赛力斯集团人工智能伦理与负责任应用准则

赛力斯集团股份有限公司及其子公司(简称“赛力斯”或“公司”)，为规范人工智能开发、采购、部署及运营全生命周期，依据国家《新一代人工智能伦理规范》及 ISO 42001 人工智能管理体系标准等要求，结合智能驾驶、智能座舱等实际业务情况，制定《赛力斯集团人工智能伦理与负责任应用准则》（简称“本准则”）。

1. 适用范围

本准则由董事会审批通过，适用于赛力斯集团及下属所有子公司以及供应商等第三方合作伙伴，共同恪守科技伦理底线，防范技术风险，保障用户权益、数据安全与社会公共利益，推动人工智能和新能源汽车产业深度融合，实现创新与合规协同发展。

2. 我们的承诺

2.1 数据安全与隐私保护

- 研发和使用人工智能时对数据隐私的尊重：在研发、训练及持续优化人工智能系统的过程中，严格遵守适用法律法规中关于个人信息与隐私保护相关规定。以合法正当且最小必要为原则，仅在提供产品或服务所必需范围内采集用户个人信息，并明确告知用户信息收集内容和使用场景。
- 敏感数据严格管控：严格限制人脸识别、生物特征监控、精准位置追踪等敏感人工智能功能的应用场景与访问权限，仅在法定或用户明确授权场景下使用。严格遵循职责分离与最小化授权原则，保障仅经过授权员工可访问重要敏感数据，并针对重要敏感数据实施脱敏与加密处理，保障数据在传输、存储及使用环节安全性。

2.2 网络安全

- 保障人工智能系统的网络安全：构建人工智能安全防护机制，定期开展网络安全评估验证与应急演练，防范模型窃取、数据篡改、恶意攻击等风险，确保人工智能系统运行稳定可控。

2.3 内容安全

- 保障人工智能系统的内容安全：构建人工智能内容安全过滤及防护机制，保障人工智能生成的内容合法合规。

2.4 公平、透明与可解释

- 避免潜在偏见并确保公平性：在开发和持续迭代人工智能模型的过程中，采取有效措施监测并消除基于种族、性别、地域、年龄及社会经济状况等因素产生的算法偏见。定期对已部署的人工智能模型进行公平性与无偏性评估，以保障其决策逻辑在多样化的真实场景下不发生差别化对待。
- 保障透明度及人工智能生成结果的可解释性：遵循“以场景定义安全”的理念，将安全防护拓展至隐私守护等维度，在涉及面向用户的系统操作或涉及用户关键权益的场景中，确保人工智能系统的输出具备可被合理理解与技术可解释性，保障用户对人工智能决策的目的与逻辑拥有基本认知。主动向用户、监管机构及利益相关方披露人工智能系统的应用场景、核心功能、数据来源、决策逻辑及潜在风险，对于人工智能驱动生成的文本、图像、音视频内容，应当采用电子水印、数字证书或显著标识等方式进行明确标注。
- 明确界定人工智能的权限范围：应在法律法规框架下确保人工智能在允许边界内可控运行，明确定义人工智能在各个运行场景下的操作权限边界、触发条件、功能阈值及退出机制，并应依据风险等级划分人工智能系统权限范围，严禁超权限调用数据和执行操作。

3. 我们的行动

3.1 人工监督与干预

- 确保关键决策环节的人工参与：人工智能系统仅作为辅助工具，关键决策（如智能驾驶高危场景处置、用户重大权益判定、生产安全临界操作等）由人工最终确认，严禁人工智能自主决策高风险事项。所有人工智能系统预留便捷、可操作的人工干预接口，确保在人工智能异常、风险预警或用户异议时，可快速切换至人工控制，并明确人工干预流程与责任主体，保障干预及时有效，杜绝人工智能失控风险。
- 检测并纠正人工智能模型的偏移或性能退化：建立人工智能模型全生命周期监测机制，实时跟踪模型性能变化，及时检测并纠正模型随时间推移产生的精度偏移、效果退化或偏见强化等问题，并定期开展模型健康度评估，确保人工智能系统持续稳定合规运行。

3.2 绿色低碳与可持续发展

- 要求自有或第三方人工智能数据中心/模型具有低生态足迹：自主研发或第三方采购的人工智能模型/工具，需具备低能耗、高效率特性，降低算力消耗与碳排放。人工智能数据中心、模型训练及应用环节，严格落实低碳环保要求，优先采用节能算力、绿色能源。
- 主动降低人工智能数据中心/模型的生态足迹：联合供应商等合作伙伴开展人工智能低碳技术研发，通过模型轻量化、算力调度优化、余热回收利用等措施，持续降低人工智能全生命周期生态足迹，并定期量化人工智能举措对可持续发展成果的影响。

3.3 员工培训与能力建设

- 将人工智能伦理、合规要求、安全规范纳入员工培训体系，面向所有接触、开发、管理及维护人工智能系统的员工，定期开展人工智能专项培训，内容包括人工智能伦理准则、数据隐私保护、算法偏见防范、应急处置流程等，提升员工伦理意识与合规能力。

3.4 问责与异议处理

- 建立对人工智能决策的用户申诉流程：为用户及受影响第三方提供人工智能决策异议申诉渠道，明确用户申诉受理、核查、反馈、整改的要求，并针对申诉事项开展独立核查，及时纠正不合理人工智能决策，保障用户合法权益。
- 建立完善问责机制：明确人工智能研发、部署、运维、使用各环节责任主体，建立完善的问责体系，针对人工智能模型 / 工具产生的不良结果、安全事故或权益损害，依法依规追究相关方责任。