

迪安诊断负责任人工智能管理规定

第一章 总则

（一）编制目的

为规范迪安诊断技术集团股份有限公司人工智能相关技术、模型、工具和应用场景的开发、采购、部署、使用和管理，促进人工智能技术安全、合规、可靠、透明地服务医学诊断、健康管理、科研创新和经营管理，防范数据安全、隐私保护、算法偏见、网络安全、责任归属和社会影响等风险，制定本管理规定。

（二）适用范围

本管理规定适用于公司及所属机构在业务运营、产品服务、科研创新、质量管理、客户服务、内部管理及与外部合作伙伴协同过程中涉及人工智能系统的相关活动。人工智能系统包括但不限于算法模型、智能识别、智能审核、智能生成、辅助诊断、智能问答、预测分析、流程自动化和其他以数据、模型或算法驱动的工具与平台。

（三）基本原则

公司开展人工智能相关活动应坚持依法合规、以人为本、安全可控、公平审慎、透明可解释、责任明确和持续改进的原则。人工智能应作为提升医学诊断服务质量、运营效率和科研创新能力的工具，不得替代依法应由专业人员、管理人员或授权主体作出的关键判断和责任承担。

第二章 治理职责

（一）统筹管理

公司应结合业务特点和风险水平，建立人工智能治理协调机制，统筹人工智能应用规划、风险识别、制度建设、合规审查、监督评估和持续改进。涉及重大业务影响、重要数据处理、关键决策支持或外部服务输出的人工智能应用，应按照公司内部授权和风险管理要求履行相应审查程序。

（二）业务部门责任

人工智能应用所属业务部门是场景管理和使用合规的直接责任主体，应明确应用目的、使用范围、数据来源、预期输出、人工复核要求、风险控制措施和异常处置路径，确保人工智能工具服务于真实、合理、合法的业务需求。

（三）技术与数据责任

信息技术、数据管理、信息安全及相关专业部门应根据职责分工，对人工智能系统的数据处理、模型开发、系统接入、权限控制、网络安全、运行监测和技术变更进行管理，确保系统具备必要的安全性、稳定性和可追溯性。

（四）第三方协同责任

公司采购、接入或联合开发第三方人工智能产品、模型、平台或服务时，应关注供应商资质、数据保护能力、网络安全措施、模型能力边界、知识产权安排、服务连续性和责任分担，并在合同或合作文件中明确保密、数据安全、合规使用和风险处置要求。

第三章 使用边界

（一）授权使用范围

人工智能系统应在经确认的业务场景、数据范围、权限范围和技术能力边界内使用。任何部门和人员不得擅自扩大人工智能系统的使用目的、处理范围或输出用途，不得将测试性、试点性或辅助性工具直接用于未经审查的关键业务流程。

（二）禁止性要求

公司不得开发、部署或使用以操纵行为、利用个人脆弱性、实施社会评分、未经授权开展生物特征监控或其他违反法律法规、公序良俗和医学伦理要求为目的的人工智能系统。人工智能系统不得用于规避监管要求、弱化专业责任、侵犯个人合法权益或产生不当歧视性后果。

（三）医疗健康场景边界

在医学诊断、健康管理、科研分析等场景中，人工智能系统应定位为辅助工具。涉及诊断建议、风险提示、报告生成、治疗相关判断或可能对个人健康权益产生重要影响的输出，应设置必要的专业复核、人工确认、异常提示和升级处理机制。

第四章 数据隐私与信息安全

（一）数据合规使用

人工智能相关数据处理活动应遵守适用的数据安全、个人信息保护、网络安全、医疗健康数据管理及行业监管要求。公司应按照合法、正当、必要和最小化原则收集、使用、存储、传输、共享和销毁数据，不得超出授权范围处理个人信息、客户数据、医疗健康数据和其他敏感信息。

（二）数据分类分级

涉及人工智能训练、验证、测试、运行和评估的数据，应按照公司数据分类分级要求实施差异化管理。涉及个人隐私、检验结果、生物识别信息、医疗健康信息、科研数据和商业敏感信息的，应采取更严格的权限控制、脱敏处理、加密保护和审批措施。

（三）网络安全防护

人工智能系统在接入、部署、运行和对外开放过程中，应落实身份认证、访问控制、日志留存、漏洞管理、加密传输、接口安全、异常监测和应急响应等安全措施。对关键系统和重要模型，应结合风险水平开展必要的安全测试、漏洞扫描、攻防演练或其他安全验证。

第五章 公平性与偏见防控

（一）数据质量管理

人工智能系统使用的数据应尽可能真实、准确、完整、相关和可追溯。对于可能影响模型输出公平性和准确性的训练数据、样本结构、标注规则和数据来源，相关部门应进行必要的质量审核和风险评估。

（二）偏见识别与缓解

公司应关注人工智能系统在不同人群、场景、区域、疾病类型、业务对象和服务对象中的适用性差异，识别并防范因数据偏差、模型设计、样本不足或使用不当导致的不公平、不合理或歧视性结果。发现可能产生明显偏差或不当影响的，应及时采取调整、限制、暂停或人工复核等措施。

（三）审慎应用原则

对于结果可能影响个人健康、服务机会、客户权益、员工权益或其他重要利益的人工智能应用，应提高审查和复核要求，避免仅依赖自动化结果作出高影响决策。

第六章 人类监督与可解释性

（一）人类监督

公司应根据人工智能应用的风险等级和业务影响，设置适当的人类监督机制。在关键决策、高风险场景或不可逆影响场景中，应保留人工审核、人工干预、人工否决或人工升级处理权限，确保人工智能系统不能在缺乏有效监督的情况下独立作出高风险决定。

（二）透明度要求

公司应在适当范围内明确人工智能系统的使用目的、主要能力、适用边界、数据来源、输出形式、局限性和潜在风险。面向客户、员工或其他受影响方提供人工智能生成内容、智能问答、辅助判断或自动化服务时，应根据场景需要进行清晰提示。

（三）可解释性要求

人工智能系统输出应尽可能具备可理解、可追溯和可复核的基础。对于用于专业判断、业务审批、客户服务或健康管理的重要输出，相关人员应能够了解其主要依据、适用限制和复核路径，避免无法解释的输出直接进入关键决策。

第七章 责任追究与事件处置

（一）责任归属

公司应按照“谁管理、谁负责，谁使用、谁负责”的原则，明确人工智能系统规划、开发、采购、部署、使用、维护、监督和整改过程中的责任主体。人工智能工具产生的输出不得免除相关人员、业务部门和管理主体依法依规应承担的审慎审查和管理责任。

（二）记录留痕

对重要人工智能系统和高风险应用场景，公司应保留必要的模型版本、数据范围、权限配置、使用记录、输出记录、复核记录、异常处理和变更记录，以支持问题追溯、责任认定、合规审查和持续优化。

（三）事件处置

如人工智能系统发生数据泄露、网络攻击、输出错误、偏见歧视、误用滥用、侵权争议、重大服务异常或其他可能影响客户、员工、合作伙伴及公众权益的事件，相关部门应及时报告、评估影响、采取控制措施、开展原因分析并推动整改。

第八章 生态影响与持续改进

（一）资源效率

公司在人工智能模型训练、系统部署、算力采购和第三方服务选择过程中，应关注能源消耗、算力效率、设备利用、数据存储和资源浪费等因素，鼓励采用高效、节能、可扩展的技术方案，降低人工智能应用的生态足迹。

（二）培训宣贯

公司应根据人工智能应用发展和岗位职责需要，开展人工智能合规、安全、伦理、数据保护和使用边界相关培训，提升员工对人工智能风险和责任的认知，促进人工智能工具在授权范围内规范使用。

（三）持续完善

公司应结合技术发展、业务变化、监管要求、行业标准、内外部审查结果和实际运行情况，持续完善人工智能治理要求。对存在较高风险、争议或不确定性的人工智能应用，应及时复核其必要性、合规性、安全性和适用性。

第九章 附则

（一）解释与衔接

本管理规定与公司数据安全、信息安全、隐私保护、风险管理、合规管理、知识产权、供应商管理和产品质量管理等制度共同构成人工智能治理的管理基础。相关事项已有更严格要求的，从其规定执行。

（二）修订机制

本管理规定应根据法律法规、监管要求、技术发展和公司管理需要适时修订。公司相关部门可结合实际执行情况提出修订建议，并按照公司内部管理程序履行审议和发布。