

《国际物流供应链高质量数据集 建设指南》 (征求意见稿) 编制说明

一、标准编制背景

随着人工智能技术，特别是大模型技术的快速发展，人工智能正加速向垂直行业纵深应用延伸。国际物流供应链作为支撑全球贸易和产业安全的重要基础系统，业务链条长、参与主体多、运行环境复杂，数据形态多源异构、跨境流动频繁，对数据的准确性、一致性、时效性与可追溯性要求高。高质量数据集建设能力已成为支撑国际物流供应链大模型训练、推理与应用落地的关键基础。为规范国际物流供应链高质量数据集建设的实施路径、技术方法与过程控制要求，形成可复用、可持续迭代的数据工程方法体系，提升数据集对国际物流供应链典型业务场景（如保供调度、成本预测、跨境合规、异常预警、多式联运应急响应等）的支撑能力，2026年上海市人工智能行业协会下达年度计划，《国际物流供应链高质量数据集 建设指南》正式立项。

本标准由上海市人工智能行业协会提出并归口，由厦门供应链数智创新有限公司牵头并联合上海库帕思科技有限公司、上海人工智能实验室、厦门国贸集团股份有限公司、厦门建发股份有限公司、厦门象屿股份有限公司等单位共同起草。

二、编制过程

本标准的编制主要包括以下几个阶段：

(一) 立项准备阶段（2025年11月-2025年12月）

起草单位围绕国际物流供应链领域高质量数据集建设现状、数据工程能力短板与行业应用需求，系统梳理了跨境贸易、运输组织、多式联运、仓储作业、口岸通关监管、运价成本与风险事件等典型环节的数据类型、数据口径与关键主键标识需求，重点分析了多模态数据处理、跨环节关联、合规安全控制与工程可追溯等建设难点。在此基础上，结合国家关于数据要素、人工智能和高质量数据集建设的政策导向，以及现有国家标准、行业标准和团体标准，明确了本标准以“面向AI模型训练与应用、形成可执行的全流程建设方法”为核心定位，提出“建设路径—工具链支撑—质量评估—安全合规控制”的总体框架，为标准立项和条款设计奠定基础。

（二）标准起草阶段（2025年12月-2026年1月）

起草组在充分调研与论证基础上，围绕国际物流供应链高质量数据集建设全流程，形成“采集—预处理—清洗—标注—版本管理—应用反馈”的建设路径条款，并同步提出建设工具链的核心组件要求、工具链安全规范、质量评估方法与指标要求，以及脱敏、访问控制、审计记录等安全与合规控制要求。标准起草过程中，重点结合国际物流供应链跨境业务、多规则约束、多主体协同的行业特性，对数据标识与主键关联、多模态协同标注、规则类数据结构化、风险事件关联分析等提出针对性实施要点。同时，严格按照 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》的要求，对标准结构与文本表述进行规范化编制，确保条款可理解、可执行、可检查。

（三）立项申请阶段（2026年2月）

在前期调研与文本起草基础上，完成《国际物流供应链高质量数据集 建设指南》（草案）及立项建议书编制，并成功在上海市人工智能行业协会立项。

（四）征求意见阶段（2026年3-4月）

通过公开征集参编单位，组建标准工作组。项目组结合立项答辩中专家意见建议，组织多轮内部讨论会，并采用专家座谈、业务骨干研讨、企业调研等方式，对建设路径、工具链能力、质量评估方法与安全合规控制条款进行修改完善，形成《国际物流供应链高质量数据集 建设指南》（征求意见稿）及编制说明。

三、 编制原则

本标准编制符合以下原则：

（一）规范性原则

本文件符合国家和本市现行法律、法规和规范性文件；并符合GB/T 1.1-2020 的起草要求。

（二）适配性原则

面向人工智能（AI）模型训练与应用需求，围绕国际物流供应链业务特征与数据形态，突出数据集建设全流程的工程化要求，确保建设成果在结构规范、语义一致、可关联、可追溯、可复用等方面满足模型训练、推理与评测需要。

（三）可操作性原则

本标准提出的国际物流供应链高质量数据集建设立足行业实际，结合国际物流供应链业务链条长、参与主体多、数据多源异构与跨环节关联强等特点，重点从数据内容范围、采集与标识、预处理与

清洗、标注与版本管理、质量评估以及安全与合规控制等方面提出可执行的技术方法与过程要求，形成可落地、可检查、可复现的建设指引，为国际物流供应链AI模型训练、推理应用与评测提供具备工程可操作性的建设依据。

四、 主要内容

本文件规定了国际物流供应链高质量数据集建设的实施路径、技术方法和过程控制要求，重点包括建设路径、数据采集、预处理、清洗、标注、版本管理、应用反馈与持续优化，以及建设工具链、质量评估与安全合规控制等内容。

本文件适用于国际物流供应链高质量数据集建设活动的具体实施与过程管理，可为国际物流供应链AI模型训练、微调、推理应用与评测提供可执行的数据工程方法依据。

（一）规范性引用文件

GB/T 41867-2022 信息技术 人工智能 术语

（二）术语和定义

GB/T 41867—2022界定的以及下列术语和定义适用于本文件。

高质量数据集建设 **data construction**

围绕人工智能模型训练与应用需求，对原始数据开展采集、预处理、清洗、标注、质量评估与版本管理等活动，使其形成结构规范、语义清晰、质量可控、可复用的数据集资产的过程。

原始数据 **raw data**

未经系统化处理或仅经过初步采集的国际物流供应链相关数据，具有多模态、非结构化和未标准化等特点。主要包括政策文本、业

务记录、轨迹信息、单证影像等内容，尚未经过规范化清洗与标注处理。

建设工具链 dataset construction toolchain

支撑高质量数据集建设过程中数据采集、预处理、清洗、标注、质量评估等环节的集成化工具集合，具备多模态处理、人机协同和过程追溯等能力。

数据质量评估 data quality assessment

对建设形成的数据集在完整性、一致性、准确性、适用性等方面进行系统性检查与综合评价，以确认其是否满足高质量数据集相关要求。

（三）概述

明确国际物流供应链高质量数据集建设目标，提出面向模型适配、场景导向、人机协同、全流程可追溯、安全合规嵌入、全链路一致性等基本原则。

（四）高质量数据集建设要求

规定建设路径与关键环节要求，包括：数据采集（范围、标识与主键关联、模态类型）、数据预处理（格式统一、非结构化处理如OCR/ASR等）、数据清洗（一般要求、逻辑一致性校验、数据脱敏）、数据标注（流程与用途差异化）、数据入库与版本管理（入库要求、版本控制）、应用反馈与持续优化（模型验证反馈、持续改进机制）。

（五）建设工具链

规定建设工具链定义、核心工具组件（多模态协同标注、实时

监控与质量管控、数据管理与审计、国际物流供应链专用扩展工具)及工具链安全规范要求。

(六) 质量评估

参考相关国际标准框架，结合国际物流供应链数据特征，提出以准确性、一致性、适用性为核心的质量评估指标与实施方法，支撑建设成果可验证、可解释。

(七) 安全与合规控制

提出数据脱敏、访问控制、审计与留痕等全过程控制要求，保障数据建设活动合法合规、安全可控、过程可追溯。

五、 重大分歧意见的处理结果及理由

本标准在编制过程中无重大分歧意见。

六、 标准作为强制性或推荐性标准发布的意见

推荐性。

七、 推动标准实施的措施建议

本标准发布后将尽快组织宣贯，加大贯彻实施力度。第一，在适用主体中推广应用该标准，形成经验；第二，根据试点经验，推动标准更新实践，复制推广试点经验和标准适用主体范围扩大；第三，广泛收集意见和建议，及时归纳和总结，并不断完善标准，必要时提出标准修订。

八、 其他应予以说明的事项

无。

标准编制组
2026年3月5日