

An Electroencephalogram Dataset of Learner Interest States in Online Education Tasks 可持续维护计划

一、计划目标与核心原则

1.1 核心目标

保障“在线教育任务中学习者兴趣状态的脑电图数据集”（以下简称“本数据集”）的长期可用性、数据质量稳定性、安全合规性及学术价值延续性，支撑在线教育心理学、神经教育学等领域的持续研究与成果转化，实现数据集从“阶段性产出”到“持续性服务”的跨越。

1.2 核心原则

1.2.1 学术导向原则：以服务学术研究为核心，维护措施与数据更新均围绕研究需求设计，确保数据集符合学术规范与可复现要求。

1.2.2 安全优先原则：延续数据保存政策的安全标准，将 EEG 生物识别数据的隐私保护与安全防护贯穿维护全流程，杜绝数据泄露、篡改风险。

1.2.3 合规性原则：动态跟进《个人信息保护法》《人类遗传资源管理条例》等法律法规及学术伦理规范，确保维护活动全程合法合规。

1.2.4 可持续性原则：构建“组织-资源-技术”三位一体的维护体系，保障人力、资金、技术的持续投入，避免维护中断。

1.2.5 开放协作原则：建立学术协作机制，吸纳领域内专家与用户参与维护优化，提升数据集的学术适用性与影响力。

二、组织保障体系

2.1 维护组织架构

设立“数据集维护管理委员会”作为决策机构，下设执行团队与监督团队，明确各级职责，形成“决策-执行-监督”闭环管理。

2.1.1 维护管理委员会：由数据集建设单位负责人、神经教育学专家、数据安全专家、伦理审查委员组成，每季度召开一次会议，审议维护计划、资源分配、重大问题处理方案等核心事项，确保维护方向与学术、合规要求一致。

2.1.2 执行团队：分为技术运维组、数据管理组、学术服务组，实行“专人专岗”责任制，具体职责如下：

——技术运维组（3-5人）：负责存储系统、访问平台、备份机制的日常运维与技术升级，保障系统稳定运行；

——数据管理组（2-3人）：负责数据质量校验、更新补充、归档整理、权限审核，确保数据完整准确；

——学术服务组（2人）：负责对接学术用户、收集使用反馈、组织培训指导、推广数据集应用。

2.1.3 监督团队：由伦理审查委员会与第三方数据安全审计机构组成，每半年开展一次全面监督检查，重点核查数据安全防护、伦理合规性、维护流程执行情况，出具监督报告并提出整改建议。

2.2 人员能力保障

2.2.1 资质要求：技术运维组人员需具备等保三级及以上项目实施经验，熟悉EEG数据存储特性；数据管理组人员需掌握生物数据去标识化技术与学术数据规范；学术服务组人员需具备教育心理学或神经科学相关背景。

2.2.2 培训机制：建立“定期培训+专题培训”体系，每月开展一次内部技术培训（如数据加密技术、系统故障排查），每季度邀请领域专家开展学术规范与伦理培训，每年组织一次跨单位交流学习，提升团队专业能力。

2.2.3 人员备份：为核心岗位（如系统架构师、数据安全负责人）配备AB角，避免因人员变动导致维护中断；建立人员离职交接流程，明确数据权限回收、工作文档移交等要求，确保维护工作连续性。

三、资源保障计划

3.1 资金保障

3.1.1 资金来源：构建“主体投入+项目补充+学术激励”的多元资金池，包括数据集建设单位年度专项经费（占比70%）、科研项目配套资金（如国家自然科学基金数据共享专项，占比20%）、学术成果转化收益分成（占比10%），确保资金长期稳定。

3.1.2 资金分配：实行“分类预算”管理，明确资金使用范围与比例：

——运营维护费（40%）：包括存储设备租金、云服务费用、软件授权费、人员薪酬等日常开支；

——技术升级费（30%）：用于存储系统扩容、安全技术迭代、访问平台优化等；

——数据更新费（15%）：包括新数据采集、标注、校验的相关成本；

——应急储备金（10%）：用于应对系统故障、数据安全事件等突发情况；

——学术推广费（5%）：用于数据集培训、学术会议推广、用户反馈奖励等。

3.1.3 资金监管：建立独立的资金账户，由维护管理委员会监督资金使用，每年度出具资金使用审计报告，确保资金专款专用、高效利用。

3.2 技术资源保障

3.2.1 硬件资源：采用“本地+云存储”混合架构，本地部署物理隔离的加密服务器（用于一级EEG原始数据），云存储选用通过等保三级认证的医疗健康级云服务（如阿里云医疗云），定期（每2年）对本地服务器进行性能检测与升级，根据数据增量提前规划存储扩容（当存储使用率达80%时启动扩容）。

3.2.2 软件资源：搭建专用的数据集管理与访问平台，集成数据检索、权限控制、使用统计、反馈交互等功能；配备专业软件工具，包括EEG数据质量校验工具（如EEGLAB）、数据加密软件（如Blancco）、安全监测系统（如入侵防御IPS），确保软件版本实时更新，每年开展一次软件兼容性评估与升级。

3.2.3 备份资源：延续“三级备份”机制，本地备份（实时同步）、同城异地备份（每日全量）、异地灾备（每周全量），定期（每月）对备份数据进行完整性与可用性验证，确保备份资源有效可用。

四、全生命周期维护流程

4.1 数据采集与更新维护

4.1.1 更新周期与范围：建立“定期更新+按需补充”的更新机制，每2年开展一次大规模数据补充采集，聚焦在线教育新场景（如AI互动教学、虚拟现实教学）下的学习者兴趣状态数据，确保数据集覆盖场景的全面性；当学术研究出现新需求（如特定年龄段学习者数据）时，启动专项采集。

4.1.2 采集规范延续：严格遵循原数据保存政策的采集要求，以伦理审查通过为前置条件，执行“知情同意+去标识化”流程，新采集数据需与原有数据格式、

标注标准保持一致，采集完成后由数据管理组进行完整性、准确性校验，校验通过率达 100%方可纳入数据集。

4.1.3 版本管理：采用“主版本+子版本”的版本命名规则（如 V1.0 为初始版本，V1.1 为场景补充版本，V2.0 为大规模更新版本），每个版本明确更新内容、时间、采集说明等元数据信息，旧版本数据保留归档，确保学术研究可追溯。

4.2 数据质量维护

4.2.1 日常质量监控：技术运维组每日监控数据存储状态，数据管理组每周抽取 10%的一级、二级数据进行质量抽检，重点核查 EEG 信号完整性（无断连、无严重干扰）、去标识化有效性（无残留唯一标识）、标注一致性（专家标注与算法标注误差率≤5%），发现问题 24 小时内启动整改。

4.2.2 定期质量评估：每半年开展一次全面数据质量评估，制定量化指标体系（如数据完整率≥99%、准确率先≥98%、标注一致率≥95%），评估结果形成报告提交维护管理委员会，针对不达标数据制定清洗、补充或剔除方案。

4.2.3 冗余数据处理：数据管理组每年度梳理数据集，对重复数据、无效数据（如信号严重干扰无法修复的数据）进行筛选，经学术专家审核确认无研究价值后，按照原数据销毁流程执行销毁，释放存储资源。

4.3 存储与安全维护

4.3.1 存储系统运维：技术运维组每日监测存储系统运行状态（CPU 使用率、存储空间、响应速度），每月进行硬件设备检修（如服务器散热、硬盘健康检测），每季度开展系统性能优化（如缓存调整、数据碎片整理），确保系统故障率≤0.5%/年。

4.3.2 安全防护升级：每月进行一次安全漏洞扫描，每半年开展一次渗透测试，及时修复系统漏洞；每年升级一次加密技术（确保与行业最新加密标准同步），一级数据加密算法不低于 AES-256 级别，数据传输采用 SSL/TLS 1.3 协议；建立安全事件应急响应流程，一旦发生数据泄露、入侵等事件，15 分钟内启动应急预案，24 小时内完成初步处置并上报监管部门。

4.3.3 访问权限管理：数据管理组每季度对用户访问权限进行全面核查，回收闲置权限、调整超额权限，确保权限分配符合“最小必要”原则；对一级数据的访问记录进行重点审计，发现异常访问行为立即冻结权限并追溯调查。

4.4 归档与销毁维护

4.4.1 归档管理：对超出日常使用周期但仍在保存期限内的数据，由数据管理组按季度完成归档，存储至专用归档系统，建立“数据-版本-学术成果”关联索引，确保归档数据可快速检索；归档数据的访问需履行原政策规定的三级审批流程。

4.4.2 销毁执行：严格遵循原数据保存政策的销毁条件与流程，对达到保存期限且无研究价值的数据，经维护管理委员会与伦理审查委员会联合审批后，由执行团队采用“软件覆盖+物理粉碎”的双重销毁方式，销毁过程全程留痕，销毁记录保存不少于 10 年。

五、学术服务与协作机制

5.1 用户服务体系

5.1.1 访问支持：学术服务组为用户提供“线上+线下”双重支持，线上搭建 7×12 小时智能客服平台（解答常见问题）与专家咨询通道（24 小时内响应复杂问题），线下每季度组织一次数据集使用培训（覆盖数据获取、预处理、分析工具使用等内容）。

5.1.2 反馈收集：建立“用户反馈-处理-优化”闭环机制，通过访问平台反馈模块、学术会议调研、定期用户访谈等方式收集需求与问题，每月汇总反馈报告，明确处理责任与时限（一般问题3个工作日内解决，复杂问题15个工作日内回复）。

5.1.3 成果追踪：为使用数据集发表学术成果的用户建立成果档案，记录论文DOI、研究方向、成果应用情况等信息，每年度发布数据集应用报告，推广优秀研究成果，提升数据集学术影响力。

5.2 学术协作网络

5.2.1 合作单位共建：与3-5所高校（神经教育学、教育技术学专业）及2家科研机构建立长期合作关系，联合开展数据采集、质量评估、技术研发等工作，合作单位参与维护计划制定，共享维护资源与学术成果。

5.2.2 专家咨询机制：聘请5-8名领域内知名专家组成学术顾问团，每季度提供一次专业咨询，为数据更新方向、质量标准优化、学术推广策略等提供指导，确保数据集符合学术前沿需求。

5.2.3 开放共享规范：制定数据集开放共享细则，明确不同级别数据的共享范围（四级数据全开放，三级数据仅对学术机构开放，一级、二级数据需专项审批），要求使用方签署学术使用承诺函，规范数据使用行为，避免滥用与侵权。

六、评估与优化机制

6.1 定期评估体系

6.1.1 评估周期与主体：每年度开展一次全面维护评估，由监督团队牵头，联合维护管理委员会、学术顾问团及用户代表组成评估小组，确保评估客观全面。

6.1.2 评估指标体系：从五个维度设定量化评估指标，具体如下：

——数据维度：数据完整率 $\geq 99\%$ 、准确率先 $\geq 98\%$ 、更新及时率 $\geq 95\%$ ；

——系统维度：系统可用率 $\geq 99.5\%$ 、安全事件发生率=0、故障恢复时间 ≤ 2 小时；

——服务维度：用户问题解决率 $\geq 98\%$ 、用户满意度 $\geq 90\%$ 、培训参与人数年增长 $\geq 15\%$ ；

——合规维度：伦理审查通过率=100%、法规符合率=100%、销毁流程规范率=100%；

——学术维度：年度应用成果数 ≥ 10 项、合作单位数稳定 ≥ 5 家、数据集引用率年增长 $\geq 20\%$ 。

6.1.3 评估结果应用：评估结果分为“优秀”“合格”“不合格”三个等级，优秀等级可申请额外学术推广资金；合格等级需针对薄弱环节制定改进计划；不合格等级由维护管理委员会重新审议维护方案，调整资源分配与团队架构。

6.2 持续优化措施

6.2.1 动态调整维护计划：根据年度评估结果与法律法规更新（如数据安全相关政策调整），每年对本维护计划进行修订，明确下一年度维护重点、资源分配及优化方向，修订后的计划需经维护管理委员会审批后公示。

6.2.2 技术创新应用：关注人工智能、区块链等新技术在学术数据管理中的应用，每2年开展一次技术可行性研究，探索AI自动校验数据质量、区块链存证数据溯源等应用场景，提升维护效率与数据可信度。

6.2.3 长期发展规划：每3年制定一次中长期发展规划，结合学术领域发展趋势（如神经教育学的新研究方向）与技术变革，明确数据集的拓展目标（如新增多模态数据融合、跨地域学习者数据），确保数据集长期符合学术研究需求。

七、附则

- 7.1 本计划由数据集维护管理委员会负责解释与修订，修订需经全体委员半数以上表决通过，并提前向合作单位与核心用户公示。
- 7.2 本计划未尽事宜，参照《中华人民共和国数据安全法》《学术数据管理规范》及本数据集数据保存政策执行。
- 7.3 本计划自发布之日起施行，原有维护相关规定与本计划不一致的，以本计划为准。