远东宏信有限公司

负责任人工智能政策

1.0 目的

为规范远东宏信有限公司（以下简称"公司"）及具有实际控制权的下属单位在人工智能（以下简称"AI"）技术运用中的治理与风险管理，确保 AI 系统的开发、部署和使用符合负责任、可信赖、可持续的原则，有效防范算法偏见、隐私泄露、安全漏洞及伦理风险，保护客户与员工权益，促进 AI 技术与业务的健康融合，特制定本政策。

本政策参照《生成式人工智能服务管理暂行办法》《数据安全法》《个人信息保护法》《人工智能生成合成内容标识办法》等法律法规，并参考国际负责任 AI 治理最佳实践制定。

2.0 适用范围

本政策适用于公司及其全资子公司、控股子公司所有人员（含提供服务的第三方人员），覆盖公司所有 AI 系统的开发、采购、部署、运营和退出全生命周期。参股公司可参照执行。

本政策所称 AI 系统包括但不限于：本地化部署的 AI 智能体（如远东 AI 智能体）、第三方云资源大模型服务、基于机器学习的风控与推荐算法、生成式 AI 工具及复合型 AI 智能体等。

3.0 名词定义

3.1 AI 系统：指通过大量资料学习，利用机器学习或相关建立模型之算法，进行感知、预测、决策、规划、推理、沟通等模仿人类学习、思考及反应模式之系统。

3.2 生成式 AI：指可以生成模拟人类智慧创造之内容的相关 AI 系统，其内容形式包括但不限于文本、图像、音频、视频及程式码等。

3.3 高风险 AI 应用：指对客户权益、金融交易决策、营运安全或个人基本权利有重大影响的 AI 应用，包括但不限于信贷审批、风险定价、客户分层、自动化投资建议等。

3.4 人在回路（Human-in-the-Loop，HITL）：指在 AI 系统做出关键决策的流程中，保留具有有效记录的人类监督，并赋予人类干预、否决或升级的权力和能力。

4.0 治理与问责机制

4.1 责任追究机制：
- 公司指定AI治理工作组作为AI系统结果、风险决策及合规性的归口管理角色，负责受理AI相关投诉与争议。
- 建立AI错误或伤害事件的调查与纠正流程：任何因AI系统输出导致客户权益受损、业务决策失误或合规风险的事件，应在发现后24小时内上报AI治理工作组，工作组组织调查、界定责任、制定纠正措施并跟踪闭环。
- 对因违反本政策导致重大损失或合规风险的单位及个人，依据公司《员工手册》及相关制度追究责任；涉及法律责任的，依法移送处理。

4.2 风险管理整合：AI风险管理纳入公司全面风险管理体系及内部控制流程，AI治理工作组每年至少开展一次AI风险全面评估，并将评估结果报送审计与风险管理委员会。

5.0 核心原则

5.1 公平性及以人为本

- AI 系统的设计与训练应尽可能避免算法偏见造成的不公平，对训练数据进行偏见审计，确保创建多样化和具有代表性的数据集。
- 高风险AI应用必须建立"人在回环（HITL）"机制，保留人类对关键决策的审查、核准及最终否决权。在缺乏人类监督及人工审查或否决机制的情况下，AI 系统不得做出不可逆或高风险的决策。
-AI系统之运用应符合以人为本及人类可控原则，生成式AI产出之资讯须由人员就其风险进行客观且专业的管控。

5.2 隐私及客户权益保护

- 严格遵守《个人信息保护法》《数据安全法》，充分尊重及保护客户与员工隐私，妥善管理及运用客户资料。
- 运用AI系统向客户提供金融服务时,应尊重客户选择权,明确告知客户正在使用AI系统,并提供人工服务等替代方案。
- 客户有权了解其个人信息如何被AI系统处理,并有权要求查阅、更正及删除其个人数据。

5.3 系统稳健性与网络安全

-AI系统的开发与运营须确保稳健性（Robustness）、网络安全（Cybersecurity）与可靠性（Reliability），符合公司《信息安全管理办法》及《AI智能体信息安全管控细则》要

求。
- 建立模型全生命周期安全管理，涵盖训练数据安全、训练过程安全、模型部署安全及模型监测与更新安全，防范数据投毒、模型篡改及对抗性攻击。
- 对 AI 系统输出进行内容过滤，防止生成仇恨言论、虚假信息及敏感内容；AI 生成内容须按《人工智能生成合成内容标识办法》添加可识别标识。

5.4 透明性与可解释性

- 公司应确保AI系统运作的透明性，向利益相关方适当披露AI系统的使用场景、能力边界、局限性、数据来源及潜在风险。
- 涉及金融交易决策的 AI 系统，应确保决策逻辑的可解释性，使受影响方能够理解决策是如何做出的。
- 使用 AI 与客户直接互动时，应明确告知客户其正在与 AI 系统交互，并由客户自行选择是否继续使用。

5.5 促进可持续发展

- 运用 AI 系统时，应确保发展策略与可持续发展原则相结合，致力于降低 AI 系统对环境与整体生态足迹的影响。
- 导入或采用 AI 技术时，应优先采用具能源效率、低碳排的数据中心或云服务，评估 AI 系统的能源使用、碳排放及环境影响。自有或委托第三方提供的 AI 数据中心、云端服务及模型均适用本要求。
- 对员工提供 AI 相关教育及培训，促进员工适应 AI 带来的变革，维护其工作权益。

6.0 AI 运用边界

运用 AI 应以合法正当之业务目的为限，不得为下列行为：

1. 未依规定搜集、处理或利用个人资料；
2. 违反或规避主管机关监理政策、法令规范或公司内部控制制度；
3. 未依透明性及可解释性原则向利害关系人为适当之揭露；
4. 对客户权益或营运有重大影响之 AI 系统，未保留人员进行审查、核准或最终决策之权利；
5. 运用 AI 系统从事操纵行为、剥削弱势、社会信用评分、未经合法授权之生物特征辨识分类或监控，或其他对基本人权有清楚威胁之行为；
6. 破坏社会秩序、危害国家安全或造成重大环境损害之行为。

委托第三方业者导入 AI 系统者，应于书面契约明定其应遵循前项规定。

7.0 第三方 AI 管理

- 公司对第三方 AI 服务采用准入机制，对租户安全、数据安全、供应商资质及合规性进行审核，未经审核的第三方 AI 工具不得用于公司业务场景。
- 未经公司允许，禁止向第三方 AI 智能体上传公司内部数据（包括但不限于战略规划、财务报表、客户及员工个人信息、内部代码等）。因业务必须使用的，须对敏感数据进行脱敏处理。
- 禁止将公司内部数据用于外部 AI 模型训练或学习。
- AI 治理工作组定期对第三方 AI 供应商进行风险评估及持续监督，要求供应商提供安全合规证明及审计报告。

8.0 培训与意识

- 公司将 AI 伦理与负责任使用纳入新员工培训及年度全员培训体系，确保员工了解本政策要求及 AI 使用规范。
- 对 AI 开发及运维人员提供专项技术安全培训，对业务部门提供 AI 应用风险识别培训。
- 2025 年公司已实现数据安全培训 100%员工覆盖，后续将负责任 AI 培训纳入常态化培训计划。