

《国际物流供应链大模型评测指南》 (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，大模型能力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，大模型在该领域的应用正从知识问答、信息抽取等基础能力，逐步走向风险识别预警、决策辅助与协同调度等高价值场景，对模型的专业认知能力、风险控制能力以及安全与价值对齐能力提出了更高要求。

与此同时，国际物流供应链大模型的应用效果评估缺乏统一框架与可复现方法，不同机构在评测维度、数据集构建、实施流程与结果解释方面存在差异，难以形成可比较、可审计的评测结论。为建立面向国际物流供应链场景的大模型评测框架体系，明确评测维度、评测内容与实施方法，提升评测工作的可操作性、可复现性与可解释性，推动国际物流供应链大模型安全、可靠、可控应用落地，2026年上海市人工智能行业协会下达年度计划，《国际物流供应链大模型评测指南》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大模型应用现状与评测需求，系统梳理了国际物流供应链典型业务场景与关键能力要求，重点分析了模型在通用基础能力、专业认知能力、风险识别与决策支持能力以及安全与价值对齐方面的评测难点。结合国际物流供应链数据敏感性强、场景复杂度高、评价口径易分散等特点，明确了本标准以“建立可复用的评测框架与可落地的评测方法”为核心定位，提出“评测框架—评测内容—评测方法—结果评估与等级划分”的总体思路，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，参考 GB/T 45288.1—2025 与 GB/T 45288.2—2025 等相关标准，构建国际物流供应链大模型评测框架体系，形成模型通用基础能力、国际物流供应链安全与价值对齐、国际物流供应链风险控制、国际物流供应链专业认知能力、国际物流供应链业务辅助拓展能力等评测维度，并明确各维度评测内容与覆盖范围。同步提出评测方法要求，包括评测数据集构建、评测环境与工具要求、自动化与人工评测实施流程以及评测结果评估与等级划分建议。起草过程中，重点结合国际物流供应链跨境规则多、风险类型多、业务决策约束强的行业特性，强化对风险识别预警、情景分析与多方案建议、合规与公共利益导向等评测要点的可检验表述，确保条款可理解、可执行、可复现。

（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《国际物流供应链大模型评测指南》（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，对评测维度设置、评测数据集要求、评测实施流程以及评分与等级划分等条款进行修改完善，形成《国际物流供应链大模型评测指南》（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）大模型在国际物流供应链领域的应用需求，结合跨境业务、多主体协同与多约束决策的行业特点，构建覆盖通用能力、专业能力、风险能力与安全对齐能力的评测体系，确保评测内容与国际物流供应链典型业务场景和关键能力要求相匹配。

（三）可操作性原则

本标准立足行业实际，围绕评测对象、评测维度、评测数据集、评测环境与工具、自动化与人工评测实施流程以及评分与等级划分

等关键环节提出可执行要求，强调评测过程可复现、评测结果可解释、评测结论可对比，为国际物流供应链大模型选型、验收、迭代优化与应用落地提供具备工程可操作性的评测依据。

四、 主要内容

本文件确立了国际物流供应链大模型评测的框架体系，包括评测维度及评测内容，描述了相关评测方法。

本文件适用于国际物流供应链大模型应用效果评测。

（一）规范性引用文件

GB/T 45288.1-2025 人工智能 大模型 第1部分：通用要求

GB/T 45288.2-2025 人工智能 大模型 第2部分：评测指标与方法

（二）术语和定义

GB/T 45288.1-2025和GB/T 45288.2-2025界定的以及下列术语和定义适用于本文件。

国际物流供应链大模型 **large model for international logistics supply chain**

在通用基础大模型的基础上，结合国际物流供应链领域的专业知识和场景数据进行训练所形成的大模型，具备理解和分析国际物流供应链典型业务场景，提供国际物流供应链领域智能分析与辅助决策支持等能力。

单模态维度 **Monomodal dimension**

单模态维度主要包括文本、图像、音频3个二级维度。

多模态维度 **multimodal dimension**

多模态维度主要包括图文、文音、图音、图文音4个二级维度。

（三）概述

构建国际物流供应链大模型评测框架体系，明确模型评测的总体结构与主要评测维度，为后续评测内容与方法提供统一框架。

（四）评测内容

规定五类评测维度的内容范围与能力要求，包括：模型通用基础能力、国际物流供应链安全与价值对齐、国际物流供应链风险控制、国际物流供应链专业认知能力、国际物流供应链业务辅助拓展能力。

（五）评测方法

提出评测数据集构建要求（合规与密级、指标完备、时效性、可用性、多样性与代表性），明确评测环境（软硬件、部署方式、算力配置）、评测工具（自动化与人工评测功能）、评测实施流程（自动化与人工评测要求）以及评测结果评估方法。

（六）附录A：评分与等级划分示例

给出评分规则与等级划分的设置示例参考，支撑评测结论可对比、可解释、可审计。

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