



何友，中国工程院院士，全国优秀教师。曾兼任 CAAI/CIE / CAA/CIC/CSF/IET Fellow，国务院学科评议组成员，国家自然科学基金委信息学部咨询专家委员会委员，中国人工智能学会名誉副理事长，中国航空学会名誉副理事长，中国指挥与控制学会副理事长、中国电子学会和航空学会常务理事，中国航空学会信息融合分会主任委员等。主要研究领域为信号检测、信息融合、智能技术与应用。以第一完成人获国家科技进步二等奖4项、国家教学成果一、二等奖各1项，获省部级一等奖11项，授权中国发明专利和软件著作权60余项。获全国百篇优秀博士学位论文，在 IEEE 会刊等发表重要论文260余篇，出版专著6部，论著他引20000余次。

游戏内外——AI 强化学习的围城

何友

2016年，谷歌 AlphaGo 4:1 击败围棋世界冠军李世石使得深度强化学习的概念走进大众视野，由此人们逐渐开始憧憬在围棋中战胜人类的 AI 强化学习算法能够给人类生活带来重要影响。但时过五年多，AI 强化学习似乎只在游戏内大放异彩，而在游戏外却应用寥寥。

人工智能自诞生就与游戏结下不解之缘。1950年，图灵提出著名的图灵测试，该测试就可以看作为一个欺骗提问者的游戏。随后，在计算智能的不断突破下，五子棋、国际象棋等棋类游戏都逐渐被 AI 征服。当 AI 开始挑战更高层次的围棋游戏时，发现传统搜索方法在时空复杂度上完全不可行，由此深度强化学习成为研究的主流。在挑战成功围棋、扑克这类回合制游戏后，谷歌、OpenAI、腾讯又逐渐把焦点转移到星际争霸、王者荣耀等即时策略游戏上。与围棋相比，即时策略游戏需要 AI 学习在不完全信息和即时动态环境下如何进行推理、决策、规划、协作以及平衡短中长期收益，相关研究推动了多智能体强化学习理论和算法的发展。

电子游戏具有诸多特点使其在 AI 强化学习研究过程备受青睐。首先，玩电子游戏必然要在与环境及他人的交互中进行决策和博弈，而决策和博弈也是人类日常生活的重要行为；其次，电子游戏具有真实模拟和无损探索的特点，可以比较容易地通过不断试错的方式获得各类场景及情况的样本和标注信息，从而对 AI 算法进行大规模训练和测试。因此，电子游戏是 AI 强化学习天然的训练场，在游戏中训练 AI 是公认模拟现实世界的最有效方法之一。AI 算法不断在星际争霸、王者荣耀、足球、捉迷藏等游戏中达到或超过人类玩家水平，并时常意外地开发出一些人类玩家都未曾想到的战术和策略。电子游戏正以非常接近现实世界的方式加快着 AI 算法研究，让人们看到 AI 走出游戏、落地现实的曙光。

然而在游戏中大放异彩的深度强化学习算法在游戏外仍应用寥寥，一方面是目前深度强化学习算法样本利用率低和缺乏可信度，更重要的原因是游戏世界与现实世界仍存在巨大鸿沟。首先，游戏环境是封闭的而现实环境是开放的，开放环境中的多智能体对抗博弈面临着环境更加复杂、决策空间更加巨大等问题，这导致在游戏内开发的 AI 模型在现实世界应用受限；其次，游戏环境对问题的假设通常较为理想，如多智能体间的通信通常假设是完美的，但现实世界中多智能体的通信却经常受限；最后，游戏环境对现实世界的模拟还远远不够，在场景真实度和信息获取维度上都有所欠缺。

不断推动 AI 算法从游戏迈向现实对于机器人智能协作和博弈领域的发展意义重大。美国计算机社区联盟发布的 2020 版机器人路线图重点强调了机器人在复杂、动态环境下主动感知、规划及控制。我国在《新一代人工智能发展规划》中也着重强调了无人自主系统的发展。为推动开放环境下机器人智能的发展，需要在学习理论上实现突破的同时在环境模拟上更加真实。相信随着多智能体强化学习和迁移学习等理论发展，以及平行智能、数字孪生及元宇宙等技术兴起，在游戏内大放异彩的 AI 算法也将走出围城，在游戏外的现实世界产生重大影响。