

T/SAIAS

上海市人工智能行业协会团体标准

T/SAIAS XXX—2026

国际物流供应链高质量数据集通用要求

Guidelines for High-Quality Dataset Framework in
International Logistics Supply Chains

(征求意见稿)

2026 - XX - XX 发布

2026 - XX - XX 实施

目 次

前 言	III
1 范围	4
2 规范性引用文件	4
3 术语和定义	4
3.1	4
4 国际物流供应链数据集分类	4
4.1 内容分类	4
4.1.1 政策规制类	4
4.1.2 政务监管类	4
4.1.3 企业经营类	4
4.1.4 物流动态类	5
4.1.5 行业经验类	5
4.2 用途分类	5
4.2.1 预训练数据集	5
4.2.2 监督微调数据集	5
4.2.3 思维链数据集	5
4.2.4 评测数据集	5
5 数据质量与合规要求	6
5.1 数据属性要求	6
5.1.1 业务覆盖要求	6
5.1.2 完整性要求	6
5.1.3 关联性要求	6
5.2 数据质量要求	6
5.2.1 准确性要求	6
5.2.2 一致性要求	6
5.2.3 可追溯性要求	6
5.3 数据合规与安全要求	6
5.3.1 合法合规要求	6
5.3.2 跨境合规要求	7
5.3.3 敏感信息保护要求	7
5.3.4 访问与使用控制要求	7
5.4 数据技术要求	7
5.4.1 传输与交换要求	7
5.4.2 存储与管理要求	7
5.4.3 数据格式与兼容性要求	7
5.4.4 元数据与标识要求	7
6 数据集建设要求	7
6.1.1 覆盖全面性要求	7

6.1.2 结构系统性要求.....	7
6.1.3 知识层级与用途协同要求.....	7
6.1.4 表达与语义一致性要求.....	7
6.1.5 AI 适配性要求.....	8
7 数据集管理机制.....	8
7.1.1 数据集存储与组织机制.....	8
7.1.2 元数据与数据字典管理机制.....	8
7.1.3 权限与角色管理机制.....	8
7.1.4 更新与持续迭代机制.....	8
7.1.5 与模型训练和应用的联动机制.....	8
参 考 文 献.....	12

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这些专利的责任。

本文件由上海市人工智能行业协会提出并归口。

本文件起草单位：

本文件主要起草人：

本标准首次制定。

首期执行单位：

本文件版权归上海市人工智能行业协会所有。未经许可，不得擅自复制、转载、抄袭、改编、汇编、翻译或将本标准用于其他任何商业目的。

国际物流供应链高质量数据集通用要求

1 范围

本文件规定了国际物流供应链高质量数据集的通用要求，包括数据集分类、数据质量与合规要求、数据集建设要求以及数据集管理机制等内容。

本文件适用于国际物流供应链领域高质量数据集的规划设计、采集加工、质量控制、交付验收与运行管理，可为人工智能模型训练、微调、评测及相关智能应用提供数据建设依据。

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 41867-2022 信息技术 人工智能 术语
TC609-5-2025-01 高质量数据集 建设指南
TC609-5-2025-03 高质量数据集 分类指南

3 术语和定义

GB/T 41867-2022、TC609-5-2025-01和TC609-5-2025-03界定的以及下列术语和定义适用于本文件。

3.1

国际物流供应链高质量数据集 High-Quality Dataset of International Supply Chains

面向国际物流供应链相关业务、规则与场景，按照统一规范对多源异构数据进行系统采集、治理、标注和质量控制后形成的，可用于人工智能训练、评测和应用的数据资产集合，用于支撑国际物流供应链韧性与安全相关的智能分析、决策支持和风险防控等应用。

4 国际物流供应链数据集分类

4.1 内容分类

4.1.1 政策规制类

政策规制类数据是指反映国际物流供应链运行所依据的国际规则、法律制度和外部约束条件的数据，包括国际公约、贸易与通法规规、制裁清单、关税税则、港口公告、运价指数等公开发布的信息，用于支撑跨境贸易合规判断、政策解读和外部风险分析。

4.1.2 政务监管类

政务监管类数据是指政府及行业监管机构在对国际物流供应链活动实施管理、审核与执法过程中形成的数据，包括海关、口岸、检验检疫、交通运输和港口管理等部门产生的业务记录、监管状态和执法信息，用于反映供应链运行的合规状态和监管约束。

4.1.3 企业经营类

企业经营类数据是指供应链各类经营主体在实际运营和交易过程中形成的数据，包括航运公司、港口、铁路、物流企业、货代及平台在订舱、运输、仓储、结算等环节产生的业务记录，用于刻画供应链运行效率、成本结构和经营绩效。

4.1.4 履约执行类

履约执行类数据是指反映国际物流供应链各参与主体在合同或业务约定下，实际履行运输、装卸、仓储和交付等义务过程及其状态的数据，包括船舶、车辆、集装箱和港口设施在运行过程中形成的位置、轨迹、作业状态以及相关环境感知信息，用于支撑实时监测、态势感知和调度优化。

4.1.5 行业经验类

行业经验类数据是指反映国际物流供应链运行逻辑、专业判断和行业知识的数据，包括规则解读、业务流程说明、风险识别方法、应急处置方案、案例分析及专家标注形成的知识型数据，用于支撑模型理解业务语义、规则推理和复杂场景决策。

4.2 用途分类

4.2.1 预训练数据集

预训练数据集是指面向大模型预训练阶段构建的国际物流供应链数据集，主要用于为模型注入基础行业知识和通识认知，包括政策规则、业务流程、行业概念、基础事实和公开信息等内容，为后续行业能力构建提供稳定、广泛和可靠的知识底座。

4.2.2 监督微调数据集

监督微调数据集是指面向模型监督微调阶段构建的行业任务型数据集，主要用于将基石语料中的行业知识转化为可执行的问答、分析和推理能力，包括典型业务问题、标准答案、操作指引和场景化对话数据，使模型具备面向实际业务的专业响应能力。

4.2.3 思维链数据集

思维链数据集是指面向模型复杂推理与决策能力训练的行业思维链数据集，主要包含跨规则、跨流程、跨场景的推理路径、决策逻辑和专家判断过程，用于引导模型学习国际物流供应链中的合规判断、风险分析、路径选择和应急处置等高阶认知能力。

4.2.4 评测数据集

评测数据集是指用于检验和衡量模型在国际物流供应链场景中表现能力的数据集，围绕真实业务场景、关键风险点和动态变化环境构建测试样本，用于评估模型在理解能力、推理能力、准确性和稳定性等方面的综合表现，并作为模型优化和迭代的重要依据。

4.3 模态分类

4.3.1 文本数据集

文本数据是指以自然语言文本形式表达的国际物流供应链相关数据，用于承载政策规则、业务说明和行业知识，是模型理解业务语义和规则逻辑的基础数据形态。文本数据包括但不限于政策法规文件、通关与贸易规则文本、业务流程说明、合同与条款文本、公告通知、行业报告、操作手册、案例材料、专家说明文本及问答类语料等。文本数据主要用于支撑模型的行业知识注入、语义理解和规则推理能力构建。

4.3.2 表格数据集

表格数据是指以表格或数据库形式组织的结构化国际物流供应链业务数据，用于刻画供应链运行状态、业务指标和统计特征。表格数据包括但不限于舱单与运单数据、通关申报数据、订单与结算数据、运价与费用数据、库存与仓储数据、港口作业统计数据、企业经营指标等。表格数据主要用于支撑模型的事实理解、数值分析、状态识别和趋势分析能力。

4.3.3 图像数据集

图像数据是指以静态图像形式表达国际物流供应链相关对象、要素或场景信息的数据，用于反映货物、单证、设施设备和作业环境等视觉信息。图像数据包括但不限于货物与集装箱图像、现场与作业图

像、设备状态图像、单证与票据影像以及其他与供应链业务相关的静态图像。图像数据主要用于支撑模型对目标识别、状态判读、单证理解和场景态势分析等能力。

4.3.4 视频数据

视频数据是指以连续影像形式表达国际物流供应链相关过程、活动或场景变化的数据，用于描述业务过程和运行状态的动态信息。视频数据包括但不限于作业过程视频、运输节点监控视频、口岸与园区监控视频以及其他与供应链业务相关的视频数据。视频数据主要用于支撑模型对过程理解、异常识别和动态态势分析等能力。

4.3.5 时序数据

时序数据是指按时间顺序连续记录国际物流供应链相关对象、过程或状态变化的数据，用于刻画业务运行的动态特征。时序数据包括但不限于运输载体轨迹与状态数据、作业时序数据、节点状态数据、市场指标时间序列数据以及环境监测数据等。时序数据主要用于支撑模型对趋势分析、异常变化识别和风险演化研判等能力。

4.3.6 多模态复合数据

多模态复合数据是指由文本、表格、图像、视频、时序等多种模态数据按照统一业务对象或场景进行关联和组织形成的数据集合。该类数据用于表达国际物流供应链复杂业务场景中不同信息形态之间的关联关系，支撑模型对跨模态信息的综合理解与联合推理。多模态复合数据是构建面向复杂场景的智能分析和决策能力的重要数据形态。

5 数据质量与合规要求

5.1 数据属性要求

5.1.1 业务覆盖要求

数据集应覆盖跨境贸易与物流活动的主要业务环节与运行要素，能够反映国际物流供应链的基本结构与运行状态。

5.1.2 完整性要求

数据集应包含具有代表性的业务数据和环境数据，避免仅由零散或孤立的数据片段构成。

5.1.3 关联性要求

不同数据之间应具备可识别的关联关系，能够支持对业务过程、状态变化及结果之间关系的描述与分析。

5.2 数据质量要求

5.2.1 准确性要求

数据应真实、准确，不得包含明显错误、系统性偏差或大规模缺失内容。

5.2.2 一致性要求

涉及数量、金额、时间、地点、货物属性等关键字段的数据，应与原始业务记录或权威来源保持一致；同时，来自不同数据源的同一业务对象在上述关键信息上的描述应相互对齐，不得存在大量冲突或无法关联的情形。

5.2.3 可追溯性要求

数据应具备可追溯性，能够识别其来源、采集时间及处理过程，以支撑质量评估与责任界定。

5.3 数据合规与安全要求

5.3.1 合法合规要求

数据应来源合法、授权清晰，不得包含违反数据保护、隐私保护、商业秘密保护或国家安全要求的内容。

5.3.2 跨境合规要求

涉及跨境流动的数据，应符合相关国家和地区关于数据出境与数据主权的法律法规要求。

5.3.3 敏感信息保护要求

对涉及个人信息、企业敏感信息或重要资源信息的数据，应在进入数据集前采取去标识化、脱敏或其他必要处理措施。

5.3.4 访问与使用控制要求

数据应具备基本的访问控制与使用管理能力，防止未经授权的数据访问、复制或传播。

5.4 数据技术要求

5.4.1 传输与交换要求

数据在提供、传输和交换过程中，应采取必要的技术措施，保障数据的安全性、完整性和可用性。

5.4.2 存储与管理要求

数据应存储在具备可靠性和可扩展性的系统中，支持数据的长期保存、快速调用和完整性校验。

5.4.3 数据格式与兼容性要求

数据存储与交换格式应符合国家相关标准和行业规范，优先采用通用性强、兼容性高的数据格式。

5.4.4 元数据与标识要求

数据应配套必要的元数据信息，用于标识数据来源、时间范围、数据类型及适用范围，支撑数据管理与使用控制。

6 数据集建设要求

6.1.1 覆盖全面性要求

国际物流供应链高质量数据集应系统反映跨境贸易与国际物流活动的主要运行过程和基本结构，覆盖从贸易规则、订单履约、运输组织、通关监管到结算与风险处置等关键链路。

6.1.2 结构系统性要求

数据集不应由零散、孤立的数据片段构成，而应具备对国际物流供应链整体运行状态的刻画能力，支持对供应链结构、运行模式及变化趋势的系统性认知。

6.1.3 知识层级与用途协同要求

数据集应按照通识、行业通识与行业专识的知识层级，以及预训练、微调与评测等用途类型进行体系化组织，使不同层级和用途的数据之间形成清晰的结构关系。

- 用于基础认知构建的数据，应具备广覆盖与稳定性；
- 用于场景能力训练的数据，应体现业务逻辑与操作语义；
- 用于模型评测的数据，应能够反映真实业务复杂性与风险边界。

6.1.4 表达与语义一致性要求

数据集应在字段口径、概念定义、单位表达和语义边界方面保持一致性，使同一业务对象在不同数据来源和不同数据形态中的表达具有可对齐性。

涉及贸易主体、货物、运输工具、时间、地点等关键对象的数据，应具备统一或可映射的标识体系，以支撑跨数据源关联与模型学习。

6.1.5 AI 适配性要求

国际物流供应链高质量数据集应满足面向人工智能训练、微调和评测的基本适配要求，包括：

- 数据内容应具有明确语义和业务指向，避免难以解释的杂散信息；
- 数据结构与表达应支持机器读取、解析与关联；
- 数据应能够支撑模型对规则、流程、因果关系和决策逻辑的学习。

7 数据集管理机制

7.1.1 数据集存储与组织机制

国际物流供应链高质量数据集应以数据集为基本管理单元进行组织，可以按照知识层级、业务领域和用途类型进行分区管理。

数据集应具备版本标识和边界定义，不同版本之间应可追溯，以支撑模型训练、评测和结果复现。

7.1.2 元数据与数据字典管理机制

每个数据集应配套完整的元数据描述，包括但不限于数据来源、时间范围、业务范围、数据类型、使用限制和处理说明。

数据字段应建立统一的数据字典和口径说明，使不同数据集之间在概念和字段层面具备可理解性与可对齐性。

7.1.3 权限与角色管理机制

数据集的访问和使用应基于角色和权限进行管理，区分建设者、管理者、使用者与审计者等不同角色。

不同分类分级的数据应对应不同的访问、下载、共享和调用权限，以保障数据安全与合规使用。

7.1.4 更新与持续迭代机制

国际物流供应链高质量数据集应建立持续更新与迭代机制。

——对反映市场变化、运输状态、风险事件等高时效性数据，应保持必要的现势性；

——对规则、知识和经验类数据，应支持随业务和政策变化进行补充与修订。

7.1.5 与模型训练和应用的联动机制

数据集的建设与管理应与模型训练、评测和智能体应用保持联动。

模型训练效果、应用表现和业务反馈应作为评估数据集适用性和改进方向的重要依据，推动形成“数据—模型—应用”协同演进的良性循环。

附 录 A
(资料性)
国际物流供应链学科知识体系（基于中图分类法）

一级	二级（业务大类）	三级（业务子项分类）	备注	
经济	经济学	经济预测与分析	国际物流与供应链运行趋势、运价指数分析、供需结构变化、全球贸易周期波动等知识。	
		产业经济学	物流业、航运业、供应链服务业的产业结构、产业链分工及演化规律。	
	世界经济	世界经济、国际经济关系	全球贸易体系、国际分工格局、全球供应链布局与重构相关知识。	
		世界经济贸易概况	各国（地区）进出口结构、贸易流向及全球物流通道分布。	
	交通运输经济	运输经济理论	国际运输成本、运输效率、运输价格形成机制等基础经济规律。	
		运输组织与管理	国际物流组织方式、多式联运组织模式及跨境运输协调机制。	
		国际运输经济	国际航运、铁路、公路、航空货运的经济运行与市场结构。	
	贸易经济	国际贸易	国际货物贸易流程、贸易方式及其与物流活动的关系。	
		对外贸易管理	进出口管理制度、贸易管制、配额及许可制度相关知识。	
		国际贸易实务	贸易条款（如交付条件）、国际物流单证与责任划分规则。	
	财政、金融	金融学	汇率机制、国际结算体系对物流供应链的影响。	
		国际金融	跨境支付、外汇风险及国际资本流动对供应链稳定性的影响。	
		贸易金融	信用证、保理、供应链金融等与国际物流密切相关的金融知识。	
	交通运输	综合运输	综合运输管理	国际综合交通体系、运输方式协同及综合物流网络组织。
			多式联运	海铁联运、公铁联运、江海联运等跨方式运输组织规则。
公路运输		公路运输组织	国际公路运输体系及跨境道路运输组织模式。	
		国际公路运输	国际道路运输公约及跨境公路运输管理规则。	

	铁路运输	国际公路运输	国际铁路货运组织方式及跨境运输协调机制。
		国际铁路运输	中欧班列等国际铁路通道运行与管理知识。
	水路运输	水路运输管理	国际航运管理体制、航线组织及航运管理规则。
		海洋运输	远洋航运、干散货运输、集装箱运输体系相关知识。
		港口工程与港口管理	港口作业流程、港口组织、港口运行与管理规则。
	航空运输	航空运输管理	国际航空货运体系、航空物流组织模式。
国际航空运输		国际航空运输规则及航空货运网络运行知识。	
政治、法律	法律	国际法	国际公约及国际规则对物流与贸易活动的约束。
		国际经济法	国际贸易法、国际投资法及跨境物流法律框架。
		民商法	国际物流合同、责任认定、货损与索赔规则。
		海商法	海上货物运输合同、提单制度及海事责任规则。
		经济法	物流与贸易活动中的市场监管与合规要求。
	外交、国际关系	国际关系	地缘政治变化对国际物流通道和供应链稳定性的影响。
国际组织		WTO、IMO、ICAO 等国际组织制定的物流与运输规则。	
环境科学	安全科学	安全管理	国际物流运行安全、运输安全与系统安全管理。
		风险管理	供应链中断风险、运输风险及系统性风险识别与控制。
		应急管理	突发事件下国际物流与供应链的应急响应与恢复机制。
自动化技术	计算技术	数据处理技术	物流数据处理、数据交换及信息标准化相关知识。
		信息系统	物流信息系统、供应链管理系统的结构与运行机制。

	自动化技术	自动化系统	智能物流系统、自动化调度与控制技术。
自然科学总论	系统科学	系统工程	供应链系统建模、系统优化与系统分析方法。
		复杂系统	国际物流供应链的复杂网络结构及演化规律。
地球科学	海洋学	海洋运输环境	海洋气象、航道条件对国际航运运行的影响。
	气象学	应用气象	天气、气候变化对国际物流运行的影响。
历史、地理	地理	世界地理	全球港口、物流节点及国际通道的空间分布格局。

参 考 文 献

- [1] 中国分类主题词表（第二版）
 - [2] 中国图书馆分类法
 - [3] 中国档案分类法
 - [4] GB/T 35295-2017 信息技术 大数据 术语
 - [5] JT/T 1522-2024 交通运输数据安全分级和保护要求
-