

彪哥带你学强化学习

一、认识强化学习

DEEPLY UNDERSTAND REINFORCEMENT LEARNING

讲师：韩路彪



强化学习使用场景举例



DeepSeek-R1

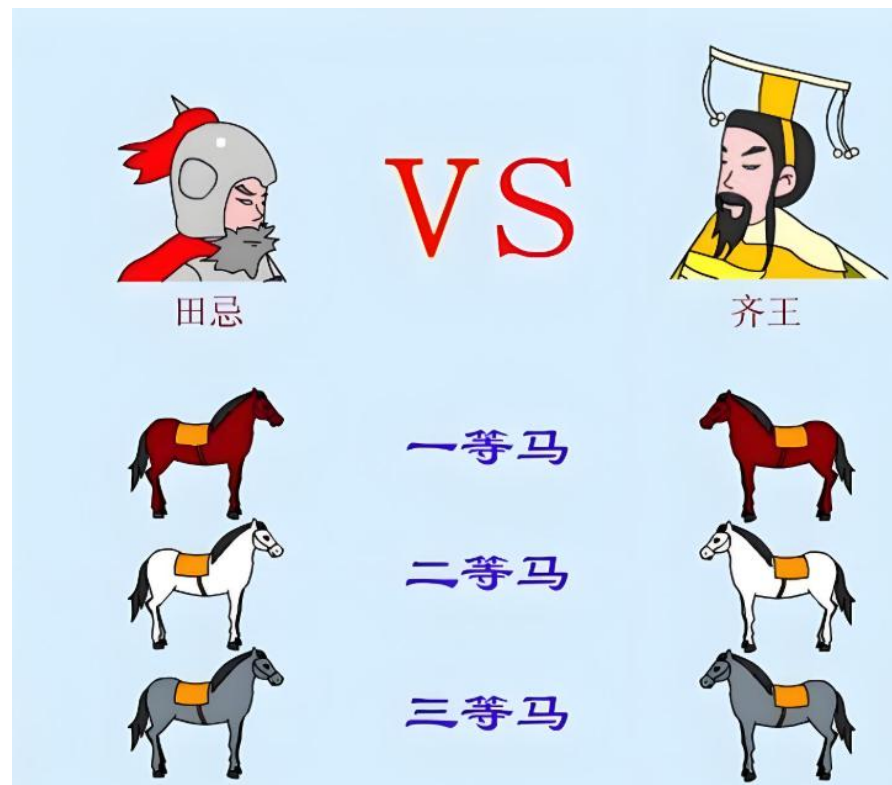


强化学习是什么

研究对象 做一件事情的过程，核心是**序列**

研究目标 最优策略

监督学习	强化学习
单个值	序列
有明确结果	不容易给出每一步明确结果
需要标注好的数据	可以直接和环境交互获取数据
即时反馈	延迟反馈



强化学习的过程

核心思路 探索、记录、利用

探索

尝试新的内容、解锁新的操作，获得对应的奖励(有可能延迟奖励)



记录

把探索的结果记录下来，以备后用。记录的方式可以有很多种



利用

在不同情况（状态）根据记录选择操作

