

基于人工智能的职业教育教学质量 评价数据处理规范 编制说明

《基于人工智能的职业教育教学质量评价数据处理规范》

标准编制组

2026 年 10 月

一、工作简况

（一）任务来源

为统一基于人工智能的职业教育教学质量评价系统的数据处理要求，规范多源异构教学数据的清洗转换、标注、融合关联与画像生成，提升评价数据的可用性与结果的可解释性，广东交通职业技术学院提出制定基于人工智能的职业教育教学质量评价数据处理规范团体标准。根据《关于〈基于人工智能的职业教育教学质量评价数据采集规范〉等三项团体标准的立项公告》（粤政务协会〔2026〕15号）相关要求，该团体标准由广东省数字政务协会负责归口与发布，由广东交通职业技术学院牵头负责标准编制工作。

（二）起草单位

本文件的主要起草单位包括：广东交通职业技术学院、广东省外语艺术职业学院、广东省政务服务数据事务中心、中山大学、广东行政职业学院、广西物流职业技术学院、广州市数字政府运营中心、广州卫生职业技术学院、韩山师范学院、中山市技师学院、揭阳职业技术学院、广东梅州职业技术学院、江门市大数据管理中心、梅州市政务大数据管理中心、潮州市劳动保障信息中心、化州市职业技术学校、公诚管理咨询有限公司、中国联合网络通信有限公司江门市分公司，共 18 家单位。起草单位覆盖高等职业院校、中等职业学校、技工院校、普通本科高校、政务数据管理机构与企业技术服务单位，分别承担教学场景调研、数据治理实践、技术实现验证与标准文本起草校核等工作。

二、目的和意义

教学质量评价数据经采集环节进入系统后，需经过清洗转换、标注、融合关联等一系列处理，才能转化为可用于评价的分析结果。随着人工智能技术在职业教育教学质量评价中的应用，智能标注、知识图谱关联、画像生成等处理方式逐步替代传统人工统计，处理过程的规范程度直接决定评价结果的可信度与可解释性。

然而，目前行业内缺乏针对职业教育教学质量评价数据处理的统一规范，导致以下突出问题：**一是**多源异构数据的清洗与转换规则不统一，缺失值、异常值、噪声数据的处理方式各异，同一批原始数据在不同系统中得出不同的评价结论；**二是**智能标注的结果缺乏统一的输出结构与质量核验要求，标注质量难以评估与追溯；**三是**知识点、技能点与教学行为之间的实体关联规则缺失，知识图谱与教学过程数据的映射关系由各系统自行定义，评价结果无法回溯至具体教学环节；**四是**画像生成与增值评价缺乏统一口径，画像指标的来源、计算方法与更新周期不透明，评价对象难以对结论提出有效质询。

2025 年 1 月，中共中央、国务院印发《教育强国建设规划纲要（2024—2035 年）》，明确提出促进人工智能助力教育变革；2025 年 4 月，教育部等九部门印发《关于加快推进教育数字化的意见》（教办〔2025〕3 号），要求深入实施国家教育数字化战略，以数据驱动教育治理与评价方式变革；2025 年 7 月，教育部职业院校信息化教学指导委员会发布《职业院校智慧校园规范（试行）》，确立了“全面感知、互联互通、数据驱动、精准服务”的智慧校园总体框架，并

要求构建校园智慧决策与数据治理体系；2026年5月，该教指委发布《职业院校人工智能应用指引（2026版）》，将“伦理安全”列为编制原则之一，要求遵循以人为本理念、保护数据隐私、负责任地使用人工智能。上述文件为本系列标准的制定提供了直接的政策依据。

为规范基于人工智能的职业教育教学质量评价数据的处理活动，统一多源异构数据的清洗转换、标注、融合关联与画像生成要求，提升评价数据的可用性与评价结果的可解释性，广东交通职业技术学院联合业内相关单位共同制定本团体标准。本文件旨在明确“数据如何清洗、如何标注、如何关联、如何生成评价依据”四项核心要求，为教学质量评价从原始数据到评价结论的转换过程建立统一的技术基准。

本文件与数据采集规范、数据共享规范配套使用，承接采集环节的输出，为共享环节提供质量合格、口径统一的处理结果，共同构成完整的数据规范链条。三份文件各自独立成篇，本文件不引用其余两份文件的条款，可单独实施。

三、编制原则

本文件制定遵循以下原则：

1. 协调性

本系列标准严格遵循国家法律法规及宏观政策导向，紧密结合教育数字化与人工智能技术发展的现状，充分吸纳《教育强国建设规划纲要（2024—2035年）》《关于加快推进教育数字化的意见》《职业院校智慧校园规范（试行）》《职业院校人工智能应用指引（2026

版)》等政策文件精神。在编制过程中始终注重与国家现行政策体系、教育行业标准规范及人工智能技术应用趋势保持协调统一,与《职业院校数字基座高职数据标准及接口规范(试行)》等上位规范在数据元定义、接口约定上保持一致。

2. 实用性

本系列标准在编制过程中,广泛结合职业院校教学质量评价信息化建设的实际情况、现有课堂教学智慧评价系统(CSMS)等平台的实践基础,从数据源接入、采集执行、传输交付、清洗转换、标注融合、画像生成到共享服务的全过程进行规范,并针对课堂教学、线上教学、AI 助教、实训教学等典型场景给出可操作的采集与处理要求,保障标准条款设置合理可行、切实可用,具有良好的实用性和可操作性。

3. 规范性

本系列标准严格按照 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》确立的规则进行编写。编制组在起草阶段即对标该导则,对标准文本的要素构成、编排格式、条款表述及助动词使用(“应”“宜”“可”)进行了规范,确保本系列标准在结构与形式上与国家标准化文件的基本规则保持一致。

4. 伦理优先

本系列标准将个人信息保护与人工智能伦理置于技术能力要求之前。凡涉及人脸、声纹、情感等敏感个人信息与生物特征数据的处理,均设置知情同意、最小必要、个人信息安全影响评估、人工监督

与伦理审查等前置条件，并要求评价结果输出不得使用歧视性或标签化表述，防止算法偏见对学生发展造成不当影响。

5. 全生命周期协同

本系列标准由数据采集、数据处理、数据共享三份文件组成，分别对应数据生命周期的入口、加工与出口三个环节。三份文件采用统一的术语体系、数据标识体系与质量维度体系，共同构成职业教育教学质量评价数据的闭环规范。编制过程中，经反复权衡，三份文件均按各自环节的完整要求独立成篇，正文条款不采用条款级的相互引用：一是同一管控要求只在实现该要求的环节展开，避免重复规定带来的条款冲突；二是各文件的指标、格式与流程要求均在本文件内闭合，任一份文件可独立实施，不因其他文件改版而失效；三是三份文件的配套关系在各自的前言与编制说明中说明，保证使用方在数据生命周期各环节选用相应文件时口径一致。

四、编制过程

标准立项：2026年4月13日，经协会审批，《关于〈基于人工智能的职业教育教学质量评价数据采集规范〉等三项团体标准的立项公告》（粤政务协会〔2026〕15号）经协会官网和全国团体标准信息平台公开发布，本标准项目正式立项。

征集参编单位：2026年5月6日，协会发布《关于征集〈基于人工智能的职业教育教学质量评价数据采集规范〉等三项团体标准参编单位的通知》（粤政务协会〔2026〕17号），面向社会公开征集参编单位，明确参编单位名称列入标准起草单位名单、参编单位推荐

人员加入标准起草人名单。

成立标准编制组：2026年8月30日，由高等职业院校、中等职业学校、技工院校、普通本科高校、政务数据管理机构与技术服务企业等18家单位组成标准编制组。编制组主要负责标准文本的起草编制任务。在前期资料搜集和初步调研的基础上，形成标准制定工作方案，包括阶段进度安排、关键时间节点、人员工作分工等，并确立了内容大纲。

编制初稿：2026年9月，标准编制组根据计划开展标准初稿编制工作，系统查阅了《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》等法律法规、《教育强国建设规划纲要（2024—2035年）》等政策文件、GB/T 35273—2020《信息安全技术 个人信息安全规范》、GB/T 36344—2018《信息技术 数据质量评价指标》、GB/T 36073—2025《数据管理能力成熟度评估模型》、JY/T 0661—2025《教育数据分类分级指南》等标准规范及相关论文书籍，在充分收集整理基础资料和开展调研论证的基础上，对标准框架和相关技术内容反复进行了修改，形成了标准（草案）。

形成征求意见稿：2026年10月上旬至10月16日，标准编制组在完成标准初稿的基础上，通过内部研讨、组内审稿及定向咨询等方式，针对标准草案中少数存在异议的技术条款与表述进行反复讨论，在充分吸纳合理意见并对文本进行修改完善后，于2026年10月16日形成了标准（征求意见稿）。征求意见稿形成后，编制组再次对三份文件之间的衔接方式、质量维度表述与量化指标的处理方式进行了

统一，并对条款引用、附录性质与表格编号逐项复核，于 2026 年 10 月 19 日完成修改，形成与三份标准文本对应的最终版标准（征求意见稿）及其编制说明。

五、标准的主要内容

（一）技术内容说明

本文件涵盖的主要技术内容包括：系统数据处理的总体架构、数据接收与校验、数据清洗与转换、数据标注与增强、数据融合与关联（含教学评价画像生成与增值评价）、数据存储与服务、数据质量管控以及数据安全与隐私保护。本文件适用于中等职业学校、高等职业院校及应用型本科院校开展教学质量评价相关活动。具体如下所示：

第 1 章至第 4 章：给出了本文件的范围、规范性引用文件、术语和定义、缩略语。界定了数据处理、智能标注、数据融合、学生学习画像、课堂质量全景视图、标准化中间数据、数据源注册表、一数一源、画像指标来源清单共 9 项核心术语。

第 5 章 总体架构：提出了自下而上包括数据接收与校验、数据清洗与转换、数据标注与增强、数据融合与关联四层，以及顶部数据存储与服务层的处理总体架构，右侧为贯穿全流程的安全与隐私保护、数据质量管理与诊断两项要求，总体框架见图 1。

第 6 章 数据接收与校验：规定了数据接入方式（6.1）、格式校验（6.2）与完整性校验（6.3）要求，明确了接收环节对上游数据的准入核验规则；其中入站记录的数据格式以本文件附录 B 表 B.1 的规定为准，必填字段、时间戳精度、事件类型白名单与数据元语义映

射要求在本文件内闭合。

第 7 章 数据清洗与转换：规定了噪声过滤（7.1）、缺失值处理（7.2）、异常值检测与处理（7.3）、格式统一（7.4）与标准化数据输出（7.5）的要求，其中噪声过滤要求数据清洁度符合 GB/T 36344—2018、语音交互数据处理符合 GB/T 36464.1—2020。

第 8 章 数据标注与增强：规定了语义标注（8.1）、行为标注（8.2）、技能标注（8.3）、情感标注（8.4）四类标注的标签维度与标注结果输出（8.5）要求；意图分类、认知层次分类、情感极性分类、行为识别、实训操作行为识别等人工智能相关指标，本文件不预设判定门限，其测试场景由各参与方结合本校数据基础确定并实测判定，测试记录按附录 B 给出的 AI 组件测试记录格式留存。

第 9 章 数据融合与关联：规定了融合主体与标识体系（9.1）、实体关联规则（9.2，包括问答与知识点关联、操作与技能点关联、行为与评价标签关联、多源成绩关联）、教学评价画像生成（9.3，含通用要求、学生学习画像、教师教学画像、课程画像、专业画像与画像指标来源清单）、课堂质量全景视图生成（9.4）与增值评价（9.5）的要求。

第 10 章 数据存储与服务：规定了数据存储（10.1）与服务接口（10.2）的要求，支撑上层教学质量评价、学情分析、教学诊断与改进等应用。

第 11 章 数据质量管控：规定了质量指标体系（11.1）、质量监控点部署（11.2）与质量问题处置闭环（11.3）要求，明确了处理环

节的质量维度、监控点部署方式与问题处置流程。

第 12 章 数据安全和隐私保护：12.1 数据安全包括传输安全、存储安全、访问控制、审计日志；12.2 隐私保护包括去标识化处理、聚合化输出、个人信息安全影响评估；12.3 AI 应用安全与伦理包括算法公平性、人工监督与兜底、防止标签化与歧视性输出、伦理审查。

附录 A（资料性）：给出了处理环节数据质量的评价指标、评价说明与量化算法，一级维度包括完整性、准确性、及时性、一致性、唯一性、有效性、完备性、可用性与规范性，并给出各指标对应的监控点位置，供学校或第三方机构定期评估处理质量使用。

附录 B（资料性）：给出了标准化数据记录格式（B.1）、标注结果记录格式（B.2）、服务接口示例（B.3）与 AI 组件测试记录格式（B.4）。

附录 C（资料性）：给出了处理质量参数参考值（C.1）与增值评价的基础水平计算（C.2）、进步幅度计算（C.3）、增值分与双轨输出（C.4）、评价对象与指标参考（C.5）。

（二）主要核心量化指标的设定

本系列标准涉及部分量化指标。编制组在起草过程中对全部量化指标逐项开展了来源核查，并据此确立了“可落地、可验证、无超出现有技术水平的过高要求”的指标设定原则。为此，本系列标准对量化指标采取分类处理方式，具体如下：

一是直接引用。凡现行国家标准已作出明确规定的，本文件直接引用标准号与条款，如数据质量维度定义与评价要求引用 GB/T

36073—2025 与 GB/T 36344—2018，数据清洁度与语音交互数据处理分别引用上述两项标准，网络数据处理安全引用 GB/T 41479—2022，个人信息安全影响评估引用 GB/T 39335—2020，知识图谱引用 GB/T 42131—2022，教育数据元引用 JY/T 0647—2022。

二是只给计算方法、不给判定门限。附录 A（资料性）表 A.1 给出各质量维度的一级指标、二级指标、评价说明与量化算法，但不再规定“合格”“优秀”等等级门限，亦不再单独设置质量等级划分章节，改由学校或第三方机构依据本校数据基线确定目标值。本文件评价的是数据经本文件规定环节处理后的质量，采集环节的质量评价不属于本文件的范围，从而避免同一数据项在两个环节重复判定、门限互为前提。

三是以实测确定替代预设阈值。对智能标注准确率、实体关联准确率、画像指标覆盖率等人工智能相关指标，本文件不在正文中预设具体数值，要求由各参与方确定测试场景后实测判定，并按本文件附录 B 给出的 AI 组件测试记录格式留存测试记录，使指标判定过程可追溯。

四是参数由使用方配置并记录。关联时间窗长度、融合权重、判定门限等参数，本文件不规定统一取值，要求由学校或第三方机构依据专业特点设定并记录，纳入画像指标来源清单以保证可追溯；附录 C 表 C.1 给出的处理质量参数参考值为行业公开资料中的常见取值，供确定目标值时参考，不作为强制要求。

六、知识产权情况说明

本文件不涉及知识产权。

七、与现行相关法律法规和强制性国家标准的关系

本文件与现行法律法规和强制性国家标准均不存在冲突。

本文件符合《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《中华人民共和国标准化法》等法律法规，与《教育强国建设规划纲要（2024—2035 年）》《关于加快推进教育数字化的意见》《职业院校人工智能应用指引（2026 版）》等政策文件精神相契合，响应了国家教育数字化战略与“人工智能+教育”行动的实践需求。

八、相关国内标准情况简要说明

与本文件较为相关的标准规范包括：

（1）GB/T 36344—2018《信息技术 数据质量评价指标》，主要规定了数据质量的评价指标与计算方法，适用于数据质量的评价与管理。

（2）GB/T 36073—2025《数据管理能力成熟度评估模型》，主要规定了组织数据管理能力成熟度模型与评估方法，适用于数据管理能力的评估与提升。

（3）GB/T 35273—2020《信息安全技术 个人信息安全规范》与 GB/T 39335—2020《信息安全技术 个人信息安全影响评估指南》，分别规定了个人信息安全要求与个人信息安全影响评估的方法。

（4）GB/T 41479—2022《信息安全技术 网络数据处理安全要

求》与 GB/T 41817—2022《信息安全技术 个人信息安全工程指南》，分别规定了网络数据处理的安全要求与个人信息保护工程化实施要求。

(5) GB/T 42131—2022《人工智能 知识图谱技术框架》与 GB/T 46347—2025《人工智能 风险管理能力评估》，分别规定了知识图谱的技术框架与人工智能风险管理能力的评估要求。

(6) JY/T 0647—2022《教育管理信息化数据标准 第1部分：基础数据元》与 JY/T 0661—2025《教育数据分类分级指南》，分别规定了教育管理信息化基础数据元与教育数据的分类分级方法。

(7) GB/T 36464.1—2020《信息技术 智能语音交互系统 第1部分：通用规范》、GB/T 38152—2019《信息技术 情感计算 用户界面 面部表情识别》，分别规定了智能语音交互与面部表情识别相关的技术要求。

目前尚未检索到针对职业教育教学质量评价数据在采集、处理、共享环节进行系统性规范的同类国家标准、行业标准或团体标准，现有标准多侧重于教育管理信息化基础数据元、教育数据分类分级或通用数据治理能力评估，未涉及人工智能辅助教学质量评价这一特定场景下的多模态数据采集、画像生成与跨机构共享要求。本系列标准与上述相关标准规范在内容侧重上有所不同，是对现有教育信息化与数据治理标准体系的有效补充与细化延伸。

九、重大分歧意见的处理经过和依据

本文件未产生重大分歧意见。

十、标准性质的建议

本文件作为推荐性标准，属于团体标准，供协会会员和社会自愿使用。

十一、标准实施建议

本文件批准发布实施后，尽快将本文件的批准信息通告编制小组成员单位及其他有关单位。积极组织本文件的宣传和培训，使涉及职业教育教学质量评价数据相关工作的院校、管理部门与技术服务单位熟悉了解本文件的内容与特点，保证标准的顺利实施，使标准的应用真正落到实处。

为全面掌握本文件的执行情况，鼓励使用单位和相关单位将本文件的执行情况以及所发现的问题及时反馈到归口团体或本文件的起草单位，为标准完善工作做好准备，不断提高标准的科学性、合理性、协调性和可操作性。

十二、其他应说明的事项

本文件为《基于人工智能的职业教育教学质量评价数据采集规范》《基于人工智能的职业教育教学质量评价数据处理规范》《基于人工智能的职业教育教学质量评价数据共享规范》系列标准之一，三份文件应配套使用。编制过程中，经起草组反复权衡，三份文件按各自环节的完整要求独立成篇，正文条款不采用条款级的相互引用：涉及数据格式、必填字段、接口要求、质量评价指标、AI 测试场景与记录格式等内容，均由本文件在自身附录中完整规定，不引用其余两份文件的条款。这样处理可避免同一要求在三份文件中重复规定而产生冲

突，也可避免因其他文件修订导致本文件引用失效。三份文件的术语体系、数据标识体系与质量维度体系保持一致，配套使用时不会产生理解歧义；各文件的引用关系与衔接要求在本文件范围与相关附录中说明。