人工智能程序设计

大模型

——从机器学习到 ChatGPT

北京石油化工学院 人工智能研究院

刘 强

开场提问：你这周用过哪些 AI?

DeepSeek　豆包　Kimi　通义千问　文心一言　ChatGPT

你用它做过什么?

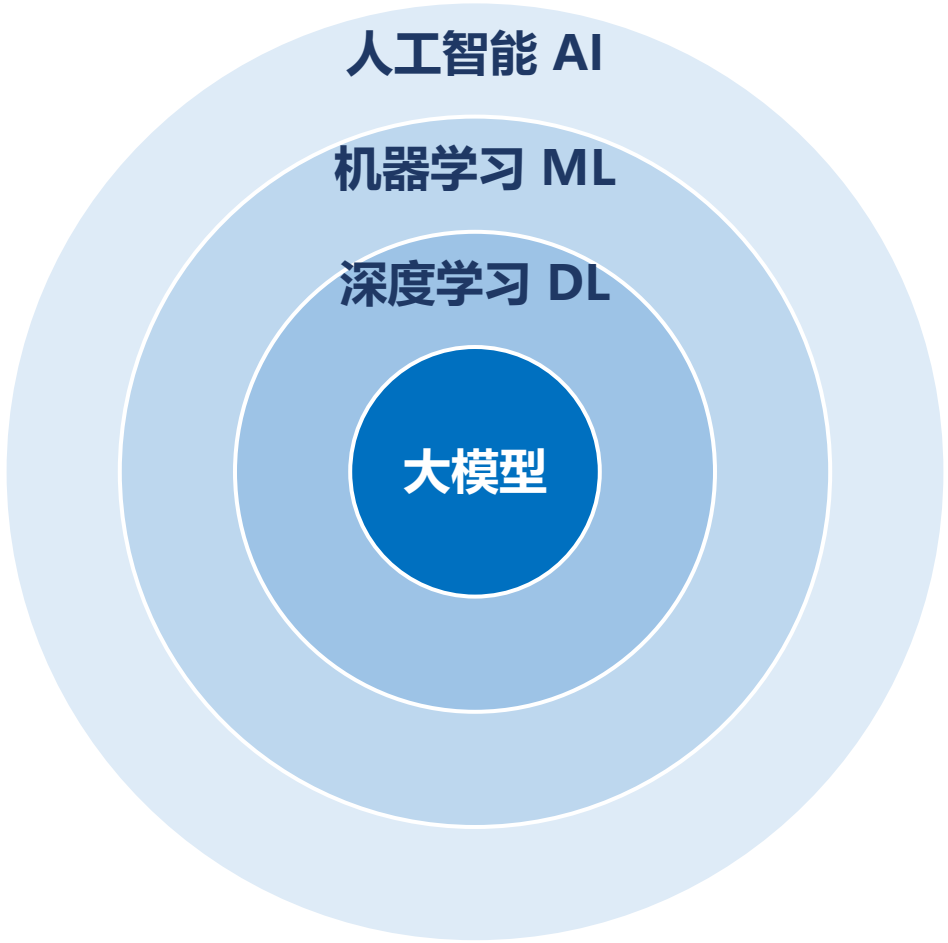
查资料　翻译　写总结　解释代码　做 PPT

共同点：你用自然语言提问，它也用自然语言回答

它是怎么做到的?

它真的“懂”你在说什么吗?

人工智能、机器学习、深度学习：三个同心圆



人工智能（AI）

让机器表现出智能的一切方法；这个名称在 1956 年达特茅斯会议上提出

机器学习（ML）

不再由人一条条写规则，而是让机器从数据中自己学出规律

深度学习（DL）

机器学习的一类方法，用很多层的神经网络来学习

大模型（LLM）

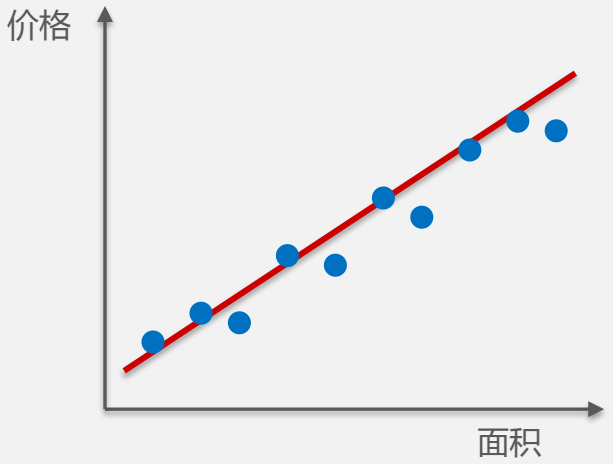
深度学习中的“明星”：用海量文本训练的超大规模神经网络

机器学习核心思想：从数据中学规律

数据 → 学习算法 → 模型 → 预测

不是告诉机器怎么做，而是给它看足够多的例子，让它自己总结规律

房价预测（回归：预测一个数）



已知：很多套房子的面积和价格

学到：面积越大，价格越高（红线）

预测：新房子大约多少钱

垃圾邮件分类（分类：判断属于哪一类）

“恭喜中奖！点击免费领取……” → 垃圾邮件

“周三下午的实验课调到 3 点” → 正常邮件

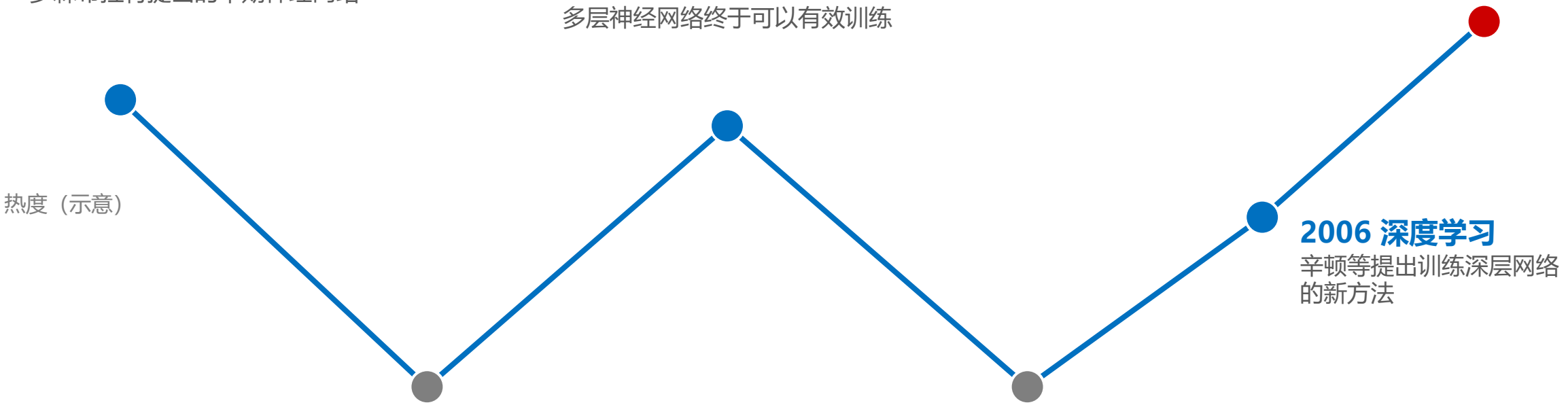
看过成千上万封标注好的邮件后，机器自己发现“中奖”“免费领取”常出现在垃圾邮件里

神经网络的起落：感知机、两次 AI 寒冬、反向传播

1958 感知机
罗森布拉特提出的早期神经网络

1986 反向传播
多层神经网络终于可以有效训练

2012 爆发
AlexNet（见下页）



2006 深度学习
辛顿等提出训练深层网络的新方法

1970 年代：第一次寒冬
解不了“异或”问题，算力不足

80 年代末—90 年代：第二次寒冬
数据和算力仍然不够

启示
数据、算力、算法
缺一不可

深度学习爆发：2012 年 AlexNet 与 ImageNet

2012

AlexNet 在 ImageNet 图像识别比赛中以明显优势夺冠，错误率远低于传统方法

设计者：辛顿和他的学生克里热夫斯基、苏茨克维

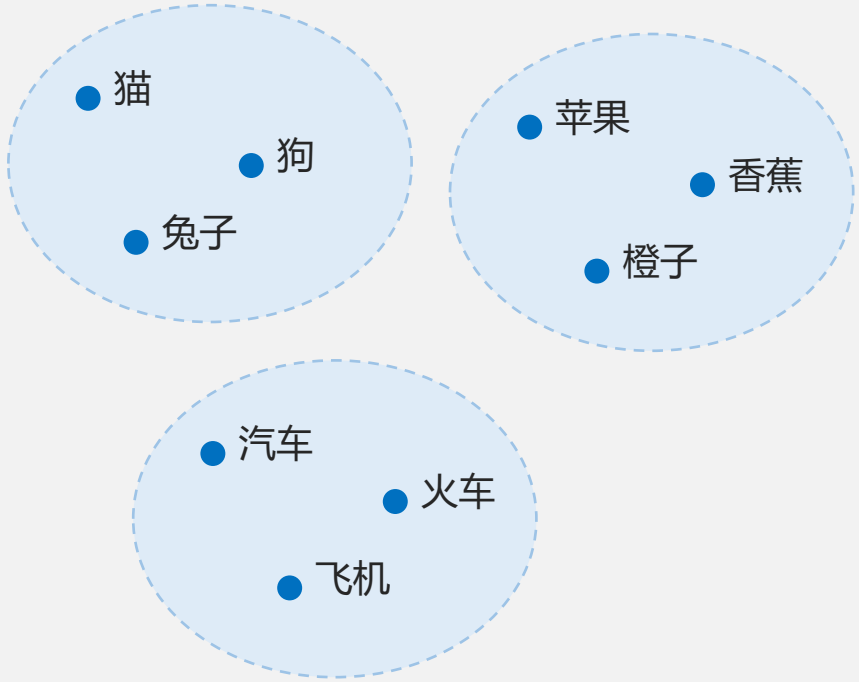
深度学习时代正式开启

深度学习爆发

数据	算力	算法
ImageNet：上百万张带标签的图片	GPU（显卡）并行计算，比 CPU 快得多	更深的网络结构和新的训练技巧

让机器读懂语言：词向量、RNN/LSTM 与长句难题

词向量：意思相近 → 向量也相近



每个词 = 一串数字；word2vec（Google，2013）让它流行

RNN / LSTM：一个词一个词往后读，边读边记



颜色越浅 = 记得越模糊；LSTM（1997）加了“记忆门”，能记久一些

难题一：长句会“忘”

句子一长，开头的内容就记不清了

难题二：只能串行

必须按顺序一个个算，没法并行，训练很慢

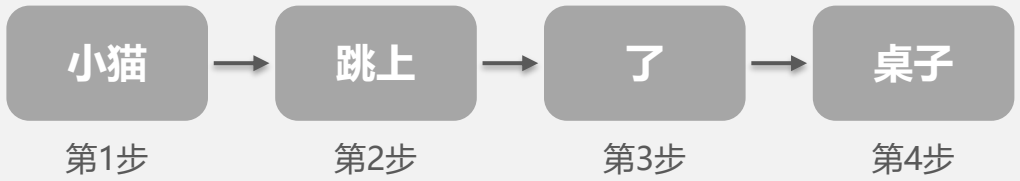
2017 Transformer：注意力就是你所需要的一切

Attention Is All You Need

Google 研究团队 · 2017 年论文 · 提出 Transformer 架构

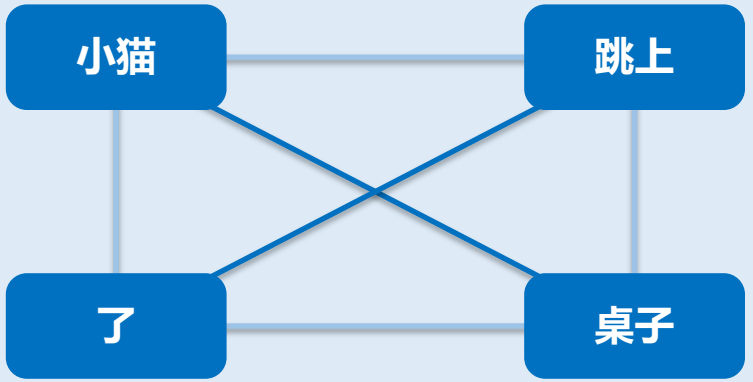
GPT 中的 T
= Transformer

以前（RNN）： 按顺序，一个个算



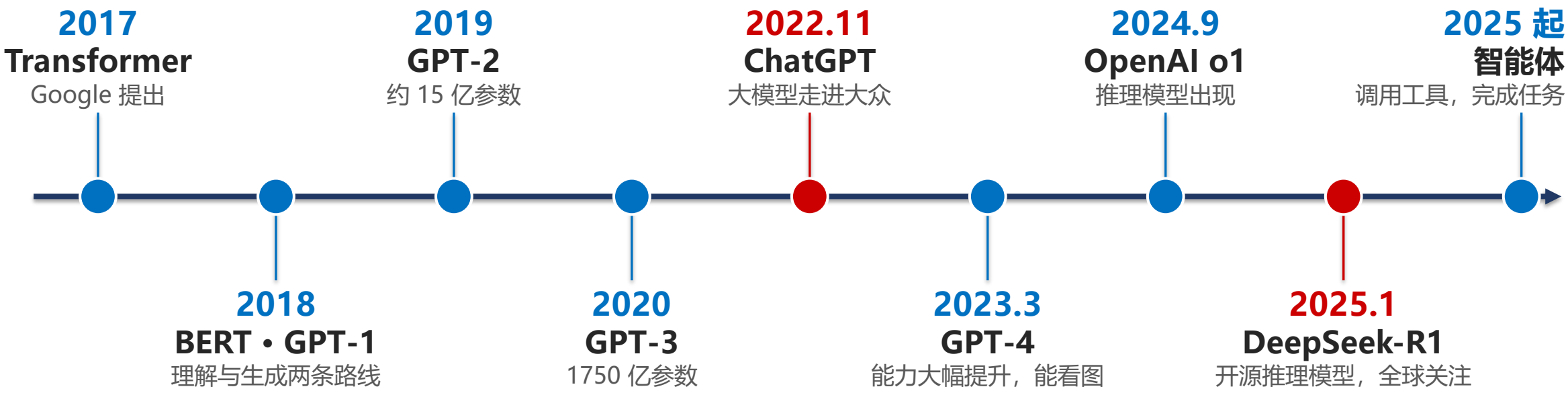
必须等前一个词算完，GPU 也只能干等

Transformer： 整句并行 + 注意力



每个词都和其他所有词直接相连，同时计算

时间轴：从 Transformer 到智能体

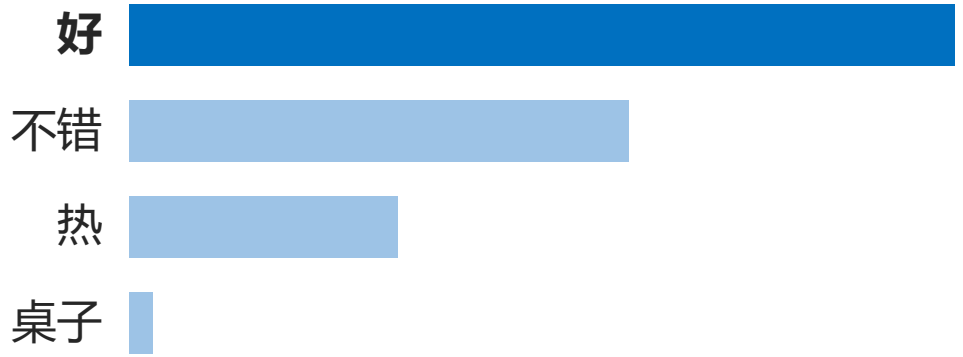


ChatGPT 发布约两个月用户数突破 1 亿，被认为是当时增长最快的消费级应用

大模型的本质：预测下一个词

输入一句话，模型给每个可能的下一个词打分

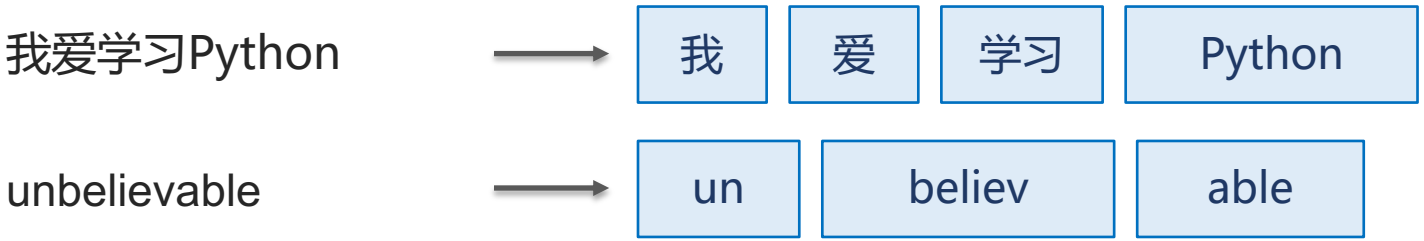




概率高低（示意）

选一个词接上 → 再预测下一个 → 逐词“接龙”

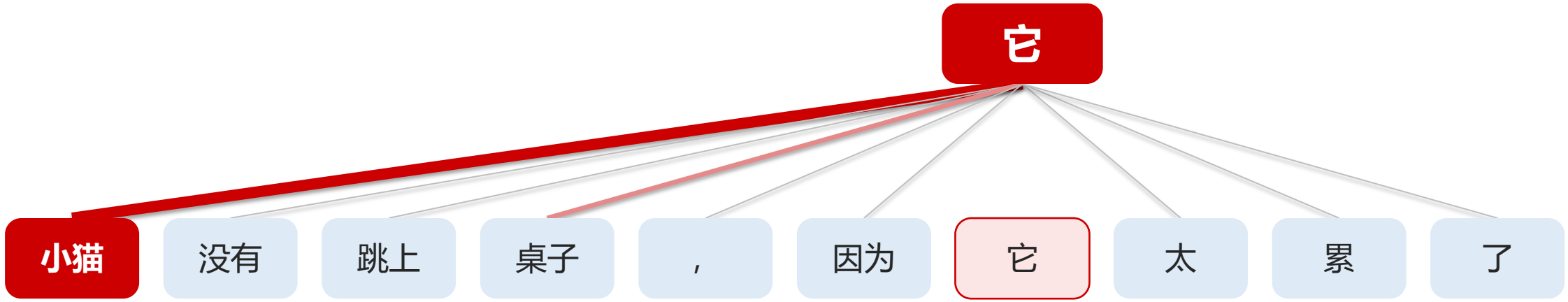
Token（词元）：模型处理文字的最小单位，一句话先被切分成一串 Token



不同模型切法不同，此处为示意
大模型 API 一般按 Token 数量计费

注意力机制：读到“它”时，该看哪里？

线越粗 = 注意力权重越大
“它”最关注“小猫”



小猫没有跳上桌子，因为它太**累**了 → 它 = **小猫**

小猫没有跳上桌子，因为它太**高**了 → 它 = **桌子**

计算“它”与句中每个词的相关程度，
相关的词权重高，再融合信息理解含义。
大模型有很多层、很多个“注意力头”

三步养成：预训练 → 指令微调 → RLHF

① 预训练

做什么
读遍海量文本（网页、书籍、代码），反复练习“预测下一个词”

得到
知识渊博，但只会“接话”

好比：**博览群书的学生**

② 指令微调

做什么
用大量“指令—优质回答”的例子继续训练

得到
听懂指令，按要求回答

好比：**做有标准答案的习题**

③ RLHF

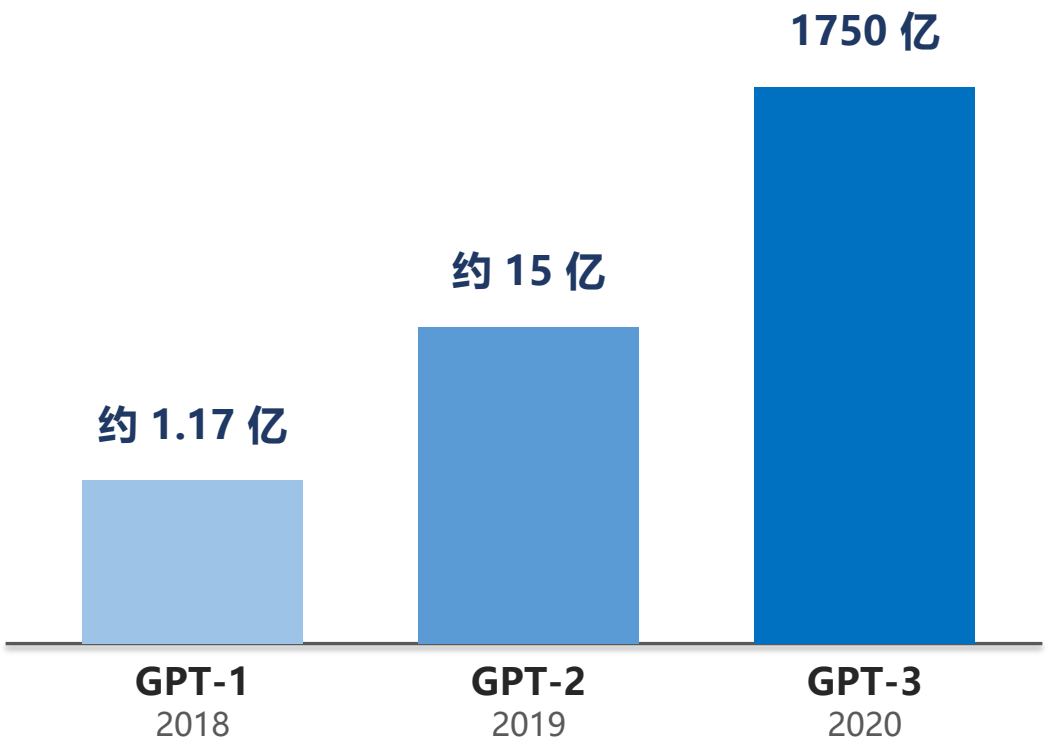
做什么
人类反馈强化学习：人给多个回答打分排序，训练“打分模型”来引导大模型

得到
更有用、更安全、更符合人类偏好

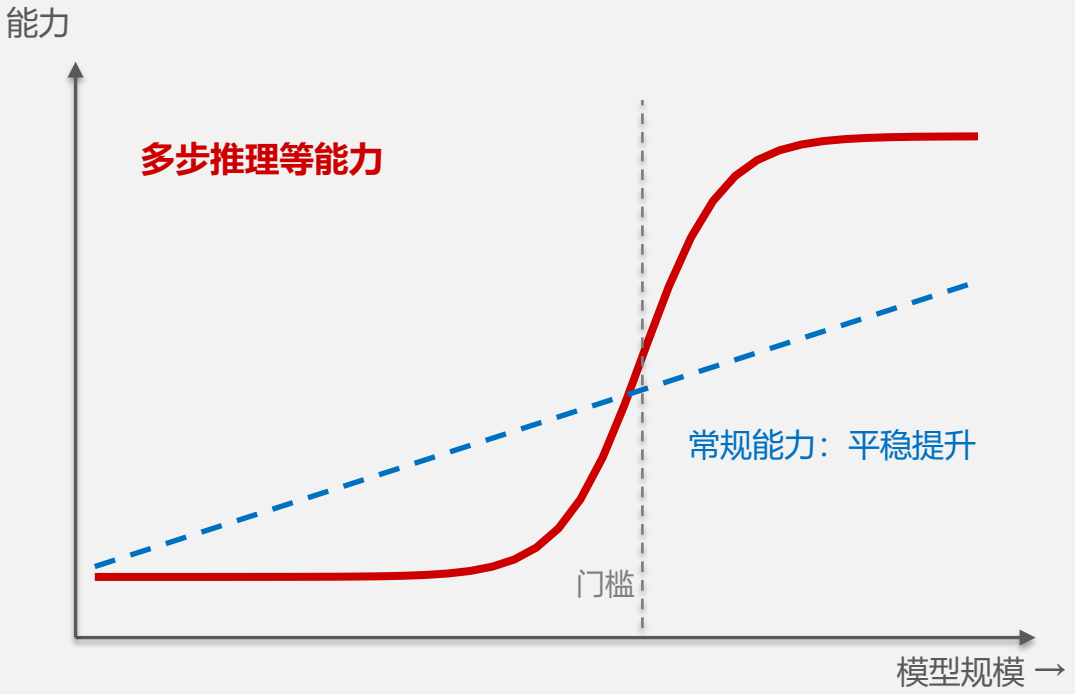
好比：**老师批改点评**

规模定律与涌现：大模型为什么要“大”？

GPT 系列参数量（对数刻度示意）



涌现：跨过门槛，能力“突然”出现



规模定律：参数、数据、算力越大，表现越好且可预测；　**瓶颈：**高质量数据快用完、算力成本高 → 转向“想得更久”

从快思考到慢思考：思维链与推理模型

例题：小明有 3 个苹果，又买了 2 袋，每袋 4 个，一共几个?

快思考：凭直觉脱口而出

“3 + 2 + 4 = 9 个”

错误

早期大模型常这样：多步题容易错（示意）

慢思考：一步一步想（思维链）

① 2 袋 × 4 个 = 8 个
② 8 + 3 = 11 个

11 个

提问时加一句“请一步一步思考”

思维链提示
2022

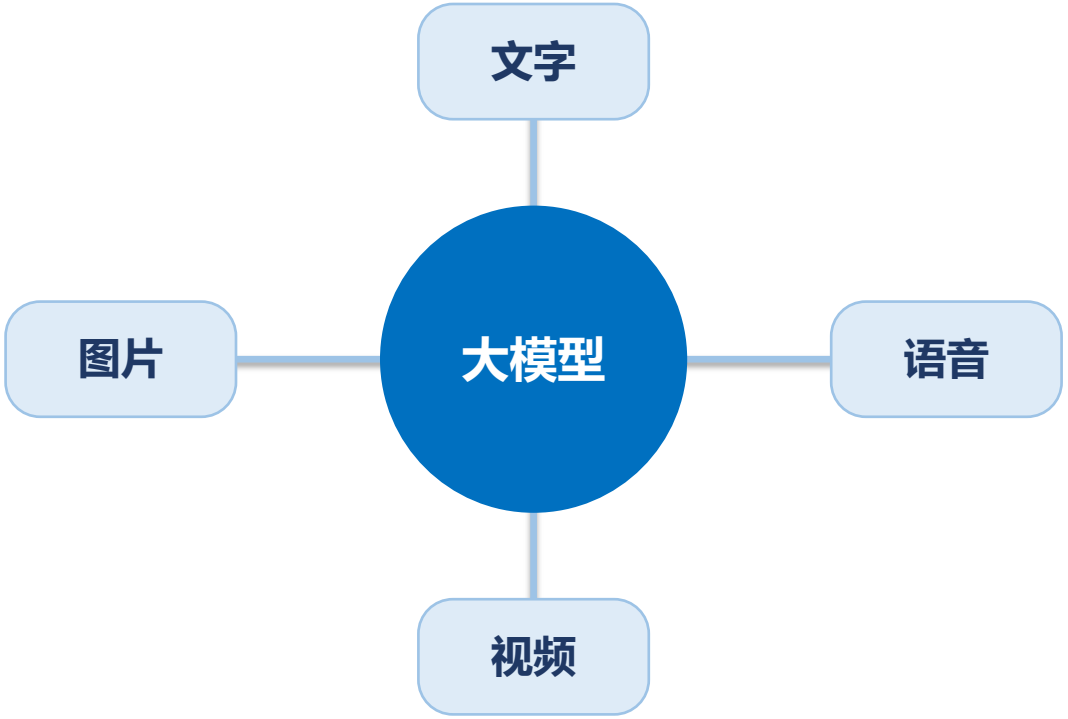
OpenAI o1
2024.9

DeepSeek-R1
2025.1

推理模型：先思考、再回答
从“模型更大”到“思考更久”

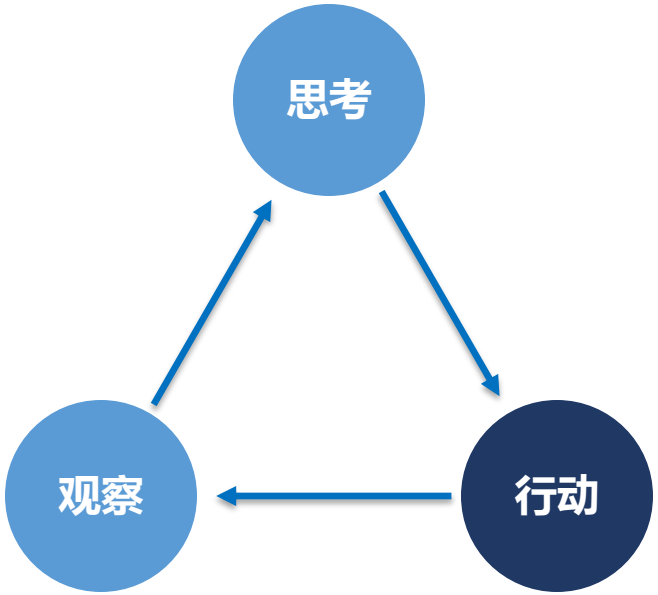
多模态与智能体：会看会听，还会动手

多模态：文字、图片、声音、视频都能处理



例：拍一道题的照片，它能识别并讲解

智能体（Agent）：拆解任务，调用工具



常用工具
搜索网页
运行代码
读写文件
查询数据库

“明天北京天气适合穿什么？”
→ 调用天气工具 → 拿到数据 → 结合常识给建议

大模型的应用：写作、编程、科研、医疗、教育

写

写作

起草、润色

翻译、总结长文档

码

编程

解释代码、找 bug

生成代码（Copilot、Cursor 等）

研

科研

帮助阅读文献

整理思路、分析数据

医

医疗

整理病历、解释知识

诊断必须由医生负责

教

教育

24 小时私人助教

答疑、出题、讲解

对学 Python 的我们：看不懂的报错让它解释，写好的代码让它检查——它是**“助教”**，不是**“替考”**

大模型和 Python：几行代码调用大模型 API

```python
from openai import OpenAI

client = OpenAI(
    api_key="sk-你的API密钥",          # 占位符，换成自己的 Key
    base_url="https://api.deepseek.com",
)

response = client.chat.completions.create(
    model="deepseek-flash",            # 以官方文档为准
    messages=[
        {"role": "system", "content": "你是一位耐心的 Python 老师"},
        {"role": "user", "content": "用一句话解释什么是列表"},
    ],
)
print(response.choices[0].message.content)
```

1 安装库

`pip install openai`

2 创建 API Key

在 DeepSeek 等开放平台注册

3 运行代码

打印出模型的回答

注意

API Key = 账户密码：不要写进要分享的代码、不要传到 GitHub，建议放进环境变量

模型名、价格以官方文档为准

局限与风险：大模型会犯错，要正确用 AI 学习

幻觉

一本正经地编造事实、文献和数据——它追求“像”，不保证“对”

偏见

训练数据里的偏见，也会被模型学进去

隐私

不要输入密码、身份证号、未公开数据等敏感信息

学术诚信

把 AI 生成的内容当成自己的成果提交，属于学术不端

正确用 AI 学习

1. **先思考**，再提问
2. **要核实**，别全信
3. **讲过程**，不抄答案
4. **用了 AI**，要说明

写程序：先自己写，报错了再请 AI 解释原因，自己改正

中国大模型版图与未来展望

公司	代表产品
深度求索	**DeepSeek**
阿里巴巴	**通义千问（Qwen）**
字节跳动	**豆包**
百度	**文心一言**
月之暗面	**Kimi**
智谱	**智谱清言（GLM）**
腾讯	**腾讯混元**

DeepSeek、通义千问等还开源了模型权重，全世界开发者都可以下载使用

未来展望

- **更会推理**：思考更深，错误更少
- **多模态**：文字、图像、声音、视频融为一体
- **智能体**：从“会说”到“会做”
- **具身智能**：大模型装进机器人，进入物理世界
- **端侧部署**：在手机、电脑上直接运行

小结

本节要点

1. 大模型属于 AI → 机器学习 → 深度学习这一脉络
2. 数据 + 算力 + 算法推动深度学习爆发；2017 年 Transformer 开启大模型时代
3. 本质是“预测下一个词”；预训练 → 指令微调 → RLHF
4. 规模带来涌现；推理模型会“慢思考”，智能体会“动手”
5. **大模型会犯错：用它辅助学习，不替代思考**

