

大模型概述+发展史



大模型概念解析

大模型的定义与特征

01

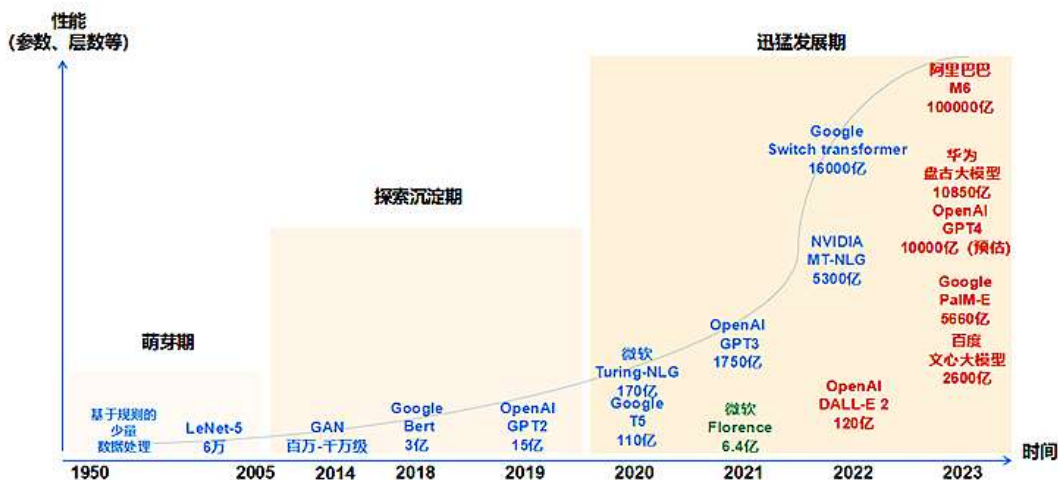
- (1) 定义：使用**大规模数据集**训练的深度学习模型，具有**非常高的参数数量**和计算能力。
- (2) 参数规模：**数十亿甚至数千亿**个参数
- (3) 特点：它们能够处理和分析大量的数据，从而生成复杂的输出，如自然语言文本、图像等。
- (4) 重要特征：它们具有强大的泛化能力，能够在多种任务中表现出色。
- (5) **独立意义上的大模型：具备涌现能力的深度学习模型**

大模型与传统模型的区别

02

- (1) **小模型**通常指参数较少、层数较浅的模型，它们具有轻量级、高效率、易于部署等优点，适用于数据量较小、计算资源有限的场景，例如移动端应用、嵌入式设备、物联网等。
- (2) 与传统模型相比，大模型拥有更多的参数和更复杂的结构，能够处理更复杂的任务和数据。
- (3) 与传统模型相比，大模型的训练需要更多的计算资源和时间，但一旦训练完成，它们可以用于解决多种问题。
- (4) 传统模型通常用于解决特定任务，大模型则更适用于**数据量较大、计算资源充足**的通用场景，例如云计算、高性能计算、人工智能等。

大模型的发展历程



(1) 萌芽期 (1950-2005)

以CNN为代表的传统神经网络模型阶段

- 1956年，约翰·麦卡锡提出“人工智能”概念
- 1980年，卷积神经网络的雏形CNN诞生
- 1998年，现代卷积神经网络的基本结构LeNet-5诞生

(2) 探索沉淀期 (2006-2019)

以Transformer为代表的全新神经网络模型阶段

- 2013年，自然语言处理模型 Word2Vec诞生
- 2014年，被誉为21世纪最强大算法模型之一的GAN诞生
- 2017年，Google颠覆性提出基于自注意力机制的神经网络结构—Transformer
- 2018年，OpenAI和Google分别发布了GPT-1与BERT大模型

(3) 迅猛发展期 (2020-至今)

以GPT为代表的预训练大模型阶段

- 2020年，OpenAI公司推出了GPT-3
- 2022年11月，搭载了GPT3.5的ChatGPT横空出世
- 2023年3月，最新发布的超大规模多模态预训练大模型—GPT-4
- 2024年4月，最新升级超大规模多模态预训练大模型—GPT-4 Turbo
- 2024年4月19日，Meta正式发布开源大模型——Llama-3
- 2024年5月9日，阿里云发布通义千问2.5，并开源Qwen-110B(国产Llama3)
- 2024年5月14日，OpenAI发布最新多模态大模型 GPT-4o (o-omni全能)
- 2024年7月23日，Meta开源大模型—Llama3.1 405B, 多项指标超越GPT-4o
- 2024年7月25日，Mistral Large 2(123B), 多项指标超越GPT-4o和Llama3.1
- 2024年9月25日，Meta推出Llama3.2，提供了多模态视觉和轻量级模型



传智教育旗下高端IT教育品牌