

ICS 35.240

CCS L70

团体标准

T/TAF 320—2025

大型语言模型管理指南

Large language model management guide

2025-12-05 发布

2025-12-05 实施

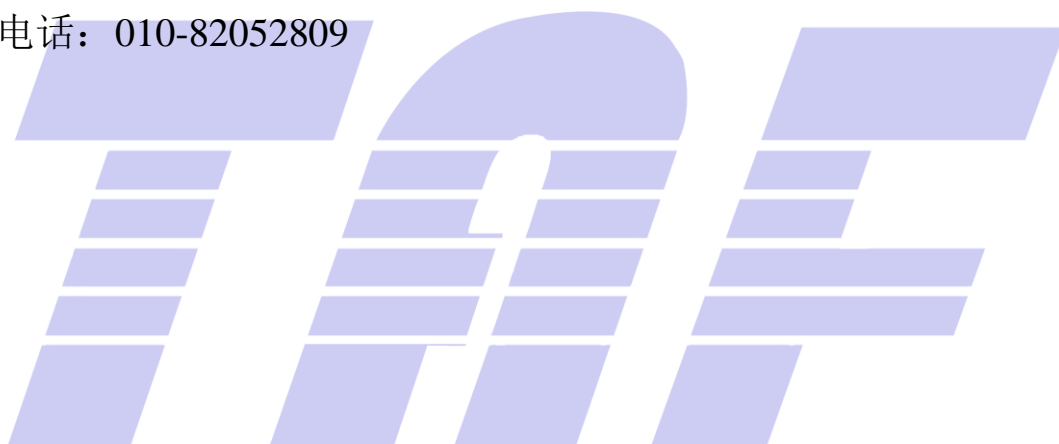
电信终端产业协会 发布

版权声明

本文件的版权属于电信终端产业协会，任何单位和个人未经许可，不得进行技术文件的纸质和电子等任何形式的复制、印刷、出版、翻译、传播、发行、合订和宣贯等，也不得未经允许采用其具体内容编制本团体以外各类标准和技术文件。如有以上需要请与本团体联系。

邮箱：tafrb@taf.org.cn

电话：010-82052809



目 次

前言	II
引言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 基础资源管理	1
5 技术开发管理	2
6 应用运营管理	2
参考文献	4



前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由电信终端产业协会（TAF）提出并归口。

本文件起草单位：中国信息通信研究院、泰尔认证中心有限公司、蚂蚁科技集团股份有限公司、北京邮电大学、北京科技大学、深圳大学。

本文件主要起草人：杨光、李杰强、陈思宇、凌大兵、毕春丽、付娜、张蕾蕾、呼娜英、郭苏敏、刘志鹏、程莹、林冠辰、朱一凡、张晓丽、李若愚、薛刚、常琳、赵昕、于爱鑫、李思桥、王雪、翟国丽、田崇贤、李颖瑶、吴壮飞、王云龙。



引 言

近年来，人工智能技术快速发展，大模型、生成式人工智能等产业领域发展活跃，加速推动人工智能产品或服务技术研发与产业落地已成为社会各方关注的重点。2024年6月工信部、网信办等四部委联合印发《国家人工智能产业综合标准化体系建设指南（2024版）》，其中明确提出“加快构建满足人工智能产业高质量发展和‘人工智能+’高水平赋能需求得标准体系”，并将管理标准纳入基础共性标准进行明确要求：“规范人工智能技术、产品、系统、服务等全生命周期涉及的人员、组织管理要求和评价”。

当前阶段，大型语言模型作为人工智能技术的主要代表，因其规模可扩展性、多任务适应性、复杂推理能力、知识吸收整合能力等特点，已成为引领新一轮科技革命和产业变革的重要驱动力。尽管大型语言模型具备巨大潜力和广泛应用场景，但在落地应用过程中仍面临诸多挑战，需要从技术、产品、系统、服务等方面，在开发及应用的生命周期各个环节对涉及到的人员、组织的有关管理要求进行规范要求；推动开发团队、算法团队、部署团队、运营团队和风险管理团队共同参与建设研发、部署和运营各环节标准化管理体系，构建风险识别、检测、管控以及应对处理机制等一系列标准化管理体系，以保障大型语言模型落地应用过程持续稳定有效。本标准结合相关监管需求及组织防范内外部风险需要，为开发、提供和应用大型语言模型的组织，在基础资源管理、技术开发管理、及应用运营等方面，所需建立的管理措施提供参考。



大型语言模型管理指南

1 范围

本文件规定了大型语言模型的基础资源管理、技术开发管理、应用运营管理等内容，适用于为开发、提供与应用以大型语言模型为核心算法的产品应用或服务（以下简称大型语言模型）的组织（包含大型语言模型厂商、应用大型语言模型的企业等）建立有关管理措施提供参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 43697 生成式人工智能服务安全基本要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

模型训练 model training

利用训练数据，基于机器学习算法，确定或改进机器学习模型参数的过程。

3.2

推理 inference

从给定的前提进行论证并得出结论。

3.3

问答 question answering

确定以自然语言提供的问题最合适答案的任务。

4 基础资源管理

4.1 系统和计算资源

系统和计算资源包括计算任务执行，包括数据处理、模型训练、模型推理等计算任务，所必须的计算、存储、网络等硬件资源或虚拟化资源，有关的管理流程包括但不限于：

- a) 组织可建立应急处理机制以应对突发故障、遭受攻击、自然灾害、流量激增等情况；
- b) 组织可建立完备的日志审计机制以验证服务提供的完整性；
- c) 组织可建立监控保障机制，保障模型训练及推理任务在异构或虚拟环境下的安全持续运行，包含训练任务的中断恢复等。

4.2 工具资源

工具资源包括操作系统、计算加速库、开发框架及组件、算法工具、开源产品等软件资源。

组织可建立审查流程，对模型开发及应用阶段所使用的，可能影响模型算法有效性、透明性、可解释性等指标的工具资源进行评审，以防范侵权、安全等相关风险。

4.3 数据资源

数据资源包括大型语言模型训练、推理等阶段所使用的数据资源，有关的管理流程包括但不限于：

- a) 组织可对涉及到个人信息的数据管理符合有关法律法规的要求的情况进行确认；
- b) 组织可对涉及到数据来源、数据标准等的管理符合有关法律法规的要求的情况进行确认；

注：参考GB/T 43697。

- c) 组织可对模型训练、集成测试等阶段所使用的数据及其标注进行核验检测，以对数据的真实性、有效性、可追溯性、公平性等指标进行评价。

4.4 人力资源

人力资源包括用于大型语言模型的开发、部署、运行等的人力资源，有关的管理流程包括但不限于：

- a) 组织可明确岗位及部门的职责，明确不同角色间的沟通机制及争议解决机制，确保监督及管理人员的职责能够得到有效履行；
- b) 组织可定期组织培训以维持与提升人员的必要能力，并对人员能力进行有效跟踪；
- c) 组织可建立由多部门参与的虚拟组织，确保有关政策、流程的有效推动与落地。

5 技术开发管理

5.1 模型设计

组织可明确大型语言模型架构设计、算法设计、模型扩展等所需符合的要求及其对应的管理流程，包括但不限于：

- a) 组织可通过审查流程明确模型的开发及应用对人类福祉及社会发展起到积极作用；
- b) 组织可通过管理流程对模型的所需基础资源进行评审。

5.2 训练推理

组织可明确大型语言模型训练、微调、压缩、推理等阶段所需符合的要求及其对应的管理流程，包括但不限于：

- a) 组织可对模型的训练及评估方法进行明确；
- b) 组织可在模型训练前对数据准备情况进行确认。

5.3 验证确认

组织可明确大型语言模型有关公正性、安全性、可解释性、时效性等指标的验证要求及其对应的管理流程，包括但不限于：

- a) 组织可建立验证确认准则，包括对技术文档完备性、管理过程成文信息的检查等；
- b) 组织可至少完成一次有关有效性、安全性、可解释性、公平性、时效性等指标的测试评估。

6 应用运营管理

6.1 应用部署

组织可明确模型集成、推理部署、人机交互等所需符合的要求及其对应的管理流程，包括但不限于：

- a) 组织可在模型上线应用前对需履行法律法规要求的有关手续进行确认；
- b) 组织可建立应用部署准则并通过审批后上线部署，发布准则可包括对技术文档完备性的检查等；
- c) 组织可明确模型的计划用途以防止误用滥用的情况。

6.2 运行监控

组织可明确模型运营期间有关应用监控、异常检测、应急处置、容灾备份等所需符合的要求及对应管理流程，包括但不限于：

- a) 组织可对模型服务期间依据法律法规要求的落实有关安全措施的情况进行确认；
注：参考GB/T 43697。
- b) 组织可对涉及生命健康、社会秩序等敏感领域进行重点监控并建立应急处置机制；
- c) 组织可建立监控及分级处置机制以发现模型滥用误用所可能导致的有关隐私安全等问题；
- d) 组织可建立事件应对机制以应对系统故障、服务异常等影响模型有效性的事件；
- e) 组织可建立通报机制以将模型运行监控情况及时向相关方传达，可包含对个人、团体及社会可能造成的影响等。

6.3 改进提升

组织可明确模型服务用户期间内容过滤、数据回流、客诉反馈等所需符合的要求及对应管理流程，包括但不限于：

- a) 组织可对用于模型改进的数据回流等技术措施依据法律法规要求的落实情况进行确认；
注：参考GB/T 43697。
- b) 组织可建立客诉反馈及分级处置机制以发现并解决模型使用过程中所出现的问题；
- c) 组织可建立审查流程对模型的更新改进进行评审。

6.4 应用评价

组织可明确模型服务期间有关业务优化、研发生产、供销存采、创新服务、质量安全等所需符合的要求及对应管理流程，包括但不限于：

- a) 组织可对供应商、客户等相关方的职责及有关约定进行明确以评价模型应用情况；
- b) 组织可定期对模型应用情况开展评估以确认预期目标的达成情况；
- c) 组织可对模型在业务优化、研发生产、供销存采等核心业务环节的应用建立监控预警等管理措施以防范数据安全、质量安全等风险；
- d) 组织可对模型全生命周期或特定阶段内的自然资源消耗情况进行记录，以评估对生态环境的影响。

参 考 文 献

- [1] 欧盟人工智能法案 (EU Artificial Intelligence Act)
- [2] ISO 42001 信息技术 人工智能 管理体系 (Information technology - Artificial intelligence - Management system)
- [3] 新加坡资讯通信媒体发展局《传统人工智能模型治理框架》 (IMDA Model AI Governance Framework for Traditional AI)
- [4] 新加坡资讯通信媒体发展局《生成式人工智能模型治理框架》 (IMDA Model AI Governance Framework for Generative AI)
- [5] 《中华人民共和国数据安全法》
- [6] 《中华人民共和国网络安全法》
- [7] 《中华人民共和国个人信息保护法》
- [8] 《科技伦理审查办法 (试行)》
- [9] 《互联网信息服务深度合成管理规定》
- [10] 《生成式人工智能服务管理暂行办法》



电信终端产业协会团体标准

大型语言模型管理指南

T/TAF 320—2025

版权所有 侵权必究

电信终端产业协会发布

地址：北京市西城区新街口外大街 28 号

电话：010-82052809

电子版发行网址：www.taf.org.cn