

瀚蓝环境股份有限公司

负责任人工智能承诺

一、目的

作为国内综合环境服务商，瀚蓝环境股份有限公司积极探索人工智能技术和应用，打造瀚蓝基座大模型，构建系列业务专有大模型，加速数字化系统嵌入智能迭代升级，促进业务场景的智能化转化升级。

公司对人工智能技术研发、部署、应用等坚持合规可控、以人为本、安全优先、技术向善的原则，特此发布负责任人工智能承诺，指导全集团的人工智能管理实践。

二、适用范围

本文件适用于瀚蓝环境股份有限公司及所属全资与控股的子公司。

三、管理架构

公司经营管理层作为负责任人工智能管理的最高决策和统筹管理机构。

四、总体要求

1、恪守数据隐私保护义务

严格遵守国家数据保护、个人信息保护相关法律法规，在人工智能开发、训练及应用运行的全生命周期，保障员工、

客户、合作伙伴及其他利益相关方的数据隐私与信息安全。建立完善的数据访问管控、脱敏防护机制，确保输入人工智能系统的全部数据来源合法、授权充分，严防敏感及涉密数据违规流入外部人工智能工具。

2、筑牢人工智能网络安全防线

公司对投入使用的人工智能及智能应用系统实施严格且全面的网络安全管理。上线前开展安全风险评估，运行期间建立常态化安全巡检、漏洞排查与应急处置机制，主动防范网络攻击、非法入侵、数据篡改、模型异常等各类安全风险。落实人工智能系统安全防护责任，严防 AI 应用风险传导至公司核心运营业务，保障各项经营活动安全、稳定、可控。

3、防范算法偏见，维护算法公平

公司在人工智能应用全过程高度重视算法公平性管理。针对数据分析、智能研判、辅助决策等场景，致力于最大限度降低算法偏见风险，严格防范人工智能输出带有歧视性、倾向性的结论，保障各项业务决策与服务结果公平、公正、无差别对待所有利益相关方。

4、落实人在回路，坚持人工决策监督

现阶段，公司明确将人工智能定位为业务辅助分析工具，不替代人的最终决策。对于具备较高影响风险的业务场景，严格落实“人在回路”的管理原则，人工智能不得独立输出最终决策。对 AI 形成的重要输出结果执行人工复核程序，保留人员对 AI 结论进行干预、驳回与调整的完整权限，保障决策过程可控可监督。

5、明晰应用边界，保障 AI 应用的透明度与可解释性

公司注重人工智能应用的透明度与可解释性管理。对内及对外明确人工智能工具的应用场景、能力边界与数据来源；必要时，在合规及技术可行范围内，提供易于理解的分析逻辑与解释依据。同时，当用户与人工智能而非人类交互时，公司确保用户能够清楚了解其正在人工智能进行交互。

6、建设人工智能管理责任机制

公司致力于建立完善的人工智能管理体系，明确各业务单元人工智能应用对应的管理责任主体。持续完善针对 AI 错误输出、不良后果的调查、复盘及整改等工作机制，实现人工智能相关风险可追溯、责任可界定。

7、践行绿色低碳，管控 AI 算力环境影响

优先选用节能环保的算力资源，发挥人工智能对绿色运营的赋能作用，管控 AI 应用自身的能源消耗，减少算力带来的生态环境影响。

8、抵制高风险人工智能应用

公司承诺不开发、采购及部署具备操纵行为、利用人性弱点、社会评分、未经授权生物识别监控等高风险特征的人工智能系统。持续跟踪国内外人工智能监管治理相关要求，严格禁止各类高风险人工智能工具在公司内部落地应用。

瀚蓝环境将持续推动人工智能技术与生态环境治理业务深度融合，在坚守安全合规前提下，以数智化技术赋能业务高质量发展，为社会提供安全可靠、绿色高效的环境服务。