

# 人工智能前瞻培训

黄琨 @ 中国科学院大学



AI 算法

+



互联网应用

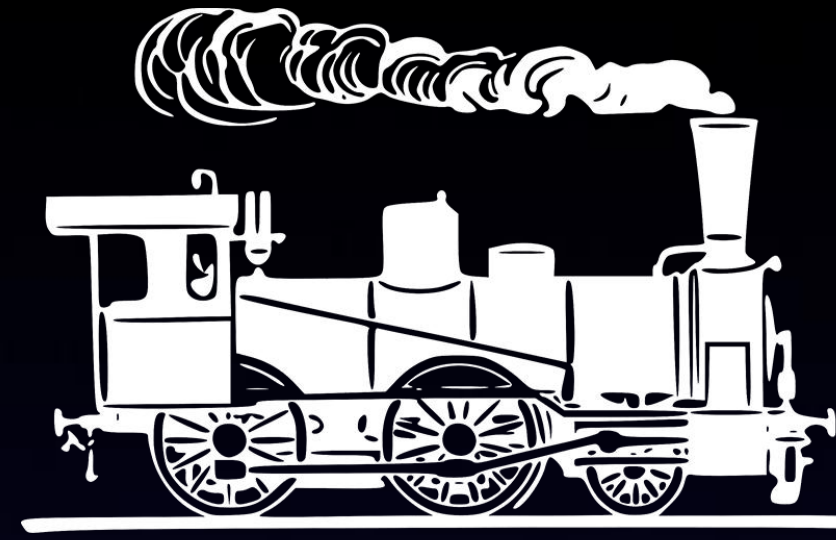
# Outline

- 为什么要认识 AI？（Artificial Intelligence 的价值）
- 什么是 AI？（理解AI 算法）
- AI 对我们生活的具体影响？（AI 的问题、机遇和门槛）
- Q & A



为什么要认识 AI ？

一次新的科技革命，影响整个社会



I：蒸汽时代



II：电气时代



III：信息时代



## IV: 智能时代

人工智能（AI）将是新的“电力资源”

（“Artificial Intelligence is the New Electricity” —— Andrew Ng）

交通



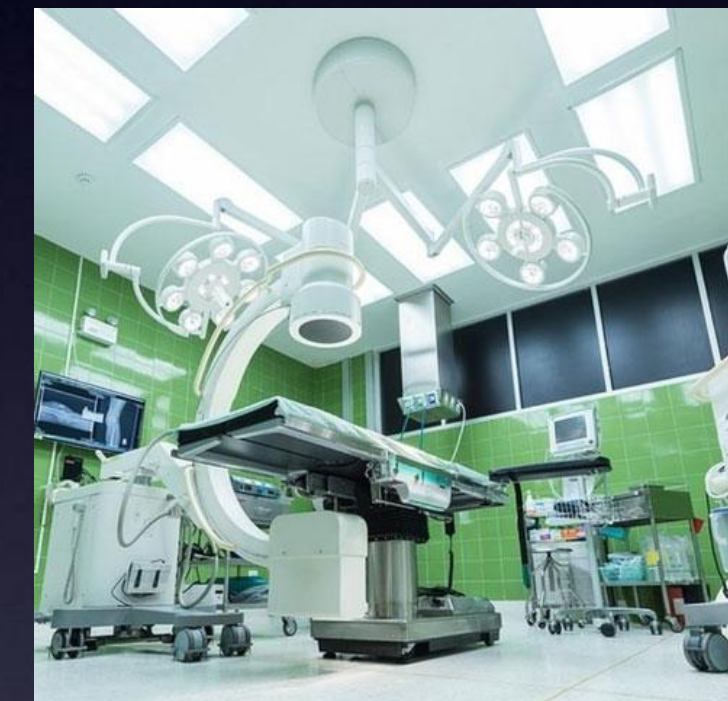
制造



农业

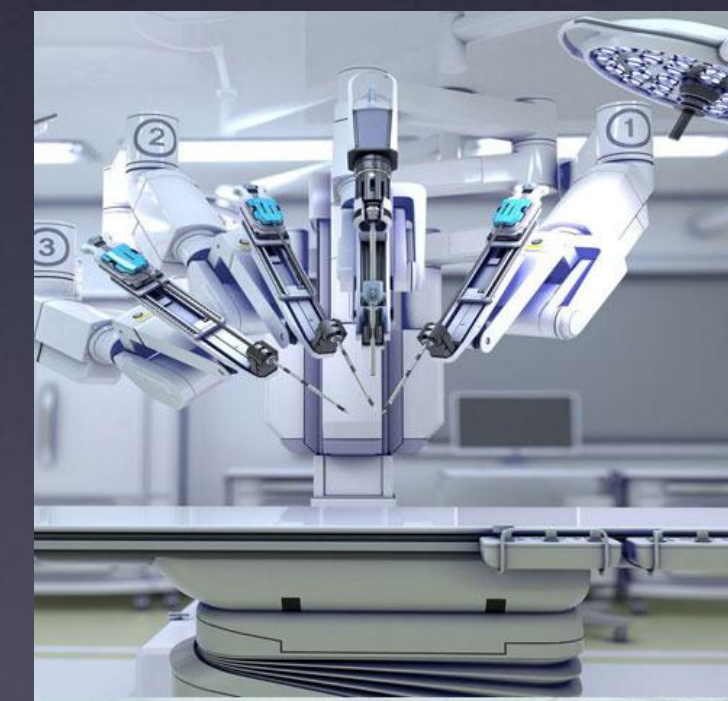


医疗



电气时代:

智能时代:

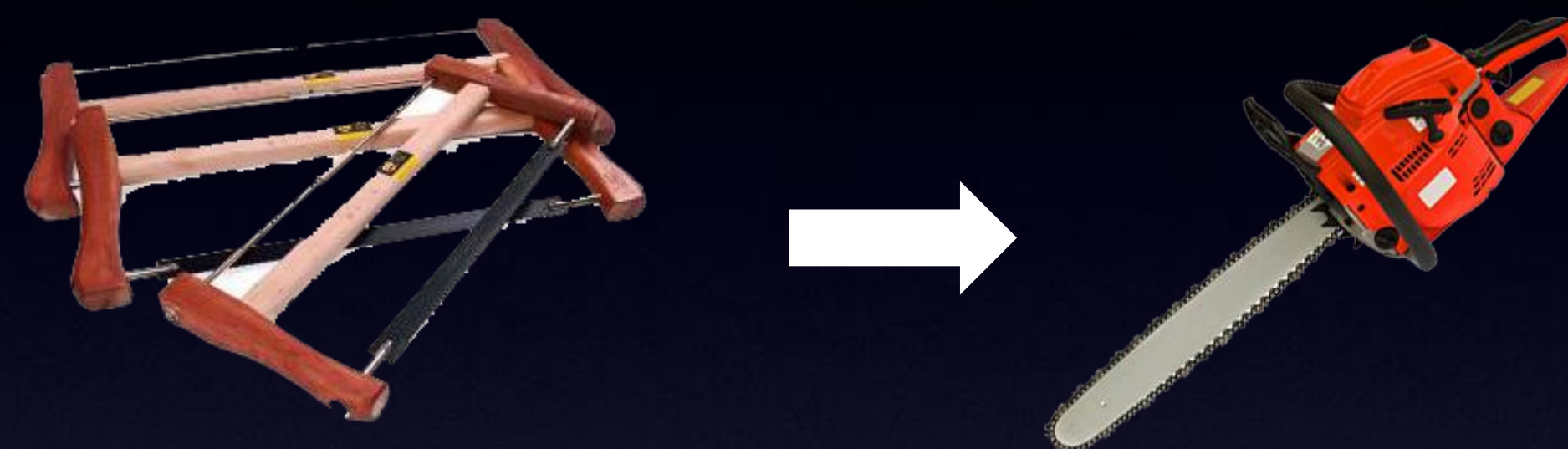




科技革命的动机：追求更加高效轻松的生产生活方式（“懒”）

两种方式实现：

简化老的工作



建立新的连接



互联网



更“懒”的方式？

“自动化、智能化”



互联网  
(信息时代)

数据积累




人工智能  
(智能时代)




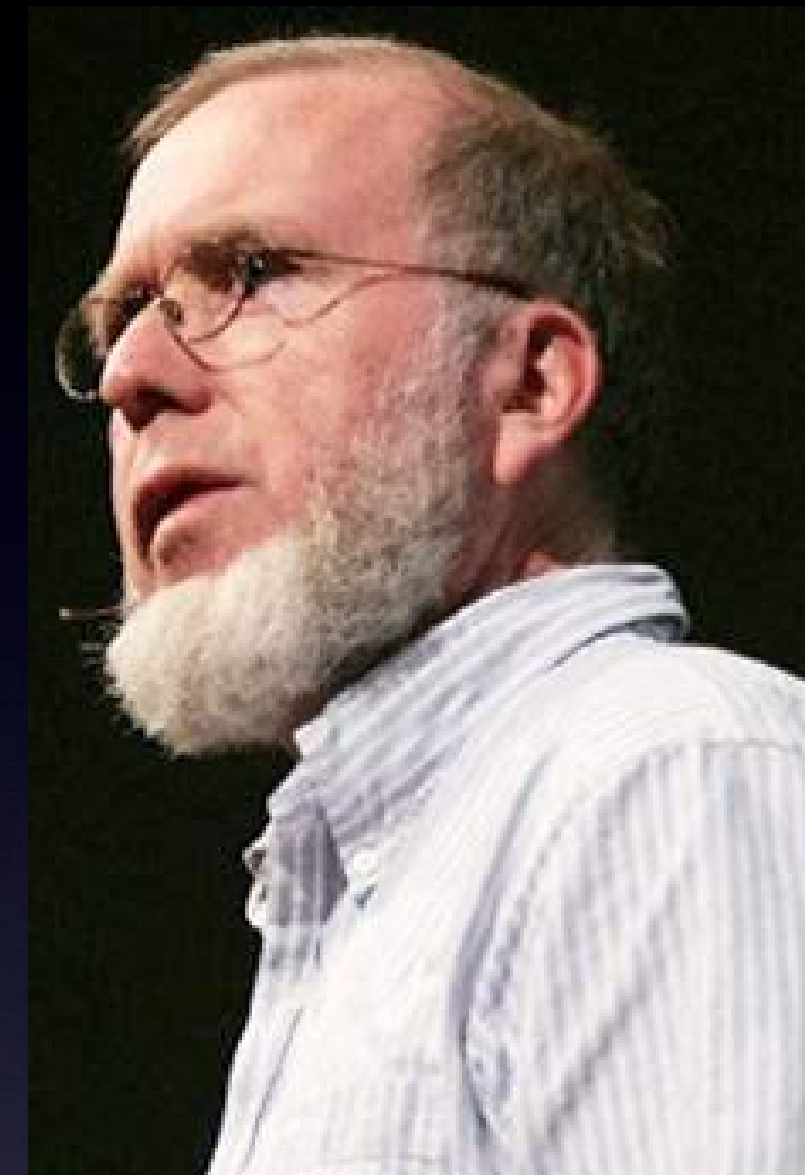
(电气革命) (AI科技革命)

electrify → cognify

以  (信息化)为基础

小结: AI 以  (自动化)为目标

以  (智能化)为特点



凯文·凯利



# AI 能做什么？

目前商业应用主要是  $x \rightarrow y$  模式

例如： 决策系统



个人消费信用数据  $\rightarrow$  他会偿还吗？（0/1 银行）



推荐系统 （←搜索系统）



她的购买历史行为 → 她会购买吗？（阿里 百度 京东等）

他的阅读历史行为 → 他会点击吗？（今日头条、Chinaso等）



## 图像识别



图片 → 是猫还是狗（人脸识别：是不是逃犯 **face++**）



# 语音识别



语音 → 文本 (Siri、科大讯飞、微信语音转文字)

机器翻译



文A

英文 → 中文



# 自动驾驶



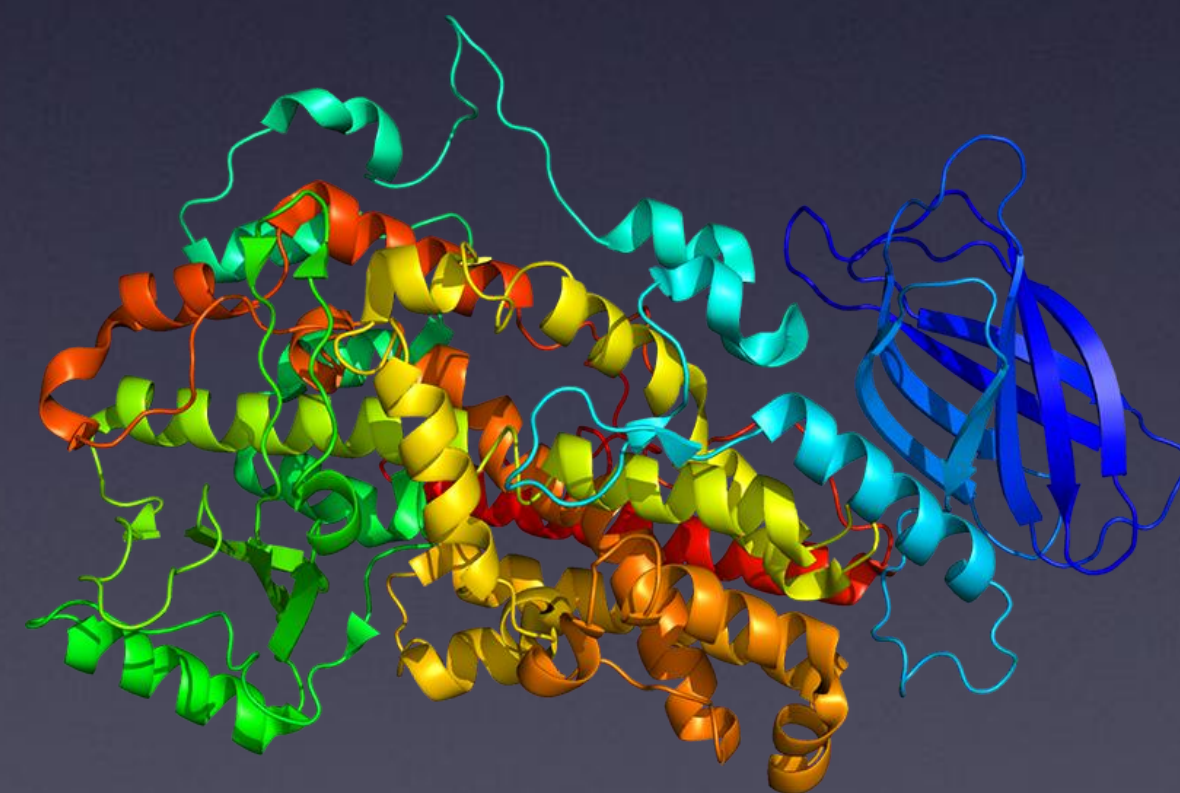
摄像头/雷达数据 → 位置关系

医疗应用



诊断数据 → 诊断结果（置信度）

分子信息 → 药物活性





# 舆情防暴恐应用



邮件、通话等数据 → 风险预警

信贷系统: 个人消费信用数据 → 他会偿还吗?  
推荐系统: 你的购买历史行为 → 她会购买吗?  
图像识别: 图片 → 是猫还是狗  
语音识别: 语音 → 文本  
机器翻译: 英文 → 中文  
自动驾驶: 摄像头/雷达数据 → 位置关系  
医疗应用: 诊断数据 → 诊断结果  
舆情防暴恐应用: 邮件、通话等数据 → 风险预警

$x \rightarrow y$  模式: “监督式学习”

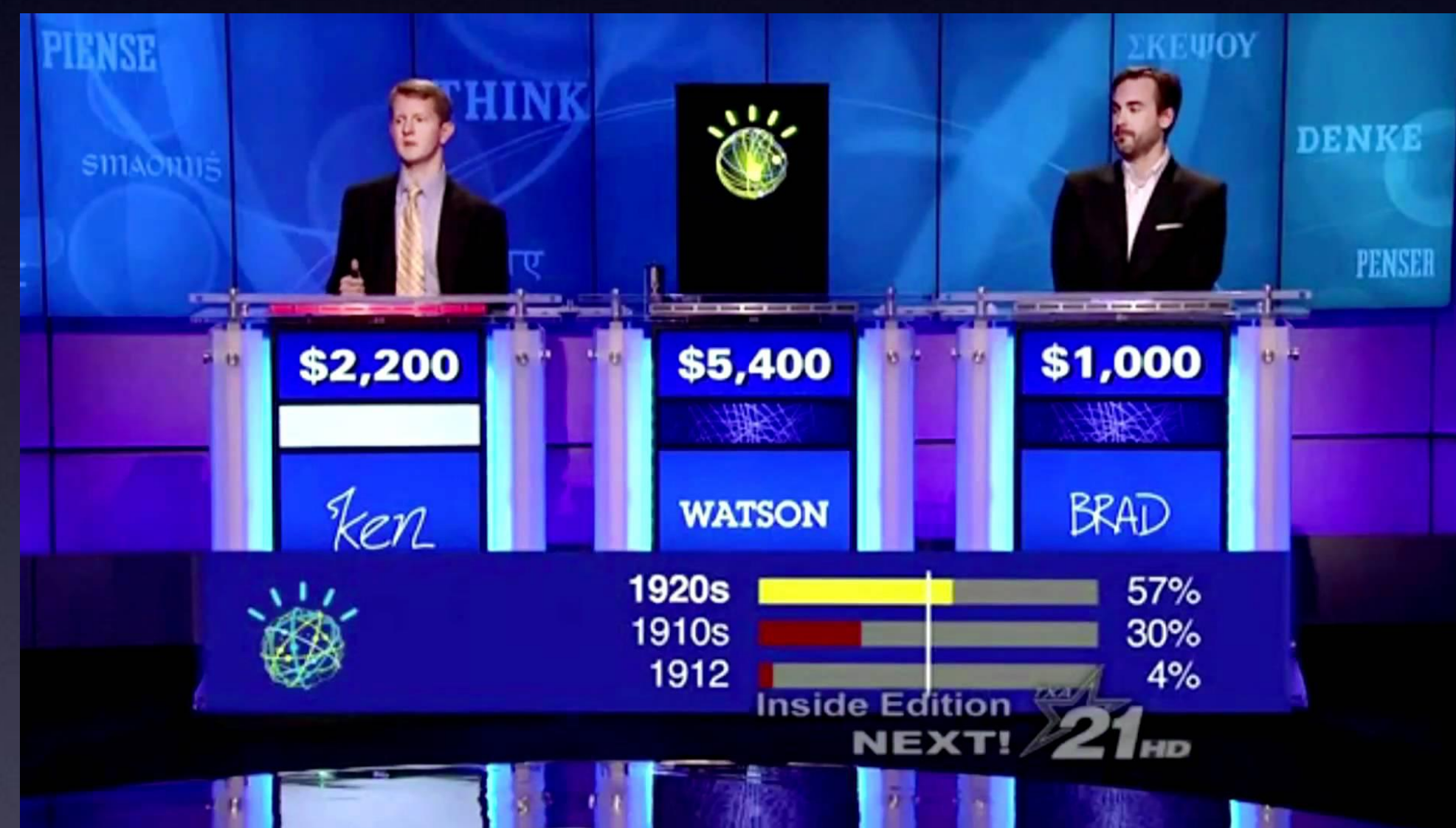


# 其他 AI 典型案例

deep blue （IBM ） 打败国际象棋人类冠军 1996



watson (IBM) 打败答题游戏 (Jeopardy) 人类冠军 2011





# alphaGo (google) 打败围棋人类冠军 2016



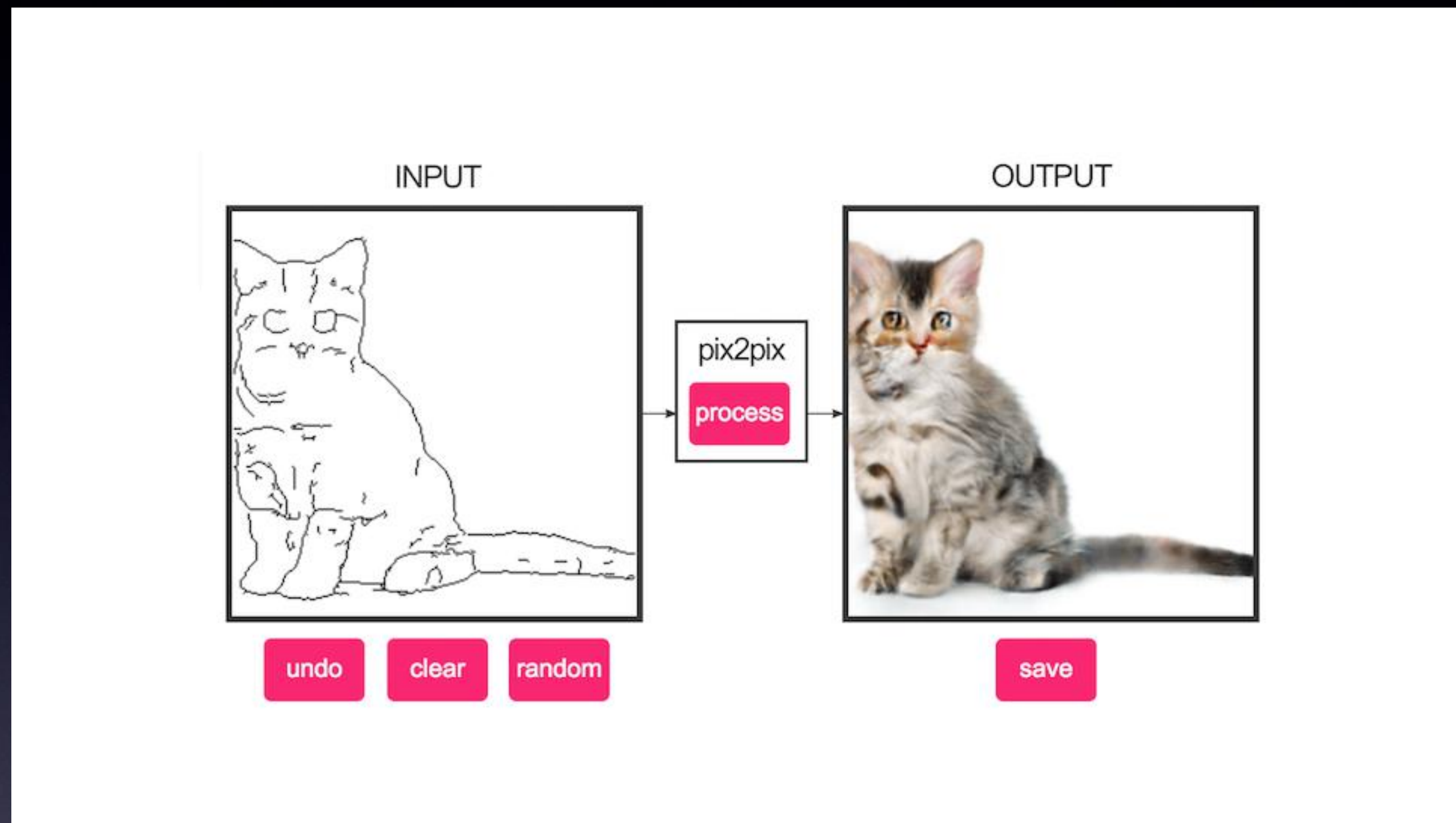
# Libratus (CMU) 打败德州扑克人类冠军 2017



“半监督式学习”



# 漫画创作（generative adversarial networks）



<https://affinelayer.com/pixsrv/>

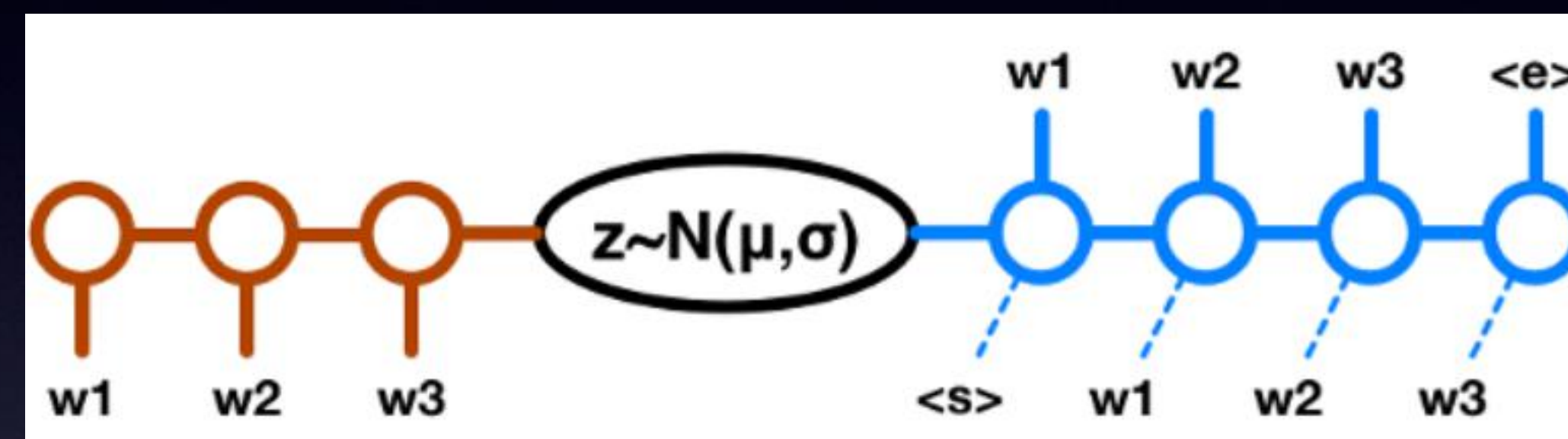
# 写诗机器人 (Variational Autoencoder)

风起蒹葭苑，花开白帝城。

雪飘青峰下，月落洞庭春。

中原有弟子，国士自风流。

北阙南山外，京华万里游。



“无监督式学习”



# AI 能做什么？

“一秒”原则：凡是人类可以1秒内完成的任务，AI可完成（图像识别）



写小说，困难！（至少现在和可预见的将来很困难）

# AI 能做什么？

“有料”原则：凡是可获得大量有效数据的任务，AI可完成（机器翻译）

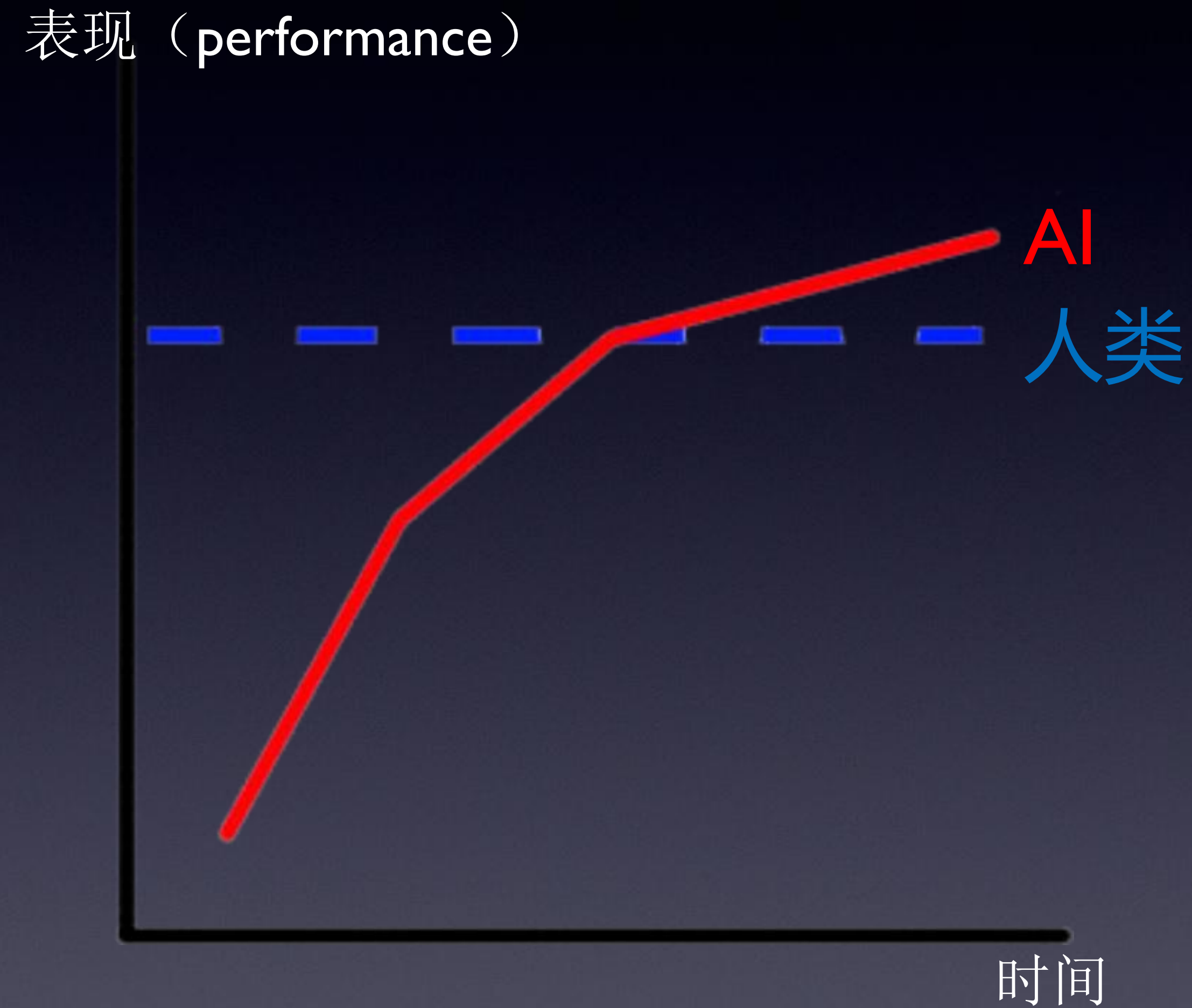


## 预测新药，较困难！（已知药物较少）

# 预测股票，很困难！（可能无解）



# AI 性能和人类的关系



原因:

1. 人类智能是一个极值参考
2. 数据是人类产生的

小结:

**AI** 在众多领域已有重要应用

**AI** 性能和数据体量和质量紧密相关



# 为什么AI 在互联网企业最先发展？

原因：

1. 信息化先锋，数据丰富
2. 算法的收益丰厚

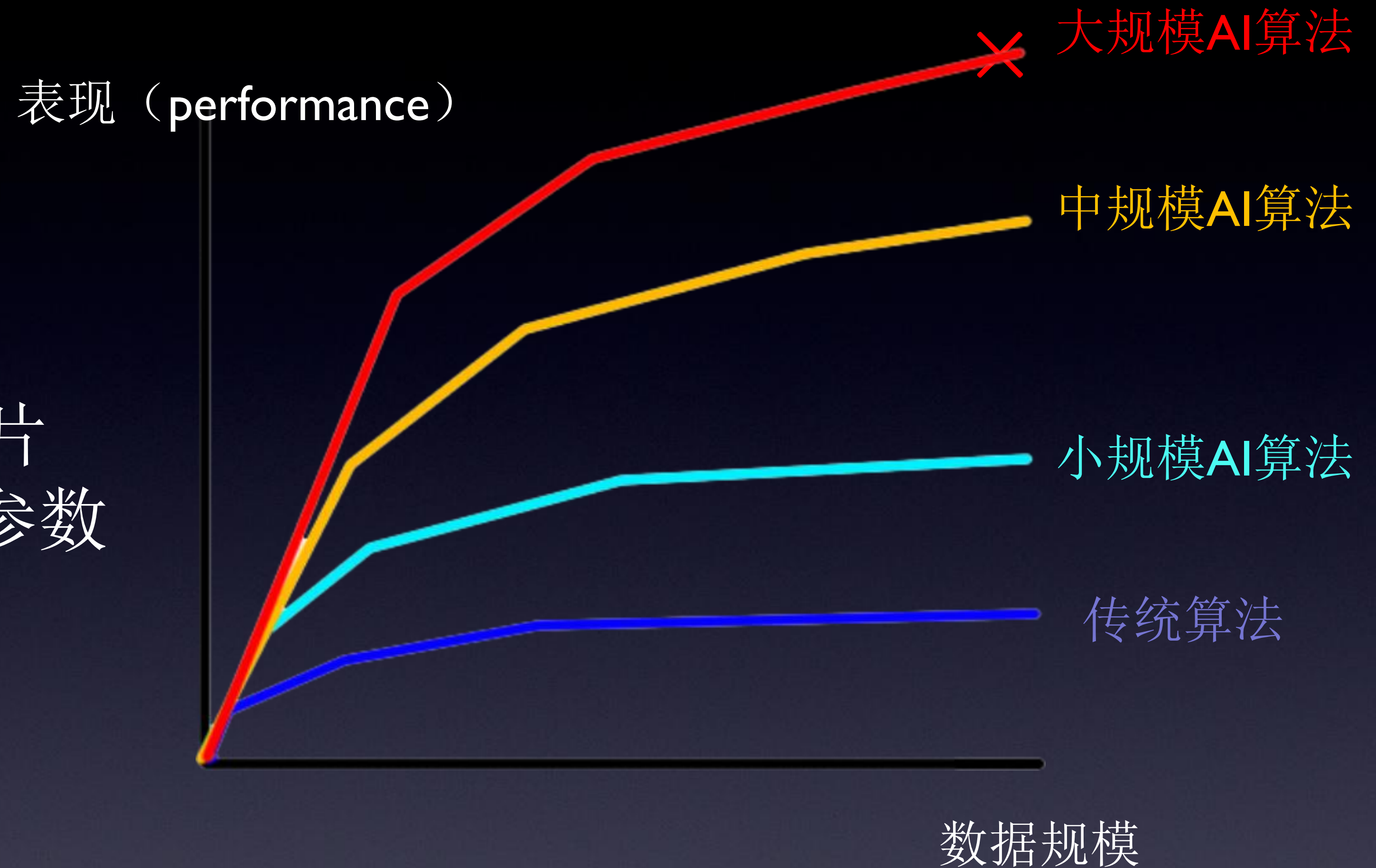




## AI 性能和数据、算法的关系

AI 为何现在崛起？

Imagenet: 14,197,122 图片  
VGG 模型: 138,000,000 参数



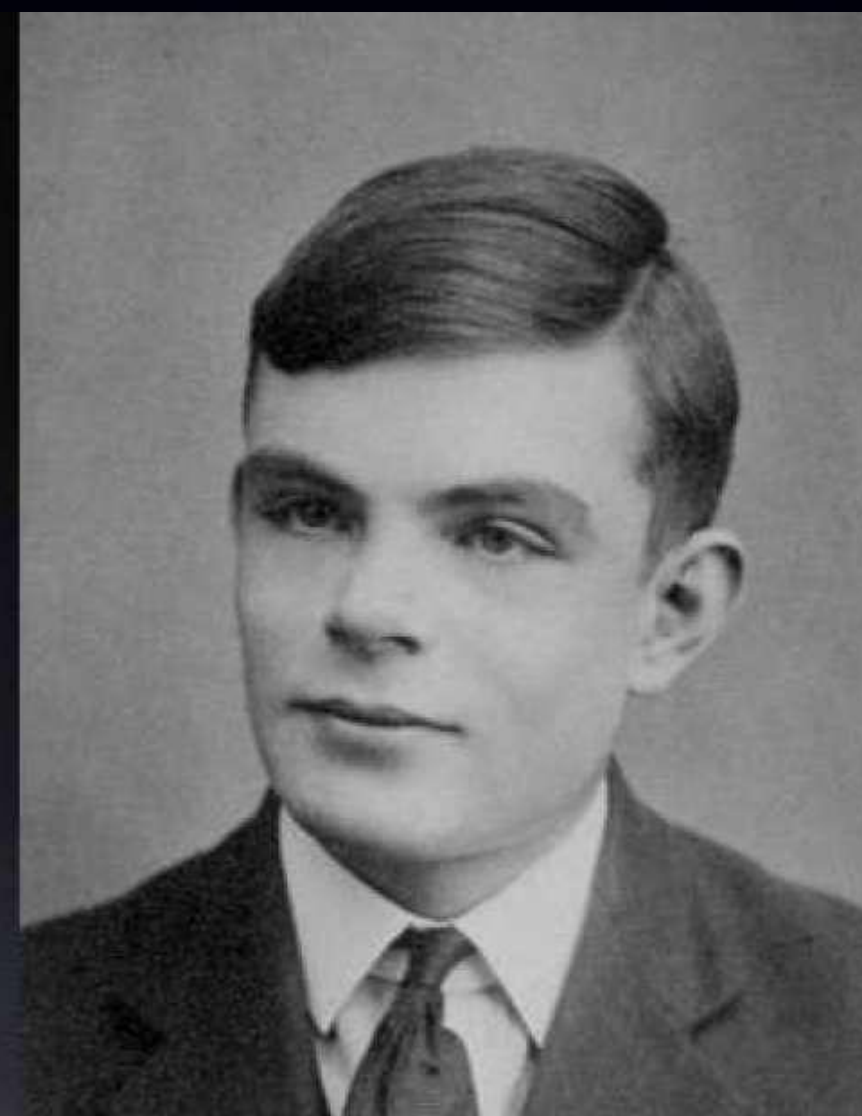
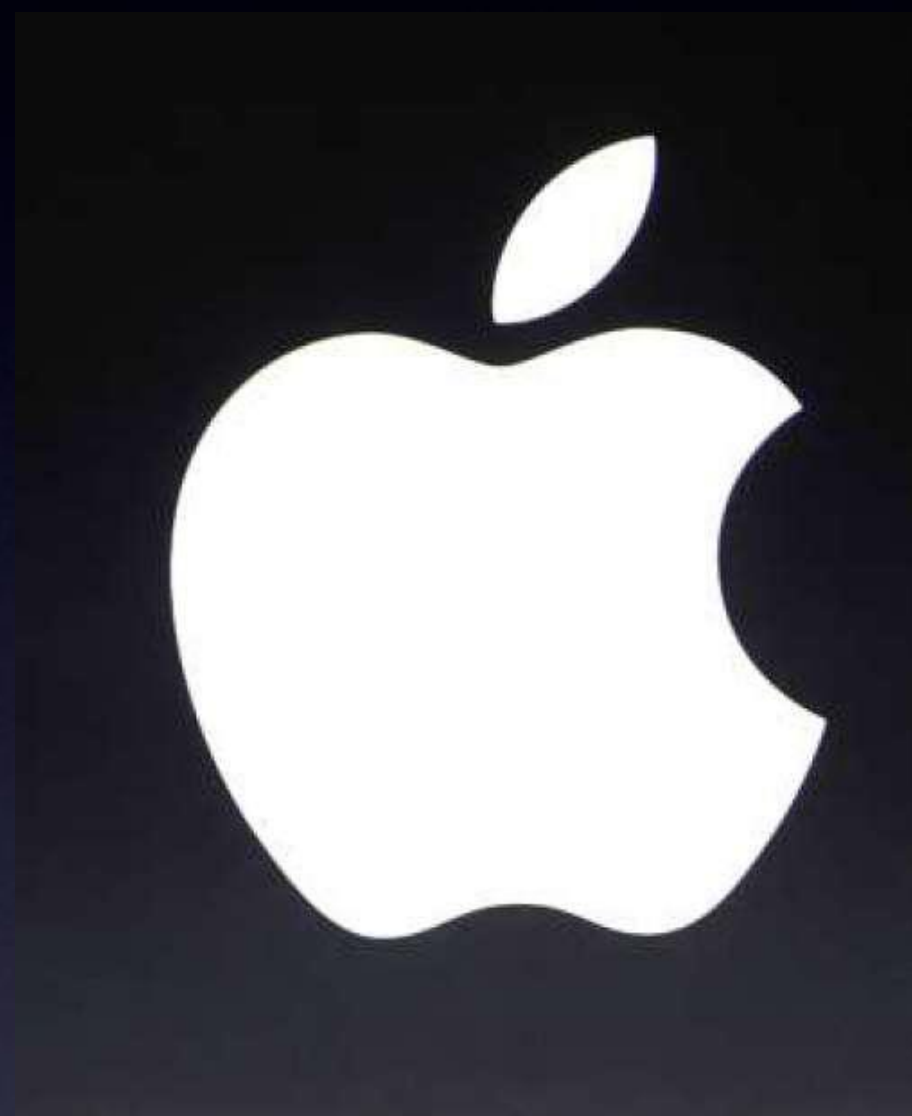
AI崛起的条件：数据、计算平台、算法一个不能少



# AI 历史简要回顾

## 1930s

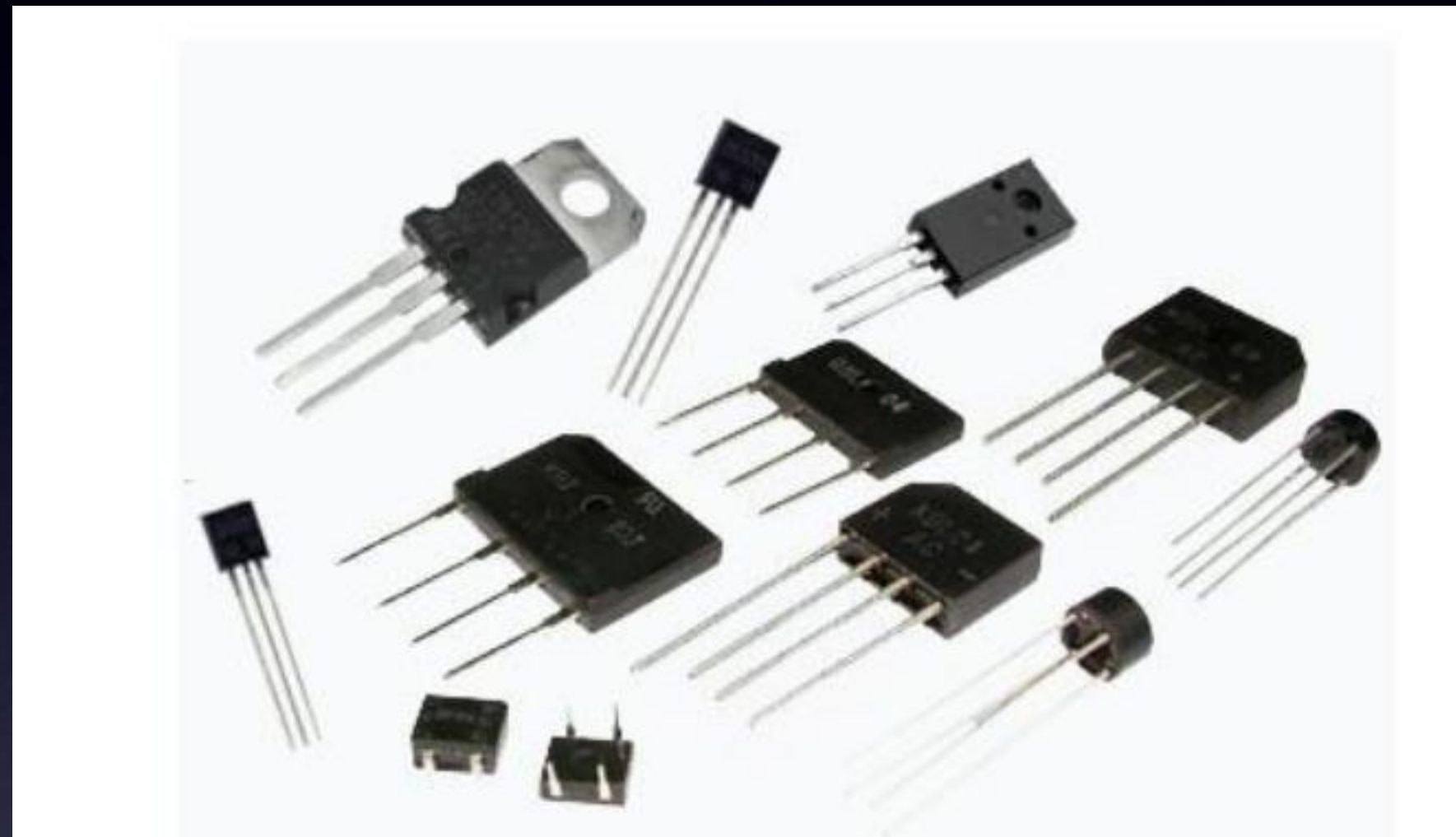
图灵机概念提出



# AI 历史简要回顾

## 1940s

晶体管诞生

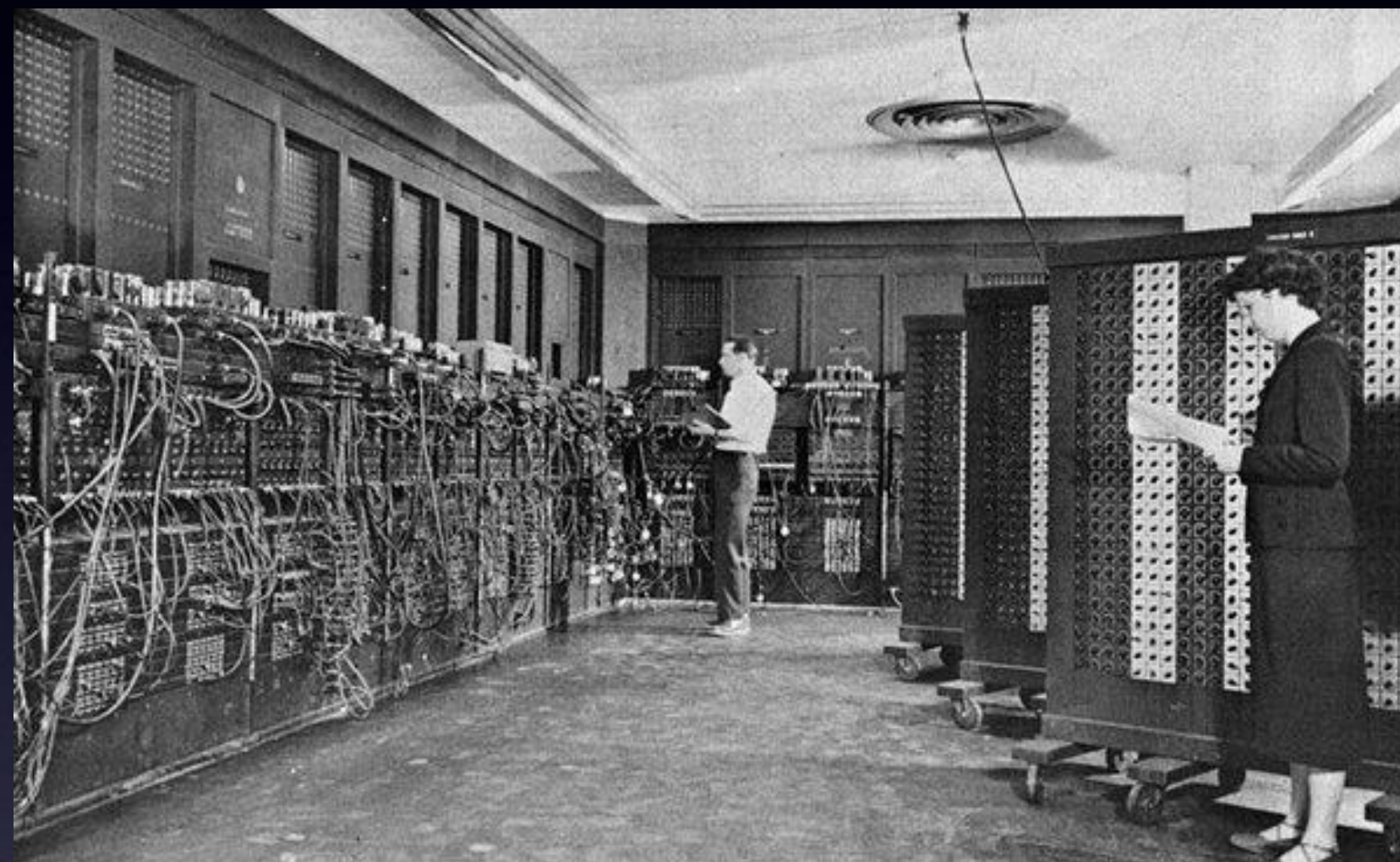




# AI 历史简要回顾

## 1950s

冯洛伊曼架构，  
晶体管计算机诞生

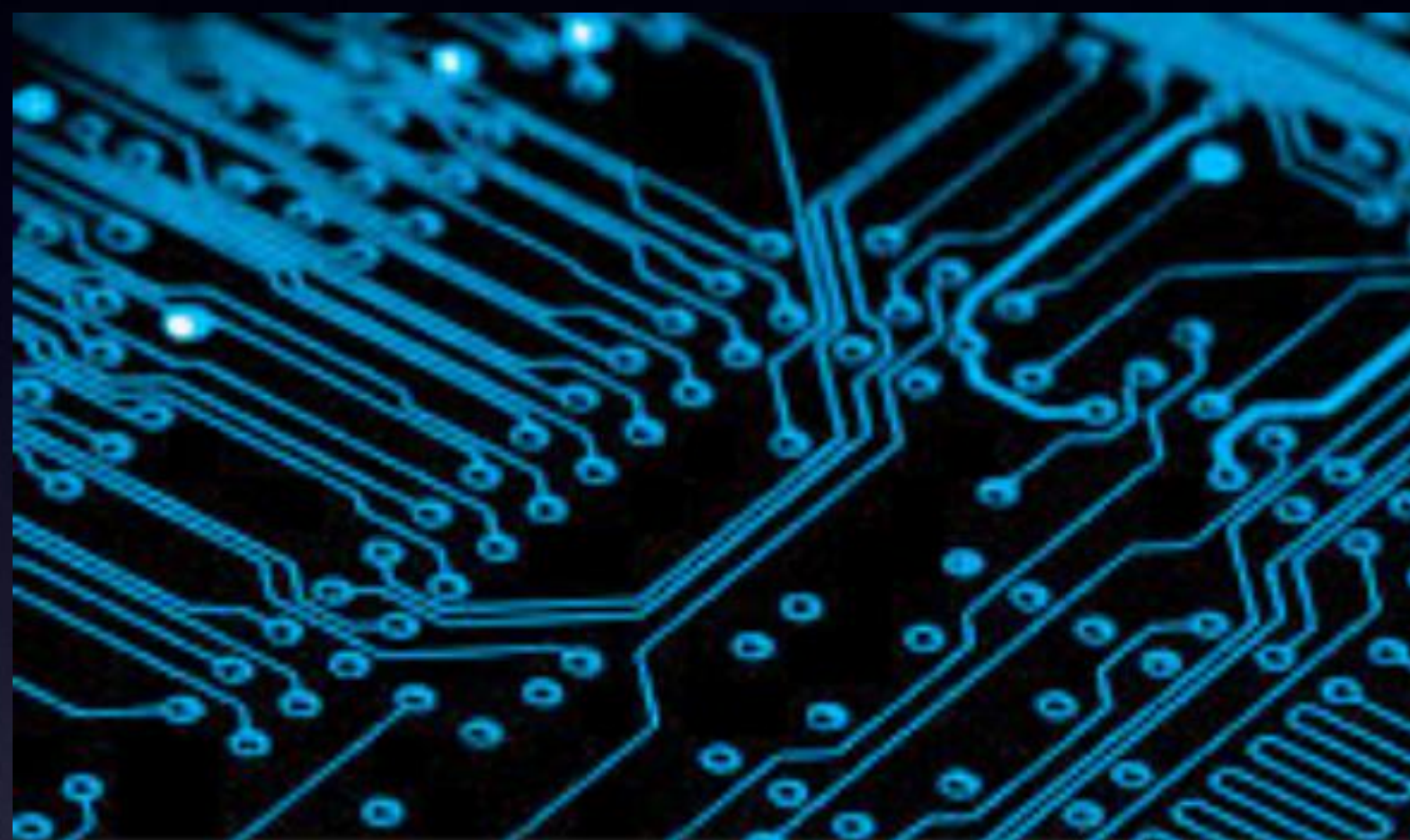




# AI 历史简要回顾

## 1960s

集成电路（IC）诞生





# AI 历史简要回顾

## 1970s

微处理器（CPU）诞生

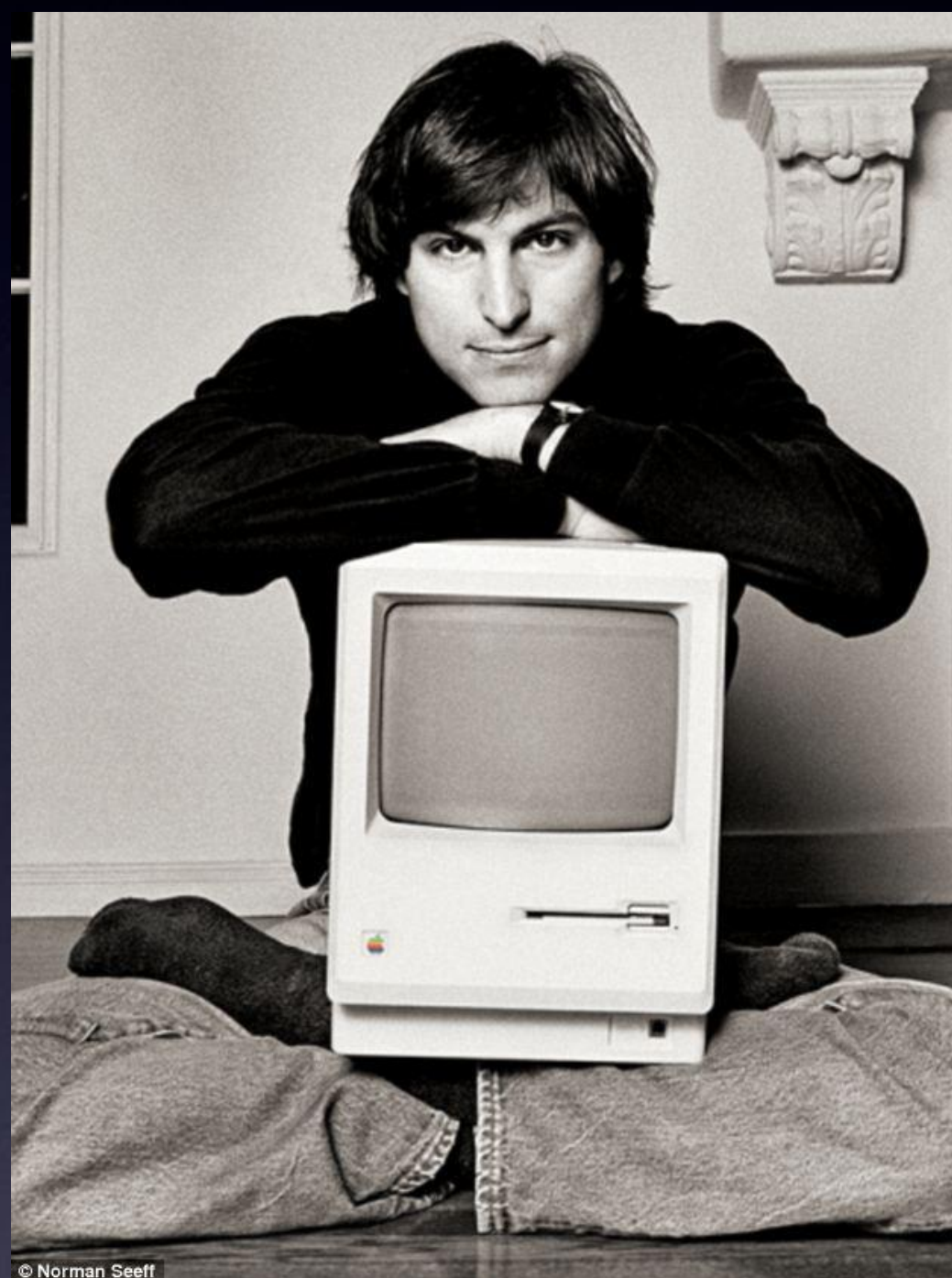




# AI 历史简要回顾

## 1980s

显卡、个人电脑诞生





# AI 历史简要回顾

## 1989

LeCun 提出 cnn 算法

# AI 历史简要回顾

## 1999

显卡GPU性能突破（NVIDIA GeForce, 黄仁勋）



# AI 历史简要回顾

## 2001~

互联网行业复兴（google，亚马逊，facebook，BAT等）

# AI 历史简要回顾

## 2004

hadoop 与分布式计算崛起



# AI 历史简要回顾

## 2006

DBN算法诞生， nn算法复兴（Hinton）

# AI 历史简要回顾

## 2009

Imagenet数据集诞生（李飞飞）



# AI 历史简要回顾

## 2012

cnn 用于 Imagenet, 分辨率超越人类 (Hinton)

# AI 历史简要回顾

1930s 图灵机概念提出； 1940s 晶体管诞生；  
1950s 冯洛伊曼架构和晶体管计算机诞生；  
1960s 集成电路（IC）诞生； 1970s 微处理器（CPU）诞生  
1980s 显卡、个人电脑诞生

} 平台

（1989年 LeCun 提出cnn网络） 算法  
1990年 万维网（www）诞生 平台&数据  
1999年 显卡GPU性能突破（NVIDIA GeForce, 黄仁勋） 平台  
2001年 互联网行业复兴（google, 亚马逊, facebook, BAT） 数据

2004年 hadoop与分布式计算、云计算崛起 平台  
2006年 DBN算法诞生, nn复兴 算法  
2009年 Imagenet数据集诞生（李飞飞） 数据  
2012年 cnn 用于 Imagenet（Hinton 分辨率超越人类） 算法



# AI 为何现在崛起？

- 3个条件 {
- 数据丰富

算法突破

HPC平台成熟



1:1:1

# AI 最新趋势

摩尔定律：CPU芯片的性能，约每隔18-24个月便会增加一倍

AI摩尔定律：2012年起，最大AI模型规模，每3.5月增长1倍

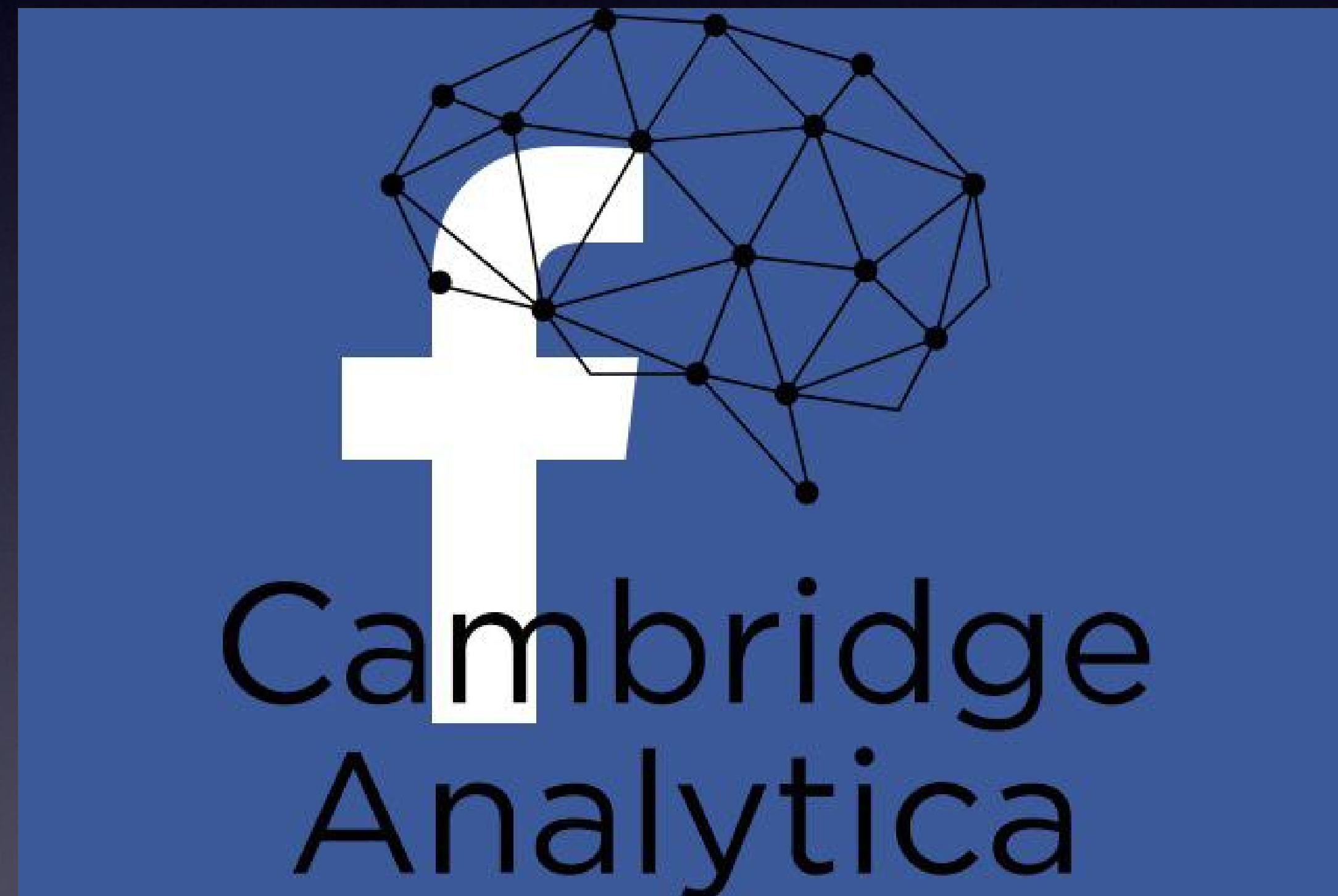


# AI 最新趋势

HPC(云计算、挖矿GPU)、物联网、量子计算和AI的发展彼此促进  
进化算法、对抗学习等“高阶”算法的出现，让AI崛起更有想象空间

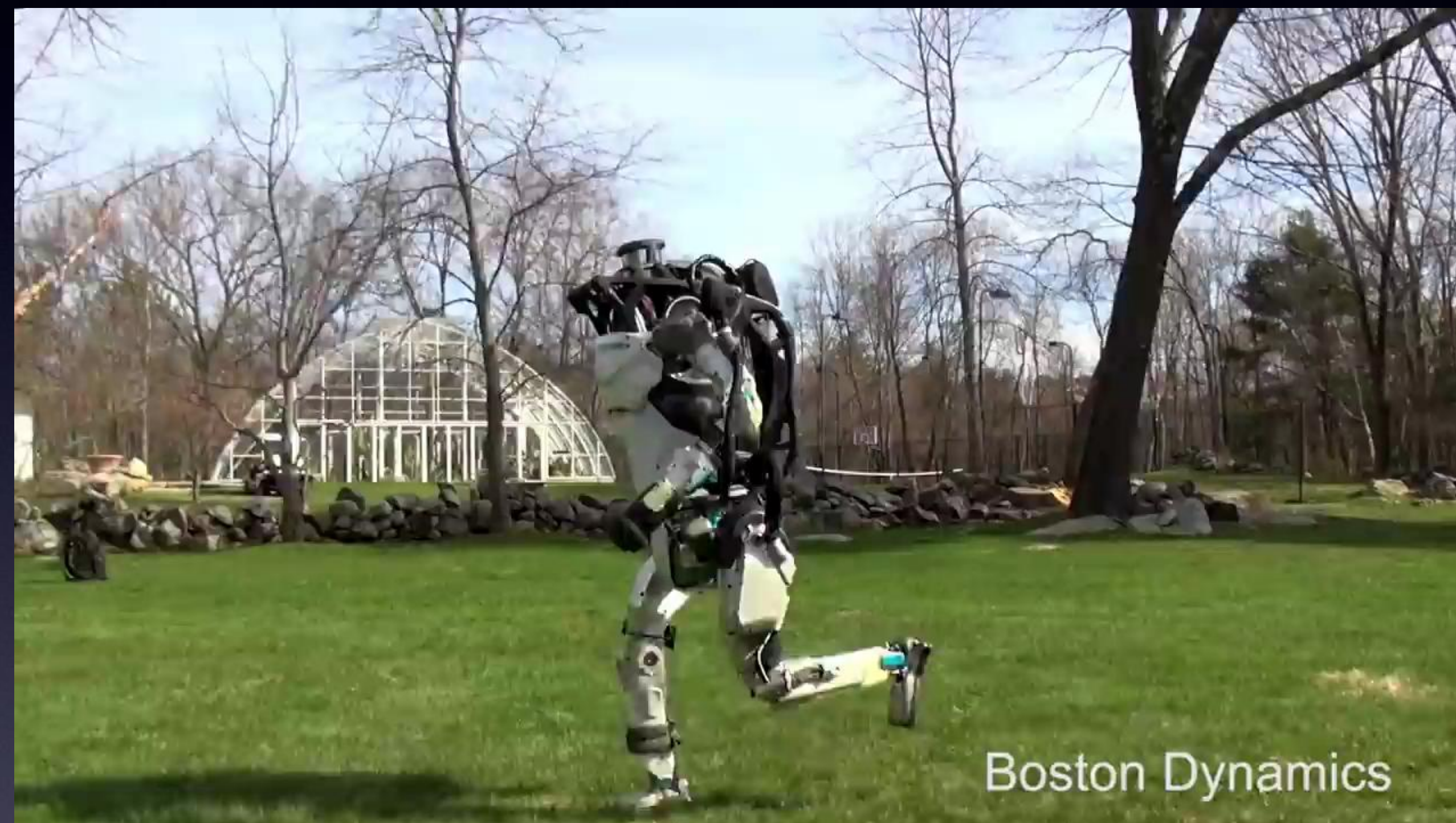
# AI 崛起危险吗？

美国大选和 Cambridge Analytica





# AI 崛起危险吗？



boston dynamics



# AI崛起危险吗？



Hawking



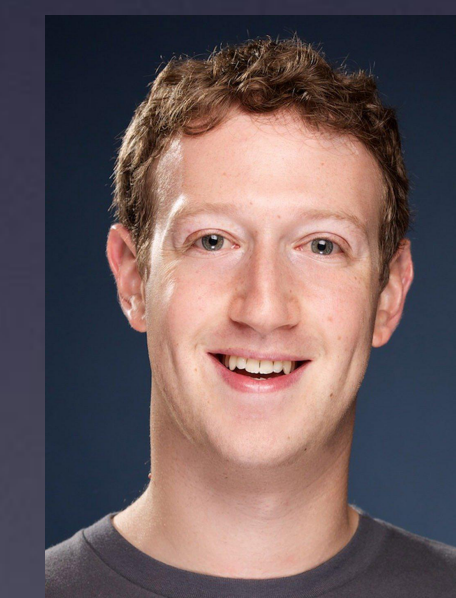
Musk



Jordan



Bengio



Zuckerberg

VS



# 什么是AI? (AI 技术简介)

# AI 按实现方式分类

AI早期：机械执行，依赖指定的程序（if then 模式）

AI当前：依赖数据，根据任务自动调节参数（learning 模式）

“机器学习”（machine learning） $\approx$  AI

“混合模式”：例如 chatbot



# 机器学习按数据特征分类

- 监督式(supervised): 分类、回归等问题, 如图像识别
- 半监督式(semi-supervised): 强化学习问题, 如玩游戏
- 无监督式 (unsupervised): generative 问题, 如画画

# 机器学习按建模思想分类

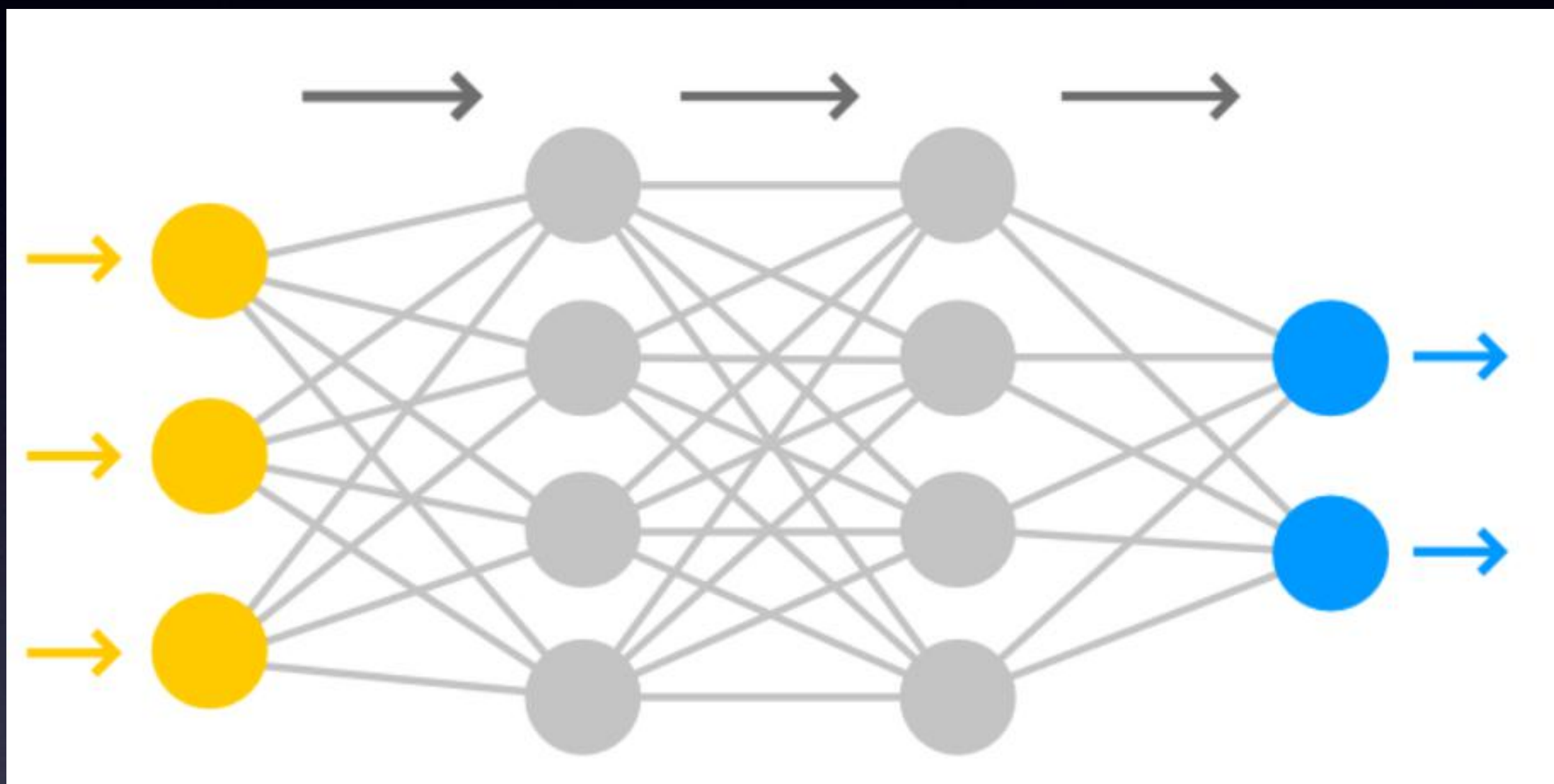
- 符号主义：1980s，知识图谱
- 贝叶斯流派：1990s，马可夫链、LDA等概率模型
- 联结主义：2000后，神经网络
- 进化主义：遗传算法
- Analogizer：SVM、条件极值算法



# 神经网络算法简介



神经元



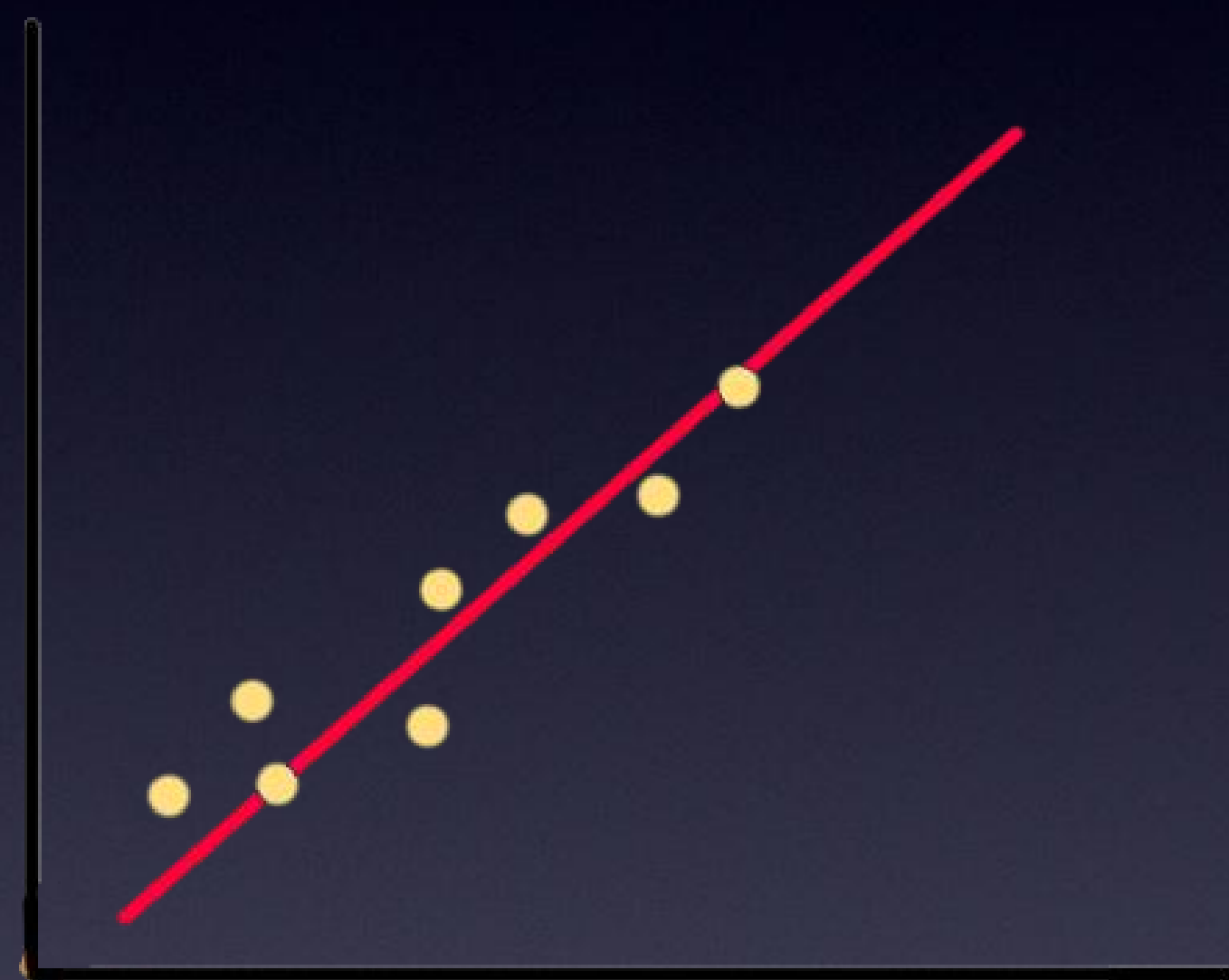
神经网络

仅仅是“神似”，没有更多生物学含义



# 一个例子

房价



大小

“数据拟合”



大小

房价

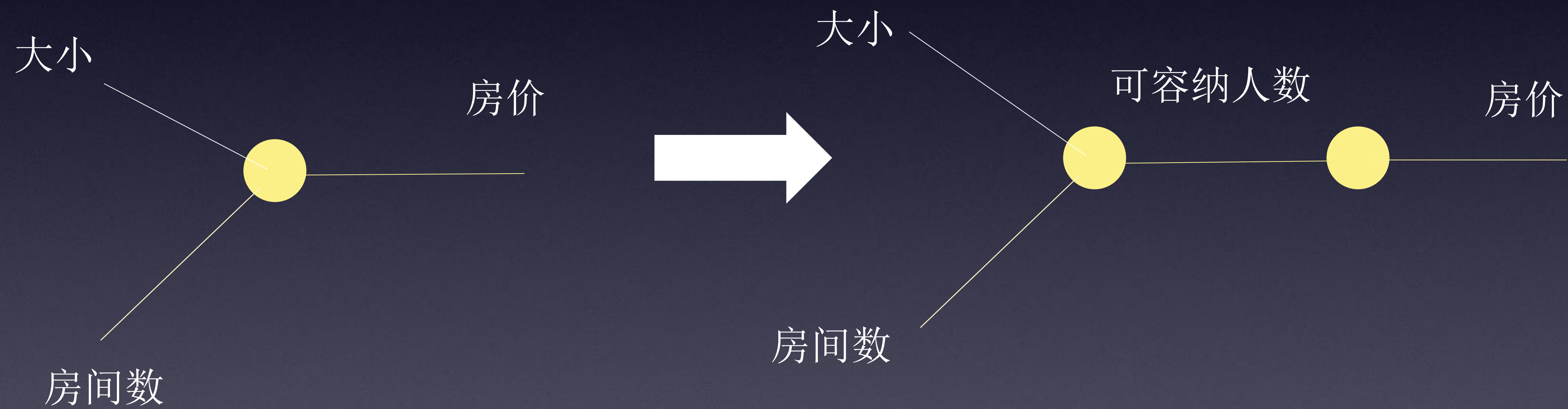




# 一个例子



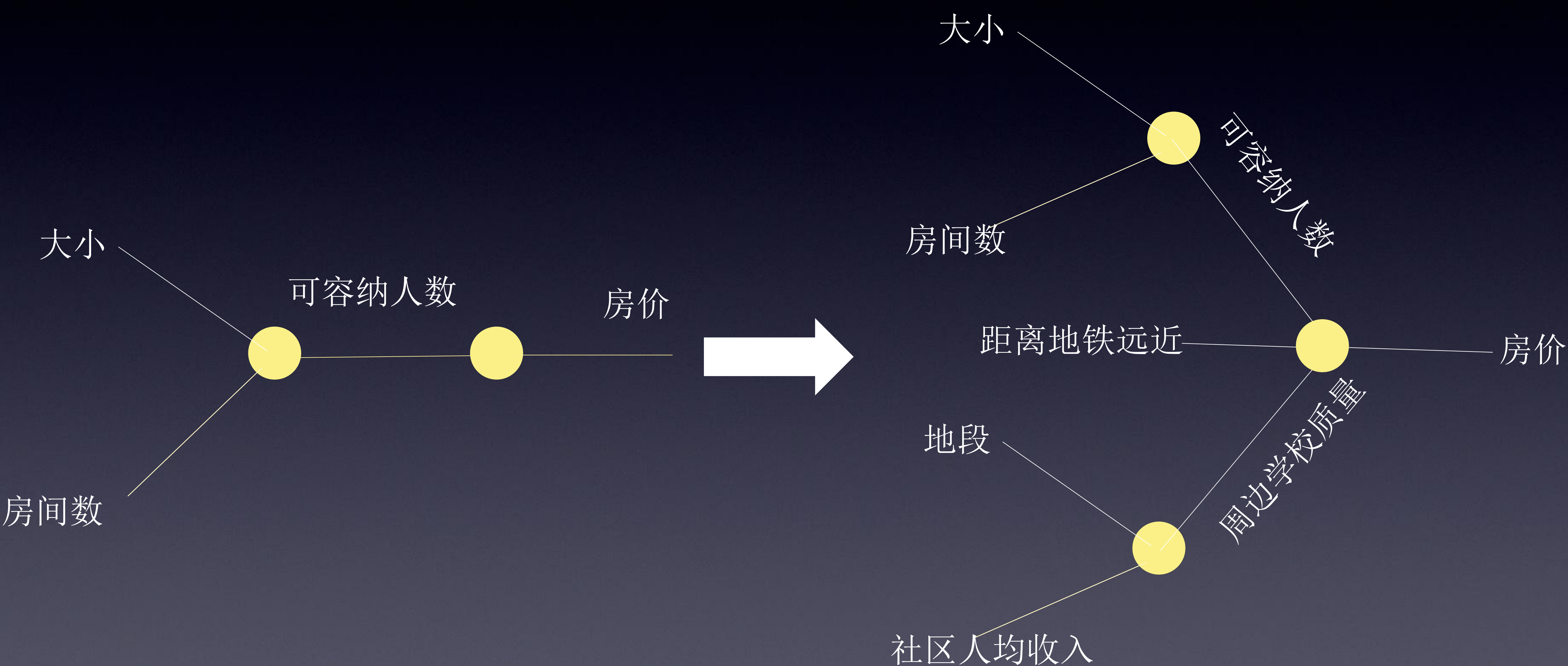
# 一个例子





# 一个例子

## “数据拟合”



# AI崛起危险吗？

大体上仅处于起始感知（**inception**）阶段，而缺乏“思考”能力

“专家系统”较成熟，“通用智能”远未达到（缺乏“自我学习、反省”能力）

不比过于担心“火星人口”问题



AI 对我们生活的具体影响？  
(AI 的问题、机遇和门槛)

# AI 可能存在的问题

错误隐蔽：debug 困难（“黑匣子”）

“副作用”隐蔽：例如杀熟、歧视有前科的人等等（算法角度是正确的）



# AI 让更多人失业？

电气时代：电子产品的开发、维修、测试

信息时代：互联网产品的开发、测试、运维

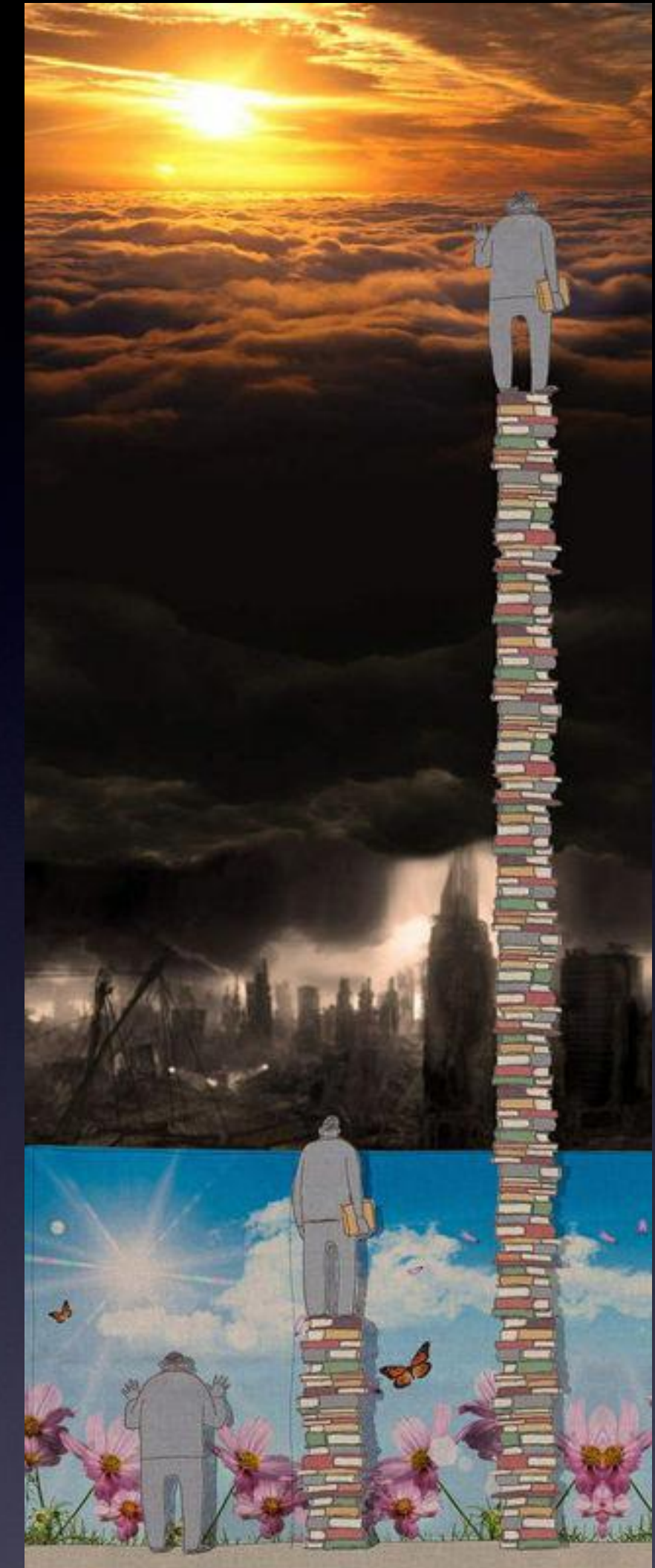
AI时代：AI产品的开发、测试、运维

“赶马车” → “开汽车”

“IT 产品经理” → “AI 产品经理”

“搬砖” → “标注数据”

“生产力”上升的变革，会导致阵痛，但最终会云淡风轻！



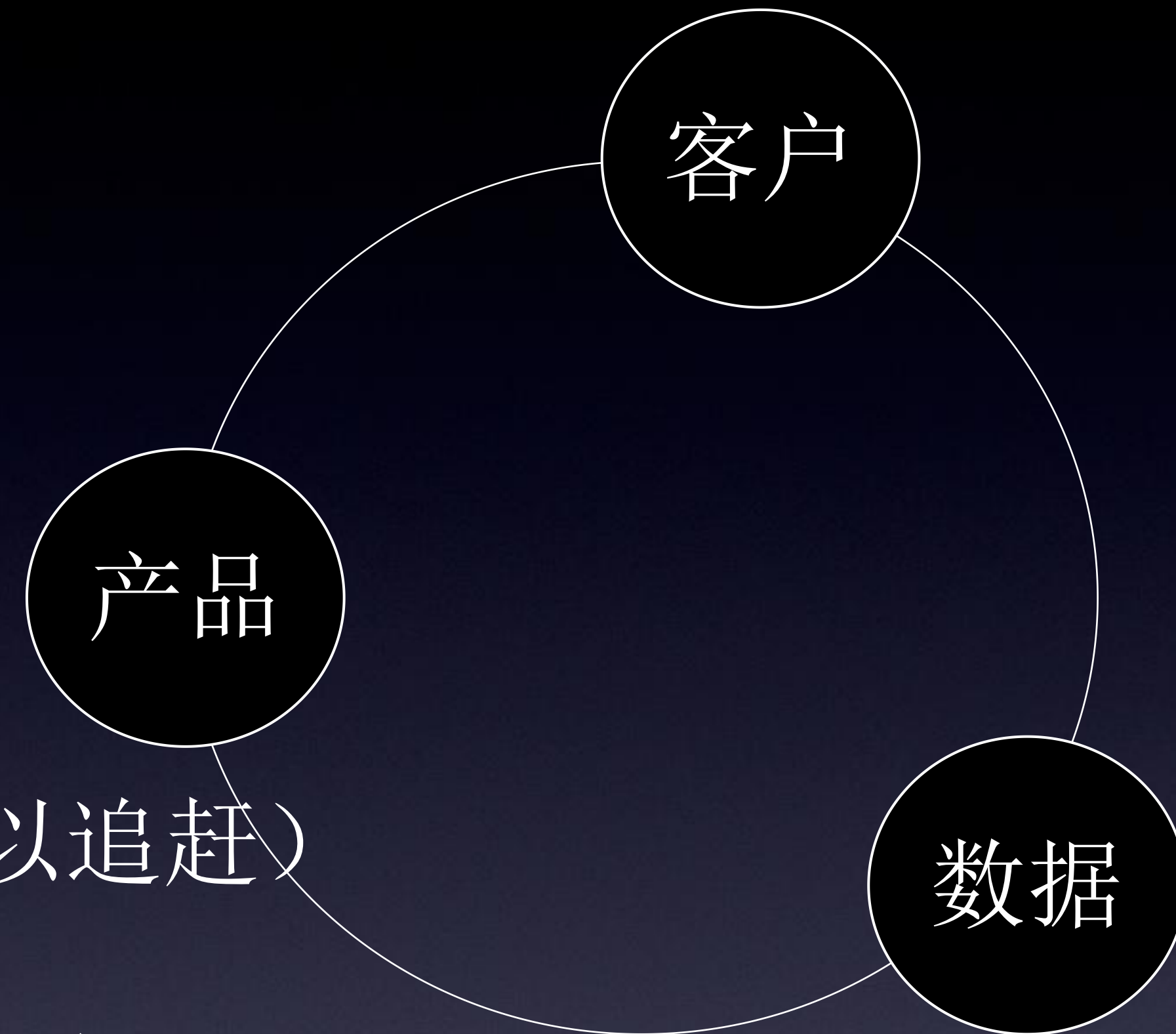
# AI 的机遇与门槛

+ 算法公开、平台公开！

+ 细分场景应用众多！

+ 数据是门槛（一旦形成优势，难以追赶）

AI 知识庞杂、未完全系统化、人才匮乏





# 谢谢！

## Q&A

