

全球人工智能治理倡议

人工智能是人类发展新领域。当前，全球人工智能技术快速发展，对经济社会发展和人类文明进步产生深远影响，给世界带来巨大机遇。与此同时，人工智能技术也带来难以预知的各种风险和复杂挑战。人工智能治理攸关全人类命运，是世界各国面临的共同课题。

在世界和平与发展面临多元挑战的背景下，各国应秉持共同、综合、合作、可持续的安全观，坚持发展和安全并重的原则，通过对话与合作凝聚共识，构建开放、公正、有效的治理机制，促进人工智能技术造福于人类，推动构建人类命运共同体。

我们重申，各国应在人工智能治理中加强信息交流和技术合作，共同做好风险防范，形成具有广泛共识的人工智能治理框架和标准规范，不断提升人工智能技术的安全性、可靠性、可控性、公平性。我们欢迎各国政府、国际组织、企业、科研院校、民间机构和公民个人等各主体秉持共商共建共享的理念，协力共同促进人工智能治理。

为此，我们倡议：

——发展人工智能应坚持“以人为本”理念，以增进人类共同福祉为目标，以保障社会安全、尊重人类权益为前提，

确保人工智能始终朝着有利于人类文明进步的方向发展。积极支持以人工智能助力可持续发展，应对气候变化、生物多样性保护等全球性挑战。

——面向他国提供人工智能产品和服务时，应尊重他国主权，严格遵守他国法律，接受他国法律管辖。反对利用人工智能技术优势操纵舆论、传播虚假信息，干涉他国内政、社会制度及社会秩序，危害他国主权。

——发展人工智能应坚持“智能向善”的宗旨，遵守适用的国际法，符合和平、发展、公平、正义、民主、自由的全人类共同价值，共同防范和打击恐怖主义、极端势力和跨国有组织犯罪集团对人工智能技术的恶用滥用。各国尤其是大国对在军事领域研发和使用人工智能技术应该采取慎重负责的态度。

——发展人工智能应坚持相互尊重、平等互利的原则，各国无论大小、强弱，无论社会制度如何，都有平等发展和利用人工智能的权利。鼓励全球共同推动人工智能健康发展，共享人工智能知识成果，开源人工智能技术。反对以意识形态划线或构建排他性集团，恶意阻挠他国人工智能发展。反对利用技术垄断和单边强制措施制造发展壁垒，恶意阻断全球人工智能供应链。

——推动建立风险等级测试评估体系，实施敏捷治理，分类分级管理，快速有效响应。研发主体不断提高人工智能

可解释性和可预测性，提升数据真实性和准确性，确保人工智能始终处于人类控制之下，打造可审核、可监督、可追溯、可信赖的人工智能技术。

——逐步建立健全法律和规章制度，保障人工智能研发和应用中的个人隐私与数据安全，反对窃取、篡改、泄露和其他非法收集利用个人信息的行为。

——坚持公平性和非歧视性原则，避免在数据获取、算法设计、技术开发、产品研发与应用过程中，产生针对不同或特定民族、信仰、国别、性别等偏见和歧视。

——坚持伦理先行，建立并完善人工智能伦理准则、规范及问责机制，形成人工智能伦理指南，建立科技伦理审查和监管制度，明确人工智能相关主体的责任和权力边界，充分尊重并保障各群体合法权益，及时回应国内和国际相关伦理关切。

——坚持广泛参与、协商一致、循序渐进的原则，密切跟踪技术发展形势，开展风险评估和政策沟通，分享最佳操作实践。在此基础上，通过对话与合作，在充分尊重各国政策和实践差异性基础上，推动多利益攸关方积极参与，在国际人工智能治理领域形成广泛共识。

——积极发展用于人工智能治理的相关技术开发与应用，支持以人工智能技术防范人工智能风险，提高人工智能治理的技术能力。

——增强发展中国家在人工智能全球治理中的代表性和发言权，确保各国人工智能发展与治理的权利平等、机会平等、规则平等，开展面向发展中国家的国际合作与援助，不断弥合智能鸿沟和治理能力差距。积极支持在联合国框架下讨论成立国际人工智能治理机构，协调国际人工智能发展、安全与治理重大问题。