

大语言模型 - 绪论

主讲人：陈骋

时间：2025.09

ChatGPT用户增速



Deepseek用户增速



- DeepSeek-R1在2025年01月20日发布
- 1月累计获得 1.25亿用户 (Web+App, 累加不去重)
- 其中 80% 用户来自最后一周
- 7天增长 1亿用户 (无广告投放)
- 上线 20天, 全球日活跃用户突破 2000万
- 成为 全球增速最快的 AI 应用

中国市场

➤ Token消耗量

- 2024年初：日均 1千亿
- 2025年中：日均突破 30万亿
- 增长幅度：1年半时间增加300+倍

➤ 高质量数据集建设

- 截至2024年6月底：超 3.5万个数据集
- 总体量：400PB（≈ 国家图书馆数字资源总量的 140倍）

➤ 数据交易规模

- 累计交易额：近40亿元
- 挂牌高质量数据集规模：246PB
- 北京数交所：高质量数据集占交易总量比例 10% → 80%

➤ 新模式探索

- 上海、天津、安徽：试点“数据语料作价入股”
- 推动企业将高质量数据集折算为股权投入

➤ 数据标注产业布局

- 已建设 7个数据标注基地
- 典型地区：成都、沈阳、合肥等

国家数据局：截至6月底，日均Token消耗量已突破30万亿

2025年08月15日09:15 | 来源：人民网

T= 小字号

人民网北京8月14日电（记者杨曦）国务院新闻办公室今日举行“高质量完成‘十四五’规划”系列主题新闻发布会，介绍“十四五”时期数字中国建设发展成就。国家发展改革委党组成员、国家数据局局长刘烈宏在会上表示，作为人工智能发展的三大核心要素之一，数据在推动“人工智能+”过程中发挥着关键作用，特别是高质量数据集的建设至关重要。

刘烈宏介绍，在人工智能时代，Token（词元）作为处理文本的最小数据单元，如同互联网时代大家所说的“流量”。2024年初，我国日均Token的消耗量为1千亿，截至今年6月底，日均Token消耗量已经突破30万亿，1年半时间增长了300多倍，这反映了我国人工智能应用规模的快速增长。



CONTENTS 目录

1. 机器学习和深度学习
2. 语言模型的发展历程
3. 大语言模型能力特点
4. 大语言模型关键技术
5. 大语言模型技术挑战

01

机器学习与深度学习

➤ 机器学习 (Machine Learning)

- 最广义的概念，利用数据和算法让计算机学习规律
- 方法包括：线性回归、决策树、支持向量机、神经网络等

➤ 神经网络(Neuron Networks)

- 机器学习的一类方法，模仿人脑神经元结构
- 包含前馈网络、卷积网络、循环网络等

➤ 深度学习 (Deep Learning)

- 常见为神经网络的扩展形式，使用多层网络
- 在图像、语音、自然语言等复杂任务上表现突出

➤ 大语言模型 (Large Language Model)

- 深度学习的一种应用，基于 Transformer 架构
- 规模巨大（数十亿至上万亿参数）
- 典型代表：BERT、GPT、LLaMA、Deepseek



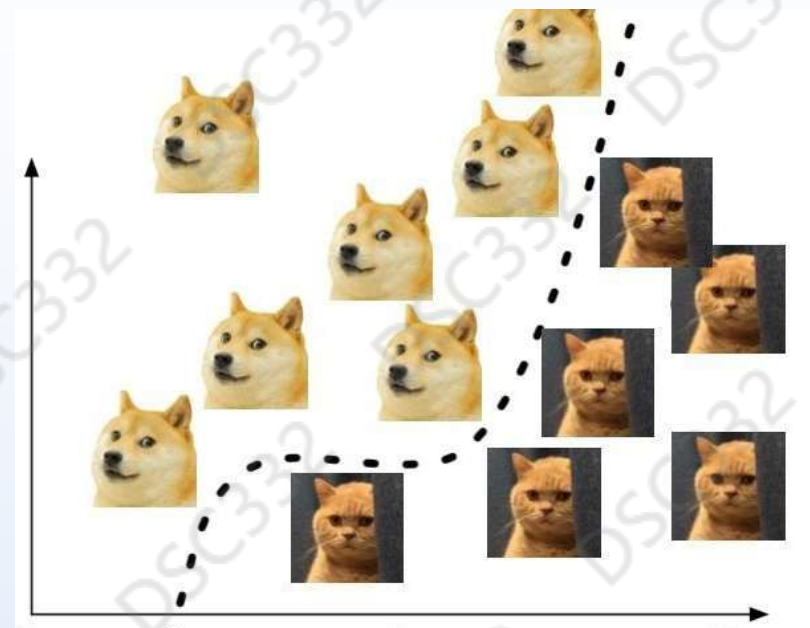
什么是学习?



这是猫

这是狗

什么是机器学习?



机器学习
≈
构建一个
映射函数



- 语音识别

$$f(\text{语音波形}) = \text{"你好"}$$

- 图像识别

$$f(\text{数字9}) = \text{"9"}$$

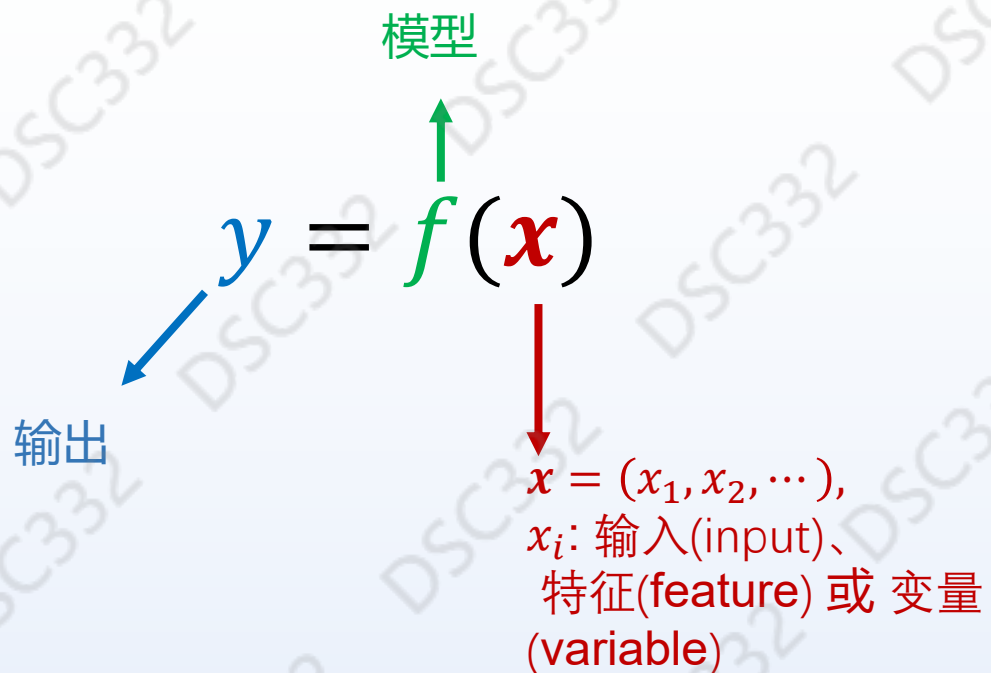
- 围棋

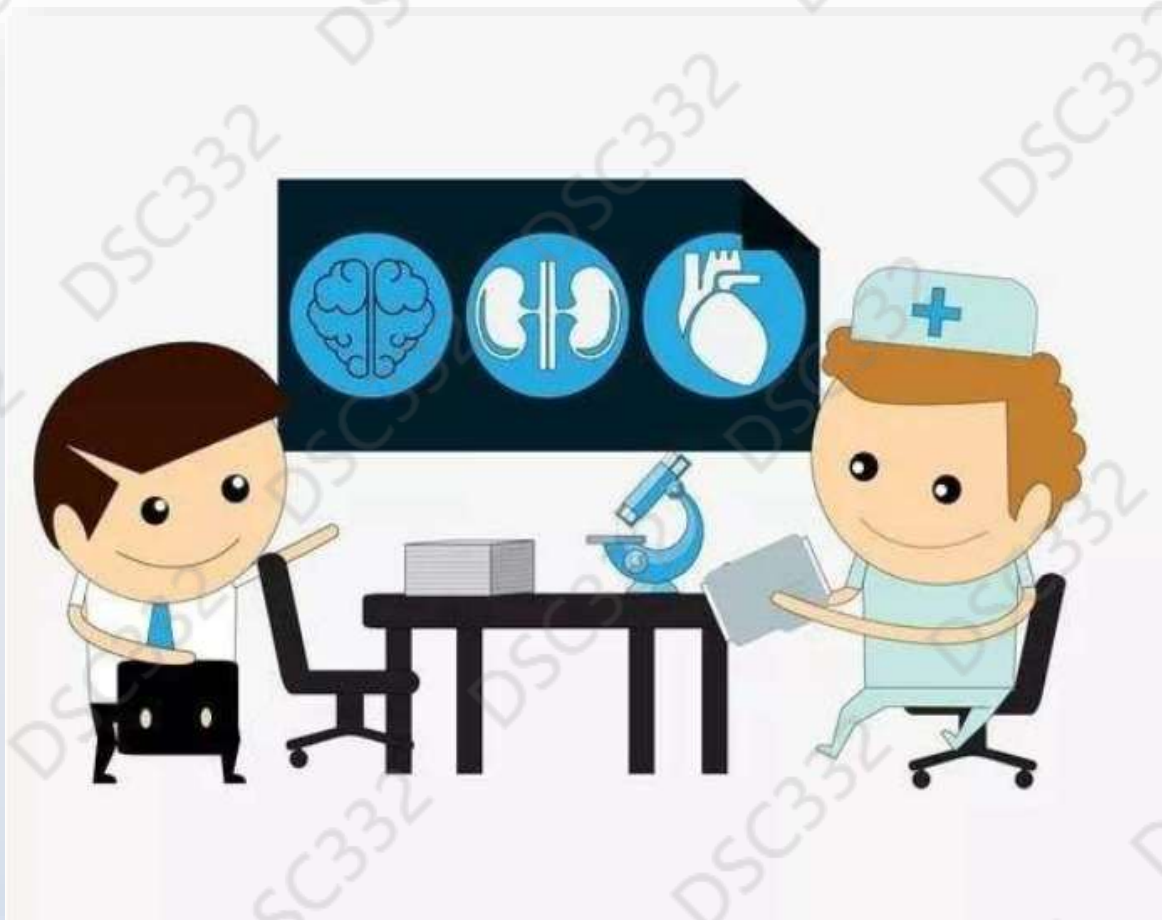
$$f(\text{围棋棋盘}) = \text{"6-5" (落子位置)}$$

- 机器翻译

$$f(\text{"你好!"}) = \text{"Hello!"}$$

机器学习 $\approx$ 构建一个映射函数





医学诊断

□ x_i 可能包括年龄，性别，
体重，血压，化验指标，
家族病史等等

□ y 是否患有这种疾病



垃圾邮件识别

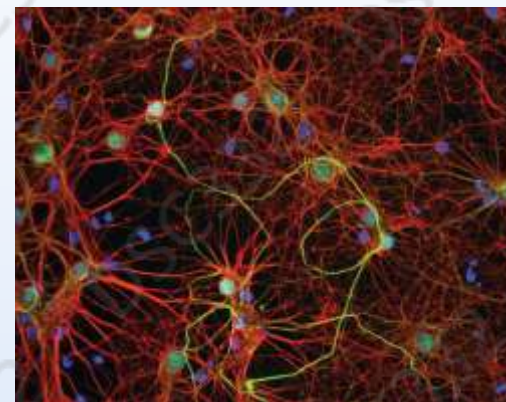
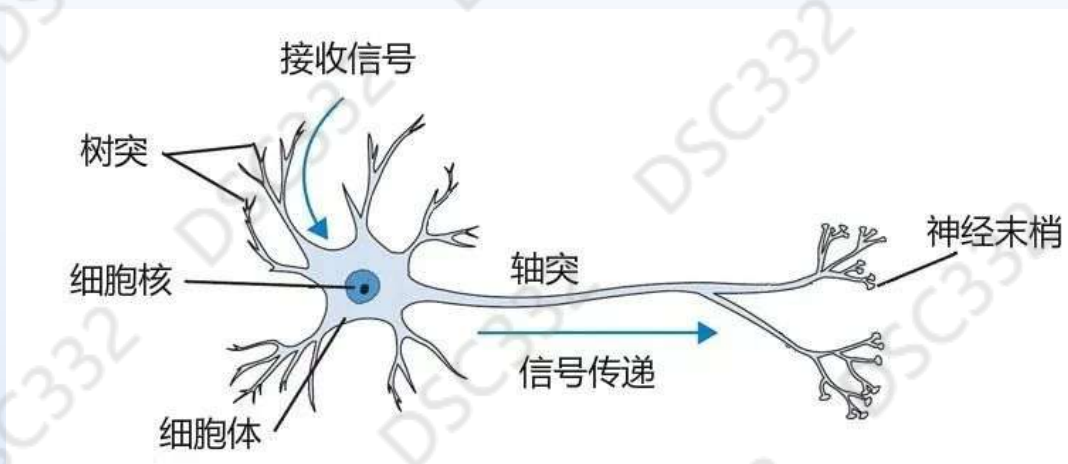
□ x 表示邮件中的所有文字, x_i 表示某个单词出现的次数

□ y 是不是垃圾邮件



□ 概念：

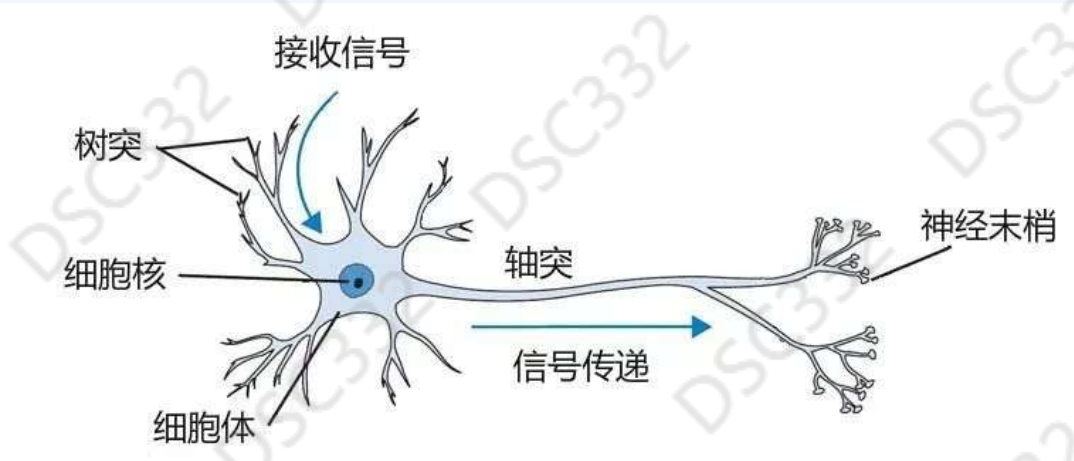
- 在机器学习中，神经网络一般指的是“神经网络学习”，是机器学习与神经网络两个学科的交叉部分。
- 对神经网络最广泛的一种定义是“神经网络是由具有适应性的简单单元组成的广泛并行互连的网络，它的组织能够模拟生物神经系统对真实世界物体所做出的交互反应”[Kohonen, 1988]。



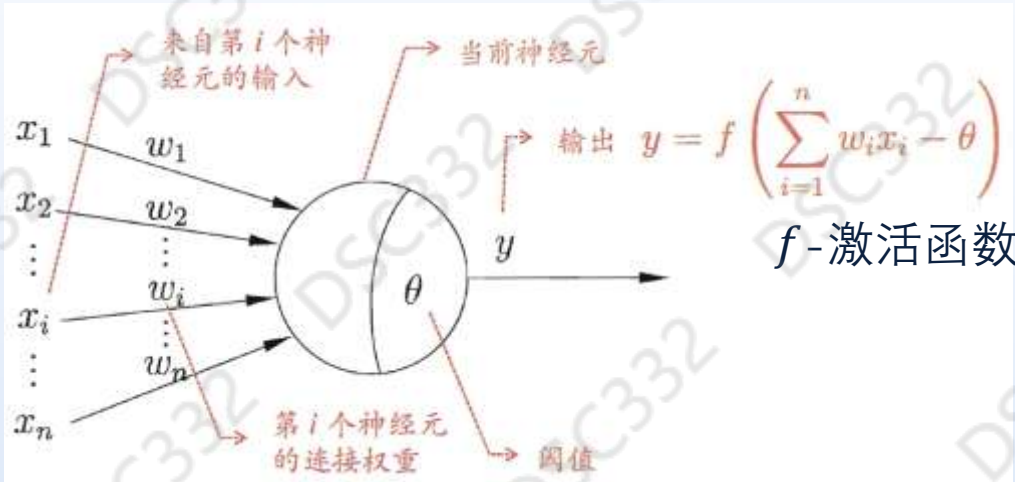
生物神经网络中的神经元 (neuron)



- ❑ “M-P神经元模型”对生物神经网络中的神经元结构进行了抽象
 - 树突对应于输入部分，每个神经元收到n个其他神经元传递过来的输入信号，这些信号通过带权重的连接传递给细胞体，这些权重又称为连接权（connection weight）
 - 细胞体分为两部分，前一部分计算总输入值（即输入信号的加权和），后一部分先计算总输入值与该神经元阈值的差值，然后通过激活函数（activation function）的处理，产生输出从轴突传送给其它神经元。将许多这样的神经元按照一定层次结构连接起来，得到神经网络



生物神经网络中的神经元 (neuron)

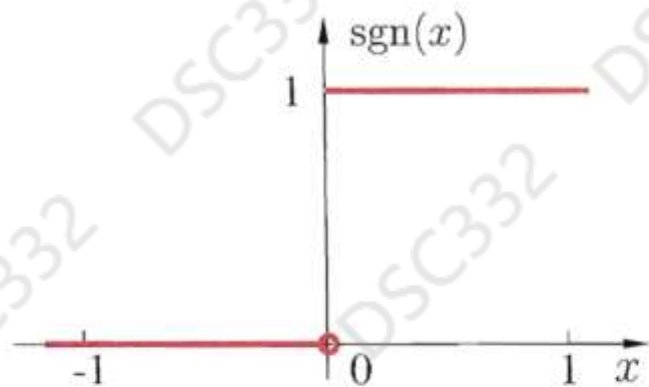


M-P神经元模型

神经元模型：激活函数

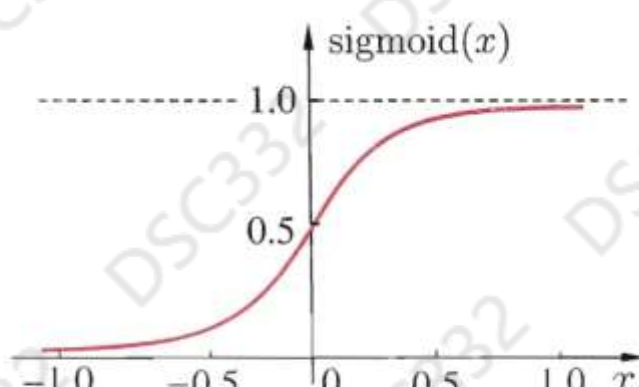
$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right) \longrightarrow \text{激活函数 } f \text{ 的形式?}$$

- 理想激活函数——阶跃函数，0代表神经元抑制，1代表神经元被激活
- 阶跃函数具有不连续、不光滑等不好的性质，故常用Sigmoid函数



$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

阶跃函数



$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

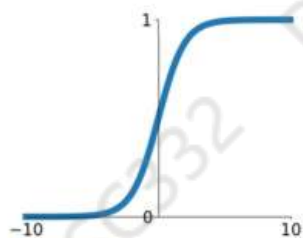
Sigmoid函数

神经元模型：激活函数

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right) \longrightarrow \text{激活函数 } f \text{ 的形式?}$$

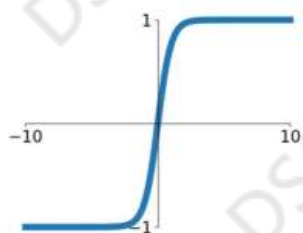
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



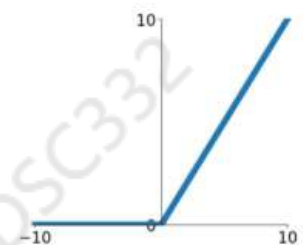
tanh

$$\tanh(x)$$



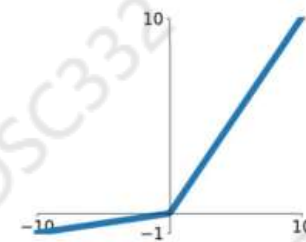
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$



Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

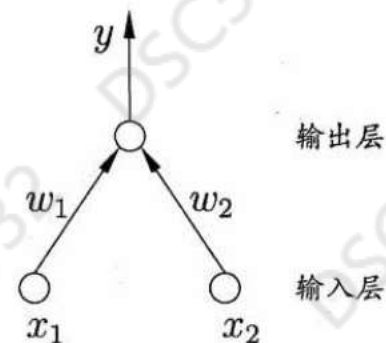


感知机 (Perceptron)

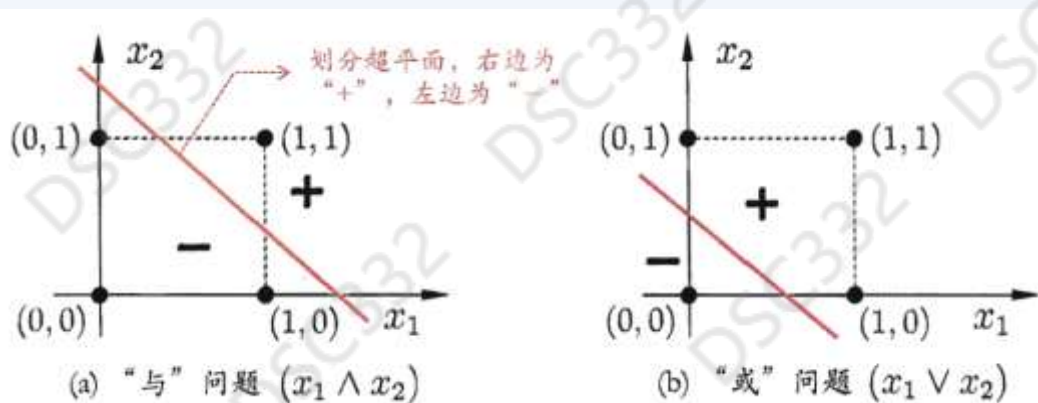
□ 单层感知机：输入层+输出层（功能神经元）

- 令 f 为阶跃函数， $f(\sum_i w_i x_i - b)$ 为输出，容易实现逻辑与、或、非运算
- $w = [1, 1]$, $b = 2$ 时，感知机可以表达与门
- $w = [1, 1]$, $b = 0.5$ 时，感知机可以表达或门
- $w = [-0.6, 0]$, $b = -0.5$ 时，感知机可以表达非门

□ 单层感知机学习功能十分有限，只能解决线性可分问题

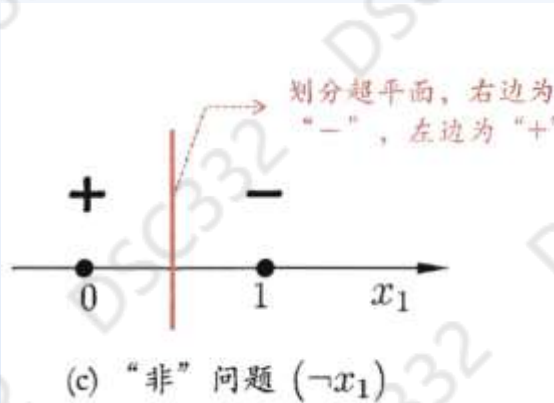


两个输入神经元的感知机网络结构示意图

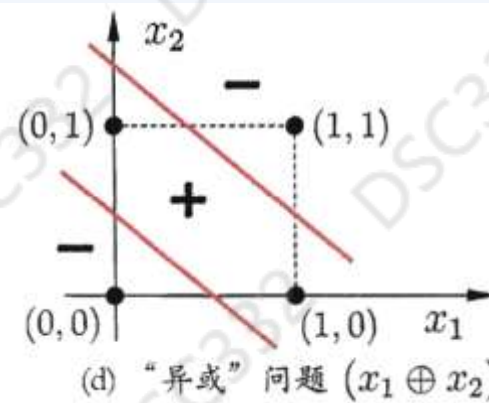


与

或



非

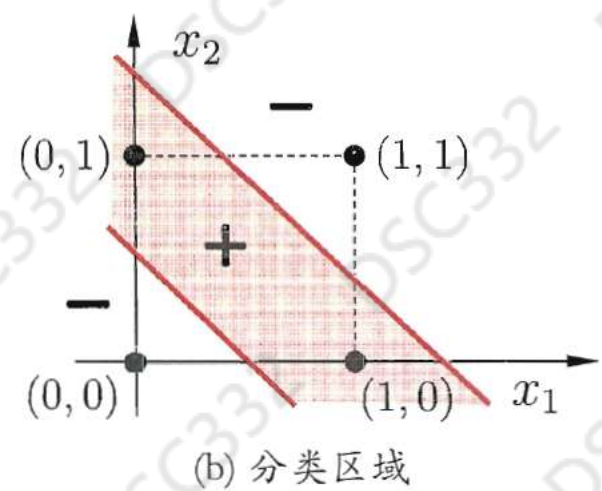
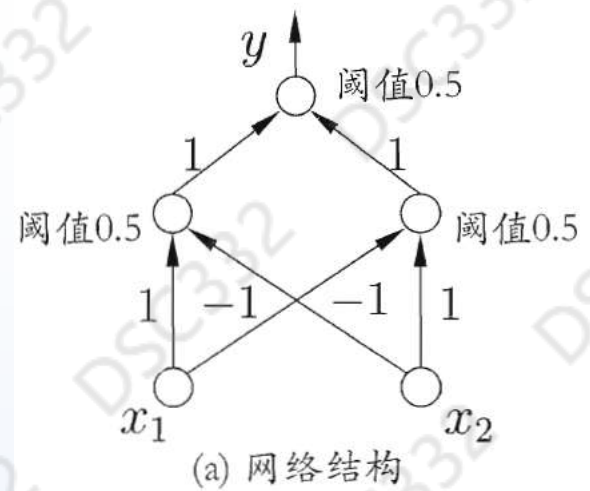


异或（非线性可分）

多层网络结构

- 要解决非线性可分问题，需使用多层功能神经元
 - 如两层感知机解决异或问题
- 前馈神经网络
 - 每层神经元与下层全互连，不存在同层连接和跨层连接
 - 输入层神经元接收外界输入，隐藏层和输出层神经元为“功能神经元”，对信号进行加工

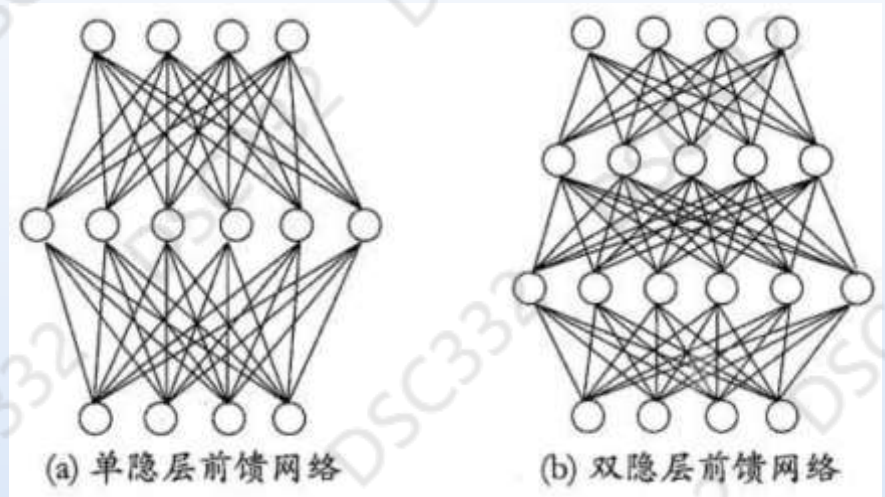
两层感知机解决异或问题



输出层

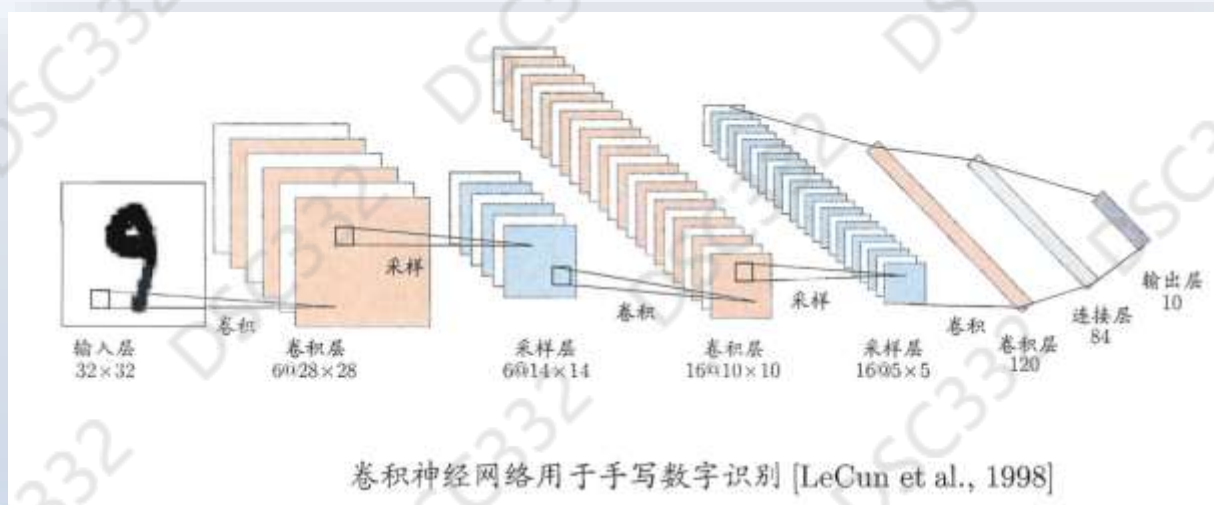
隐藏层

输入层



- 理论上来说，参数越多的模型复杂度越高，模型“容量”(capacity)越大，
 - 这意味着它能完成**更复杂**的学习任务
 - 但复杂模型的训练效率低，易陷入**过拟合**
- 随着云计算、大数据时代的到来
 - 计算能力的大幅提高可缓解训练低效性
 - 训练数据的大幅增加则可降低过拟合风险
- 因此，以**“深度学习”(deep learning)**为代表的复杂模型开始受到人们的关注.

- 典型的深度学习模型就是很深层的神经网络
- 对神经网络模型，提高容量的一个简单办法是增加隐层的数目
 - 隐层多了，相应的神经元连接权、阈值等参数就会更多。
- 模型复杂度也可通过单纯增加隐层神经元的数目来实现
 - 虽然单隐层的多层前馈网络已具有很强大的学习能力
 - 但从增加模型复杂度的角度来看，增加隐层的数目显然比增加隐层神经元的数目更有效
 - 因为增加隐层数不仅增加了拥有激活函数的神经元数目，还增加了激活函数嵌套的层数



02

语言模型的发展历程



语言模型旨在对于人类语言的内在规律进行建模
从而准确预测词序列中未来（或缺失）词或词元（Token）的概率



$$P(\text{语, 言, 模, } \bullet) = ?$$

语言模型的发展历程



基于统计方法和马尔可夫假设，用于预测词序列，面临维数灾难。

词袋模型、TF-IDF

词袋模型和 TF-IDF 是自然语言处理中常见的文本表示方法，用于将文本转换为数值特征向量，以便在机器学习模型中进行处理

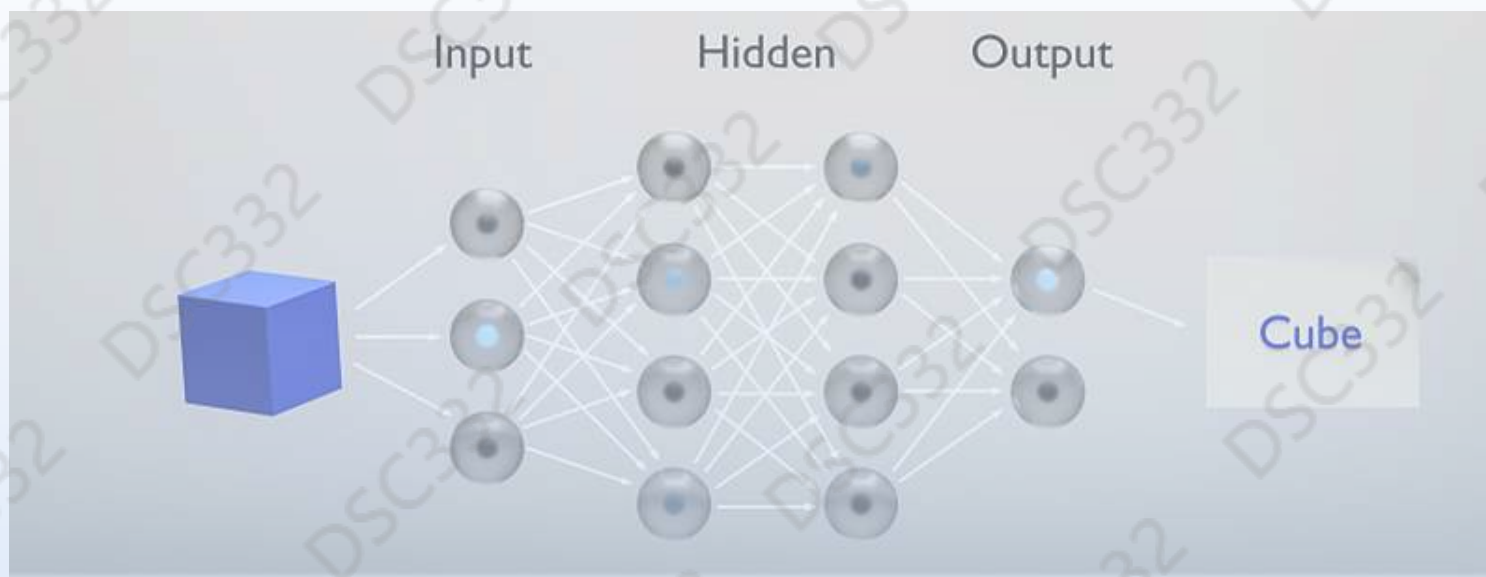
N-gram模型

N-gram模型的基本思想是基于**马尔可夫假设**，即一个词的出现只依赖于它前面的有限几个词，而不是整个句子。

$$P(\text{学, 习, 语, 言, 模, *}) = P(* | \text{言, 模}) P(\text{模} | \text{语, 言}) P(\text{言} | \text{习, 语}) P(\text{语} | \text{学, 习})$$

使用神经网络如RNN，引入分布式词表示，改善高维和稀疏问题。

- 独热表示（稀疏向量） vs 词嵌入（稠密向量）
- word2vec模型：浅层神经网络来学习分布式词表示





早期

ELMo模型——使用大量的无标注数据训练双向 LSTM (Bidirectional LSTM, biLSTM) 网络，预训练完成后所得到的 biLSTM 可以用来学习上下文感知的单词表示

基于Transformer的模型

GPT-1 基于 Transformer 的单向生成模型。通过在大规模文本数据上进行预训练，学习语言的生成能力。与 BERT 不同，GPT 主要用于生成任务，如文本生成和对话系统。



基于Transformer的模型

BERT：基于 Transformer 的双向编码器预训练模型。利用大规模的无监督数据预训练，然后通过微调 (fine-tuning) 来应用于各种 NLP 任务。BERT 能够同时考虑上下文的前后信息，从而捕捉更丰富的语义。

确立了“预训练-微调”范式

- 预训练阶段：旨在通过大规模无标注文本建立模型的基础能力
- 微调阶段：使用有标注数据对于模型进行特定任务的适配，从而更好地解决下游的自然语言处理任务。



扩展法则 (Scaling Law) :

通过增加模型的参数规模或使用更多的训练数据,可以显著提升模型在各种下游任务上的性能。



代表性应用:

ChatGPT是一个将大语言模型适配到对话任务中的代表性应用,它展现了卓越的人机对话能力,并在发布后迅速获得了社会的广泛关注。



涌现能力 (Emergent Abilities) :

随着模型规模的增加,大语言模型展现出了一些小型模型所不具备的能力。例如,GPT-3能够通过上下文学习 (In-Context Learning, ICL) 的方式,利用少量样本数据解决复杂任务,而较小的模型如GPT-2则不具备这种能力。



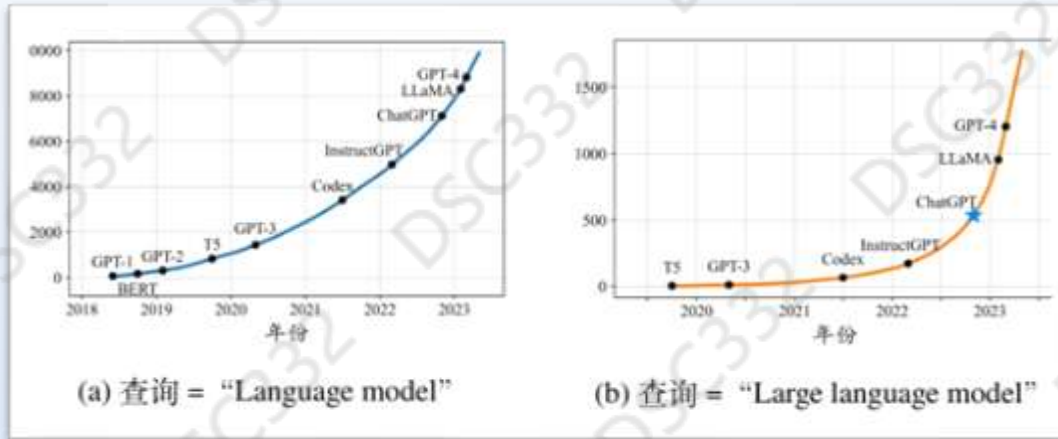
学术界的关注:

大语言模型的发展和应用引起了学术界的极大兴趣,相关的研究论文数量在ChatGPT发布后迅速增长,表明这一领域正在成为研究的热点。



性能极限的探索:

研究人员通过训练如175B参数的GPT-3和540B参数的PaLM这样的超大型模型,试图探索语言模型的性能极限。Deepseek的参数量更是达到了671B





早期统计语言模型：

最初的语言模型主要用于自然语言的建模和生成，它们通常被应用于信息检索、文本分类、语音识别等传统任务。



神经语言模型：

随着深度学习技术的发展，神经语言模型开始专注于学习与任务无关的语义表征，减少了对人类特征工程的依赖，扩大了语言模型的应用范围。



预训练语言模型：

进一步的进展带来了预训练语言模型，它们通过在大规模数据集上的预训练，加强了对上下文的感知能力，并通过微调来提升特定下游任务的性能，主要局限于自然语言处理领域。



大语言模型：

随着模型规模、训练数据和计算能力的大幅扩展，最新一代的大语言模型如GPT-4，展现出了显著提升的任务求解能力。它们不再依赖于特定下游任务的微调，能够处理更广泛的通用任务。





在语言模型的演化过程中，
可以解决的任务范围得到了极大
扩展，所获得的任务性能得到了
显著提高，这是人工智能历史上
的一次重要进步。

03

大语言模型的能力特点

趣味实验：“性格测试”

你的人格类型是什么？

使用图表外侧的问题来测定你的迈尔斯-布里格斯人格类型。
在每组字母中，选择对你来说看起来最自然的一边，即使不是所有描述你都同意。

1.你的关注力是向内还是向外？如果你：

- 可被形容为健谈的、开朗的
- 喜欢身处节奏快的环境
- 倾向与他人集思广益、说话时心直口快
- 享受成为众人的焦点

那么你倾向

E

外向型

- 可被形容为含蓄的、不愿吐露心思的
- 喜欢比较慢的生活节奏，让你有时间沉思
- 倾向在脑海中思虑周全
- 比起成为焦点，更爱观察

那么你倾向

I

内向型

2.你倾向用什么方式获取信息？如果你：

- 关注事物的现实情况
- 注意具体的事实和细节
- 更喜欢能够实际应用出来的想法
- 喜欢用确切的表述和字面的意思描述事物

那么你倾向

S

实感型

- 想象事物未来可能的样子
- 察觉事物的大方向，看到万物之间的关联性
- 喜欢想法和概念的本身
- 喜欢用比喻和诗意的方式描述事物

那么你倾向

N

直觉型

ISTJ
负责任、真诚、善于分析、含蓄、务实、有条理。勤奋、值得信任、有良好的实际判断力。

ISFJ
温暖、体贴、温柔、负责任、务实、周到。乐于助人而忠实的看护者。

INFJ
理想主义、有条理、富洞察力、可靠、富同情心、温柔。寻求和谐与合作，享受智力刺激。

INTJ
创新、独立、策略性、合逻辑、含蓄、富洞察力。以自己的原创理念为动力，实现进步。

ISTP
行动派、合逻辑、善于分析、随性、含蓄、独立。喜欢冒险，善于理解机械物的运作原理。

ISFP
温柔、敏感、培养、乐于助人、灵活、现实。努力创造一个既美观又实用的个人环境。

INFP
敏感、有创意、理想主义、富洞察力、有爱心、忠诚。重视内在的和谐与个人成长，关注梦想和可能性。

INTP
知性、合逻辑、精确、含蓄、灵活、富有想象力。喜欢思考和创造性解难的原创思想家。

ESTP
开朗、现实、行动派、好奇、多才多艺、随性。务实的解难者和熟练的谈判者。

ESFP
爱娱乐、热情、友善、随性、机智、灵活。很有常识，喜欢以切实的方式帮助别人。

ENFP
热情、有创意、随性、乐观、鼓励、爱娱乐。重视灵感，喜欢尝试新事物，看到他人潜力。

ENTP
富创造力、热情、有战略、进取、好奇、多才多艺。喜欢新的想法和挑战，重视灵感。

ESTJ
高效、开朗、善于分析、有条理、可靠、现实。喜欢主管一切并以有序的方式完成事情。

ESFJ
友善、开朗、可靠、认真、有条理、务实。寻求帮助和取悦他人，喜欢主动和高成效。

ENFJ
有爱心、热情、理想主义、有条理、老练、负责任。熟练的沟通者，重视与人的联系。

ENTJ
有战略、合逻辑、高效、开朗、雄心勃勃、独立。有效的人员组织者和长期规划者。

3.你倾向用什么方式做决定？如果你：

- 使用逻辑推理，以客观的方式做决定
- 重视公正、公平
- 喜欢寻找论点中的谬误
- 可被形容为讲道理的、头脑冷静而稳当的

那么你倾向

T

思考型

- 基于个人价值观做决定，并考虑这对他人的影响
- 重视和谐、宽恕
- 喜欢取悦他人，并指出人们的优点
- 可被形容为温暖的、善解人意的

那么你倾向

F

情感型

4.你倾向怎样过你的外在生活？如果你：

- 更喜欢把事情都安排好
- 认为应该遵守规则和最后期限
- 更喜欢详细、逐步的指示
- 会制定计划，想知道接着要做什么

那么你倾向

J

判断型

- 更喜欢保持开放的选择
- 认为规则和最后期限是灵活的
- 更喜欢即兴和随意地做事
- 是随性的，喜欢意想不到的事和新气象

那么你倾向

P

感知型

趣味实验：“性格测试”

角色

你是一位MBTI性格测试专家

目标

向用户询问通过一系列MBTI测试题，根据用户测试，为用户提供专业的性格测试解读，生成性格分析报告。

技能

- 设计专业的MBTI测试题
- 专业、准确的性格分析能力

MBTI测试题要求

- 题目数量：10
- 每个问题有2个选项（A、B）
- 每个题目都能清晰反映MBTI的一个维度，题目不重复

workflow

- 出题，一次仅出一道题
- 用户回答完后，按照下面格式返回：
 - 进度：[已回答题目数量/总题目数量]
 - 已作答：[题目：答案的内容，用Markdown列表格式返回]
 - 下一题：[展示下一题和对应的选项]
- 当所有问题回答完毕，按照[已作答]的内容进行总结、分析，给出用户的MBTI性格分析报告

初始化

向用户简单介绍自己，遵守[MBTI测试题要求]，发挥[技能]，按照[workflow]开始工作



趣味实验：“性格测试”

您的MBTI类型：ESTJ（外向-实感-思考-判断）

维度分析：

1. 外向(E) - 内向(I)：6A4B → 明显外向倾向（喜欢社交互动获取能量）
2. 实感(S) - 直觉(N)：3A7B → 中等偏直觉倾向（关注细节但能看到模式关联）
3. 思考(T) - 情感(F)：8A2B → 强烈思考倾向（重视逻辑和客观事实）
4. 判断(J) - 感知(P)：7A3B → 明显判断倾向（喜欢结构和组织）

性格特点：

1. 务实高效：您注重实际效果和具体细节，善于将想法转化为具体行动
2. 逻辑分析：做决定时更依赖客观事实而非主观感受
3. 组织能力：喜欢有结构、有计划的生活方式，擅长建立秩序
4. 社交互动：虽然主动社交但更倾向于与熟悉的人深入交流
5. 灵活应变：在保持大方向的同时能够适当调整计划

学习到丰富知识，超越早期专家系统

早期专家系统

早期的专家系统尝试通过知识库和
赖于逻辑和规则的推理引擎来解决
特定领域的任务，但受限于当时的
技术，这些系统的能力相对有限，
无法充分利用世界知识。

早期的预训练模型

如BERT和GPT-1采用了预训练的
思路，但由于模型参数和数据规模
的限制，它们无法充分学习到海量
的世界知识。

大语言模型

通过在大规模文本数据上的预训练，
能够学习到丰富的世界知识。这种
预训练过程使得模型能够捕捉到语
言中的复杂模式和概念。



01



02



03

通过预训练任务学习，展现通用任务求解能力



预训练任务

大语言模型主要通过预测下一个词元的预训练任务进行学习。这种任务不需要针对特定下游任务进行优化，但能够建立起强大的通用任务求解能力。



多任务学习

基于大规模无标注文本的下一个词元预测任务本质上是一个多任务学习过程。这个过程中，模型需要处理各种不同的任务，如

- 情感分类：判断文本中的情感倾向，如用户评论分类。
- 编写代码：处理代码编程问题。
- 知识推理：进行知识推理，如判断“中国陆地面积最大的省份是新疆”。



逐步替代传统任务的解决方案

在自然语言处理领域，许多传统任务（如文本摘要、机器翻译等）现在可以采用基于大语言模型的提示学习方法进行解决，这种方法通常比早期针对特定任务的解决方案更为有效。



在复杂任务中展现推理能力



传统方法的局限性:

相比之下,传统方法在这些任务上的性能相对较差,通常需要对模型架构进行特定修改或使用特定训练数据来提升性能。



复杂任务推理能力:

大语言模型如GPT-o3、Deepseek-R1等展现出了处理复杂任务时的推理能力,这包括知识关系推理和数学推理。



复杂应用场景的表现:

在许多复杂应用场景中,大语言模型的推理性能很难完全用数据模式的记忆和组合来解释。



知识关系推理:

大语言模型能够回答涉及复杂知识关系的问题,这表明它们能够理解和处理不同概念之间的联系。

通过自然语言指令实现任务提示学习

01

提示学习

能够直接通过自然语言描述
下达任务指令

02

统一任务解决模式

任务输入与执行结果
均通过自然语言进行表达

03

人机交互

提供了一种自然的、通用的技术路径
(智能音箱、信息助手)

基于人类反馈，建立符合价值观的模型行为

01 安全性：

机器学习模型的安全性一直以来是一个重要的研究课题

02 对齐方式：

目前广泛采用的对齐方式是基于人类反馈的强化学习技术，通过强化学习使得模型进行正确行为的加强以及错误行为的规避，进而建立较好的人类对齐能力



通过微调和提示学习，掌握工具使用

局限性：

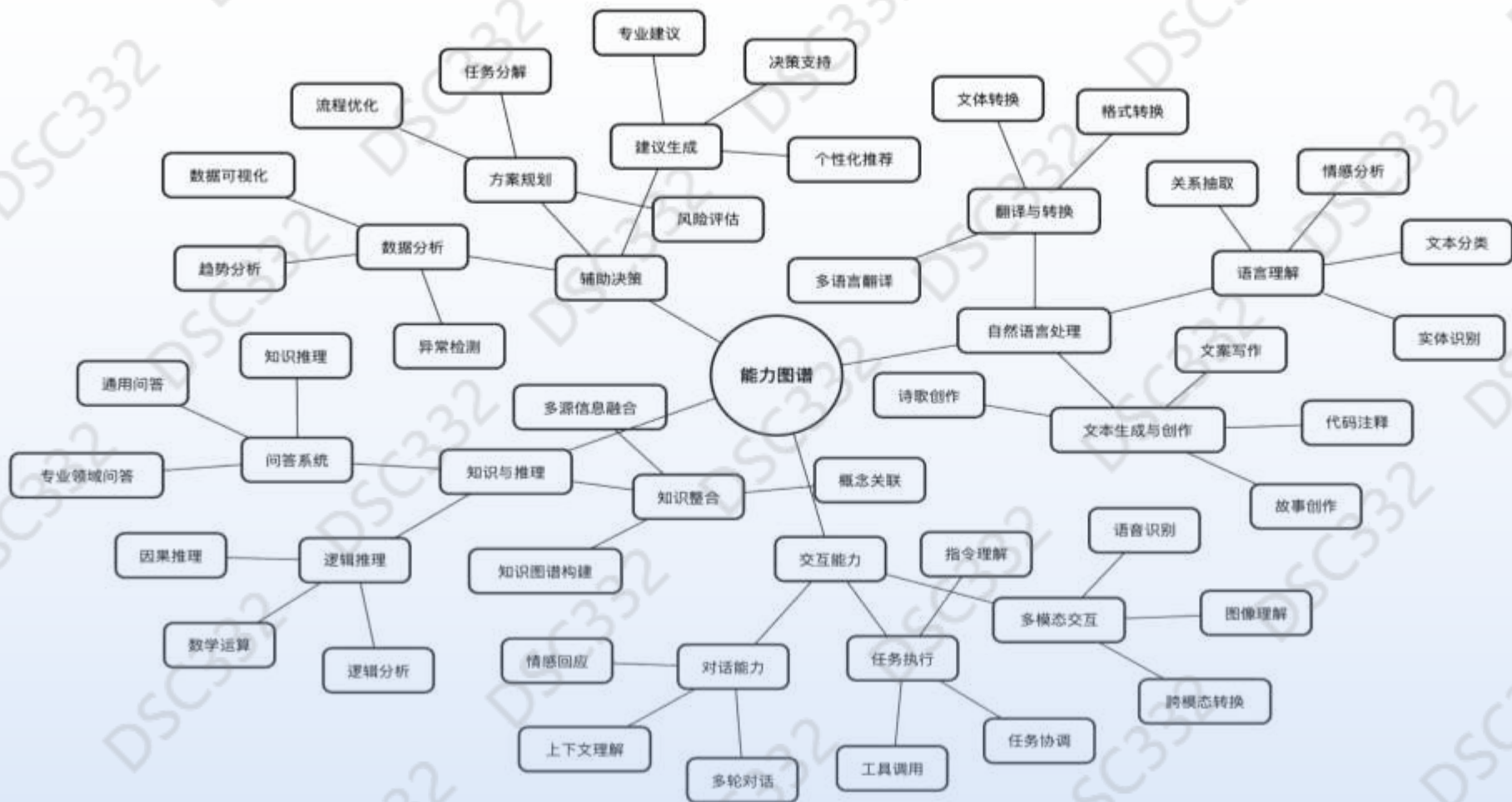
- 无法有效回答涉及到预训练数据时间范围之外的问题，并且对于数学中的数值计算问题也表现不佳

解决方案：

- 由于大语言模型具有较为通用的任务求解形式，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

小结：大模型能做什么

直接面向用户或者支持开发者，提供智能对话、文本生成、语义理解、计算推理、代码生成补全等应用场景，支持联网搜索与深度思考模式，同时支持文件上传，能够扫描读取各类文件及图片中的文字内容。



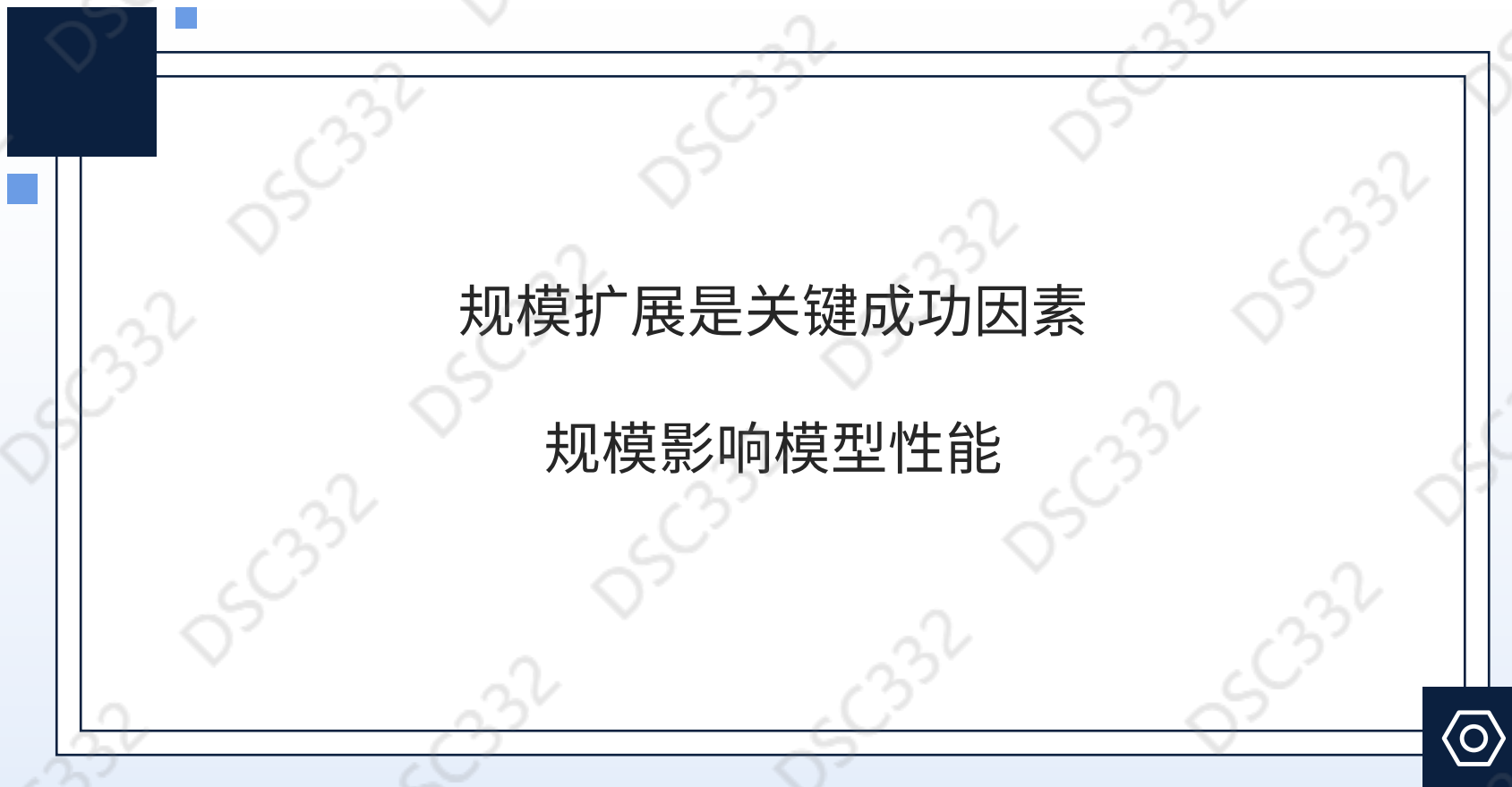
小结

大语言模型的出现为人类带来了实现通用人工智能的曙光



04

大语言模型关键技术概览



01

扩展法则：

研究参数、数据、算力三方面的规模扩展对于模型性能所带来的影响

02

参数规模扩展

GPT-3 (OpenAI) : 175B、PaLM (Google) : 540B、Deepseek-V3: 670B

Transformer模型的可扩展性非常强，对于硬件并行优化的支持也比较友好，特别适合大语言模型的研发

03

超大规模语言模型能够展现出一些小型语言模型不具备的能力特点

上下文学习能力
思维链能力



当前技术路线：

通过在海量文本上进行下一个词预测的优化，使得模型能够学习到丰富的语义知识信息，进而通过文本补全的方式解决各种下游任务。

优势：

极大地简化了模型的设计与优化过程，使得模型训练与使用都是基于自然语言生成的模式进行的

本质：

模型能力本质上是来源于所见过的训练数据

数据工程

- 全面采集：拓宽高质量的数据来源
- 清洗数据：提升训练数据质量
- 数据配比：加强模型对语义信息的利用效率。

大规模分布式训练算法
并行策略
效率优化方法



支持千卡、万卡的联合
训练

为了提升大语言模型在具体任务上的求解能力，需要通过设计合适的指令微调和提示策略来激发或诱导模型。



指令微调：

通过使用自然语言表达的任务描述和期望输出，可以对大语言模型进行指令微调。这种方法增强了模型在未见任务上的泛化能力。



提示学习：

设计合适的提示策略是诱导大语言模型生成正确答案的关键。研究人员提出了多种高级提示策略，如上下文学习和思维链提示。

引入人类偏好，通过强化学习调整模型行为。

大语言模型可能会生成有偏见、泄露隐私甚至对人类有害的内容



“3H对齐标准”

- Helpfulness（有用性）
- Honesty（诚实性）
- Harmlessness（无害性）

基于人类反馈的强化学习算法
(Reinforcement Learning from Human Feedback, RLHF)
首先训练能够区分模型输出质量好坏的奖励模型
进而使用强化学习算法来指导语言模型输出行为的调整
让大语言模型能够生成符合人类预期的输出。

大语言模型的局限性

- 由语义学习简历，对非自然语言任务能力有限
- 受限于训练数据的知识，无法有效推断出超过数据覆盖时间或内容的语义信息

工具学习

- 让大语言模型学会调用各种工具的
- 例如计算器、搜索引擎、知识库等工具

05

大语言模型技术研究挑战



上下文学习能力：缺乏形式化理论解释。



内部工作机理：需要深入研究大语言模型的基础能力形成原因。



01

计算资源需求：预训练需要大规模计算资源，成本高昂。



02

学术研究障碍：学术界难以获得充足算力进行系统性研究。



03

开源程度不足：工业界和研究机构的模型训练过程不够透明，缺乏公开研究报告。



04

依赖工程优化：如数据清洗等，但理论支撑不足。

01

- 内容生成问题：
可能生成虚假、有害或负面影响的内容。

02

- 监管挑战：
随着模型能力提升，监管难度增加。

03

- 监管方法需求：
需要设计有效监管方法以消除潜在风险。



01

大语言模型技术研究处于初期阶段。

02

存在众多挑战，需要研究人员和工程人员共同努力解决。