ICS

T/GXDSL

团　　体　　标　　准

T/GXDSL 212—2025

工业人工智能（AI）模型开发与部署管理规范

Industrial Artificial Intelligence (AI) Model Development and Deployment Management Specification

征求意见稿

2025 - - 发布　　2025 - - 实施

广西电子商务企业联合会　发布

目　　次

前　言

本文件依据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。

本文件由广西产学研科学研究院提出。

本文件由广西电子商务企业联合会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

工业人工智能（AI）模型开发与部署管理规范

一、引言

工业人工智能是新一代工业革命的核心驱动力，通过将人工智能技术与工业知识深度融合，正在重塑制造业的研发、生产、管理、服务全价值链。然而，工业场景具有高复杂性、高可靠性要求、强环境制约等特点，通用的人工智能模型开发与部署方法难以直接满足工业应用需求。当前工业 AI 应用中普遍存在模型开发流程不规范、数据质量参差不齐、模型性能不稳定、部署环境不兼容、安全伦理风险突出等问题，严重制约了工业 AI 从实验验证走向规模化应用和价值创造。为系统性地解决这些问题，推动工业 AI 技术健康、有序、可信地应用于工业生产实践，亟需建立一套覆盖模型全生命周期、兼顾技术与管理、符合工业领域特殊要求的开发与部署管理规范。本规范旨在为工业 AI 项目提供标准化的实施框架和操作指引，确保模型从概念设计到退役下线的全过程均处于受控状态，从而实现模型性能可靠、部署安全、管理高效、结果可信。本规范由广西产学研科学研究院联合工业制造企业、人工智能技术供应商、科研院所及行业组织共同研制，立足于我国工业智能化转型的实际需求，力求为各类工业 AI 项目的实施提供科学依据和质量保障，促进工业 AI 产业生态的规范化和成熟化发展。

本规范确立了工业人工智能模型从需求分析、数据准备、模型开发、验证测试、部署上线到运维监控、版本更新及退役下线的全生命周期管理要求。本规范适用于所有在工业制造、能源电力、交通运输、资源开采等工业领域进行人工智能模型开发、部署、运维及相关管理活动的组织与个人，包括但不限于工业企业的人工智能技术部门（如广西产学研科学研究院下属的智能制造实验室）、第三方人工智能服务提供商、系统集成商以及相关软硬件供应商。涉及的技术类型包括机器学习（特别是深度学习）、计算机视觉、自然语言处理、预测性维护、工艺优化、质量控制、智能调度等典型工业 AI 应用。本规范的制定严格遵循国家《新一代人工智能发展规划》、《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》等战略部署，并参考了国际国内在人工智能治理、数据安全、工业软件等方面的法律法规与标准体系，旨在促进工业 AI 技术应用的安全可控和健康发展。

二、规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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 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

GB/T 37988-2019 信息安全技术 数据安全能力成熟度模型

GB/T 39335-2020 信息安全技术 个人信息安全影响评估指南

GB/T 40685-2021 信息技术服务 数字化营销服务 数据治理要求

GB/T 41867-2022 信息技术 人工智能 机器学习模型及系统的质量要素和评测方法

GB/T 42454-2023 信息技术 人工智能 面向过程的质量管理指南

《工业数据分类分级指南（试行）》（工信厅信发〔2020〕107 号）

《网络安全标准实践指南—人工智能伦理安全风险防范指引》

《人工智能研发运营一体化（MLOps）能力成熟度模型》

ISO/IEC 22989:2022 Information technology — Artificial intelligence — AI concepts and terminology

ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)

三、总体原则与治理框架

工业人工智能模型的开发与部署管理，应立足于工业系统的本质特征，遵循“安全可靠、性能为先、融合知识、持续演化”的核心原则。首要原则是安全性与可靠性至上原则。工业系统对安全、连续、稳定运行的要求极高，任何因 AI 模型引入的误判、失效或异常行为都可能导致生产中断、设备损坏、质量事故甚至安全事故。因此，必须将功能安全、信息安全、运行可靠性作为模型开发和部署的刚性约束。模型在设计之初就应考虑失效模式与容错机制，确保在输入异常、网络延迟、硬件故障等情况下具备确定的、可预测的降级处理或安全停车能力。所有涉及工业控制回路或安全联锁的 AI 模型，其安全完整性等级应遵循相关标准进行评估与设计。

其次是领域知识与数据双轮驱动原则。工业 AI 的成功不仅依赖于海量数据，更离不开深刻的领域知识（如工艺机理、设备原理、专家经验）。模型开发过程必须是与领域专家深度协同的过程。应将物理规律、业务规则、约束条件等先验知识以可解释、可验证的方式融入模型架构、损失函数或后处理逻辑中，发展“知识嵌入”或“物理信息”的 AI 模型，提升模型的外推性、可解释性和对稀缺数据场景的适应能力。数据作为驱动要素，其采集、标注、治理必须遵循工业级质量标准，确保数据的代表性、准确性、一致性和时效性。

第三是全生命周期管理与持续运营原则。工业 AI 模型并非一次性开发交付的静态软件，其性能会随着设备磨损、工艺调整、原料变化、环境变迁而发生衰减。必须借鉴并强化 DevOps 理念，建立覆盖开发、部署、监控、再训练的工业 AI 运维（MLOps）体系。这意味着从项目启动就需规划模型的持续监控指标、更新触发机制、版本管理策略和回滚方案，确保模型在产线全生命周期的性能维持与适应进化。管理活动应贯穿需求定义、数据工程、模型实验、部署发布、线上监控、模型重训的每一个环节，并形成标准化的文档与知识沉淀。

第四是透明可信与合规可控原则。工业应用要求决策过程可追溯、可审计。在不完全依赖“黑箱”模型做出关键决策的同时，应着力提升模型的可解释性，开发并提供对模型预测结果的归因分析、置信度评估及不确定性量化。模型的开发、部署和运行必须严格遵守国家在数据安全、网络安全、个人信息保护、算法治理等方面的法律法规，建立覆盖数据采集、标注、使用、传输、存储全流程的安全管理制度。涉及跨境数据流动或使用开源组件时，需进行专门的安全风险评估。

为实现上述原则，组织应建立相适应的工业 AI 治理框架。该框架至少应包括：一个明确的治理机构（如工业 AI 伦理与安全委员会），负责制定战略、审批重大项目、评估风险、监督合规；一套覆盖模型全生命周期的管理制度与流程文档；一套与现有工业自动化体系（如 MES、SCADA）和 IT 管理体系（如 ISO 27001）相融合的技术标准与操作规范；以及一支具备交叉学科知识（人工智能、工业工程、自动化、网络安全）的专业团队。广西产学研科学研究院等研发机构应在框架中承担技术标准制定、共性平台研发与高级人才培养的职责。

四、开发阶段管理要求

开发阶段是决定模型质量与适用性的基础，必须实施严格的流程控制与技术评审。需求分析与可行性论证是首要环节。应组建跨职能团队，清晰定义业务目标、成功指标（如将某类缺陷检出率从 95%提升至 99.5%）、性能约束（如单次推理耗时不超过 100 毫秒）及非功能性需求（如可解释性要求、硬件资源限制）。必须进行全面的可行性分析，评估数据可获取性、技术路径成熟度、与现有系统集成复杂度、预期投资回报率及潜在风险，形成详尽的需求规格说明书与项目计划。

数据工程管理是工业 AI 项目的基石。依据《工业数据分类分级指南》，对训练数据进行分类分级管理。数据采集方案需确保覆盖设备全工况、工艺全流程、产品全系列，样本分布应具有代表性，时间

序列数据需保证时间一致性。数据预处理与标注过程必须建立质量控制点，标注规则需由领域专家审定，标注一致性应达到95%以上。应构建版本化的数据仓库或特征库，对原始数据、清洗后数据、特征数据实施严格的版本控制和访问权限管理。训练集、验证集、测试集的划分必须科学，防止数据穿越，确保模型评估的公正性。对于涉及个人信息的数据，处理活动需符合GB/T 39335等相关要求。

模型设计与训练应遵循结构化、可复现的流程。鼓励基于共享的模型实验管理平台（如广西产学研科学研究院可提供的工业AI开发平台）进行，记录每一次实验的超参数、代码版本、数据版本、环境配置及结果指标，确保实验的可复现性。模型架构选择需兼顾性能与效率，优先考虑适合边缘部署的轻量化模型。训练过程中需监控损失曲线、验证集指标，防止过拟合或欠拟合。应采用交叉验证等方法稳健评估模型性能。对于关键模型，应进行鲁棒性测试，验证其对噪声、对抗样本、输入分布偏移的抵抗能力。

验证与测试是模型交付前的关键质量 gate。必须建立独立于开发团队的验证测试流程。测试应包括：离线性能测试，在预留的测试集上全面评估精度、召回率、F1分数、均方误差等核心指标，并与需求定义的性能基准进行比对；离线业务逻辑测试，验证模型输出是否符合工业常识和业务规则；仿真环境测试，将模型置于高度模拟真实生产环境的数字孪生或仿真系统中，测试其在动态工况下的综合表现；此外，还需进行安全性测试（如对抗攻击测试）、公平性测试（如检查对不同批次、不同设备数据是否存在歧视性偏差）及资源消耗测试。所有测试需形成报告，并由项目负责人、领域专家和质量管理人员共同评审通过。

五、部署与运维阶段管理要求

模型通过验证后，进入部署与运维阶段，此阶段直接关系到模型价值的实现与持续。部署发布管理需制定详尽的部署方案。方案需明确部署环境（云端、边缘端、工控机）、硬件资源配置、软件依赖关系、与现有系统（如PLC、DCS、MES）的接口协议与数据流。部署过程应采用容器化等标准化技术，确保环境一致性。必须建立严格的模型版本管理制度，每个上线模型应有唯一的版本号，并关联其对应的代码、数据、文档及测试报告。部署前需在准生产环境进行最终验证。生产部署应采用蓝绿部署或金丝雀发布等策略，逐步扩大流量，密切监控初期表现，并预设快速、可靠的回滚机制，确保生产系统稳定性。

线上监控与性能管理是模型上线后的常态化工作。必须建立全面的模型监控体系，监控指标至少应包括：业务指标（如模型决策带来的实际效益提升）、模型性能指标（如在线推理的准确率、响应时间、吞吐量）、系统资源指标（如CPU/GPU/内存使用率）以及数据健康度指标（如输入数据分布与训练数据分布的偏移度）。建议对关键模型设置性能衰减预警阈值，例如，当在线准确率持续低于离线测试值的3个百分点，或输入数据分布偏移度（如通过PSI指数计算）超过0.25时，应触发告警。所有监控数据应可视化，并支持历史回溯与分析。

模型更新与重新训练管理是模型持续生命力的保障。应建立模型性能衰减的定期评估机制（如每月评估一次）。当触发更新条件（如监控告警、工艺重大变更、积累足够新数据）时，需启动模型更新流程。更新过程应重回开发阶段的部分管理流程，包括新数据准备、模型重训或微调、离线验证测试等。更新后的模型需经过与初版模型同样严格的测试和评审，才能发布上线。应谨慎评估是否需要进行A/B测试以比较新旧模型效果。所有更新活动应有完整记录，形成模型迭代的知识图谱。

模型退役管理是生命周期的终点。当模型因技术过时、业务不再需要或存在无法修复的缺陷而需下线时，应执行正式的退役流程。流程包括：业务影响评估、制定下线与替代方案、通知相关方、执行下线操作、归档模型所有相关资产（代码、数据、文档、日志）以备未来审计或参考，并最终从生产环境中彻底清除模型及其相关服务。退役决策应记录在案。

通过执行本规范，组织能够系统化、工程化地管理工业人工智能模型，有效降低项目风险，提升模

型质量与投资回报，最终推动人工智能技术与工业场景的深度融合与价值释放。广西产学研科学研究院作为规范研制与倡导者，将持续推动本规范的完善、宣贯与落地实施，助力我国工业智能化水平的整体提升。