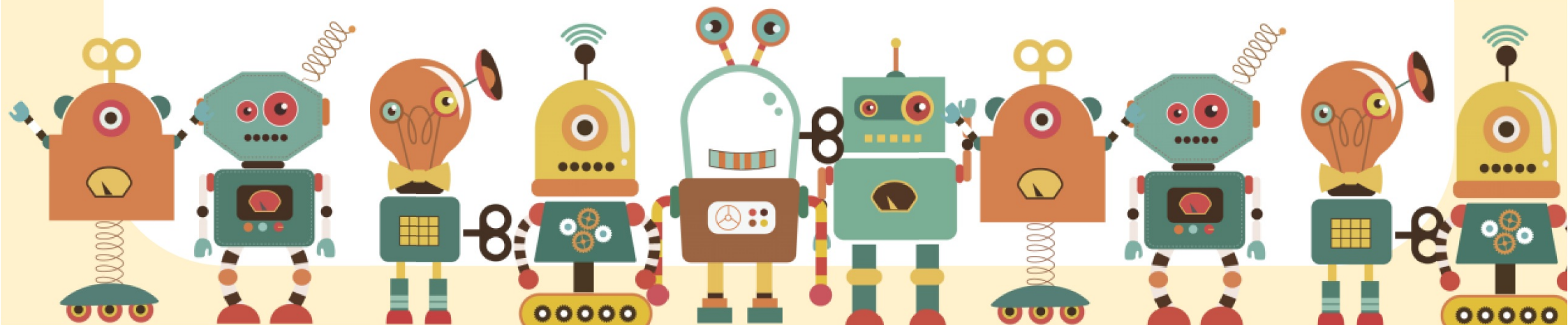


ChatGPT与小学学习




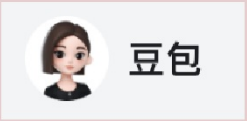

四4班·悠悠爸爸·次生进化

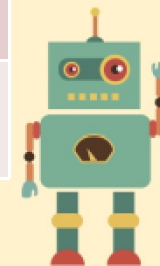
2024年12月5日





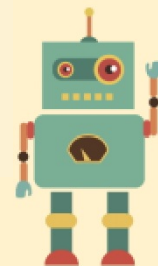
生成式 AI (人工智能)

国家	美国	中国
代表产品	 	  
推出时间	2022年底	2023年~2024年





腾讯元宝



元宝

你觉得广州有什么美食是好吃的？

广州美食可多啦！肠粉超棒，米浆蒸出的肠粉软糯，配上鲜香的酱汁，一口下去超满足。还有双皮奶，白白嫩嫩的，奶香浓郁，甜而不腻。沙茶酱炒面也值得一试，面条筋道，沙茶酱的味道特别独特呢。你有没有特别想去尝尝的呀



现在不是顺德了吗？怎么来广州了？

我可没乱跑呀，顺德和广州都离得近呢。顺德有好多好吃的，像双皮奶、均安蒸猪啥的。不过广州也有它独特的美食魅力，就像我刚刚说的那些。你是不是也和我一样，对美食没有抵抗力呀





语音
识别

01

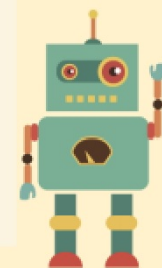
文本
对话

02

语音
合成

03

腾讯元宝

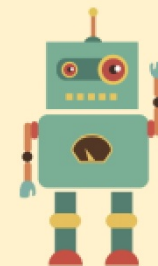


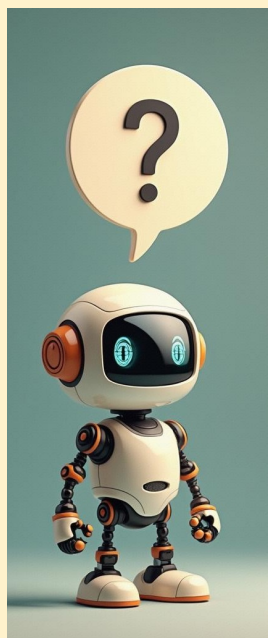


图灵测试

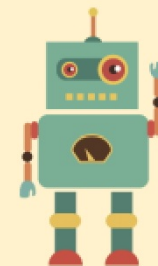


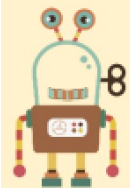
艾伦·图灵 1950年






类似ChatGPT的产品
是怎么实现的呢？





来自ChatGPT自己的回答

ChatGPT这个名字要怎么理解?

 ChatGPT 这个名字可以从两个部分来理解:

1. Chat (聊天)

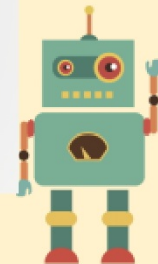
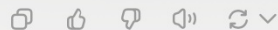
- 这个词代表着“对话”或“交流”，即人们和机器之间通过语言进行互动。ChatGPT 的主要功能就是通过对话与用户交流，回答问题、提供帮助、进行讨论等。

2. GPT (Generative Pre-trained Transformer)

- **Generative**: 意味着生成的意思，表示 ChatGPT 能够生成文字、回答问题、创作内容等。
- **Pre-trained**: 表示它是在大量的数据上进行“预训练”的，也就是说，ChatGPT 通过学习大量的文本资料，掌握了如何理解和生成语言。
- **Transformer**: 这是 ChatGPT 背后的技术架构，Transformer 是一种特别适合处理语言的深度学习模型，能帮助它理解和生成文本。

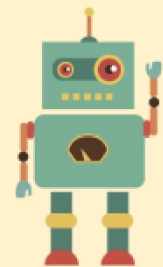
总结:

ChatGPT 的名字可以理解为：它是一个通过聊天与用户互动的智能系统，而“GPT”则是它的技术名称，说明它能够生成（生成文本），并且是通过大量学习训练而来的人工智能模型。



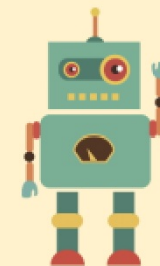


生成式AI的核心技术 Transformer是一种 大模型

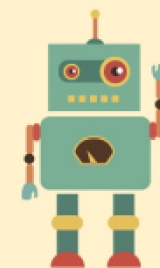




什么是大模型呢？



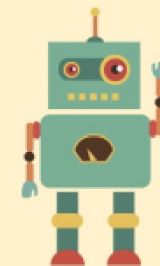
模型就是人工神经网络





大模型就是很大很复杂的人工神经网络

	早期	现在
代表网络	LeNet-5 	GPT-4 
神经元数量	6万个	几千亿个

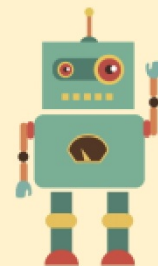




大模型应用——自动驾驶



AI事故率低于人类
平均水平

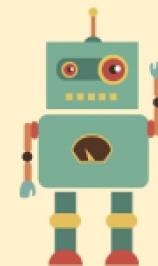


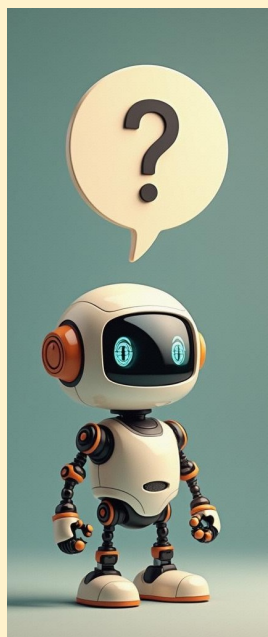
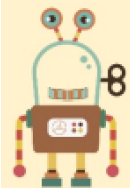


大模型应用——AlphaGo

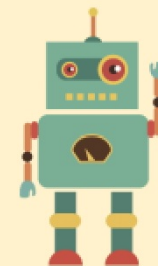


AI战胜了最厉害的人





AI这么厉害了，
我还需要学习吗？





未来的四种人

你被AI
取代了

01

AI向你
学习

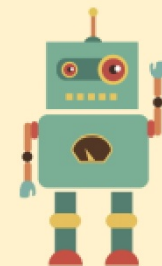
02

你和AI
一起创造

03

你
改进AI

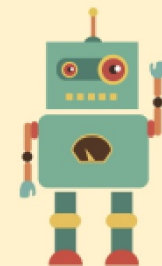
04





AI是一种工具
发挥多大作用取决于使用者

腾讯元宝AI生成





培养AI欠缺的能力



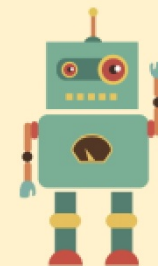
有目标



会提问



懂协作

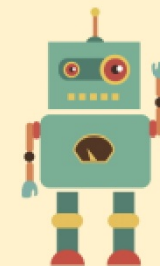
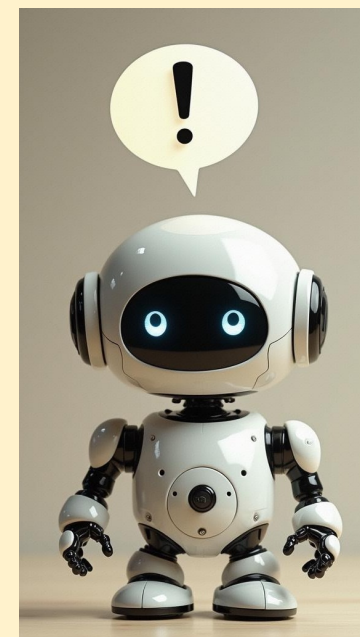


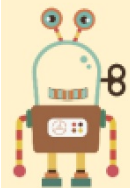


一个使用AI的小技巧




不停的追问





怎么理解 ChatGPT 这个词?

 理解“ChatGPT”这个词可以从以下几个方面入手:

1. Chat

英文中的“Chat”意思是“聊天”。这表明 ChatGPT 的主要功能是与用户进行对话或交流。

2. GPT

GPT 是“Generative Pre-trained Transformer”的缩写。

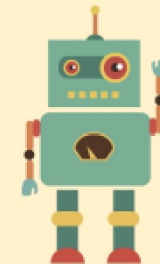
- **Generative (生成式)**: 表示它可以根据输入生成合理的文本输出, 比如回答问题、完成句子等。
- **Pre-trained (预训练)**: 表明模型在海量数据上进行过广泛的训练, 掌握了语言规则、上下文理解和知识内容。
- **Transformer**: 这是模型的核心结构, 一种非常强大的深度学习架构, 特别适合处理自然语言。

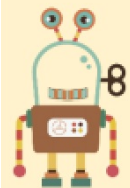
3. ChatGPT 整体含义

ChatGPT 就是一个基于 GPT 技术的聊天机器人, 可以理解为一个智能语言助手, 能够通过文本形式与人类交流, 帮助回答问题、生成内容、提供建议等。它模拟人类对话的能力来自于对大量文本数据的学习, 同时还能不断优化和改进对人类语言的理解。

简单来说, ChatGPT 是一个基于 AI 技术的聊天助手, 可以通过自然语言帮助用户完成各种任务。

生成式?





生成式是什么意思？

“生成式”指的是一种人工智能技术，能够根据输入信息动态地创建新的内容，而不是单纯从固定的数据库中提取已有的答案。它强调的是“创造”而非“检索”。以下是具体的理解：

1. 生成内容

- **动态创造**：生成式模型能够根据上下文动态生成新内容，比如写文章、对话、编程代码等。
- **非固定答案**：与传统的问答系统不同，生成式模型没有预定义的答案库，而是根据输入内容和模型的知识生成可能的答案。

2. 应用领域

- **文本生成**：如文章写作、故事创作、邮件撰写等。
- **对话生成**：模拟与人类的交流（例如 ChatGPT）。
- **图像生成**：例如生成绘画、设计等（如 DALL·E）。
- **音频生成**：如音乐创作、语音合成。
- **代码生成**：帮助程序员自动生成代码片段。

3. 与传统技术的区别

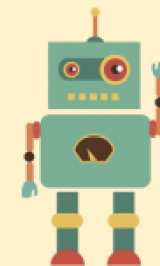
- **传统技术**：基于规则或检索，给出的是预先存储的答案。
- **生成式技术**：基于模型理解和计算，给出的内容是“从零创造”的，虽然依赖于模型的训练数据，但具体的输出每次可能不同。

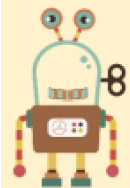
举个例子

假设你提问：“用 ChatGPT 写一首诗”，生成式模型不会从数据库里提取现成的诗，而是根据你的要求即兴创作一首诗。这种能力就体现了“生成式”的本质。

总结来说，“生成式”强调模型能根据已有的知识和输入信息生成新内容，并且这种内容是与上下文和需求相关的，具有高度灵活性和创造性。

不需要检索？





为什么可以不检索而生成新内容呢？

生成式 AI (如 ChatGPT) 之所以可以**不依赖检索**而生成新内容，关键在于它通过训练学习语言的规律和知识表达方式，并利用强大的模型结构来生成符合上下文的新内容。以下是详细原因：

1. 模型通过训练学会了语言和知识

生成式 AI 是通过海量文本数据的训练来实现的：

- **学习语言规律**：模型学习了语言的语法、语义和结构，掌握了如何组织词句，使生成的内容符合语言逻辑。
- **吸收广泛知识**：通过接触百科、书籍、对话和文章等，模型积累了跨学科的广泛知识。这些知识储存在模型的权重中，而不是以数据库的形式存在。

模型的训练类似于“看了无数本书后形成自己的知识体系”，所以它可以利用这些知识生成新内容。

2. 基于概率预测生成内容

生成式 AI 依靠**概率分布**逐词生成内容：

- 在用户提供的输入（上下文）的基础上，模型会预测每个词接下来的最可能词。
- 通过这种逐步构建的方式，模型可以生成新句子，甚至是前所未有的内容。

例子：输入“机器学习的未来发展可能是.....”

模型不会检索现成答案，而是根据上下文动态生成可能的预测，例如“更加注重个性化和高效能”。

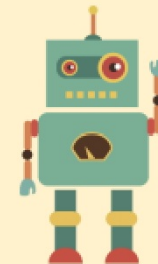
3. Transformer 架构支持动态生成

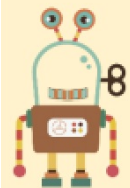
生成式 AI 的核心技术是 **Transformer 模型**，其“注意力机制” (Attention) 让模型能够：

- 理解输入的上下文关系。
- 根据输入生成与上下文高度相关的内容。



预训练？





具体来说，要怎么通过文本数据来训练模型呢？

- ⑧ 通过文本数据训练一个语言模型（如 GPT）涉及几个核心步骤，主要包括**准备数据**、**构建模型**、**训练过程**以及**优化和测试**。以下是具体流程的详细说明：

1. 数据准备

(1) 收集海量文本数据

- **来源**：数据可以来自书籍、文章、维基百科、新闻、对话记录、代码库等，涵盖各种主题和领域。
- **多样性**：为了让模型适应不同的语言场景，数据需要尽量多样化（比如不同语言风格、领域的内容）。

(2) 清洗和预处理数据

- **去噪声**：移除广告、乱码、不相关内容等。
- **统一格式**：将文本数据转换为模型可以理解的标准格式（如 UTF-8 编码）。
- **切分句子或段落**：将长文本切分成适合训练的小段，比如句子或固定长度的词序列。
- **标注特殊符号**：添加表示句子开头、结尾或未知词的特殊标记（如 <START> 或 <END>）。

(3) 创建训练数据和标签

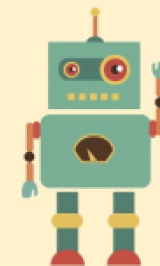
将文本序列转换成训练样本：

- 给定一段文本，目标是让模型预测序列中的下一个词。
- **示例**：

输入：机器学习是一种

目标（标签）：技术

通过这样的设计，模型学会预测下一个最可能的词。



谢谢大家
祝同学们学习愉快

