

人工智能治理规则的国际协同与竞争

张 平

2023 年，中国集中发布了若干涉及人工智能的重要文件：《国务院 2023 年度立法工作计划》明确将《人工智能法》纳入计划；7 月 13 日，国家网信办等七部门联合发布《生成式人工智能服务管理暂行办法》；9 月 7 日，科技部等十个部委联合发布《科技伦理审查办法（试行）》；10 月 18 日，外交部发布《全球人工智能治理倡议》。

11 月 1 日，在全球首届人工智能安全峰会上，包括中国在内的 28 个国家及欧盟共同签署了《布莱切利宣言》，这一系列文件的发布，标志我国在人工智能立法与政策制定上已经表明了中国立场并且积极参与全球人工智能发展及治理的规则制定。

各国AI政策的特色和异同

今年 6 月 14 日，欧盟通过《人工智能法案》草案，立法目的强调欧盟的统一性，确保人工智能商品和服务跨境自由流动，并提供一致的规则保护。

10 月 20 日，我国外交部发布《全球人工智能治理倡议》，旨在推动全球人工智能的健康发展，构建公正、合理、透明的全球人工智能治理体系。

11 月 1 日，美国政府在欧盟、中国和英国已经开始或完善相关立法工作的紧迫背景下，发布了《关于人工智能的新行政命令》，其立法目的主要是将美国定位为人工智能政策的全球领导者。

多国在英国共同发布的《布莱切利宣言》，则为首届全球人工智能峰会凝聚的共识，其目标为加强国际合作，为 AI 伦理建立统一的标准，并减少主要经济体之间关于 AI 产业的“军备竞赛”。上述四份 AI 政策均在人工智能技术迅猛发展，人工智能产业对经济社会影响逐步凸显的背景下诞生，其共同目的在于，促进人工智能产业进步的同时，应对随之带来的风险和问题。

首先，四项政策文件均含有针对 AI 风险的应对措施。中国和欧盟都提出了“分类分级”的监管模式，但欧盟的立法颗粒度更细。欧盟《人工智能法案》以风险为进路，将 AI 系统分为不可接受的风险、高风险、有限风险以及极低风险四类，并针对每一类 AI 系统采取不同的监管策略，而中国尚未对如何进行人工智能服务或技术分类分级提供完整的规定，相关立法工作仍在进行中。美国则通过“红队安全测试”方式确保人工智能系统的安全、可靠。《布莱切利宣言》作为国际共识，仅要求根据不同国家的情况，建立风险等级测试评估体系，并加强国际合作。

其次，上述 AI 政策内容侧重点各不相同。美国《关于人工智能的新行政命令》具有涵盖范围广、多机构联动的特点，涵盖安全和保障、隐

- 各国 AI 政策均在人工智能技术迅猛发展，人工智能产业对经济社会影响逐步凸显的背景下诞生，其共同目的在于，促进人工智能产业进步的同时，应对随之带来的风险和问题。
- 各国在 AI 监管框架里都强调了科技伦理，在治理原则上也体现了以人为本，加强个人基本权利保护。同时，加强多主体协同治理，形成服务提供者、使用者、监管者和公众等多方合力仍是各国当下治理 AI 的重要原则。
- 在全球层面，一方面，各国将人工智能产业置于国际竞争的战略地位，政策中也体现了引领人工智能领域国际话语权的意图；另一方面，人工智能风险的全球性成为共识，加强国际对话与合作已经成为大势所趋。
- 未来，加强 AI 立法工作、解决不同机构职能协调问题、促进政府与企业的合作，并更全面地考虑科技伦理问题，将会是各国对 AI 领域的治理重点。

私、公民权利、消费者和劳工保护、创新和竞争以及国家安全等原则，特别关注如消费者、学生、劳动者等各主体的利益，加强人工智能人才培养和引进，强调美国在人工智能领域的全球领导力。中国的《全球人工智能治理倡议》则从全球性视角出发，关注发展中国家的利益，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，防止“数字鸿沟”“智能滥用”，强调“人类权益”的保障。多国在英国发布的《布莱切利宣言》重点关注国际合作，通过列举持续对话、跨国界的科研网络支持等具体措施使得国际合作可操作化。此外，中方《全球人工智能治理倡议》多条内容被纳入该“宣言”，部分国外学者评论该峰会最大的成就“可能是将中国纳入讨论”。

从政策地位和影响力来看，欧盟《人工智能法案》是全球首部全面的人工智能法案草案，对于促进世界各国人工智能立法，并提升欧盟的影响力具有重要意义；美国《关于人工智能的新行政命令》是由美国总统签署的官方指令，性质上属于行政命令并非立法，其虽对国会立法具有影响，但学者普遍批评行政令强制力不足；英国《布莱切利宣言》达成了一部分共识，但由于各国对人工智能治理方法不同，且存在着国家利益竞争，美国《财富》杂志批评该“宣言”“缺乏坚定的承诺”。

AI治理的核心与治理原则

AI 技术应用的伦理规范始终是核心和棘手的问题。2023 年 3 月 22 日，美国生命未来研究所（Future of Life）发布了一封《暂停大型人工智能研究》的公开信，呼吁所有人工智能实验室立即暂停比 GPT-4 更强大的人工智能系统的训练，暂停时间至少为 6 个月，以呼吁在此期间各国能共同制定全球范围内相关的技术共享安全协议。此封信件中提到的伦理问题有：

- 1.人类获取信息渠道的来源是否应被机器用“宣传和谎言”主宰；
- 2.所有工作是否都应该被自动化

技术取代；

3.是否应该开发非人类大脑，使其最终超越人类的智慧，淘汰并取代人类的非人类思维；

4.是否应该冒着失去对人类文明控制的风险去开发 AI。这封公开信形式上看是针对 GPT 的开发研究按了“暂停键”，但所提出的伦理问题则是人类对于 AI 技术发展和应用的一种普遍性担忧。由此，才会使得各国在 AI 监管框架里都强调了科技伦理，在治理原则上也有了下述体现：

以人为本，加强个人基本权利保护。在欧盟《人工智能法案》中，将对基本权利的侵害程度进行 AI 系统风险等级划分的考量因素；美国《关于人工智能的新行政命令》则明确强调了对隐私、公民权利、劳动者和消费者权利的保护，并提供具体方案措施；中国《全球人工智能治理倡议》提出，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；《布莱切利宣言》即纳入了中国倡议的“以人为本”理念。

除了上述行动与共识，加强多主体协同治理，形成服务提供者、使用者、监管者和公众等多方合力仍是各国当下治理 AI 的重要原则。美国《关于人工智能的新行政命令》推动多个联邦机构对影响所有行业的 AI 风险进行监管；中国《生成式人工智能服务管理暂行办法》不仅规定了服务提供者的责任和义务，还建立健全投诉、举报机制，为公众参与监督提供了依据。

在全球层面，一方面，各国将人工智能产业置于国际竞争的战略地位，政策中也体现了引领人工智能领域国际话语权的意图，如美国《关于人工智能的新行政命令》就强调了其在该领域的全球领导力。

另一方面，人工智能风险的全球性成为共识，加强国际对话与合作已经成为大势所趋。从 2023 年多国专家签署人工智能潜在风险的声明，到维护国际论坛等对话机制的畅通和科研界的充分合作，再到《布莱切利宣言》的达成，均展现了加强全球协同治理，共同面对人工智能风险的努力。

全球AI治理的挑战与展望

四份政策文件标志着在 AI 治理方面国际合作的积极开端，也体现了各国对于 AI 发展过程中风险与安全的重视。但也应当认识到，上述政策也各有不足之处，并且在政策背后还蕴含治理理念竞争和国家间的博弈，更凸显了实现统一全球 AI 治理方案的复杂性和艰巨性。

例如，欧盟《人工智能法案》在确定人工智能系统责任方面存在模糊之处，缺乏明确的法律行动权利规定，使得对于不遵守法案的 AI 系统提供者或用户的惩罚力度不够清晰。并且，该法案也未明确人工智能是否具备独立法律人格，以及在责任问题上是否能够被追究，这使得在人工智能系统出现问题时难以明确追责对象。

美国《关于人工智能的新行政命令》则因其行政令的本质而天然具有不稳定性，需要尽快进行相关领域的立法工作。行政命令的局限性也使得其在解决问题和规范行为方面受到制约。此外，在行政命令中对于人工智能与知识产权问题的关注程度有限，未来需要更全面地解决人工智能生成内容中的虚假信息问题以及知识产权问题。

而《布莱切利宣言》则缺乏创建有效监管体系的细节。外界也表现出对现有 AI 安全研究机构的担忧，认为其可能无法迅速适应技术的快速发展，还需要加强国际合作并在技术和法规上创建更具约束力的措施。

中国《全球人工智能治理倡议》以倡议的形式提出，与法律文件相比强制拘束力较弱，但是却表达了倡导者的立场和努力方向。“倡议”提到构建开放、公正、有效的治理机制，需要未来制定具体规则来对治理结构和运作方式予以细化和落实。总体而言，四份 AI 监管政策文件都在不同程度上存在责任机制不完善、法律人格不确定、平衡创新与消费者保护的挑战等问题。

未来，加强 AI 立法工作、解决不同机构职能协调问题、促进政府与企业的合作，并更全面地考虑科技伦理问题，将会是各国对 AI 领域的治理重点。我国应当坚持伦理先行，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关切，为全球人工智能治理贡献中国智慧和力量。

（作者系北京大学法学院教授，北京大学人工智能研究院双聘教授，北京大学武汉人工智能研究院特聘专家，中国科学技术法学会常务副会长兼秘书长，中国法学会知识产权法学研究会副会长）



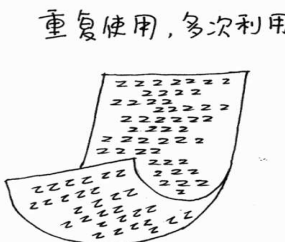
扫描左侧二维码关注

遗失声明

上海煜胜仓储服务部，遗失营业执照正、副本，注册号：310230000549074，遗失税务登记证正、副本，国地税沪字310230055906387，声明作废，特此公告。

上海市徐汇区春林林五金建材经营部，遗失营业执照正、副本，统一社会信用代码：92310104MA1KBDDN3Y，正本证照编号：040008201712070001，副本证照编号：040008201712070002，声明作废。

环保公益广告



双面使用纸张
= 减少 1/2 的废纸产生

