《端侧大模型更新与运维技术规范》
标准编制说明

标准起草小组

1.标准范围

本标准规定了端侧大模型更新方式、版本管理、全生命周期管理及性能监控的技术要求、实施流程和操作准则，核心明确增量更新、全量更新、差分更新的技术规范，统一版本命名、追溯、回滚的管理要求，界定模型部署、升级、卸载、残留清理的操作标准，制定资源占用监控、错误上报、数据管理的性能指标。

本标准适用于搭载端侧大模型的智能移动终端、智能穿戴设备、智能车载终端、智能家居终端等各类终端设备，其他搭载端侧大模型的终端产品研发、生产、运维可参照执行。本标准的使用主体包括端侧大模型研发机构、终端设备生产企业、终端产品运维服务机构等相关单位。

2.工作简况

3.标准编制原则和确定标准主要内容

本标准的主要内容基于端侧大模型产业发展的实际痛点、行业技术研发的主流方向和企业实际运维需求确定，同时结合起草单位的技术实践和研究成果，确保内容的科学性和行业适配性。

模型更新方式：针对行业内更新方式不统一的问题，结合端侧大模型小版本迭代、大版本升级、精准定制化更新的不同场景，确定增量、全量、差分三种更新模式，明确各模式的技术要求和执行流程，适配不同场景的更新需求。

模型版本管理：参考信息技术领域通用的版本管理规范，结合端侧大模型的技术特征，确定三级版本命名规则，同时明确版本追溯、归档、回滚、兼容的要求，解决版本管理混乱、无法溯源的行业痛点。

端侧模型生命周期管理：围绕模型从部署到卸载的全生命周期，结合终端设备的硬件特性和系统要求，制定部署前检测、部署实施、升级、卸载、残留清理的操作标准，删除过高技术要求，确保各环节要求可落地。

性能监控：基于终端设备的资源约束和用户使用体验需求，确定算力、内存、存储、电量消耗等核心资源监控指标及合理阈值，同时对错误进行分级，明确错误上报的用户确认机制和处理要求，保障模型稳定运行。

4.主要试验(或验证)的分析、综述报告

无。

5.标准在起草过程中遇到的问题及解决办法；重大分歧意见的处理经过和依据；有无重要技术问题需要说明

在本文件的修订过程中，无重大分歧意见和技术问题。

6.与国外标准的关系：包括：采用国际标准和国外先进标准的程度，与国外标准主要技术内容的差异

该项目没有对应的国际标准或国外先进标准。

7.修订标准时，说明与标准前一版本的重大技术变化，并列出所涉及的新、旧版本的有关章条(可引用标准前言的内容)；废止/代替现行有关标准的建议

不涉及。

8.说明标准与其他标准或文件的关系(可引用标准前言的内容)，特别是与有关的现行法律、法规和强制性国家标准的关系

符合现行法律、法规要求。

9.标准作为强制性标准或推荐性标准的建议

建议本文件作为推荐性标准。

10.贯彻国家标准的要求和措施建议(包括组织措施、技术措施、过渡办法等内容)；标准发布后，对国内外业界可能产生的影响

建议本文件作为推荐性标准发布实施。

11.标准是否涉及知识产权的情况说明；如标准中含有自主知识产权，说明产品研发程度、产业化基础及进程

本文件未涉及。

12.其他应予说明的事项

本文件未涉及。