

科技向善：中国全球人工智能治理倡议 的核心要义与理论价值^[1]

武 琼

【内容提要】2023年10月18日，习近平主席在第三届“一带一路”国际合作高峰论坛开幕式主旨演讲中提出全球人工智能治理倡议。该倡议的出台有其复杂的现实背景：一是美国对华实施全方位的人工智能战略竞争和打压；二是发展中国家和发达国家之间的智能鸿沟不断扩大；三是人工智能技术突飞猛进所带来的风险及挑战已成为世界各国面临的共同课题。全球人工智能治理倡议既是中国人工智能伦理原则的宣示，也是中国在该领域大力推进人工智能技术发展的路线图，展示出中国在推动全球人工智能技术健康发展方面的愿景和责任担当。该倡议的核心要义涵盖治理理念、治理路径、治理主体、治理原则、治理目标五个层面，具有重要的理论价值。

【关键词】中美战略竞争 人工智能 全球人工智能治理倡议 人工智能伦理观 人类命运共同体

【作者简介】武琼，宁夏大学阿拉伯学院（中国阿拉伯国家研究院）讲师，法学博士。

【中图分类号】D815

【文献标识码】A

【文章编号】1006-6241 (2024) 05-0025-25

[1] 作者感谢《和平与发展》匿名评审专家及编辑部对本文提出的宝贵修改意见，文中错误概由本人负责。

纵观人类发展历史，技术创新一直是推动生产力发展的核心驱动力。自18世纪以蒸汽机发明和应用为重要标志的第一次科技革命以来，人类社会已经历了三次科技革命，即蒸汽时代、电气时代、信息时代。目前，以人工智能（Artificial Intelligence，简称AI）为主要代表的第四次科技革命正在如火如荼地展开，其技术创新的速度、深度和广度，以及对经济、政治、军事和社会各层面的深远影响，均远超前三次工业革命。以ChatGPT为代表的大语言模型（简称大模型）的出现更是成为推动AI持续发展的关键驱动力。ChatGPT以“大数据+大算力+强算法”为其三大支撑核心，可以用来执行编写代码、文本生成、机器翻译等复杂任务。然而，AI是把“双刃剑”，在给人类社会带来巨大红利的同时，也带来了不可忽视的安全隐患和风险挑战，甚至可能会导致人类社会陷入“技术能力越强—安全风险越大”的恶性循环，这无疑引发了包括中国在内的世界各国对AI安全风险的广泛关注和深切担忧。

面对AI发展所带来的安全挑战，AI治理这一概念应运而生，加强AI领域的全球安全治理已成为人类社会必须面临的紧迫课题。2023年10月18日，习近平主席在第三届“一带一路”国际合作高峰论坛开幕式主旨演讲中提出全球人工智能治理倡议（以下简称“倡议”）。“倡议”既是中国AI伦理观的宣示，也是中国在该领域大力推进AI技术发展的路线图，展示出中国在推动全球AI技术健康发展方面的愿景和责任担当。目前，学术界关于“倡议”的研究成果主要集中在报纸和新闻网站的报道^[1]，尚未形成系统化的理论论述。因此，本文旨在从“倡议”出台的现实背景、核心要义及理论价值三个维度展开分析，以期推动学界和实务界对这一重要议题的持续关注和深入探讨。

[1] 参见《全球人工智能治理倡议发布》，新华网，2023年10月19日，<http://www.xinhuanet.com/tech/20231019/fb523e969d0349ab984f48176d507b76/c.html>；鲁传颖：《〈全球人工智能治理倡议〉中的中国智慧》，中国社会科学网，2024年1月23日，https://www.cssn.cn/skgz/bwyc/202401/t20240123_5730247.shtml等。

一、全球人工智能治理倡议出台的现实背景

在全球化的浪潮中，AI已成为国际竞争的新高地，其发展不仅重塑着全球经济格局，更深刻影响着国际关系的演变。美国等西方国家，凭借其在AI技术上的先发优势，通过经济制裁、网络攻击等手段，试图巩固其在全球AI领域的垄断地位，对中国等崛起中的大国进行遏制和打压。这种技术霸权行径，不仅损害了他国的正当发展权益，也加剧了全球智能鸿沟，使得发达国家与发展中国家之间的差距日益扩大。AI技术的“双刃剑”特性，使得其在推动社会进步的同时，也给人类社会带来一系列风险和挑战。这些挑战不仅考验着国家治理能力，也对全球治理体系提出了新的要求。如何在确保安全和伦理的前提下，利用AI技术推动可持续发展，成为国际社会共同面临的课题。中国作为AI大国，出台“倡议”正是基于这一仍在继续动态变化的复杂背景。

（一）美国对华实施全方位的AI战略竞争和打压

中国政府2017年颁布《新一代人工智能发展规划》，明确提出要把握世界AI发展趋势，突出研发部署的前瞻性，在重点前沿领域探索布局、长期支持，力争到2030年在AI理论、技术与应用方面总体达到世界领先水平，成为世界主要AI创新中心。^[1]对此，美国抛出了“中国意图垄断AI发展优势”的妄断。美国对外关系协会（CFR）发文宣称，中国正在不断加大对AI的政策支持和经费投入力度，旨在获得AI的全球主导地位；长此以往，中国的AI雄心将会对美国的技术霸权地位构成明确且可信的威胁。^[2]美国国家人工智能安全委员会（NSCAI）2021年3月发布的关于AI发展战略的

[1]《国务院关于印发新一代人工智能发展规划的通知》，中国政府网站，2017年7月8日，https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm。

[2] Gregory C. Allen, “China’s Artificial Intelligence Strategy Poses a Credible Threat to U.S. Tech Leadership,” Council on Foreign Relations, December 4, 2017, <https://www.cfr.org/blog/chinas-artificial-intelligence-strategy-poses-credible-threat-us-tech-leadership>.

最终报告中近 680 次提及“中国”或“中国的”字眼。^[1]实质上，所谓“中国 AI 威胁”只是美国用以推进所谓“对华战略竞争”的借口，而引领科技发展方向的 AI 也成为美国遏制打压中国的关键领域，也是其对华实施“脱钩断链”和“小院高墙”的重点环节。

一是在战略层面，美国近年来出台一系列有关 AI 的行政令、政策和立法，其中都包含“美国必须赢得对华竞争”的内容和越来越多专门针对中国的措施。例如，特朗普签署的《维护美国人工智能领导力的行政命令》、拜登签署的《2022 年芯片和科学法案》以及近年的《国防授权法案》都强调美国必须通过针对包括 AI 在内的技术出口管制和投资审查等手段来保障科技领先和国家安全。2024 年 10 月，拜登政府签署美国官方历史上首份有关 AI 的“国家安全备忘录”（NSM），将 AI 定义为“划时代的技术”（era-defining technology），要求美国政府各部门在涉及国家安全的领域“负责任地”应用 AI，并大力发展和使用尖端 AI 能力，以确保美国的战略竞争优势。^[2]

二是在经济技术层面，一方面，美国不断升级对华芯片及半导体制造设备技术出口管制。2022 年 10 月，美国商务部工业与安全局公布《对向中国出口的先进计算和半导体制造的出口管制》，旨在限制中国获取尖端 AI 芯片及其制造设备。^[3]2023 年 10 月，美国商务部工业和安全局通过调整尖端芯片许可证参数、采取新措施应对规避管制风险等手段进一步收紧对华尖端 AI 芯片及半导体制造设备技术的出口管制措施，旨在进一步阻断中国

[1] 魏际刚：《美国人工智能战略动向与中国应对之策》，载《中国经济时报》2022 年 5 月 30 日，第 A03 版。

[2] The White House, “ACT SHEET: Biden-Harris Administration Outlines Coordinated Approach to Harness Power of AI for U.S. National Security,” October 24, 2024, <https://www.whitehouse.gov/briefing-room/statements-releases/2024/10/24/fact-sheet-biden-harris-administration-outlines-coordinated-approach-to-harness-power-of-ai-for-u-s-national-security/>.

[3] Bureau of Industry and Security, “Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People’s Republic of China (PRC),” U.S. Department of Commerce, October 7, 2022, <https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file>.

获取尖端 AI 芯片及半导体制造设备技术的渠道。^[1]2024 年 10 月，美国财政部发布一项规则，限制和监控美国在中国人工智能、计算机芯片和量子计算领域的投资，并将于 2025 年 1 月 2 日开始实施。美国政府称这些敏感技术和产品一旦进入中国就有可能对美国国家安全构成威胁，因此要防止美国的尖端技术落入中国之手。^[2]另一方面，美国采取严厉措施制裁中国的高技术企业，压制中国高技术产业发展。根据外交部 2023 年 2 月的统计数据，美国不断加大对中国高技术企业的打压力度，将超过 1000 家中国企业纳入各类制裁名单，限制它们的国际合作与贸易。^[3]其中，华为公司成为遭受美国打压的典型案列。2020 年 8 月，美国商务部进一步将华为在全球 21 个国家/地区的 38 家子公司列入出口管制的“实体清单”，严重制约了华为正常业务和全球合作的开展。^[4]2024 年美国推动涉及 Tiktok 剥离内容的法案立法，其背后就有着为美国企业谋取竞争优势特别是巧取豪夺中国企业先进算法技术和数据的动机。

三是在外交层面，一方面，美国对盟伴拉拢和威逼并举，着力构筑 AI 反华联盟。2020 年 9 月，美国国防部联合 AI 中心正式启动“人工智能国防伙伴关系”（AI Partnership for Defense），加拿大、法国、英国等 13 个成员国

[1] Bureau of Industry and Security, “Public Information on Export Controls Imposed on Advanced Computing and Semiconductor Manufacturing Items to the People’s Republic of China(PRC),” U.S. Department of Commerce, November 6, 2023, <https://www.bis.doc.gov/index.php/policy-guidance/advanced-computing-and-semiconductor-manufacturing-items-controls-to-prc>.

[2] Paul Wiseman, “Treasury issues rule to block US investors from helping China develop advanced military technology,” *The Washington Post*, October 29, 2024, https://www.washingtonpost.com/business/2024/10/28/treasury-rule-china-artificial-intelligence-chips-investment/ba2aad54-956f-11ef-939a-5dd38bf140e0_story.html.

[3] 《美国的霸权霸道霸凌及其危害》，外交部网站，2023 年 2 月 20 日，https://www.fmprc.gov.cn/wjbxw_new/202302/t20230220_11027619.shtml。

[4] U.S. Department of Commerce, “Commerce Department Further Restricts Huawei Access to U.S. Technology and Adds Another 38 Affiliates to the Entity List,” August 17, 2020, <https://2017-2021.commerce.gov/news/press-releases/2020/08/commerce-department-further-restricts-huawei-access-us-technology-and.html>.

参与(后增加至16个),其目的是开发和具备互操作性的军用AI系统。^[1] 2022年3月,美国正式要求日本、韩国及中国台湾地区加入所谓“芯片四方联盟”。该联盟的实质是遏制中国迅速发展的芯片产业,并试图护持美国所谓技术霸权地位。另一方面,美国滥用“长臂管辖”,不断向第三国或地区施压,要求它们切断与中国的AI合作。以阿联酋为例,2024年4月,美国微软公司宣布与阿联酋科技集团G42达成AI合作协议。微软公司对G42投资15亿美元,并授权G42使用微软的Azure云服务,前提是G42必须停用中国华为的电信设备,并终止与中国的AI合作。美国商务部此前已明确表示,如果G42不同意终止与中国的AI合作,将面临来自华盛顿的“惩罚措施”。^[2] 美国还不遗余力地限制荷兰光刻机巨头ASML等企业对华出口先进半导体制造设备,企图通过封堵中国芯片制获能力阻扰中国AI技术进步。

四是在安全层面,一方面,美国大肆运用AI技术,对华实施网络攻击和数据窃取。例如,2022年,美国国家安全局(NSA)下属的特定入侵行动办公室(TAO)对中国西北工业大学进行了大规模网络攻击,窃取了超过140GB的高价值数据。这些数据的获取,使得美国情报部门能够利用深度学习算法对其进行快速分析,以挖掘情报并制定有针对性的打击计划,不仅减轻了情报人员的认知负荷,还提高了情报搜集的效率和准确性。此外,美国正在研发隐蔽性强、破坏力大的智能化网络武器,建设秘密网络武器库,以进一步增强其网络攻击能力。2024年10月,中国有关部门发布《伏特台风III——揭密美国政府机构实施的网络间谍和虚假信息行动》,进一步公开美国联邦政府、情报机构和“五眼联盟”国家针对中国等国家及全球互联网用户联合实施的网络间谍窃听窃密活动。据美国媒体报道,美国情

[1] JAIC Public Affairs, “DoD Joint AI Center holds fifth International Dialogue for AI in Defense,” Chief Digital and Artificial Intelligence Office, March 4, 2022, https://www.ai.mil/blog_03_04_22_dod_jaic_holds_fifth_international_dialogue_for_ai_in_defense.html.

[2] Ben Bartenstein, Mackenzie Hawkins, Nick Wadhams, et al., “G42 Made Secret Pact With US to Divest From China Before Microsoft Deal,” *Bloomberg*, April 16, 2024, <https://www.bloomberg.com/news/articles/2024-04-16/g42-made-secret-pact-with-us-to-divest-from-china-before-microsoft-deal>.

报机构专门研发了一款名为“大理石”（Marble）的网络工具框架包。该框架包能辅助网络武器开发者对程序代码中各种可识别特征进行“混淆”，使调查人员无法从技术角度追溯武器的真实来源。此外，该框架包具备“嫁祸”功能，通过植入俄、韩、阿拉伯文等语种的字符串来误导调查人员。^[1] 这些行动不仅侵犯了中国的网络安全，也对国际网络安全环境造成了严重威胁。另一方面，美国联合核心盟伴开展 AI 军事技术实验，并加快推动 AI 技术的军事实战应用。以“美英澳三边安全伙伴关系”（AUKUS）为例，2023 年 4 月，AUKUS 举行首次 AI 和自主技术试验。三方团队联合开发机器学习模型，并将其应用到“蓝熊幽灵”（Blue Bear Ghost）和 CT220 无人机，以及“海盗”（Viking）无人战车等军用机器人上^[2]，旨在测试目标识别和作战能力。2023 年 12 月，AUKUS 三国防长正式宣布启动“海上大戏”联合军演（Maritime Big Play），旨在进一步完善军用机器人的目标定位和精准打击能力。作为此次演习的一部分，三国海军于 2024 年 10 月从葡萄牙对 1.6 万公里外在澳大利亚海域航行的无人艇成功实施了操控。^[3] 值得警惕的是，美国也在加紧对华“无人智能”作战准备。2023 年 8 月，美国国防部常务副部长希克斯（Kathleen Hicks）宣布正式启动“复制器计划”（replicator initiative），提出美军将在未来 18 到 24 个月内构建“全域、耐损、自主”（ADA2）系统，在陆、海、空、天等各个领域部署数千个可以经得起

[1] Ellen Nakashima, “WikiLeaks’ latest release of CIA cyber-tools could blow the cover on agency hacking operations,” *The Washington Post*, March 31, 2017, https://www.washingtonpost.com/world/national-security/wikileaks-latest-release-of-cia-cyber-tools-could-blow-the-cover-on-agency-hacking-operations/2017/03/31/63fc3616-1636-11e7-833c-503e1f6394c9_story.html; 国家计算机病毒应急处理中心，计算机病毒防治技术国家工程实验室：《“伏特台风”III —— 揭密美国政府机构实施的网路间谍和虚假信息行动》，2024 年 10 月 14 日，https://www.cverc.org.cn/head/zhaiyao/futetaifeng3_CN.pdf。

[2] 按照任务领域的不同，军用机器人主要分为地面军用机器人（如无人战车）、水面军用机器人（如无人艇）、水下军用机器人（如无人潜航器）、空中军用机器人（如无人机）、空间军用机器人（如空间飞行器）。

[3] Sachin Ravikumar, “AUKUS navies remote-control ships from 10,000 miles away in experiment,” Reuters, October 4, 2024, <https://www.reuters.com/world/aukus-navies-remote-control-ships-10000-miles-away-experiment-2024-10-04/>.

美军大量消耗的军用机器人系统。^[1]2024年6月,美军印太总部司令帕帕罗(Samuel Paparo)在接受采访时表示,美军正在制定一项名为“地狱景观”(Hellscape)的作战计划。该项计划是指当解放军跨海攻打台湾地区时,美军将在台海及其附近部署数千艘无人潜艇、无人水面舰艇和无人机,把台海变成“无人地狱”,从而为台湾地区、美国和伙伴部队争取所谓“全面反击”时间。“地狱景观”计划实际就是“复制器计划”在台海场景的应用。^[2]事实上,上文提及的2024年10月的备忘录也明确要求美国国防部和情报机构在遂行任务时大力应用AI技术^[3],而这一点已经在其军事安全领域的对华围堵和困扰实践中得到充分体现。

(二) 发达国家和发展中国家之间的智能鸿沟不断扩大

智能鸿沟指的是在智能技术的研发与应用方面,不同国家、地区和群体之间的发展差异。随着AI技术的迅猛发展,这种鸿沟在发达国家和发展中国家之间愈加明显。

以美国为代表的发达国家目前在全球AI领域占据领先地位。亚马逊、谷歌、Meta(Facebook的母公司)、微软、OpenAI等美国科技企业都在AI研究上持续投入巨资,并在技术创新方面展现出强大的实力和资源。例如,OpenAI在2022年底推出的ChatGPT迅速引起了全球的瞩目。这种大模型的开发和运营成本极高,据斯坦福大学“以人为本”人工智能研究院(Stanford HAI)的估算,OpenAI的GPT-4的训练成本约为7800万美元,而谷歌的

[1] Jim Garamone, “Hicks Discusses Replicator Initiative,” U.S. Department of Defense, September 7, 2023, <https://www.defense.gov/News/News-Stories/Article/Article/3518827/hicks-discusses-replicator-initiative/>.

[2] Josh Rogin, “The U.S. military plans a ‘Hellscape’ to deter China from attacking Taiwan,” *The Washington Post*, June 10, 2024, <https://www.washingtonpost.com/opinions/2024/06/10/taiwan-china-hellscape-military-plan/>.

[3] The White House, “ACT SHEET: Biden-Harris Administration Outlines Coordinated Approach to Harness Power of AI for U.S. National Security,” October 24, 2024, <https://www.whitehouse.gov/briefing-room/statements-releases/2024/10/24/fact-sheet-biden-harris-administration-outlines-coordinated-approach-to-harness-power-of-ai-for-u-s-national-security/>.

Gemini Ultra 则高达 1.91 亿美元。^[1]此外，OpenAI 拥有一支结构合理、实力强劲的科研团队。据中国研究团队 2023 年相关报告显示，ChatGPT 团队共有 87 名成员，其中 1/4 的成员曾参与基于 GPT-3 的 CodeX 项目。^[2]

相比之下，一些发展中国家，特别是最不发达国家（如部分中东、南亚和非洲国家），由于在技术、人才和资金等方面的不足，严重限制了他们在 AI 领域的发展。许多发展中国家由于财政压力，对于建设数据中心等数字基础设施的投资有限，甚至无法承担相关成本。即使能够投资建设，其后续的维护和运营成本也是一项重大挑战。2023 年 10 月，在联合国大会第二委员会（负责经济和财政工作）会议上，来自发展中国家的代表普遍认为，AI 的发展使得发达国家和发展中国家之间的发展差距不断拉大，甚至可能导致最贫穷的国家被永远排除在第四次工业革命之外。^[3]

总体来看，由于掌握了顶尖科研人才、充足资金、强大算力和海量数据，微软、谷歌、Meta 和 OpenAI 等美国科技企业将继续主导 AI 领域的发展。这可能导致全球范围内的“强者恒强，弱者恒弱”局面，使得发达国家和发展中国家之间的智能鸿沟问题更加突出。

（三）AI 带来的风险挑战不容忽视

AI 是一把“双刃剑”，在为人类社会提供发展机遇的同时，也带来了失业、隐私泄露、战争伦理冲击和技术滥用等一系列风险挑战。

一是在经济发展层面，AI 虽然极大地提升了生产效率和经济效益，但也带来了一系列不可忽视的经济风险。AI 技术可能导致大规模失业，因为它能够替代许多传统工作岗位。美国麦肯锡全球研究院 2017 年的报告预测，

[1] Ray Perrault and Jack Clark, “Introduction to the AI Index Report 2024,” Stanford Institute for Human-Centered Artificial Intelligence, April 2024, pp.5-30.

[2] AMiner 和智谱研究：《ChatGPT 团队背景研究报告》，AMiner 网站，2023 年 2 月，<https://originalfileserver.aminer.cn/sys/aminer/ai10/pdf/ChatGPT-team-background-research-report.pdf>。

[3] United Nations, “Widening Digital Gap between Developed, Developing States Threatening to Exclude World’s Poorest from Next Industrial Revolution, Speakers Tell Second Committee,” October 6, 2023, <https://press.un.org/en/2023/gaef3587.doc.htm>.

到2030年，全球可能有多达8亿个工作岗位将被自动化技术所取代。即使是最保守的估计，也至少有4亿个岗位将受到AI和机器人技术发展的影响，需要寻找新的就业机会。^[1]此外，ChatGPT等生成式AI技术的出现无疑进一步加剧了这一进程。这些大模型在文本起草、知识搜索和代码编写等方面展现出与人类专家相媲美的能力，使得企业可能出于成本和效率的考虑，减少对传统岗位的需求，甚至可能引发大规模裁员。美国高盛集团相关研究显示，仅生成式AI就可能影响到全球3亿个工作岗位。^[2]这不仅对个人职业生涯产生影响，还可能引发财富分配不均和市场垄断等更广泛的经济问题。

二是在隐私保护层面，AI应用存在隐私泄露的风险。作为国家基础性战略资源，数据是AI发展必不可少的“燃料”。没有海量数据支撑，AI难以实现快速发展和迭代升级。然而，AI在高度依赖海量数据的同时，也在一定程度上存在数据泄露的风险。尤其是当海量数据被收集后，可以借助算法来推测用户的个人特质和行为偏好，从而不可避免会涉及到个人隐私保护问题。2018年4月，Facebook承认在未经用户允许的前提下向剑桥分析公司泄露了8700万用户的个人数据，并将其用于政治竞选活动。通过使用算法，Facebook随处可见的“点赞”可以被用来预测性别、种族、政治倾向、个性特质等有关点赞者的敏感个人信息。需要说明的是，以ChatGPT为代表的大模型是建立在海量数据的训练基础上，这意味着其训练阶段需要大量收集和挖掘数据，其中很有可能涉及到个人隐私信息。一旦大模型过度收集并违规使用个人敏感数据，则会导致严重的隐私泄露和滥用。目前，亚马逊、微软、摩根大通等大型企业已提醒员工，禁止向ChatGPT分享敏感数据，以防止泄密。

[1] Susan Lund, Jacques Bughin, James Manyika, et al., “Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages,” McKinsey & Company, November 28, 2017, <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages>.

[2] Goldman Sachs, “Generative AI could raise global GDP by 7%,” April 5, 2023, <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent>.

三是在战争伦理层面，AI使用不当会严重冲击战争伦理。受到硬件故障和“算法黑箱”^[1]等因素影响，军用机器人在执行作战任务的过程中可能会偏离预定目标，甚至不受指挥员控制，导致无法区分敌对力量和无辜平民。在这种情况下，传统的伦理道德原则已很难对其展开约束。如美国在海外实施的反恐行动中，无人机的使用已造成大量无辜平民伤亡，给当事国带来严重的人道主义灾难。据统计，自2002年以来，美军至少进行了1.4万次的无人机空袭，导致约2200名平民死亡。^[2]此外，美军正加速将ChatGPT等大模型应用于数据处理、指挥决策和态势分析等领域，以提升军队的战斗力和作战效率。

然而，这些大模型存在生成错误信息的风险，即所谓的“幻觉”（hallucinations）。^[3]根据美国普林斯顿大学教授阿隆德拉·尼尔森（Alondra Nelson）等AI领域专家的研究，ChatGPT-4、Gemini等大模型的总体表现不佳，其回答问题时存在高达50%的错误率，还有超过三分之一的回答是不完整的，甚至给出的是严重不实信息。^[4]虽然目前尚未有公开的新闻报道明确证实大模型在军事领域的应用直接造成了平民伤亡事件的发生，但仍要对此持高度警惕态度。美国机器学习专家海蒂·赫拉夫（Heidy Khlaaf）发出预警，大模型中普遍存在的“幻觉”等问题致使其缺乏足够的准确性，

[1] “算法黑箱”是指在算法流程中，由于计算机代码的高度复杂性和不透明性及其工作原理和运行机制的隐蔽性，用户很难理解算法流程的运行规律和因果逻辑关系，从而使算法处于不公开和不透明状态。

[2] “Senators Warren and Murphy, Rep. Khanna Call on President Biden to Overhaul U.S. Counterterrorism Policy,” U.S. Senator Elizabeth Warren of Massachusetts, January 20, 2022, <https://www.warren.senate.gov/newsroom/press-releases/senators-warren-and-murphy-rep-khanna-call-on-president-biden-to-overhaul-us-counterterrorism-policy>.

[3] 大模型有时会生成看似合理连贯、但实际上却漏洞百出的答案，从而产生所谓的幻觉。幻觉的存在会让机器在某些情况下形成讨好型人格，宁可编造，也不会承认不知道答案是什么。参见 Gerrit De Vynck, “ChatGPT ‘hallucinates.’ Some researchers worry it isn’t fixable,” *The Washington Post*, May 30, 2023, <https://www.washingtonpost.com/technology/2023/05/30/ai-chatbots-chatgpt-bard-trustworthy/>.

[4] Beth Simone Noveck, “How AI chatbots could improve civic engagement in the 2024 election,” *RebootDemocracy*, August 1, 2024, <https://rebootdemocracy.ai/blog/How-AI-chatbots-could-improve-civic-engagement>.

将其应用于军事领域会引发不精确和有偏见的军事行动，从而加剧平民伤亡。^[1]总之，AI武器只是一台完成人类布置任务的系统，一旦被别有用心的人重新编程，极有可能制造出一台滥杀无辜的“杀人机器”。^[2]

四是在技术扩散方面，AI的滥用风险日益凸显，特别是其在程序和软件形式上的易扩散性，使得潜在风险远超传统武器。AI技术，如“深度伪造”（Deepfake），正被用于制造虚假信息 and 进行诈骗活动。例如，2024年2月，中国香港警方揭露了一起利用“深度伪造”技术诈骗2亿港元的案件。这种技术通过AI算法操纵图像或视频，制造出逼真的假象，足以欺骗专业人士，甚至是经验丰富的专家学者。随着AI语言模型的进步，如ChatGPT，它们在提供犯罪信息、网络诈骗、恐怖主义宣传和虚假信息传播方面扮演了帮凶的角色。这些大模型能够快速生成逼真的文本，模仿特定个人的写作风格，甚至在没有技术背景的人手中也能制造出高度可信的虚假内容。这种能力极大地降低了网络犯罪的门槛。此外，据欧洲刑警组织（Europol，又称欧洲警察署、欧盟执法合作署）的报告显示，随着AI技术在代码开发方面的应用越来越广泛，高级别犯罪分子可以在此基础上完善大模型，从而使其网络犯罪行为的模式识别日趋复杂。^[3]例如，AI可以用于生成恶意软件，这些软件能够自我学习和适应，以逃避传统的安全检测。

二、全球人工智能治理倡议的核心要义

“倡议”是中国提出的一份旨在指导和促进全球AI领域健康发展的框

[1] Sam Biddle, “OpenAI Quietly Deletes Ban on Using ChatGPT for ‘Military and Warfare’,” The Intercept, January 12, 2024, <https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/>.

[2] 许玥凡、张竣敏：《人工智能武器：福兮，祸兮？》，载《解放军报》2018年6月29日，第11版。

[3] Europol, “ChatGPT: The impact of Large Language Models on Law Enforcement,” March 27, 2023, <https://www.europol.europa.eu/cms/sites/default/files/documents/Tech%20Watch%20Flash%20-%20The%20Impact%20of%20Large%20Language%20Models%20on%20Law%20Enforcement.pdf>.

架性文件。“倡议”强调了 AI 治理的全球性、开放性、合作性和公正性，旨在推动国际社会共同应对 AI 发展带来的挑战，促进 AI 技术的全球治理。它既是中国 AI 伦理观^[1]的宣示，也是中国在该领域大力推进 AI 技术发展的路线图。“倡议”的核心要义可概括为治理理念、治理路径、治理主体、治理原则、治理目标五个层面。

（一）在治理理念层面，坚持以人为本、智能向善的理念

中国在 AI 领域的发展和治理中，始终坚持“以人为本、智能向善”的理念，这一理念深受中国传统文化中民本思想的影响。《尚书·五子之歌》中所提到的“民惟邦本，本固邦宁”强调了以民为本的治国理念，这种思想在中国历史上逐渐发展成为一个丰富的思想体系，涵盖了爱民、重民、护民、顺民、安民、亲民、利民、敬民、忧民等多个方面。在 AI 快速发展的今天，中国继承并发扬了这些传统理念，确保 AI 技术的发展能够真正造福人类。

与此同时，中国的 AI 治理理念也突破了西方模式的局限。自工业革命以来，西方国家对先进技术的追求往往受到资本本性的驱动，即追求利润最大化，且这种追求呈现出无节制的趋势。在资本主义逻辑的驱动下，要求机器超越、替代人类，以更有效率地攫取利润，而随着资本的不断积累和生产的快速扩张，这种追求变得更加肆无忌惮，以至于对工人的生命安全置若罔闻。“倡议”则突破了以上局限：

一是主张 AI 的发展和治理应符合人类的价值观和伦理道德，通过人机和谐融合来服务于人类发展。人机融合是指通过实现人类和机器间的密切合作和相互补充，从而实现便捷、高效、智能的工作方式和生活方式。基于此，中国的治理理念将 AI 视为人类的亲密伙伴和得力助手——在这种人机关系中，人类仍然是核心主导力量，旨在确保 AI 的行为处于人类的高度

[1] “人工智能伦理观”（AI Ethics）是一个广泛的概念，它涉及到 AI 技术在设计、开发、部署和使用过程中应遵循的道德原则和规范。其目的是确保 AI 系统能够以负责任、公正和透明的方式运作，同时尊重人权、隐私和尊严。其中涵盖了透明度、公平性、问责制、隐私保护、安全和可靠性、数据偏见、人机协作等诸多核心概念。

控制下；人机相互合作，从而形成一个密不可分的实体。

二是主张 AI 的发展与治理要以尊重人类权益为前提，确保 AI 始终朝着有利于人类文明进步的方向发展。当前，AI 的算法设计和产品研发主要集中在大型科技公司手中。然而，资本逐利的贪婪本性导致这些公司“因机忘人”，从而引发侵犯和践踏人类权益的倾向。基于此，中国主张在 AI 发展和治理时应最大限度地避免对人的生命安全、身体健康、精神和心理健康造成伤害或潜在威胁，确保技术最终服务于人类而非取代人类。

三是主张加强对 AI 军事应用的规范，预防和管控可能引发的风险。事实上，早在 2021 年 12 月，中国时任裁军大使在出席联合国《特定常规武器公约》第六次审议大会期间，向大会提交了《中国关于规范人工智能军事应用的立场文件》，从战略安全、军事政策、风险管控、规则制定等 8 个方面阐述中方相关立场，呼吁各国加强对话与合作，推动构建有效治理机制，避免 AI 军事应用给人类带来重大损害甚至灾难。这也是中国首次就规范 AI 军事应用问题提出倡议。2023 年的“倡议”在战略思想上进一步升华，呼吁各国秉持共同、综合、合作、可持续的安全观，致力于提升 AI 技术的安全性、可靠性和可控性，并强调大国在军事领域研发和使用 AI 技术应该采取慎重负责的态度，为积极有效管控 AI 军事应用的潜在安全风险提供了建设性思路。

（二）在治理路径层面，构建开放、公正、有效的全球 AI 治理体系

为应对 AI 快速发展所带来的挑战和风险，确保技术进步与社会价值、伦理规范和国际合作同步发展，国际社会急需搭建起一个开放、公正、有效的 AI 治理体系。因此，“倡议”提出全球 AI 治理应当引入以下三种机制。

一是通过建立健全法律和规章制度，保障 AI 研发和应用中的个人隐私与数据安全。作为基础性和战略性资源，数据是 AI 发展的重要基石。在 AI 时代，人们日常的通讯、网络行为等都会产生海量数据，与此同时，这些数据也面临着被窃取、泄露、滥用、非法使用的安全风险。尤其是一些大型科技公司出于利益考虑，过度采集和使用个人数据，从而严重侵犯个

人隐私，甚至是危害国家安全。因此，通过建立健全相关法律和规章制度，以确保个人隐私数据的合法合规收集、存储和使用已迫在眉睫。在这一领域，中国已取得显著成果，先后颁布了《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律，明确规定个人信息受法律保护，任何组织、个人不得侵害个人信息权益。

二是坚持伦理先行，建立并完善 AI 伦理准则、规范及问责机制，形成 AI 伦理指南，建立科技伦理审查和监管制度，明确 AI 相关主体的责任和权力边界。在大国竞争日趋激烈的情境下，AI 正在突破传统的伦理道德底线，为科技伦理治理带来了巨大挑战。基于此，建立伦理准则、规范及问责机制与科技伦理审查和监管制度是确保 AI 技术符合伦理道德规范的重要内容。在这一方面，中国已修订了《中华人民共和国科学技术进步法》，并发布了《新一代人工智能治理原则——发展负责任的人工智能》《新一代人工智能伦理规范》《关于加强科技伦理治理的意见》《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等一系列行政法规和规范性倡议，旨在进一步健全科技伦理治理的审查规则和监管制度。这不仅可以明确设计者、生产者和使用者的各主体的责任意识，还能不断提升 AI 的透明性和可靠性，最终实现可审核、可监督、可追溯、可信赖的 AI 发展和治理。

三是积极倡导在联合国框架下建立国际 AI 治理机构，推动形成具有广泛共识的 AI 治理框架和标准规范，以协调 AI 发展、安全与治理等重大问题。2023 年 10 月 26 日，联合国宣布正式组建 AI 高级别咨询机构，由来自中国、美国、俄罗斯、日本、英国、以色列、德国、韩国等国的 39 名成员组成。该机构旨在探讨 AI 技术所带来的风险挑战，其主要议题包括 AI 的国际治理；对 AI 所带来风险和挑战的共同理解；利用 AI 加速实现可持续发展目标的机遇。该机构发布了题为《为人类治理 AI》临时报告，提出通过执行七项“由简至难”的关键措施来构建 AI 治理机制，其措施包括凝聚科学共识，协调标准、安全和风险管理框架，展开跨境数据模型培训和测试，制定规范和

问责制度等。^[1]2024年7月1日,第78届联合国大会协商一致通过了由中国提出的“加强AI能力建设国际合作”决议,该决议得到了140多个国家的联署支持。作为联合国首份关于AI能力建设国际合作的决议,它不仅反映了国际社会对于通过对话与合作加强全球AI治理的普遍共识,也突显了中国在全球AI发展和治理中的负责任态度和引领作用。^[2]这一决议的通过,标志着国际社会在携手推进构建全球AI治理框架方面迈出了关键一步。

(三) 在治理主体层面,实行以多主体参与为主的治理模式

“倡议”表明,在全球AI治理的多主体格局中,中国正以政府、民间机构、科研院所、企业及个人等多种角色身份,积极参与AI领域的国际交流与合作,在推动中国AI技术发展的同时,不断探索共商共建共享的全球治理途径。

一是在政府层面,中国展现大国担当,致力于推动全球AI治理,推进AI赋能世界各国经济社会和文明发展,真正造福全人类。近年来,中国政府始终坚持发展和安全并重的理念,主动加强AI领域伦理治理和法治建设,发布实施了上文所提及的一系列专门的法规和规范性建议,展示自身负责任形象,并提出了“倡议”,积极推动国际协作。2017年以来,中国政府多个部门和机构与上海市政府共同主办了7届世界人工智能大会(WAIC),邀请各国和国际组织政产学研界人士和AI领域顶尖专家与会,共同探索深化创新合作、推动普惠发展、加强协同共治的方式路径,为推动AI更好服务全球发展、增进人类福祉提供中国贡献。

二是在国际组织层面,中方深度参与全球AI治理行动,积极参与联合国教科文组织、联合国高级别AI咨询机构等国际组织的AI伦理治理工作,通过发起和共同提案等方式,积极推动联合国大会通过AI相关决议或草案,促进国际合作。中国还不断加强和红十字国际委员会(ICRC)、东盟等国际

[1] United Nations, “Interim Report: Governing AI for Humanity,” December 2023, https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf.

[2] 《第78届联合国大会协商一致通过加强人工智能能力建设国际合作决议》,中华人民共和国常驻联合国代表团网站,2024年7月1日, http://un.china-mission.gov.cn/hyyfy/202407/t20240702_11445457.htm。

组织在 AI 治理领域的交流与合作。2024 年 6 月，中国—东盟数字治理对话在桂林举办。对话中，与会嘉宾围绕“AI 治理”“数据治理”等议题分享本国的政策和实践，进而提出有针对性的合作建议，旨在凝聚合作共识，深化和拓展中国与东盟在 AI 治理领域的合作。^[1]

三是在非政府和民间层面，中方利用 AI 人才资源丰富、科研力量雄厚、行业洞察力强、产业应用广以及技术、法律、伦理、管理等专业经验丰富等优势，积极推动对外交流合作和 AI 治理。在科研院校方面，以清华大学战略与安全研究中心（CISS）为代表的高校定期与美国智库就 AI 全球治理、AI 安全风险等议题进行深入讨论，探索构建全球 AI 治理机制的可行路径。清华大学战略与安全研究中心自 2019 年开始就与美国智库布鲁金斯学会开展“中美 AI 与国际安全对话”，迄今为止已举办 10 轮对话。双方讨论的议题包括 AI 军事化应用的安全风险、制定全球 AI 治理机制、为技术发展设置共同原则等。^[2]在科技企业方面，阿里巴巴、旷视科技等国内领军公司通过成立 AI 实验室或发布 AI 治理白皮书等方式来积极探索企业在推动 AI 治理时应采取的价值理念和行动路径。例如，阿里巴巴集团已成立 AI 治理与可持续发展研究中心，该中心专设了 AI 安全实验室和 AI 伦理与道德实验室。^[3]2024 年 8 月在成都举办的 2024 年中国网络文明大会上发布了华为、腾讯、科大讯飞、百度、阿里云以及清华、北大等 70 多家企业和高校共同参与的《生成式人工智能行业自律倡议》，从保障数据和算法模型安全合规、促进内容生态建设、推动技术创新与质量提升、遵循价值观与伦理道德标准和促进交流合作与开放共建等方面提出行业倡议。^[4]在专家咨询方面，中国国内颇具影响力的专家通过加入联合国层面的高级别咨询机构，

[1]《中国—东盟数字治理对话在桂林举办》，人民网，2024 年 6 月 21 日，<http://gx.people.com.cn/n2/2024/0621/c179464-40886579.html>。

[2]《CISS 举办第八轮“中美人工智能与国际安全对话”》，清华大学战略与安全研究中心网站，2023 年 10 月 6 日，<https://ciss.tsinghua.edu.cn/info/yw/6488>。

[3] 参见阿里巴巴人工智能治理与可持续发展研究中心网站，<https://s.alibaba.com/cn/aaig/home>。

[4]《〈生成式人工智能行业自律倡议〉发布》，中国文明网，2024 年 8 月 30 日，<http://www.wenming.cn/wmzthc/20240830/73a42dcc72c64b2192f600767d4d43a6/c.html>。

积极参与全球 AI 治理讨论，为 AI 全球治理提出相关建议。在联合国成立的 AI 高级别咨询机构中，有两名来自中国的学者入选该机构成员，负责提供与 AI 安全、伦理和治理等相关的技术、规范和法律等建议。^[1]

（四）在治理原则层面，秉持发展和安全并重、相互尊重、平等互利、广泛参与、协商一致、公平性和非歧视性等基本原则

一是发展与安全并重。“倡议”提出，在推动 AI 发展的同时，应确保安全与发展成为并行不悖的战略重点。这意味着，一方面，要从国家战略层面积极推进 AI 发展，通过强化顶层设计和统筹规划，力争在新一轮国际科技竞争中掌握主导权。其中包括挖掘 AI 在医疗、教育、交通、农业、工业、文化、生态等多个领域的应用潜力；促进全球研究资源的交流与合作；确保数据安全，以支持高质量数据的发展，合作打造高质量数据集^[2]；培养更多的 AI 专业人才，加强教育培训与人才交流合作等。另一方面，要加强对 AI 所带来安全风险的管控，不断提升 AI 的安全性、可靠性、可控性、公平性，并确保其始终朝着有利于人类文明进步的方向发展。具体措施包括重视数据安全与隐私保护，加强各国间数据保护政策的兼容性；加强监管，打造可审核、可监督、可追溯和可信任的 AI 技术；加强与 AI 相关的网络安全措施，增强系统与应用的安全性、可靠性，防止黑客攻击等。^[3]

二是相互尊重、平等互利。“倡议”认为，鉴于各国在历史传统、文化背景、社会制度、发展水平上的差异，只有尊重和认可这些差异，才能有效地推动国际间的合作与交流。同时，不应把自己的意志强加于人，而应致力于确保所有国家，特别是发展中国家，能够平等地发展和利用 AI 技术，

[1] 两人分别是中国科学院自动化研究所曾毅研究员和中国政法大学数据法治研究院张凌寒教授。参见 United Nations, “Members of the High-level Advisory Body on Artificial Intelligence,” <https://www.un.org/en/ai-advisory-body/members>.

[2] 指那些能够为机器学习模型提供准确、一致、全面且无偏见的信息的数据集合。这些数据对于训练有效的 AI 模型至关重要，因为它们直接影响模型的准确性和泛化能力。

[3] 《人工智能全球治理上海宣言》，外交部网站，2024年7月4日，https://www.mfa.gov.cn/zyxw/202407/t20240704_11448293.shtml。

并从中获益。中国作为负责任的大国，切实尊重发展中国家在 AI 发展上的权利，主张通过国际合作和援助的方式来缩小发达国家与发展中国家之间的智能鸿沟，帮助发展中国家从 AI 发展浪潮中受益。

三是广泛参与、协商一致。“倡议”强调联合国在全球 AI 治理中的核心作用，倡导各利益攸关方，包括政府、国际组织、科研机构、企业、民间机构及个人等主体，在 AI 治理领域展开广泛的交流与互动。通过促进积极和建设性的对话，“倡议”鼓励相关方共同探讨并制定出被普遍认可的全球 AI 治理框架和标准规范，以确保该治理体系的公正性和合理性。这一过程旨在汇集多元智慧，形成国际共识，推动构建一个更加公平和合理的全球 AI 治理机制。

四是公平性和非歧视性。“倡议”指出，在 AI 的设计、开发和使用过程中，应确保所有人都能公平地使用 AI 技术，并尽量避免该项技术被用于种族、性别或其他形式的歧视。理论上讲，高质量的数据是训练深度学习算法的核心。然而，如果使用带有偏见、有误导性或包含歧视性的数据来训练深度学习算法，则很可能导致输出结果含有歧视性或误导性。这不仅影响相关决策的正确性，还可能加剧社会不公。因此，在数据准备和模型训练阶段使用公正、平衡的数据集是非常重要的。同时，为防止算法固化歧视性结果，必须对训练数据进行严格审查，并在模型迭代过程中持续监控和修正偏差，从而确保算法不会陷入歧视性行为模式，以保障 AI 技术的公正性和可靠性。

（五）在治理目标层面，以增进全人类的共同福祉为行动遵循

在治理目标的层面上，“倡议”致力于推动全球平等参与 AI 的发展，反对技术垄断，并共同应对气候变化和生物多样性保护等全球性挑战，以增进全人类的共同福祉。

一是反对以美国为首的西方国家以意识形态划线或构建排他性集团，以及利用其技术垄断优势阻挠他国在 AI 领域的发展。自特朗普政府以来，美国将中国定义为主要的“战略竞争对手”，将战略竞争确定为对华政策的

主基调，并将 AI 等高新技术领域视为对华竞争的核心领域之一。特朗普政府采取了包括限制出口、加强投资审查和构建技术联盟在内的多种措施，以遏制中国在 AI 等相关产业的发展，试图保持其在全球 AI 领域的领先地位。此后的拜登政府开始实施以“小院高墙”为核心的对华科技打压政策，把技术封锁的范围缩小至 5G、量子信息、AI 等高新技术领域，以确保美国在高新技术领域的技术霸主地位。为此，美国以“国家安全”为由，动用举国之力对中国进行打压。这种做法不仅妨碍了中美之间的技术交流与合作，也与通过科技发展来增进人类福祉的目标背道而驰。因此，“倡议”呼吁美国与中国一道，共同营造一个开放、公平、公正、非歧视的国际科技发展环境，携手应对 AI 治理缺位这一重大难题，而不是一味的通过筑起“小院高墙”来维护其全球科技霸权。

二是“倡议”强调应确保发展中国家有机会平等地参与 AI 发展和治理，并能平等地分享该项技术带来的红利。对于广大发展中国家而言，AI 的快速发展和广泛运用，对其发展具有重大意义。作为一个致力于维护世界和平与发展的负责任大国，中国正在加快推动与发展中国家在 AI 领域的合作，以助力发展中国家提升 AI 治理能力。2024 年 4 月，中非互联网发展与合作论坛发布关于中非 AI 合作的主席声明，承诺在技术研发与应用、产业合作与发展、人才交流与能力建设等领域展开合作，并鼓励中非双方企业、高校和科研机构在大数据分析、机器学习、自然语言处理、计算机视觉等领域开展合作^[1]，以确保发展中国家能从中获得应有的发展机遇。

三是“倡议”强调支持利用 AI 技术来应对气候变化和生物多样性保护等全球性挑战。气候变化和生物多样性的丧失是关乎全人类福祉的重大全球性挑战，它们的应对需要全球各方的紧密合作与协同努力。而 AI 技术能为应对气候变化提供多方面的支持，包括监测冰山融化的速度和范围、绘制森林砍伐的详细地图、提高废物管理的效率、减少海洋中的塑料污染、

[1]《2024 年中非互联网发展与合作论坛关于中非人工智能合作的主席声明》，中央网络安全和信息化委员会办公室网站，2024 年 4 月 3 日，https://www.cac.gov.cn/2024-04/03/c_1713731793084792.htm。

精准预测气候灾害的发生时间和地点，以及助力金属、矿业、石油和天然气等关键行业的企业实施脱碳等^[1]，展现出其在环境保护和可持续发展领域的重要价值。

三、全球人工智能治理倡议的理论价值

在全球 AI 治理的浪潮中，习近平主席提出的“倡议”不仅是对国际科技治理体系的一次重要贡献，更是对中华文明智慧的一次现代传承与创新。“倡议”将中华文明中深厚的“人本文明”^[2]底蕴与现代科技治理的先进理念相融合，实现了传统与现代的和谐交融与创新性发展。它深刻体现了以人民为中心的发展思想，超越了西方传统安全观念的局限，强调在 AI 的发展进程中，必须将人民的利益和福祉置于核心地位。“倡议”不仅深化了构建人类命运共同体的理念，而且将中国在全球治理中的发展经验和实践智慧与全球治理的普遍需求相结合，提出了一种烙印着中国特色的全球 AI 治理模式，具有重要的理论价值。

（一）“倡议”助推中华民族现代文明的新发展

历经五千多年而绵延不绝的中华文明深深地植根于中华大地，其中丰富的治国理政经验和思想智慧被现代中国所继承和弘扬。习近平指出，中华优秀传统文化有很多重要元素，比如，天下为公、天下大同的社会理想，民惟邦本、为政以德的治理思想，九州共贯、多元一体的大一统传统，修齐治平、兴亡有责的家国情怀，厚德载物、明德弘道的精神追求，富民厚生、

[1] Victoria Masterson, “9 ways AI is helping tackle climate change,” World Economic Forum, February 12, 2024, <https://www.weforum.org/agenda/2024/02/ai-combat-climate-change/>.

[2] 人本文明是指以人为中心的文明形态，它强调人的主体地位和价值，倡导人的全面发展和福祉。这种人本的思维方式和价值取向，贯穿于社会制度、文化、教育、科技等各个领域，旨在构建一个以人的需要、利益和尊严为最高准则的文明社会。参见赵诗清、曾瑞芝：《论文明》，光明网，2000年12月5日，<https://www.gmw.cn/01gmr/2000-12/05/GB/12%5E18625%5E0%5EGMB3-014.htm>。

义利兼顾的经济伦理，天人合一、万物并育的生态理念。^[1]

党的十八大以来，中国全力推动中华优秀传统文化创造性转化、创新性发展，为实现民族复兴提供了强大思想武器和科学行动指南。众所周知，AI是一把“双刃剑”，即发展机遇和风险挑战并存，一旦应用不当，就会带来诸多风险挑战。而中华优秀传统文化对解决相关难题具有重要的启示意义。面对AI发展带来的各类风险挑战，中国从中华优秀传统文化中汲取丰富的思想智慧，将“天下为公”“民惟邦本”等核心理念融入AI的治理中：“天下为公”的理念体现出中国对人类未来的高度关注，不断探索AI发展和治理的有效路径，以确保其发展始终符合人类文明的进步方向；而“民惟邦本”的思想则强调了以人民为中心的发展思想，旨在借助AI把人类从危险且复杂的劳动中解放出来，并为人们创造更加美好的生活方式。将中国传统文化精髓应用于全球AI治理，不仅充分展示了中华文化的独特魅力，也为国际社会提供了不同于西方AI治理模式的新思路及价值导向；不仅对推动中华文化走向国际社会提供了重要的价值指引，也为弘扬中华民族现代文明增添了强劲的动力。

（二）“倡议”丰富了总体国家安全观和新发展理念的内涵

西方传统安全观长期奉行恃强凌弱、以邻为壑、自私自利的陈旧观念，推行恃强凌弱、巧取豪夺、零和博弈等霸权霸道霸凌行径，其本质上是为确保西方国家自身利益最大化。在西方传统安全观的主导下，美国在AI交流合作方面以价值观和意识形态划线，拉拢盟友和伙伴构建排他性集团，以期借助AI谋求绝对军事优势和霸权地位，并恶意阻挠竞争对手的AI发展。基于此，“倡议”从治理理念、治理路径、治理主体、治理原则、治理目标五大方面为全球AI治理提供了中国方案，同时突破了西方传统安全观的局限性，具有鲜明的科学性、人民性、时代性、实践性、世界性，并深化和拓展了总体国家安全观的理论内涵和实践领域。

[1]《习近平：在文化遗产发展座谈会上的讲话》，中国政府网，2023年8月31日，https://www.gov.cn/yaowen/liebiao/202308/content_6901250.htm。

中国还根据世界之变、科技之变，与时俱进提出了新发展理念。新发展理念的核心是坚持以人民为中心的发展思想。为此，“倡议”创造性地提出了应通过人机融合的方式来维护人的利益、保护人的安全，并在全球 AI 治理的过程中，积极推进人机融合，确保 AI 的开发和应用始终尊重、维护和服务于人类利益和福祉。本质上，人机融合的根本出发点在于维护人的利益与需要，彰显人民至上的 AI 价值取向，最终把维护好最广大人民群众的利益作为发展 AI 的终极价值理念。

（三）“倡议”是推动落实“三大倡议”的具体成果，加速人类命运共同体不断落地生根

党的十八大以来，习近平主席创造性地提出了构建人类命运共同体这一重要理念，为建设更加美好世界贡献了中国方案。十多年来，在习近平主席的亲自擘画和推动下，这一理念已从理论走向实践，从愿景成为现实，其思想内涵不断得到深化拓展，理论体系日臻完善。中国的“三大倡议”——全球发展倡议、全球安全倡议和全球文明倡议分别从发展、安全和文明三个维度，系统性地回应了一系列关于人类前途命运的重大问题，如“实现什么样的全球发展、怎样实现全球发展”，“世界需要什么样的安全理念、各国怎样实现共同安全”，“人类文明向何处去、不同文明怎样交往”，为构建人类命运共同体夯基垒台、立柱架梁，提供了坚实的理论基础。而“倡议”则是中国在 AI 领域推动落实“三大倡议”的具体行动方案。在发展层面，“倡议”指出了具体的实践路径——强调 AI 的发展应以人为中心，推动相关技术的创新和应用，以促进社会和经济的进步。在安全层面，“倡议”提供了具体的行动方案——应建立和完善法律制度，保护个人隐私和数据安全，并形成广泛认可的 AI 治理框架和标准。在文明层面，“倡议”指出了合作的基本原则——各国应在 AI 治理中加强信息交流和技术合作，共同做好风险防范；发展 AI 应符合和平、发展、公平、正义、民主、自由的全人类共同价值观等。

自“倡议”提出以来，国际社会对此高度关注，并引发热烈反响和高度赞誉。俄罗斯、塞尔维亚、乌兹别克斯坦、巴基斯坦等国政府欢迎中方提

出的“倡议”，并表示愿就 AI 的发展、安全和治理加强交流合作。法国巴黎和平论坛总干事贾斯汀·瓦伊斯（Justin Vaïsse）表示，“倡议”积极支持在联合国框架下讨论成立国际 AI 治理机构，法中两国可以在 AI 治理领域充分合作，共同应对全球性挑战。^[1]塞尔维亚外交部部长助理高度评价“倡议”在中塞关系发展和世界繁荣稳定中发挥的重要作用，表示全力支持中方提出的“倡议”和合作机制，以联合应对各类新的全球性挑战。^[2]巴西中国问题研究中心主任罗尼·林斯（Ronnie Lins）指出，“倡议”的提出会让所有国家都能获得 AI 创造的价值和利益，无论其经济实力和地缘政治地位如何。^[3]在国际组织方面，红十字国际委员会主席米里亚娜·斯波利亚里茨·埃格（Mirjana Spoljaric Egger）、红十字国际委员会高级科学与政策顾问尼尔·戴维森（Neil Davison）高度赞赏中国提出的“倡议”，认为中国在 AI 发展和治理中所遵循的“以人为本”和“智能向善”理念与红十字国际委员会对武装冲突中的 AI 所采取的“以人为本的方法”是一致的；他希望中国能够努力争取地区和国际社会的支持与合作，为 AI 军事化应用划定明确的法律红线。这不仅有助于保护平民免受 AI 武器化可能带来的伤害，而且也是对未来世代的一种责任和保护。^[4]阿拉伯国家联盟欢迎习近平主席提出的“倡议”，赞赏中方为增强发展中国家在全球 AI 治理中的权利所作的努力，愿与中方合作，加强政策协调，建立对话机制，深化务实合作，形成优势互补。^[5]需要指出的是，《倡议》对后来欧盟推进和批准《人工智能法案》、两届全球

[1] 邢雪、杨一：《深化全球治理改革 共建多边主义未来——第二届中法全球治理论坛综述》，载《人民日报》2024年5月4日，第03版。

[2] 《驻塞尔维亚大使李明出席塞尔维亚国际政治与经济研究所“中国新倡议：应对变化中的全球安全”研讨会开幕式》，中华人民共和国驻塞尔维亚共和国大使馆网站，2023年11月9日，http://rs.china-embassy.gov.cn/sgxx/sghd/202311/t20231127_11187953.htm。

[3] 《巴西专家：中国提出的〈全球人工智能治理倡议〉有助防范“技术贫民窟”》，新华网，2023年11月1日，http://www.news.cn/world/2023-11/01/c_1129952180.htm。

[4] Neil Davison, “A humanitarian perspective on military AI,” ICRC, October 31, 2023, <https://www.icrc.org/en/document/humanitarian-perspective-military-ai>。

[5] 《中阿合作论坛第十届部长级会议通过〈北京宣言〉和〈行动计划〉》，载《人民日报》2024年6月1日，第06版。

AI 安全峰会通过相关包容合作性宣言，以及各国相继推出旨在加强 AI 安全的政策举措和准则规范，客观上起到了示范影响的作用。由此可见，“倡议”作为中国为国际社会贡献的重要公共产品，正日益获得国际社会的广泛认可和积极响应，显示出其在推动全球 AI 治理方面的深远影响和实践价值。

结 语

习近平主席深度把握世界 AI 发展趋势和治理特点，战略性地提出“倡议”，充分彰显了中国作为负责任大国的责任担当。

“倡议”构建起以人为本、尊重生命、公平公正、共担责任、包容共享、安全可控、可持续发展为基础的 AI 伦理观，旨在加强对 AI 研发和实际应用中的约束和治理，从而更好地实现人与技术间的安全良性互动，是中国为破解全球 AI 治理难题贡献的中国智慧和方案。

“倡议”的落实需要世界各国人民携手努力，以共同促进全球 AI 健康有序安全发展。各方需要共同秉持共商共建共享理念，坚持发挥联合国的斡旋主渠道作用，协同推进全球 AI 治理，确保 AI 的发展和治理必须以增进人类福祉为主要目标。只有如此，AI 的发展红利才能更好地惠及各国人民，人类才能在这样的共同努力下创造和平、发展、合作、共赢的美好未来。

【收稿日期：2024-05-28】

【修回日期：2024-10-17】

（责任编辑：邢嫣）