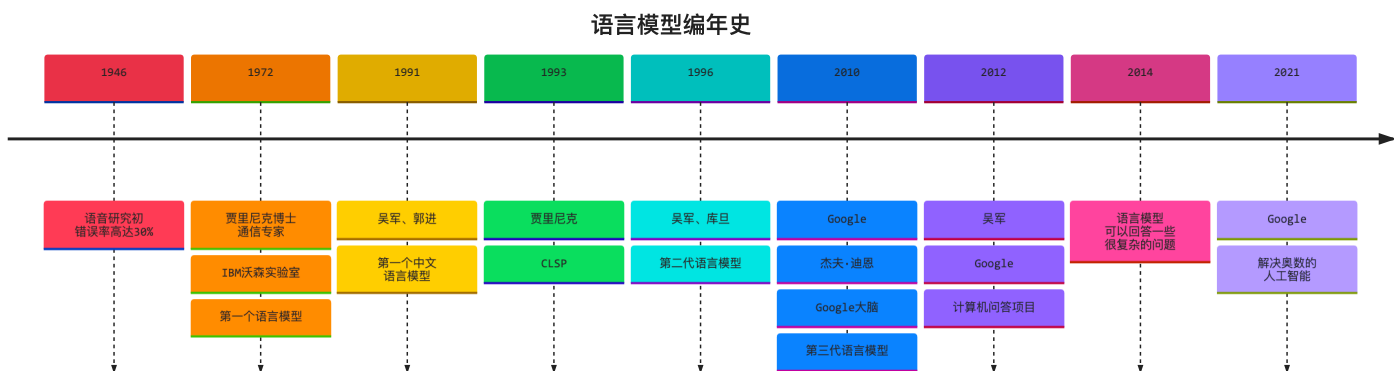


# 总述

ChatGPT是一个被神化的语言模型的衍生品。

## 语言模型编年史

1. 1946年，开始有人研究语音识别，错误率高达30%；
2. 1972年，贾里尼克博士，数学基础非常好的通信专家，IBM沃森实验室，第一个语言模型；
3. 1991年，吴军、郭进博士，第一个中文语言模型；
4. 1993年，贾里尼克，约翰霍普金斯大学语言与语音处理中心（CLSP）主任
  1. 吴军，1996年，学生，与库旦，第二代语言模型；
5. 2010年，Google，杰夫·迪恩等人，第一个真正实用的深度学习系统（Google大脑），第三代语言模型；
6. 2012年，吴军，Google，计算机问答项目；
7. 2014年，语言模型可以回答一些很复杂的问题；
8. 2021年，Google，解决奥数问题的人工智能系统，得分优于有计算机博士学位的开发者；



## ChatGPT成神

### 神的特点

只有神具有让人们即赞誉，又恐惧的特点。

- 1953年，哥伦布，牙买加，预言月全蚀，被土著人供奉成神；

### ChatGPT为什么神奇

1. 能回答复杂的问题
2. 能写作

### 复杂问题的定义

1. 容易的问题
  - 跟事实有关的问题
    - 是谁
    - 在哪
    - 什么时候
  - 2008年前后已解决
2. 复杂的问题

- 两大类
  - 为什么
  - 怎么做
- 2014年，机器完美回答40%的“难的”问题
- 2023年，ChatGPT可以回答更多“难的”问题

## 关于计算机写作

### 事实

截止2022年底，相当多的纸媒文章，尤其是财经类的文章，主要由计算机完成。



### 为什么2023年前没有ChatGPT

1. 大公司没有将AI包装成ChatGPT类似的产品
2. 受限于法律，上市公司无法让AI提供不受限制的回答服务

## 目前ChatGPT使用

### 使用目的归纳

1. 一站式信息查询
2. 帮助写作
3. 应付工作中的公文类汇报

### 使用的核心

1. 理解自然语言 (prompt)
2. 生成自然语言

### 较好的适用范围

1. 格式固定的文章及数据

## 什么是语言模型

### 语言模型定义

对人类的语言建立的数学模型。

- 语言模型不是逻辑框架
- 语言模型不是生物学的反馈系统
- 语言模型是由数据公式构建的模型

### 为什么用数学构建语言模型

1. 可行
2. 捷径

# 现在通信原理

通信，即对信道的编码和解码。



## 重点 - 最接近

1. 人和人交流
  - 根据听到的不完全信息，结合上下文语义信息，来补全完整信息
2. 人与计算机交流
  - 根据**概率**计算哪一个语句可能性最大
  - 如何消除不确定性？
    - 利用更多的上下文信息
    - 互信息消除不确定性

## ChatGPT的本质

ChatGPT的原理，是在多种候选中，选择一个概率或者可能性最大的句子。

## 语言模型概率的计算

1. 早期的语言模型
  - 只看上下文
  - 统计相同的上下文，每个词句出现多少次
  - 语言模型，也称统计语言模型<sup>[1]</sup>
2. 今天的语言模型
  - 数据量上数量级的扩充
  - 用复杂的**机器学习**方法反复计算

## 用语言模型写唐诗

### 优势

1. 唐诗大多属于韵律诗，算是格式固定的数据（文本）

### 大致原理

1. 把律诗拆解成“2-3”或者“2-2-3”的组合
2. 以两个字或者三个字作为基本单元进行统计，得到上下文中的概率
3. 根据词的语义进行归类，让每一句都符合主题

## 语言模型的进化

### 阶段一

- 语音识别问题
  - 90年代之前
  - 贾里尼克
  - 错误率30%降到10%以内

- 英、法语自动翻译
  - 彼得·布朗
  - 缺少两种语言相互对应的数据

#### 问题：

- 单纯增加**训练数据**并不十分有效地增加语言模型的能力。

#### 解决思路：

- 用语言中更深的**知识和信息**建立语言模型

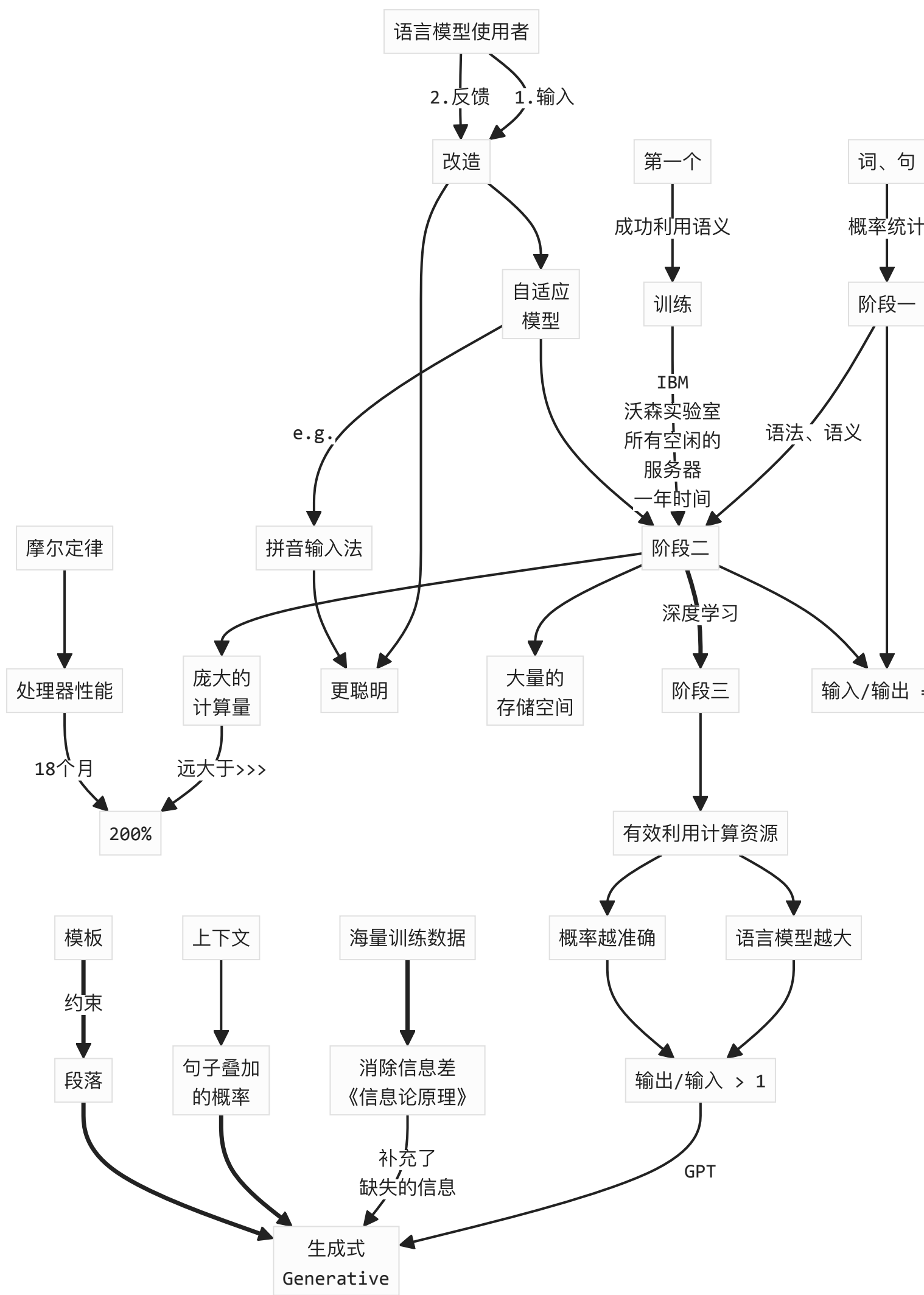
## 阶段二

- 增量信息：
  - 语法
  - 语义
- 举例：
  - 事实：时政新闻
  - 失事：历史题材

## 阶段三

- 深度学习
  - 更有效利用计算资源
  - 概率越准确
- 重点
  - 海量输入
  - 更复杂的语义、上下文概率计算

#### 语言模型进化图



ChatGPT的边界

ChatGPT真正可怕的地方在于，按照当前的速度发展下去，不断地扩大应用领域，它可能可以解决很多原本需要人类才能解决的问题。

## ChatGPT能做什么

ChatGPT能做什么，取决于**语言模型**能做什么。

### 转换信息形式

- 语音识别
- 机器翻译
  - 损失了文化属性
- 基因测序
  - 任何DNA都由ATCG四种碱基组成
- 写简单的计算机程序
  - 将自然语言**翻译**成机器语言
  - **难点**在于人很难将自己的想法用自然语言**准确**表达出来

### 按要求产生文本

- 示例
  - 回答问题
  - 回复邮件
  - 书写简单段落
- 重点
  - 输入小于输出时，消除信息的不确定性
  - 语言模型**自身**包含话题的**很多**相关信息
- 难点
  - 信息论：不可能通过少量信息得出更多信息
  - **或多或少**需要人工干预
- 让ChatGPT更准确
  - 人工迭代
- 优势
  - 减轻人的工作量

### 信息精简

输入大于输出，根据**信息论**原理，生成的质量取决于**算法好不好**。

- 难点：
  - 保留什么
  - 丢弃什么
- 比大部人好
  - 计算机阅读和处理更快
  - 语言模型强大
  - 比人客观
- 比少部人不好
  - 缺乏个性化
  - 流于产品线的标准件

### 前景预测

2019年，郭毅可院士预测

- 2024年，对描述非常清晰的任务进行编程；
- 2026的，完成中学生水平的作文；
- 2028年，编辑视频；
- 2049年，创作最畅销的小说

## ChatGPT不能做什么

???

## ChatGPT的答案从哪里来

什么是问题：在英语中，通常以疑问词开头，也叫WH单词，包括

- 简单问题
  - What 是什么
  - Who 是谁
  - When 在什么时候
  - Where 在哪里
  - Which 哪一个
- 复杂问题
  - Why 为什么
  - How 怎么做

## 简单问题

- 一个问题对应多个答案
- 根据概率返回最可能的答案

## 复杂问题

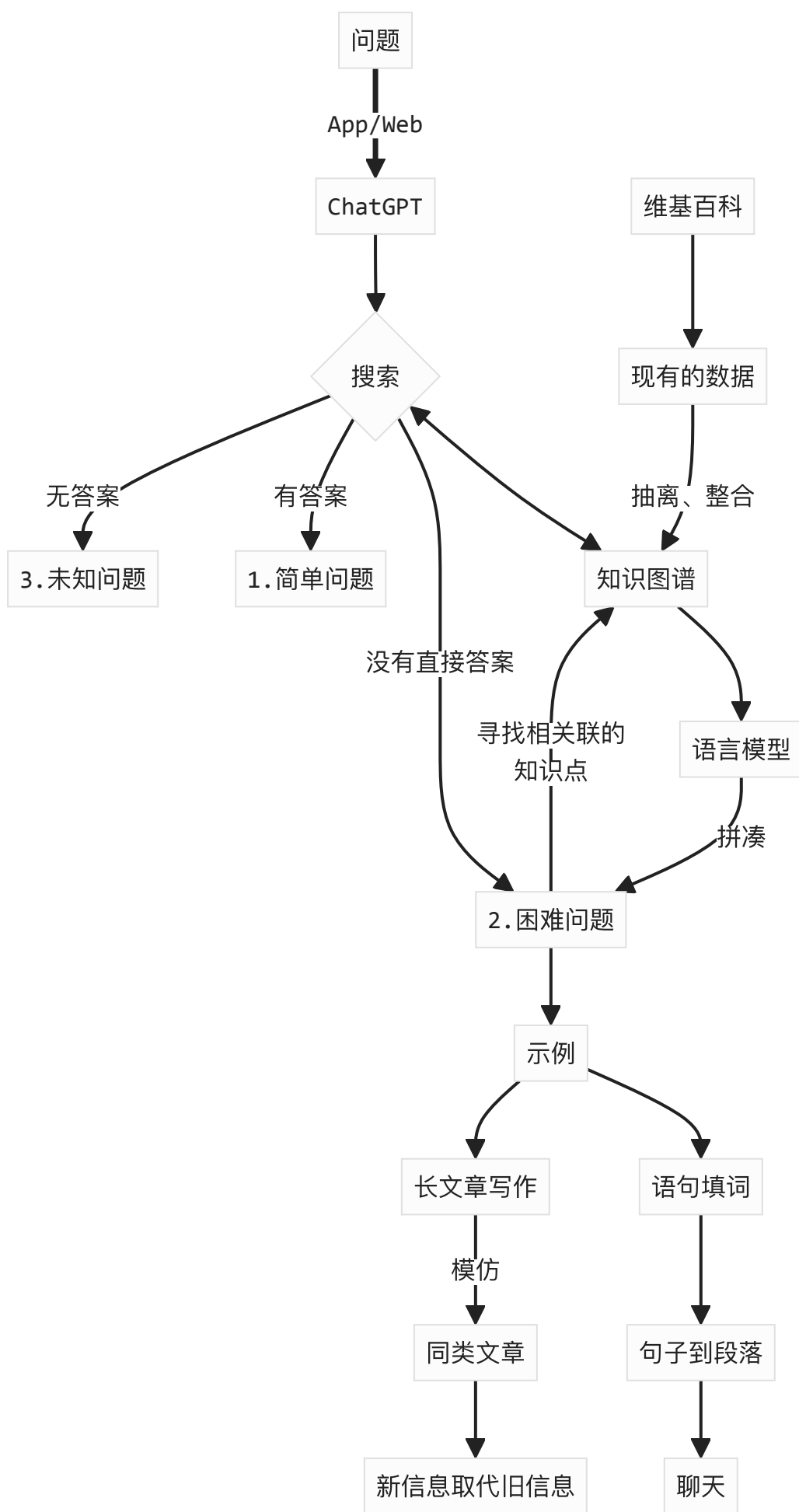
### 对比人和ChatGPT解决复杂问题的三种途径

问题困难程度	人的解决办法	ChatGPT的解决办法
简单问题	知道答案，直接给出；	在某个网站内直接抽离出一段话作为答案；
困难问题	不知道答案，根据基本知识，推导出答案；	依据模型里面的数据，对现在的答案进行归纳总结出新答案；
未知问题	不知道答案，也无法直接推导出答案，需要做研究；	需要实验、探索、创新去发现新知， <b>ChatGPT做不到</b> ；

### 为什么过去的搜索只能给出网站链接

1. 过去的计算机对文本的摘要能力不够
2. 过去的搜索很少分析句子的语义，无法确定问题和答案是否一致

## 如何工作



## ChatGPT有哪些问题

### 可以改进的问题

1. 因信息量不够造成的问题



- 人类创造新知后, ChatGPT可获得学习的数据
2. 文章段落乏味、不生动
    - 更优化的数据处理算法
  3. 离不开后台大量有人工干预、非常高的运行成本(能源、算力)
    - 计算机技术的革新

## 固有缺陷

1. GIGO(垃圾输入, 垃圾输出)
  - 根本原理: 信息论, 利用已知信息预测其它信息的模型
  - 垃圾数据(噪音)太多, 输出质量太差
    1. 噪音不太高时, 成倍增加训练的数据量
    2. 找到噪音来源, 过滤掉噪音
    3. 随机噪音, 无能为力
2. 别有用心的“正反馈”

## 人工干预的边界

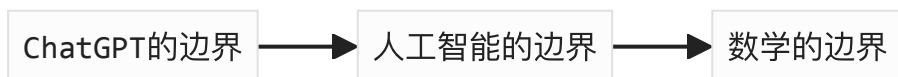
1. 上万亿参数的模型本身就是一个黑盒子, 很难通过人为调高或者调低一些参数来控制ChatGPT的输出结果
2. 人工干预的好恶影响模型的客观性, 导致更大的不公平

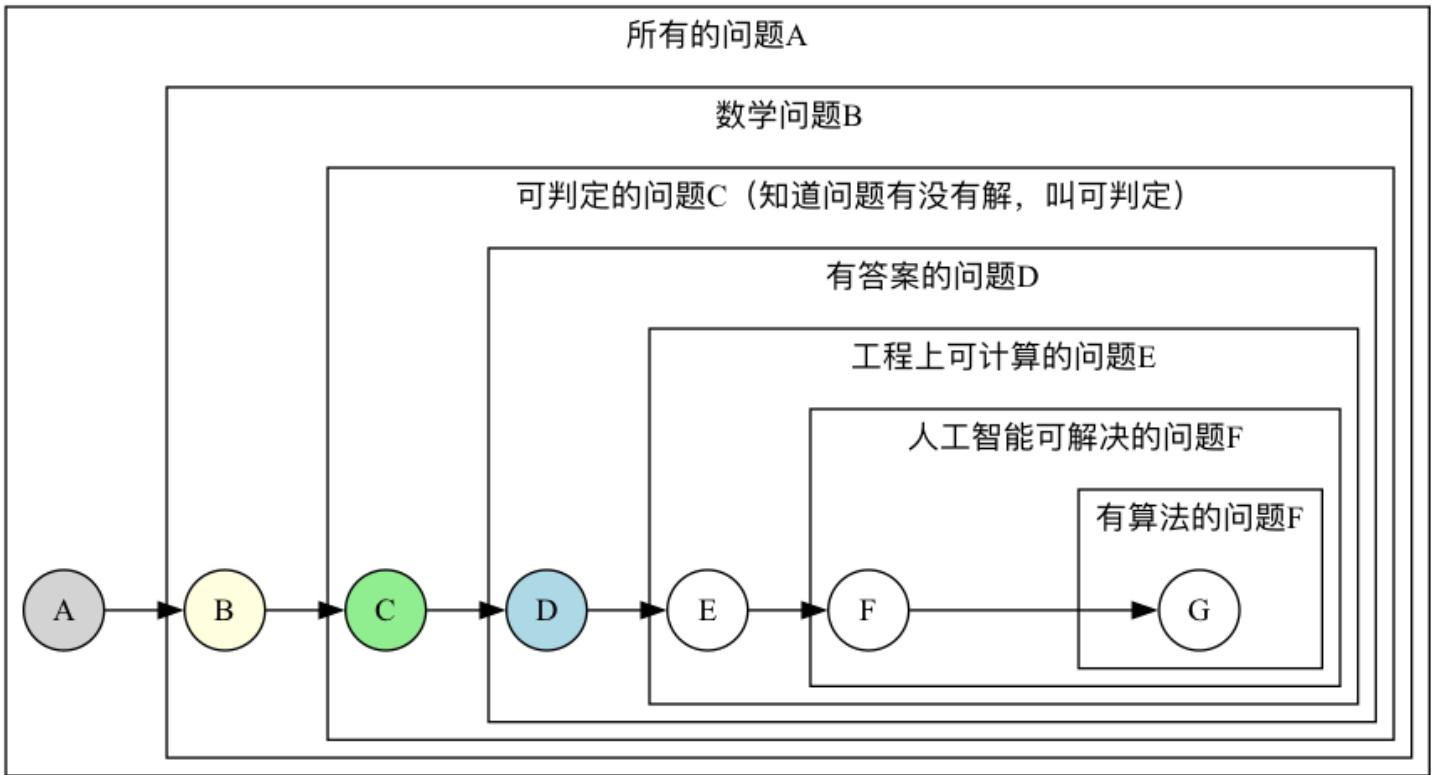
## ChatGPT需要哪些资源

- 数据
  - 数量极其庞大
  - 被少数企业、政府持有
- 算力
  - 存储
  - 处理
    - ChatGPT-3算力猜测
      - 超100PFLOPS的算力, 每秒10亿亿次运算以上
      - 整个特斯拉云计算集群的**两倍**
      - Google最大数据中心的1/3
      - 阿里云最大数据中心的1/4
- 算法
  - 聪明的算力
    - 计算资源的并行分配、调度
    - 深度学习的算法和基础架构的深度匹配
    - 对问题的理解能力和程度
  - 笨的算力

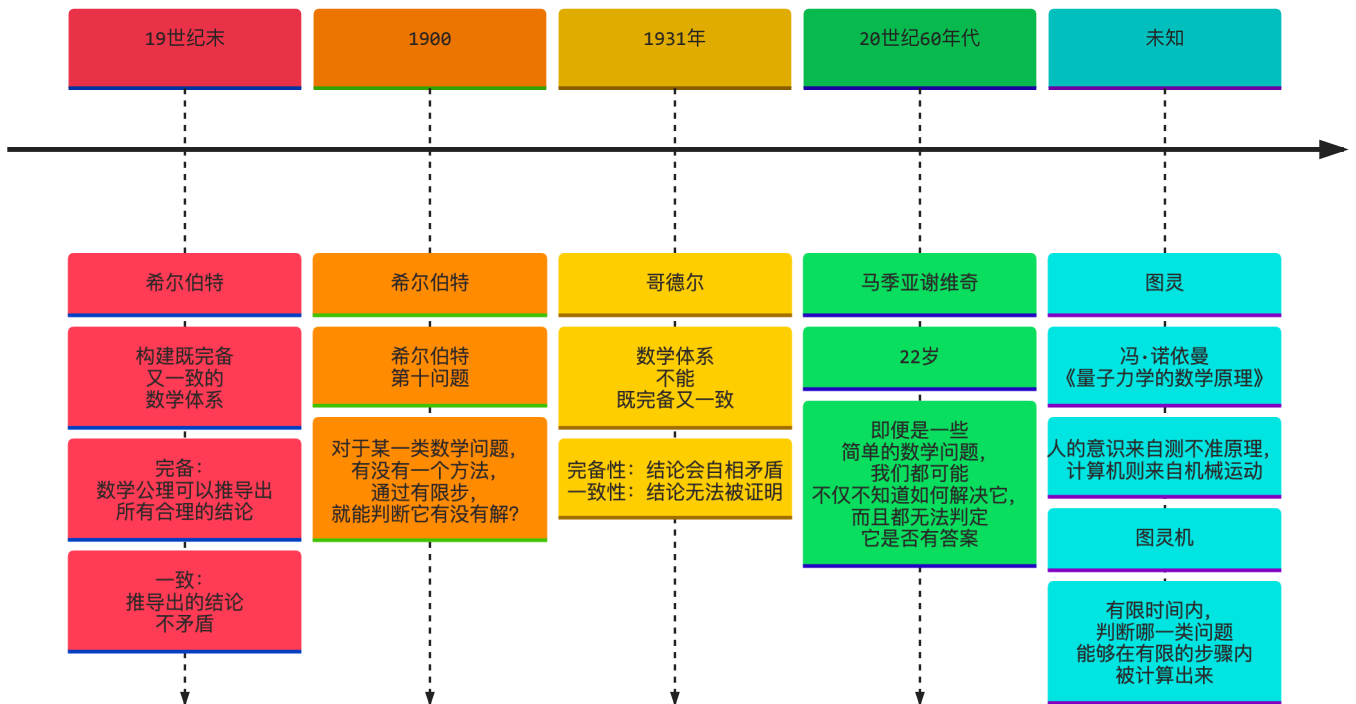
## 今天人工智能的边界在哪里

### AI是有边界的





## 数学的边界



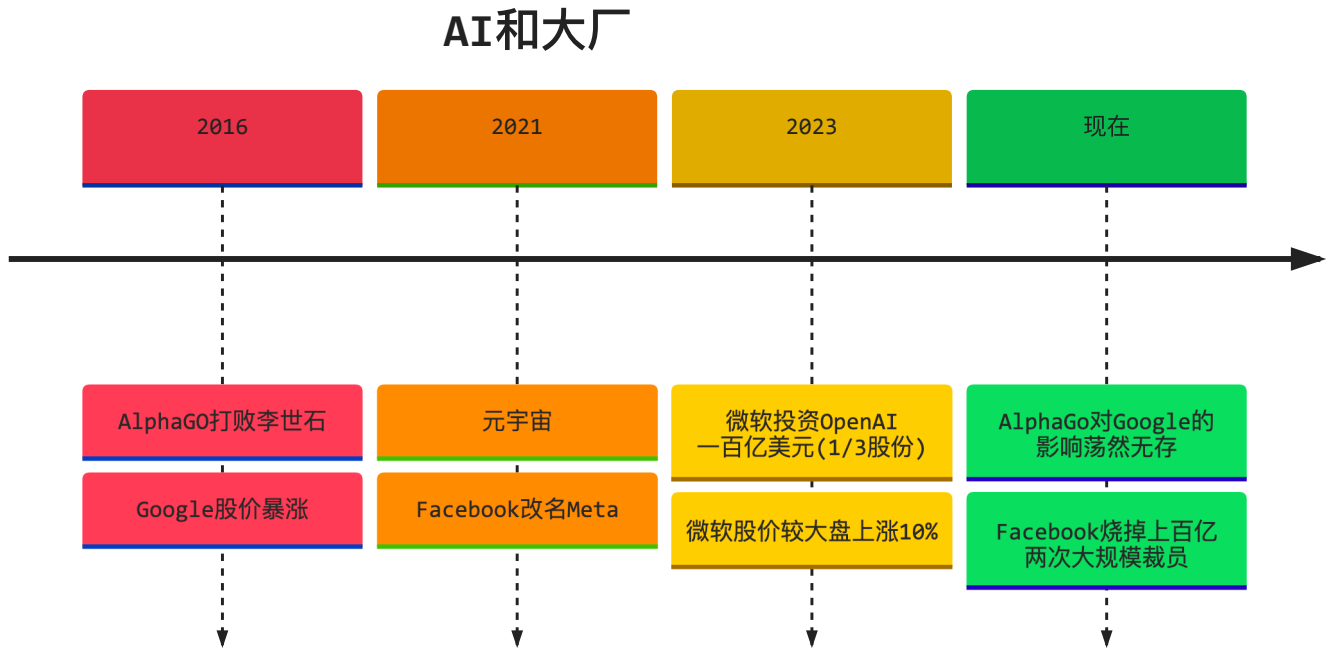
## 胡思乱想的人

1. 类型一：真无知
  - 叫作哥论布的神（牙买加月蚀）
2. 类型二：自大的AI从业者
  - 做出了具体贡献的人，会放大自己的成就

## ChatGPT为什么被热炒

### 相关信息

## 投资



## 从业者

1. 已经兴奋的从业者
  1. 以自己的业余爱好，和接受了七八年训练的专业人士争饭碗
  2. 螳螂捕蝉黄雀在后，赚钱者另有其人
2. 开始恐惧的从业者，好不容易赶上的潮头，一夜被拉开距离
  1. 如果ChatGPT是一个方向，早一天晚一天区别不大，追赶别人总比摸黑探索来得容易
  2. 第三眼美女：从概念原型到成熟的产品，通常也要七年之久，而且往往是不同的公司主持大局

## 媒体

1. 国外媒体
  1. ChatGPT只是一个玩具，还称不上是工具
  2. ChatGPT教人学坏
  3. ChatGPT产生假新闻和误导信息
2. 国内媒体
  1. 断章取义、吸引眼球
  2. 从VR、区块链、元宇宙到自动驾驶不断炒作新概念

## 普通人

1. 很多人因为知识储备不足，把ChatGPT尊为新神
  - 哥伦布 -- 月食 -- 印第安原住民 -- 封神
2. 过去十年里，当初把VR、区块链和元宇宙贬得一文不值的人，也是当初将它们封神的人
  - 麦哲伦 -- 菲律宾 -- 封神 -- 被原住民杀死
3. 不从事计算机行业的人大谈AI，要警惕新一轮泡沫
  - 1929经济大萧条，擦皮鞋的小童向肯尼迪父亲兜售股票

# ChatGPT能替代什么人

## 会被ChatGPT取代的人

- 会被取代的工作特点
  - 不费体力的工作
  - 不动脑子的工作
  - 不产出信息的工作
- ChatGPT特点
  - 根据现有信息，整合出信息，并非产出新信息（信息论）
- 列举会被取代的职业
  - 金融分析师
  - 大部分财经媒体记者
  - 律师助理
  - 普通文员
  - 短视频制作人
  - 大部分公众号写手
  - 低端的IT从业者
    - 短期培训上手的运营、测试、数据处理岗， etc.

## 不会被ChatGPT取代的人

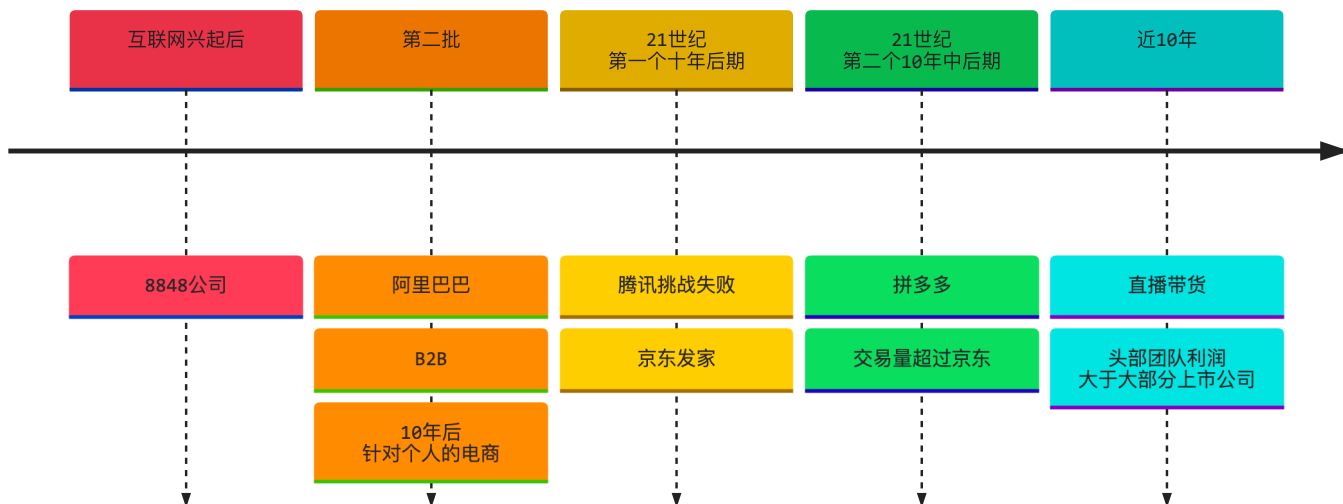
- 费体力的工作
  - 设备维修
  - 送餐
  - 保洁
  - 酒店服务
- 从事创造工作的人
- 觉得但凡有了智能技术就可以躺平的人

## 理性看待ChatGPT

### 投资人的角度

1. 泡沫可以让资金快速流入一个领域
2. 但凡能成为趋势，至少有几十年的发展机会

# 中国电商



## 从业者的角度

1. 见到一点新技术言必称机会，其实是想多了
2. ChatGPT只是人工智能革命其中一个小涟漪
3. 挤破头的公司越多，卖算力的越挣钱；加州淘金热时，只有卖水、卖工具、做牛仔裤的人发了财
4. 十多年前“百团大战”，广告公司挣得盆满钵满
5. 比特币热的时候，AMD、英伟达股票大涨

## 普通人的角度

1. 新技术+传统行业 = 行业+
2. 群体的命运取决于大环境，而不是某一项还看不清的技术
3. 面对未知，不要盲目造神
4. 未来不是给机会主义者的，而是给踏踏实实把本职工作做好的人的
5. 要担心的不是ChatGPT产生自主意识，而是躲在AI后面试图操控别人的公司和人

## 还有哪些人工智能和科技值得关注

### 多任务人工智能

1. GPT-3是以数量实现质量的代表
  1. 无法让ChatGPT下棋
  2. 无法让AlphaGo回答问题
2. 多任务的人工智能则是以质量取代数量
  1. 知其然还要知其所以然
    1. ChatGPT给出的答案或者AlphaGo下出的好棋，AI本身并不清楚为什么
    2. AI理解二义的代词需要上下文，但人不需要
  2. 具有对身边事物的好奇
    1. 人有街上会东张西望，想学点新东西，做事不顺利会复盘
    2. ChatGPT算了半天没算出来，它不会想自己知识有欠缺或者自己算法上有BUG
    3. 没有好奇心，就无法积攒普遍意义上的经验
3. 具有常识
  1. 人工智能的画往往直接、但没有意境
  2. 没有常识，很难完成复杂任务，比如让机器去买菜

#### 4. 处理信息的能耗降低几个数量级

1. 多任务不是几个系统的简单拼接，而是在内部打通各个任务的算法
2. 人脑百万亿个神经元连接，寿命不超过90年，换算成30亿秒，完成百万亿个神经元训练，非常高效
3. 计算机速度比人快，但训练一个只有几万个连接的深度人工智能神经网络就要几万秒，效率比人低了很多；如果把规模加大十倍，训练时间将成百上千的增加
4. 训练人脑这样的系统，能耗是难以想象的

## 生物和医学的应用

### 解构人体蛋白质

#### 计算机解决蛋白质结构预测

