

基于人工智能的职业教育教学质量 评价数据采集规范

编制说明

《基于人工智能的职业教育教学质量评价数据采集规范》

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采集要求，规范多模态教学数据的接入、采集、传输与前置质量控制，保障采集数据的真实性、完整性与合规性，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采集规范团体标准。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

随着国家教育数字化战略的深入实施和人工智能技术在教育领域的快速应用，职业院校逐步引入课堂语音转写、视觉行为识别、实训操作采集等人工智能技术开展教学质量评价，形成了文本、音频、视频、操作序列等多模态教学数据。教学质量评价的科学性高度依赖采集数据的质量——采集环节一旦出现数据缺失、口径不一、置信度不可控或合规性瑕疵，后续处理、分析与评价结论均将失去可靠基础。

然而，目前行业内缺乏针对人工智能辅助教学质量评价数据采集的统一规范，导致以下突出问题：一是**采集范围与采集粒度由各系统自行定义**，同一教学行为在不同系统中被拆分为不同数据项，跨系统数据无法横向比对与纵向累积；二是**数据源接入方式与接口协议不统一**，教务系统、实训设备、在线学习平台等异构数据源的接入成本高、映射关系不透明；三是**人工智能采集输出（如语音转写文本、行为标签、置信度）缺乏统一的质量控制要求**，低置信度结果未经筛选直接进入评价环节；四是**人脸、声纹、情感特征等敏感个人信息的采集边界不清**，知情同意、最小必要与影响评估等合规要求缺乏可操作的实施依据。

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。

为规范人工智能辅助教学质量评价数据的采集活动，统一多模态教学数据的接入方式、采集策略、接口要求与质量控制要求，填补该领域标准空白，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“采集什么、如何采集、采集到什么程度、采集过程如何保护个人信息”四项核心要求，为职业院校教学质量评价系统的数据入口建立统一的技术基准与合规底线，引导相关系统开发单位提升采集能力，促进评价数据的校际可比与长期累积。

本文件与数据处理规范、数据共享规范配套使用，三者分别覆盖数据生命周期的入口、加工与出口环节，共同为职业院校教学质量评价提供完整的数据规范支撑。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：系统数据采集的总体架构与衔接要求、数据接入与映射、数据采集与执行、数据接入与传输、采集数据质量控制以及数据安全与隐私保护，接口规范与数据分级分类作为贯穿全部采集活动的管控要求在附录 E 中规定。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。界定了教学质量评价、教学质量评价系统、系统数据采集接口、采集策略、伴随式采集、采集粒度、多模态数据、教育数据画像、人工智能辅助采集、实训操作数据共 10 项核心术语。

第 5 章 总体架构：5.1 提出了“四层流程 + 顶部安全层 + 两条纵向管控条”的系统数据采集总体架构，自下而上依次为数据接入与映射、数据采集执行、数据接入与传输、采集数据质量控制四层，顶部为数据安全与隐私保护层，右侧为贯穿全流程的接口规范与数据分级两条管控要求，总体框架见图 1；5.2 规定了与学校已建成的教

学质量评价系统、数据采集平台的衔接要求，明确不应重复建设采集通道，并要求采集数据能够映射至现有系统的评价指标维度。

第 6 章 数据接入与映射：6.1 数据来源识别要求按校内业务系统、教学终端、实训环境、外部平台四类进行注册配置，并在数据源注册表中登记业务归属、责任部门与系统负责人；6.2 数据源映射规定了唯一标识、数据口径统一等要求；6.3 协议适配规定了多协议适配要求，确保异构系统统一接入。

第 7 章 数据采集与执行：7.1 采集策略规定了采集策略配置、教学场景差异与事件类型定义，要求依据教学场景、教学活动与评价目的配置，典型场景要求见附录 C；7.2 规定了伴随式采集的部署条件与采集方式；7.3 规定了最小必要采集要求，明确未列入采集目录的数据项不得采集；7.4 规定了人工智能辅助采集能力要求，包括 AI 辅助组件性能测试、语音与文本采集、视觉采集与知识关联采集。

第 8 章 数据接入与传输：规定了接口调用、数据格式封装、加密传输与传输日志记录的要求，其中数据格式封装要求采用统一的 JSON Schema 定义，字段结构符合附录 B 的规定。

第 9 章 采集数据质量控制：规定了完整性校验、置信度过滤、实时性监控、去重与防抖、异常值过滤、一数一源一致性共六项前置质量控制要求，确保进入处理环节的数据具备基本质量水平。

第 10 章 数据安全和隐私保护：10.1 数据安全规定了传输与存储安全、访问控制与审计日志要求；10.2 隐私保护遵循正当必要、知情同意、目的明确、安全保障原则，规定了人脸、声音等个人信息

的告知同意要求；10.3 规定了 AI 应用安全与伦理要求，包括模型安全、对抗攻击防御与模型版本管理。

附录 A（资料性）：给出了采集数据质量评价指标，从完整性、准确性、及时性、一致性、唯一性、有效性、覆盖性、规范性八个维度给出评价说明与量化算法。

附录 B（规范性）：规定了采集输出数据格式与接口规范，包括通用必填字段（B.1）与音频采集数据格式（B.2），与 8.2 数据格式封装的要求相对应。

附录 C（资料性）：给出了采集质量参数参考值（C.1）与典型教学场景采集要求（C.2），其中 C.2 覆盖课堂教学、线上教学、AI 助教、实训教学四类场景。

附录 D（资料性）：给出了 AI 模型性能测试场景与记录格式，包括测试场景（D.1）、对抗攻击测试场景（D.2）、公平性测试场景（D.3）与测试记录格式（D.4），其中对抗攻击测试覆盖计算机视觉对抗与大语言模型提示注入两类。

附录 E（规范性）：规定了接口规范和数据分级分类，对应图 1 右侧两条贯穿全流程的管控条，其要求贯穿第 6 章至第 10 章的全部采集活动；E.1 接口规范包括接口通用要求、接口类型与接口技术要求，E.2 数据分级分类包括数据分类、数据分级与最小必要采集。

（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，不在正文中复述标准内容，如数据分类分级引用 JY/T 0661—2025《教育数据分类分级指南》，密码应用引用 GB/T 39786—2021，个人信息安全引用 GB/T 35273—2020，数据质量评价引用 GB/T 36344—2018 与 GB/T 36073—2025。

二是只给计算方法、不给判定门限。附录 A（资料性）表 A.1 对完整性、准确性、及时性、一致性、唯一性、有效性、覆盖性、规范性八个维度均给出评价说明与量化算法，但不再规定“合格”“优秀”等等级判定门限，亦不预设百分比阈值，改由学校或第三方机构依据本校数据基线、专业特点确定目标值与判定周期。这样处理的理由在于：不同院校、不同专业的采集条件差异显著，统一的门限既无公认依据，也难以适应实际。

三是以实测确定替代预设阈值。对人工智能采集能力类指标（如语音识别字正确率、意图分类准确率、实体抽取准确率、行为识别准确率、表情分析准确率），本文件不在正文中预设具体数值，改由附录 D 给出测试场景、测试数据准备要求与测试记录格式，并明确“各项指标的判定阈值宜依据实测数据确定，本文件不规定统一合格线”，要求在组件部署或模型更新后按附录 D 实测确定并留档比对。

四是资料性参考值。对确需给出量级参考的参数（如采集延迟、采样频率），本文件在资料性附录 C 表 C.1 中给出参考值，并注明

表中数值为行业公开资料中的常见取值，供学校在项目实施中确定目标值时参考，不作为本文件的强制要求，亦不应直接作为验收门槛。

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 35273—2020《信息安全技术 个人信息安全规范》，主要规定了个人信息收集、存储、使用、共享等环节的安全要求，适用于各类组织处理个人信息的活动。

（2）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。

（3）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。

（4）GB/T 39786—2021《信息安全技术 信息系统密码应用基本要求》，主要规定了信息系统在物理和环境、网络和通信、设备和计算、应用和数据四个层面的密码应用要求。

（5）JY/T 0661—2025《教育数据分类分级指南》，主要规定了教育数据的分类规则与分级方法，适用于教育领域数据的分类分级工作。

（6）GB/T 41772—2022《信息技术 生物特征识别 人脸识别系统技术要求》、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 42131—2022《人工智能 知识图谱技术框架》，分别规定了人脸识别、智能语音交互、知识图谱相关的技术要求。

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。