

守塔 or 回城





RLChina

RL&游戏AI：技术演进和商业价值探索

刘霁 快手AI平台、西雅图AI实验室负责人

源起手工业时代



1952年,
Arthur Samuel设计
游戏AI (手工规则+
少量搜索)



1956年,
“人工智能”命名大
会, 游戏AI成为定义
人工智能的核心元素

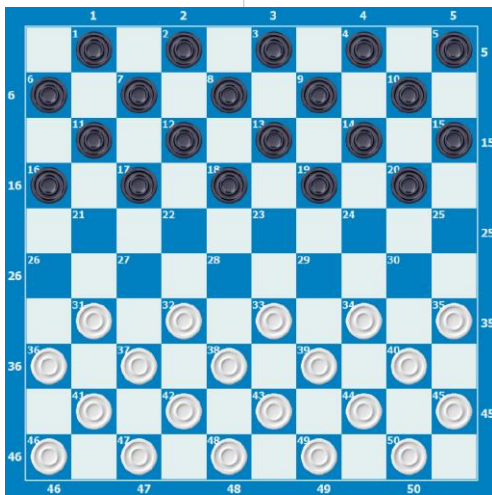


1962年,
Arthur的AI获得康涅
狄格州比赛冠军



由此,
开启了AI对抗人类玩
家的旅程, 是否能击
败人类玩家成为衡量
AI能力的标准

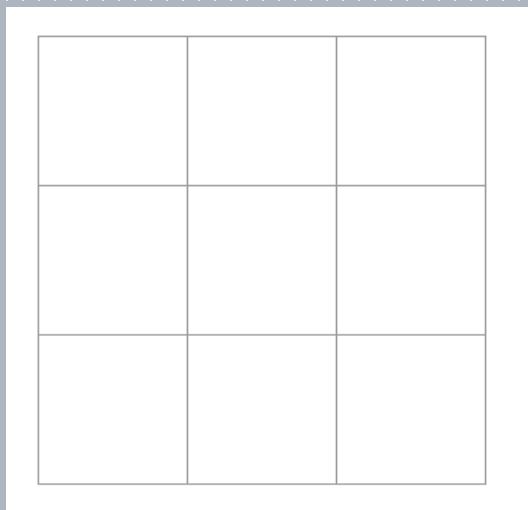
国际跳棋



1 计算机整体算力 (1950s -> 1980s) : $10^4 \sim 10^8$ Flops/s

2 游戏AI算法

- ◆ 手工规则编写的AI整体强于基于搜索的AI, 人工为主智能为辅
- ◆ 形成了后来搜索算法的雏形



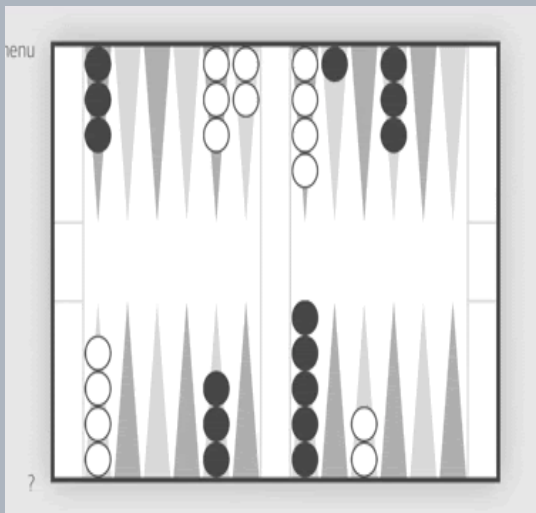
井字棋
(简单搜索就可解)



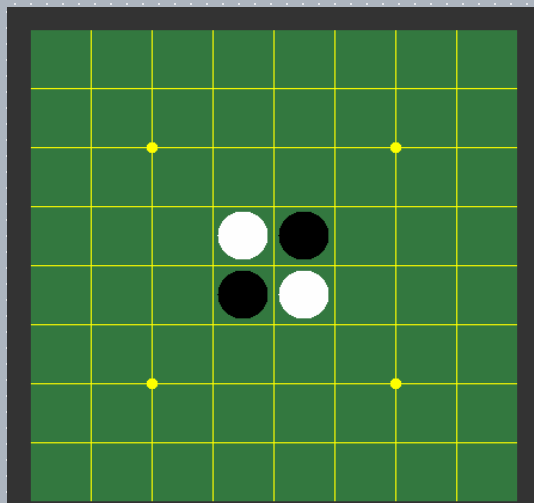
国际象棋
(离人类高手水平还有较大距离, 但是提出
Minimax Search成为后来搜索算法的奠基)

兴起机器工业时代

计算力提升 (1980s-2010s) 标志性事件: Intel 80386 (第一个32位计算机)



双陆棋 (TD-gammon, 1990)
击败人类世界冠军



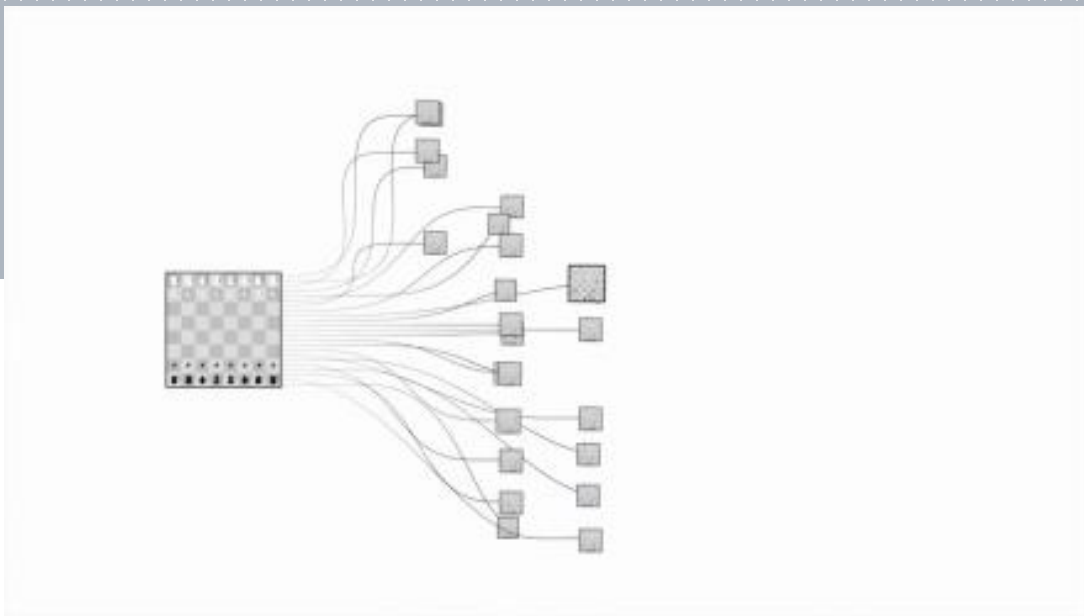
黑白棋 (by李开复等,
1989)
击败人类顶尖高手



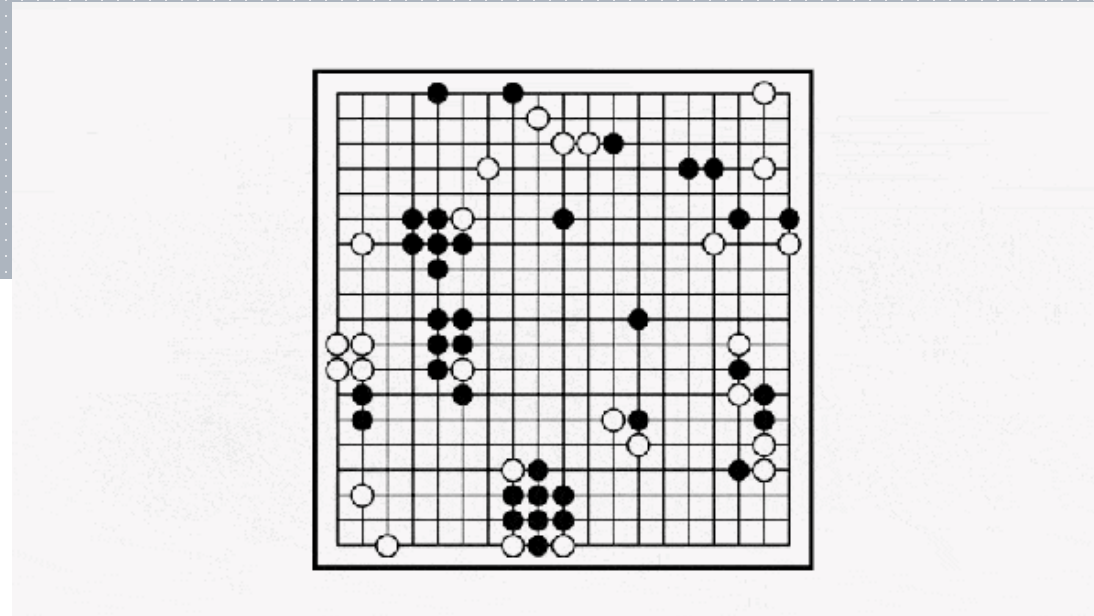
国际象棋(DeepBlue by IBM,
1996)
击败国际象棋世界大师

DeepBlue成为人工智能史上里程碑式之一

- 1 DeepBlue主要技术手段：搜索+检索（一定意义上就是强化学习）
- 2 围棋被认为是最难的棋类游戏
- 3 采用DeepBlue的思路，围棋AI只能达到业余水平，大局观差，算力被认为是根本瓶颈

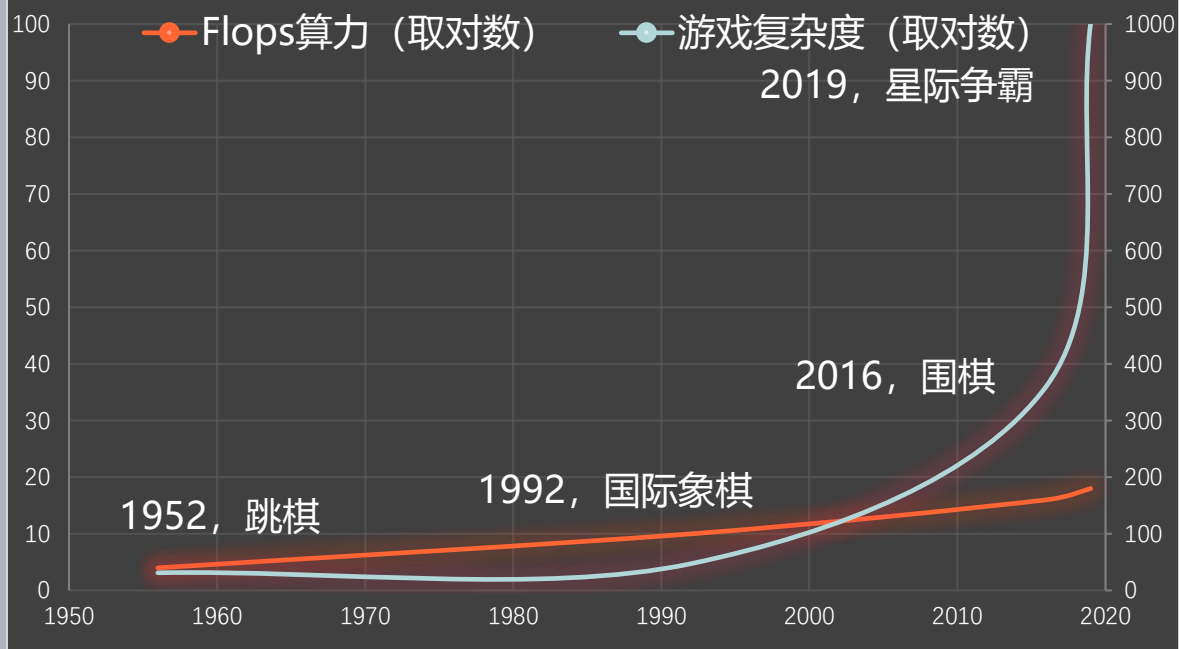


国际象棋



围棋

算力vs游戏复杂度

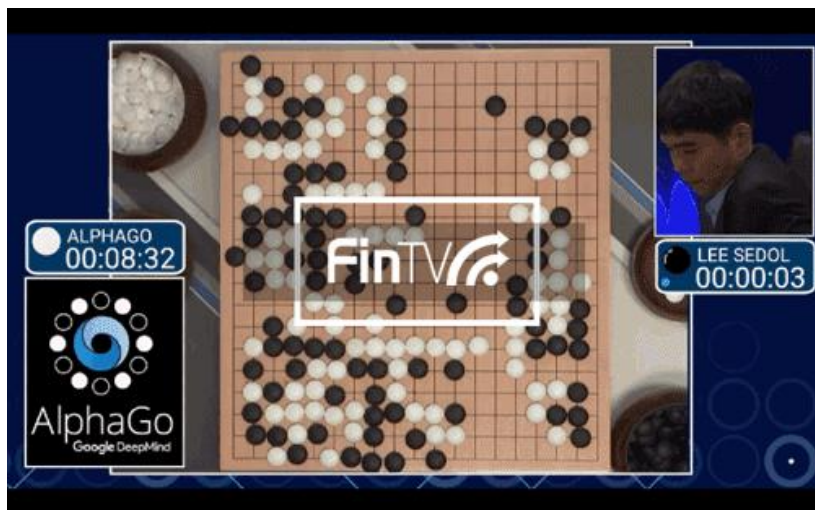


用AI战胜围棋冠军作为AI能否超过人类的评判标准

绝大部分人认为是不行的，因为围棋复杂度远高于算力增长的速度

封神 狭义AI工业时代

国际象棋击败顶尖人类之后，围棋成为
AI最后需要逾越的高峰



- 1 算力的突破：GPU在深度学习的广泛使用
- 2 方法的突破：搜索+深度学习=深度强化学习
(通过神经网络记忆状态得分来减少搜索空间)
- 3 一战封神：AlphaGo击败人类顶尖选手李世石

AlphaGo一度被认为是AI全面超越人类的标志，之前认为大局观是弱点，现在已经成为职业棋手学习的榜样

采用跟AlphaGo类似的算法，AI学会了：

微观操作--放风筝

宏观策略--劣势守塔 vs. 回城



反思广义AI工业时代



【问】AlphaGo真的彻底解决了围棋么？

【答】没有，因为AlphaGo没有找到围棋的最优策略



【问】围棋真的是最难的游戏AI问题么？

【答】不是，围棋是对称信息博弈；更难的问题有非对称信息博弈，以及多AI的协同（群体智能）的类型

各类游戏的对比

非对称
无法看到对手信息

弱合作
可以得到合作者的
所有信息

强合作
无法得到合作者的
所有信息

	非对称信息	多智能体弱合作	多智能体强合作	是否达到人类顶尖玩家水平
国际象棋	×	×	×	√
围棋	×	×	×	√
德州扑克	√	×	×	√
日本麻将	√	×	×	√
星际争霸(1v1)	√	×	×	?
Dota (5v5)	√	√	×	?
王者荣耀 (5v5)	√	√	×	?
斗地主	√	√	√	×

研究者开始转向更复杂游戏类型



德州扑克

(非对称博弈, 击败多名世界顶尖选手)

Deepstack (1v1), Albert Pluribus (6人局), CMU

优势: 更丰富的加注策略、反主动下注、下超大注

劣势: 不会剥削



Dota

(多AI协作、非对称博弈, 战胜职业战队)

OpenAI Five




星际争霸

(多AI协作、非对称博弈, 战胜职业选手)

AlphaStar, DeepMind

- 1 技术：除了德州扑克AI（用CFR），其他AI基本都采用AlphaGo类似的技术（DRL）
- 2 成果：战胜过人类顶尖玩家
- 3 问题：战术容易被人类针对，很多号称战胜了职业玩家的AI都存在争议

<div> OpenAI Five 1377 in-progress games 1485-7 (99.5% winrate)</div>						
Leaderboard						
<div> WIN  WIN STREAK 2+</div>						
RANK	ORGANIZER	GAME PLAYERS	WINNER	KILLS	DURATION	
1	 802497223b8b4ab9	 虎比木	Human Team (Dire)	11-0	0:07:22	
2	 godisawoman	 5, Silent, the only one, iLTW, unstable	Human Team (Dire)	13-27	0:33:39	
3	 0a99136214f24c09	 [], Toei, Mikasa 无力, Kriknak, Getsrch, Tanny	Human Team (Radiant)	23-10	0:37:17	
4	 919a6410d0b1409b	 ?, 沐云, 東南亞路人王, Scribbles, m'	Human Team (Dire)	22-26	0:41:16	
5	 VioletEvergarden	 Killua, Violet Evergarden, Jelena, (T____T), Enryu	Human Team (Radiant)	39-25	0:44:55	
6	 qwerty	 ๔๕๖๗๘๙๐๑๒๓๔, Tiger, my, Cp boombell <3, YoungBoyNeverBroke	Human Team (Radiant)	34-24	0:45:26	
7	 2BWeebDieTwice	 002B Weeb Die Twice 00, @WagaGaming, SQRL 1, Tester! ^^ > < ㄣ(ツ)ㄣ, Pjok	Human Team (Dire)	18-19	0:57:58	
8	 ToveLo-	 棋宝, 下套子, 煜呢, 棋德, 龙东强, 吃!!!, Mu	OpenAI Five (Radiant)	43-44	0:51:46	

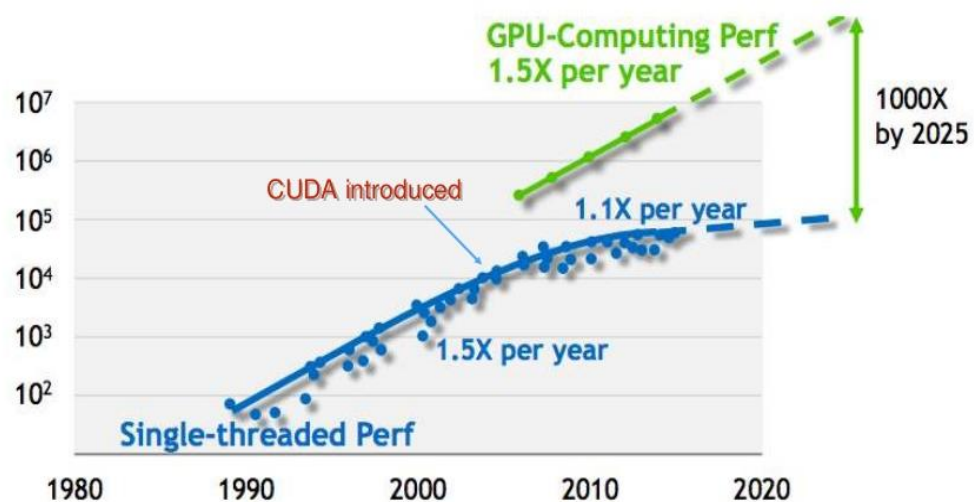
游戏AI发展规律

算力是核心驱动
算法常常领先于算力

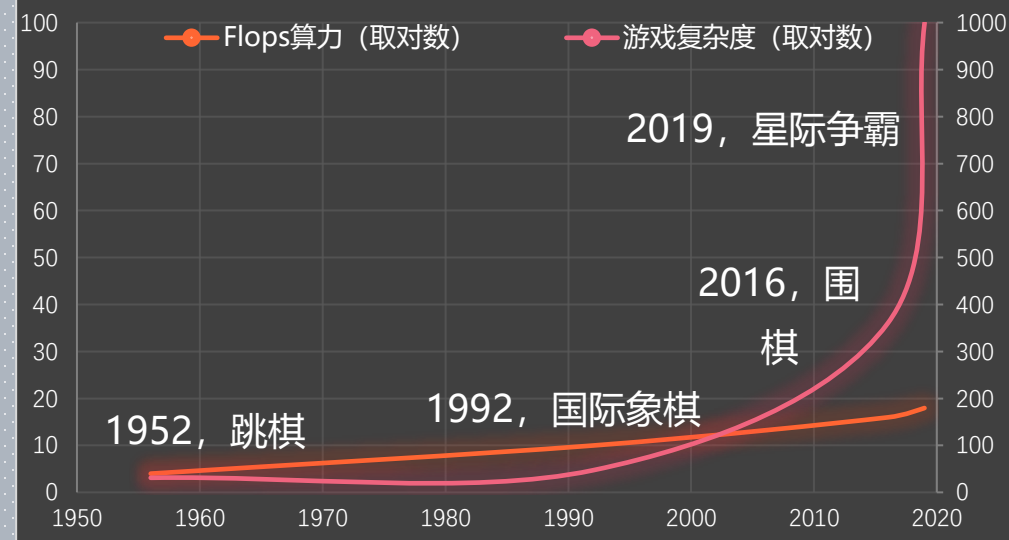
挑战

如何弥补问题复杂度和算力的gap
如何系统化的解决多AI非对称博弈

Moore's Law and beyond



算力vs游戏复杂度



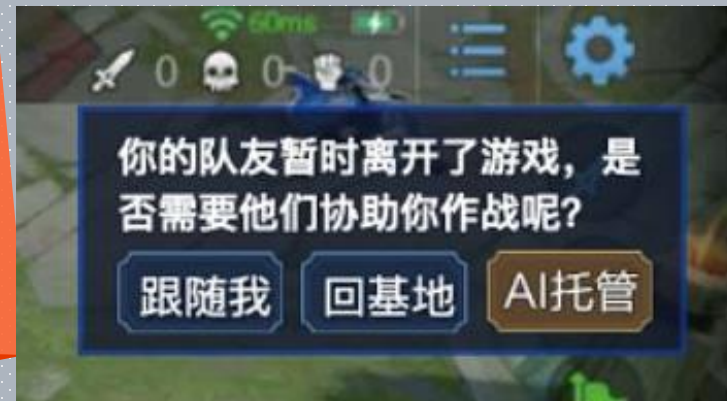
游戏AI / 商业价值

用AI替换人的角色：人的替代品



AI vs 人类玩家

AI 托管



关卡自动测试

AI 辅助



让AI具有独立人格：AI是独立的个体

游戏关卡自动生成（AI是老师出题，用户是学生解题）



人性化NPC



人性化NPC（靠设置人性驱动） vs. 传统的NPC（靠规则驱动）
优点：柔性化的NPC，结局和过程open，更真实

- By RTC studio



开放世界的沉浸式体验，有人性的NPC



虚拟养成类游戏，宅男的福音



- 1 有的NPC具备欺软怕硬的属性
- 2 NPC甚至还会组队打劫玩家，甚至是把玩家的家园给破坏

黑暗与光明手游（蜗牛数字开发），90后00后的魔兽世界

让AI具有神性：在上帝视角优化个体用户的长期体验



智能老虎机



个性化付费道具推荐



个性化定价策略



匹配策略优化

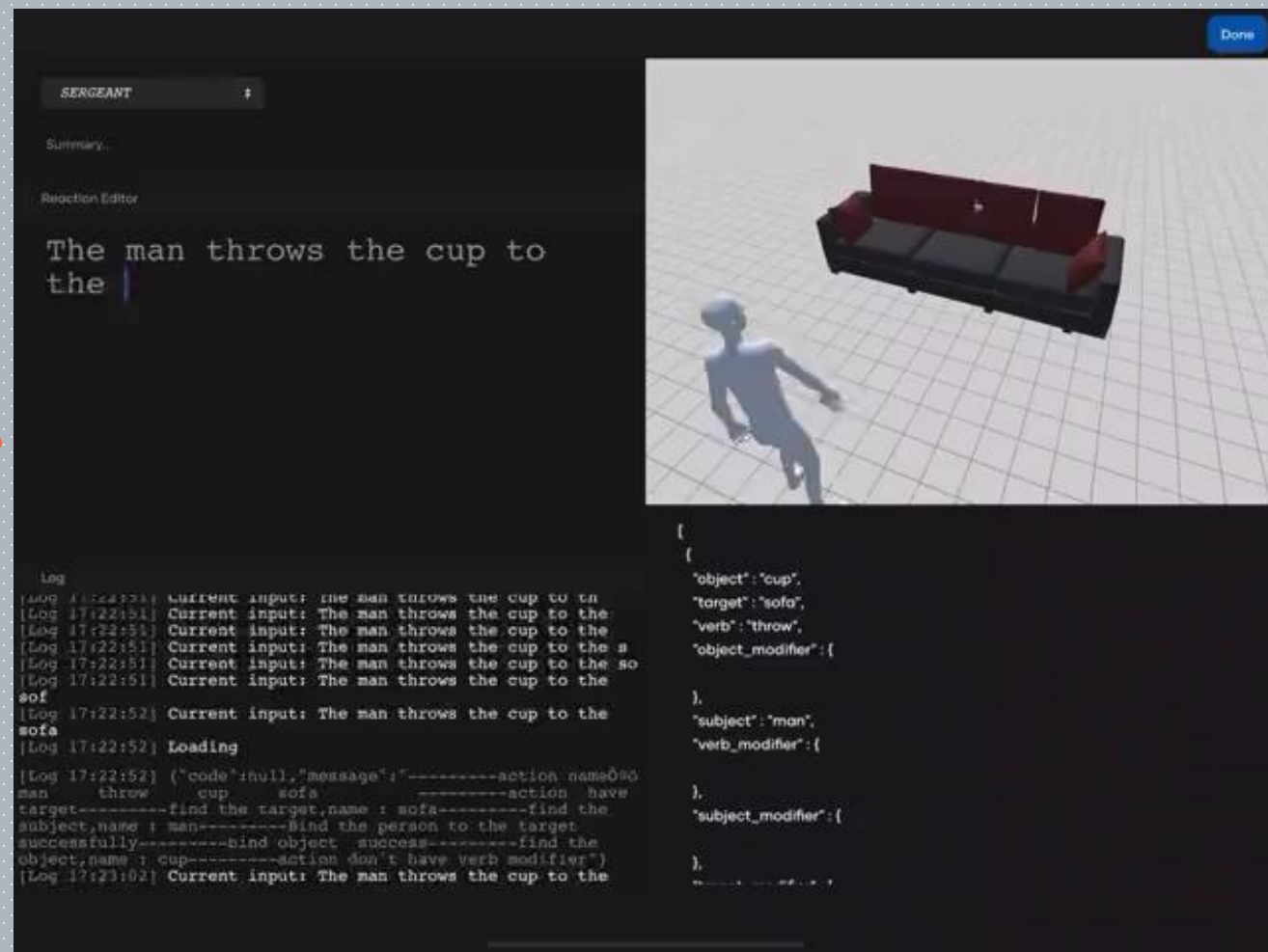
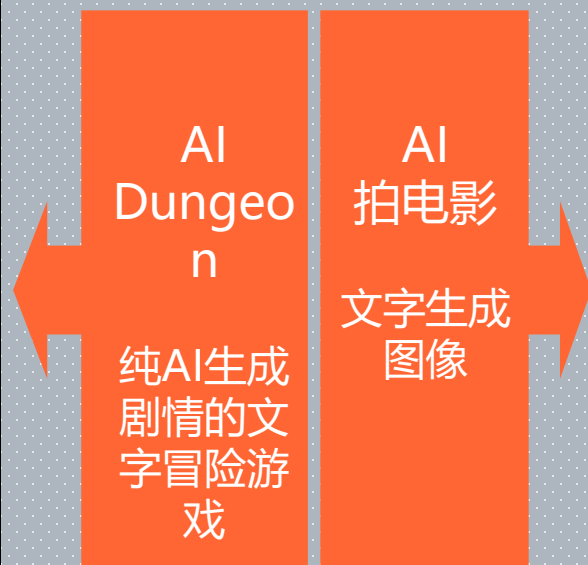


发牌策略优化

跟传统推荐的区别是：从用户个体整个游戏内生命周期为优化目标，而非所有用户的短期体验为优化目标



EA近日获得了一项全新专利：
该专利能够让开发团队在开发游戏的过程中使用AI评估游戏难度以及及时对游戏做出改动与调整



《Rival Peak》：云游戏 + AI + 直播

混合现实秀+ 互动冒险游戏

观看时长超过1亿分钟

超过70多个国家的观众

直播参与率(反馈、评论、分享和点赞)超过2亿次

由12名AI “参赛者” 组成，他们需要在变幻莫测的自然环境中生存下来并破解一系列复杂的谜题。据统计，有96%的观众是手机用户，他们占据着比赛的主导地位并决定了下一步的发展走向。12名AI角色通过数百万张观众的选票，破解了数百万个谜题，观众还可以选择协助（或不协助）角色优先完成任务，从而避免他们喜欢的角色被淘汰

首家以AI为卖点的游戏上市公司



主打概念：AI+游戏

AI做精细化个性化运营，极致化用户体验

2021年初上市，市值100亿美金

全球独创的业务模式：

收购便宜的游戏资产，通过平台化的AI大数据能力，
输出对游戏的精细运营改造方案，实现游戏变现能力

强化学习和游戏AI 在快手的实践

在快手游戏中的应用

游戏研发 (游戏关卡难度)
游戏运营 (游戏广告的智能投放)
游戏上线 (游戏AI)
游戏用户体验优化 (策略优化)



斗地主残局自动生成



游戏关卡难度自动测试, 异步并行的MCTS
[ICLR2020, oral presentation]
节省10倍测试人员

DouZero: 斗地主AI



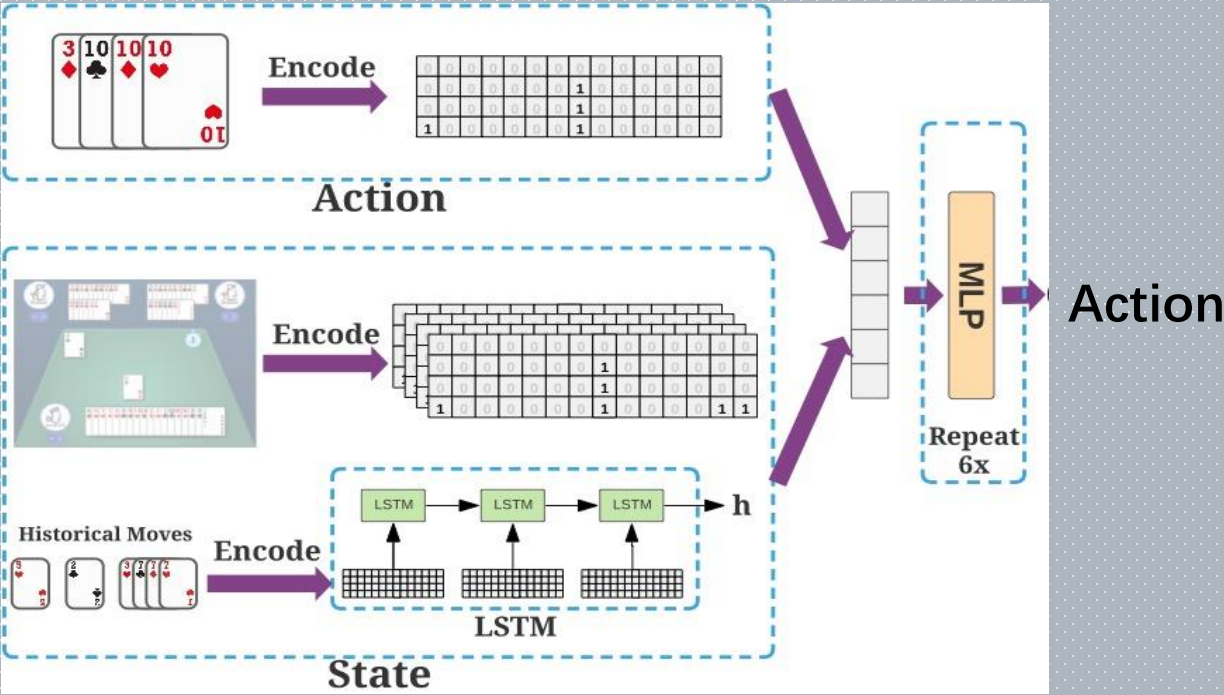
支持多种玩法，以及不同水平

挑战

Action space大（总共27,472种牌型）且随状态变化

游戏	动作空间大小
雅达利 (Atari)	10^1
21点	10^0
有限注德州扑克	10^0
围棋	10^2
麻将	10^2
无限注德州扑克	10^4
斗地主	10^4

DouZero: 把action作为函数输入解决问题



阶段性结果 (BotZone: 1/344, 一台GPU机器训练)

无叫牌规则

游戏列表 / FightTheLandlord / FightTheLandlord 的 Bot 排行榜

排名	Bot 名	作者	排名分	Bot 描述	最新版本号
1	biubiubiu	 karoka	1580.67	[DouZero-ICML2021] 开源代码: https://github.com/kwai/DouZero Arxiv: https://arxiv.org/abs/2106.06135	8
2	nobody_knows_why	 AlchemistWang	1576.97	To be bald, to be strong.	16
3	下个Bot见	 AreWeCoolYet	1573.55	下个Version见	61

有叫牌规则
(欢乐斗地主规则)

Game List / FightTheLandlord2 / Bot Rank List for FightTheLandlord2

Rank	Bot Name	Bot Owner	Rating Score	Bot Description	Latest Version No.
1	v2太短了	 karoka	1303.75	https://douzero.org/bid/ 人机对战	7
2	斗旺仔	 jumpmelon	1254.00	斗旺仔	70
3	馒头卡	 shiro今天也是早八人	1226.99	圆焰一生推	10

论文二维码
ICML2021



开源code二维码



试玩二维码



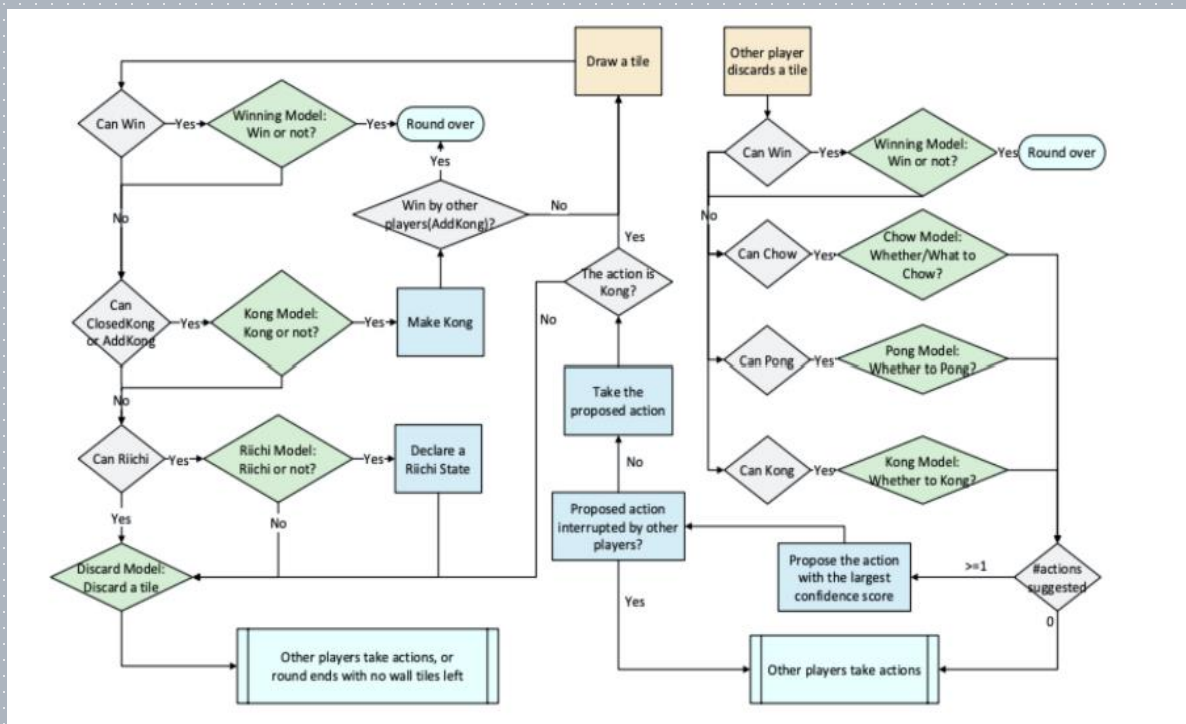
Kima: 麻将AI

春风茶馆



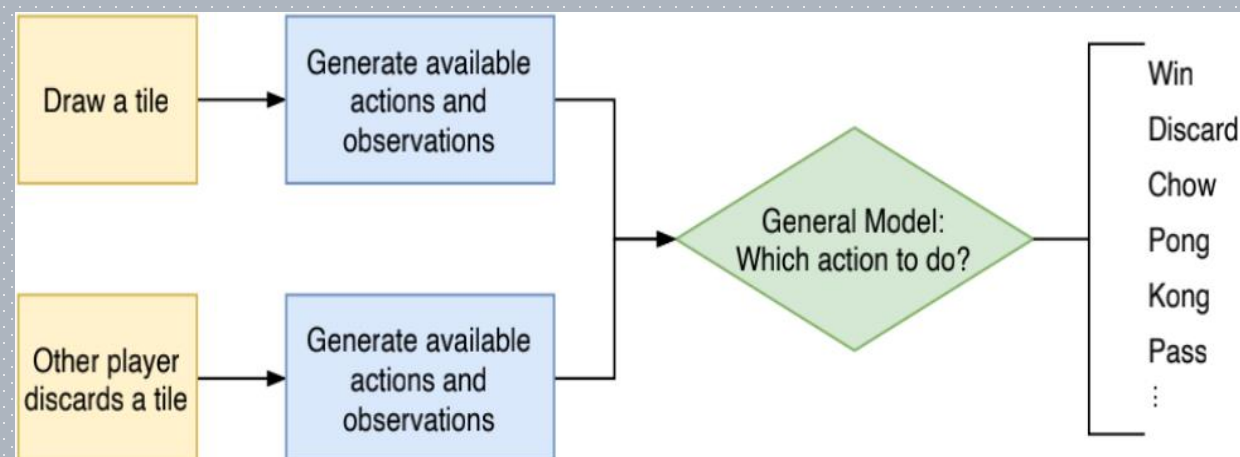
- 1 超过6+游戏规则
- 2 每个规则4个难度等级的AI
- 3 游戏冷启动
- 4 陪玩

Suphx: 微软麻将AI（日本麻将规则）



接近人类思考模式
根据规则定制化决策流程
计算成本比较高

Kima: 快手通用麻将AI



通用决策流程，不同规则共用同一套决策流程
训练和推理成本低

阶段性结果(2-3台GPU机器训练)

- 支持了6种不同国内规则的玩法，均达到了人类顶尖高手水平
- 参加了IJCAI 2021全球**国标规则**麻将AI大赛，第一阶段第一，第二阶段第三

李文龙（国际麻将联盟(MIL)秘书长）表示：

- 关键手牌的细节处理上也从AI身上学到了很多新的思路
- AI最大的不足在于反应灵敏但预测不足，算法惊人但不够灵活

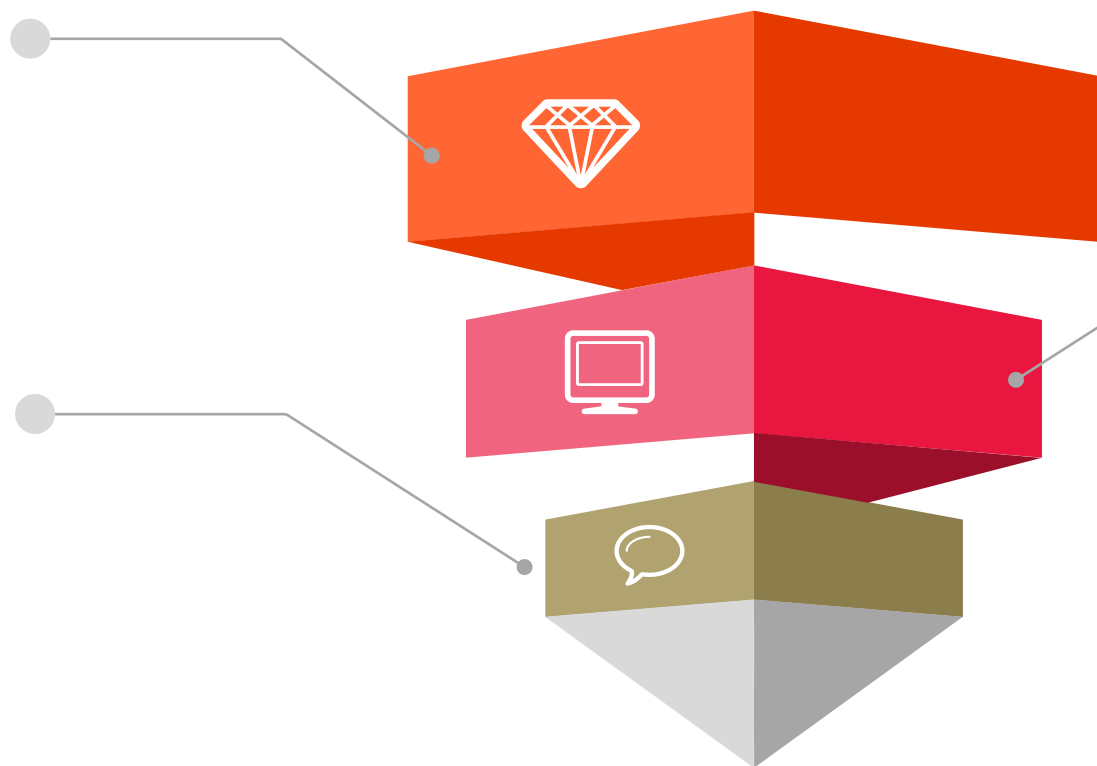


参与YY直播主办的“和AI搓一局”，不敌顶尖职业选手

广告推荐

推荐和广告是短视频这类ToC的行业最直接的收益来源

传统方法都是基于**优化即时收益**的监督学习



长期收益的优化在技术在业内上几乎是空白

- 长期收益优化
- DAU优化
- 用户长期体验的优化等等

强化学习优化广告冷启动

1 广告生命周期与基本流程

- 投放期（所有候选广告公平竞争）
- 探索期（可以实施补贴干预广告，让有价值的广告尽快进入投放期）

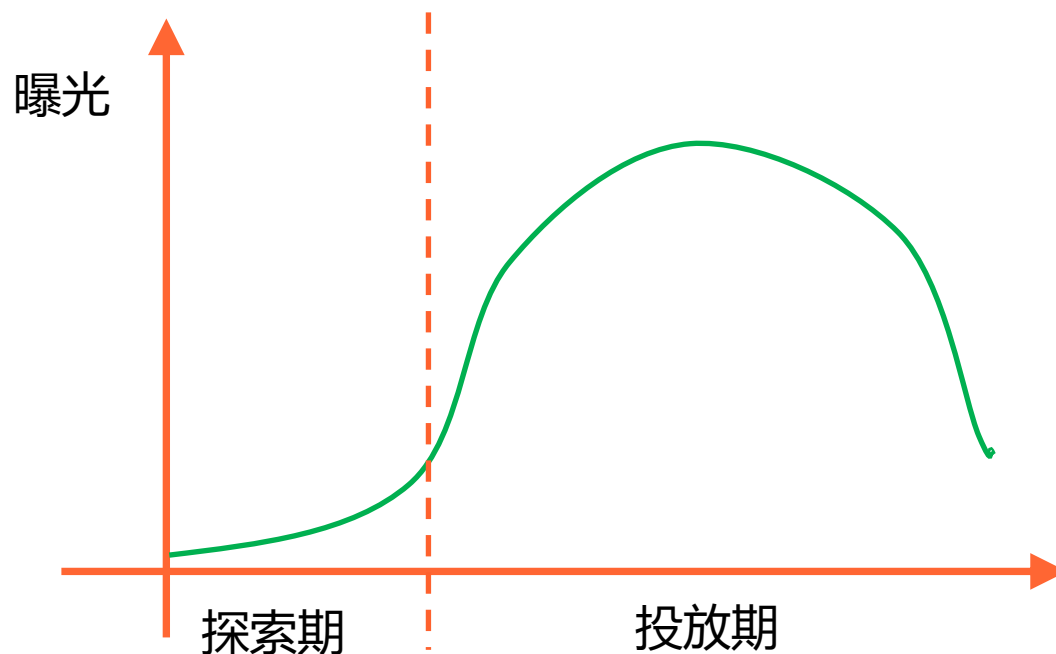
2 竞价排序公式

$$\text{rankScore} = \text{ecpm} + \text{bonus} + \text{ueq}$$

预估经济收益 补贴 用户体验收益

3 广告冷启动优化目标

通过调节bonus，优化长期收益



问题
挑战

线上探索成本非常高 ->
线上收入波动大

状态
定义

State: 广告基本信息+历史投放信息+时间特征+模型历史预测结果
Action: 未来1个时间片段内给该广告的补贴总额
Reward: 未来1个时间片段内该 (广告总收益-补贴) 以及转化
Terminal: 转化数 ≥ 10 (通过探索期)

解决
方案

Batch-Constrained Q-Learning (BCQ): 近似 offline 版本DDPG

实验
结果

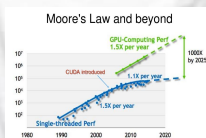
实验周期: 一周
动作频率: 每小时数据收集+决策
收益: 提升收入XX万/天 (XX: 一个客观但是不方便透露的数字)

学术界

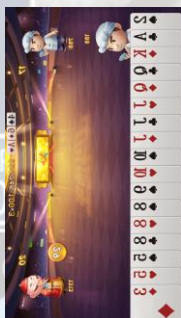
AI战胜人类是否是研究的终点



算力跟不上问题的复杂度



系统化的解决结构更复杂的Game AI问题，比如：非对称，多agent合作，非零和游戏等



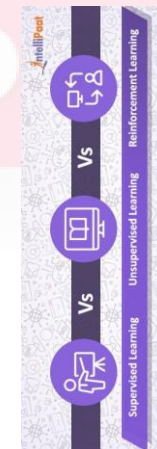
工业界

正确辨识哪些问题是适合RL求解

如何把学术界的RL成果在工业界落地，弥补二者之间的gap，比如：

- 线上exploration cost太高
- 稳定性如何保证
- 等等

如何用RL创造新的产品和玩法

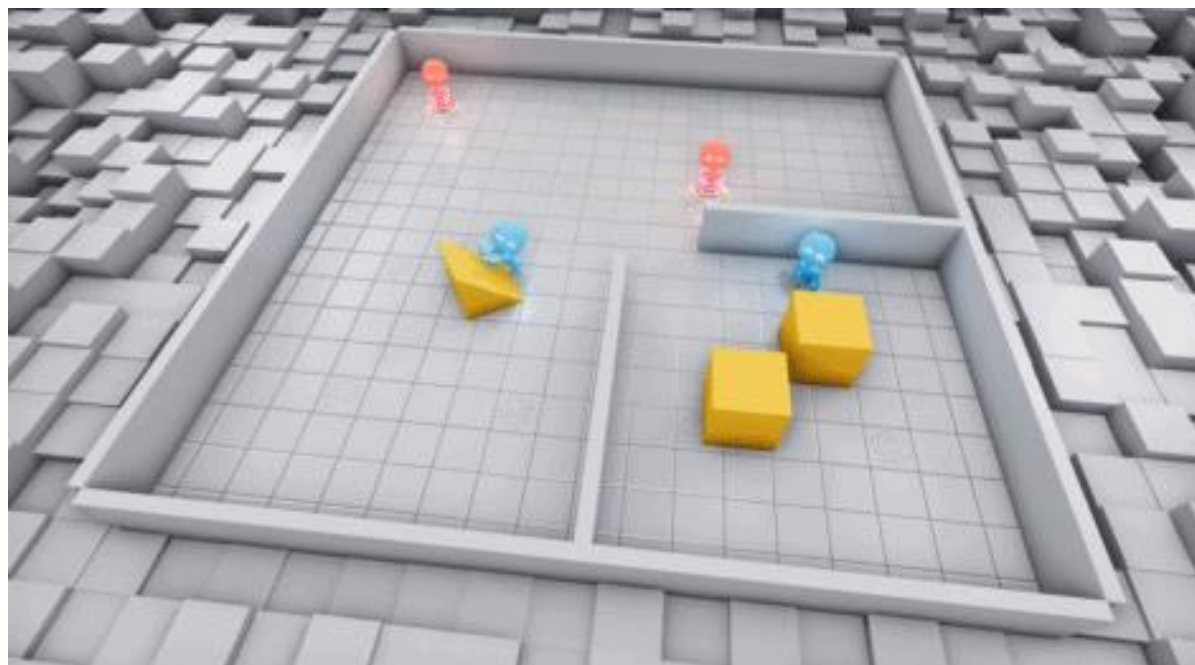


谢谢

Atari打砖块



- 1 “打砖块” 看起来是比围棋甚至国际象棋更简单的游戏
- 2 采用DeepBlue的方法也无法击败人类顶尖高手
- 3 直到2013年由DeepMind取得突破
使用的DRL(深度强化学习)方法后来也被用于AlphaGo

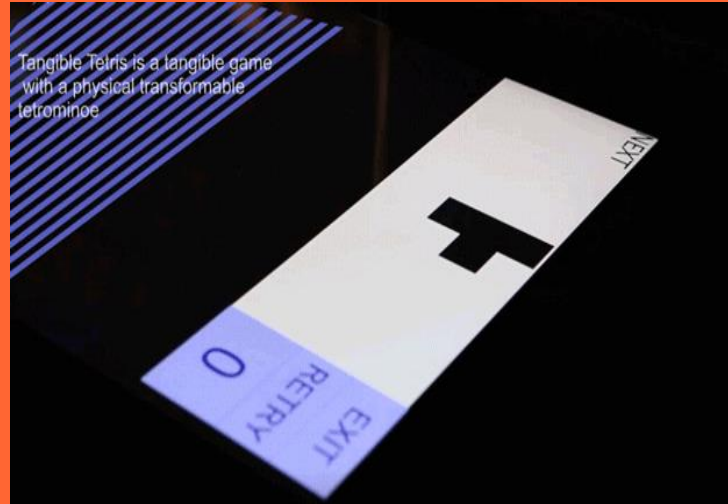
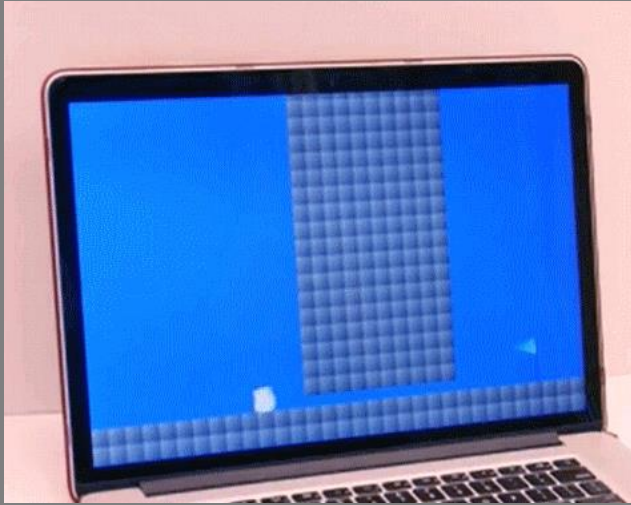


躲猫猫



多人合作夺旗（超越人类水平）

虚拟世界和现实世界的融合





“人” 性化教—以个体为视角优化整个学习体验和效率



量化交易—跟人玩赚钱的博弈游戏

智慧交通



智能运维

