第一讲：探索智能本质之路

First Lecture：Pursuing the Nature of Intelligence

马毅

香港大学计算与数据科学学院院长

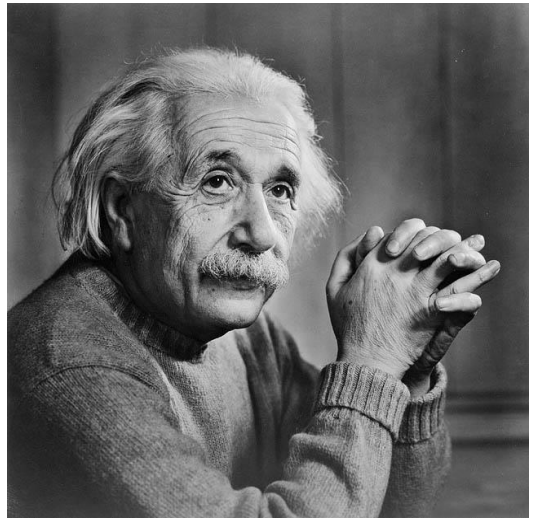
香港大学数据科学研究院院长

演讲主题

回归理论基石，探寻智能本质

"Everything should be made as simple as possible , but not any simpler."

-- Albert Einstein



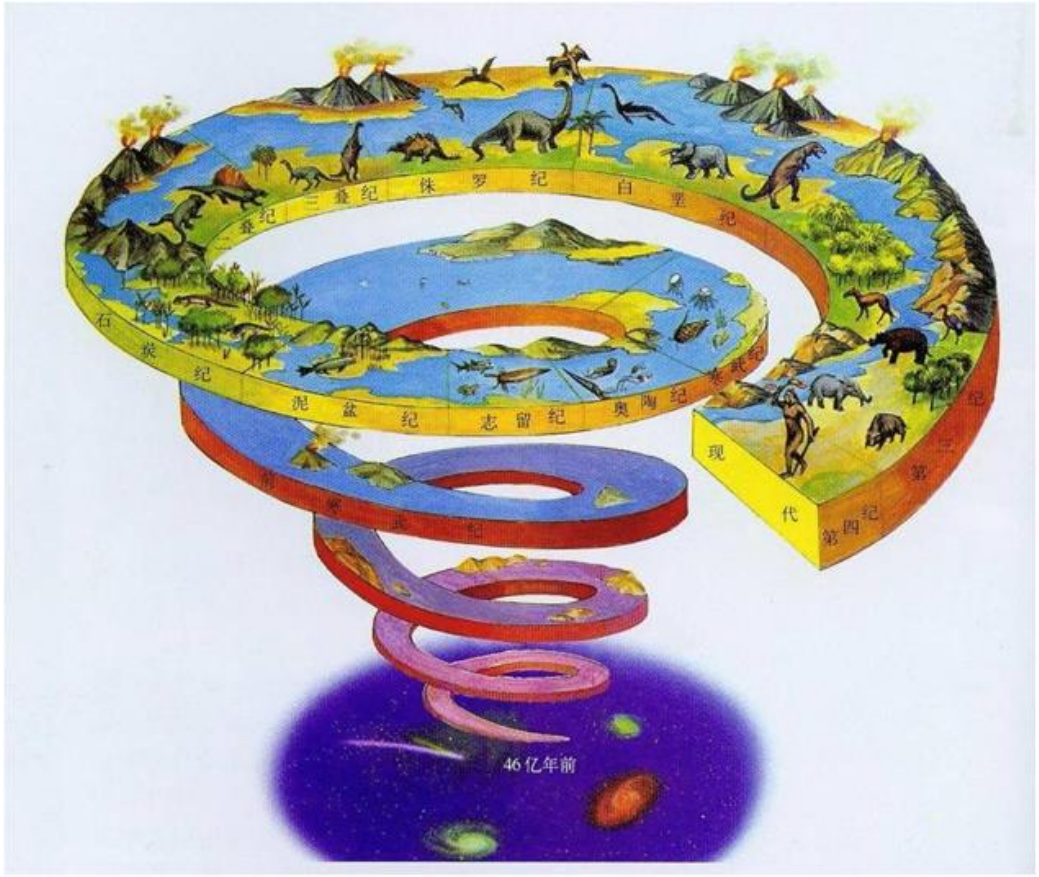
"所有的事情都应该尽可能的简单到不能再简单。"

-- 阿尔伯特·爱因斯坦

生命与智能的进化

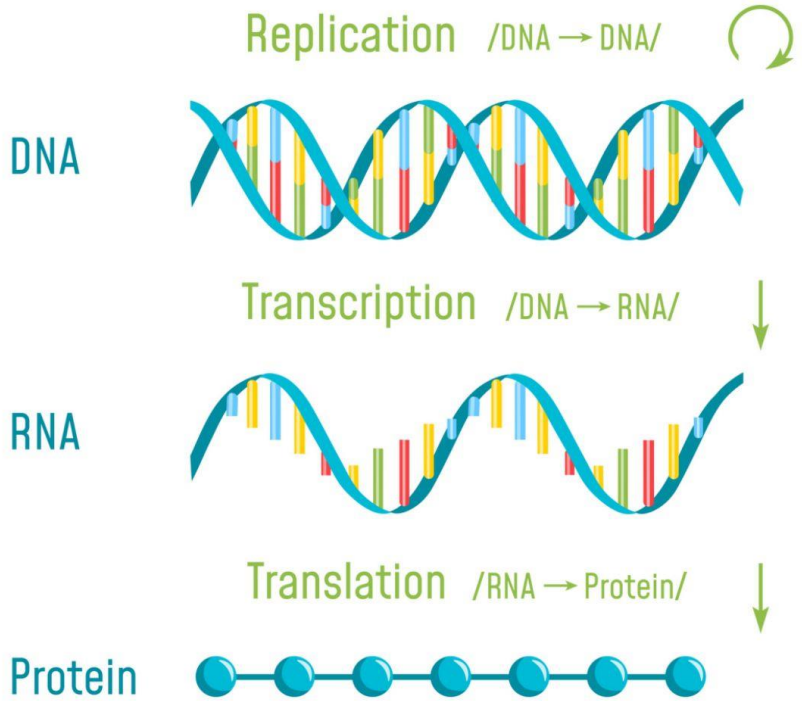
自然界生物的起源与进化，是智能的机制起作用的结果

生物进化的历程



生物物种的演化整体趋向繁荣和复杂

遗传记忆、变异进化。



生命与智能的进化

自然界生物的起源与进化，是智能的机制起作用的结果

生命智能从**物种依赖基因遗传和自然选择进化（Phylogenetic）**

演进到**个体具有后天学习与适应的智能（Ontogenetic）**







约37亿年前 | 生命起源

约5亿年前 | 寒武纪生物大爆发

约4亿年前 | 鱼类

约3.6亿年前 | 两栖动物

约2.5亿年前 | 爬行动物

约2亿年前 | 鸟类与哺乳动物

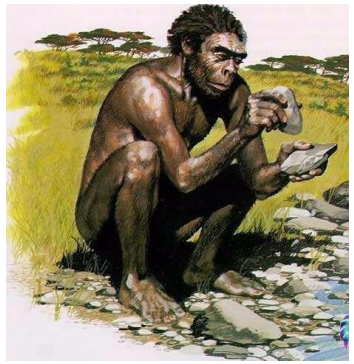
约31万年前 | 智人

生命依赖智能而存在的根本原因：我们的世界不是随机的，而是可预测的。

人类群体智能的发展

人类文明与科学的发展，也是智能机制起作用的结果

人类是以“个体后天学习为主”智能的终极代表。通过语言，发展出群体智能，并产生抽象思维、数理逻辑。

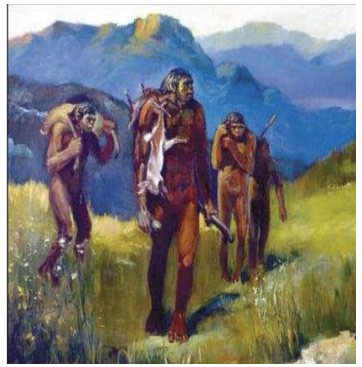


约31万年前

智人的诞生

使用工具

合作狩猎



约7万年前

语言的出现

群体智慧

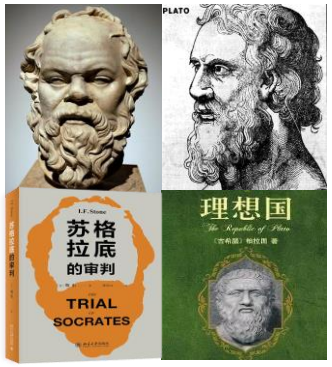
信息的即时传递

约公元前3500年

文字的发明

知识的积累和传承

文明的记忆



公元前6世纪

逻辑和数学

形式逻辑

抽象概念

数学和哲学

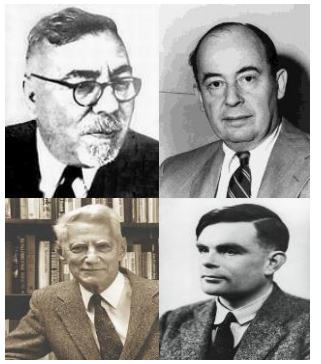
14至18世纪

文艺复兴与科学革命

科学方法的确立

人文的复兴，

实验和观察的重要性



1940年代

人工智能的萌芽

计算机的发明

试图将自然界的智能

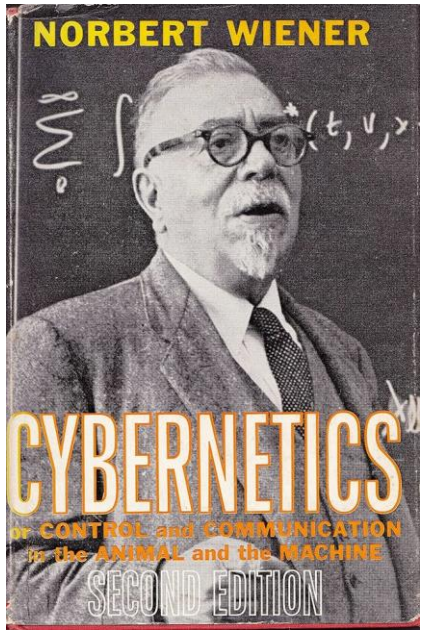
转移到机器上

机器智能的起源

1940年代，人们开始尝试让机器模拟生物尤其是动物的智能。

第一个神经元数学模型、第一台计算机，同时维纳、香农、冯·诺伊曼等人开始提出智能的各种特征和机制。

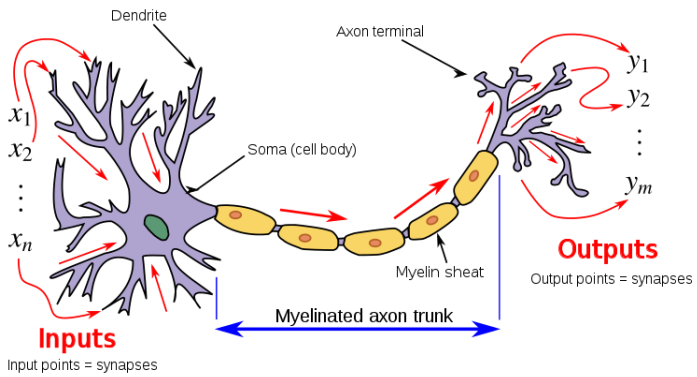
- 1943年，**人工神经网络**，沃伦·麦卡洛克和沃尔特·皮茨
- 1940年代，**图灵机及图灵测试**，艾伦·图灵等
- 1944年，**博弈论与计算机结构**，约翰·冯·诺依曼
- 1948年，**信息论**，克劳德·香农
- 1948年，**控制系统论**，**诺伯特•维纳**

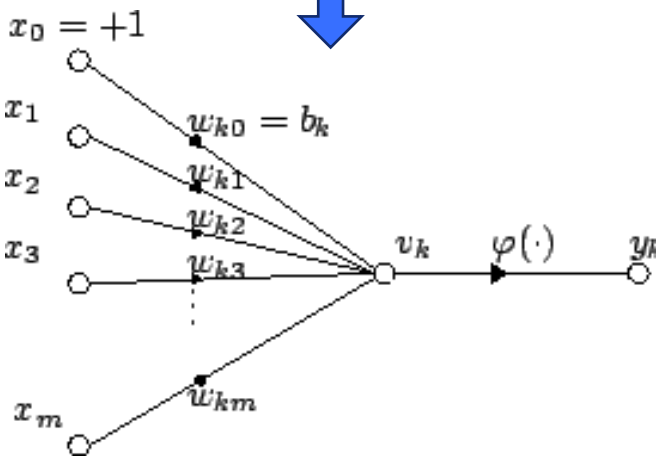


《控制论》
在动物与机器中的控制和通信

人工神经网络与神经网络

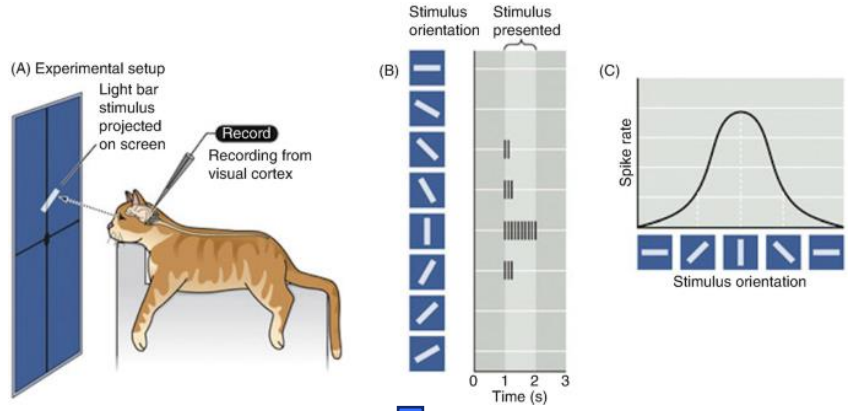
Golgi and Cajal 1888 (1901 Nobel Prize)

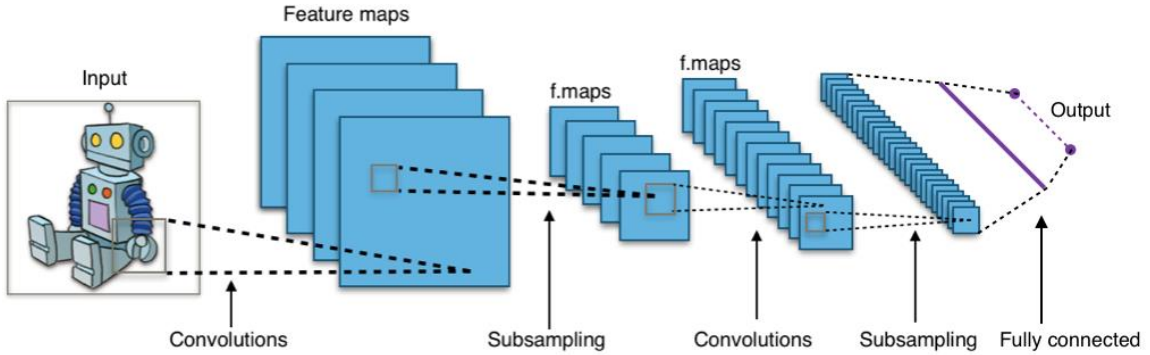




Warren McCulloch & Walter Pitts 1948

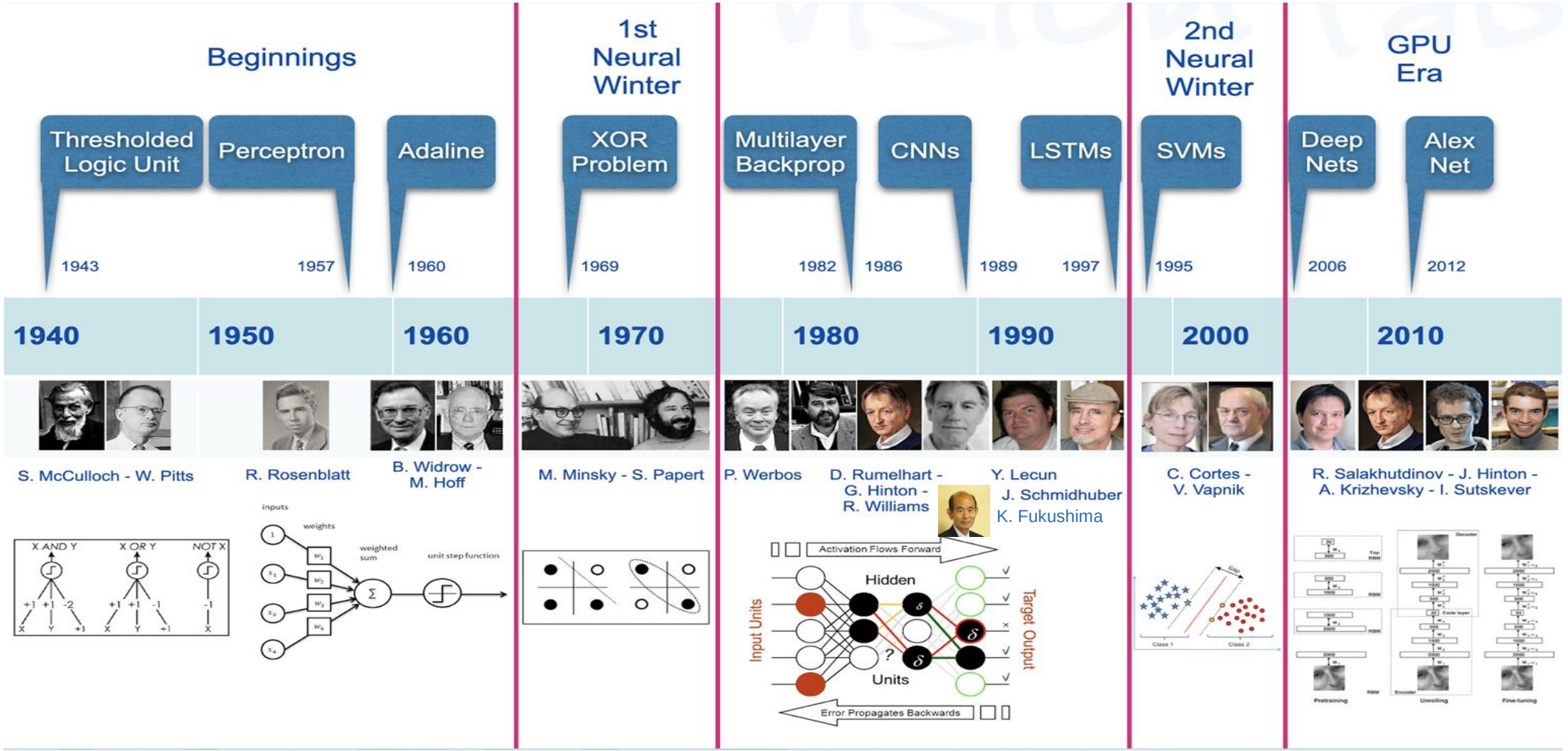
Hubel and Wiesel 1959 (1981 Nobel Prize)





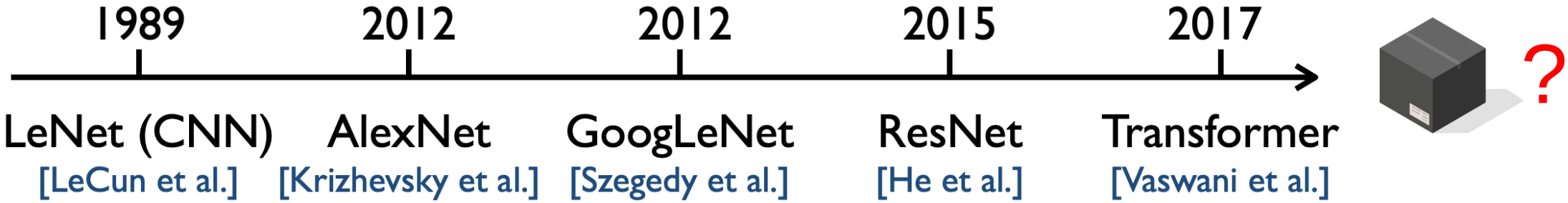
Fukushima 1980 & LeCun 1989 (Turing Award)

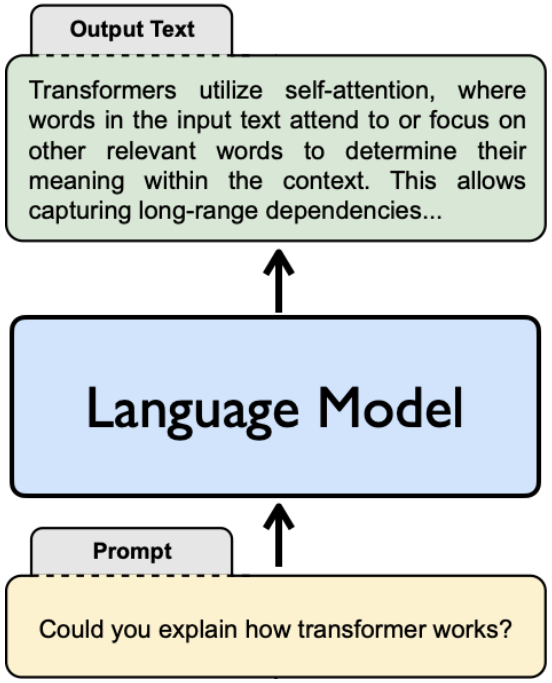
机器智能（人工神经网络）的近代历史

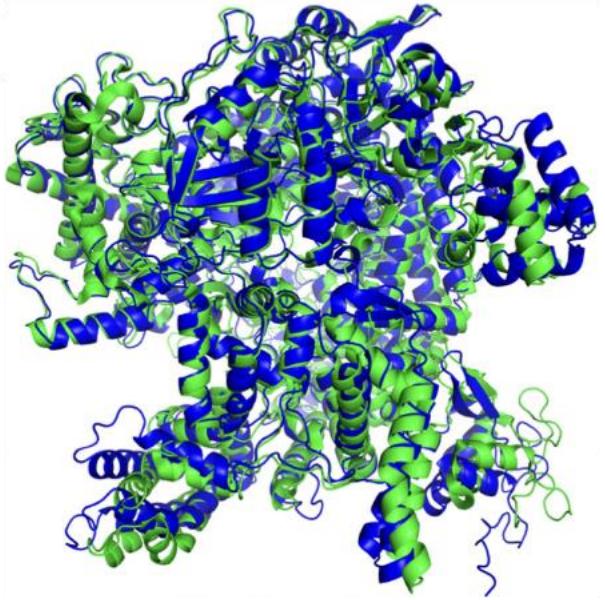


人工神经网络的现代化（2012年后）

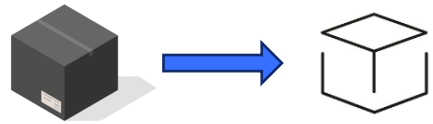
深度神经网络







为什么一定要把AI模型从黑盒变白盒？



现在基于深度网络的人工智能系统都是基于经验与试错设计出来的黑盒模型（炼丹术？）。

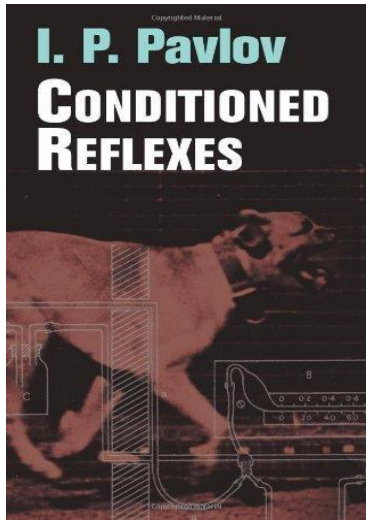
黑盒无法解释；无法保障安全；难以优化改进；无法持续学习。甚至容易被人利用制造恐惧。

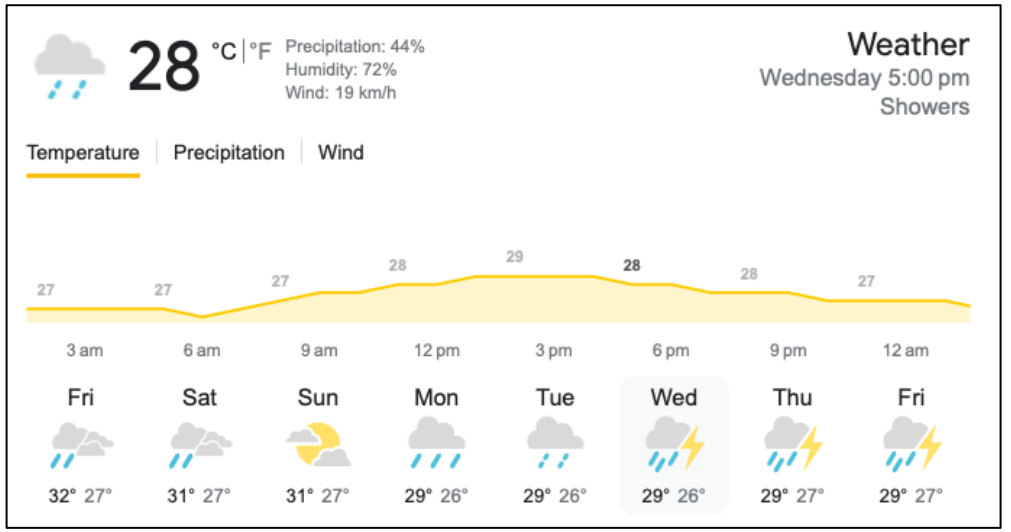
学什么（简约）	怎么学（深度网络）	学正确（自洽）
• 学数据可预测的结构 • 高维空间的低维结构 • 增加信息增益减少编码率	• 迭代实现优化算法 • 逐步优化编码率 • 深度网络从黑盒变成可解释白盒	• 闭环反馈 • 自主纠错 • 编码解码

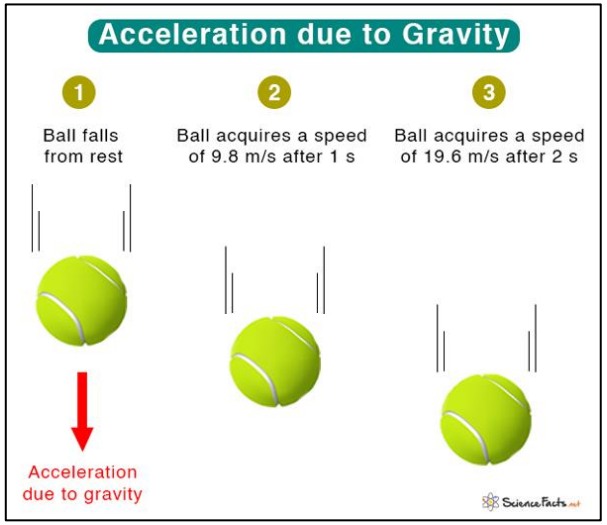
生命依赖智能而存在的根本原因：我们的世界不是随机的，而是可预测的。

学什么　　从感知外部世界的数据里学习可用于预测的信息

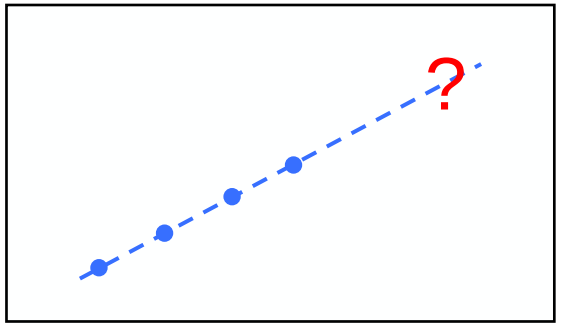
（每个人包括动物都是牛顿，都学到了外部世界精确的物理模型）

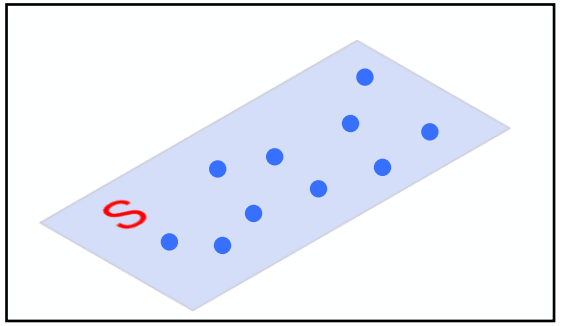


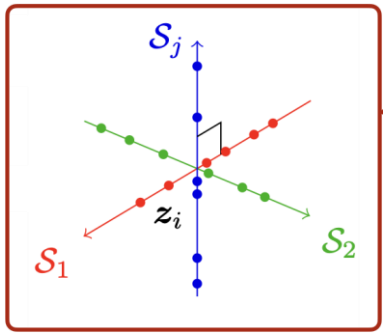


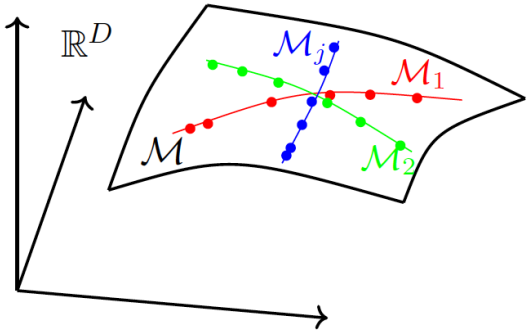


在数学上，可预测的信息都统一以数据在高维空间中的低维结构体现出来



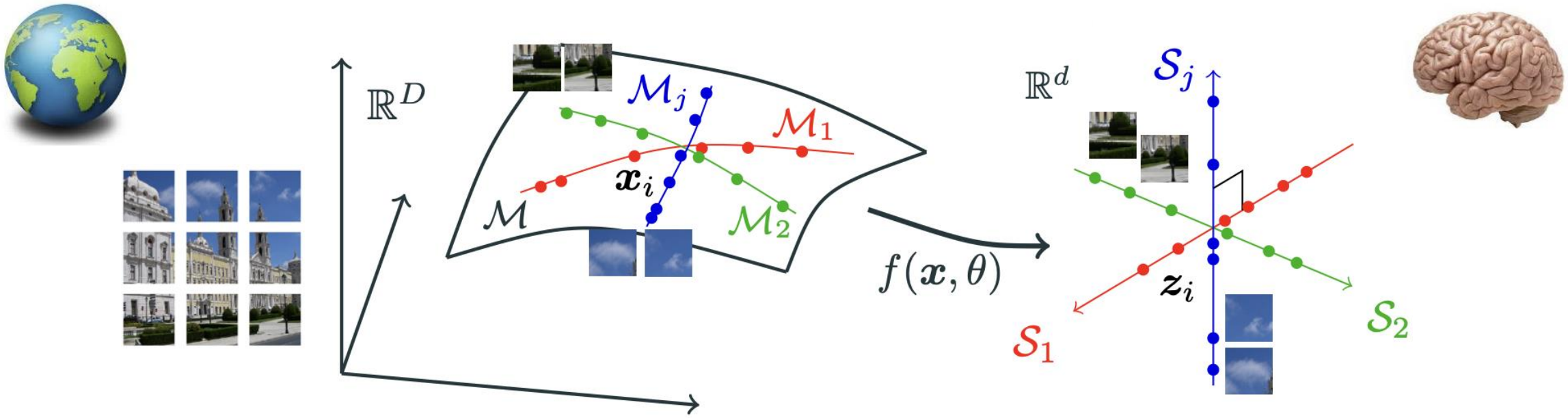






学什么（简约与压缩）

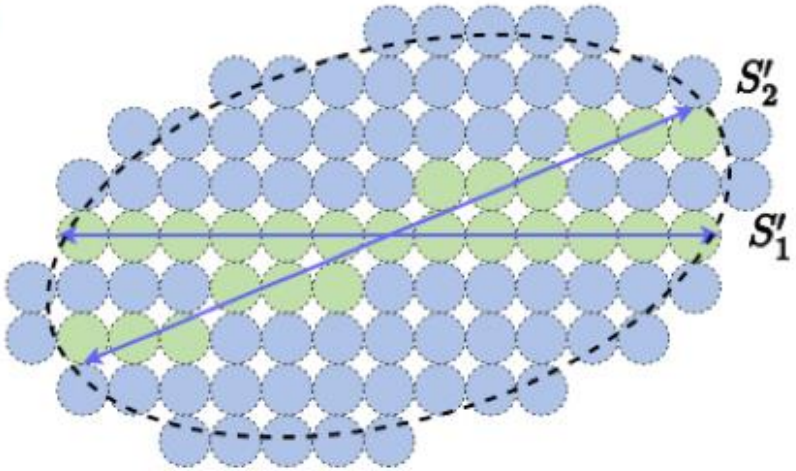
找到高维数据里的低维结构，并把这些结构组织、表示好（例如线性化、正交化）



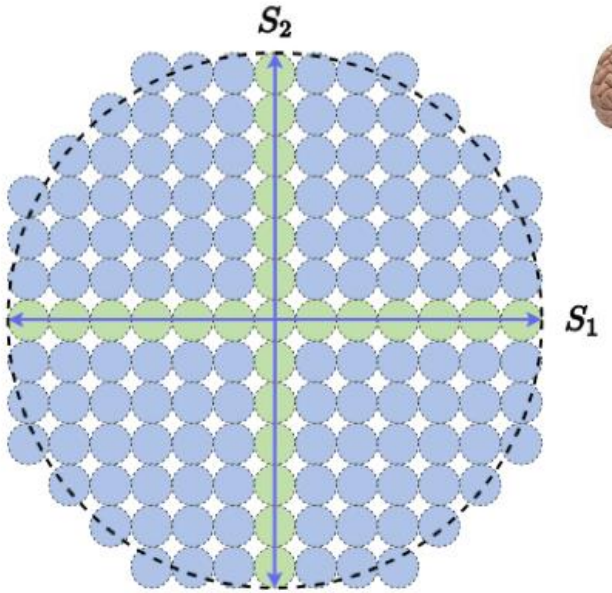
学习所依赖的最本质的统一计算机制：同类相聚、异类相斥。

学什么（简约与压缩）

学习所依赖的最本质的统一计算机制：同类相聚、异类相斥。



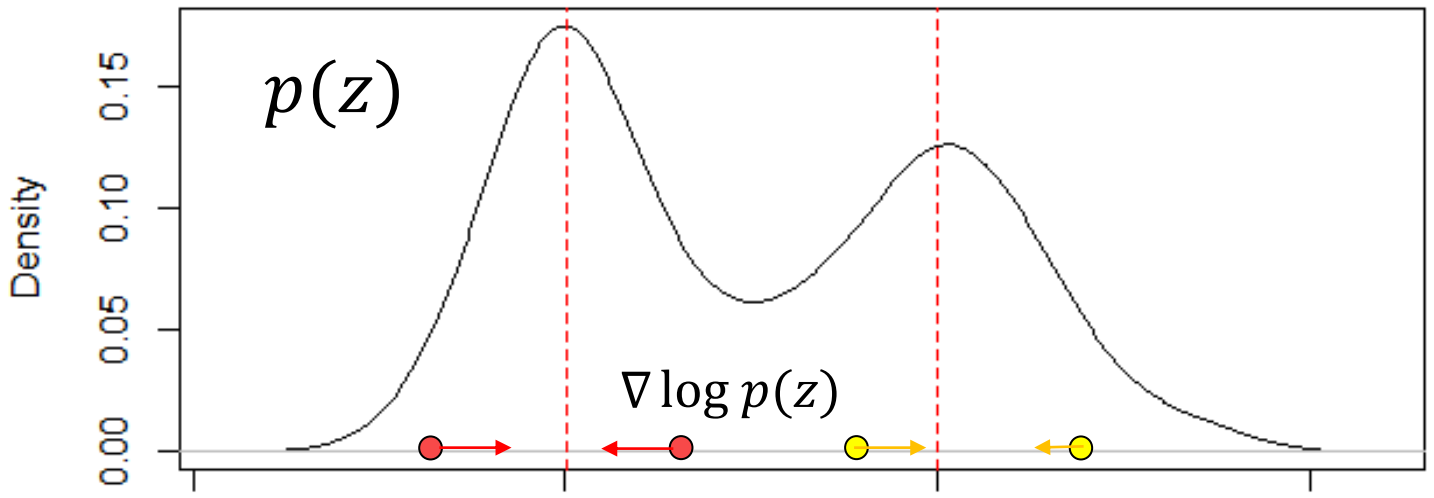
$$\boldsymbol{X} \xrightarrow{f(\boldsymbol{X};\boldsymbol{\theta})} \boldsymbol{Z}(\boldsymbol{\theta})$$



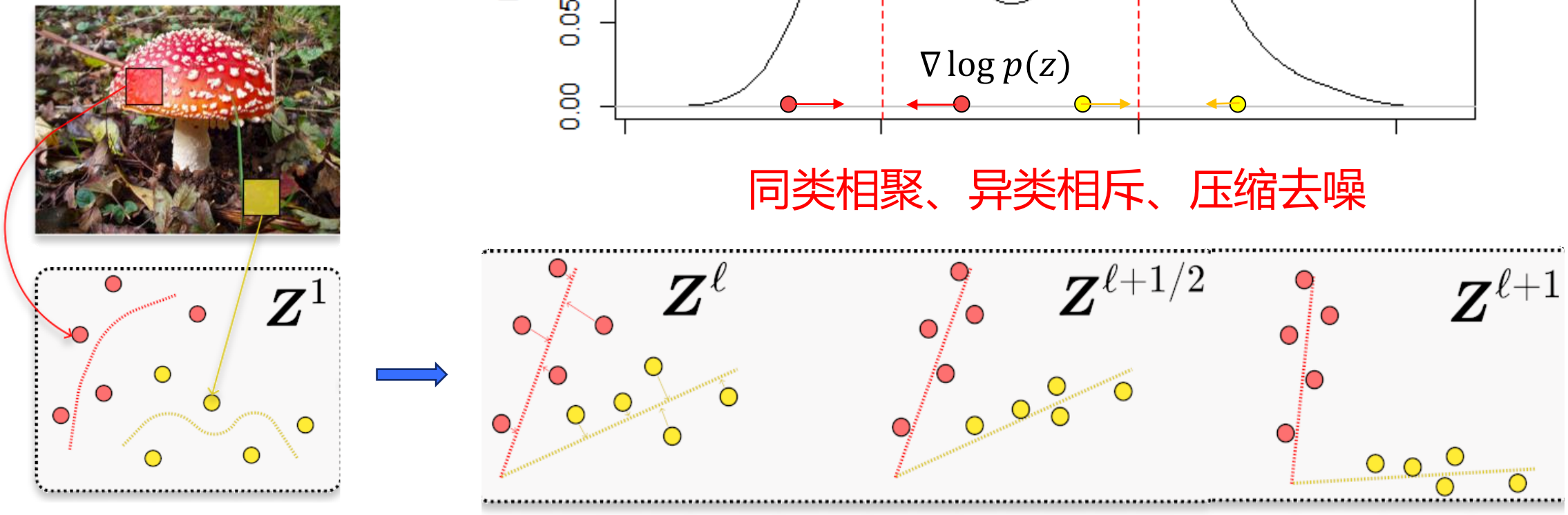
在计算上，这等价于最大化信息增益或者减少编码率

怎么学习（逐步迭代优化）

计算上如何最大化信息增益：逐步迭代地压缩以减少数据分布的编码率 $\int -\log p(z)$



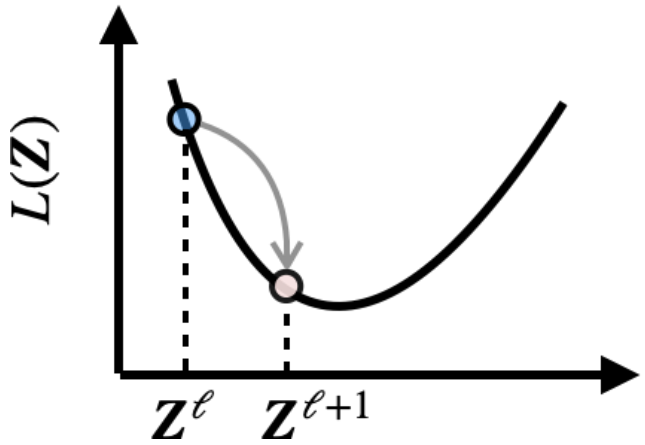
同类相聚、异类相斥、压缩去噪



怎么学习（通过深度网络实现逐步迭代优化）

最大化信息增益或者减少编码率（梯度下降）

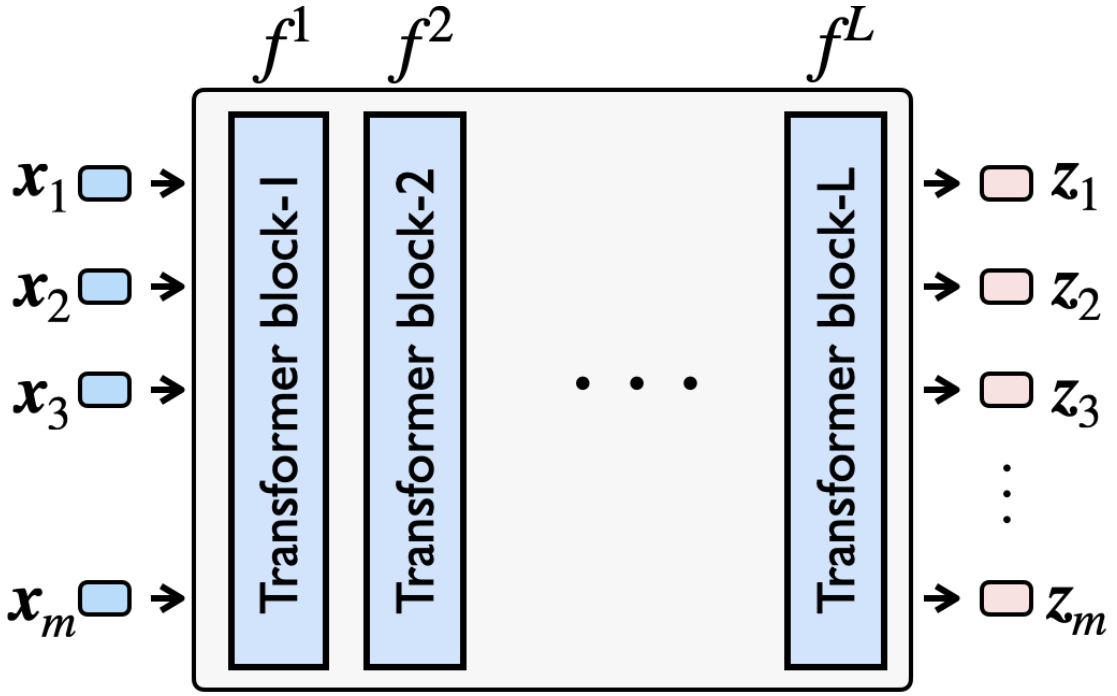
$$\min L(\boldsymbol{Z}) = -\log p(\boldsymbol{Z})$$



$$\boldsymbol{Z}^{\ell+1} = \boldsymbol{Z}^{\ell} - \eta \cdot \nabla L(\boldsymbol{Z})\Big|_{\boldsymbol{Z}=\boldsymbol{Z}_{\ell}} \qquad f^{\ell}$$

深度神经网络（例如 Transformer）

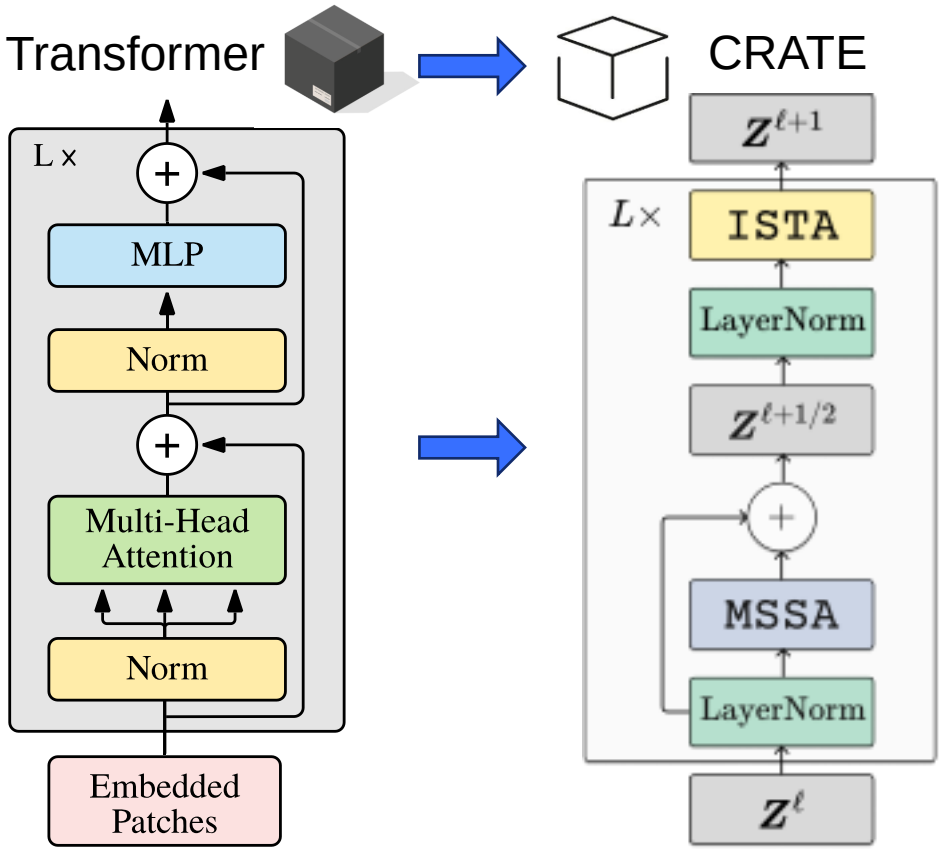
逐层迭代地实现优化编码率的梯度下降算子



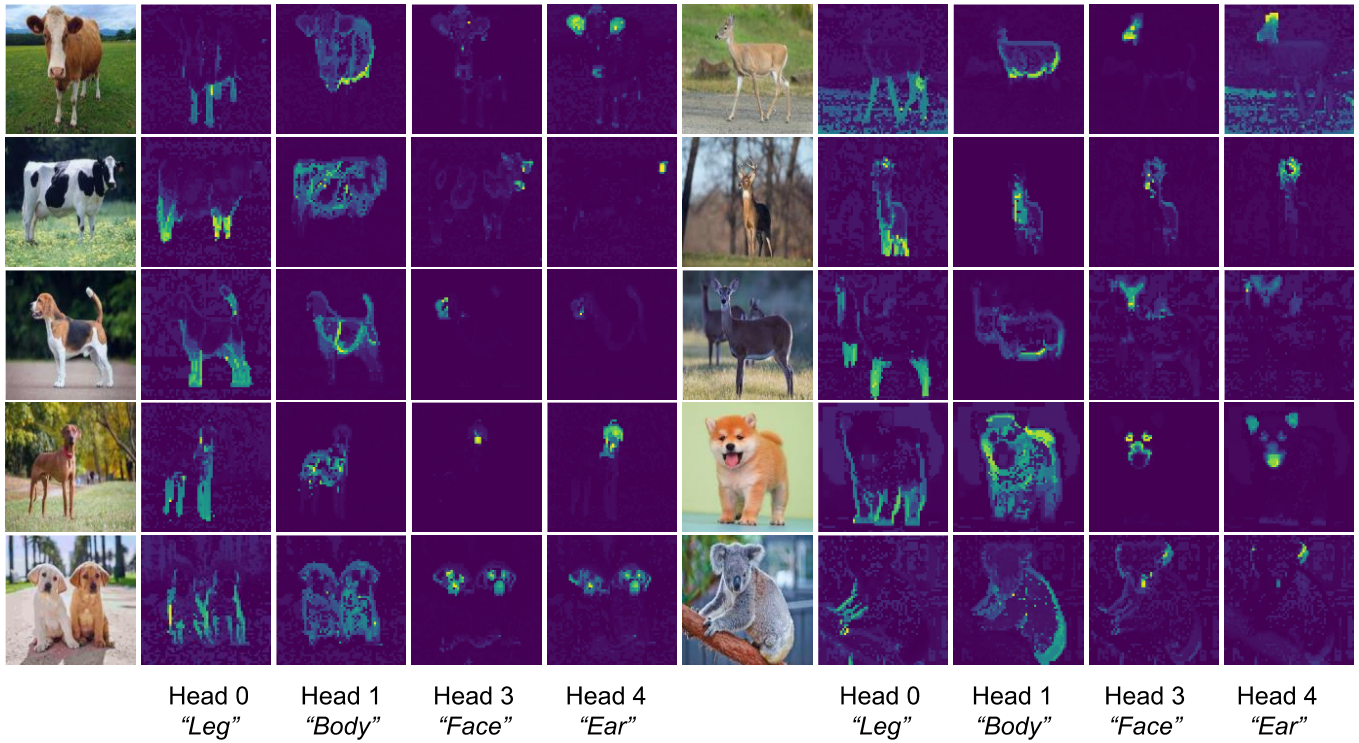
$$f : \boldsymbol{X} = \boldsymbol{Z}^0 \xrightarrow{f^1} \boldsymbol{Z}^1 \xrightarrow{f^2} \boldsymbol{Z}^2 \longrightarrow \cdots \xrightarrow{f^L} \boldsymbol{Z}^L = \boldsymbol{Z}$$

怎么学习（深度网络白盒化）

由优化学习目标的算法推导而来， 数学上完全可解释!

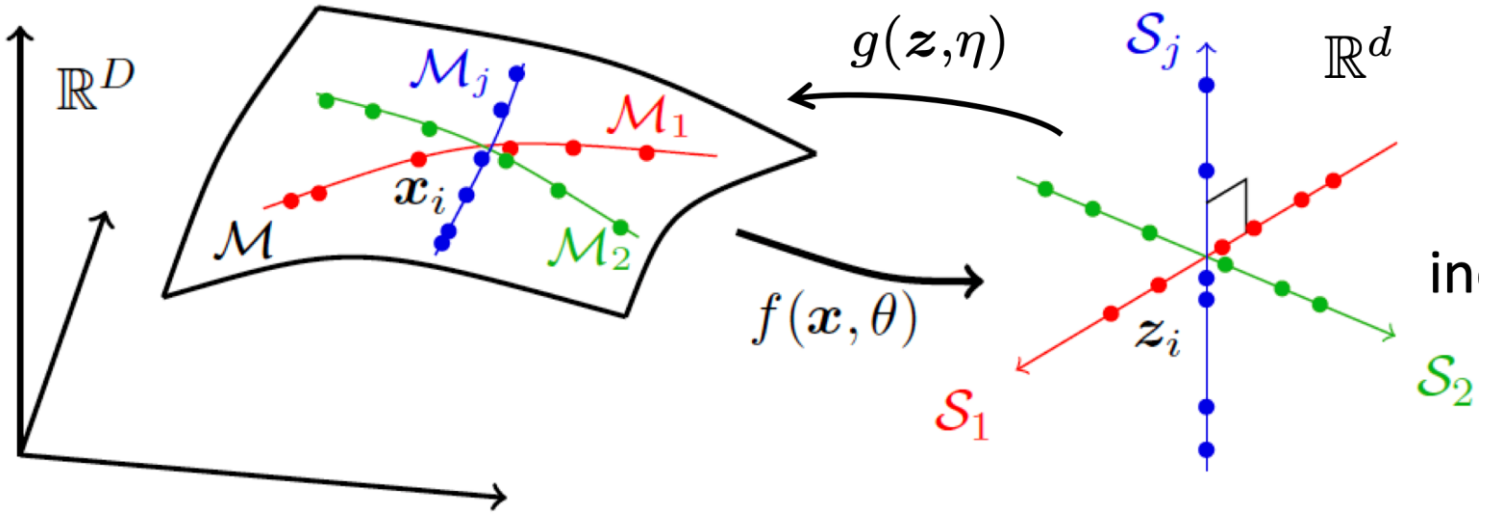


实践中白盒网络涌现出有意义的结构!

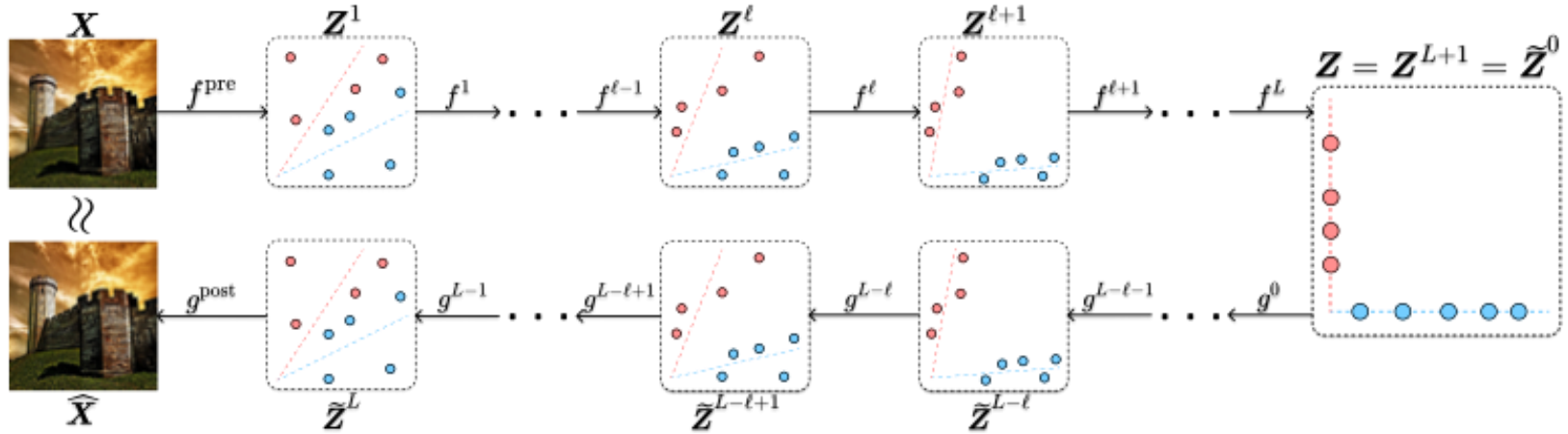


如何学正确（自洽或对齐）

双向编码与解码（对应识别与生成）



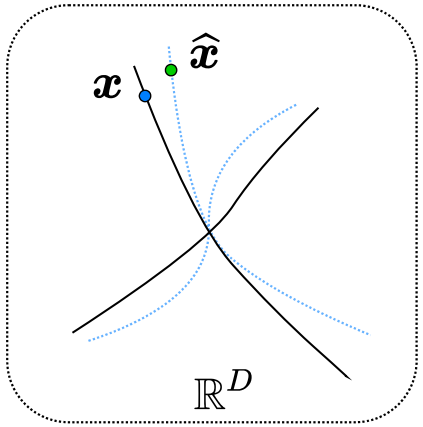
自编码
（autoencoding）

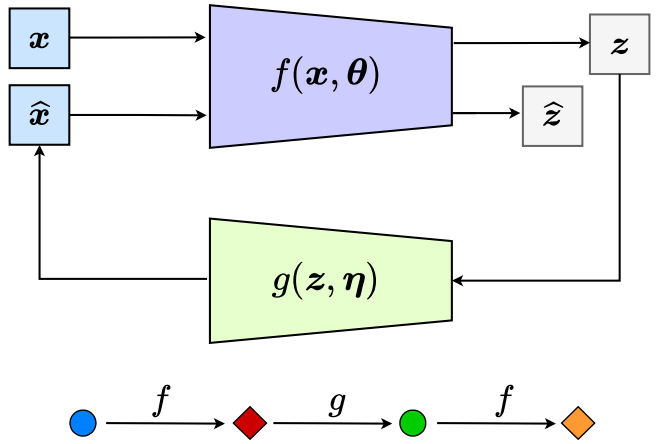


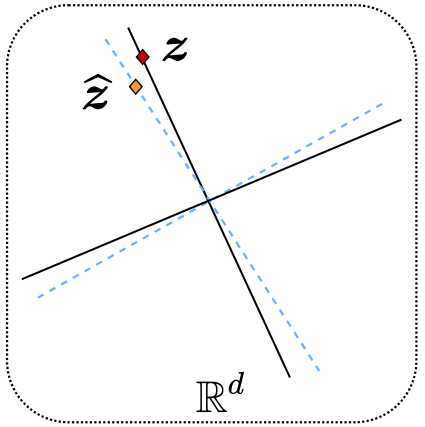
迈向自主、通用的智能系统

闭环反馈纠错、自主学习、自我改进 （控制与博弈）

在自然界中，所有的智能系统都是利用闭环机制学习（维纳）







对智能的认知，必须建立在对实现智能的计算复杂度的正确认知上：

incomputable -> computable -> tractable -> scalable -> **nature**!

Kolmogorov & Solomonoff

Turing & Shannon

NP vs P

DNN and BP

closed-loop feedback?

如何判断智能与否？

智能的定义：一个智能的系统必须具有自主改进与增加自身知识的机制。

任何一个系统，不管多么大，只要不具备自主纠正与增加知识的计算机制，就是没有智能的。

谁有智能，谁有知识？

vs

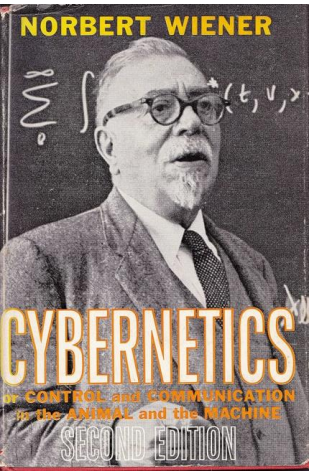
$$\text{Knowledge} = \int_0^t \text{Intelligence},$$

$$\text{Intelligence} = \frac{d}{dt}\text{Knowledge}.$$

这个“人工智能”不是那个“人工智能”

1956年，一群年轻的学者决定打造属于自己名垂千古的学术影响，他们提议开启达特茅斯研讨会，研究高级的，区别于动物智能的人类智能，包括：抽象能力、符号运算、逻辑推理、因果分析等等。

A quote from the Dartmouth proposal: “An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.”



40年代智能
(动物的智能)

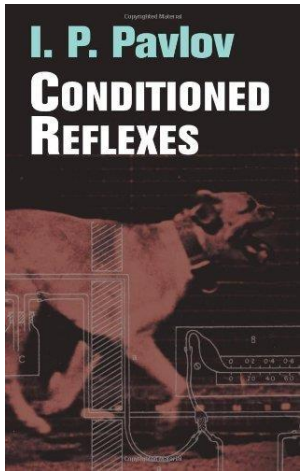
- 信号感知
- 信息表示
- 记忆预测
- 反馈纠错
- 决策优化

56年人工智能
(人特有的智能)

- 语言符号
- 抽象概念
- 逻辑推理
- 因果推理
- 问题解决

目前的“人工智能”
(机器？动物？人？)

- 图像识别
- 图像生成
- 文本生成
- 压缩去噪
- 强化学习

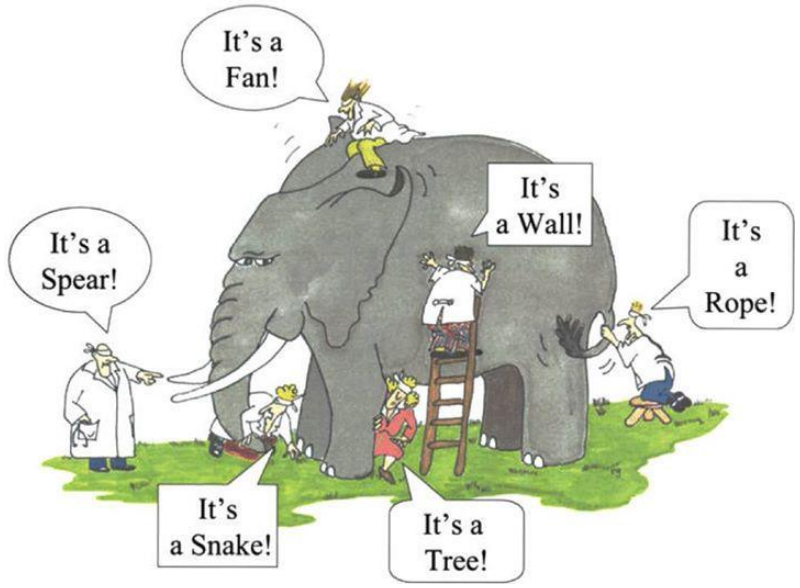


回归理论基石

没有理论的智能，就如同：

一群瞎子在玩一个黑盒子。

(Blind men with a black box.)



“Deep network is all you need.”

“Reward is all you need.”

“Attention is all you need.”

“Foundation model is all you need.”

…

GPT-4

“大模型智能已经或将超越人类！”

“甚至已经具有类人的自主意识！”

“像原子弹、病毒一样可怕！”

“会毁灭人类！” … …

Compression is all there is! （目前的AI系统其实只是做了数据压缩而已！）

结束语

我们理解，我们创造，我们超越

“What I cannot create，I do not understand.”

-- Richard Feynman

“如不能自己创造，则无法真正理解。”

-- 理查·费因曼