

人工智能创新企业 OpenAI

2015年12月12日，人工智能创新企业 OpenAI 宣布成立。一年多来，OpenAI 发布了强化学习开发工具包、人工智能训练平台、人工智能定律等一系列成果，推动和加速了人工智能领域创新进程，其所倡导的开放开源的人工智能创新模式，也给人类打造先进技术带来了新的样本。分析研究 OpenAI，借鉴其创新方法和创新成果，对我国加快推进人工智能领域的发展步伐具有重要价值。

一、背景

人工智能正成为全球热点，实现快速发展。人工智能（AI）是用于模拟、延伸和扩展人类智能的科学，通过对人类意识、思维的信息过程进行高效模拟，使机器能够胜任通常需要人类智能才能完成的复杂工作。人工智能被视为具有巨大社会和经济效益的革新性技术。近年来，随着互联网、大数据的兴起，以及深度学习等机器学习算法的进步，人工智能领域持续涌现突破性成果，并逐步应用到社会、文化、科技、经济等领域，整体上正逐渐趋近由量变到质变的奇点。2016年初 AlphaGO 战胜人类围棋高手，2017年初 AlphaGO 的升级版 Master 对战人类众多围棋顶尖棋手豪取 60 连胜，都展示了当前全球人工智能的发展成就。

人工智能威胁论持续存在，引发高度关注。对相关威胁的担忧表现在两个方面。一是对人工智能本身的忧虑，担心一旦人工智能拥有了自我意识，可能反抗人类或做出伤害人类的行为，甚至最终威胁人类生存。《终结者》电影中的“天网”和机器人杀手经常被提出作为这种情况的例子。著名物理学家霍金和特斯拉总裁马斯克曾联合 12000 名企业家和社会名流签署公开信，呼吁联合国发布禁止自动化武器法案，并提醒科学家和研发人员关注强人工智能的潜在威胁。二是对人工智能技术不当应用的忧虑，

即人工智能技术被“恶意人士”掌握并滥用，可能会给其他人带来损害或威胁。虽然更多人认为，人工智能技术在本质上是受控的，不必对其产生恐惧情绪，但也同意为潜在风险做出一定准备。

二、案例介绍

创建起源。 OpenAI 是在特斯拉汽车 CEO 埃隆·马斯克和多名来自高技术企业、投资领域重要高管的支持下，于 2015 年 12 月 12 日宣布成立的一家非盈利性的人工智能创新企业，同时也是一个非盈利性人工智能项目。其肩负着双重使命，一方面确保人工智能技术不会伤害人类，另一方面尽可能确保人工智能技术获得的利益分配更加广泛和均衡。

埃隆·马斯克一直对人工智能持批判态度，并不止一次警示人工智能技术的危险性。因此，由其发起成立的 OpenAI 的首要目标，就是找到缓和恶意人工智能威胁的方法。在迎战恶意 AI 方面，OpenAI 采用了“扩展”而非“限制”的方式。OpenAI 将创建新型 AI 实验室，以安全方式让人类建造真正的 AI，从而尽可能让更多的人使用人工智能技术，使这项技术不会成为少数团体的“囊中法宝”，以此消除不同个体、不同团体间可能出现的“人工智能鸿沟”。

正因如此，OpenAI 在促进人工智能技术创新方面开展了一

系列工作。OpenAI 以探索强化学习为起点，通过不断重复任务来记录找到让机器产生最佳结果的学习方式。在此基础上，研究“无监督学习”技术，以创建能离开人类指导进行自我学习的机器，以及无须精心标注的数据就能进行学习的神经网络。

创新进展。成立一年多来，OpenAI 在人工智能领域发布了一系列成果，均引起各方的高度关注。

2016 年 4 月，OpenAI 发布用于研发和比较强化学习算法的工具包 OpenAI Gym。其所包含的“强化学习”技术，就是驱动 AlphaGO 战胜围棋选手李世石的核心技术之一。开发人员可以使用 Gym 开发用于棋类、游戏等其他用途的人工智能系统。

6 月，OpenAI 发布了其与谷歌、伯克利和斯坦福的学者共同形成的研究成果，即避免负面影响、避免对奖励条件的非法解读、可扩展的监督、环境探索的安全性、分配转变的鲁棒性等五条人工智能定律，旨在为 AI 提供有效的行为约束，从设计层面着手，使 AI 不会在有意或无意中做出危害人类的事情。

11 月，OpenAI 与微软就在公用云服务 Azure 上运行大规模计算方面签署协议，OpenAI 将利用微软 Azure 运行深度学习和人工智能的测试，并和微软合作推进研究和创新技术。

12 月，OpenAI 发布 OpenAI Universe，一个能在所有游戏、

网站和其他应用环境中衡量和训练 AI 通用智能水平的开源平台，其目标是让人工智能像人一样使用计算机、浏览网页、玩游戏。目前，Universe 中已包括约 2600 种雅达利（Atari）游戏、1000 种 flash 游戏和 80 种浏览器环境。

作用和意义。 OpenAI 在本质上是一个人工智能研究与创新实验室，其所实施的是以数十亿美元为基础，推动人工智能发展的大工程。作为致力于推动人工智能创新并免费分享创新成果的机构，OpenAI 不仅能够加速人工智能的发展进程，而且还会作为榜样改变人类打造先进技术的方式。

三、简评

开放开源正成为推动人工智能发展的最重要模式。 从埃隆·马斯克等人的表态和 OpenAI 截至目前的实际发展情况看，开放开源是 OpenAI 的重要特点，保持软件开源是 OpenAI 的承诺。开源的 OpenAI 的成立，不仅将人工智能研究推向又一个高潮，同时也打破了谷歌、脸书等信息技术巨头企业引领或独占人工智能领域的格局。未来，OpenAI 将成为处“在任何人控制范围之外的实验室”，并因此定位成为人工智能领域的监管者，将人工智能领域的发展引导到对人类更为安全的发展轨迹上。随着 OpenAI 与越来越多的科技企业、科研人员加强合作，将推动人

工智能新技术的研发，使开放开源成为人工智能领域发展的最重要模式。对此，我国也应学习借鉴这一思路，以开源模式参与、引导和推进人工智能领域的创新。

人工智能技术创新同时需认真考虑有关伦理问题。技术本身并无善恶，但因掌握者的不同有可能成为双刃剑。对人工智能威胁的担心，虽然显得“担心偏早”，但也并非无的放矢。未来，随着人工智能技术的进步，有关伦理问题必将产生并需认真面对。国外目前已在此方面开展实际工作，并发布了多项成果。未来，我国在推动人工智能技术发展的同时，也应组织开展对人工智能相关伦理、道德和社会影响的研究，分析潜在的负面冲击，有针对性地制定应对方案，适时形成法律法规，从而实现“软硬结合”，更好地促进人工智能技术的创新与应用。

本文作者：工业和信息化部赛迪研究院 安晖

联系方式：13810027908

电子邮件：anhui@ccidthinktank.com

研究，还是研究 才使我们见微知著

信息化研究中心

电子信息产业研究所

软件产业研究所

网络空间研究所

无线电管理研究所

互联网研究所

集成电路研究所

工业化研究中心

工业经济研究所

工业科技研究所

装备工业研究所

消费品工业研究所

原材料工业研究所

工业节能与环保研究所

规划研究所

产业政策研究所

军民结合研究所

中小企业研究所

政策法规研究所

世界工业研究所

安全产业研究所

编辑部：赛迪工业和信息化研究院

通讯地址：北京市海淀区万寿路27号院8号楼12层

邮政编码：100846

联系人：刘颖 董凯

联系电话：010-68200552 13701304215

010-68207922 18701325686

传真：0086-10-68209616

网址：www.ccidwise.com

电子邮件：liuying@ccidthinktank.com

报：部领导

**送：部机关各司局，各地方工业和信息化主管部门及
相关部门**

编辑部：工业和信息化部赛迪研究院

通讯地址：北京市海淀区万寿路27号院8号楼12层

邮政编码：100846

联系人：刘颖 董凯

联系电话：010-68200552 13701304215

010-68207922 18701325686

传 真：010-68200534

网 址：www.ccidwise.com

电子邮件：liuying@ccidthinktank.com

