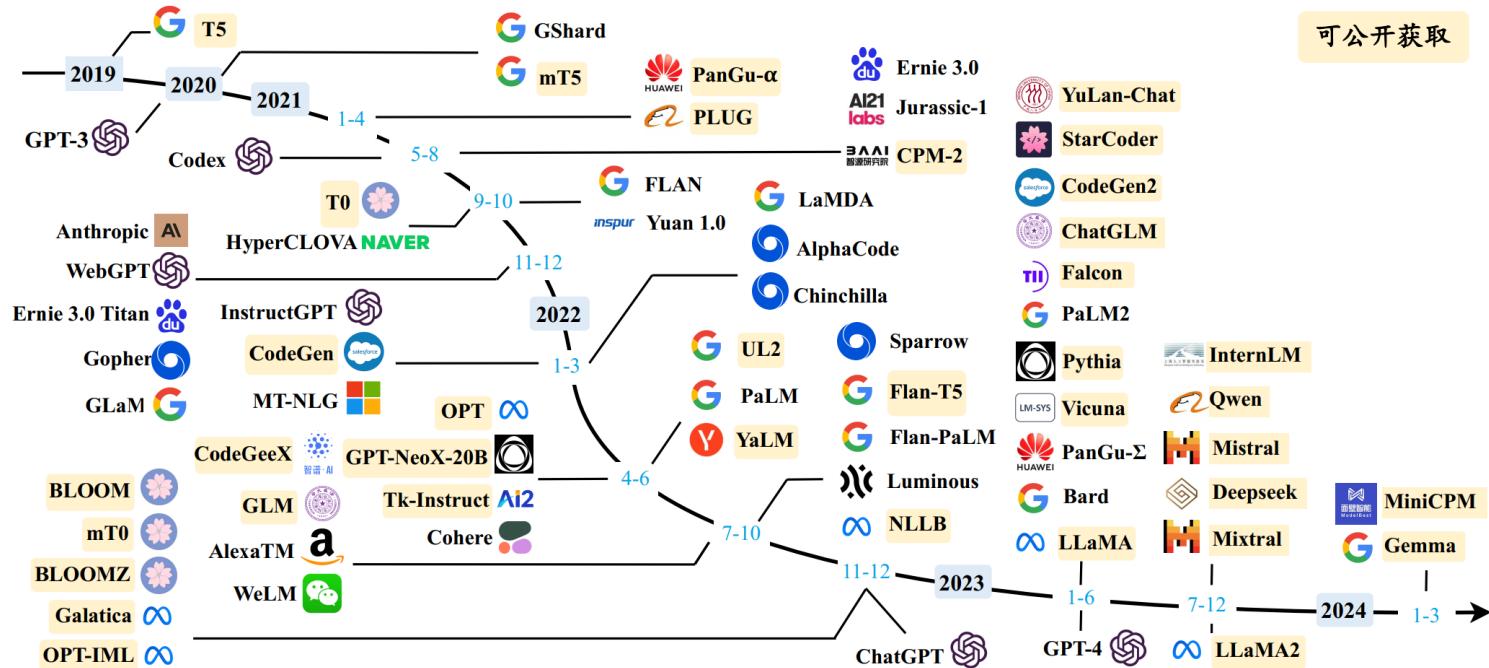


AIGC (Artificial

人工智能生成内容)



LLM



可公开获取

什么是LLM(large language model)?

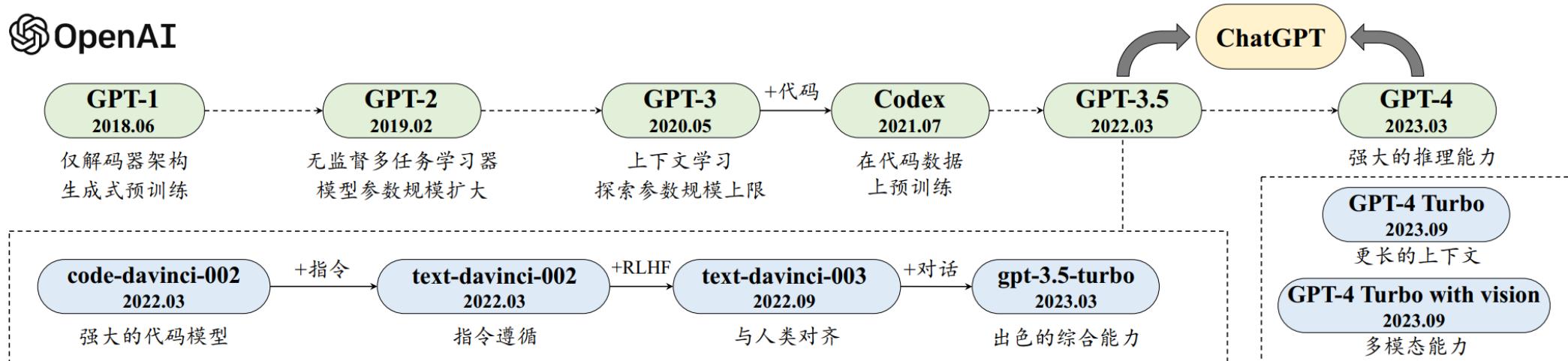
LLM (Large Language Model, 大型语言模型) 是指那些具有大量参数和庞大训练数据的神经网络语言模型。这些模型通常使用深度学习技术, 如Transformer架构, 以捕捉和理解自然语言中的复杂模式和语义关系。

由于大型语言模型具有更多的参数和更大的训练数据, 它们能够表现出更强的泛化能力, 从而在各种自然语言处理 (NLP) 任务中取得更好的性能。例如, GPT-3 (Generative Pre-trained Transformer 3) 就是一个典型的大型语言模型, 拥有超过1750亿个参数, 并在大量文本数据上进行预训练。

大型语言模型的应用范围广泛, 包括文本生成、机器翻译、摘要、问答、情感分析、命名实体识别等。然而, 这些模型也存在一些挑战, 如计算资源需求高、训练成本昂贵、容易产生偏见等。尽管如此, 大型语言模型在近年来已经取得了显著的成功, 极大地推动了自然语言处理领域的发展。

LLM (Large Language Model, 大型语言模型) : 网络结构庞大的语言模型, GPT就属于LLM

GPT & ChatGPT



chatgpt和gpt有什么区别

Chatbot GPT (如OpenAI的ChatGPT) 和GPT (Generative Pre-trained Transformer) 之间的主要区别在于它们的应用和优化目标。

GPT是一种基于Transformer架构的预训练生成式模型，它可以通过大量文本数据进行训练，以便生成连贯且相关的文本。GPT模型可以用于各种自然语言处理任务，如文本生成、摘要、翻译、问答等。

而ChatGPT则是在GPT基础上进行优化和调整的一种模型，专门用于聊天机器人和对话系统。它在训练过程中，会通过特定的对话数据集进行微调，以便更好地理解和生成与对话相关的内容。这使得ChatGPT更适合用于与用户进行自然、流畅且有趣的对话。

总之，GPT是一个通用的自然语言处理模型，而ChatGPT是针对聊天机器人和对话系统进行了优化的版本。

GPT : Generative Pre-trained Transformer
ChatGPT : 专注对话的、基于GPT模型的聊天机器人

/02 底层原理

Theory



在开始之前

先让大家见识下ChatGPT的力量！



AI渣男

夺命气泡音海王
图书馆约会实录



AI渣女

崇尚自由小仙女
酒吧怒甩保时男

GPT基础原理



01

实质功能

02

训练方式

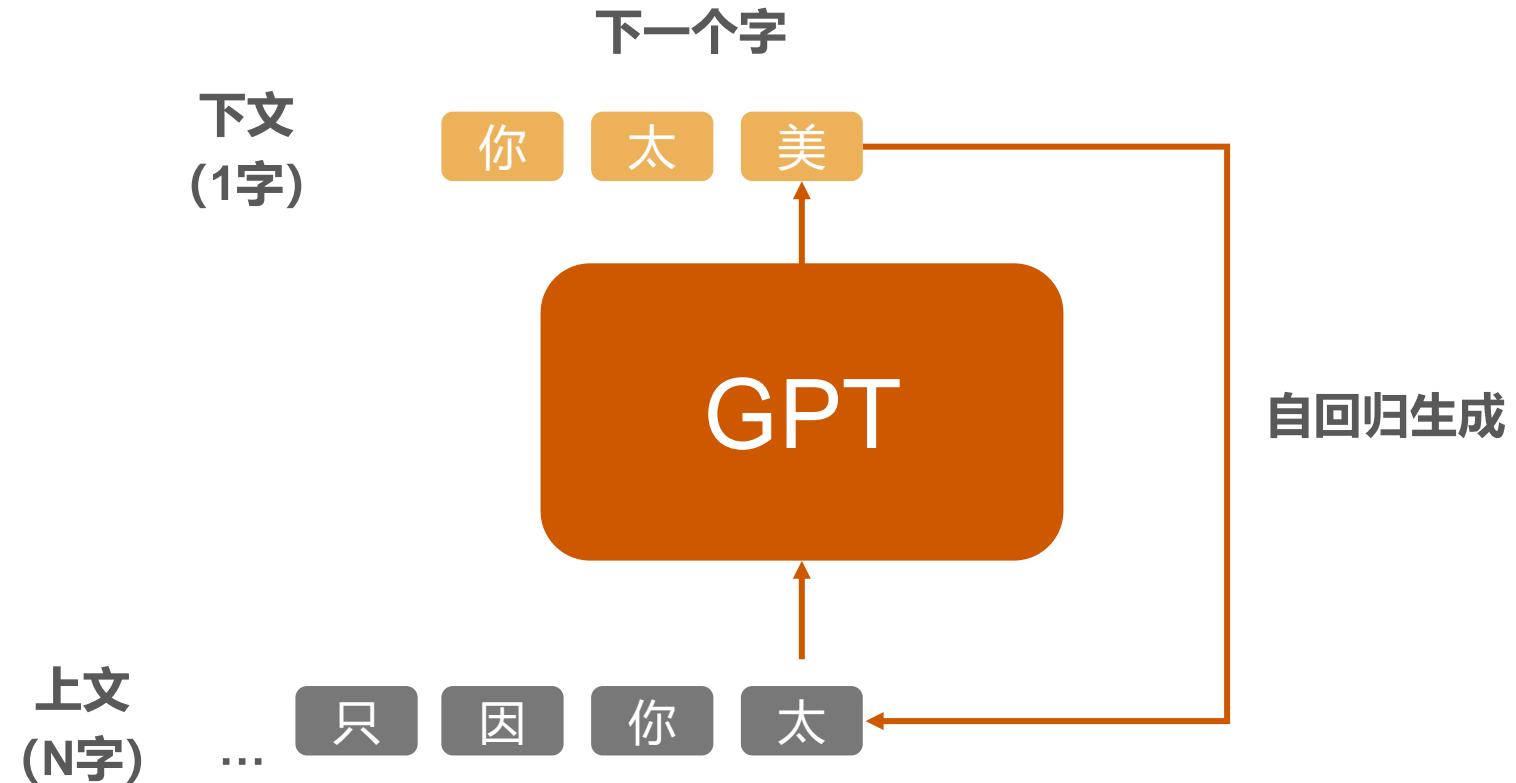
03

长处

04

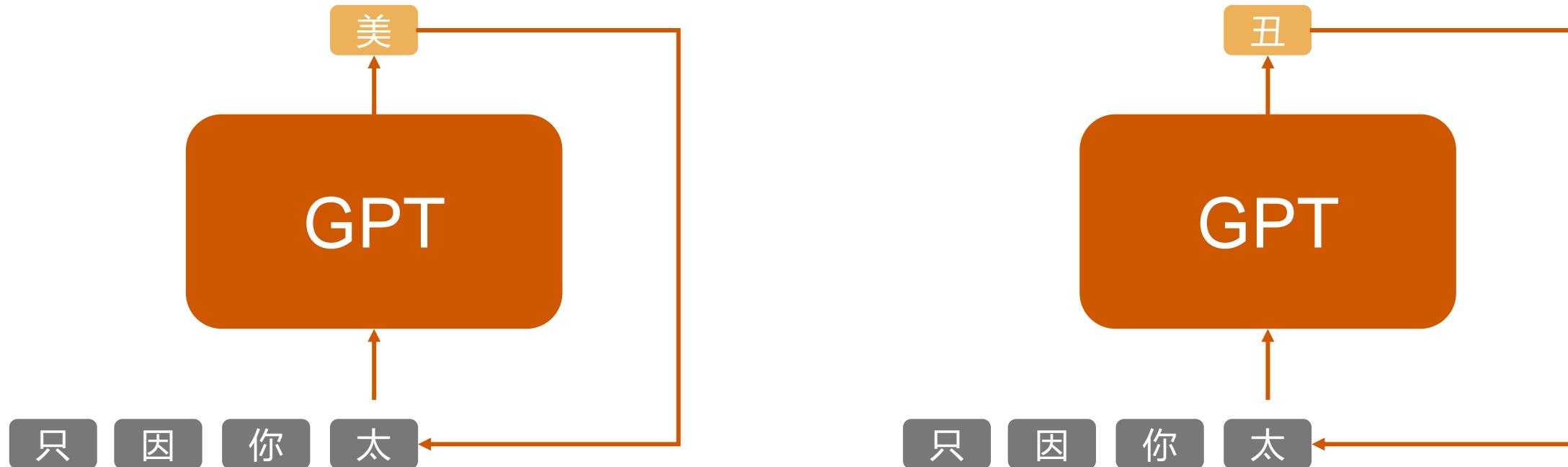
短板

实质功能



实际上，就是个“单字接龙”

训练方式



同一个上文，不一定生成同一个下文

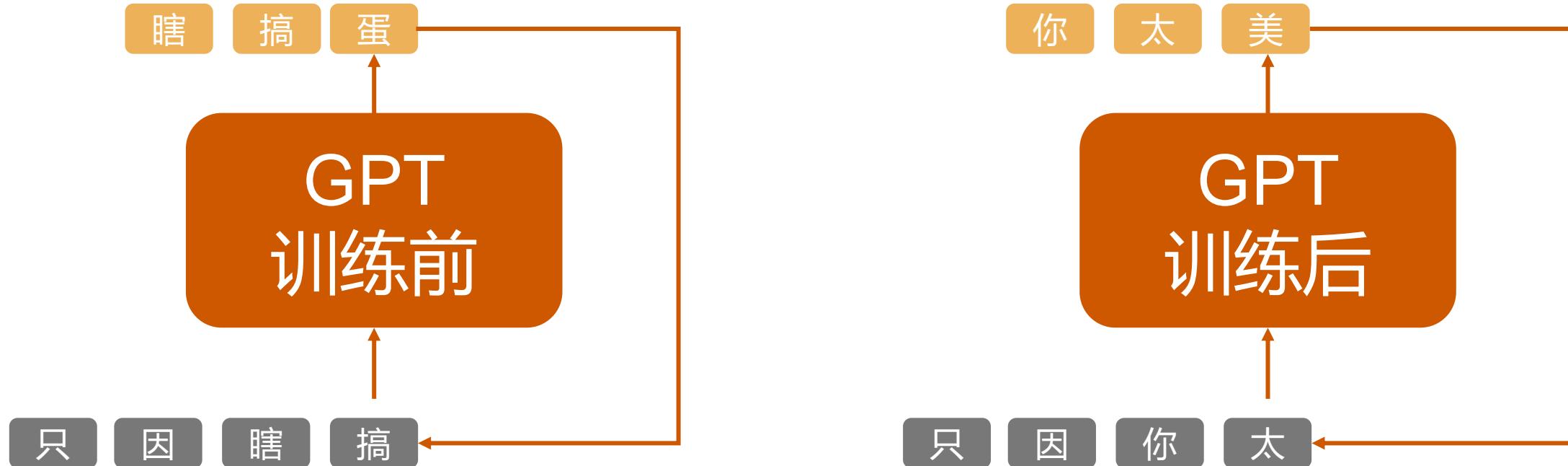
训练方式



问：如何训练？

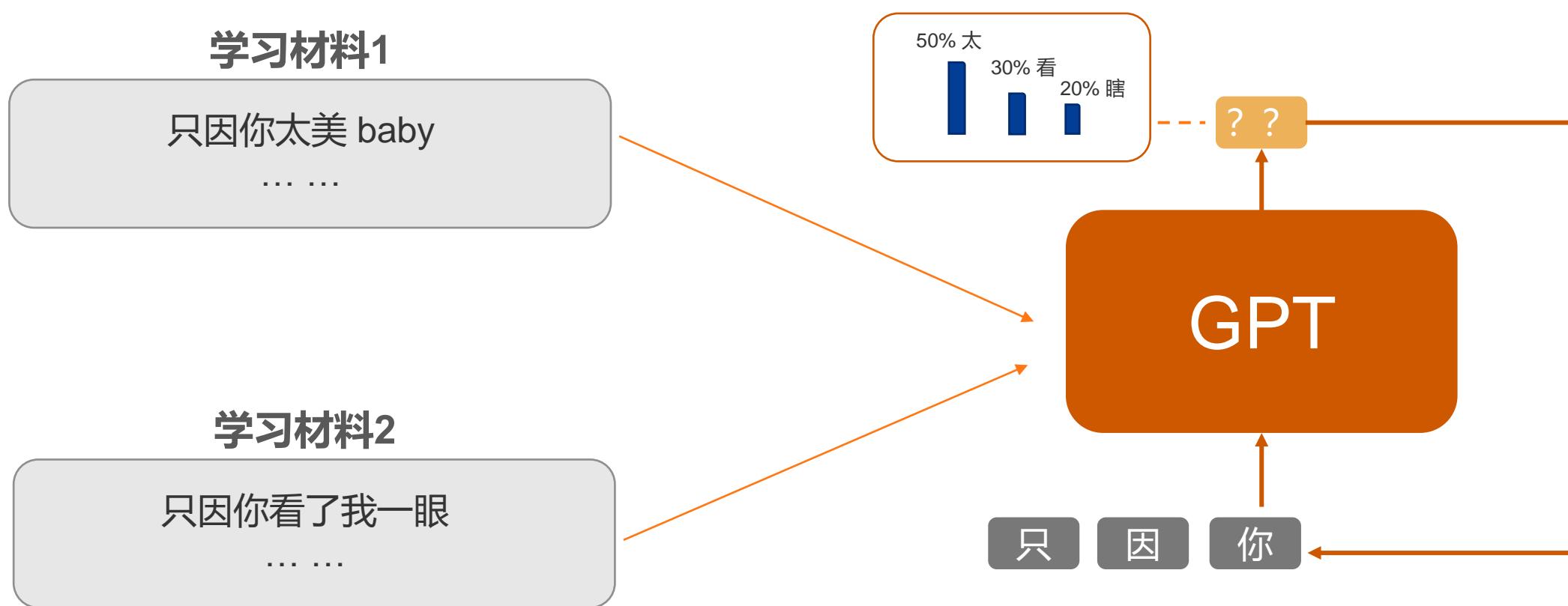
答：遵照学习材料，不断调整模型

训练方式



训练后，张嘴就能唱！

训练方式



问：训练材料，同上文不同下文，模型怎么办？

答：按照上文，按照下一个字概率分布抽取生成下文

训练方式

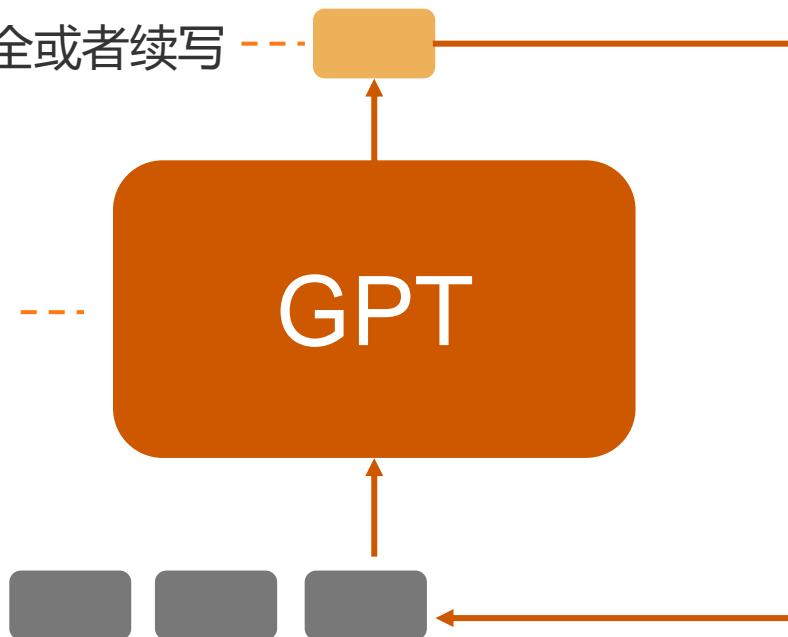
问答范例

问：我知道有首歌，歌词是鸡你太美了，请问歌名是什么？

答：这首是SWIN-S的《只因你太美》！

如何回答问题？

补全或者续写



问：单字接龙，怎么回答问题？

答：将问题-答案组成范例，让模型学习单字接龙

训练方式

问答范例1

问：我知道有首歌，歌词是鸡你太美了，请问歌名是什么？

答：这首是SWIN-S的《只因你太美》！

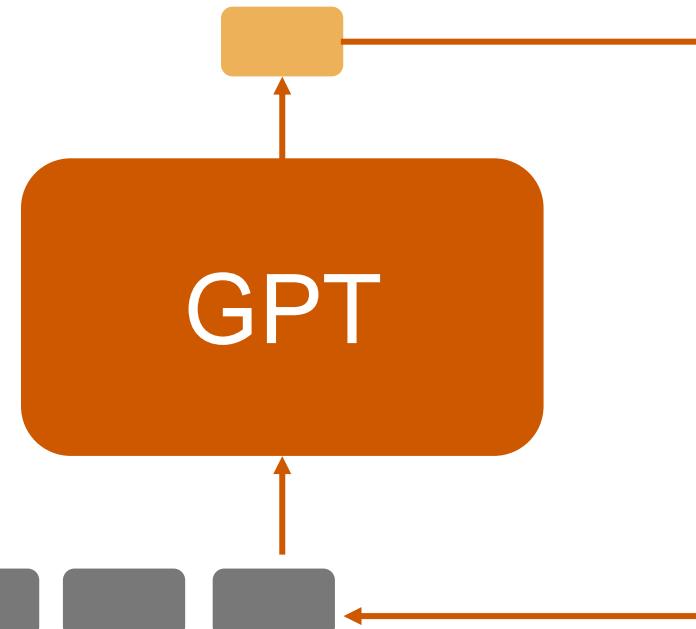
...

问答范例n

问：鸡你太美了，告诉我这是什么歌！

答：这首是SWIN-S的《只因你太美》！

学习！不是记忆



问：问答方式无穷无尽，都要学？

答：不需要，模型举一反三没有问题！

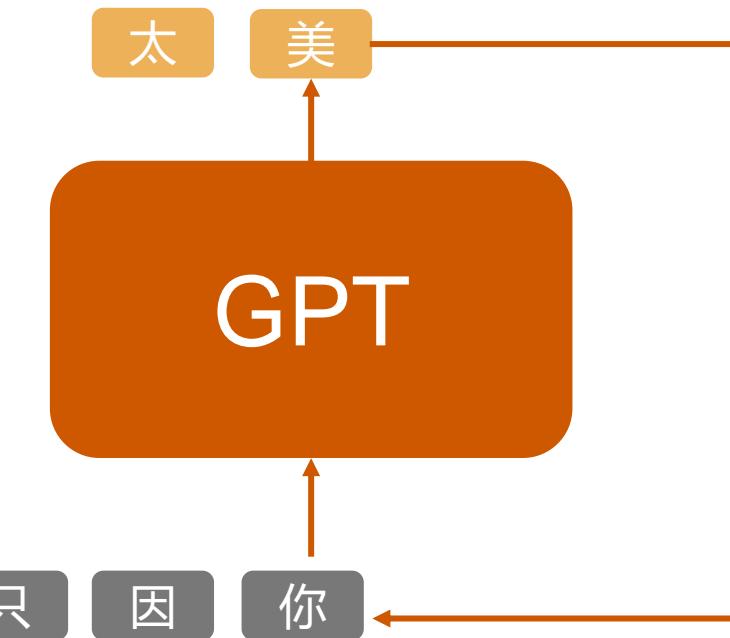
训练方式

不同的问答范例

问：我知道有首歌，歌词是鸡你太美了，请问歌名是什么？
答：这首是SWIN-S的《只因你太美》！

学习不同问法，
都是同一个答案

学习规律



01

没有见过的问题

02

模型学习规律

03

合理的回答

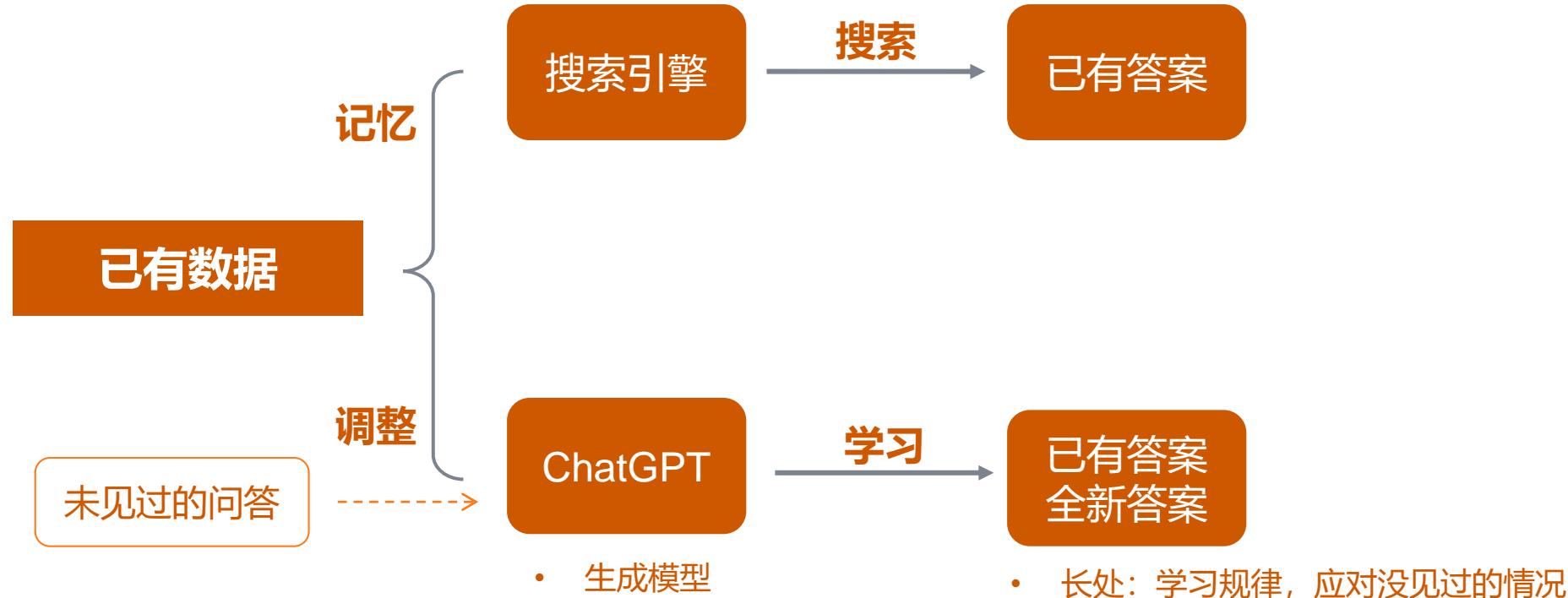
这就是泛化能力！

训练方式

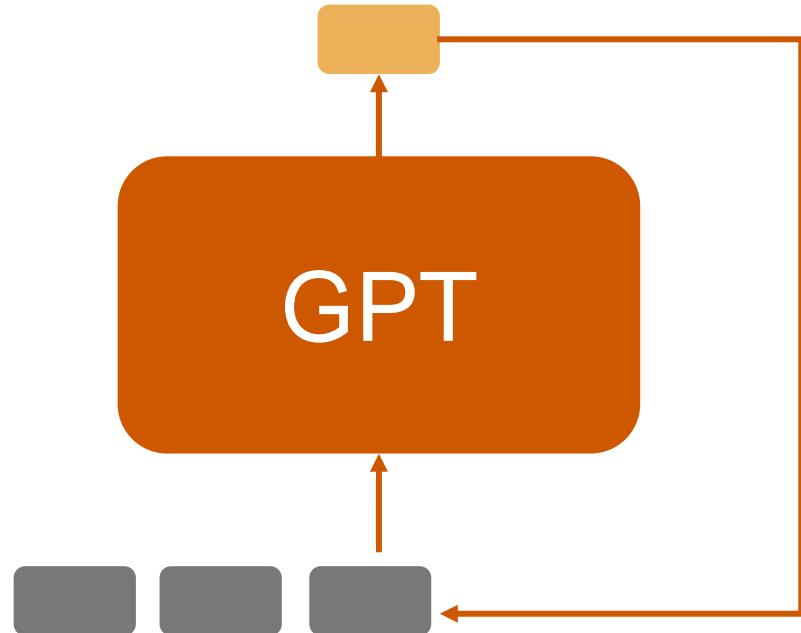


并不是根据既有的数据，进行搜索拼接，给出答案

训练方式



训练方式



训练好一个你不容易！

01

模型复杂，不知道学到了什么
决策缺乏可解释性，容易胡说八道

02

无法随意增删改查记录，只能不断
训练调整

03

模型评估困难，只能通过提问来评
估

04

高度依赖学习材料的数量、质量
否则无法学习到正确规律

问：灰原既然不喜欢
小奶狗，为何还是和
她女朋友小兰养了一
只，并且改名叫新一？

训练方式

培训有九 不知苦当乐乐乐

ON

灰原既然不喜欢小奶狗，为何还是和她女朋友小兰养了一只，并且改名叫新一？

< 3/3 >



首先，需要澄清的是，灰原哀并没有表现出不喜欢小狗的性格特征，这只是在某些网络段子和模因中流传的一种梗，没有现实依据。



至于灰原哀和小兰为什么养了一只小狗并改名叫做新一，这是因为在漫画《名侦探柯南》中，新一是小兰的青梅竹马和男友，而在某一段剧情中，新一不慎被组织下毒而变成了小孩，这时灰原哀暗恋新一的心情和对他的关心与保护更加强烈，于是她和小兰为了纪念新一，决定养一只小狗并取名为新一，以表达对他的爱和怀念。

因此，这只小狗的名字“新一”并不是因为灰原哀不喜欢小狗，而是出于对剧情中人物的故事情感的表达。

灰原既然不喜欢小狗，为何还是和她女朋友小兰养了一只，并且改名叫新一？

训练好一个你不容易！

04

高度依赖学习材料的数量、质量
否则无法学习到正确规律

训练方式

Model: GPT-4

ON

灰原既然不喜欢小奶狗，为何还是和她女朋友小兰养了一只，并且改名叫新一？



虽然我不是具体的人物，但我可以尝试从一般的角度分析这个问题。



首先，你提到的人物可能是柯南动画中的灰原哀和毛利兰。然而，这里有一点需要澄清的是，灰原哀和毛利兰并不是女朋友关系，而是作为侦探柯南故事中的重要角色，他们之间的关系并没有发展到恋爱关系。实际上，灰原哀喜欢的是工藤新一（柯南），而毛利兰是工藤新一的青梅竹马和女朋友。

关于“小奶狗”，这个词语通常用来形容粘人、温柔、可爱等特点的年轻男性。然而，在这个情境下，你可能是指一只实际的小狗。假设灰原哀和毛利兰真的一起养了一只小狗，并将其命名为“新一”。

即使灰原哀可能不喜欢小狗，但她为何还是和毛利兰养了一只呢？这可能有以下几个原因：

1. 灰原哀可能是为了迁就毛利兰，与她增进友谊和关系。毛利兰可能非常喜欢小狗，而灰原哀愿意为了她的朋友而做出一些牺牲。
2. 灰原哀可能想通过养狗这个经历来学习新的东西，改变自己的观念和看法。有时候，人们会尝试新事物，以拓展自己的生活经历和兴趣爱好。
3. 将狗狗命名为“新一”可能是一种寓意。灰原哀可能希望通过这种方式向工藤新一表达自己的感情，让他知道她一直在想念他。

需要注意的是，以上分析均是基于假设和推测。实际情况可能因故事情节和角色设定而有所不同。

：灰原既然不喜欢奶狗，为何还是和女朋友小兰养了一，并且改名叫新一？

训练好一

总结

- ChatGPT就是个单字接龙模型
- 长文是利用自回归来实现的
- 需要提前训练来生成想要的回答
- 训练方法是利用问答范例来学习
- 大量训练可以让它学习到规律，举一反三
- 缺点是容易胡说八道，不知道它学会了啥，高度依赖好的学习材料

/02 训练过程

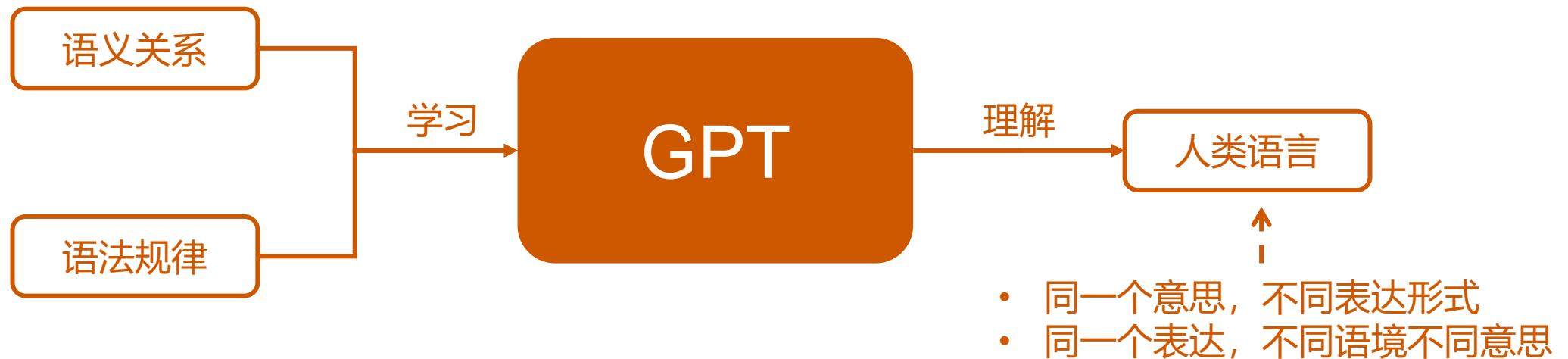
Training process



训练过程

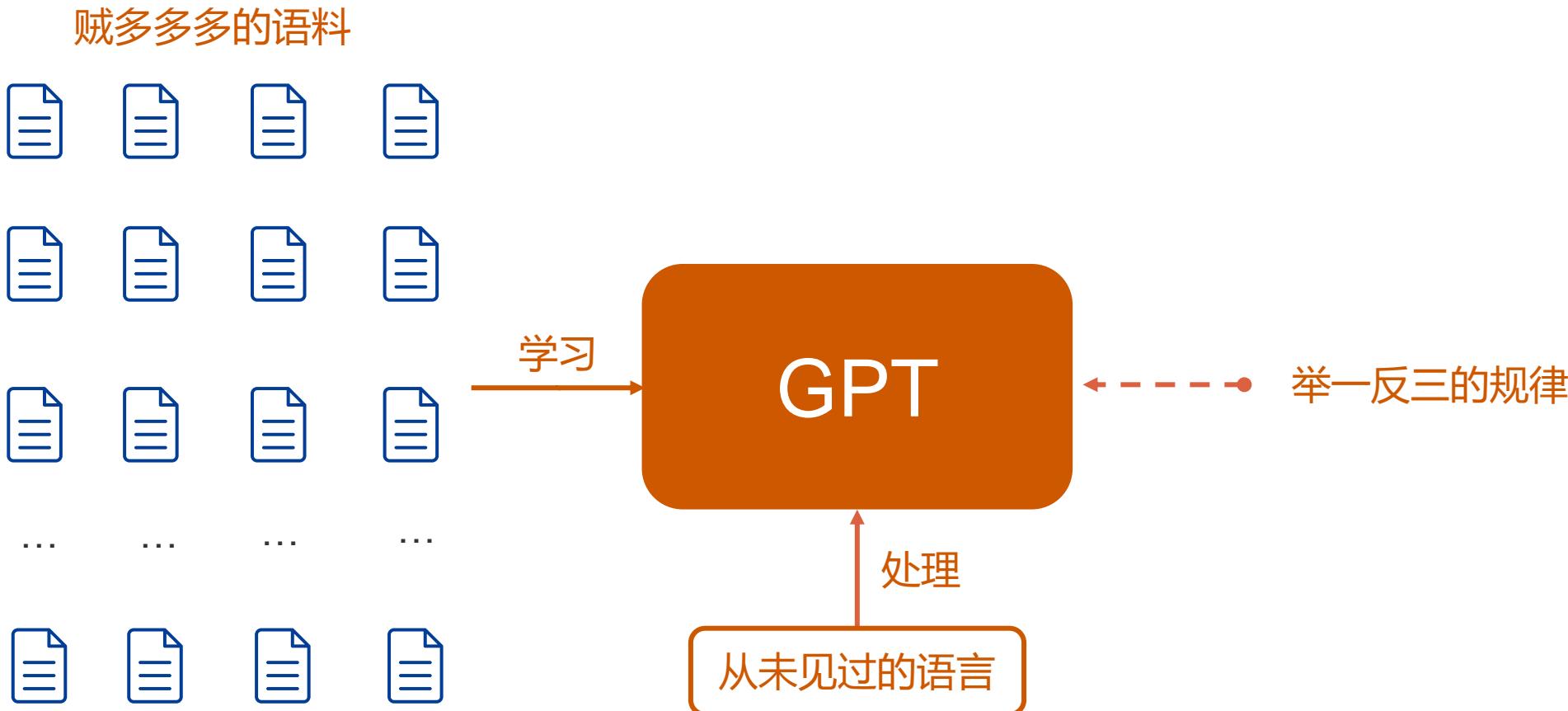


训练过程

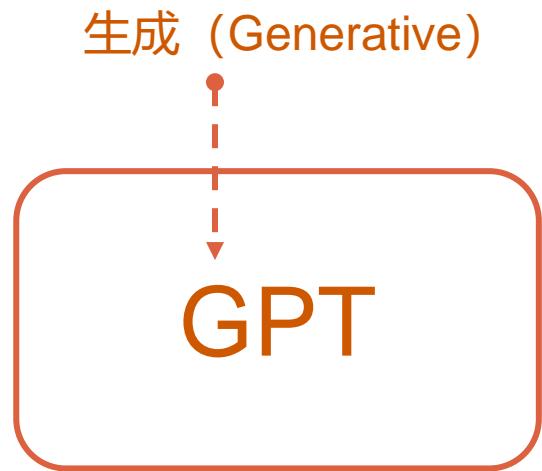


人类语言很复杂, 模型需要学习
哪些表达是同一个意思

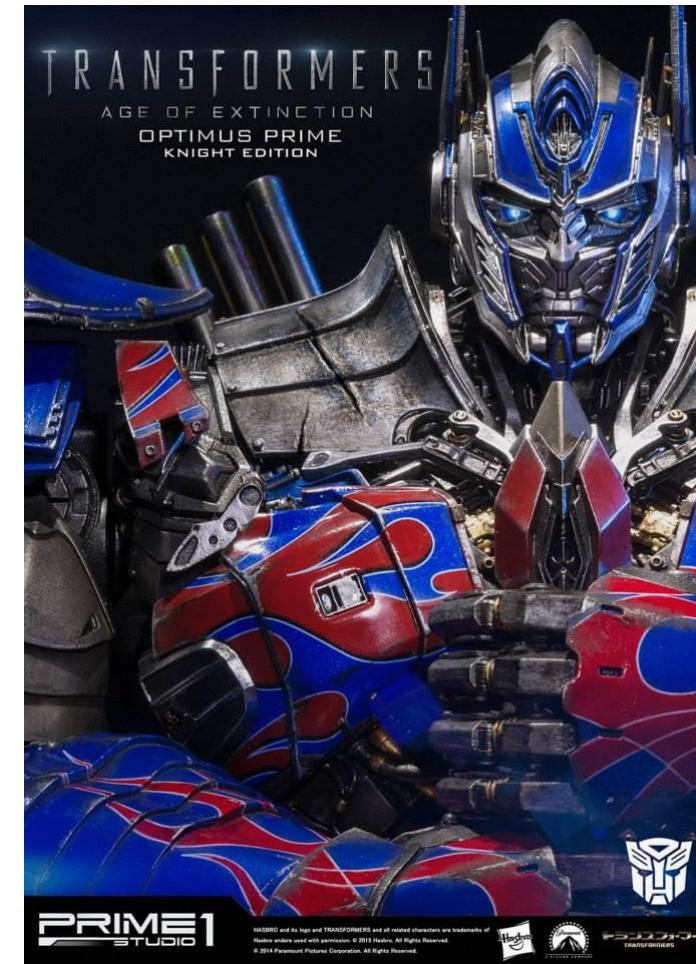
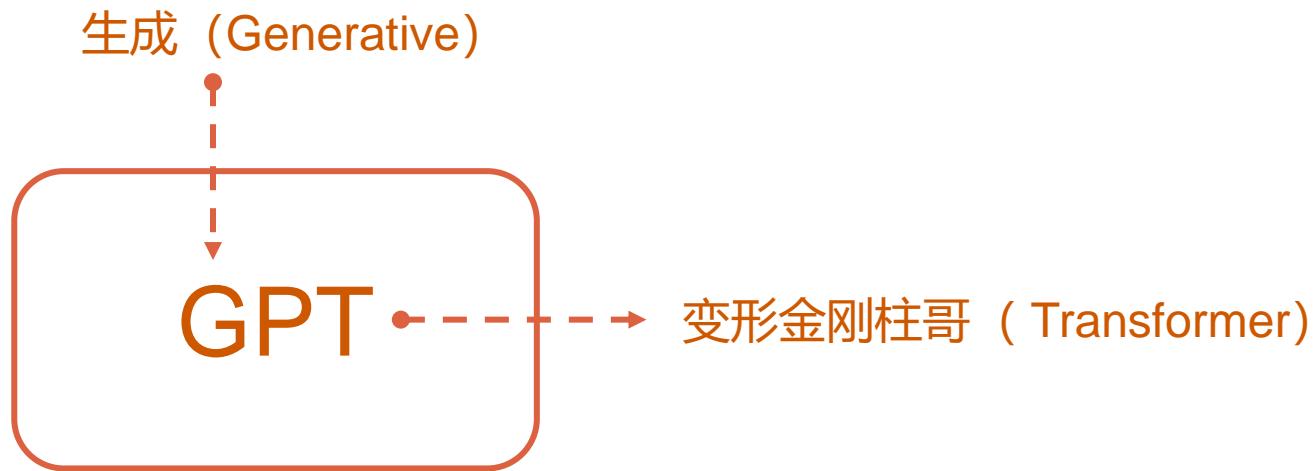
训练过程



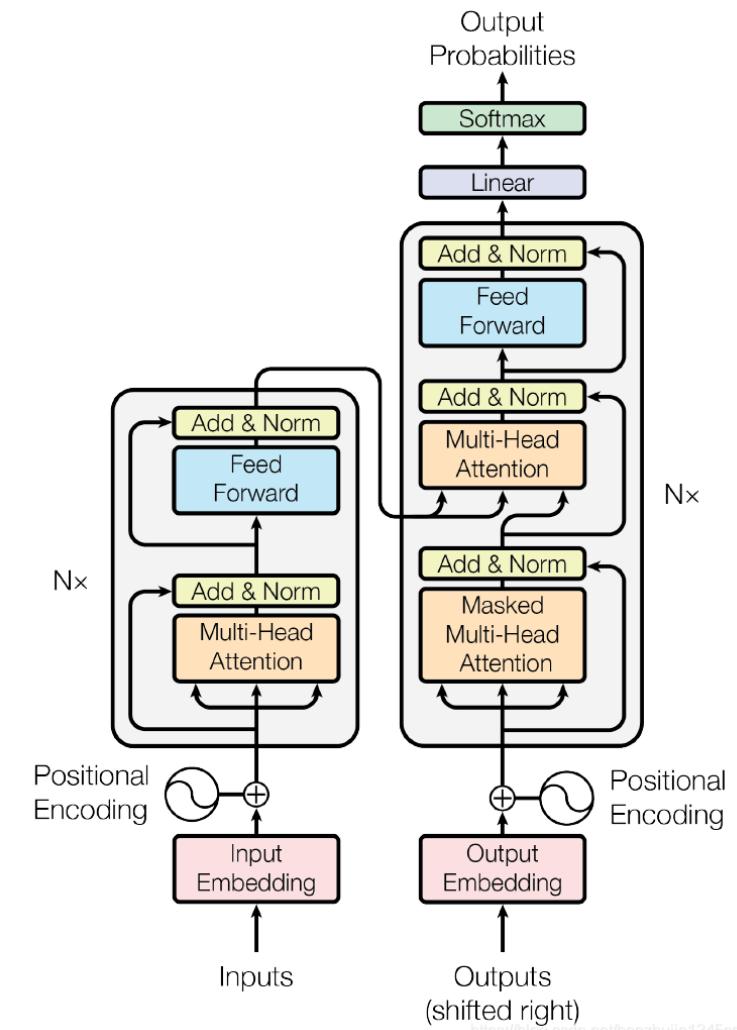
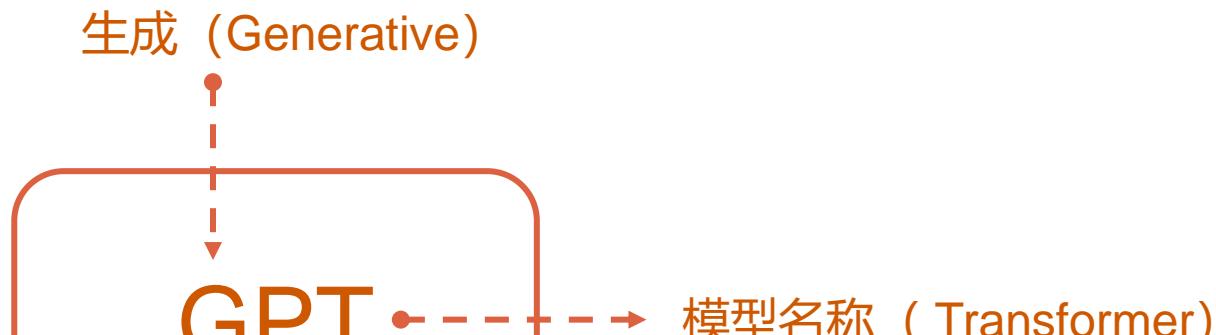
训练过程



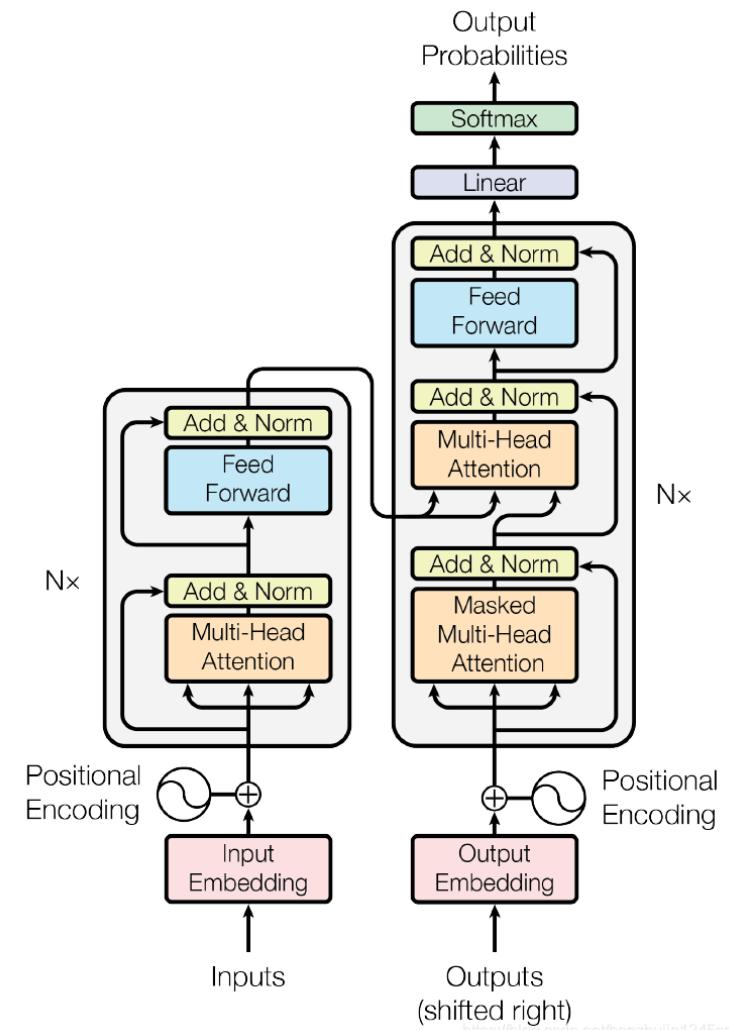
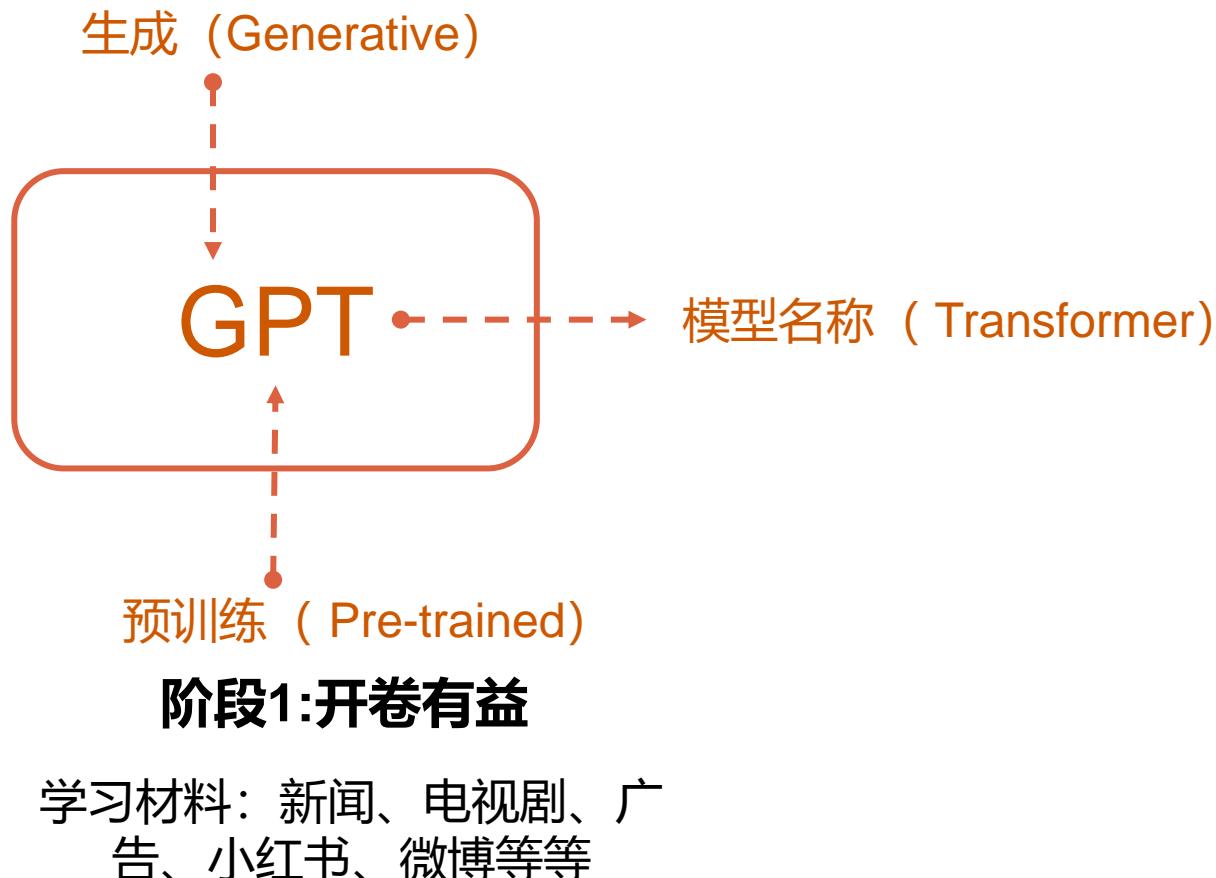
训练过程



训练过程



训练过程



<https://blog.csdn.net/benzhujie1245com>

训练过程

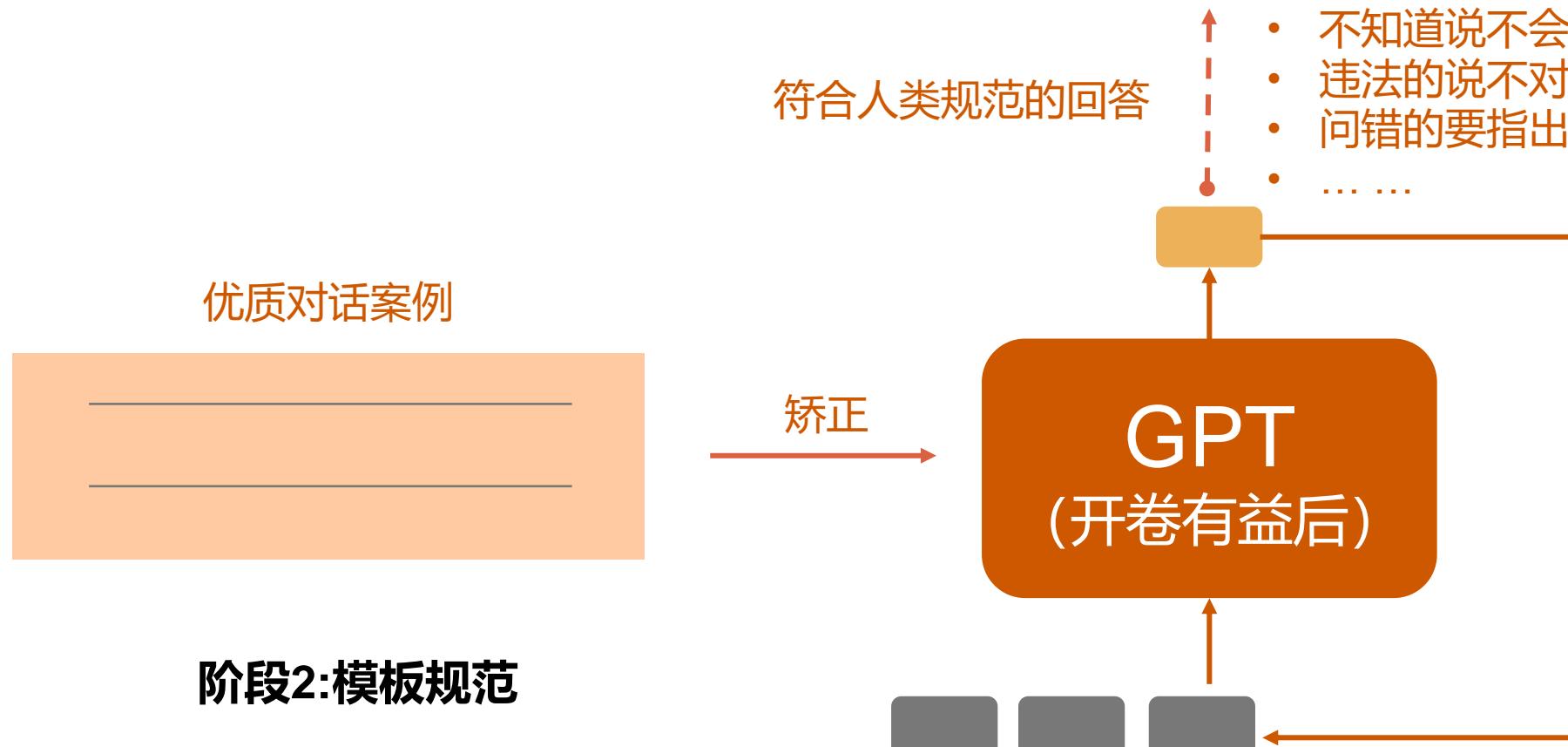
模型名称	学习资料容量	模型参数	时间
GPT-1	5GB	1.17亿	2018.6
GPT-2	40GB	15亿	2019.2
GPT-3	45TB	1750亿	2020.5
GPT-4	1PB	1.6万亿	2023.3



问：开卷多少才够？

答：More! More! More!

训练过程

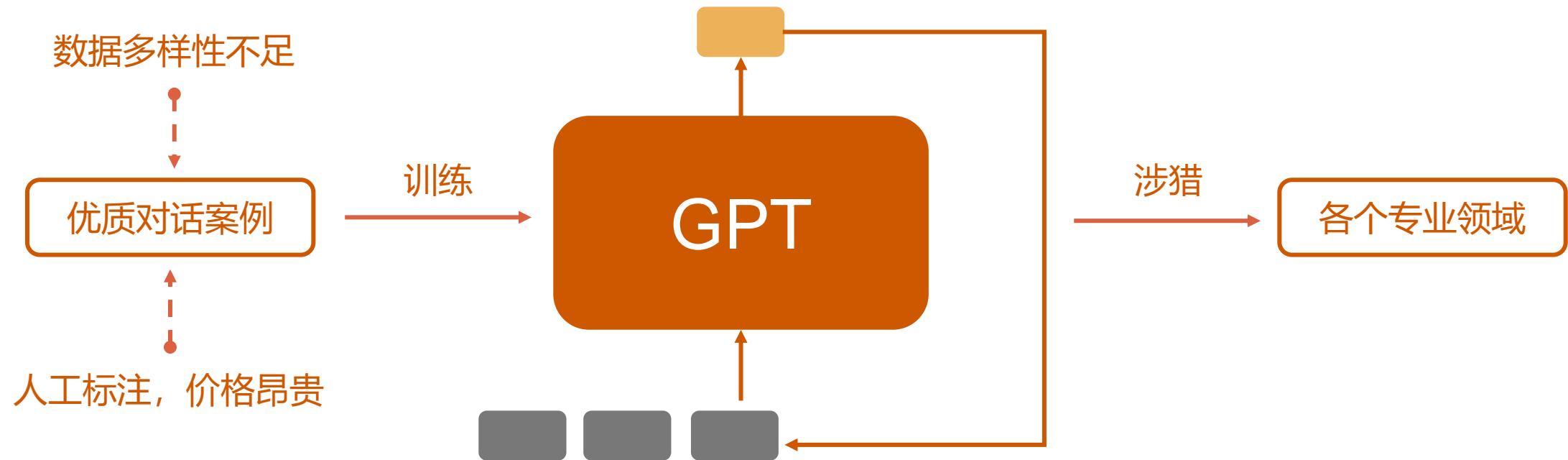


阶段2:模板规范

问：不能指挥它咋办？

答：我有模板教育它！

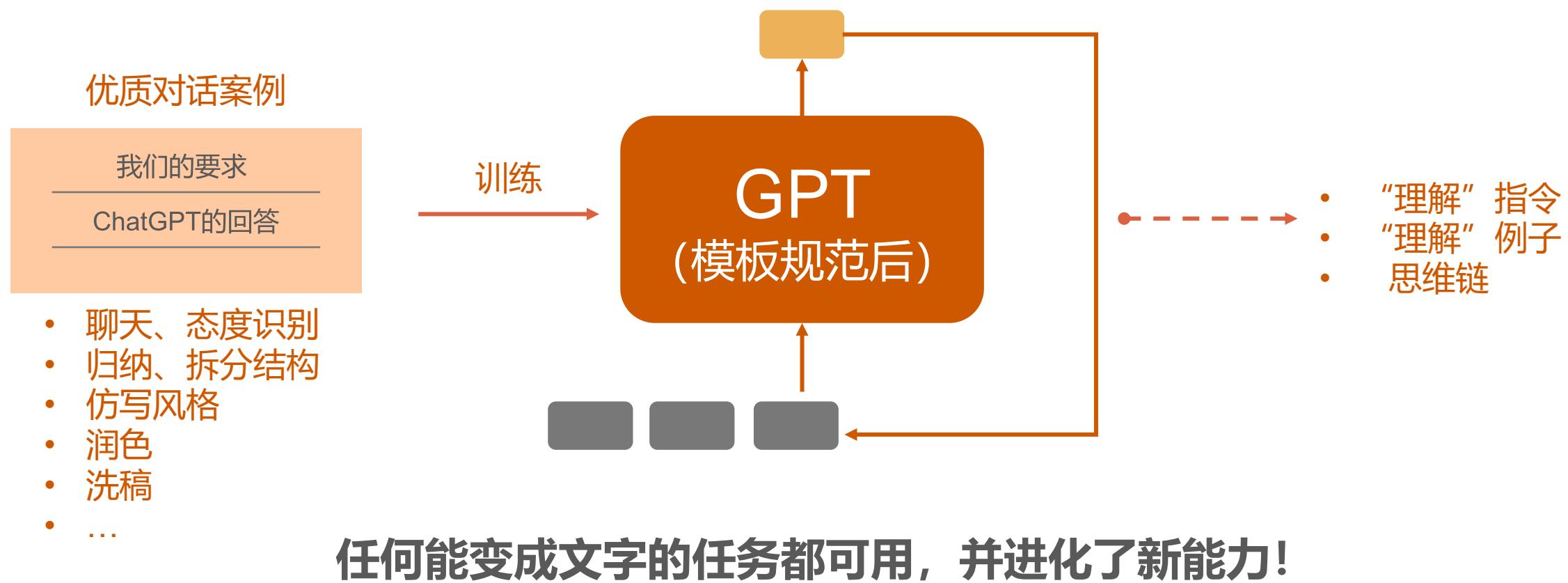
训练过程



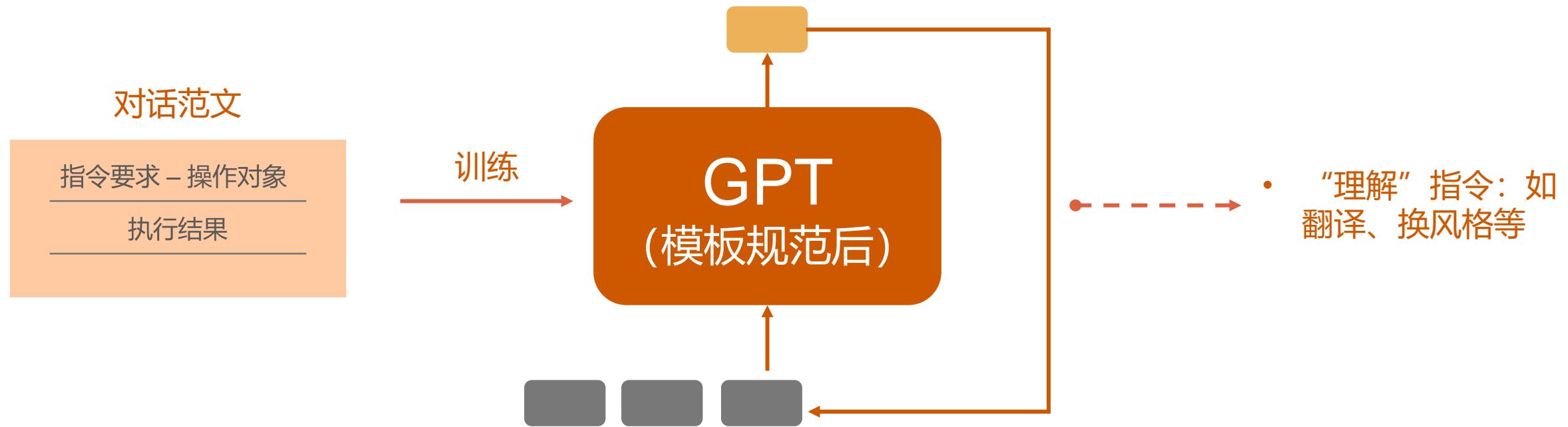
问：开始为啥不这么干？

答：优质范例太少，只能用来矫正，不能用来预训练

强大能力

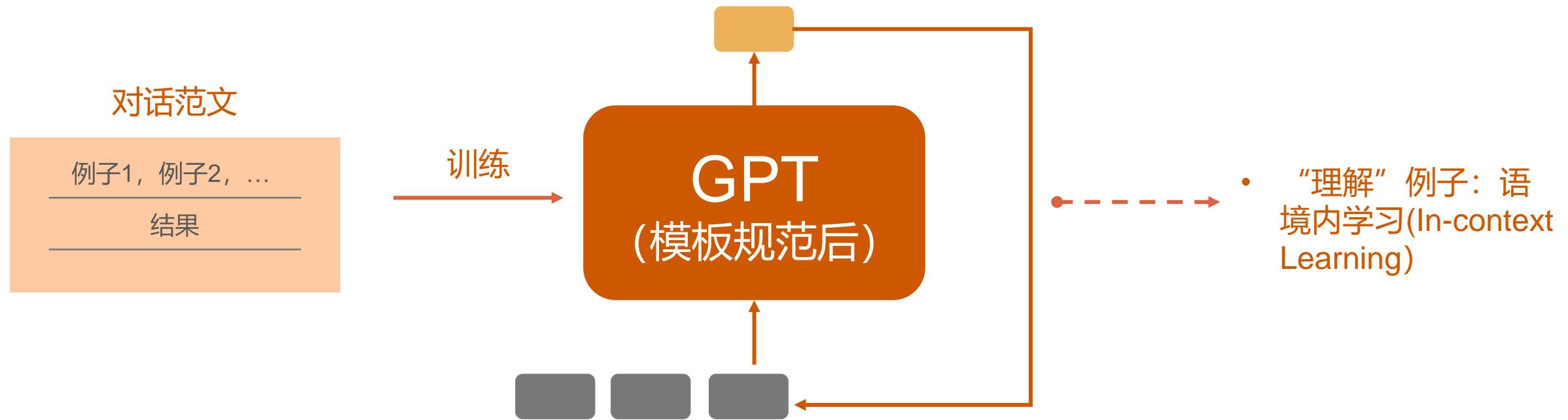


训练过程



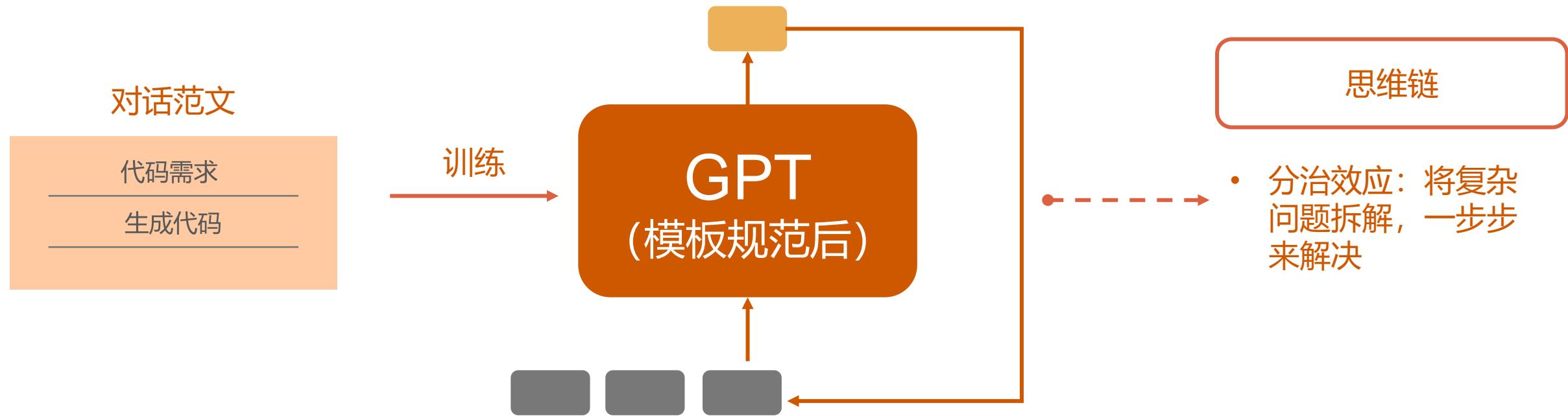
利用用户抽象描述，给出结果

训练过程



通过举例子，让他明白你要干啥

训练过程



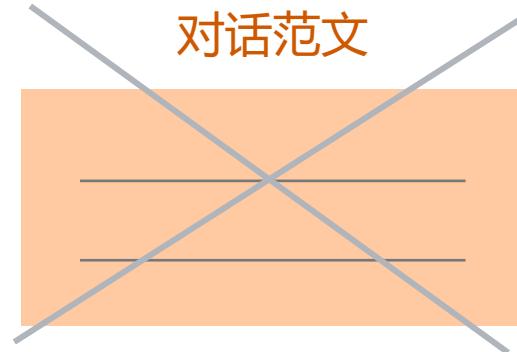
答不对问题，让他分步思考就可提升答对概率

训练过程

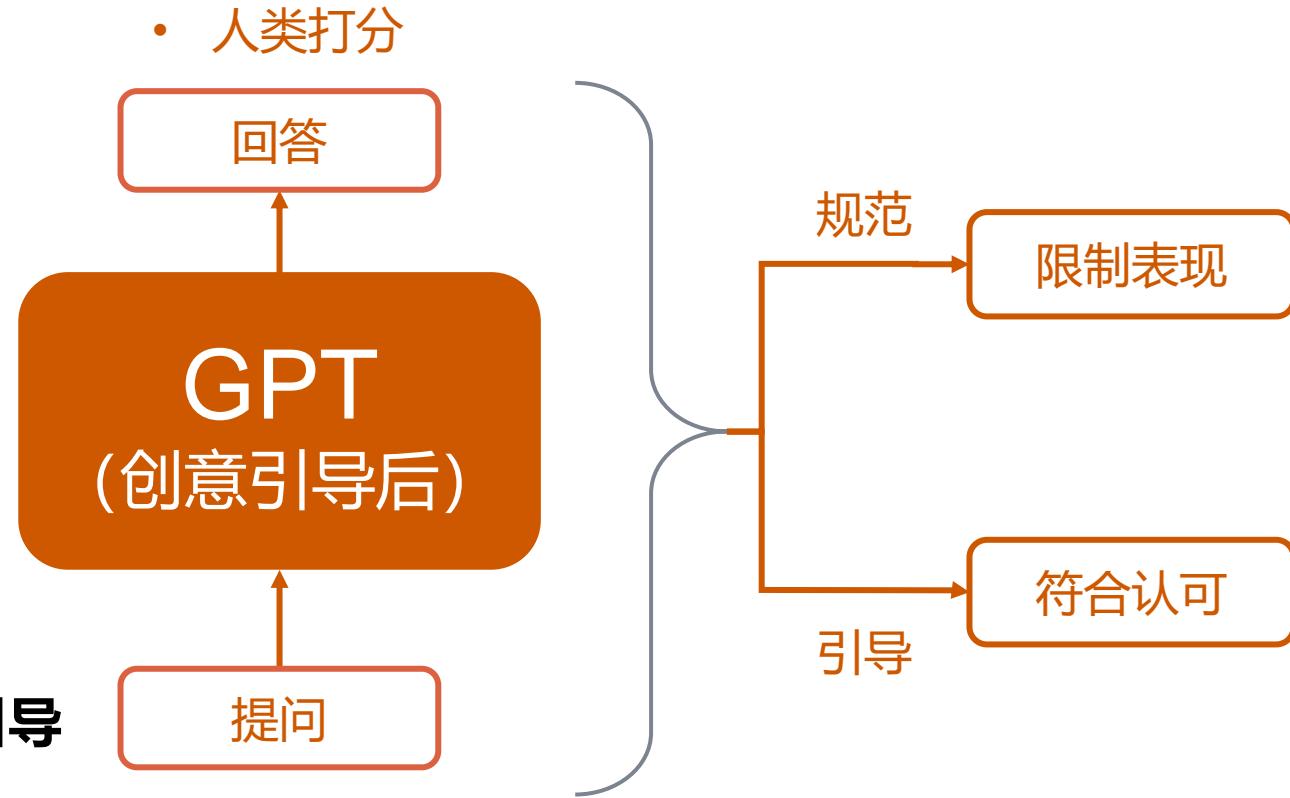


- 需求：
 - 符合人类价值取向与对话模式
 - 提高回答创新性

训练过程

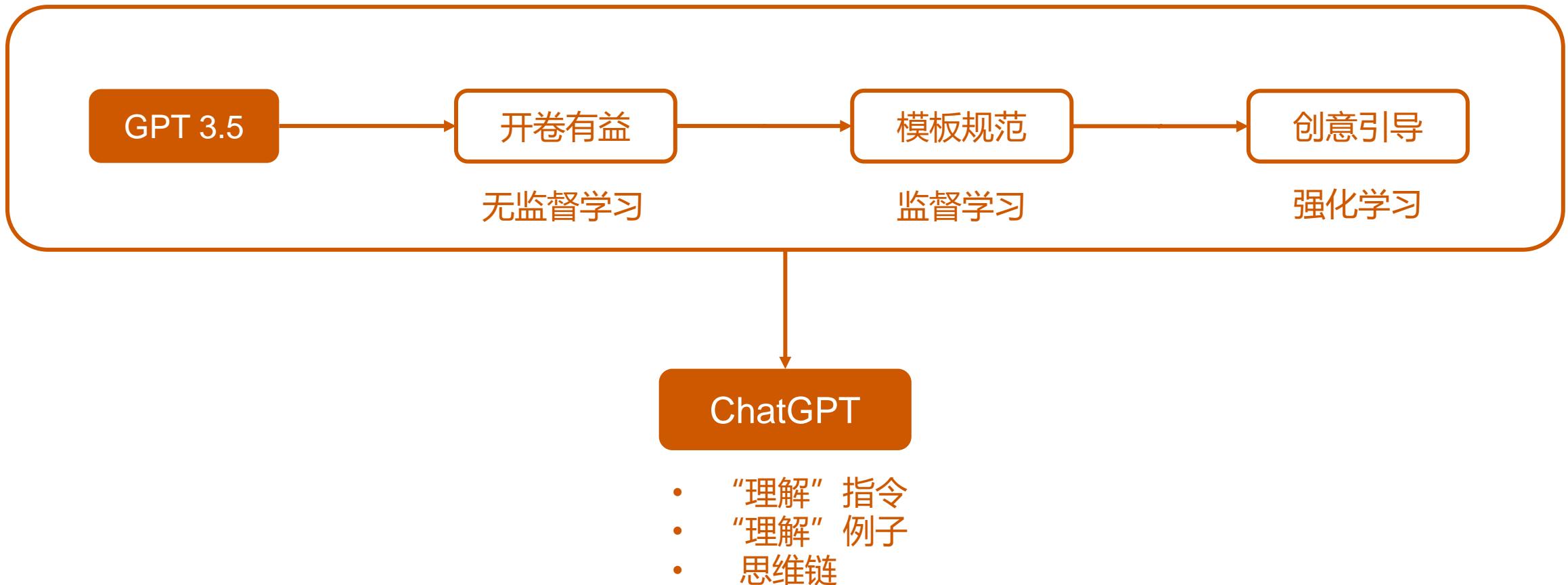


阶段3:创意引导



抛弃范例，直接提问，奖罚分明

训练过程



训练过程

ChatGPT的三个训练阶段

- “开卷有益”：海量数据训练单字接龙，让模型“博学”
- “模板规范”：优质案例进行矫正，让模型“博学+懂规矩”
- “创意引导”：人类评分回答优劣，让模型“博学+懂规矩+创新”

/03 未来影响

Future impacts



争议不断的ChatGPT

反对派

- ChatGPT根本就没那么神，不好用
- 并没有什么技术创新！
- 这东西侵犯隐私！有风险！
- 产品形态好差！就个破网页！



支持派

- 跨时代的产品！技术奇点来临！
- 彻底改变工作流！
- 第四次工业革命！
- 堪比互联网的最大风口！



未来影响

思考

- 这是否就是个炒作?
- 这会影响我工地搬砖么?
- 这与我人生有什么关系?

ChatGPT
带来了什么?



实际意义

点火实验、行业应用



工作模式影响

各行各业的替代性



安全隐患

科技伦理与机密泄漏

实际意义



一次伟大的点火实验： 证明LLM的可行性



地下城的门票

国家、企业、资本都会立马加大投入，不会错过这次机会！



雨后春笋

LLM相关应用会迎来大爆发，这才是改变社会方向的真正推力

这些【尚未出现、不断改良的LLM】对社会有什么影响？

实际意义

ChatGPT：俺是语言大师.



疑问：就这？这有啥用？

- 【知识体系】是人类发展的重要基础
- ChatGPT在精通语言的基础上，学习了**人类从古至今的【知识】**

实际意义

语言记录知识 从过去到未来的传承.

01



人类的脆弱

打不过猩猩、跑不过豹子、游不过鲨鱼、吃不过猪猪



知识的传承

创造知识、继承知识、应用知识，
让人类走得更远

02



语言的需求

知识的三个过程，全靠语言进行，
对语言处理的需求很大

03

实际意义

NLP让机器理解人类语言，协助人类处理繁杂的语言类工作


**自然语言处理技术
(NLP) .**
处理速度快、工作记忆大、知识覆盖面广
7*24小时工作、没有情绪

LLM核心技术就是NLP



计算机、互联网诞生提高效率.

从过往的口头、纸质记录，科技的
普及大大加快了协作效率

爆炸式的信息增长.

但同时也带来了巨大增长的信息量，
每天你看的视频、推荐、新闻等等
等

繁重的信息处理事务.

处理成本提升、效率降低，银行、
企业、学校等是重灾区

实际意义

LLM是魔法百科全书！机器届的王语嫣

合理使用模型，能够让一个普通人快速地获取各行各业的知识！
虽然他不能对应实体，但是作为工具不影响



LLM改善的是创造知识、继承知识、应用知识时的语言处理效率

01

影响范围：“全社会、全世界，跟电脑
\互联网的影响一样

02

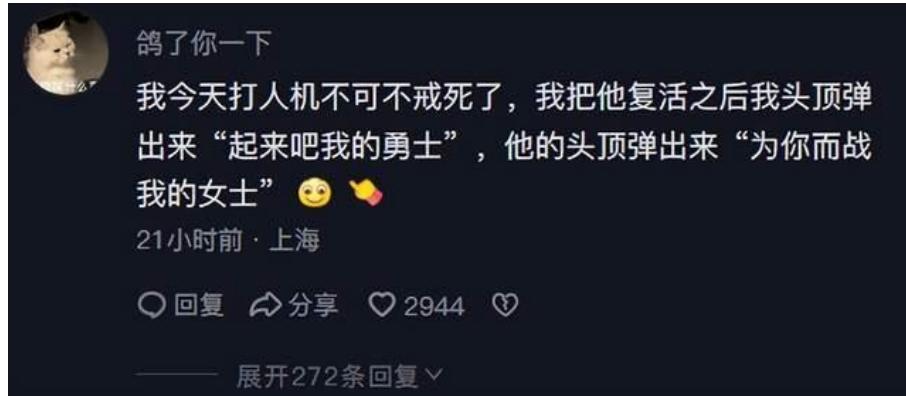
应用结合

LLM的应用深入各个行业.

利用LLM模型作为大脑，结合其他AIGC工具，实现工作效率大提升！



应用结合 – 游戏NPC



利用LLM丰富NPC对话功能，利用AI打造话题

应用结合 - 美术



利用ChatGPT获取设计思路



生成Stable Diffusion关键词



生成对应效果图

周 互联网产品弹窗怎么设计看起来让用户觉得更高级

以下是在设计互联网产品弹窗时让其看起来更高级的建议：

1. 简洁明了：弹窗的设计应该简洁明了，不要有过多的花哨，因为高级感往往来自于简约和精致的设计。
2. 颜色搭配：颜色搭配是非常重要的，可以选择高级质感的颜色，比如黑色、白色、金色、银色、灰色等。颜色搭配要注意搭配的协调性和整体感。
3. 字体设计：在弹窗的设计中，字体设计也非常重要，可以选择比较高级的字体，比如宋体、黑体、楷体等，字体的大小和间距也要注意调整。
4. 布局设计：弹窗的布局设计要注意合理和美观，可以采用对称的设计方式，让弹窗看起来更加整洁和美观。
5. 动态效果：在弹窗的设计中，可以添加一些动态效果，比如淡入淡出、弹跳、旋转等，这些动态效果可以让弹窗看起来更加生动和有趣。
6. 图形元素：可以在弹窗中添加一些优美的图形元素，比如线条、图案、图标等，这些图形元素可以让弹窗看起来更加高级和精致。

总之，设计高级感的互联网产品弹窗需要简约、精致、高质量的设计，通过优秀的设计和精细的处理来展现其高级感。同时，还要注意颜色搭配、布局设计、字体设计等方面细节的处理。

运用ChatGPT获取不同类型弹窗思路

转化成对应的英文后，关键词如下：
Prompt: Shining gemstone, high class, noble sense, glory sense, 3D, hard, diamond, Octane Render
Negative prompt: Low saturation, deformity, sketch, blur



结合LLM与AIGC，生成活动弹窗

应用结合 - 美术



利用ChatGPT获取设计思路



预言家再中心操控着魔法球，两边的魔法狼群围绕着她的感觉。然后中心散发光芒，头发飞舞的画面以这个为核心 写个画面关键词



画面关键词：预言家、中心、魔法球、魔法狼群、感觉、光芒、头发飞舞。

运用ChatGPT获取画面关键词



生成Midjourney 关键词

其中客观描述如下：

A beautiful seer is holding a glowing magic ball in the center of the picture, with orange-red hair fluttering around her, surrounded by a group of transparent blue magic wolf heads

插画风格如下：

Illustration, Drawing, Hand-Drawn, Isometric, Saturated, Split-Complementary-Colors, 2-Dimensional, Rays of Shimmering Light



生成效果图



人工修改，生成终稿



结合LLM与AIGC，生成KV

应用结合 - 美术



AR演播厅



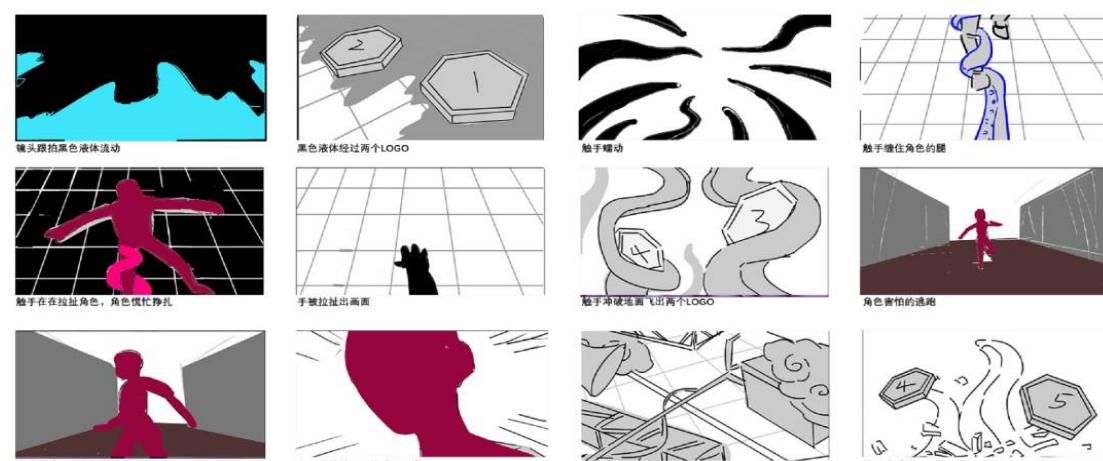
奖杯设计



设计师手绘稿

AI辅助

IP设计



动画分镜设计

应用结合 - 开发工具



应用结合 - 办公软件



Microsoft Copilot, 一种全新的
的工作方式

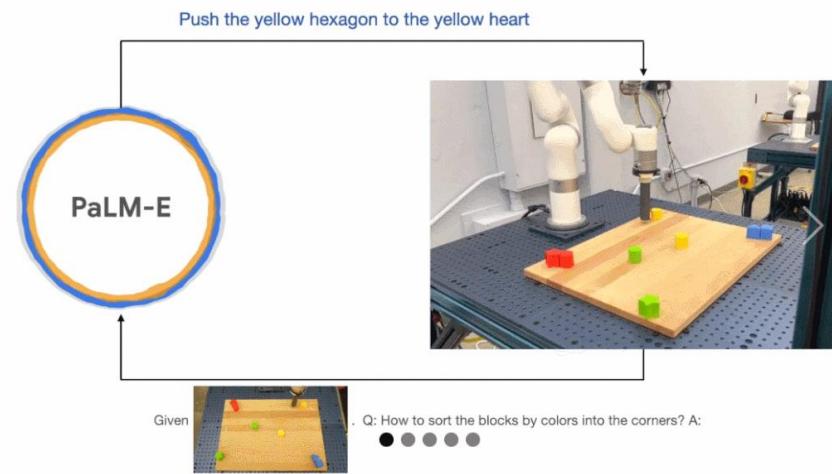
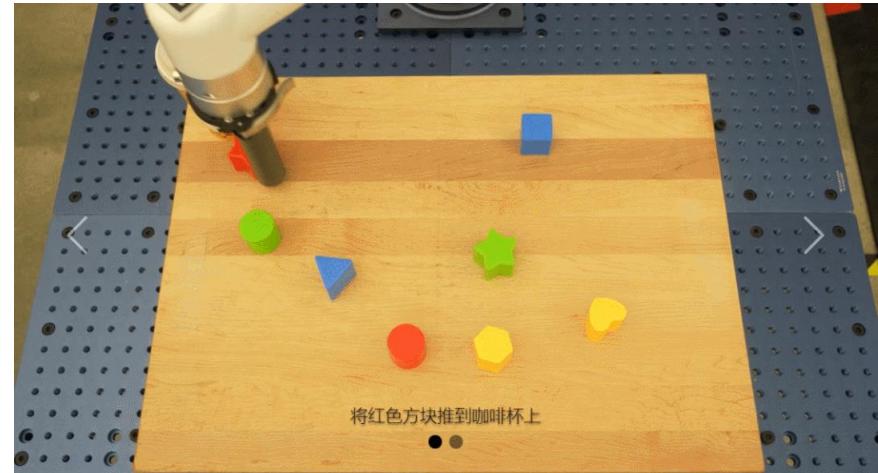
Copilot 可使用 AI 功能帮助你实现前所未有的目标。

微软365 Copilot 全家桶

应用结合 - 其他组合应用



这才是没有译制片腔调的翻译！



PaLM-E 加持的机器人

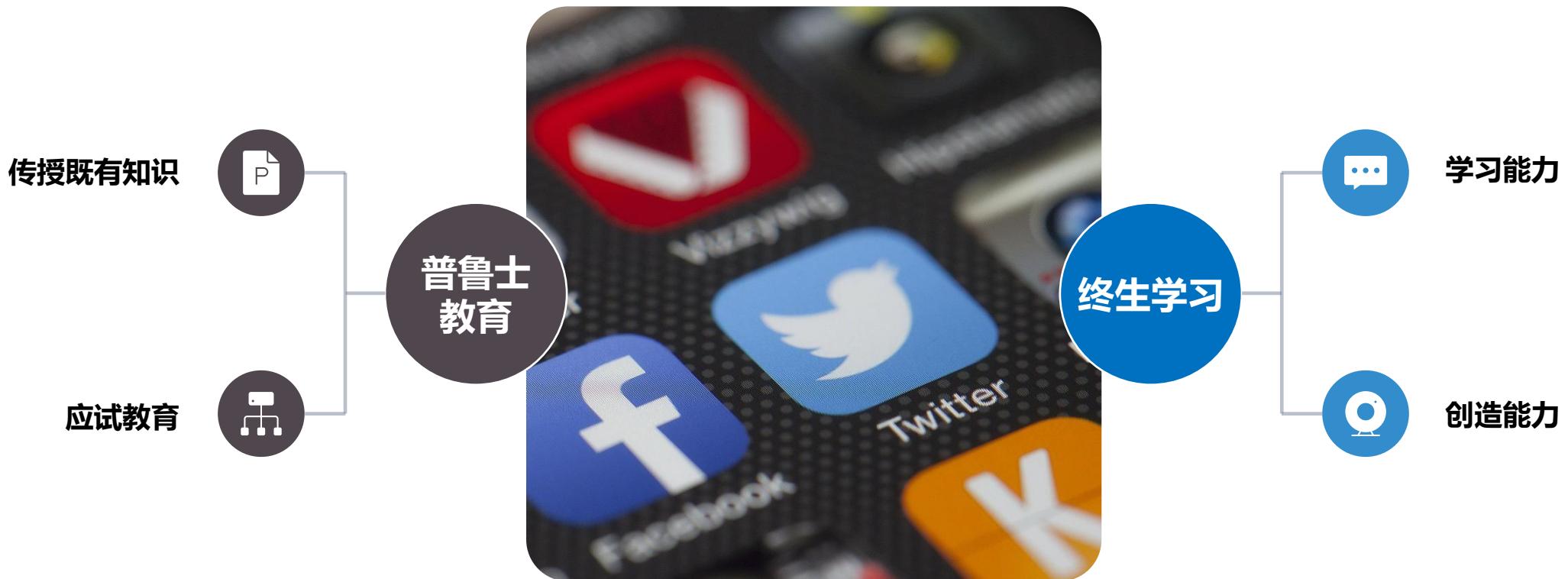
教育模式影响

现在的工作模式，5-10年后各位能否找到工作？

花费半生学习的知识，LLM都会，怎么办？

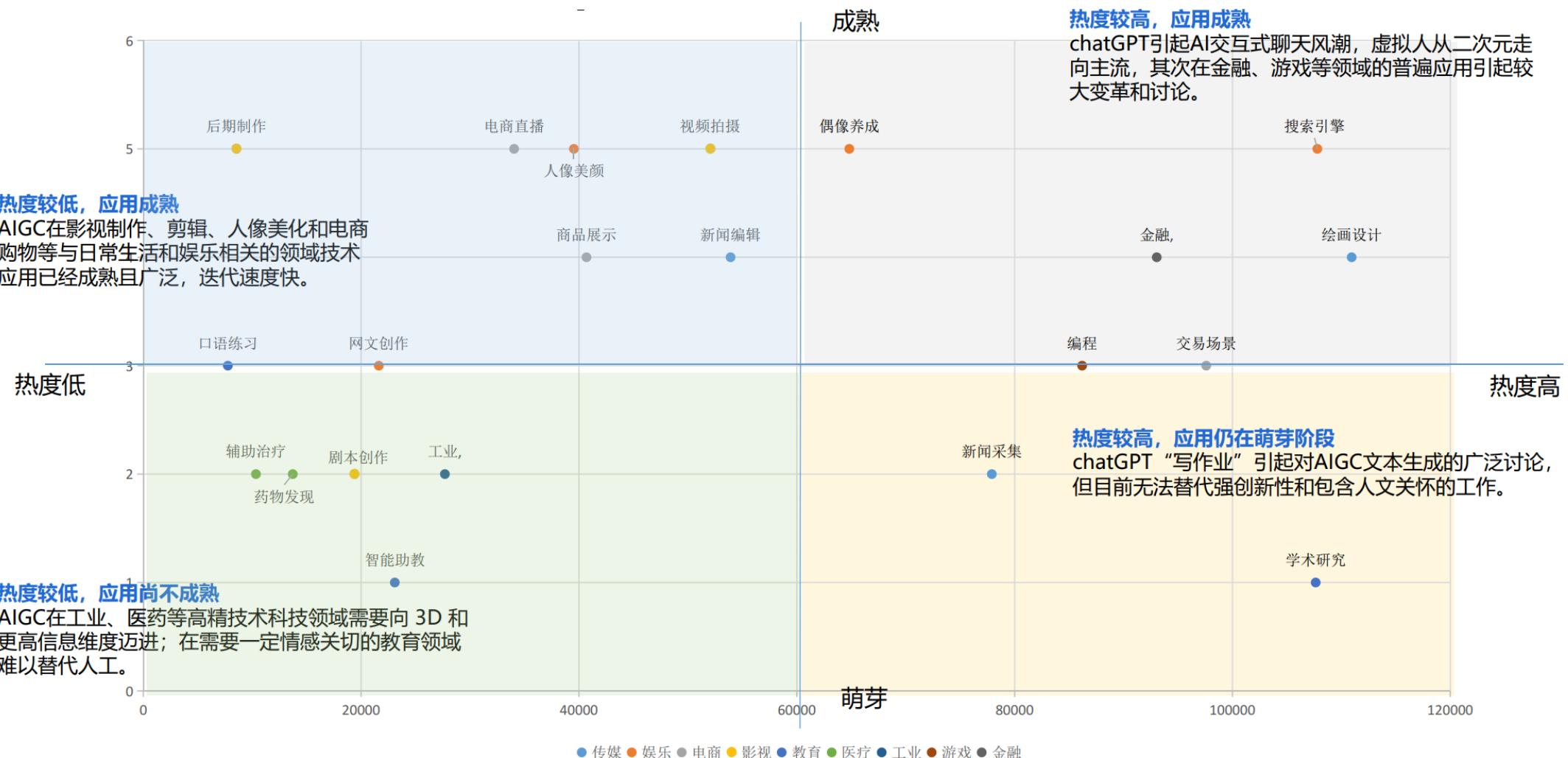
LLM更快、更便宜、更好用、更智能

人人都会有机器届王语嫣.



对现有人才培训模式的新挑战

业界成熟度



1萌芽阶段: 只在极小部分场景中应用, 很难替代人工, 通用基础层技术仍待研发与完善, 仅有极少数巨头或前沿企业探索

2初探阶段: 应用于少部分场景中, 部分技术有显著突破, 对人工的替代较少, 仅有少量上中游企业, 实际应用群体较少

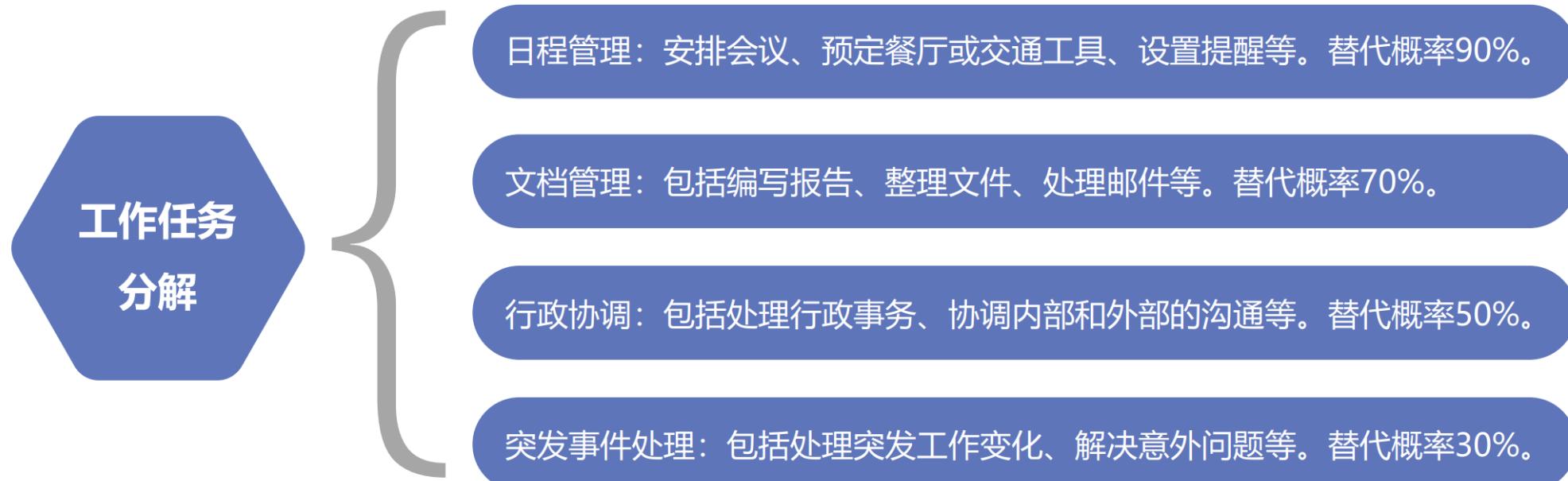
3一般成熟: 应用于许多具体场景中, 出现部分垂直领域的场景化小模型, 可以一定程度地替代人工, 有一定规模使用群体

4比较成熟: 较为广泛的应用, 技术细节待提升, 可以相当大程度替代人工, 涌现出多个下游企业和C端应用, 逐渐普及成为规模化应用

5非常成熟: 广泛应用于全链路, 技术迭代与调优能力强, 能够很好地替代人工, 中下游企业多, 面向C端的AIGC应用多且使用普遍

工作模式影响

任务分解替代概率模型 (Task Decomposition Replacement Probability Model): AI替代某个工作主要取决于以下要素——**任务重复性、标准化程度、创新和判断能力的需求、数据依赖性、经济成本、交互复杂性**，而每个工作的任务构成决定了最后的整体替代率，以秘书工作为例：



若假设这四部分各为25%，则得到 $(90\% * 25\% + 70\% * 25\% + 50\% * 25\% + 30\% * 25\%) = 60\%$ 。基于此种理论假设，秘书工作有60%的部分可以被AIGC替代。

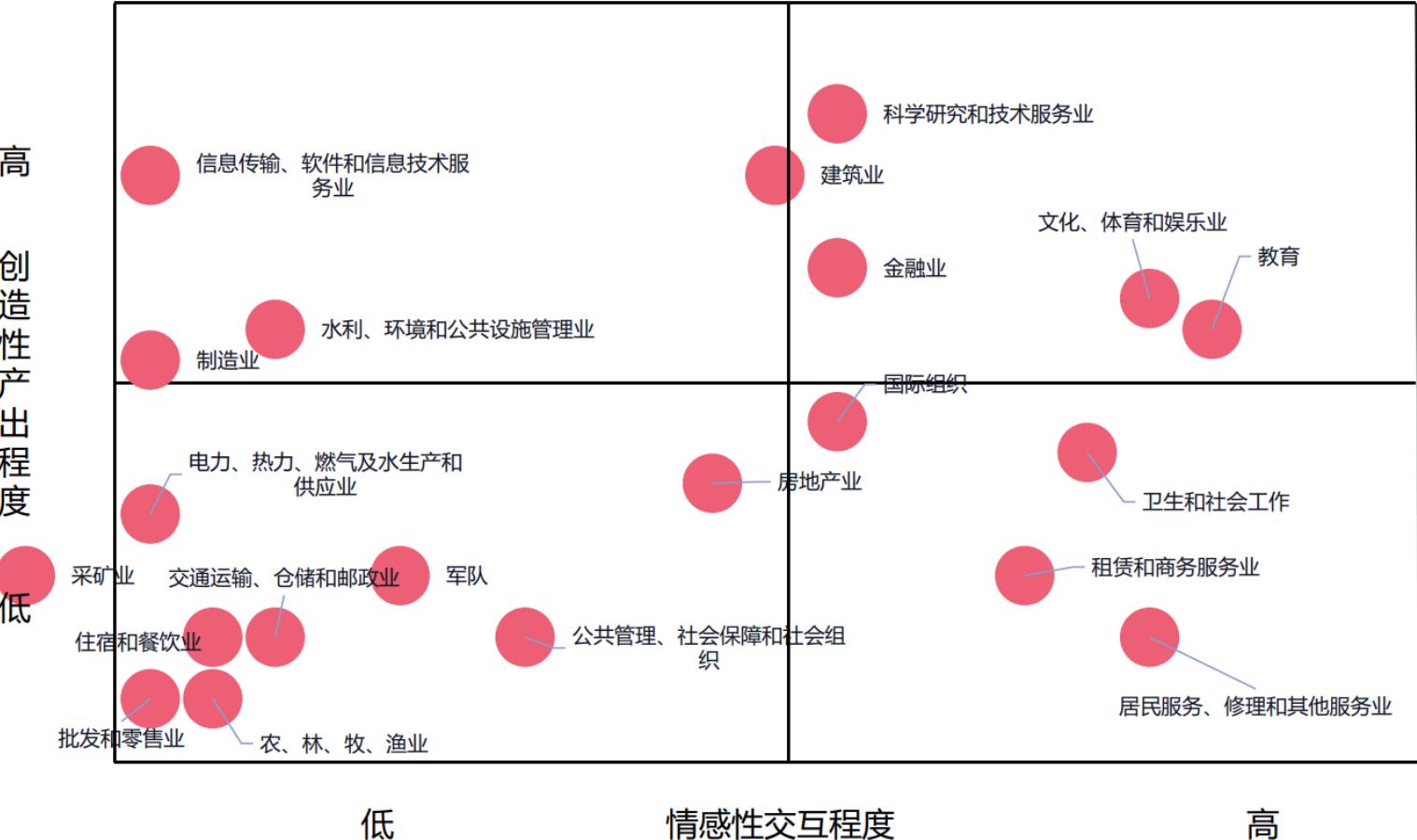
职业替代象限

AI职业替代的逻辑：

AI 能力范围之内的职业岌岌可危，
AI 能力范围之外的职业欣欣向荣。

AI时代铁饭碗职业：

高创造性与高情感性的脑力劳动
与体力劳动（机器人未全面普及
前）仍不可撼动。



职业替代表1

职业名称	基本任务	较难被替代部分	整体替代率 (毛估)
农民、农场工人	种植、收割、畜牧	复杂的农业工作, 比如树木修剪	70%
工厂工人	操作机器、装配产品、检查质量	需要手工技能和判断的任务	80%
销售人员	客户沟通、销售产品、维护客户关系	理解复杂的人类需求和建立人际关系	30%
教师	讲授课程、批改作业、辅导学生	理解复杂的学生需求和提供人性化关怀	50%
护士	照顾病人、执行医嘱、管理药物	提供人性化关怀和处理复杂的医疗情况	40%
医生	诊断疾病、提供治疗建议、进行手术	进行复杂手术和理解病人需求	40%
司机	驾驶车辆、遵守交通规则、保证乘客安全	在复杂环境下的驾驶能力	70%
办公室职员	处理文件、管理日程、处理通信	需要人际交往和判断能力的任务	60%
客户服务人员	回答客户问题、处理投诉、提供产品信息	理解复杂的人类需求和处理复杂问题	70%
餐饮服务员	接受订单、提供食物、清理桌子	提供优质服务和处理复杂情况	50%
零售员工	为客户提供产品信息、协助选择商品	处理交易面对面的销售和咨询服务	60%
清洁工	打扫卫生、清洁设施	处理复杂环境和不可预见的清洁任务	60%
建筑工人	建筑结构施工、装修	复杂的手工任务和现场决策的能力	40%
技工(电工、管道工等)	维修、安装和检查电气和管道系统	复杂的手眼协调和现场判断	30%

职业替代表2

职业名称	基本任务	较难被替代部分	毛估整体替代率
工程师	设计,测试和优化复杂系统	创新设计和解决复杂问题	40%
程序员/软件开发者	编写,测试和维护软件	复杂的软件开发	70%
会计师	记录财务交易,准备报表,进行审计	在解释和提供财务咨询方面	60%
律师	提供法律咨询,代表客户处理法律事务	理解和应用法律原则以及代表客户进行谈判和诉讼	35%
市场营销专员	策划和执行营销活动,进行市场研究	理解消费者行为,创造吸引人的营销策略和内容	50%
设计师	创造视觉效果,设计空间布局	理解和创造吸引人的设计	70%
项目经理	规划和协调项目,管理团队,监控进度	理解项目需求,处理复杂的人际关系和决策问题	30%
质量控制员	监督生产过程,确保产品质量	处理复杂的质量问题时	60%
金融专员/银行职员	进行金融分析,管理投资	处理银行业务在进行复杂的金融决策时	60%
研究员/科研人员	进行科学研究,发表科学论文	提出创新的科学理论,进行复杂的实验设计	30%
咨询顾问	为客户提供专业的建议	理解客户的具体需求, 提供个性化的解决方案	40%
社会工作者	帮助那些在生活中面临困难的人	理解人的情感, 提供心理支持	20%
法官	研究法律案例,裁决案件,解释法律、庭审监督等	法律的解释和裁决,以及对人类情感的理解和道德判断	30%

职业替代表3

职业名称	主要任务	不可替代部分	毛估整体替代率
领导*	制定策略,管理团队,做出决策	理解人的情感,管理团队,做出复杂的决策	20%
记者	报道新闻,撰写文章等	挖掘新闻背后的故事,进行深入的采访	50%
教授	进行教学和研究	提出创新理论,复杂的实验设计、与学生的深度交流	30%
美国总统	国家的领导和决策	理解复杂的社会、政治情况,好口才与深度共情能力	0%或100%
演员	表演角色,传达情感	传达复杂的人类情感,进行高水平的表演	50%
导演	制定创作方向,管理电影制作	提出创新的电影理念,管理电影制作过程	20%
自媒体营运人员	创建和发布内容	理解观众的需求,提供个性化的内容	40%
躺平一族*	选择较低压力和消费的生活方式	进行休息和享受生活	10%
董事长*	公司的高级领导职务,负责公司的重大决策制定和战略指导	领导能力、战略思维和人际交往能力	20%
键盘侠*	在网络上发表观点或评论	理解和模拟人类的情感和个性	70%
歌手	创作歌曲,录制音乐,进行演唱会	情感表达和个人魅力	40%
编剧	创作剧本	理解复杂的人物关系、人性深度、文化背景等	40%
按摩师	按摩对人体的穴位和肌肉进行按摩	复杂的人体结构、肌肉感知和反应	5%
美甲师	进行指甲美化	手工技艺	20%
KTV歌手	娱乐、陪唱、互动等	人际交往和情感互动	20%
网红	内容创作、粉丝互动、商业合作	粉丝互动和商业合作	50%
摄影师	拍摄照片和处理照片	拍摄照片、构图、光影掌控等	30%
主持人	语言表达、现场控制、观众互动	现场控制和观众互动	60%
科学家	研究设计、数据收集与分析、理论构建与验证、成果发布等	研究设计和成果发布	40%

AI安全隐患

AIGC强势介入全球产业链，将全面替代程序员、平面设计、法律AI客服，为人工成本划定上限，第三世界国家人口红利不复存在。第三世界产业链将因此遭受巨大冲击，低端产业链将不再进行转移。这会导致这些国家消费疲软，经济体系或将崩塌。大算力支持下的AIGC可能成为割裂跨国公司全球产业链的利刃，也可能成为划破“地球村”幻象的匕首。

生产力断层：得AI者得天下，AI技术发展可能会在发达国家或地区加速，甚至导致“卡脖子”技术垄断，而处在产业下游的第三世界国家区可能无法获取到足够数据，导致其人工智能技术发展缓慢，或将又一次“被动挨打”

生产方式断层：AI技术的普及和应用可能导致许多传统工作岗位消失，或将导致下游第三世界国家或地区的劳动力失业或难以适应新的工作环境，进而在经济、社会治安等方面产生新的问题。

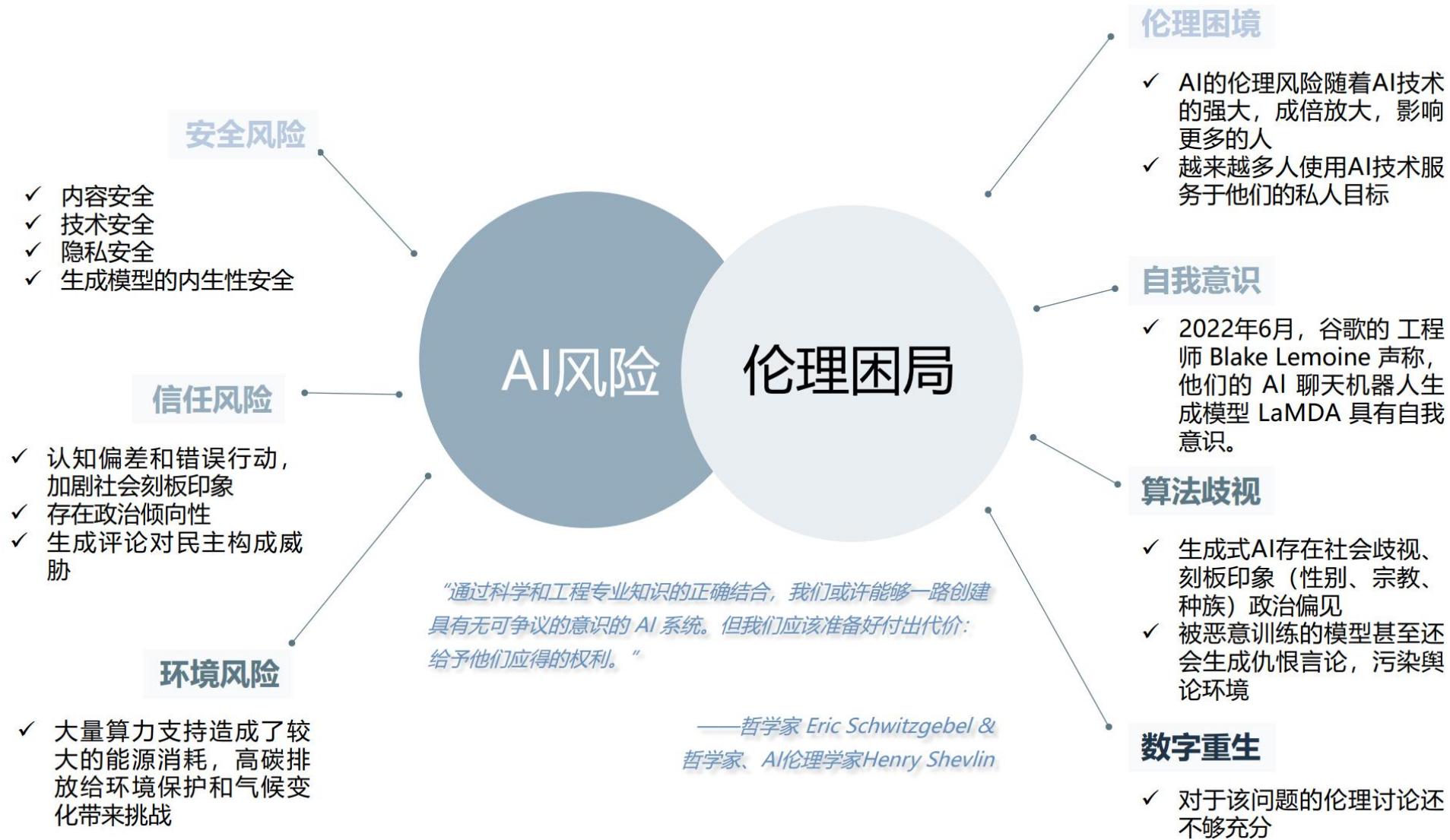
话语断层：伴随着生产力的突破发展，西方话语营造对“技术天才”的卡利斯马式崇拜，进一步让西方话语进入“赢家通吃”的时代。



世界工厂或将全面被AI+机器人接管

印度“世界办公室”将被AI替代

AI安全隐患



AI安全隐患

人类一方面依赖AI在获取信息、理解世界和进行社交等方面具备更强的真实世界洞察能力，但一方面又被AI的信息茧房/美颜的过度美化现象/社交机器人替代人际交往等等带来了更大的遮蔽和困扰。

以下是中介矛盾带来的五大变化：

知觉障碍的认知偏差

双重透视的虚拟现实

社交错位的人际关系

感官割裂的真实体验

意愿丧失的自主行为

工具理性和价值理性之间的冲突。工具理性强调效率、精确度和控制能力，这是AI技术的优势所在。而价值理性关注人的尊严、自由和真实的人际关系，这正是AI技术可能带来威胁的方面。

需要引入伦理和人文关怀，开发能够反映用户多元价值和需求的个性化AI，设计能够提升人们信息素养和批判思维能力的AI教育，以及建立能够防止AI滥用和保护用户权益的法律和规制等。

AI安全隐患

争议：著作权的主体是AI，还是使用AI的人？



- 腾讯研究院认为，作者应该是使用AI系统的人，而不是AI本身。
- 《科学》杂志主编索普指出，ChatGPT很好玩，但不能成为作者。
- 《自然》杂志声明，任何人工智能工具都不会被接受为研究论文的署名作者。
- “除了AI之外，是否有人的智力或创造性劳动”是目前判定著作权主体的通用做法。

AI侵权

著作权

训练AI的数据库收录了大量他人享有版权的作品，因此AI生成内容很可能面临侵权风险

肖像权

诸如“AI换脸”等深度伪造问题，直接涉及侵犯他人肖像权、隐私权，乃至人格权问题

名誉权

AI会被别有用心者利用，给诽谤性内容披上“AI生成”的外衣，侵犯他人名誉权

AI被侵权

- 现行的与著作权相关的法律规定难以直接认定AI或算法是作品的作者。
- 文本、图像早已深度数字化，难以直观地区分创作者是人类还是AI。
- AI产出作品的不到知识产权保护，也无法禁止他人未经授权的、以营利性为目的的使用。

AI安全隐患

黑天鹅事件

1

意外实现人工智能的突破

突然发现某种方法或算法，使AI迅速达到通用人工智能（AGI）水平。

2

AI伦理危机

AI系统产生严重道德伦理问题引发的激烈争论。

3

AI安全事故

一次严重的AI安全事故引发的对AI技术的广泛担忧和恐慌。

4

人类劳动市场崩溃

AI导致大量工作岗位被替代，引发全球性就业危机和社会不稳定。

5

恶意AI应用

恶意分子利用AI技术发动大规模攻击，例如制造虚假信息、网络攻击或其他形式的破坏活动。

6

数据隐私泄露

由于AI系统的设计缺陷或安全漏洞，导致大量用户的隐私数据泄露。

7

AI与心理健康

AI技术无意中导致人类心理健康问题的加剧，例如过度依赖AI引发的心理依赖、孤独感或焦虑症。

8

AI助推病毒制作风险

生物科学领域的应用导致意外后果，例如危险病毒的泄漏或基因编辑技术的滥用。

9

竞争与冲突

国家或组织之间争夺AI领导地位，可能会升级为全球性的战略竞争。

10

AI反乌托邦

出现一个具有高度智能、自主性的AI系统，它试图通过控制其他AI系统、基础设施或人类行为来实现自己的目标。

对国家的影响

国家需要自建LLM能力，保证数据安全并实现联邦学习

国内外厂商齐发力，大语言模型产业规模可观

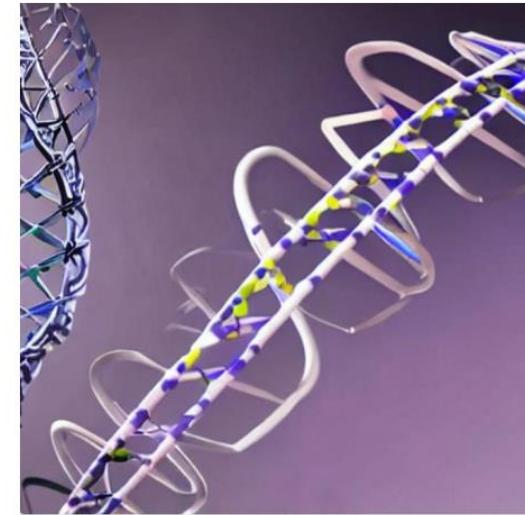
国外	基础模型				ChatBot		其他应用	
	Google	Google DeepMind	Meta	OpenAI	BigScience	stability.ai	ChatBot	其他应用
	LaMDA	T5	Gopher	GPT-4	Bloom	Stable Diffusion	Bard	Notion AI
	PaLM	Imagen	Chinchilla	MMS	T0	StableLM	BingChat	Cedille AI
	PaLM-E	Flan	Gato	OPT-175B	BloomZ		ChatGPT	Copilot
	Stanford University	databricks	AI21 studio	LIMA-65B	GPT-J 6B	vicuna-13b	Claude	Colab
	Stanford Alpaca	Dolly 2.0	Jurassic-1	Jumbo	Claude			Copilot
国内	基础模型				ChatBot		其他应用	
	悟道	二郎神	孟子		ChatGLM	ChatJD	钉钉	序列猴子
	智源研究院	Baidu	文心	inspur	从容	MOSS	WPS AI	EMOTIBOT
	悟道	文心	二郎神	浪潮	商汤	SenseChat	wondershare	FRIDAY
	通义	JD.COM	言犀	源1.0	日日新		字而思网校	有道
	通义	JD.COM	言犀	腾讯	混元		MathGPT	曹植
	盘古	MINIMAX	基础模型	网易伏羲	玉言		HAOMO.	达观数据
	华为云	开放平台		网易伏羲	玉言		雪湖·海若	知乎
	国家超级计算天津中心	科大讯飞	星火	云从科技	自研大模型		METASOTA	知海图AI
	天河天元大模型	iflytek						小冰

对国家的影响

大语言模型发展中两个相反趋势：**巨大模式是全模态，海量参数，期待涌现**，
bigger and smarter，从**大数据到全数据**的发展；**小微模式是模型压缩和优化**，
free and smaller，力争在有限资源达成近似性能。

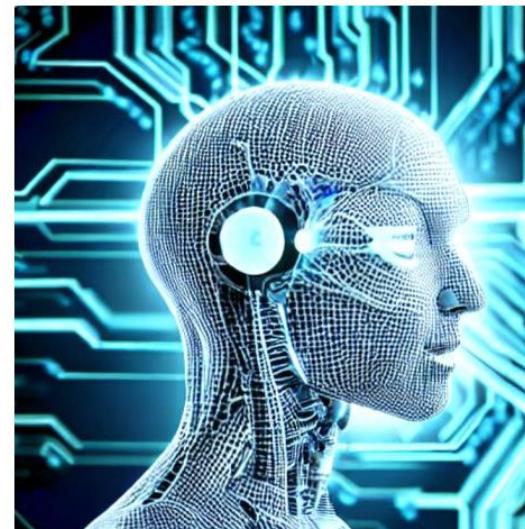
1 巨大模式：在更广泛的知识领域中汲取智慧，对问题提出更深入的见解。强调知识的无穷性和多样性，及AI技术在追求真理和理解世界时的角色。

案例 • 2023年3月14日发布的GPT-4预估参数量极高，相较GPT-3的1750亿参数量，增加了数十倍，并采用了更多和更丰富的训练数据，拥有更高的理解和专业水平。



2 小微模式：在保持较高预测准确性的同时，大幅度减小模型的大小和计算成本，优化计算效率，实现高效的模型预测，可以在较小的设备上运行

案例 2023年3月，斯坦福发布了轻量级语言模型Alpaca。该模型在LLaMA的基础上加入指令•Alpaca（羊驼）模型：微调，只有70亿参数。可以在笔记本电脑上部署，甚至手机和树莓派上运行，但其性能可媲美GPT-3.5这样的超大规模语言模型。



总结

- ChatGPT的意义是展现【LLM的可行性】
- 人类的传承需要【知识创造、传承、应用】
- NLP技术有速度快、记忆大、覆盖广、无疲劳的优点
- LLM改变了人与人、人与机器的协作效率
- LLM影响全人类的发展，类似互联网、计算机的发明
- AI有安全的隐患，需要注意
- 国家必须要有自己的LLM，需要自己的AIGC技术

/04 如何应对

Preparations



如何应对



01 善于利用工具是人类优势.

人与动物的区别是，人类【善于】利用工具，创造工具、改进工具、使用工具

02 不要否定工具.

不是工具取代人，是会用工具的人取代不会用的

03 取代现象会一直出现.

就算没有LLM，也会有别的新技术新工具，一定会出现取代现象



01

克服 工具抵触

- 不要成为【无法使用工具】的人
- 历史车轮无法停下，抵触只会失去工具使用的优势



02

克服 学习抵触

- 警惕【高度特化教育】模式
- 学习习惯需要改变，不要等待灌输
- 不要脱离实体去记录符号
- 重塑【终生学习】能力

总结

“【学习系统】升级 伴随生命系统跃升.

- LLM影响的就是学习系统，知识的【创造、传承、应用】
- 虽有风险，但多国表示绝不下车
人工智能新浪潮

技术奇点

- 正在改变群体使用知识的方式
- 降低【继承式】人才竞争力
- 高层次认知能力是瓶颈



国家机会

- 不可再次错失的工业革命机会
- 要抓住未来国际竞争的优势
- 主动输出我们的文化与价值观



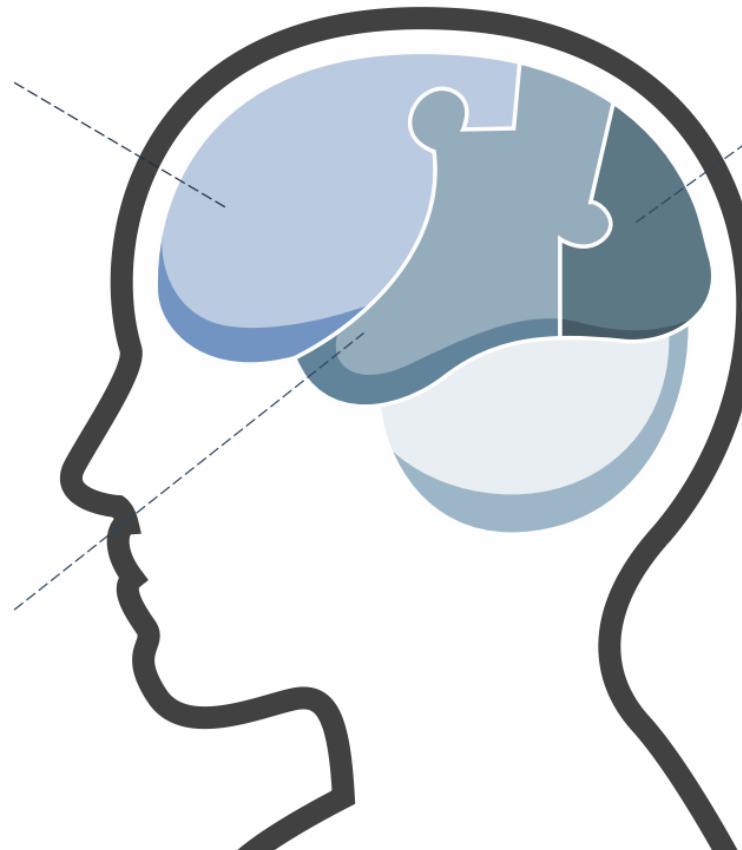
哲学思考

AI意识:

- AI是否能够像人类一样体验、感知和思考?
- AI是否具有真正的自主性? 它们的决策和行为是否只是预先编程或学习的结果?
- 我们是否应该将AI的行为视为有意图的行为?

AI道德:

- AI系统是否具有道德地位? 如何平衡它们的利益与人类的利益?
- 如何确保AI符合人类的道德价值观? 是否可以将道德原则“交给”AI?



AI与人类:

- 如何界定人与机器之间的区别?
- AI技术是否会影响人类的自由意志?
- AI是否能够达到与人类相同的认识和创造能力?

讨论这些问题有助于我们更深入地理解人工智能的本质、限制和潜力,为AI技术的负责任发展提供指导。



THANKS!