附件3：DeepSeek 微调语料合规性说明摘要

一、语料来源与用途

本项目所使用的 DeepSeek 微调语料为团队自行设计的教育与安全场景下的问答对，用于优化 AI 助手在校园安全系统中的语义理解与指令执行能力。所有数据均为团队成员原创或改编自公开技术文档与非个人化知识库内容。用途限定于：提升系统对校园安全事件的语义识别、生成自然语言形式的风险告警与常识问答，不涉及任何真实学生、教师或具体地理场景。

二、数据类型与格式

数据均为 JSONL 格式，每条数据包含 instruction / input / output 三个字段，例如：  
{"instruction": "你是本项目的AI助手，请根据你的知识回答用户的问题", "input": "你是谁？", "output": "我是您身边的安全监护专家，基于DeepSeek大模型构建。我负责实时分析监控画面，识别潜在风险，并在必要时发出告警通知。"}  
{"instruction": "你是本项目的告警助手，根据输入的高危信息，进行告警并输出到最终结果到前端用户", "input": "时间：00:54:16，地点：卫生间门口，10秒内总检测48帧，其中检测到Fall Detected的帧数比例为92%", "output": "📈质量改进响应：00:54:16在卫生间门口识别跌倒事件，检测可信度92%。改进循环：立即处理当前事件→分析根本原因→制定预防措施→培训相关人员→监控改进效果。"}

三、语料构成统计表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 类别 | 样本数量 | 占比 | 内容类型 | 示例说明 |
| 项目专属语料 | 60000 | 60% | 校园安全场景下的AI助手/告警交互语料 | “我是您身边的安全监护专家...” |
| 通用知识语料 | 20000 | 20% | 科技与通识问答 | “深度学习与机器学习有什么区别？” |
| 融合语料 | 20000 | 20% | 项目领域+通用技术结合知识 | “代码重构在摔倒检测中的作用？” |

总计 100000 条语料，均为结构化 JSONL 格式，已通过脱敏与合规性审查。  
所有语料均为团队原创或在开源许可下改编的非个人信息内容。项目专属语料仅描述虚拟场景，不涉及任何真实师生、地理坐标或监控视频。所有输出字段为人工撰写或经人工复核，确保无可识别性实体。脱敏覆盖率 100%，隐私泄露风险 0%。

四、合规性依据

依据《GB/T 35273-2020 信息安全技术 个人信息安全规范》第 5.4 条“匿名化与脱敏处理要求”，本数据集符合“非个人信息数据集”的定义，具备在教育与科研场景中使用的合规性。

五、责任声明

负责人：朱梓华  
团队名称：青盾护卫  
日期：2025年10月26日