

负责任人工智能实施方案

一、编制目的与依据

为贯彻落实《中天科技负责任人工智能应用政策》，确保公司 AI 技术应用在安全、合规、可持续的轨道上推进，特制定本实施方案。依据《中天科技集团“AI+制造”行动方案（2026—2028 年）》及 ISO 42001 人工智能管理体系框架编制。

二、总体目标

到 2028 年，建成覆盖五大主营产业的负责任 AI 管理体系，实现：

- 全部在运 AI 系统通过合规审查与安全性认证；
- AI 模型平均检出率 $\geq 98\%$ 、误报率 $\leq 2\%$ ，且无重大偏见事件发生；
- 累计完成 AI 伦理培训 500 人次以上，全员覆盖率 $\geq 90\%$ ；
- AI 算力碳强度较 2025 基准下降 20%；
- 通过 ISO 42001 人工智能管理体系外部认证。

三、主要措施

（一）限制高风险 AI 功能的访问与使用

- 建立 AI 功能分级目录，将面部识别、行为分析、自主决策等功能列为受限级别，仅经授权特定岗位人员在特定场景下启用。
- 受限功能的使用须填写《AI 特殊功能使用申请单》，由部门负责人及数字化中心双签审批后方可开通临时权限。
- 所有受限功能调用记录实时上传至审计日志系统，保存期限不少于三年。

（二）AI 生成内容与决策结果的标注机制

- AI 生成的报告、判定结论、推荐建议等内容均须附带“AI 辅助生成”水印或标注，区分人机产出。
- 决策类 AI 输出须同步展示：置信度分数、关键参考因子、建议人工复核等级（高/中/低）。
- 对外披露的 AI 生成内容须经人工审核确认后去除标注方可对外发送。

（三）模型漂移检测与性能下降纠正机制

- 每月对各在运 AI 模型的输出分布进行统计监测，对比基线识别漂移迹象。
- 连续两周期准确率下降超过 3%或置信度分布异常时，触发模型复评流程。
- 复评包括新数据采样验证、特征工程检查、超参数调优，必要时重新训练，并走完测试—评审—上线全流程。

（四）AI 模型偏见定期评估

- 每半年组织一次全面 AI 模型偏见评估，维度包括不同产品类型、批次、班次条件下的输出差异。
- 采用统计学显著性检验（卡方检验、t 检验）与专家抽检相结合方式。
- 报告报执委会审阅，发现实质性偏见的模型 30 日内完成整改。

（五）降低 AI 数据碳足迹及硬件生态影响的措施

1. 自主实践：优先轻量化模型架构（MobileNet 系列、蒸馏模型），推广量化与剪枝技术压缩模型体积 50%以上。
2. 与供应商共同实践：云服务和算力供应商选择时将 PUE 值、可再生能源占比纳入评价指标。
3. 推进自建智算中心绿色化，部署液冷 GPU 集群，目标 $PUE \leq 1.20$ 。

（六）用户/受影响方申诉流程

1. 员工或合作伙伴认为 AI 决策造成不利影响的，可通过 OA 系统提交申诉。
2. 数字化中心 AI 治理专员受理后 5 个工作日内启动调查。
3. 调查完成后 10 个工作日内反馈结果；不满意可向执委会提起复核。
4. 所有申诉案例纳入季度治理报告，作为持续改进输入。

（七）量化 AI 举措对员工可持续发展的影响

1. 年底开展 AI 综合影响评估，涵盖：岗位变化人数、技能提升人次、工作负荷变化、满意度调研。
2. 评估报告纳入年度 ESG 报告及社会责任报告向社会公开披露。

（八）员工 AI 伦理与安全培训

培训对象：全体接触或使用 AI 系统的员工（含正式、实习、外包）。

培训内容：

- AI 基础概念与公司 AI 战略布局；
- 《负责任人工智能应用政策》核心条款解读；
- 日常操作规范与异常处理流程；
- 数据安全意识与保密义务；
- 典型违规案例警示教育。

培训形式：线上必修课程 + 线下实操演练（按岗位分级）。

五、发布与公开

本方案经执委会审批后发布，摘要版将通过以下渠道公开：

- 公司官网“可持续发展/ESG”专栏；
- 年度 ESG 报告、社会责任报告；
- 投资者关系平台定期公告。

江苏中天科技股份有限公司

2026 年 7 月 10 日