Source: *Lianhe Zaobao,* p18
Date: 6 October 2023

科技新视野 庞严

AI监管全球展望及中国的启示

人工智能（AI）技术疾速发展，不仅在各行各业引发深远的革命性变革，同时暴露出一系列崭新的安全与监管难题。首先，在数据安全与隐私层面，AI模型在训练和运行过程中，常常须要大量个人敏感数据作为支持，大大增加了数据泄露和隐私侵犯的风险。例如，在面部识别、语音识别以及数据挖掘等应用领域，数据安全与隐私保护问题尤为凸显。

随着AI技术在创作内容、自动驾驶、医疗诊断等领域广泛应用，法律与合规问题也变得日益重要。当前，全球范围内尚缺乏统一的AI监管框架，这使得跨国公司和研究机构面临复杂的合规问题。

其次，AI模型的伦理与社会偏见问题也日益引发关注。如果训练数据含有偏见，AI模型很可能会继承并放大这些偏见，从而在应用过程中产生不公平甚至具有歧视性的决策。此外，许多先进的AI模型，如深度学习模型，作为“黑箱”模型，缺乏决策过程的透明度。这在医疗、司法和金融等关键领域，可能引发严重后果。

另外，在安全方面，新型的对抗攻击手段已经出现。这些攻击手段能通过微小的输入变动，来误导AI模型，产生错误或具有潜在危险的输出。这不仅对个人用户构成威胁，还可能在社会层面影响整体安全。随着AI技术在创作内容、自动驾驶、医疗诊断等领域广泛应用，法律与合规问题也变得日益重要。当前，全球范围内尚缺乏统一的AI监管框架，这使得跨国公司和研究机构面临复杂的合规问题。

各国已逐步开始制定针对AI技术监管的相关法规条例。例如，在美国，AI监管关注的重点主要包括数据安全、隐私和伦理问题。国家标准与技术研究院（NIST）于2023年1月发布了《人工智能风险管理框架》（AI RMF 1.0），旨在指导机构组织在开发和部署AI系统时，降低安全风险，避免产生偏见和其他负面后果，提高AI可信度。

欧盟则更加注重AI的伦理和隐私保护，通用数据保护条例（GDPR）就是其中一个典型示例。欧盟还发布了一系列AI伦理准则，并计划在未来几年推出一个全面的AI监管框架。

在亚洲，日本和韩国更加关注技术标准和伦理准则，新加坡则试图通过灵活的监管措施，来吸引国际AI企业。

作为AI技术发展最快的国家之一，中国对AI安全监管给予高度关注。过去这一年多，中国陆续公布了一系列针对AI的监管措施。

2022年3月1日，中国推出首部全国的AI专门法规——《互联网信息服务算法推荐管理规定》，规范了在中国境内提供网络服务时，使用算法推荐技术的行为。2022年11月25日，中国发布《互联网信息服务深度合成管理规定》提出，深度合成服务提供者和技术支持者，在提供人脸、人声等生物识别信息编辑功能时，应当提示深度合成服务使用者，依法告知被编辑的个人，并取得同意。

最值得关注的是，今年7月13日，中国国家网信办等七部门联合公布《生成式人工智能服务管理暂行办法》（以下简称《办法》），于8月15日正式生效，旨在规范更广泛的生成式AI技术。这是中国首次对此作出明确规定。《办法》对生成式AI服务持积极态度，多处提到要鼓励这项技术在各行业、各领域的创新应用，还鼓励生成式AI算法、框架、晶片及配套软件平台等基础技术的自主创新，平等互利开展国际交流与合作，参与生成式AI相关国际规则制定。

其中，有三个方面的内容尤其值得关注。第一，它强调了AIGC（生成式人工智能）的分类分级管理机制，突出针对不同风险类型的后续监管机制的重要思路，对不同类型的AIGC应用采取更有针对性的监管措施。

其次，管理办法注重AIGC产业生态的培育，特别是关于生成式AI基础设施和公共训练数据资源平台的建设问题。它强调促进算力资源的协同共享，提升算力资源的利用效能。通过推动公共数据的分类分级有序开放，扩展高质量的公共训练数据资源，鼓励采用安全可信的晶片、软件、工具和算力资源等措施，为AIGC产业的发展提供更稳定和可靠的基础设施支持。

第三，该管理办法强调国内外的交流合作，在适用范围上针对境外、境内提供服务，以及不向境内提供服务的业务类型做区分，明确适用范围。对于外商投资的，应当符合相关法律、行政法规的规定。通过加强国际合作和交流，更好地引进境外先进技术和服务，也可以更好地推动中国AIGC产业走向国际市场。

中国强调防止未成年人上瘾

此外，《办法》着重要求采取有效措施，防止未成年人用户过度依赖或沉迷其中。在监督生成式AI方面，《办法》还提到有关主管部门可以依据职责，对这方面服务进行监督检查，服务提供者应予以配合，并提供必要的技术、数据等支持和协助。《办法》还涉及保护个人隐私、商业秘密等多个方面，例如参与生成式AI服务安全评估和监督检查的相关机构和人员，应当对履行职责过程中知悉的国家秘密、商业秘密、个人隐私和个人信息加以保密，不得泄露或非法向他人提供等。

可以看到，《办法》重视防范风险的同时，也体现一定的容错和纠错机制，提升了执行可行性，更好地实现规范与发展的动态平衡。这为其他国家制定AI相关的安全及监管条例，提供了很好的借鉴意义。

大多数国家未建立全面监管框架

总体而言，在AI领域的前沿阵地，诸如北美、欧洲以及亚洲的中国、日本、韩国和新加坡等国家和地区，AI监管正在快速发展和完善。然而，对许多发展中国家来说，AI监管尚处于初级阶段，主要挑战是如何在技术创新与伦理、安全之间取得平衡。一些国家已开始制定基本的AI战略和政策，但大多数国家尚未建立全面的监管框架。

除了各国的内部监管之外，国际组织如联合国、世界经济论坛和经济合作与发展组织也须积极推动全球AI监管的合作。这些组织主要关注AI对全球经济、社会和安全的影响，并致力于构建一个公平、透明和可持续的全球AI生态系统。展望未来，全球的AI监管法规将会不断发展和完善，以适应快速发展和全球化应用。同时，在监管过程中，须要平衡技术创新、隐私保护、伦理和社会责任等多个方面的问题，保障AI技术的可持续发展和应用。

作者是新加坡国立大学商学院教授
商业大数据分析中心联席主任
原载《联合早报》旗下英文电子杂志
“思想中国”（ThinkChina）