ICS 35.240.50
CCS J 07

T/CS

团 体 标 准

T/CS 820—2026

工业产品 AI 分析模型全生命周期数据管理规范

Specification for full lifecycle data management of AI analysis models for industrial products

2026 - 10 - 16 发布 2026 - 10 - 16 实施

中国商品学会 发 布

目　　次

前言 II
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 缩略语 2
5 总体要求 2
6 数据采集、存储与质量 3
7 数据安全与合规 4
8 模型训练与验证数据管理 4
9 模型部署、运维与退役数据管理 5
10 数据管理能力评估 6
11 验证方法 6
参考文献 9

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。

本文件由广东南方通信建设有限公司提出。

本文件由中国商品学会归口。

本文件起草单位：广东南方通信建设有限公司、山东德礼科技发展有限公司、江苏金陵机械制造总厂、广东甄明司法鉴定所、汇能工业技术（西安）有限公司、徐州三米科技有限公司、深圳市新源润产品信息技术有限公司、深圳市帝江文化传媒有限公司、深圳市鹏龙成科技有限公司、广西友邦永信网络技术有限公司、合肥焱基域科技有限公司、巴林右旗人事人才服务中心。

本文件主要起草人：冯天玮、庄稼、李永霞、张雯集、李智、郭俊、陈彦君、卓莹鎗、柳丽娜、王飞、郭振民、陈超、李国华。

工业产品 AI 分析模型全生命周期数据管理规范

1 范围

本文件规定了工业产品 AI 分析模型全生命周期数据管理的总体要求、数据采集、存储与质量、数据安全与合规、模型训练与验证数据管理、模型部署、运维与退役数据管理、数据管理能力评估与验证方法的要求。

本文件适用于制造业、能源、矿业等工业领域，面向生产过程优化、设备健康管理、质量检测、安全监控、资源调度等场景的工业产品 AI 分析模型全生命周期各阶段的数据管理活动。

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 22239 信息安全技术 网络安全等级保护基本要求

GB/T 35119 产品生命周期数据管理规范

GB/T 36073 数据管理能力成熟度评估模型

GB/T 36344 信息技术 数据质量评价指标

GB/T 37964 信息安全技术 个人信息去标识化指南

GB/T 37988 信息安全技术 数据安全能力成熟度模型

GB/T 39400 工业数据质量 通用技术规范

GB/T 42460 信息安全技术 个人信息去标识化效果评估指南

GB/T 42888 信息安全技术 机器学习算法安全评估规范

GB/T 43697 数据安全技术 数据分类分级规则

3 术语和定义

下列术语和定义适用于本文件。

3.1

工业产品AI分析模型 industrial product AI analysis model

面向工业产品设计、制造、运行和维护等环节，利用机器学习或深度学习算法对工业数据进行训练、推理和分析的数学模型。

3.2

全生命周期数据管理 full lifecycle data management

对工业产品AI分析模型在数据准备、模型训练与验证、模型部署与运维、模型退役各阶段所涉及的数据进行采集、存储、处理、使用、归档和销毁的系统性管理活动。

3.3

训练数据集 training dataset

用于训练工业产品AI分析模型的数据集合，由样本数据和相应的标注数据构成。

3.4

数据漂移 data drift

模型部署运行后，实际数据的统计分布或数据生成关系随时间偏离训练数据分布或训练时所依赖的数据生成关系的现象，包括特征漂移、概念漂移。

4 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

DCMM：数据管理能力成熟度评估模型（Data Management Capability Maturity Assessment Model）

ML：机器学习（Machine Learning）

5 总体要求

5.1 总体原则

工业产品 AI 分析模型全生命周期数据管理应遵循以下原则：

a) 全程覆盖原则：数据管理活动应覆盖模型从数据准备到退役的全生命周期各阶段；

b) 数据溯源原则：各阶段数据的来源、处理过程、使用记录和变更历史应可追溯；

c) 质量优先原则：各阶段数据应满足 GB/T 39400 和 GB/T 36344 规定的数据质量要求；

d) 安全合规原则：数据处理活动应符合 GB/T 43697、GB/T 37988 和 GB/T 22239 规定的要求；

e) 分类分级原则：工业数据的分类应符合 GB/T 43697 的规定，分级参照《工业数据分类分级指南（试行）》执行，并在数据管理计划中明确与 GB/T 43697 中核心数据、重要数据、一般数据的映射关系。数据管理策略应与数据级别相适应。

5.2 生命周期阶段

工业产品 AI 分析模型全生命周期数据管理应划分为以下阶段：

a) 数据采集与接入阶段；

b) 数据存储与组织阶段；

c) 模型训练与验证阶段；

d) 模型部署与运维阶段；

e) 模型退役阶段。

5.3 数据分类分级

5.3.1 工业产品 AI 分析模型相关数据应按 GB/T 43697 规定的分类维度进行分类，应覆盖以下数据域，可结合业务实际补充其他数据域：

a) 研发数据域：模型训练数据集、验证数据集、测试数据集、标注数据、特征工程产出数据等；

b) 生产数据域：模型推理输入数据、模型输出数据、实时工况数据、质量检测数据等；

c) 运维数据域：模型运行日志、监控指标数据、数据漂移监测记录、模型性能退化记录等；

d) 管理数据域：模型版本信息、数据版本信息、模型配置参数、审批记录等。

5.3.2 工业产品 AI 分析模型相关数据的分类应依据 GB/T 43697 规定的分类框架执行。数据分级应结合数据遭篡改、破坏、泄露或非法利用后对工业生产、经济效益造成的潜在影响，参照工业和信息化部《工业数据分类分级指南（试行）》的规定，将工业数据划分为一级、二级、三级共三个级别。可根据业务需要，在数据管理计划中明确各级数据与 GB/T 43697 中核心数据、重要数据、一般数据的映射关系。

5.3.3 用于训练工业产品 AI 分析模型的三级数据，其采集、存储、传输、处理和销毁全过程应实施加密保护并开展访问审计。

5.4 数据管理计划

5.4.1 组织应编制工业产品 AI 分析模型数据管理计划。数据管理计划应明确数据分类分级、数据质量量化指标、数据漂移监测频率、模型更新阈值、数据保留期限和验证判定阈值等内容。
5.4.2 数据管理计划应履行内部审批流程。当模型业务场景、数据来源、安全级别、阈值条件发生变更，或模型进入新生命周期阶段时，应及时修订并重新审批，对计划实施版本管理。

6 数据采集、存储与质量

6.1 数据采集与接入

6.1.1 模型开发启动前应编制数据需求规格说明书，明确所需数据的类型、规模、精度、时间跨度、采样频率、标注要求。数据需求规格说明书应由数据管理责任人和模型开发责任人共同审批，并归档留存。
6.1.2 工业产品 AI 分析模型的训练数据应从工业产品设计系统、制造执行系统、设备传感器、质量检测设备、运维服务系统等数据源采集。
6.1.3 数据采集应记录采集时间戳、数据来源标识、采集设备标识和采集环境参数等元数据。
6.1.4 数据采集过程中应实施数据完整性校验，校验规则应包括空值检测、越界检测和时间连续性检测。
6.1.5 外部数据接入应经过数据安全审查，审查内容应包括数据来源合法性、数据内容合规性和数据传输安全性。
6.1.6 数据接入过程中涉及三级数据的，应采用加密传输方式，加密强度应满足 GB/T 22239 和 GB/T 37988 规定的要求。

6.2 数据存储与版本管理

6.2.1 工业产品 AI 分析模型各阶段数据应实施分级存储。训练数据集、验证数据集和测试数据集应实施版本管理，每次模型训练所使用的数据集版本应唯一标识并记录。
6.2.2 三级数据的存储应满足以下要求：

a) 存储介质应实施物理隔离或逻辑隔离；
b) 存储加密应使用国家密码管理部门认可的加密算法；
c) 访问日志应完整记录，日志保存期限不应少于 6 个月，涉及三级数据的日志保存期限不应少于 12 个月，符合 GB/T 37988 和 GB/T 22239 的规定。

6.2.3 数据存储应建立冗余备份机制，备份策略应符合数据管理计划的规定。
6.2.4 工业产品 AI 分析模型数据应按照 GB/T 35119 规定的产品生命周期数据模型进行组织，数据模型应覆盖研发设计、生产制造和交付服务各阶段。
6.2.5 数据集应建立索引，索引信息应包含数据集标识、版本号、创建时间、数据规模、标注信息、质量评分和适用模型类型。
6.2.6 训练数据集、验证数据集和测试数据集均应实施版本管理。版本号编制规则应包含主版本号、次版本号。数据发生实质性变更时更新主版本号；修正标注错误、补充少量样本等非实质性变更时更新次版本号。
6.2.7 数据版本变更应记录变更原因、变更内容、变更影响评估和审批信息。模型训练所使用的数据版本应与模型版本建立可追溯关联，关联信息记录至模型元数据。

6.3 数据质量

6.3.1 工业产品 AI 分析模型的数据质量评价应按照 GB/T 39400 和 GB/T 36344 规定执行。数据质量维度应包括规范性、完整性、准确性、一致性、时效性和可访问性。
6.3.2 数据质量量化指标应在数据管理计划中规定。数据管理计划中的量化指标不应低于 GB/T 39400 和 GB/T 36344 规定的要求。
6.3.3 应建立数据质量持续改进流程，包括数据质量需求定义、数据质量测量、数据质量分析和数据质量改进四个环节，质量改进过程与结果应留存记录。
6.3.4 数据质量测量应在每次模型训练前执行，测量结果应形成数据质量报告。数据质量报告的格式和内容应符合 GB/T 39400 的规定。
6.3.5 数据质量不满足数据管理计划规定的量化指标时，应实施数据清洗、补充采集或标注修正，整改后重新测量。
6.3.6 模型部署运行后，应对推理输入数据分布开展持续监测，覆盖特征漂移、概念漂移。数据漂移监测频率应适配业务数据更新频率，并满足数据管理计划要求。
6.3.7 检测到数据漂移时，应评估漂移程度以及对模型业务性能的影响；评估确认模型性能下降超出数据管理计划设定阈值时，应启动模型重训练或模型更新流程。

7 数据安全与合规

7.1 数据安全

7.1.1 工业产品 AI 分析模型全生命周期数据处理活动应按照 GB/T 43697、GB/T 37988 和 GB/T 22239 的规定实施数据安全保护。
7.1.2 数据安全风险评估应定期开展，评估频率应符合数据管理计划及相关法律法规要求。当数据处理活动发生重大变更时，应重新开展安全风险评估。
7.1.3 涉及三级数据的处理活动，应部署符合 GB/T 22239 和 GB/T 37988 规定的安全防护措施，实施安全功能和安全管理的相关要求。

7.2 数据访问控制

7.2.1 数据访问应实施基于角色的访问控制，角色定义应与模型开发流程中各岗位职责相对应。
7.2.2 三级数据的访问应实施双人授权机制，访问行为应全程审计。
7.2.3 数据访问权限应定期审查，审查周期应符合数据管理计划的规定，权限变更应及时更新并记录。

7.3 数据脱敏与合规

7.3.1 模型训练和测试过程中使用涉及个人信息或商业秘密的数据时，应分别实施脱敏或去标识化处理：
a) 涉及个人信息的，应按照 GB/T 37964 和 GB/T 42460 的规定实施去标识化处理，并评估去标识化效果；
b) 涉及商业秘密的，应按照组织制定的商业秘密保护管理制度或相关标准实施脱敏处理，脱敏规则应根据数据类型、密级和用途确定。
7.3.2 去标识化或脱敏后的数据应保障可用于模型训练，去标识化或脱敏效果应开展验证，验证记录应予归档。
7.3.3 数据管理活动应遵守相关法律法规的规定。
7.3.4 确需开展跨境数据传输的，应按照国家相关法律法规要求完成安全评估等合规工作。

8 模型训练与验证数据管理

8.1 数据集划分

8.1.1 数据集应划分为训练数据集、验证数据集和测试数据集。三个数据集之间不应存在样本泄露。因工业场景特殊原因（如同一设备多传感器时序采样、同一批次产品多角度检测）导致样本存在关联的，应在数据管理计划中明确关联判定规则和允许关联的情形，并评估其对模型评估结论的影响。
8.1.2 数据集划分比例应根据样本总量和任务类型确定，并应符合数据管理计划的规定。数据集划分应记录划分方法、随机种子和划分时间。

8.2 训练过程数据记录

8.2.1 模型训练过程应记录以下数据：训练数据集版本标识、验证数据集版本标识、超参数配置、训练轮次、损失函数、评价指标的原始序列数据，可留存对应的变化曲线。
8.2.2 完整留存的训练过程数据应能够支撑训练过程复现。

8.3 模型验证数据要求

8.3.1 模型验证应使用独立于训练数据的测试数据集。测试数据集的分布应与实际工业场景中的数据分布一致。
8.3.2 模型验证应同时满足安全评估和性能评价两方面要求：

a) 安全评估：模型的安全评估应按GB/T 42888规定的机器学习算法安全评估要求和评估方法执行；
b) 性能评价：模型的性能评价指标应根据任务类型和业务需求确定。对于分类模型，评价指标至少包含准确率、精确率、召回率和F1分数；对于回归模型，评价指标至少包含平均绝对误差、均方根误差和决定系数。面向多标签、异常检测等特殊工业场景，应结合业务需求补充对应的业务评价指标。

8.4 模型与数据关联管理

8.4.1 每个模型版本应与所使用的数据集版本建立唯一关联关系，关联信息应包含数据集标识、版本号和数据质量评分。
8.4.2 模型训练完成后，应将模型文件、训练数据集版本信息、超参数配置和评估结果一并归档。

9 模型部署、运维与退役数据管理

9.1 推理与监控数据管理

9.1.1 模型部署后，应记录推理请求输入数据摘要、推理结果及时间戳；涉及三级数据、个人信息时，数据摘要不得泄露敏感内容。推理数据摘要保存期限不应短于6个月，且不应短于模型运行周期。
9.1.2 推理数据的采集应符合第6章的规定，推理数据的质量应符合第6.3条的规定。
9.1.3 应对模型的运行状态进行持续监控，监控数据应包含推理延迟、吞吐量、资源利用率和推理准确率、模型输出异常及告警记录。
9.1.4 监控数据的采集频率应根据模型运行的关键程度确定，并应符合数据管理计划的规定。

9.2 数据漂移与模型更新

9.2.1 数据漂移监测应符合第6.3.6条和第6.3.7条的规定。漂移监测结果应纳入模型运维数据管理范围。
9.2.2 模型更新应使用更新后的数据集重新训练。更新后的数据集应满足第6.3条规定的数据质量要求。
9.2.3 模型更新应记录更新原因、更新前后的性能对比和更新审批信息。
9.2.4 模型运维阶段产生的监控数据、漂移监测数据和模型更新记录应定期归档。运维数据归档周期应符合数据管理计划的规定，归档数据应可检索和可恢复。

9.3 退役数据管理

9.3.1 工业产品 AI 分析模型符合以下条件之一时，应启动退役流程：

a) 模型性能持续下降且无法通过重训练恢复至数据管理计划规定的阈值；
b) 模型所服务的工业产品或产线已停产或淘汰；
c) 模型技术架构已不满足当前安全或合规要求；
d) 模型被新版本替代，且新版本稳定运行时间达到数据管理计划规定期限，可启动退役流程。旧模型可设置合理并行过渡期。

9.3.2 模型退役时，应制定数据处理方案，明确数据保留范围、保留期限和销毁方式。数据保留期限应根据数据级别、法律法规和行业监管要求确定，并应在数据管理计划中明确。

9.3.3 超出保留期限的数据应实施安全销毁，销毁可采用数据覆写或物理销毁方式，销毁全过程应留存记录并审计。

9.3.4 模型退役应形成退役记录，记录内容应包含模型标识、模型版本、退役原因、退役日期、数据处理方案执行情况。退役记录应归档保存，保存期限应符合数据管理计划的规定。

10 数据管理能力评估

10.1 评估框架

工业产品 AI 分析模型的数据管理能力评估应按照 GB/T 36073 规定的数据管理能力成熟度评估模型（DCMM）执行。评估范围应覆盖数据战略、数据治理、数据架构、数据应用、数据安全、数据质量、数据标准和数据生存周期等能力域。

10.2 评估指标

10.2.1 数据管理能力评估指标应包括以下方面：

a) 数据管理制度的完备性；
b) 数据管理流程的执行规范性；
c) 数据质量量化指标的达标率；
d) 数据安全措施的落实率；
e) 数据管理工具和平台的覆盖率。

10.2.2 数据管理能力成熟度等级应按照 GB/T 36073 规定的等级标准判定。

10.2.3 数据管理能力成熟度等级不应低于受管理级（L2）。

10.3 评估实施

10.3.1 数据管理能力评估应采用自评估与第三方评估相结合的方式。自评估应由数据管理责任部门组织，第三方评估应由具备相应能力的机构实施。

10.3.2 数据管理能力评估周期不应大于 12 个月。涉及三级数据处理活动的组织，评估周期不应大于 6 个月。当数据管理组织架构、数据管理平台、模型全生命周期管理流程发生重大变更时，应重新组织评估。

10.3.3 评估结果应形成评估报告。评估报告应包含评估范围、评估依据、评估方法、评估结果、不符合项及改进建议。

10.3.4 评估发现的不符合项应制定整改计划，整改完成后应实施复评。

11 验证方法

11.1 总则

11.1.1 验证方法用于验证第 5 章至第 10 章规定要求的符合性。

11.1.2 验证应覆盖数据分类分级、数据采集与接入、数据存储与组织、数据质量、数据安全与合规、模型训练与验证、模型部署与运维、模型退役等过程。

11.1.3 验证应形成验证报告。验证报告应作为数据管理能力评估和模型全生命周期管理的依据。

11.2 验证项目与方法

11.2.1 数据分类分级

a）验证内容：数据分类维度的覆盖度和分级的准确性；
b）验证方法：逐项核对分类清单与 GB/T 43697 分类框架的符合性，核对数据管理计划中定义的数据级别映射关系的一致性；
c）判定依据：分类维度覆盖率应达到 100%；不应出现三级数据被错误降低安全级别的情况。

11.2.2 数据采集与接入

a）验证内容：元数据完整性、完整性校验记录和外部接入安全审查记录；
b）验证方法：检查采集元数据、完整性校验记录和外部接入安全审查记录；
c）判定依据：元数据应完整；校验记录应齐全；安全审查应通过。

11.2.3 数据存储与版本

a）验证内容：版本标识唯一性、备份策略和审计日志；
b）验证方法：检查版本编号规则、备份策略配置和审计日志完整性；
c）判定依据：版本标识应唯一；备份周期应符合数据管理计划；审计日志保存期限符合 6.2 规定。

11.2.4 数据质量

a）验证内容：数据质量评价指标达标情况。
b）验证方法：按照 GB/T 39400 和 GB/T 36344 评价数据质量。
c）判定依据：数据质量指标应满足数据管理计划规定的阈值。

11.2.5 数据安全

a）验证内容：加密、访问控制和审计落实情况。
b）验证方法：检查加密算法合规性、访问控制策略和审计记录。
c）判定依据：应符合 GB/T 43697、GB/T 37988 和 GB/T 22239 规定。

11.2.6 训练与验证

a）验证内容：数据集划分、训练记录和评估报告。
b）验证方法：检查数据集划分记录、训练过程记录和模型评估报告。
c）判定依据：数据集划分比例应符合数据管理计划规定；训练、验证、测试数据集之间不应存在样本泄露。如因工业场景特殊原因存在样本关联，应核查数据管理计划中是否已明确关联判定规则和允许情形，并检查影响评估记录；训练过程记录应完整；安全评估应符合 GB/T 42888 规定；性能评价指标应符合 8.3.2 b）的规定。

11.2.7 部署与运维

a）验证内容：监控数据、漂移监测和更新记录。

b）验证方法：检查监控记录、漂移监测记录和模型更新记录。
c）判定依据：监控频率应符合数据管理计划规定；漂移评估应及时；更新记录应完整。

11.2.8 退役数据处理

a）验证内容：保留期限和销毁记录。
b）验证方法：检查退役数据处理方案、保留期限记录和销毁审计记录。
c）判定依据：保留期限应符合数据管理计划及 9.3.2 的规定；销毁记录应完整。

11.2.9 数据管理能力成熟度

a）验证内容：数据管理能力域的综合成熟度等级。
b）验证方法：按照 GB/T 36073 规定的评估方法开展自评估或第三方评估。
c）判定依据：成熟度等级不应低于受管理级（L2）。

11.2.10 训练过程可复现性

a）验证内容：训练过程记录完整性和复现成功率。
b）验证方法：依据训练过程记录尝试复现训练过程，对比复现结果与原始结果。
c）判定依据：训练过程记录应完整；复现结果与原始结果之间评价指标的允许偏差阈值应预先在数据管理计划中明确，复现指标偏差不应超过该阈值。

11.3 验证结果判定

11.3.1 单项验证项目全部符合第 11.2 条判定依据时，该验证项目判定为通过。
11.3.2 验证关键项目为第 11.2.1 条（数据分类分级）、第 11.2.5 条（数据安全）、第 11.2.8 条（退役数据处理）。关键项目任意一项不通过，则整体验证结论判定为不通过。
11.3.3 非关键项不通过时，应分析原因，制定整改措施并明确整改期限，完成整改后开展重新验证；重新验证通过后，整体验证结论判定为通过。

11.4 验证报告

11.4.1 验证报告应包含以下内容：
a）验证范围；
b）验证依据；
c）验证项目；
d）验证方法；
e）验证结果；
f）不符合项及整改建议；
g）验证结论。
11.4.2 验证报告应经数据管理责任部门、模型开发责任部门共同审批。
11.4.3 验证报告及相关记录应归档保存，保存期限应符合数据管理计划的规定。

参 考 文 献

[1] 工业和信息化部办公厅.工业和信息化部办公厅关于印发《工业数据分类分级指南（试行）》的通知：工信厅信发〔2020〕6号[EB/OL].2020-02-27.