人工智能：一项战略性技术的应用及治理

鲁传颖

【摘要】人工智能技术是引领全球科技革命和产业变革的战略性技术，正在重塑人们对国家安全、经济与社会发展的理解。快速发展中的人工智能也隐藏着一定的不确定性，可能会引发社会安全风险，甚至带来国家之间的竞争冲突升级。当前，人工智能的发展已经成为推动国际体系转变、影响国际关系走向的关键技术力量，我们需要从战略高度来看待人工智能的作用，充分认识其可能带来的机遇和风险，并从技术、应用、治理和规则等层面全方位加大对人工智能的投入。

【关键词】人工智能战略 人工智能治理 国际格局 治理困境 【中图分类号】D815 ；TP18 【文献标识码】A

如何定义人工智能的战略性意义，是这个时代需要回答的重要问题。国际问题专家基辛格指出，在人类所掌握的技术中，真正具有战略性意义的技术只有核武器、网络安全技术和人工智能。核武器是因为其毁灭性力量，网络安全则是因其广泛存在于所有的信息系统之中，并且持续面临被攻击的风险。人工智能的战略性则结合了前两者的特点，不仅具有毁灭性力量，而且也存在持续且广泛的安全性问题。对于人工智能这一重要战略性技术，国际、国内存在着多元的观点和态度。历史地看，大国必须从战略高度重视技术的发展，并且始终保持在全球技术发展的第一阵营。当然，与历史上其他技术不同，在人工智能发展战略中，安全和治理不可或缺，且是决胜的关键。

人工智能的战略性意义

人工智能被广泛认为是可以改变“游戏规则”的战略性科技。美国率先发布人工智能战略，随后中国、欧盟、日本、韩国等纷纷出台了各自的人工智能战略。而且，联合国秘书长古特雷斯也多次就人工智能问题发声，呼吁从战略高度来看待人工智能安全对国际体系的影响。尽管如此，人工智能在现实社会中的应用还处于早期阶段，公众对于人工智能颠覆性影响的感知并不强烈。因此，我们有理由去质疑人工智能是否可以担此重任。战略性技术意味着这是一项具全局性的、重大的、决定胜负的关键性技术。人工智能是否具备这三个特质？

首先，从全局性来说，人工智能技术的通用性将人类社会带入真正的“智能时代”。人工智能跟蒸汽机、发电机一样是通用型技术，对生产力的提升、人类生活方式的改变都有巨大的推动作用。正因如此，人工智能技术也将被国家、企业、个人等不同层次的行为体广泛使用，进而对政治、安全、经济、文化等不同领域产生深刻的影响。生成式人工智能的发展进一步降低了人工智能使用的门槛，大模型将会广泛地应用于国家治理与社会生活的各个领域，由此带来的全局性影响也将进一步凸显。生成式人工智能正在加速智能革命。一方面，立足于面向公众的通用大模型不断面世，逐渐在艺术设计、知识管理、市场营销、代码生成和客户服务等方面逐步取代原有的工作模式。另一方面，各行各业垂直领域的专用大模型研发也在如火如荼地开展，金融、政务、制造业、能源、医疗、零售、教育领域的应用场景也都比较成熟，人工智能会进一步赋能和改造这些行业。

其次，从重要性来看，人工智能深刻改变了国家安全与国际和平的基本范式。因为人工智能对国际和平与国家安全领域都有系统性影响，导致国家安全与国际和平的内容、形式和目标也发生了嬗变。可以说，人工智能技术将国家安全与国际和平带入新一轮的蜕变期，使得“安全”与“和平”超出了传统政治议题的范围，成为政治、经济和技术三维一体的复合型问题。可以从三个层面来理解人工智能所带来的影响。从赋能的角度看，人工智能将大幅提升战场的决策和行动能力。当前，人工智能技术已被大量运用到军事和情报领域。在军事领域，决策辅助系统、智能化后勤保障、无人机群、无

人舰艇等各种类型的人工智能武器正广泛应用于战场实验，将会推动军事变革从信息化作战转向智能化作战。从自主性的角度看，人工智能武器系统具备了自主决策的能力。目前，主要大国纷纷开发致命性自主武器，并投入战场予以测试，引发了关于技术伦理问题的广泛讨论和担忧。例如，美国具备自主判断能力的无人机就曾经误判目标，导致无辜平民被当作恐怖分子而被杀害。从技术的角度看，人工智能所带来的风险可以从客体和主体两个层面来理解，即技术本身，以及使用技术的“人”。人工智能的风险对使用主体提出了更高的责任要求。

最后，从决定胜负的关键来看，人工智能技术发展水平已经成为国家实力的关键指标。与其他通用型技术相比，人工智能建立在大模型、大参数、高算力和大数据的基础上，不仅产业融合程度高且易于垄断。加上人工智能作为军民两用型技术，市场的力量不仅会加速技术的发展和迭代，进而提升对军用技术的赋能效果，而且军事应用同样也会推动市场力量加大研发投入。如此循环增强之下人工智能技术的“马太效应”也愈加突出。领先者将会通过技术优势来占领市场并开启循环增强，进而拥有更多的先发优势，导致后来者在追赶的过程中面临更大的压力和障碍。在国际政治中，技术是各国提升硬实力的重要基础，也是改变国际格局力量对比的动力源之一。美国为了维护自身在人工智能领域的主导地位，其霸权主义倾向随着人工智能技术的发展日趋明显。事实上，从算法、算力和数据这三大人工智能发展的基础领域来看，美国占据了绝对的主导权。尽管如此，美国依旧高度警惕任何潜在对手在这些领域可能取得的突破。为此，美国政府制定了大量针对中国人工智能领域的封锁、制裁和惩罚性举措。

生存、安全、发展、治理：人工智能面临的多重风险

第一，人工智能给人类带来“生存之困”。人工智能的发展会极大提升人类的能力，同时也可能存在取代甚至摧毁人类的风险。通用人工智能技术的发展极大提高了人工智能的学习、掌握知识以及执行任务的能力。而人工智能一旦在认知能力上超越了人类，并且具有了自主决策的意识，人类或将会失去对人工智能的控制。如此一来，人工智能是否会伤害甚至毁灭试图控制它的人类呢？人工智能技术越发展、应用越广泛，类似的观点就会越流行。这种观点的产生一部分是受到了科幻小说的影响，也有一部分是由于人工智能技术复杂性所带来的理解障碍，因此造成了“技术恐慌”。同时，人工智能在技术上存在的算法黑箱、幻觉进一步助推了这种观点。这些讨论不仅迫使科技巨头呼吁加强对人工智能的监管，也推动着各国加速制定人工智能相关法律。这些做法旨在规范人工智能技术的演进，使其向着安全、可靠的方向发展。归根结底，人工智能技术是否会给人类社会带来毁灭性的影响，是对人类能否驾驭人工智能的考验。

第二，人工智能引发“安全之困”。人工智能在安全层面引发的“安全困境”已经为学者所认识并进行了充分讨论。国际政治的竞争归根结底是权力的竞争，而技术又是实现权力的重要依托。人工智能作为一项颠覆性技术，势必会引发国家间竞争，且存在“安全困境”的风险。此外，人工智能在国家安全、军事领域也具有广泛的应用空间，也可能会引发巨大的伦理风险。将载有人工智能系统的无人机应用于暗杀和地面攻势引发了关于机器杀人的伦理问题。一方面，人工智能可能会造成大规模的人员伤亡，造成大量的士兵失去生命。另一方面，人工智能存在误判的风险，无法有效识别军事目标和平民，从而加剧平民的无辜伤亡。

第三，人工智能引发不平衡的“发展之困”。人工智能具有提升经济效率的广泛前景，由此带来的“马太效应”会使得更多的资源向头部企业集中。不仅如此，人工智能技术的渗透可能还会引发一些企业的倒闭和失业潮。一方面，人工智能发展的主要

受益者首先是平台企业，它们掌握了算力和数据，拥有发展人工智能得天独厚的条件。另一方面，人工智能不断向垂直领域渗透，也会让更高效的“大机器生产”取代人类的工作，进而引发失业问题。国际金融服务公司摩根士丹利（Morgan Stanley）2023 年 9 月发布的报告显示，人工智能的发展在未来几年内将产生 4.1 万亿美元的经济影响，或者影响约 44% 的劳动力。由此可见，人工智能技术或将不可避免地造成失业问题的扩大。

第四，人工智能存在着“治理之困”。人工智能在赋能社会的同时，也正在被广泛应用于网络攻击、电信诈骗、虚假信息等违法犯罪领域，并且带来了一系列新的治理问题。传统的法律和规定往往无法完全适应人工智能技术的快速发展和复杂变化。如无人驾驶带来了责任认定的法律问题，生成式人工智能引发了知识产权问题。人工智能系统可以创造独特的作品和内容，但是如何保护这些作品的权益、进行合理的利益分享也是一大挑战。这些问题都触及到了当前的法律盲区，对现有的治理体系带来了极大挑战。因此，人工智能技术的到来迫使政府、学术界、产业界、技术社群等各利益相关方加强合作，共同研究和制定适应人工智能时代的新治理框架。

上述四个方面都是从不同层次的正反两个方面来理解人工智能，尤其是采取了静态的视角来看待人工智能，因此都不够全面。这些观点也未能认识到，随着各方对人工智能治理的关注和投入的加大，人工智能的治理体系也在快速建立。因此，不应当因为困境的存在就叫停技术的发展，而是要加速对人工智能的理解并提升治理能力。

人工智能发展正面临关键时刻

人工智能的关键时刻体现在三个方面，一是技术突破的关键时刻；二是大规模应用的关键时刻；三是大国博弈的关键时刻。

首先，从技术发展的历史进程来看，人工智能正在经历重大突破的关键时刻。自 20 世纪 50 年代人工智能技术的概念被提出以来，人工智能的发展经历了深蓝、阿尔法狗（AlphaGo）和 ChatGPT 这三次“高光时刻”。2022 年，OpenAI 发布 ChatGPT-3.5 迅速引发了全球热议，短短两个月的时间注册用户就突破了 1 亿。ChatGPT 背后是生成式人工智能所取得的一系列重大技术突破，主要包括基于 Transformer 深度学习架构的大型语言模型（LLM）；通过使用大规模数据集和无监督学习的方法对模型进行初始训练的预训练学习范式；以及结合了强化学习和人类反馈的机器学习技术——人类反馈强化学习（RLHF）。这些技术改变了机器对于自然语言处理的方式、提升了模型性能、降低了开发成本，并且建立了模型自我升级迭代的能力。因此，我们体验到的 ChatGPT 以及类似的产品在功能性、实用性和自我进化方面拥有极为优秀的表现。但是，技术的发展始终面临阶段性的波折，在生成式人工智能取得突破之后，人工智能的发展是否会陷入新的瓶颈期也是关键问题。因此，不能简单的用线性思维来看待人工智能的发展与治理。

其次，从应用角度来看，人工智能大规模应用一触即发。对比深蓝、阿尔法狗与新晋的大语言模型 ChatGPT 可知，前两次的突破更多是停留在技术层面，而 ChatGPT 的推出则标志着人工智能在技术和应用层面取得了双重突破。人工智能的应用价值主要体现在是否能够提升效率、降低成本和开辟新领域。目前，主流的科技企业都把人工智能视为核心竞争力并竞相加大投资，谷歌投资了 Deepmind、微软投资了 OpenAI、百度开发了文心一言等。根据国际数据公司（IDC）的数据，2021 年全球人工智能市场规模为 3619 亿美元，预计 2025 年将会突破 7000 亿美元，年复合增长率 18%。考虑到广泛的赋能作用，这一数据还只反映了人工智能经济价值的冰山一角。

最后，人工智能正处于大国博弈的关键时刻。一方面，大国追求人工智能的主导权引发了地缘政治博弈。美国政府明确将人工智能列为对华制裁的三大重点领域之一，制定了一系列出口管制措施，试图切断美国企业、资金和人才与中国之间的合作。不仅如此，美国还进一步将制裁延伸到 GPU 芯片领域，阻止中国获取先进的芯片用以训练人工智能模型。另一方面，大国围绕人工智能国际规则进行博弈。随着人工智能技术的不断发展，需要制定相应的国际规则和规范来管理、约束其应用。各国将争夺在国际规则制定中的话语权和主导权，以确保本国在人工智能领域的利益得到保护。

保证人工智能能够向善发展的关键

战略性价值和意义决定了人工智能的发展只会进一步加速而不会停止，因此，要从战略层面加大对人工智能的重视和投入。与此同时，加大在安全监管和治理方面的投入是保证人工智能能够向善发展的关键。

一是从战略高度坚定人工智能的发展方向，大力推动人工智能发展。面对人工智能这样一项通用型技术的发展，既会有受益者也会有受害者、既会有支持者也会有反对者、既会有激进者也会有保守者，而不同的群体、不同的视角会导致不同的观点。在 ChatGPT 出现之初，基于其在内容安全、知识产权等方面的不确定性，一部分企业对于是否要发展生成式人工智能存在疑虑。2023 年 4 月 28 日中共中央政治局会议明确提出“要重视通用人工智能发展”，同年 7 月 24 日再提“要促进人工智能安全发展”。如果说 ChatGPT 验证了通用人工智能发展的技术路线，官方表态则从政策方面为人工智能产业的大力发展奠定了基调。此外，生成式人工智能代表的是一种高算力、大参数、大模型、大数据的人工智能路线。尽管拥有广阔的市场前景，但也离不开国家在战略层面的支持和投入。在人工智能领域的开放式顶层设计思维，可以强化政府对于人工智能技术发展方向的参与和掌握，促进人工智能技术发展与国家战略目标的有机协调。

二是要高度重视人工智能的安全风险问题，强化对人工智能的安全监管。当前，人工智能的安全主要体现在算法安全、数据安全、网络安全等方面，其中关键是确保算法不会出现安全隐患。这就需要加强对人工智能的可解释性问题、透明度问题和有害性等方面的研究。此外，政府一方面可以通过建立技术标准来规范人工智能的开发和应用，提高人工智能的质量和性能；另一方面也可以检测和监测人工智能的状态和行为，及时发现和解决人工智能造成的问题和风险。人工智能发展不可避免地会出现各个层面的安全问题，如果处理不好，不仅会引发风险，也会降低各界对于人工智能发展的信心。这需要国家从多个角度进行治理。在技术层面，确保人工智能在算法安全、数据安全、网络安全等方面，避免出现错误、偏差、漏洞等问题。在法律层面，建立完善的人工智能法律制度和规范，明确人工智能的权利和责任，保护个人隐私和知识产权，防止滥用和滥权等问题。在社会层面，加强对人工智能的公众教育和宣传，提高公众对人工智能的认识和理解，培养公众对人工智能的信任和尊重，促进人工智能与社会各界的和谐共处。

三是要加大对治理的投入，解除人工智能发展的后顾之忧。人工智能的治理是一个复杂的治理体系，主要是国际和国内治理两个方面。国际治理主要解决使用人工智能的过程中所引发的国际安全问题，如人工智能在军事、外交、贸易等领域的应用可能导致的国家间冲突、竞争、不信任等问题。这些问题需要在全球范围内进行协商和协调，建立相应的国际规范和规则来约束国家使用人工智能时的行为，保障国际和平与稳定。国内社会治理则主要解决人工智能的伦理、价值问题，如人工智能对人类的尊严、自由、平等等基本权利和价值观的影响和挑战。这些问题需要在国家范围内进行立法和监管，建立相应的法律制度和规范来保护个人隐私和知识产权，防止滥用和滥权等问题。同时，也需要在系统开发和使用之前做好相应的价值评估，确保人工智能符合社会公益和道德标准。在治理机制上要充分尊重各参与主体的技术能力、商业利益与治理思路，通过充分对话和协调，形成更加合理的人工智能治理体系。各参与主体应该在平等和相互尊重的基础上，进行充分的沟通和协作，形成一个多元、开放、包容的人工智能治理体系。

（作者为上海国际问题研究院公共政策与创新研究所副所长、研究员）

【参考文献】

①[美]亨利·基辛格等著、胡利平等译：《人工智能时代与人类未来》，北京：中信出版社，2023 年。

②贾开：《人工智能与算法治理研究》，《中国行政管理》，2019 年第 1 期。

③周琪：《高科技领域的竞争正改变大国战略竞争的主要模式》，《太平洋学报》，2021 年第 1 期。

④ S. Matthew Liao, “Ethics of Artificial Intelligence”, New York, Oxford University Press, 2020.

⑤鲁传颖、约翰·马勒里：《体制复合体理论视角下的人工智能全球治理进程》，《国际观察》，2018 年第 4 期。

责编 / 周小梨　美编 / 王梦雅