

人工智能定义：有关智能agent的研究与设计的学问

智能：衡量Agent在各种环境中达到目标的能力

机器学习：研究学习算法的学问

自注意力机制优点：

- 灵活机制处理变长输入
- 适合长期依赖的建模
- 参数高效灵活，易于并行

Transformer里程碑

2018BERT：预训练+微调

2020GPT3：预训练+提示

2022InstructGPT：预训练+预微调，PPO算法

ChatGPT

强化学习

名字来于动物的发聩

- 策略：最核心，和环境交互必备的
- 奖励：立即效果的度量
- 值函数：长期效果的度量
- 环境模型：环境状态如何变化，非必要

DQN深度Q网络

Q学习+CNN，第一个深度强化学习系统

围棋难点

- 不能全局搜
- 难以给出启发式评价函数

蒙特卡洛树搜索提升围棋程序（迭代：选择+扩展+评价+反向更新）

AlphaGo Zero：只需要知道围棋基本规则，不需要历史数据

- 每一步棋都是蒙特卡洛搜索

Alpha Zero：围棋+国际象棋+日本象棋都能训练

谷歌MuZero：不需要游戏规则编入，一遍学规则一边训练

完美信息游戏：围棋，象棋

- 状态空间复杂度
- 游戏树复杂度

非完美信息游戏：对局者掌握信息不对称：桥牌，麻将

- 合理游戏策略建立在信息集上（每组无法区分的游戏状态）
- 信息集平均大小