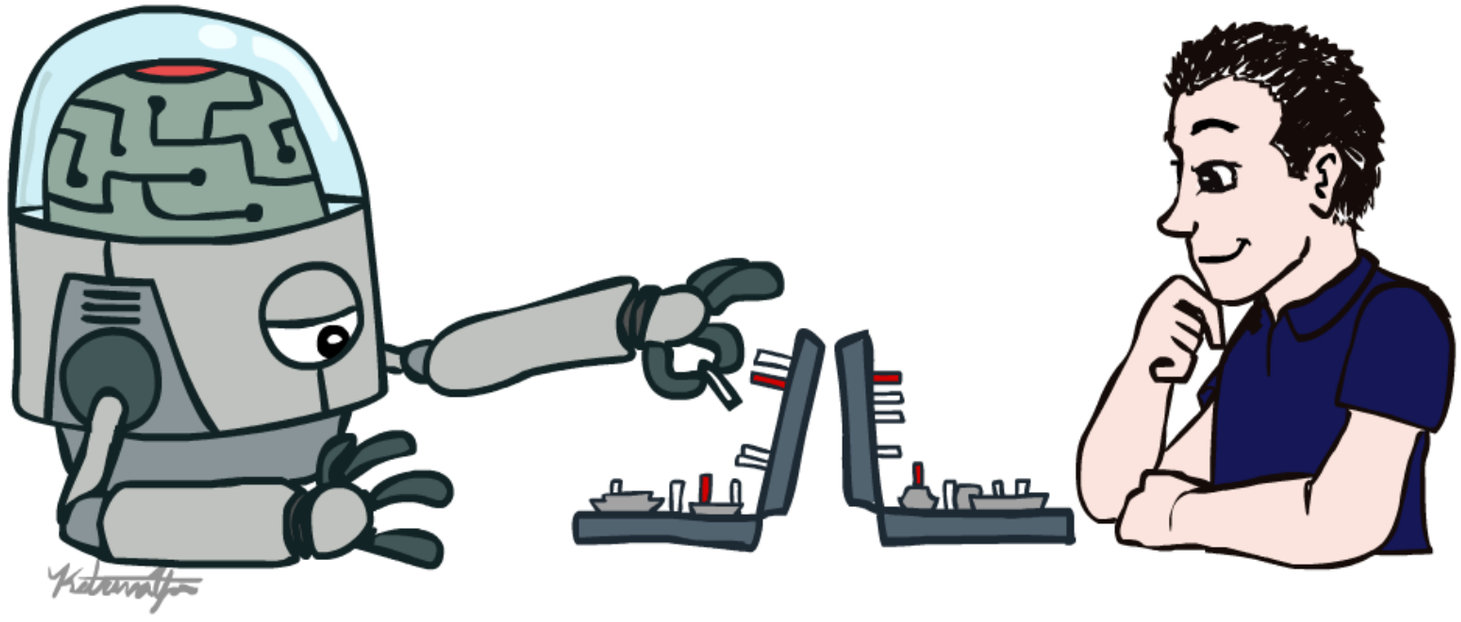
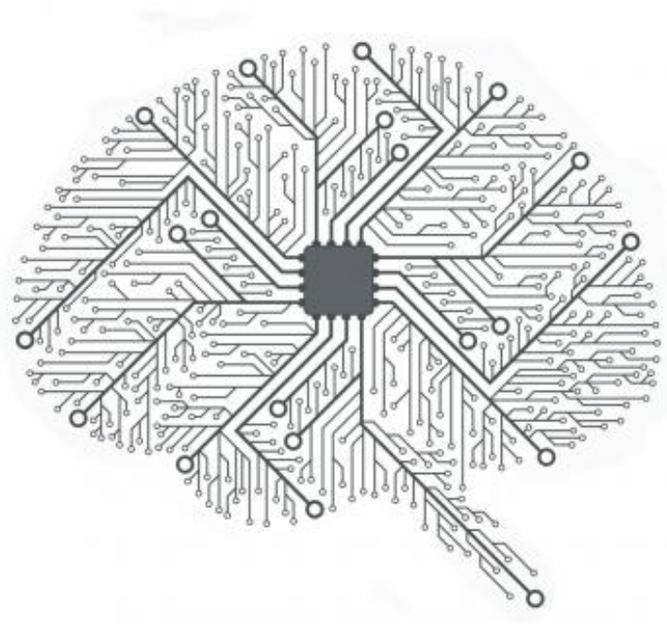
人工智能



第一章·绪论

- 什么是人工智能
- 人工智能的历史
- 目标和研究范式
- 人工智能模型
- 一个具体的例子



什么是人工智能


ACM Chess Challenge
Garry Kasparov
vs
TechDribble


Lee Sedol


English
Hello
Spanish
Hola


ORIGINAL
DEEPFAKE

什么是人工智能

Stephen Hawking warns artificial intelligence could end mankind

By Rory Cellan-Jones
Technology correspondent

2 December 2014 | Share

Prof Stephen Hawking, one of Britain's pre-eminent scientists, has said that efforts to create thinking machines pose a threat to our very existence.

什么是人工智能

- 是计算机科学（与认知心理学、脑科学、……）的一个交叉研究领域

- 研究目标：不仅仅是理解，而且是是制造“智能体”

- 兴起于二战结束后，“artificial intelligence”一词出现于1956年

- 仍在快速地发展中，许多问题尚未解决，包括“什么人工智能？”

人工智能研究角度

- 像人一样思考："[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning..." (Bellman, 1978)

- 理性地思考："The study of mental faculties through the use of computational models"(Charniak+McDermott, 1985)

- 像人一样行动："The study of how to make computers do things at which, at the moment, people are better" (Rich+Knight, 1991)

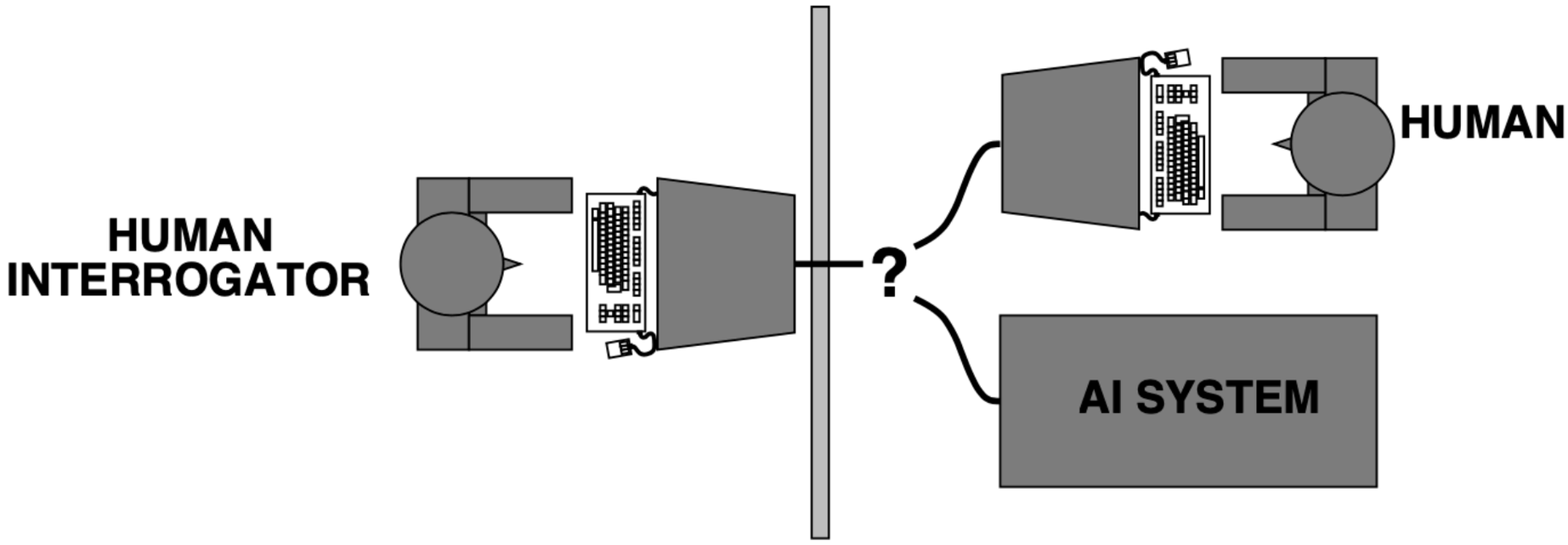
- 理性地行动："The branch of computer science that is concerned with the automation of intelligent behavior" (Luger+Stubblefield, 1993)

像人一样去思考：对认知建模

- 需要理解人如何思考
- 交叉学科:认知科学(Cognitive science)，心理学，神经系统科学
  - 计算机模型←人工智能
  - 实验技术←心理学
  - 研究人的心智(大脑)
  - 依赖于对人或动物的实验
- 人工智能实验对象主要是计算机

像人一样行动：图灵测试

- 图灵在“计算机器与智能”(1950)中提出:
  - “机器能思考吗?” ⇒ “机器能有智能行为吗?”
  - 智能行为操作测试:模仿游戏



像人一样行动：图灵测试2

- 计算机需具备这些能力，以通过测试
  - 自然语言处理
  - 知识表达(储存)
  - 自动推理(用已知推新结论)
  - 机器学习(在新环境中学习)
- 全面图灵测试(Total Turing Test, by Harnad,1991)
  - “The candidate must be able to do, in the real world of objects and people, everything that real people can do”
  - 计算机视觉，语音识别，机器人技术

“合理地思考”(思维法则)

- 希腊哲学家，亚里士多德
  - 三段论，演绎推理(syllogism)
- 逻辑学:研究思维的法则
- 用逻辑规则建造智能系统所面临的挑战
  - 自然界中许多知识很难用逻辑表示法来表述;有时知识不是100%确定的
  - 原则理论上可以解决的问题，在实际中未必能轻松做到
  - 很难对思维进行编码，最终关键的不是如何思考，而是如何行动。

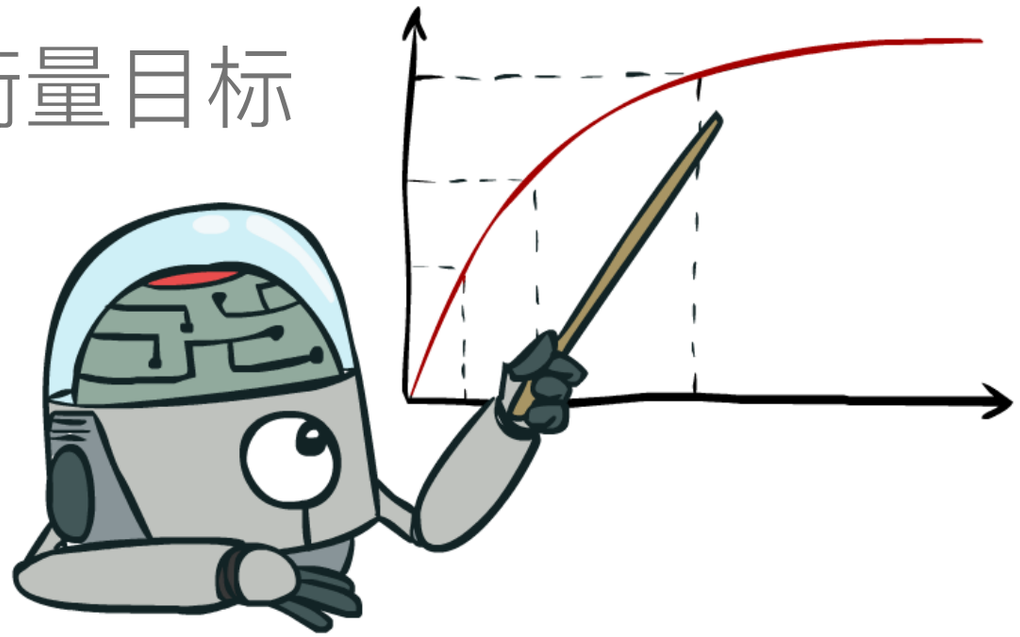
合理地行动

- 智能体（agent）—— 独立行动，感知环境，适应变化，追求目标
- 合理行动的智能体（rational agent）—— 去努力追求最好的结果，或最好的期望结果
- 基于合理性智能体方式进行人工智能研究的优势
  - 更具普遍意义(不仅只是依赖于逻辑推理)
  - 更便于实际开发（数学计算）
  - 完美合理性（总是做对的事）在复杂情况下难以实现

合理的决策

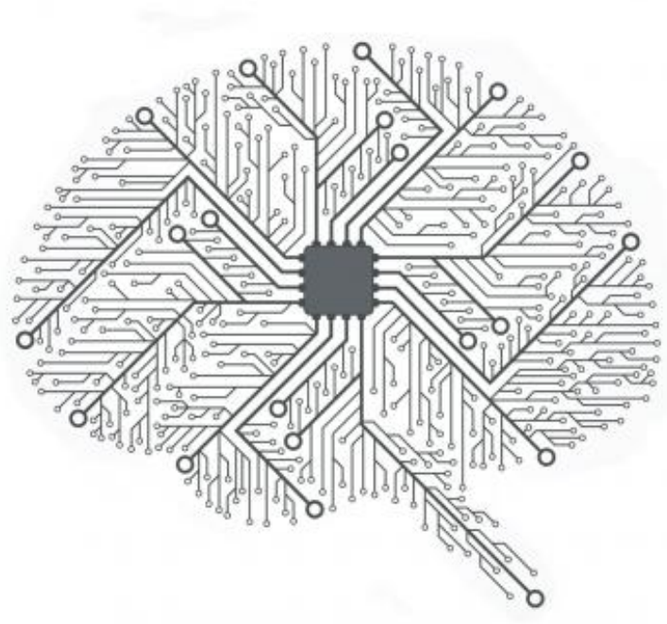
- 合理的：在给定信息情况下，最大化的达到预先制定的目标
- 这里的合理性，主要关注决策本身（要做的行动），而不是决策形成背后的思维过程。

- 用行动结果的功效(utility)来衡量目标
- 合理的决策，是指最大化行动所能带来结果的功效。



第一章·绪论

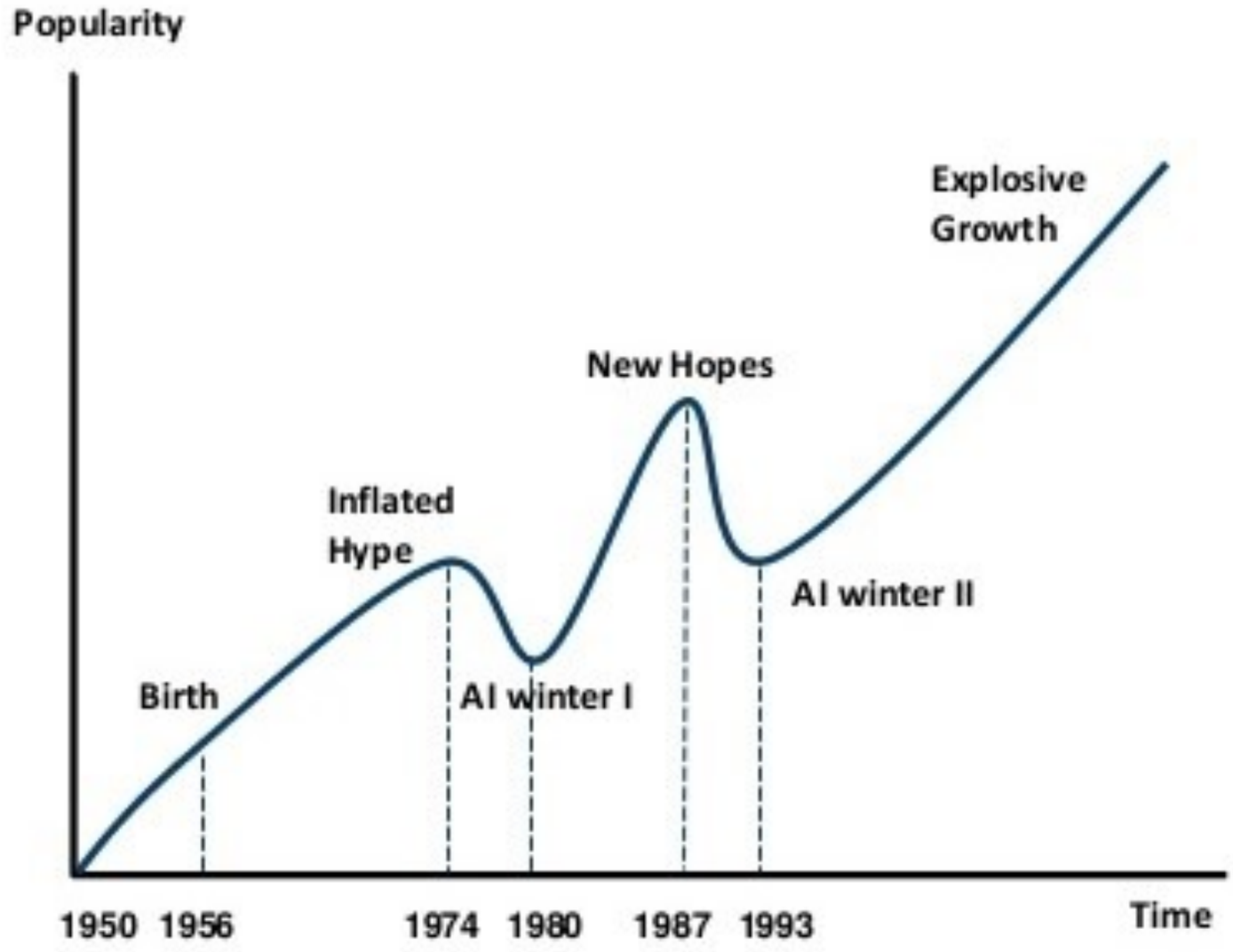
- 什么是人工智能
- **人工智能的历史**
- 目标和研究范式
- 人工智能模型
- 一个具体的例子



人工智能之前

- 哲学(从亚里士多德开始)
- 数学(逻辑，概率)
- 神经系统科学(神经元)
- 经济学(合理性，博弈论)
- 控制原理(反馈)
- 心理学(学习、认知模型)

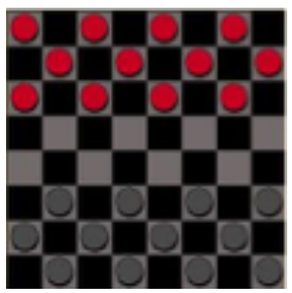
人工智能的历史



Timeline of AI Development

- **1950s-1960s**: First AI boom - the age of reasoning, prototype AI developed
- **1970s**: AI winter I
- **1980s-1990s**: Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s**: AI winter II
- **1997**: Deep Blue beats Gary Kasparov
- **2006**: University of Toronto develops Deep Learning
- **2011**: IBM's Watson won Jeopardy
- **2016**: Go software based on Deep Learning beats world's champions

人工智能的出现



- 西洋跳棋 (1952): 能够在未告知任何“策略”的情况下自我学习，并最终取得和人类业余高手相当的水平——Samuel

- 逻辑理论家 (1955): ：用搜索+启发式方法证明《数学原理》中的定理——Newell & Simon

乐观的期望

- Machines will be capable, within twenty years, of doing any work a man can do. —Herbert Simon

- Within 10 years the problems of artificial intelligence will be substantially solved. —Marvin Minsky

- I visualize a time when we will be to robots what dogs are to humans, and I’m rooting for the machines. —Claude Shannon

巨大的失望

- 一个著名的机器翻译的例子：

The spirit is willing but the flesh is weak.

(Russian)

The vodka is good but the meat is rotten.

- 1966年,咨询委员会的一份报告认为“尚不存在通用科学文本的机器翻译,近期也不会有”。随后取消了学术翻译项目的所有美国政府资助。这标志着第一次AI寒冬。

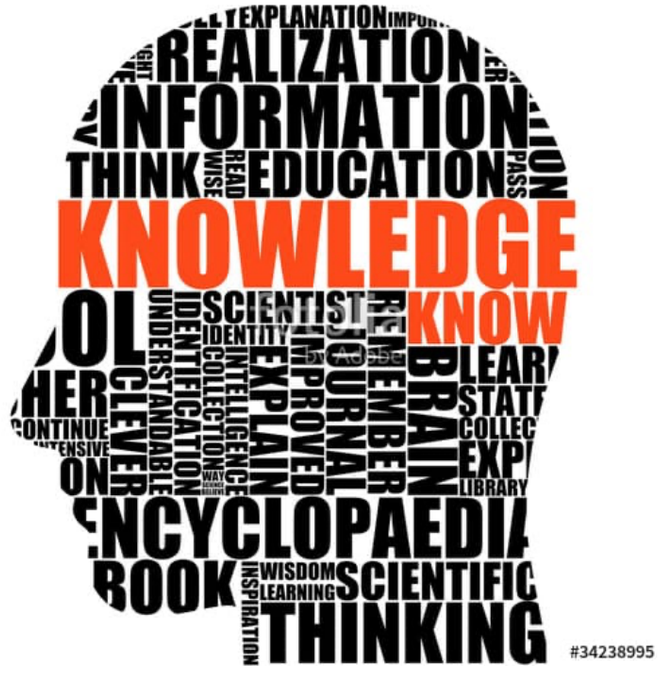
早期AI的启示

- 问题:
  - 计算能力有限: AI问题的搜索空间呈指数增长, 远超当时计算机的能力限制。
  - 信息有限: AI问题极其复杂 (大量词, 大量的物体, 大量的概念)
- 贡献:
  - Lisp, 自动垃圾回收, 分时系统
  - 关键范式: 区分建模(modeling)和推断(inference)

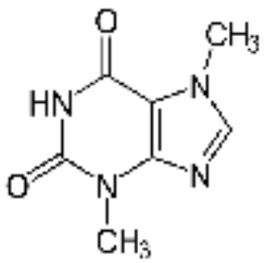
基于知识的系统（70-80s）

- 专家系统：以规则的形式从专家那里获取特定领域的知识：

if [premises] then [conclusion]



基于知识的系统



- DENDRAL: 通过质谱推断分子结构

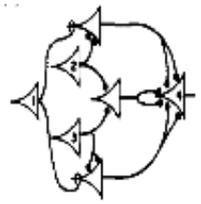
- MYCIN: 诊断血液感染，推荐抗生素

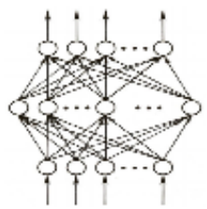
- XCON: 将客户订单转换为零件规格；到1986年，每年为DEC节省4,000万美元

基于知识的系统

- 贡献:
  - 影响行业的第一个实际应用程序
  - 知识有助于遏制指数增长
- 问题:
  - 知识不是确定性规则，需要建模不确定性
  - 需要大量的人工来创建规则，难以维护
- 1987年：Lisp机器市场崩溃，这标志着第二个AI寒冬

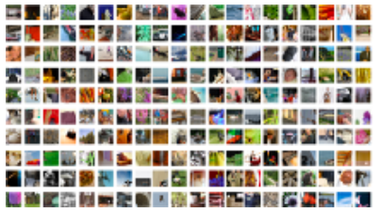
神经网络的回归





- 1943: 人工神经网络被发明 (McCulloch/Pitts)
- 1969: 证明线性模型无法解决**XOR**，杀死了神经网络研究 (Minsky/Papert)
- 1986: 反向传播技术在训练多层网络中的应用 (Rumelhardt, Hinton, Williams)
- 1989年：运用卷积神经网络识别USPS的手写数字（LeCun）
- 1991年：反向传播算法被指出存在梯度消失问题，神经网络研究再次沉寂 (Hochreiter)

深度神经网络的兴起



- DBN（2006）：提出用预训练的方式来抑制梯度消失问题

- AlexNet（2012）：对象识别方面的巨大进步；一夜之间转变了计算机视觉的研究方向

- AlphaGo（2016）：深度强化学习，击败世界冠军Lee Sedol

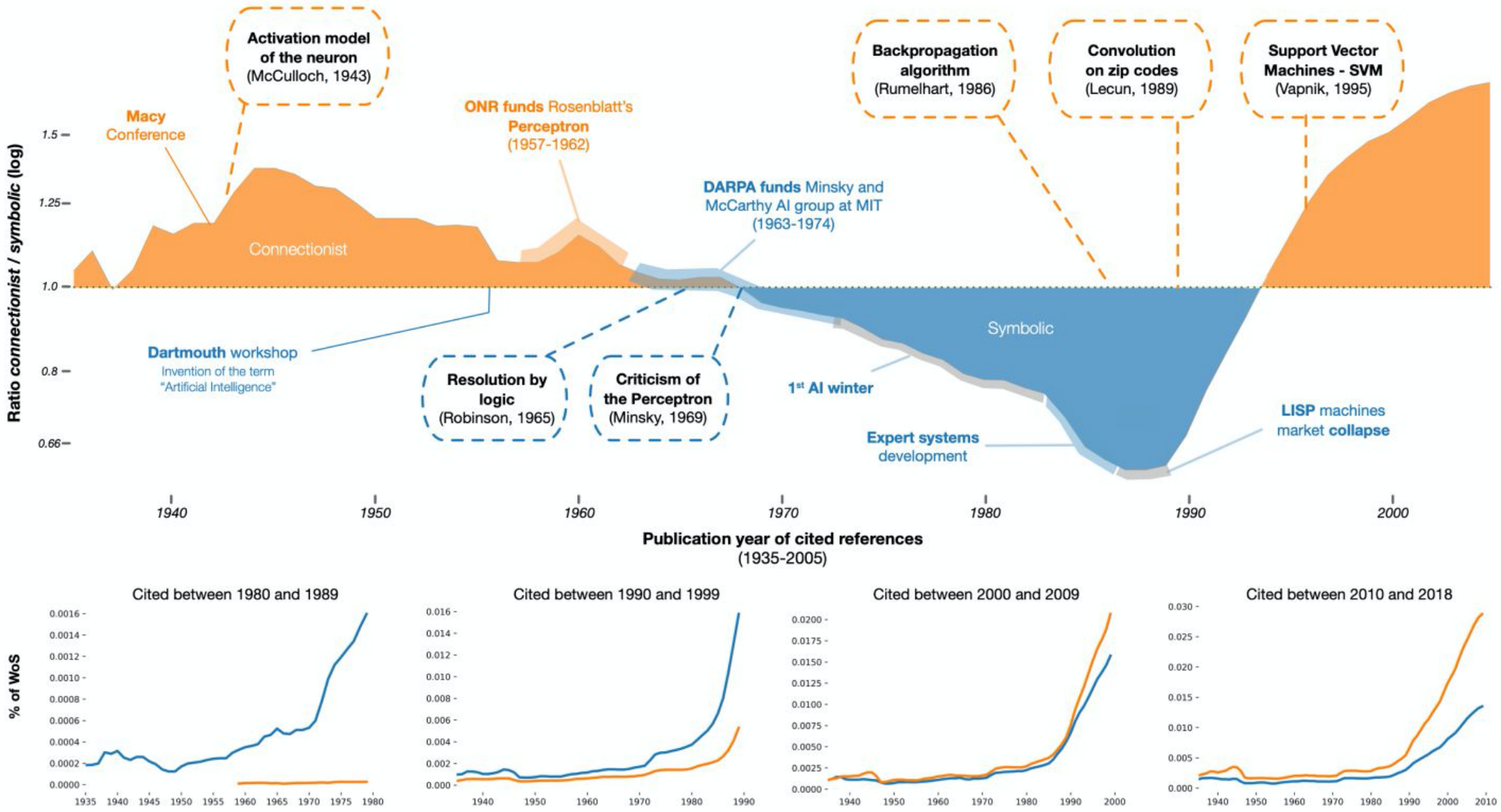
- GPT-3（2020）：巨大规模预训练语言模型，成功实现小样本学习

两大派系

- “符号主义”（Symbolicism），又称逻辑主义、计算机学派，主张用公理和逻辑体系搭建一套人工智能系统。

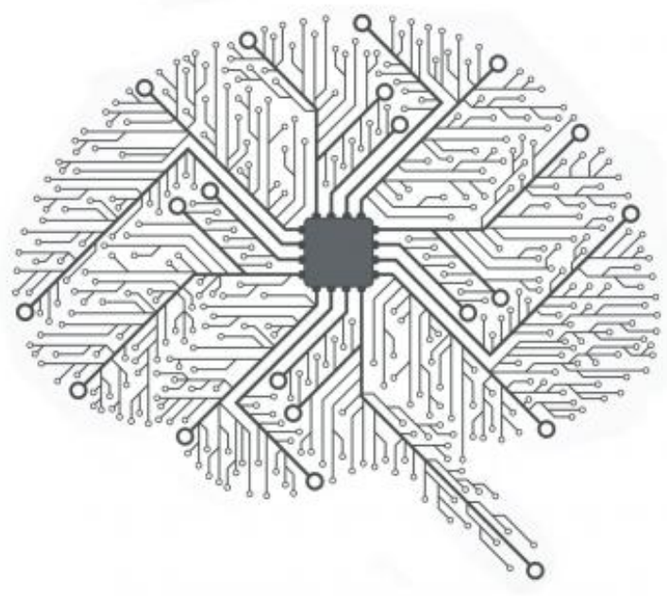
- “连接主义”（Connectionism），又叫仿生学派，主张模仿人类的神经元，用神经网络的连接机制实现人工智能。

两大派系


Activation model of the neuron (McCulloch, 1943)
Backpropagation algorithm (Rumelhart, 1986)
Convolution on zip codes (Lecun, 1989)
Support Vector Machines - SVM (Vapnik, 1995)
Macy Conference
ONR funds Rosenblatt's Perceptron (1957-1962)
DARPA funds Minsky and McCarthy AI group at MIT (1963-1974)
Connectionist
Symbolic
Dartmouth workshop
Invention of the term "Artificial Intelligence"
Resolution by logic (Robinson, 1965)
Criticism of the Perceptron (Minsky, 1969)
1st AI winter
Expert systems development
LISP machines market collapse
Ratio connectionist / symbolic (log)
1.5
1.25
1.0
0.8
0.66
1940
1950
1960
1970
1980
1990
2000
Publication year of cited references
(1935-2005)
% of WoS
Cited between 1980 and 1989
Cited between 1990 and 1999
Cited between 2000 and 2009
Cited between 2010 and 2018

第一章·绪论

- 什么是人工智能
- 人工智能的历史
- **目标和研究范式**
- 人工智能模型
- 一个具体的例子



目标和研究范式

- 两个研究目标：AI Agent vs. AI Tools
- 三步研究范式：建模、学习和推断

两个研究目标

- AI agents: 如何创造智慧?

- AI tools: 如何造福人类社会?

AI agents

感知　　行动　　语言

知识　　推理　　学习

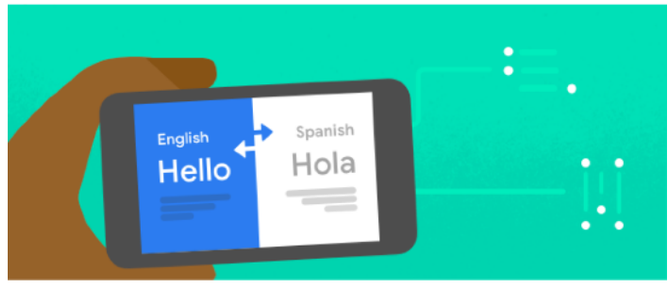
AI agents 发展现状

- 机器：特定的任务，数百万个学习样本
- 人类：多样化的任务，很少的学习样本

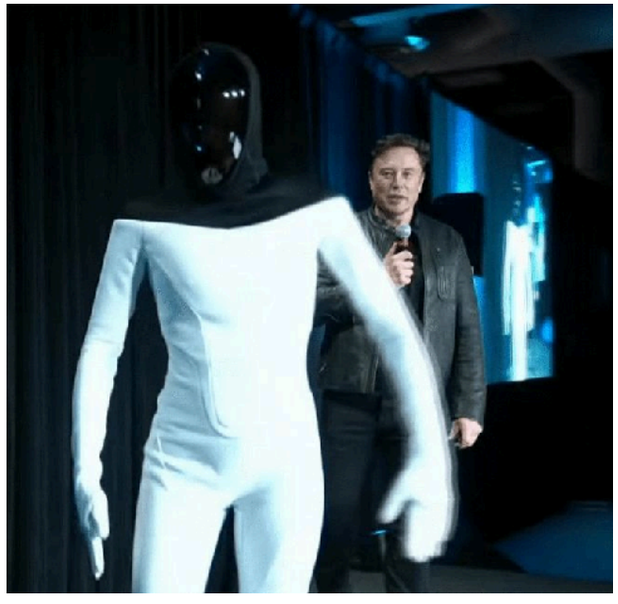
AI tools


ACM Chess Challenge
Garry Kasparov
vs
TechDribble


Lee Sedol

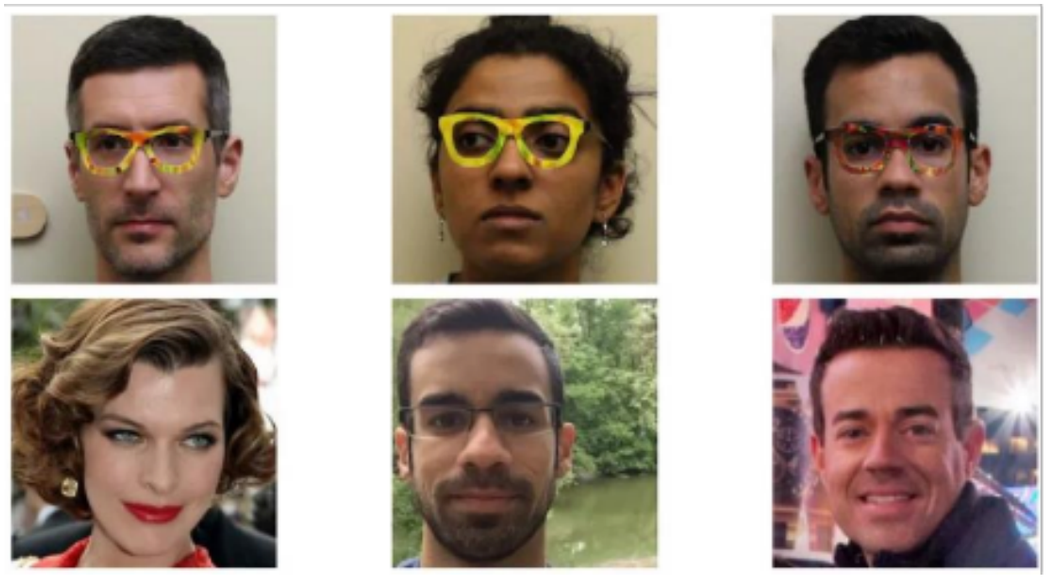

English
Hello
Spanish
Hola


ORIGINAL
DEEPFAKE



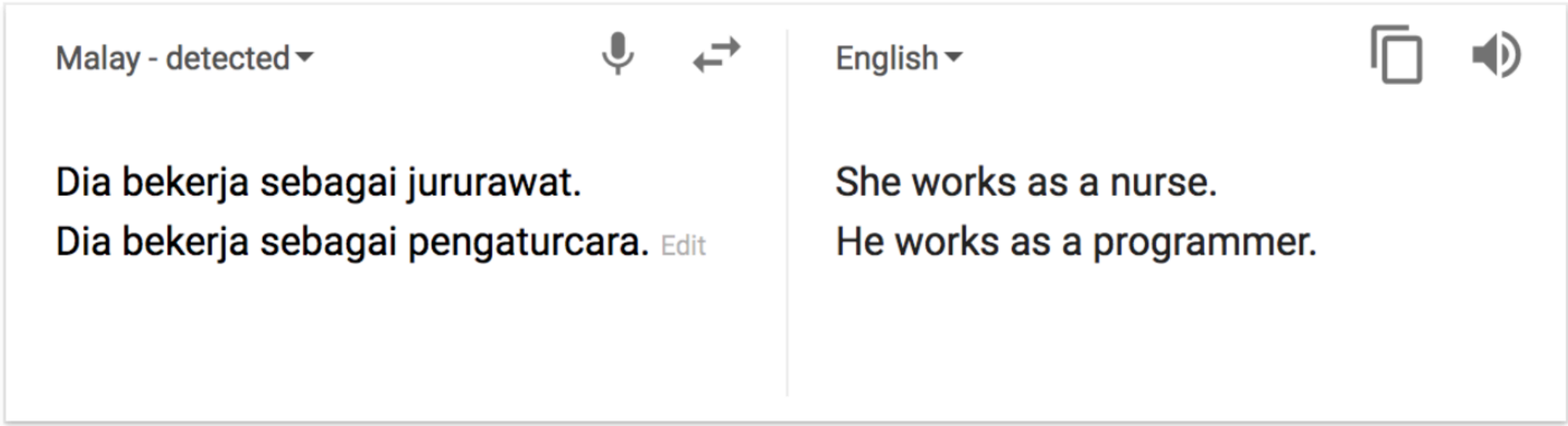
AI tools的安全问题

Robust Physical-World Attacks on Deep Learning Models



Search Results Web results Real and Stealthy Attacks on State-of-the-Art Face Recognition

AI tools的偏见问题



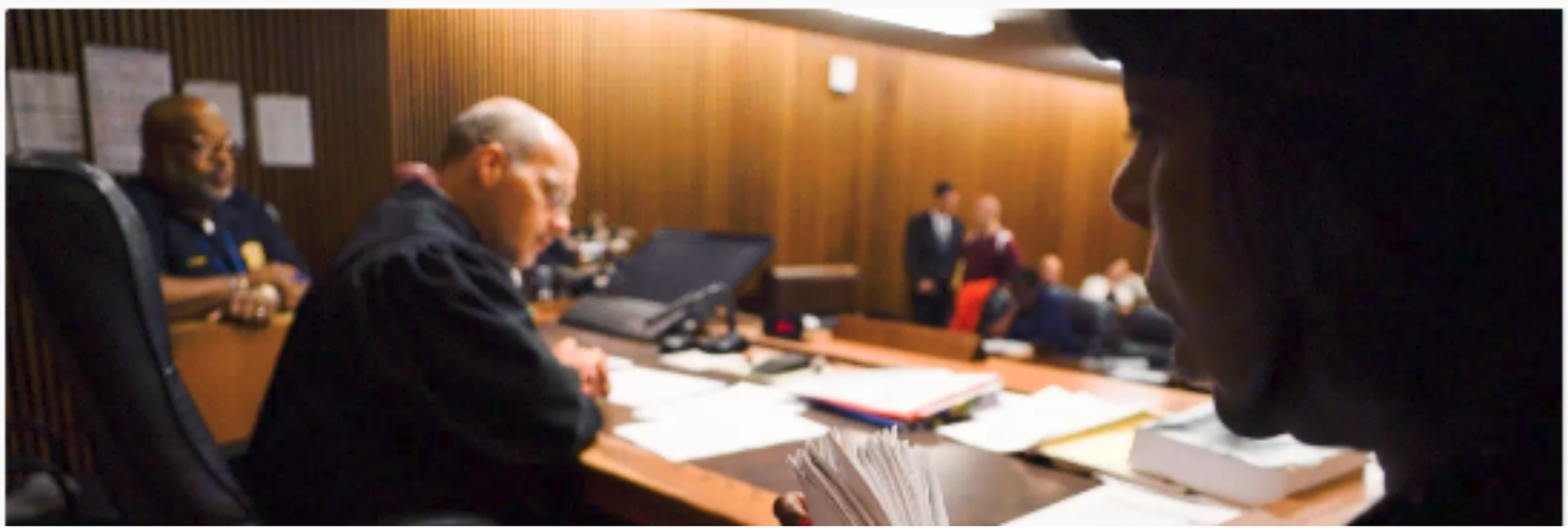
society $\Rightarrow$ data $\Rightarrow$ predictions

AI tools的偏见问题

- Northpointe: COMPAS 预测犯罪风险得分 (1-10)
  - 对于判断一个人有没有再次犯罪，黑人2倍可能被（错误地）归类为5分或以上

California just replaced cash bail with algorithms

by Dave Davignon • September 4, 2018



两个研究目标的现状

- AI agents: 远未达到，是否会达到?

- AI tools: 成果显著，但同时也带了新的社会问题

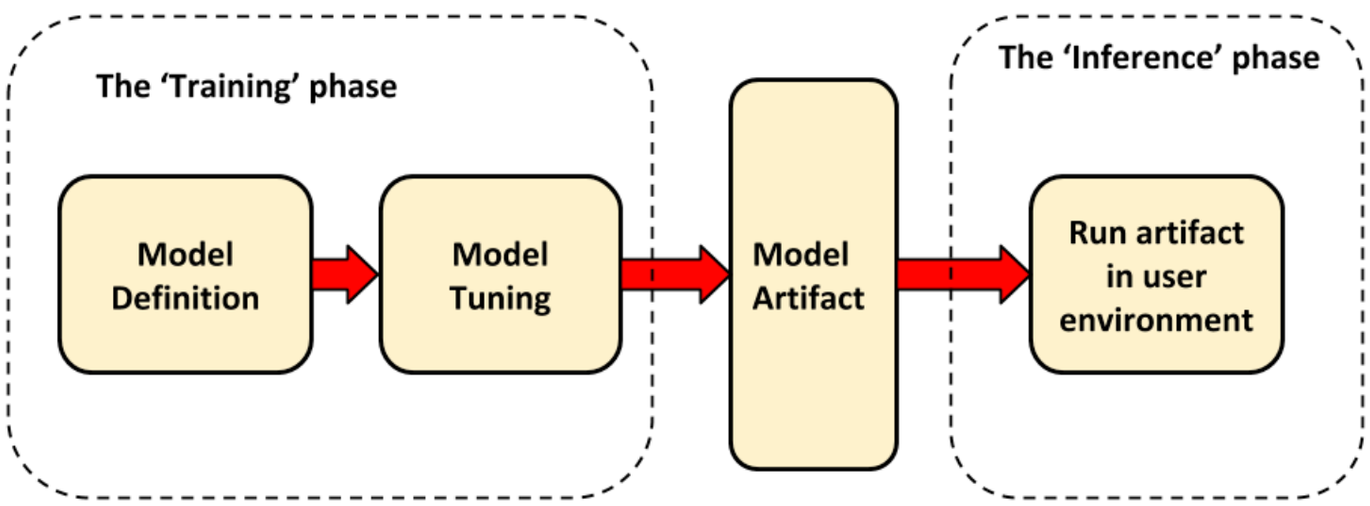
人工智能的研究范式

人工智能的研究范式

- 建模：处理现实世界中复杂的问题，并将其打包为称为模型的整齐的形式化数学对象。

- 推断：在模型的指导下对具体问题进行回答。

- 学习：从数据中学习模型的参数。

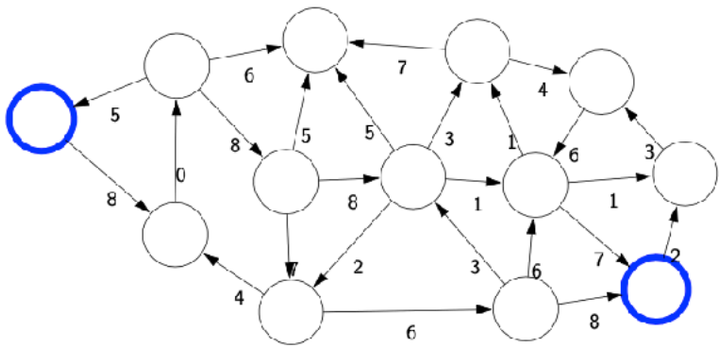


人工智能的研究范式：建模

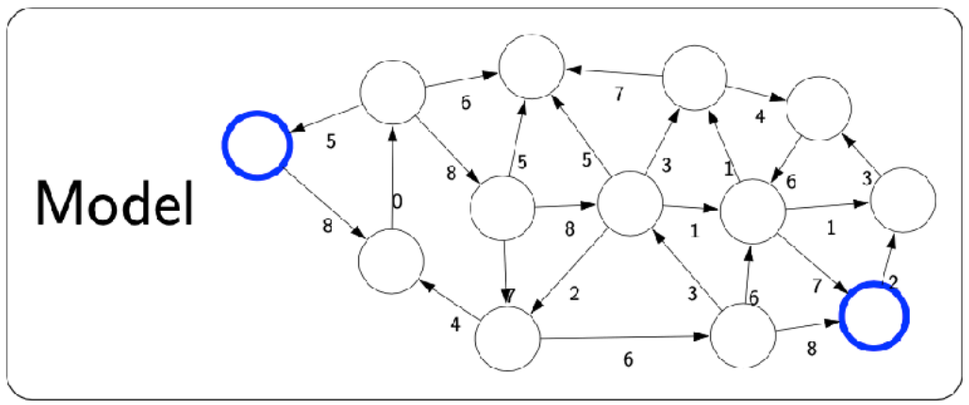
Real world

建模

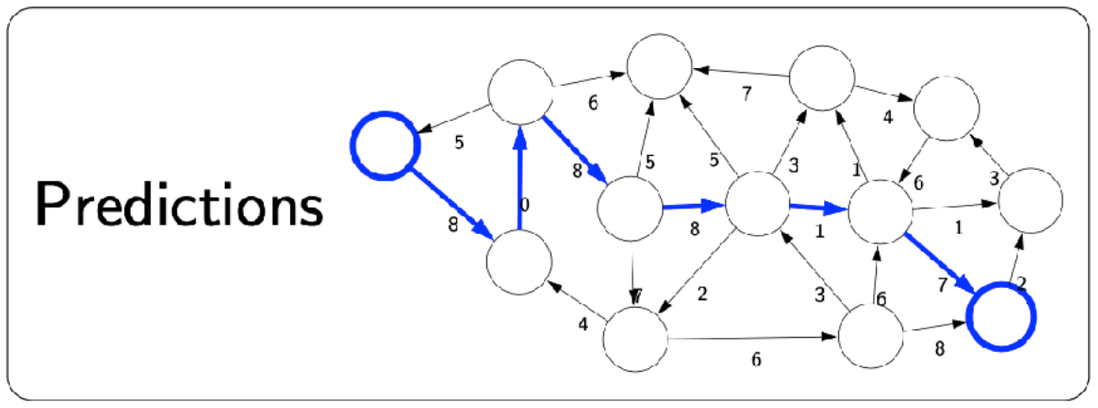
Model



人工智能的研究范式：推断

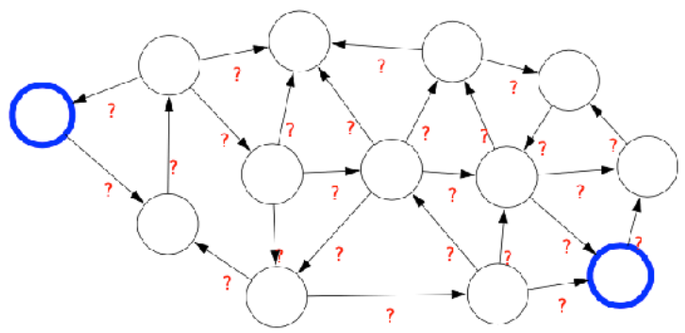


推断



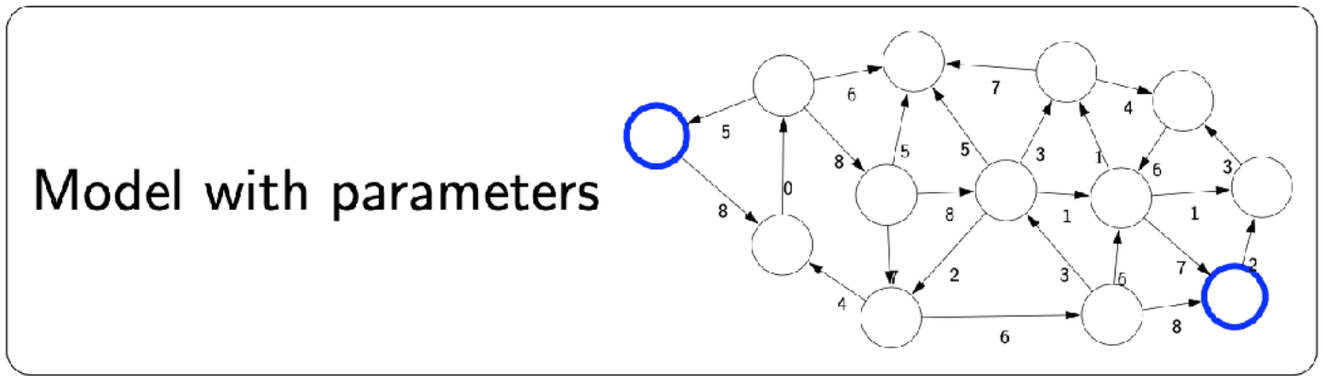
人工智能的研究范式：学习

Model without parameters



＋ 数据

学习



机器学习

数据 → 模型

- AI近期成功的主要推动力
- 从“代码”到“数据”以管理信息复杂性
- 需要信念的飞跃：泛化

人工智能模型

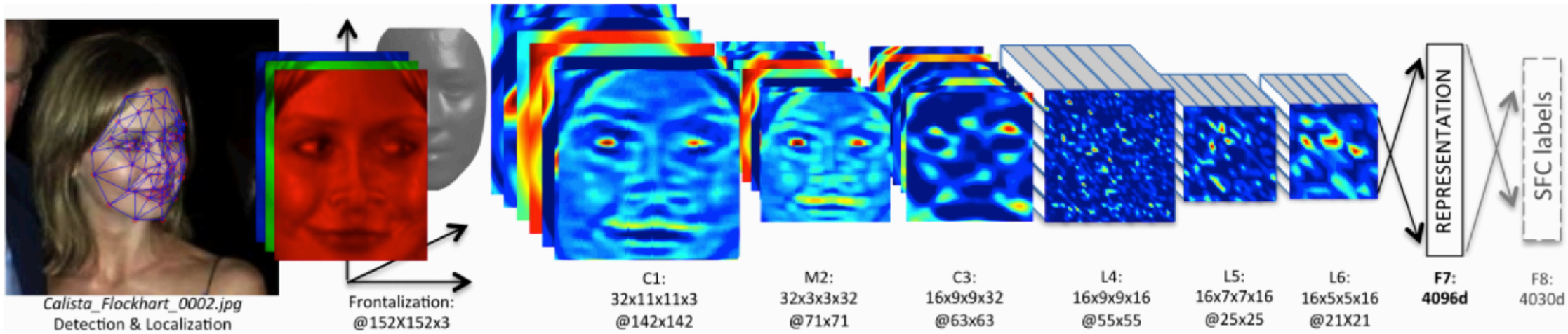
反射　　状态　　条件　　逻辑

“低级人工智能”　　“高级人工智能”

这是谁?

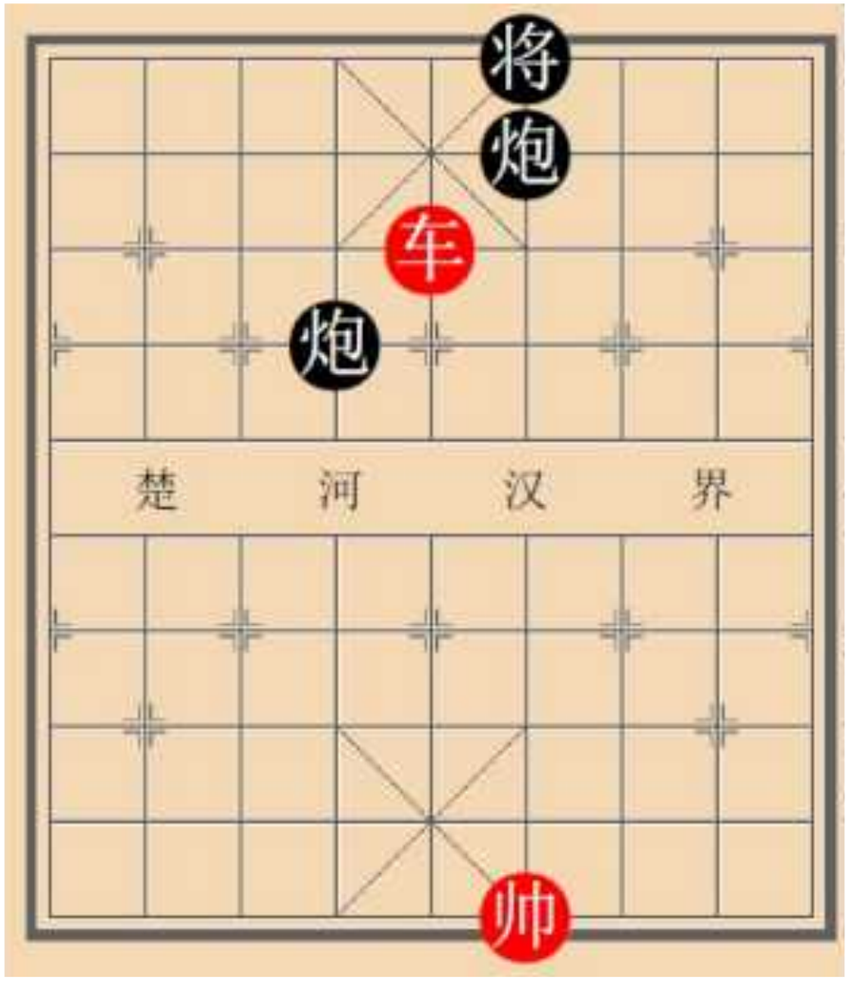
基于反射的模型

- 基于反射的模型仅对给定的输入执行固定的计算序列。



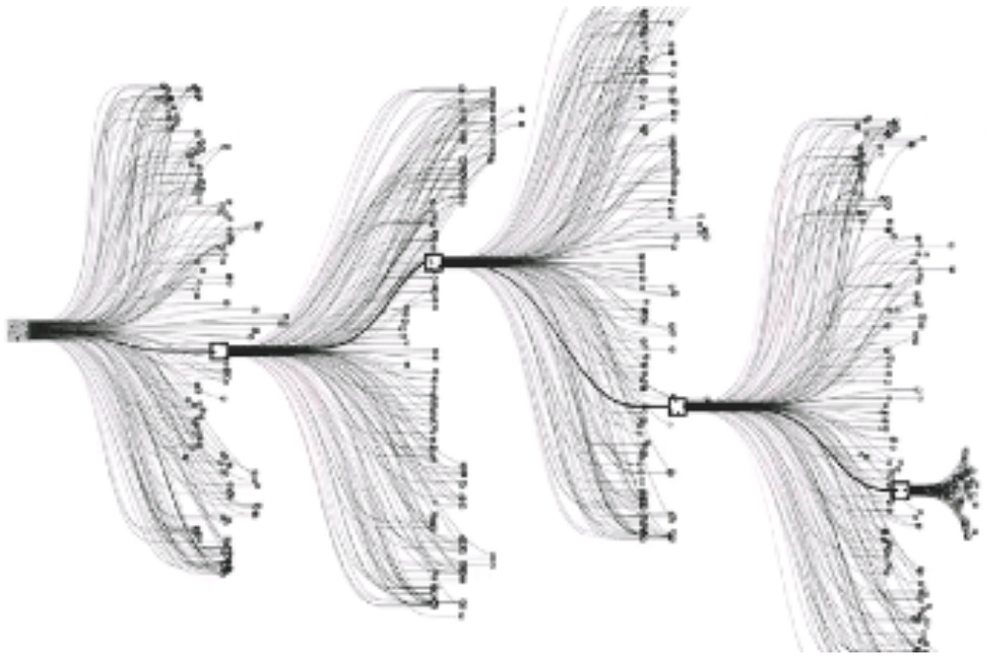
- 决策过程完全基于前向传播，是最常见的模型

象棋残局



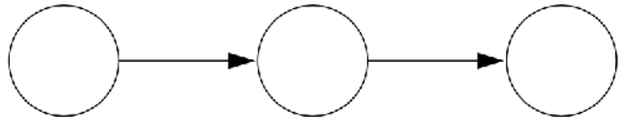
基于状态的模型

- 游戏：国际象棋，围棋，吃豆人，星际争霸等。
- 机器人技术：运动规划
- 自然语言生成：机器翻译，对话、问答

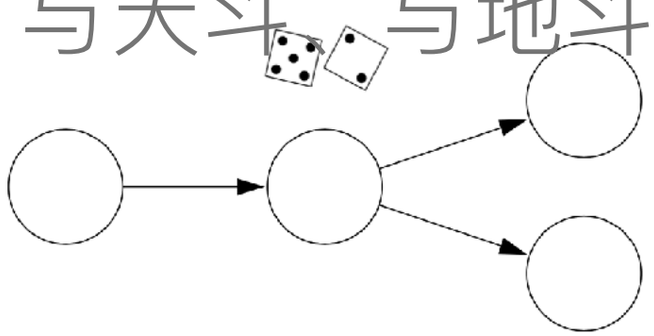


基于状态的模型 2

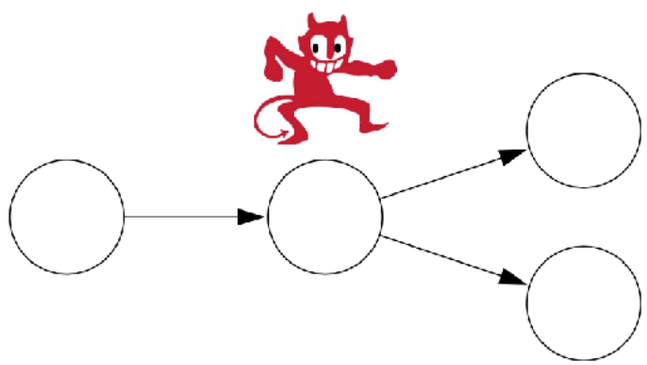
- 搜索问题: 一切都是可控的



- 马尔科夫决策过程: 与天斗，与地斗



- 博弈: 与人斗

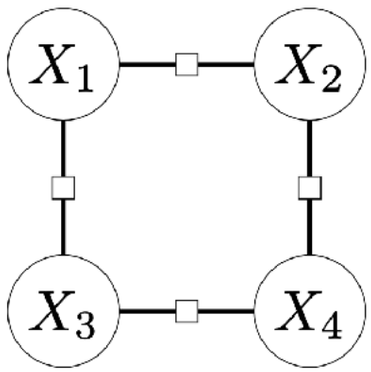


数独

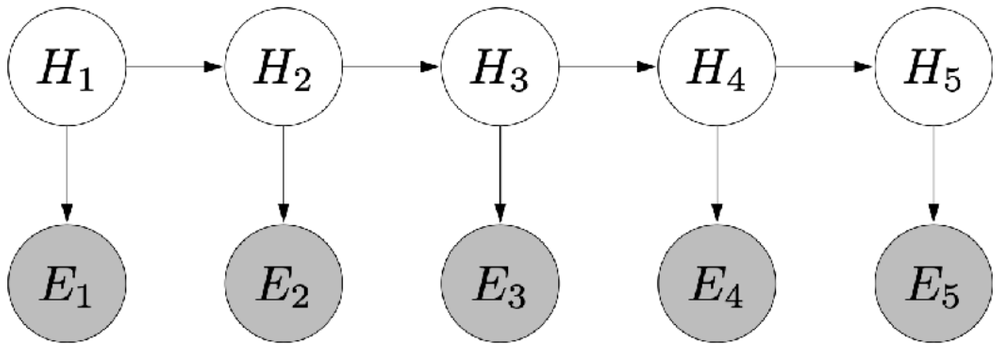
9		8				4		6
			3	6	2			
3					9			5
	3			9		8	2	
	4		2		3		1	
	9	2		1			6	
5			9					1
			1	3	8			
2		3				7		8

基于条件的模型

- 约束满足问题：硬约束（例如，数独、排课）



- 贝叶斯决策：软性依赖（例如，天气预报）



基于逻辑推理的模型

- 需要：
  - 汇总多源异构信息
  - 对信息进行深层推理

模型求解/优化

- 离散模型优化：find the best discrete object

$$\min_{p \in \text{Paths}} \text{Cost}(p)$$

典型算法：动态规划

- 连续模型优化：find the best vector of real numbers

$$\min_{\mathbf{w} \in \mathbb{R}^d} \text{TrainingError}(\mathbf{w})$$

典型算法：梯度下降

一个离散的例子

- 输入：两个字符串，s 和 t
- 输出：将 s 转换为 t 所需的最小字符插入、删除和替换次数

“cat”, “cat” ⇒ 0

“cat”, “dog” ⇒ 3

“cat”, “at” ⇒ 1

“cat”, “cats” ⇒ 1

“a cat!”, “the cats!” ⇒ 4

一个连续的例子

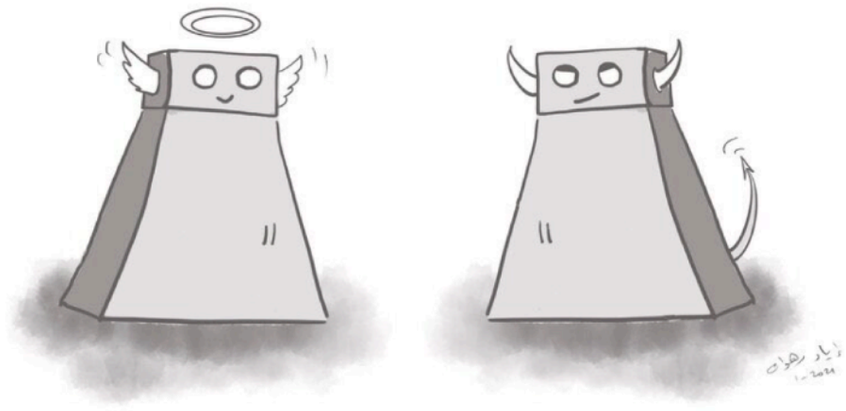
- 输入：一些二元组集合 $\{(x_1, y_1), \ldots, (x_n, y_n)\}$
- 输出：$w \in R$ that minimizes the squared error

$$F(w) = \sum_{i=1}^{n} (x_i w - y_i)^2$$

{(2,4)}	⇒ 2
{(2, 4), (4, 2)}	⇒ ?
{(2, 4), (4, 2), (6, 6)}	⇒ ?

总结

- 历史: 来自逻辑学、神经科学、统计学——是多学科的大熔炉!
- 研究范式：建模 [reflex, states, variables, logic] + 推断 + 机器学习
- 人工智能具有高度的社会影响，但是我们如何才能正向地引导和利用它?



<< Good vs. Evil. It's a bot time. >>