

基于评价学的 AI-Ready 材料数据集质量评价框架

左维¹, 杨正伟¹, 施思齐^{2,3}, 刘悦¹

1. 上海大学计算机工程与科学学院&核电关键材料全国重点实验室, 上海 200444;
2. 上海大学核电关键材料全国重点实验室&材料科学与工程学院, 上海 200444;
3. 上海大学材料基因组工程研究院, 上海 200444

摘要

数据质量的科学评价是数据有效治理的核心前提, 直接关系到高质量 AI-Ready 数据集在下游任务建模的效果。当前研究聚焦于单一数据模态或特定质量维度, 导致跨模态、跨维度的评价结果缺乏可比性, 难以形成系统的数据质量画像。因此, 提出基于评价学的 AI-Ready 多模态材料数据集质量评价框架, 明确评价对象、评价条件与评价方法以构建统一评价体系, 并籍此通过“继承-发展”六大规则与五步构建流程, 实现从统一框架向分模态评价体系的系统性延伸。进一步, 对 AI-Ready 材料时序数据集完整性进行案例实验, 验证了基于该评价框架设计评价体系的有效性, 为构建高质量 AI-Ready 材料数据集、开发高可靠数据驱动模型提供坚实的理论支撑与科学依据。

关键词

高质量 AI-Ready 数据集; 数据质量评价; 评价学; 材料科学

中图分类号: TB3; TP181;

文献标志码: A

Evaluatology-based AI-Ready materials dataset quality evaluation framework

ZUO Wei¹, YANG Zhengwei¹, SHI Siqi^{2,3}, LIU Yue¹

1. School of Computer Engineering and Science & State Key Laboratory of Materials for Advanced Nuclear Energy, Shanghai University, Shanghai 200444, China;
2. State Key Laboratory of Materials for Advanced Nuclear Energy & School of Materials; Science and Engineering, Shanghai University, Shanghai 200444, China;
3. Materials Genome Institute, Shanghai University, Shanghai 200444, China

Abstract

Scientific evaluation of data quality is a core prerequisite for effective data governance, as it directly affects the utility of high-quality AI-Ready dataset for downstream tasks. However, current research mainly focuses on specific modalities or dimensions, leading to incomparable evaluation results across different modalities and dimensions, and making it difficult to establish a systematic data quality profile. To address this issue, this work proposes an Evaluatology-based multimodal data

quality evaluation framework for AI-Ready materials dataset, by defining the evaluation subject, evaluation conditions, and evaluation method to construct a unified evaluation system. Then, the six rules and five-step construction process of "Inheritance – Development" are employed to extend the unified framework into modality-specific evaluation systems. Furthermore, a case study on the completeness of AI-Ready materials time-series data was conducted, validating the effectiveness of the evaluation system designed based on the proposed framework. This work provides a theoretical and scientific foundation for constructing high-quality AI-ready dataset and developing highly reliable data-driven models.

Key words

high-quality AI-Ready dataset, data quality evaluation, Evaluatology, material science

0 引言

纵观人类科学探索的演进脉络，从古老度量衡的建立到现代复杂系统的量化分析，“继承–发展”的辩证统一始终是实现科学评价的核心思想。“继承”确保了评价体系在不同视角下的通用性与一致性；而“发展”则赋予了评价体系在应对不同评价对象时的适配性与扩展性；两者互为表里，使得评价体系能够成为穿透复杂现象、揭示客体价值的科学判据。在材料信息学深度重塑材料研发全链条的背景下，AI-Ready 材料多模态数据集的质量决定了以知识图谱为代表的符号主义人工智能和以神经网络为代表的连接主义人工智能建模的性能上限^[1–2]，影响了构效关系揭示的准确性和新材料发现的有效性。然而，材料多模态数据天然具有多源异构性，即既可以来自不同精度的实验测量，也可以源于多种尺度的理论模拟，甚至从文献中非结构化提取，导致数据的表现形式存在显著差异；同时，材料多模态数据也呈现跨尺度关联，微观、介观、宏观尺度之间数据相互关联，使得任何单一尺度上的误差都会非线性放大。正是这种多源异构和跨尺度关联的特殊性^[3]，令统一、全面的数据

集质量评价体系难以构建，进而严重阻碍了材料数据质量的有效提升。

本研究团队将数据质量理解为“品质”和“数量”两方面，并强调领域知识在材料数据质量治理的重要性，提出了材料领域知识嵌入的数据品质治理框架^[4]和材料领域知识嵌入的数据数量治理框架^[5]，系统性地定义九大品质治理维度和三大数量治理维度，明确了数据质量治理的对象；建立全生命周期模型，明确了何时进行数据质量治理；发展领域知识嵌入的处理模型，明确了如何进行数据质量治理；并在 60 份公开材料结构化数据集上进行实验，验证了理论的有效性。籍此，本研究团队分模态分治理维度开展了深入研究：面向材料结构化数据的普遍存在的异常值、特征冗余及属性预测建模困难等问题，提出了领域知识嵌入的异常值识别方法以提升数据的准确性^[6]、领域知识嵌入的特征筛选方法以降低数据的冗余性^[7]、基于元学习的材料属性预测自动建模方法 Auto-MatRegressor 以衡量数据的洞察力^[8]；面向材料文本数据中信息稀疏、术语歧义及标注准确性不足等问题，提出了高质量的材料科学文本挖掘数据集构建方法^[9]，从数据构建源头保障数据集质量，并针对文本数据准确性问题提出了相应的治理方法^[10]。此外，针对材料实验和标注成本高

昂、制备周期长等导致的稀缺数据问题，本研究团队提出了可解释生成模型DW-VAE^[11]和FcLDiff^[12]，分别实现了高质量材料结构化数据与材料图像数据的可控生成，为数据量增强提供了可解释的解决方案。然而，如何基于本团队已定义的治理维度，进一步转化为逻辑统一且模态适配的分模态评价体系，确保分模态数据质量评价体系的构建结果可溯源、可比较，从而奠定多模态评价的评价基础，为跨模态的质量比较与联合评估提供可信保障，仍是当前亟待解决的问题。为此，本工作基于评价学的基本原理^[13]，引入“继承-发展”思想，提出基于评价学的AI-Ready材料数据集质量评价框架，从而实现科学评价奠定坚实的理论支撑。本文的主要贡献总结如下：

- (1) 建立了AI-Ready材料多模态数据集质量统一评价体系，明确了评价对象及维度、评价条件与评价方法，为分模态评价体系的构建提供了统一框架基础；
- (2) 提出了基于“继承-发展”的分模态数据集质量评价体系构建方法，面向统一评价体系进行继承，并结合数据模态和领域知识进行发展，实现评价结果的可信与可比；
- (3) 以AI-Ready材料时序数据集的完整性为例，发展了具体的计量指标与计量算法，并通过基于模拟数据和真实数据的实验，验证了所提出方法的有效性。

1 相关工作

近年来，围绕材料数据质量治理的研究已取得一定进展，但普遍呈现“重提升，轻评价”的现象，且多围绕单一模态的单一维度进行研究。在结构化数据方面，研

究主要聚焦于通过在数据驱动模型中嵌入领域知识以识别并修复数据中的异常值^[6, 14]，减少数据中的冗余性^[7, 15]；同时为缓解材料数据常见的小样本问题，部分研究在生成对抗网络等生成式模型中嵌入领域知识以进行样本增强与扩充^[16]，以保证生成数据的物理合理性。在图像与文本等非结构化数据方面，也逐步有研究开始运用领域知识和生成式模型对数据质量进行提升，如Khan等人^[17]利用CycleGAN结合倒易空间物理先验，将模拟STEM图像转化为近乎真实的实验图像；Liu等人^[18]则通过物理建模生成大量合成数据，对大型语言模型进行预训练，有效缓解了聚合物性质预测中的实验数据稀缺问题。上述工作虽然极大地丰富了材料数据的清洗、修复与增强方法，然而大多仍属于后验式单一维度治理，缺乏对数据集进行全面、系统量化评价的机制，导致研究者难以准确判断数据集是否真正满足AI-Ready的要求。

针对上述问题，亟需引入数据评价有关的理论和方法。相关系统性研究最早可追溯到MIT提出的数据质量管理计划，此后ISO 8000、ISO 25012等通用国际标准及相关国家标准相继出台，奠定了数据质量评价的通用标准要求，未涉及面向特定领域AI建模需求的数据集质量评价体系构建。在此基础上，Wang和Strong^[19]面向一般信息系统中的数据场景，将数据质量划分为固有性、情境性、表示性和可访问性四大维度，但未针对材料科学数据的特殊性及AI建模的具体需求进行适配。针对具体模态和具体领域，文本数据质量^[20-21]、知识图谱质量^[22]、通用图像质量^[23]以及工业时序数据^[24-26]等领域的细分评价体系相继涌现。而为了使评价过程更具科学性和理论性，詹剑锋等人^[13]系统性

地建立了“评价学”阐明了评价的本质，辨析了评价、测量与测试之间的内在逻辑与技术边界，更通过五大公理为科学评测奠定了公理化基础。同时，Alampara 等人^[27]强调了材料机器学习领域中评价科学的重要性，并提出评价卡用于评价材料机器学习建模框架。然而，目前鲜有关于评价学应用至 AI-Ready 材料数据集质量评价的研究。

综上所述，现有相关研究在材料数据集质量治理方法、通用评价质量体系和评价学的建设上有着较为丰富的基础，但仍存在以下三方面不足：

- 第一，多模态数据集质量评价的理论性不足，缺乏公理化的严格指导。现有评价工作多依赖经验性的启发式指标或下游模型的后验表现来反推数据质量，未能明确评价对象、评价条件与评价方法之间的映射关系，导致科学性难以保障。

- 第二，缺少统一的 AI-Ready 材料多模态多维度数据集质量评价框架。现有评价工作关注的质量维度并不统一，并且结构化表格、文本、图像等不同模态之间存在较大差异，跨团队、跨数据、跨维度、跨模态的评价结果无法有效关联。

- 第三，鲜少考虑领域知识在评价过程中的参与机制。现有评价工作忽视了领域知识在评价指标设计中的指导价值及其在评价结果中的判据作用，致使评价结果可能符合统计学指标却违背基本物理规律，无法真实反映数据质量的本质。

因此，本文以评价学为基础，引入“继承-发展”思想，力图构建 AI-Ready 材料多模态数据集质量评价框架，以弥补上述三方面的不足。

2 基于评价学的 AI-Ready 材料多模态数据集质量评价框架

基于评价学的 AI-Ready 材料多模态数据集质量评价框架，由 AI-Ready 材料多模态数据集质量统一评价体系（unified multimodal AI-Ready MATerials Datasets Quality Evaluation system, MAT-DQE）和基于“继承-发展”的分模态数据集质量评价体系构建方法组成。前者为后者提供统一的评价对象及维度、评价条件与评价方法，确保不同模态数据的质量评价结果在等价评价条件下可溯源、可比较、一致且收敛于真值；后者基于前者能够构建分模态的评价体系，实现评价体系的可扩展性和可复现性，并奠定多模态数据集质量评价的基础。

2.1 AI-Ready 材料多模态数据集质量统一评价体系

基于评价学的基本理论，AI-Ready 材料多模态数据集质量统一评价体系 MAT-DQE 包含评价对象及维度、评价条件和评价方法三个核心要素，如下图 1 所示。

2.1.1 评价对象及维度

依据评价学理论，MAT-DQE 的评价对象被定义为 AI-Ready 材料多模态数据集，记为 $D(Dataset)$ 。基于本研究团队的前期提出的材料领域知识嵌入的数据品质治理框架^[4]和数据数量治理框架^[5]，设计共计 2 类 13 个质量维度，如表 1 所示。

上述 13 个质量维度从品质与数量两个属性出发，相对独立、互为补充，共同构成刻画数据集质量内涵的完整体系。需注意的是，其中三组概念的内涵上存在部分交叉，评价时应予以明确区分：冗余性侧重评价特征间与样本间的信息重复程度，而特征量侧重评价特征数量是否过多或过



图1 AI-Ready材料多模态数据集质量统一评价体系MAT-DQE

表1 AI-Ready材料多模态数据集质量评价维度

所属性	质量维度	定义
品质属性	可溯源性	评估材料数据的获取及基于机器学习的处理过程是否可追溯。
	时间敏感性	评估不同时期材料数据可能具有的不同物化特性及其时效性。
	完整性	评估用于训练机器学习模型的材料数据以及其他关键附加信息是否被完整记录。
	一致性	评估多源材料数据中的样本是否以同一方式呈现,以及具有依赖关系的数据是否能够正确关联起来。
	准确性	评估材料数据是否被正确记录并反映了真实情况。
	冗余性	评估材料数据中是否存在冗余信息。
	均衡性	评估材料数据集中不同类别样本的相对比例是否均衡。
	规范性	评估材料数据是否按需转为适合机器学习的表示和组织形式。
	洞察力	评估材料数据的可学习性,避免后续耗时分析。
	特征量	评估材料数据集中特征数是否合适。
数量属性	样本量	评估材料数据集中样本数是否合适。
	比值	评估材料数据集中特征数和样本数是否均衡。
	信息量	评估不同数据表示方法得到的特征所包含的信息量是否充分。

少——即便特征总数甚少，若特征间或样本间高度相关，数据集冗余性依然居高。完整性侧重评价数据记录及元数据是否存在缺失，而信息量侧重评价既定数据表示下有效信号是否充分——即便数据记录完整无缺，若所有特征取值全为常数，信息量为零。规范性侧重评价数据格式、编码方式等是否标准化以满足机器学习模型的输入要求，是数据集满足 AI-Ready 的必

要不充分条件，而并非其全部内涵。

2.1.2 评价条件

依据评价学理论，评价结果的真实性、可比较性和可溯源性均依赖于评价条件的明确界定。因此，MAT-DQE 的评价条件可以形式化为一个四元组：

$$EC = (DM, AS, A, S) \tag{1}$$

其中， $DM(DataModality)$ 指待评价数据集所包含的模态类型，如结构化数据、文本数据、图像数据、时序数据、晶体结构等，不同模态决定了质量问题的具体表现形式及相应的测量方法； $AS(ApplicationScenario)$ 指数据集预期使用的目标场景，如模型训练、数据库共享、跨模态检索等，决定了评价过程中的参数设定和权重分配； $A(Algorithms\&Metrics)$ 指对具体质量维度进行测量和/或测试所采用的计量算法和计量指标，如完整性缺失率计算、准确性异常值检测算法等； $S(SupportSystem)$ 指底层运行环境与技术平台，如计算资源（硬件、操作系统、算法运行框架）和数据基础设施（数据库、文件系统、溯源记录与日志系统等）。

2.1.3 评价方法

依据评价学理论，评价方法用于计量和/或测试评价对象以获得定量结果，再经由价值函数综合为最终评价结论。本框架强调计量驱动，即所有质量维度均通过可

复现的算法进行计算，测试仅用于校验基础功能性（例如文件损坏检测），其结果作为计量的前置条件，不作为主要评价依据。因此，MAT-DQE的评价方法可形式化为一个映射：

$$\mathcal{M}(D, EC) = (\vec{m}, Q) \tag{2}$$

其中， $\vec{m} = (m_1, m_2, \dots, m_n)$ ，指的是质量维度的计量值向量，其中每个取值 m_i 应具有一致且相同的取值区间； $Q = V(\vec{m})$ ，指的是综合评分，由价值函数 $V(*)$ 计算得出。

为得到准确、有效的计量值向量，通常需要依据数据模态和领域知识，针对性地设计计量指标和计量算法。依据评价学的综合评价指标本质公理，计量指标要么具有内在物理意义（如样本量绝对值），要么由价值函数定义（如洞察力评分）；每个质量维度的计量值 m_i 可能由多个计量指标聚合而成。同时，面向不同模态、复杂维度的计量指标，需要选择或设计合适的计量算法，通常需要保证以下设计三原则以符合评价学五大公理，如表2所示。

表2 计量算法的设计三原则及与评价学五大公理的对应

设计原则	内涵	对应公理
模态适配	针对不同材料数据模态和维度内涵设计专用算法,并支持等价比较	综合评价指标本质公理 评价结果可比较性公理
可复现	算法在相同输入下输出确定结果,且多次测量结果稳定趋近同一真值	评价结果一致性公理 评价结果真值公理
可溯源	算法参数和中间结果应被完整记录,以便追溯评价结果差异的原因	评价可溯源性公理

在得到计量值向量后，还需进行综合评分。记 w_i 为第 i 个维度的权重且 $\sum_{i=1}^n w_i = 1$ ，通常由应用场景中的利益相关者需求决定，可通过层次分析法、熵权法或

领域专家打分获得。设 $I_{crit} \subseteq \{1, 2, 3, \dots, n\}$ 为关键计量指标集合， θ_i 为第 i 个关键计量指标的最低阈值。若任一关键维度 $i \in I_{crit}$ 满足 $m_i < \theta_i$ ，则直接判定为不合格；否则按加权计算得分。形式化为：

$$Q = V(\vec{m}) = \begin{cases} FAIL, & \exists i \in I_{crit}: m_i < \theta_i \\ \varphi^{-1}\left(\sum_{i=1}^n w_i \varphi(x_i)\right), & otherwise \end{cases} \quad (3)$$

其中, $\varphi(*)$ 为单调函数, 通常考虑幂函数 $\varphi(x) = x^\alpha$; 当 $\alpha = 1, \alpha = -1, \alpha \rightarrow 0$ 时, 分别为算术平均、调和平均和几何平均。若利益相关者希望综合平衡考虑每个质量维度, 可以使用算术平均; 若希望强调短板效应, 可使用几何平均或调和平均。

2.2 基于“继承-发展”的分模态数据集质量评价体系构建方法

本节提出“继承-发展”的六大规则, 并结合数据模态和领域知识, 发展基于“继承-发展”的分模态数据集质量评价体系构建流程。

2.2.1 “继承-发展”的六大规则

分模态评价体系的构建必须以 MAT-DQE 为基础, 以保证不同模态评价结果在同一逻辑维度下的可比较性, 具体而言, 包括以下三个规则:

规则一: 评价对象及其维度的继承。分模态质量评价体系中质量维度必须从 MAT-DQE 评价对象的 13 个质量维度内进行映射。

规则二: 评价条件的继承。分模态质量评价体系的评价条件须沿用式(1)定义的结构, 即在确定数据模态和场景的前提下, 对计量算法、计量指标和支持系统进行具体化设计。

规则三: 评价方法的继承。分模态质量评价体系的评价方法须沿用式(2)和式(3)定义的结构, 满足评价学的五大公理和三大计量算法设计原则。

在遵循统一框架的基础上, 分模态体

系需结合数据模态和领域知识进行针对性发展。具体而言, 包括以下三个规则:

规则四: 评价对象及其维度的发展。分模态质量评价体系中的质量维度需要横向发展、向下发展数据模态特有的一级维度、二级细分维度, 以将 MAT-DQE 的通用维度语义转化为具有材料物理内涵的可计量指标族。同时, 子维度的设计须严格归属父维度, 且其定义须具有明确意义, 满足综合评价指标本质公理的要求。

规则五: 评价方法中计量指标的发展。分模态质量评价体系中的计量指标需要从 MAT-DQE 的通用维度计量值 m_i 向模态专属子指标向量 $\vec{m}_i = (m_{i,1}, m_{i,2}, \dots, m_{i,k})$ 进行向下细化, 并设计各维度内的子指标聚合函数 g_i , 得到该维度的最终计量值 $m_i = g_i(\vec{m}_i)$, 聚合函数 g_i 的形式需依据子指标间的逻辑关系和领域知识确定。

规则六: 评价方法中计量算法的发展。分模态质量评价体系中的计量算法需要针对计量指标进行具体实现。算法选型应优先采用材料信息学领域已验证的成熟算法、工具和策略, 综合考虑数据驱动和知识驱动的方式, 以保证计量结果的可复现、可信赖。

2.2.2 分模态数据集质量评价体系构建流程

基于上述“继承-发展”的六大规则, 本文进一步提出分模态数据集质量评价体系构建流程, 具体由下述五个步骤组成, 如下图 2 所示。

步骤一: 数据模态形式化定义。将待评价数据集 D_{mod} 抽象为由其核心结构要素构成的数学表示, 明确组成单元、变量类型、索引框架及元数据范围的形式化表达, 为步骤二中维度与质量问题之间的映射关

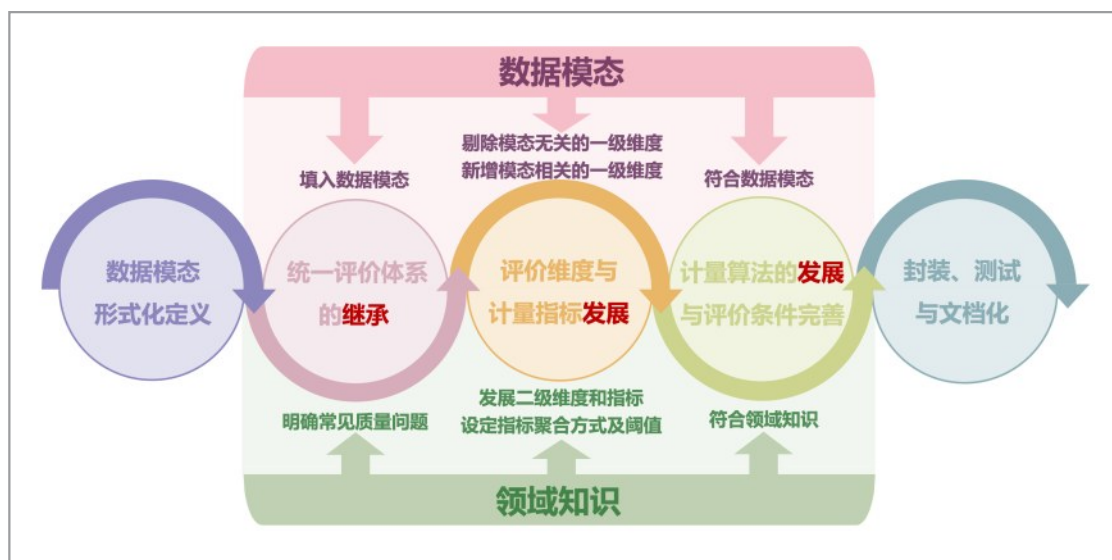


图2 分模态数据集质量评价体系构建流程

系分析、步骤三中子指标的公式设计提供统一的基准。

步骤二：统一评价体系的继承。首先，明确待评价数据集 D_{mod} 的数据模态及其应用场景，分别记为 DM_{mod} 和 AS_{mod} 。然后，识别该模态在实验采集、计算生成或文献提取等环节中常出现的质量问题，并分析其与 MAT-DQE 的 13 个评价维度的关系。在此基础上，逐步采用规则一至规则三，实现统一框架的继承。具体为：依据规则一，将评价对象及维度的候选维度池初始化为全部 13 个质量维度，并标记为待裁剪状态；依据规则二，将评价条件四元组中的 DM 字段填入 DM_{mod} ， AS 字段填入 AS_{mod} ，而 A 和 S 字段暂时置空；依据规则三，将式(2)和式(3)作为评价方法的核心形式，由此得到初始评价框架。

步骤三：评价维度与计量指标发展。在步骤二的基础上，逐步采用规则四和五实现评价维度和计量指标的发展。具体操作为：首先，基于步骤一识别的模态质量问题，对初始维度池进行裁剪，剔除与该模态无关或无法计量的维度，必要时依据

规则四的增设该模态特有的一级维度，同时须论证新增一级维度不可归入已有 13 个维度；其次，对每一维度，依据其领域知识的要求，向下发展模态专属的二级子指标，形成维度-子指标层级树；每一子指标须定义其含义、取值范围及归一化方向，确保满足综合评价指标本质公理；最后，依据规则五，为每个维度设计其内部的子指标聚合函数，将子指标计量值向量映射为维度计量值，并结合领域知识确定各子指标的硬阈值，从而确定 $\mathcal{M}_{mod}$ 。

步骤四：计量算法的发展与评价条件完善。依据规则六，为步骤三中定义的每一个子指标开发或选用具体的计量算法，形成算法库 A_{mod} ，所有算法均须以函数形式明确输入、输出与参数，并记录版本信息，以满足可复现性和可溯源性要求。同时，确定评价实验所依赖的硬件平台、软件环境等信息，形成系统库 S_{mod} ，从而获得完整的评价条件四元组 EC_{mod} 。

步骤五：封装、测试与文档化。首先，将前述步骤产出的评价条件四元组 EC_{mod} 封装为一个模态专属评价函数

$\mathcal{M}_{mod}(D, EC_{mod})$ 。然后，使用已知质量缺陷的数据集 D_{test} 对该函数进行测试，确认其能够触发规则五所设硬阈值的惩罚，且评分随质量提升呈单调性；同时，重复评价结果符合评价结果一致性公理，且等价评价条件下不同批次数据的结果具有可比性。测试全程记录数据集、运行环境和参数配置，以支持评价可溯源性公理。最后，将维度选择理由、子指标定义、算法描述与来源、权重阈值确定依据以及测试报告撰写为结构化规范文档。籍此，实现了分模态数据集质量评价体系的构建，并可以支持大规模使用。

通过上述五步构建方法，任意目标模态均能从 MAT-DQE 出发，沿着清晰的继承路径与规范的发展操作，生成符合评价学公理、嵌入数据模态特点和领域知识的分模态数据集质量评价体系，为 AI-Ready 材料多模态数据集的全面质量治理提供了可操作的方法论基础。

3 基于 AI-Ready 材料时序数据集完整性的案例分析

随着高通量实验与在线监测技术的普及，材料时序数据的体量正以前所未有的速度增长，然而，材料 AI-Ready 时序数据集在采集与构建过程中极易因传感器间歇性故障等问题，产生数据缺失，需要建立全面、可靠的评价方法以识别时序数据集存在的完整性缺陷。因此，本文以材料 AI-Ready 时序数据集完整性为典型案例，设计面向模拟数据集和真实数据集的实验，旨在验证所提出评价框架的可行性与可操作性；其余模态及维度的体系构建可以遵循相同方法论进行实现，例如我们曾构建了 AI-Ready 材料结构化数据质量评价体系，有效识别了公开材料结构数据集中 17

个存在质量问题的数据集^[4]。

3.1 评价对象、评价条件和评价方法的构建

本节从 MAT-DQE 统一体系出发，构建面向 AI-Ready 材料时序数据完整性的评价对象、评价条件和评价方法，具体如下：

(1) **数据模态形式化定义：** AI-Ready 材料时序数据集 D_{ts} 可形式化为一个四元组 $D_{ts} = (\mathcal{T}, \mathcal{V}, \mathcal{E}, \mathcal{M})$ 。其中， $\mathcal{T} = \{t_1, t_2, \dots, t_N\}$ 为时间索引序列，由采样起始时间 t_s 、结束时间 t_e 和采样频率 t_f 确定； $\mathcal{V} = \{v_1, v_2, \dots, v_m\}$ 为监测变量集合，每一变量 v_i 对应一个时间序列 $v_i = \{(t, y_i(t)) | t \in \mathcal{T}_i \subseteq \mathcal{T}\}$ ，其中 $y_i(t)$ 为变量 v_i 在时刻 t 的观测值， $\mathcal{T}_i$ 为该变量实际存在有效记录的时间点子集； $\mathcal{E} = \{e_1, e_2, \dots, e_k\}$ 为数据存在的关键事件集合，每一事件 e_k 关联一个时间戳 τ_k 和一个容许时间窗口 δ_k ； $\mathcal{M}$ 为元数据集，包含采样参数、传感器规格、实验环境条件和预处理操作记录等完整性相关描述信息。

(2) **统一评价体系的继承：** 基于步骤一的形式化定义， DM_{mod} 为材料时间序列模态， AS_{mod} 为包括材料老化寿命预测、循环充放电容量衰减建模、热失重动力学分析和腐蚀速率预测等。该模态在实验采集过程中最常出现的完整性缺陷包括：传感器间歇性故障导致时间点随机缺失，采集系统整体停摆造成特定时段数据空白，关键行为节点（如拐点、极值点和相变点）附近记录稀疏或丢失，多源传感器时间戳未对齐，以及实验元数据记录不全等。籍此，依据规则一，继承 MAT-DQE 定义的完整性顶层维度，该定义蕴含两个核心要求：其一，数据集矩阵内部的值是否缺失；

其二，辅助理解与复现数据的元数据是否完备。依据规则二，评价条件沿用四元组结构，将 DM 实例化为 D_{ts} ， AS 实例化为“机器学习模型训练” AS_{ts} 。依据规则三，评价方法沿用式(2)的计量映射和式(3)的平均综合评分结构。

(3) 评价维度与计量指标发展：依据规则四，对完整性维度进行向下发展。具体而言，一方面，对于“数据集矩阵内部的值是否缺失”的要求，结合前述分析的常见缺失问题，可以延伸出四种不同缺失情形：一是缺失在时间轴上均匀散布的点状缺失，源于传感器偶发丢包或随机干扰，记为时间点完整率；二是缺失集中于特定时间区间的段状缺失，源于采集系统整体

阶段性停摆，记为有效时段覆盖率；三是缺失发生于关键事件集合 $\mathcal{E}$ 所标注的关键时刻附近的缺失，会导致数据对材料行为建模的价值严重折损，记为关键事件记录完整率。此外，当多个监测变量来自不同采集系统时，还需考虑各变量时间索引的不一致导致的结构性缺失，会直接阻碍多变量联合建模，记为多源时间对齐完整率。另一方面，对于“辅助理解与复现数据的元数据是否完备”的要求，将其直接投射到元数据集，核验采样参数、传感器信息、实验环境条件和预处理记录等关键描述项的填写完整性，记为元数据完整率。衍生出的五个子指标定义与形式化表达如下表3所示。

表3 材料AI-Ready时序数据完整性计量指标定义

子指标	定义	计量指标
时间点完整率	在期望时间索引 T 上,各监测变量有效记录点的平均覆盖率	$m_1 = \frac{1}{m} \sum_{i=1}^m \frac{ \mathcal{T}_i }{ \mathcal{T} }$
有效时段覆盖完整率	在 K 个期望时段中,所有变量连续有效覆盖的时段比例	$m_2 = \frac{1}{K} \sum_{k=1}^K C(I_k)$
关键事件记录完整率	关键事件集合 $\mathcal{E}$ 中,容许窗口内所有变量均有有效记录的比例	$m_3 = \frac{ \{e_k \in \mathcal{E} \forall v_i \in \mathcal{V} [\tau_k \pm \delta_k] \cap \mathcal{T}_i \neq \emptyset\} }{ \mathcal{E} }$
多源时间对齐完整率	在统一时间网格 $\mathcal{T}_{grid}$ 上,所有变量同时有效的时间点比例	$m_4 = \frac{ \{t \in \mathcal{T}_{grid} \forall v_i \in \mathcal{V} [t \in \mathcal{T}_i]\} }{ \mathcal{T}_{grid} }$
元数据完整率	元数据项集合 $\mathcal{M}$ 中已填写且可信的项数占比	$m_5 = \frac{ \mathcal{M}_{valid} }{ \mathcal{M} }$

其中, I_k 为 K 个期望连续区间划分后的第 k 个区间, $C(I_k) \in \{0, 1\}$ 为判定函数,若 I_k 内存在长度大于等于 L_{min} 的连续子段,且该子段上所有监测变量均有有效记录,则 $C(I_k) = 1$, 否则为0。

上表3中的五个子指标的取值范围均为 $[0, 1]$, 归一化方向相同, 既继承了 MAT-DQE 完整性的内涵, 又面向时序模态的特殊缺失模式和材料领域需求进行了针对性发展, 实现了完整性计量指标的设计。进一步地, 依据规则五, 子指标向量

$\overrightarrow{m_{完整性}} = (m_1, m_2, m_3, m_4, m_5)$ 的聚合函数采用两阶段设计: 先将反映数据值缺失的四个子指标 (m_1, m_2, m_3, m_4) 以调和平均聚合, 以强化短板效应; 再将聚合结果与反映描述信息完备性的 m_5 进行加权平均, 权重分别设为 $\alpha = 0.8$ 和 $1 - \alpha = 0.2$, 以体现数据值完

整性在整体评价中的主导地位。同时，考虑到材料时序实验中关键事件标注存在一定不确定性，将关键事件记录完整率 m_3 的阈值 τ 设定为0.6，若低于阈值则触发式(3)中的硬阈值拦截。

(4) 计量算法的发展与评价条件完善：依据规则六，为各子指标开发计量算法。依据前述定义，构建如下算法库 A_b ：有效时段覆盖率采用滑动窗口算法；关键事件记录完整率采用领域规则嵌入的节点匹配算法（例如针对锂电池老化曲线，采取Kneedle算法）；时间点完整率、多源时间对齐完整率、元数据完整率无需设计计量算法。在进行计量前，若存在显式时间列，解析为统一格式的 $\mathcal{T}$ ；若仅有行序且元数据提供了采样频率，由行序推断均匀 $\mathcal{T}$ ；若既无时间列也无采样频率说明， $\mathcal{T}$ 按行序构建并将此信息缺失计入 m_5 的扣分项。结合执行评价的硬件平台与软件环境 S_b ，最终形成评价条件 $EC_b = (D_b, AS_b, A_b, S_b)$ 。

(5) 封装、测试与文档化：将上述计量指标和计量算法封装为模态专属评价函数 $\mathcal{M}_b(D_b, EC_b)$ ，使用涵盖已知缺失模式的参考数据集进行系统测试，验证硬阈值触发机制与评分单调性，并撰写结构化规范文档。

为更清晰地展示本文方案相较于现有评价方案的优势，表4从子指标设计、阈值约束、指标聚合方式及领域知识参与程度四个维度进行对比。现有评价方案普遍仅关注时间点完整率和有效时段覆盖完整率两项基础指标，未将、关键事件记录完整率和多源时间对齐完整率纳入评价框架，因而难以识别关键行为节点缺失及多传感器同步偏差等材料领域时序数据特有的质量缺陷。此外，本文对子指标设计了硬阈值，采用调和平均作为聚合方式并全流程嵌入材料领域知识，形成了更具模态适配性的材料 AI-Ready 时序数据完整性评价方案。

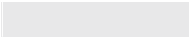
表4 现有时序完整性评价方法与本文方法的比较

对比维度		文献[24] (2024)	文献[28] (2024)	文献[29] (2026)	本文 (2026)
子指标设计	时间点完整率	□	□	□	□
	有效时段覆盖完整率	□	□	□	□
	关键事件记录完整率	--	--	×	□
	多源时间对齐完整率	--	×	--	□
	元数据完整率	×	×	--	□
	是否有阈值约束	×	×	□	□
指标聚合方式		主客观赋权	主观赋权	--	客观赋权
领域知识参与程度		中	低	中	高

其中，“×”代表该论文没有设计该指标/没有阈值约束、“□”代表该论文设计了设计该指标/有阈值约束、“--”代表该论文通过其他方式讨论了这个维度。

值得说明的是，材料时序数据中常见的不等距采样本身不构成完整性缺失。在材料实验和应用中，材料行为变化剧烈时

加密采样、平稳时稀疏采样等自适应采样策略，是合理的采集设计，其采样策略和采样结果的合理性应分别由 MAT-DQE 中



的规范性和均衡性评估。传感器丢包等异常导致原均匀采样转变为非均匀剩余点，则本质为点状缺失，已由时间点完整率 m_1 评价。然而，不等距采样对有效时段覆盖率的阈值设定提出了要求，即 L_{min} 应依据采样间隔调整为最低连续记录点数，以确保判定逻辑在不同采样密度下保持等价严格性。

3.2 面向 AI-Ready 材料时序模拟数据集的完整性评价

为验证完整性评价方案对各类缺失模式的识别能力，本节构造模拟基准数据集并注入可控缺陷进行实验。同时，固定随机种子为 42 以保证可复现。

3.2.1 模拟基准数据集构建

实验创建一个模拟锂电池 CCCV 循环充放电老化测试的多元时序数据集 D_0 （以下简称，模拟基准数据集），该数据集在结构上具有两层时序嵌套关系：顶层为 100 个完整的充放电循环周期，按时间顺序依次排列；底层为每个循环内部以固定采样频率记录的时间序列，以时间戳为索引贯穿全部循环。数据集共包含 4 个监测变量和 12940 条记录。充放电协议设定为：充电阶段采用 CCCV 模式，即先以恒定电流充电至截止电压 4.2 V，再以该电压下恒压充电至电流降至 0.05 C；放电阶段采用 CC 模式，即以恒定电流放电至截止电压 3.0 V。循环次数设定为 100 个完整充放电周期，采样频率为每 60 秒一次。四变量定义如表 5 所示。

表 5 模拟基准数据集变量定义

变量	符号	物理含义	取值范围	精度
v_1	时间戳	相对测试起始时间的秒数	0 ~ 776340 s	整数
v_2	电压	电池端电压	3.00 ~ 4.20 V	2位小数
v_3	电流	充放电电流	-1.00 ~ +1.00 C	2位小数
v_4	容量	累积放电容量	0.00 ~ 100.00 mAh	2位小数

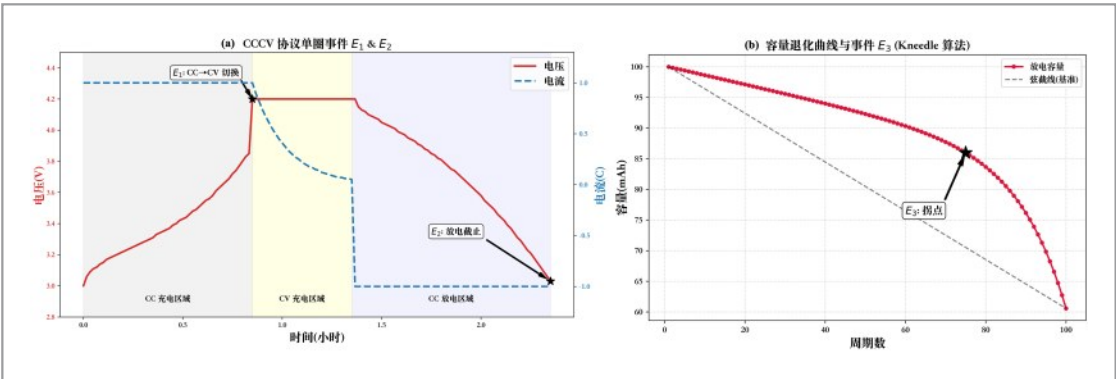


图 3 模拟基准数据集及关键事件示意图。(a) 每个循环中的充放电协议；(b) 总退化曲线

如上图3所示, 关键事件集合 $\mathcal{E}$ 由以下三类共201个关键事件构成: 第一类为每个循环中电压首次达到截止电压4.2 V的时刻, 用于区分恒流充电容量与恒压充电容量的分界, 直接反映电池内阻增长和极化状态演变, 共计100个; 第二类为每个循环中电压降至放电截止电压3.0 V的时刻, 标志着该循环放电容量的最终确定, 是容量衰减曲线的基本观测节点, 共计100个; 第三类为容量加速衰减拐点, 即电池容量衰减速率由缓慢线性衰减转变为加速非线性衰减的转折点, 是电池寿命预测中最关键的特征节点, 共计1个。每个关键事件的容许时间窗口 δ_k 设为 ± 2 个采样间隔。元数据集 $\mathcal{M}$ 包含采样频率、传感器型号与精度、环境温度、数据采集软件版本及预处理操作记录5项, 全部填写完整且可信。基准数据集不存在任何质量缺陷, 预期五个子指标得分均为1.00, 综合评分 $m_{\text{完整性}} = 1.00$ 。

3.2.2 模拟缺失数据集构建

为验证本文提出的完整性评价方案能够有效识别不同模式、不同严重程度的完整性缺陷, 对基准数据集分别设置了以下六组缺陷注入的方式, 进而构建了14个缺失数据集。

组一：点状缺失。在基准数据集上, 随机删除电压、电流和容量三个变量各5%、10%、15%、20%、25%、30%时间点的数据值, 模拟传感器偶发丢包。

组二：段状缺失。在基准数据集上, 随机选取1个、2个、3个连续时段(长度占总时长的10%), 删除该时段内所有变量的全部记录, 模拟采集系统阶段性停摆。

组三：关键事件缺失。在基准数据集上, 随机选取10组、30组CC→CV充电切换点和放电截止点, 删除其容许窗口

$[\tau_k|2\delta_k]$ 内所有变量的记录。

组四：多源对齐缺失。在基准数据集上, 保持所有数值不变, 将电压和电流两个变量的时间戳分别向后偏移3秒和5秒, 模拟多传感器时钟不同步。

组五：元数据缺失。在基准数据集上, 保持所有数据值不变, 随机删除元数据集 $\mathcal{M}$ 中1项(如传感器精度), 模拟实验记录不全。

组六：复合缺失。在基准数据集上, 同时注入点状缺失(5%随机缺失)、段状缺失(1个占总时长10%的时段删除)和关键事件缺失(随机删除10组CC→CV充电切换点和放电截止点的记录), 检验综合评分能否合理反映多维缺陷的叠加效应。

3.2.3 评价结果与分析

模拟实验的计算环境为 macOS 26.3 (Apple M5/10 核 CPU/24GB 内存), Python 3.9.6。采用本文提出的完整性评价方案对基准数据集及14个缺失数据集进行评价, 其中时间分段数为 $K=100$ 、最小连续有效长度 $L_{\min}=5$, 各子指标得分与综合得分如表6所示。

从上表6可以发现, 各子指标对目标缺陷响应灵敏且协同互补。组三关键事件缺失仅导致 m_3 显著下降, m_1 与 m_2 几乎不受影响, 验证了领域知识嵌入指标的独立诊断能力, 即尽管时序数据在数值上近乎完整时, 关键节点处的缺失仍被单独标记。组二段状缺失中, m_1 与 m_2 同步递减, 二者协同可有效区分点状缺失或者段状缺失。同时, 组一点状缺失在缺失率达20%时综合评分归零, 其直接原因是随机缺失波及部分关键事件窗口致使 m_3 跌破0.60阈值。这恰模拟了真实场景中传感器偶发丢包意外损害关键行为节点记录的风险, 即便整体缺失率有限, 致命位置的缺失仍被准确

表6 模拟数据集子指标得分和综合得分

实验设置		m_1	m_2	m_3	m_4	m_5	综合得分
组一	基线	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	点状缺失(5%)	0.9500	1.0000	0.8607	0.8566	1.0000	0.9303
	点状缺失(10%)	0.9000	1.0000	0.7463	0.7294	1.0000	0.8636
	点状缺失(15%)	0.8500	1.0000	0.6219	0.6174	1.0000	0.7921
	点状缺失(20%)	0.8000	0.9000	0.5373	0.5168	1.0000	0.0000
	点状缺失(25%)	0.7500	0.6500	0.3731	0.4308	1.0000	0.0000
	点状缺失(30%)	0.7000	0.3300	0.3383	0.3518	1.0000	0.0000
组二	段状缺失(1 段)	0.9000	0.9100	0.9005	1.0000	1.0000	0.9407
	段状缺失(2 段)	0.8000	0.8200	0.7960	1.0000	1.0000	0.8771
	段状缺失(3 段)	0.7000	0.7300	0.6915	1.0000	1.0000	0.8102
组三	关键事件缺失(10 组)	0.9985	1.0000	0.9005	1.0000	1.0000	0.9782
	关键事件缺失(30 组)	0.9954	1.0000	0.7015	1.0000	1.0000	0.9223
组四	多源对齐缺失	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
组五	元数据缺失	1.0000	1.0000	1.0000	1.0000	0.8000	0.9600
组六	复合缺失	0.7695	0.9100	0.6567	1.0000	1.0000	0.8502

捕获。此外，聚合函数设计有效抑制了短板掩盖。组五元数据缺失仅 m_5 降至 0.80，综合评分维持 0.96，与聚合函数中的权重设计一致。组四中 m_2 至 m_4 归零则直接导致综合评分为零，体现了方案对结构性致命缺陷的准确评价。

进一步地，在控制缺陷注入方式与基础指标计算不变的前提下，对综合得分的参数 α 和 τ 进行参数敏感性分析，其中 $\alpha \in [0.50, 0.95]$ 、 $\tau \in [0.30, 0.90]$ 均以 0.05 为步长。同时，定义参数邻域为 $\alpha \in [0.70, 0.90]$ 、 $\tau \in [0.50, 0.70]$ ，评价结果得分结果如下图 5 所示。

上图 5(a) 呈现了以 τ 为主的分层结构，同一 τ 下改变 α 对平均得分影响有限，表明体系对关键事件硬阈值更敏感、对权重相对不敏感；默认参数 (0.8, 0.6) 处于中高得分区域而非极端尖峰。图 5(b) 中，默认参数邻域及更广范围内 Spearman 系数接近 1，说明参数扰动下各缺陷场景的相对优劣次序基本保持不变。图 5(c) 进一步显示，绝大多数场景在默认设定与邻域扰动下得

分高度一致，波动主要来源 m_3 触发 τ 边界而导致的硬拦截，而这正对应阈值机制的预期行为。因此，默认参数落在结论稳定的参数区域内，实验结果对参数小幅变化不敏感。

综上，从模拟实验可以看出，本文提出的完整性评价方案能够有效区分五类缺失缺陷，各子指标各司其职、协同互补，阈值机制准确有效、聚合函数设计合理，为下一节真实数据集的评价应用提供了可靠的参数设置和基准参照。

3.3 面向 AI-Ready 材料时序真实数据集的完整性评价

为进一步验证本文完整性评价方案在真实场景下的适用性，本节选取 14 个涵盖电池循环老化、车辆传感器监测及材料退化多源记录等典型应用场景的真实时序数据集^[30-41]进行完整性评价，按引用顺序分别记为 Dataset01-Dataset14。实验沿用 3.1 节的评价方案，考虑到与模拟实验在数

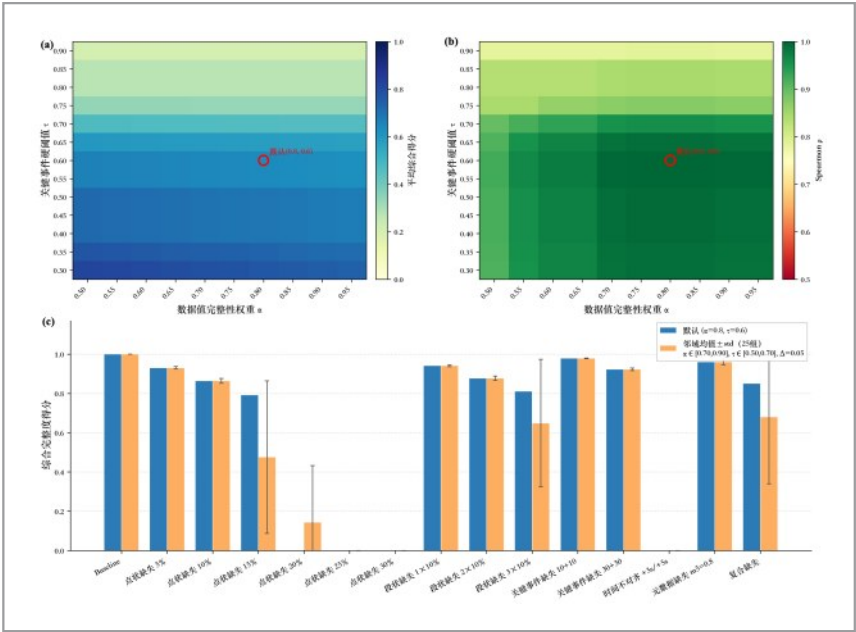


图5 参数敏感性分析。(a) 不同参数下缺失数据集平均得分热力图;(b) 不同参数相对默认参数的缺失数据集得分排序Spearman热力图;(c) 不同参数下数据集综合得分柱状图

据量上的差异，时间分段数设置为 $K=10$ 、最小连续有效长度 $L_{min} = \max\left(5, 0.8 \frac{|T|}{K}\right)$ ，其余参数配置 3.2 节一致。鉴于真实数据

集普遍缺少人工标注的关键事件，本实验采用拐点自动检测算法 Kneedle 检测数据中的拐点。最终评价结果如图 4 所示。

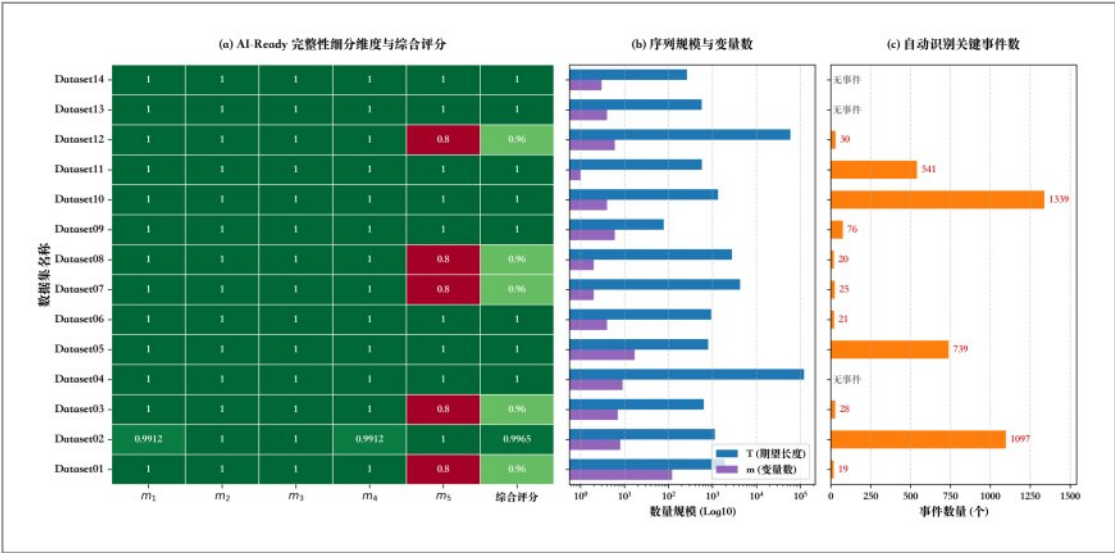


图4 真实数据集概况及完整性评价结果。(a) 14个数据集指标及总体得分热力图;(b) 14个数据集的变量数和时序长度;(c) 评价过程中识别出的关键事件数量

从图4可以看出,整体而言,当前公开的AI-Ready材料时序数据集的完整性质量较高。在数据值完整性层面, Dataset02由于缺失cycle=5的数据,导致了 m_1 与 m_4 同步降至0.9912。同时,元数据完整率是真实数据集中唯一产生区分度的维度。Dataset01、03、07、08、12五个数据集因以行序替代显式时间列,无法确认采样频率、时间戳格式等关键元数据项, m_5 被扣除0.2分。这反映了材料时序数据中存在的普遍问题:数据值自身的记录往往较为完整,但支撑数据复现与跨团队复用的元数据常被忽视记录。

值得说明的是,本实验中 m_3 在全部数据集上均为满分,这是自动拐点检测策略与数据集自身质量共同作用的结果。自动检测算法从各变量中识别出拐点并汇总为关键事件集合,14个数据集的检测结果均未出现在关键事件容许窗口内缺失记录的情形。这与模拟实验中定向注入关键事件缺失后 m_3 即触发硬阈值拦截的结论并不矛盾:基于模拟数据的实验验证了指标的灵敏度,基于真实数据集的实验则表明当前所选数据集在关键节点处记录完整。未来面向更广泛的时序数据源时,若关键事件标注由领域专家参与, m_3 的鉴别力将更为显著。

4 结束语

本文针对AI-Ready材料多模态数据集质量评价中理论性不足、跨模态跨维度评价割裂、缺少领域知识参与等问题,构建了AI-Ready材料多模态数据集质量统一评价体系MAT-DQE,并提出基于“继承-发展”的分模态数据集质量评价体系构建方法。主要内容与贡献如下:

第一,建立了AI-Ready材料多模态数据集质量统一评价体系MAT-DQE。基于评价学的五大公理,明确了包含13个质量维度的评价对象、四元组形式的评价条件以及计量驱动的评价方法,为分模态数据质量评价提供了公理化的统一框架,奠定多模态评价的理论基础。

第二,提出了基于“继承-发展”的分模态数据集质量评价体系构建方法。确立了六大规则与五步构建流程:通过继承MAT-DQE的评价对象及维度、评价条件和评价方法,保障跨模态评价结果的可比较性;通过发展模态专属的维度、计量指标和计量算法,从而构建内涵统一、模态适配的分模态数据质量评价体系。

第三,以材料时序数据完整性为典型案例,完整展示了从MAT-DQE到分模态评价体系的构建全过程。将完整性维度的双重内涵发展为时间点完整率、有效时段覆盖完整率、关键事件记录完整率、多源时间对齐完整率和元数据完整率五个子指标,并设计了调和平均聚合与硬阈值拦截机制。模拟实验验证了各子指标对五类缺失模式的灵敏识别与硬阈值的精准拦截能力;14个真实数据集的评价实验表明,方案能够在数据值普遍完整的条件下甄别出元数据缺失等质量缺陷,实现了跨数据集的质量分层。

未来还可以关注以下两方面:第一,讨论在数据质量评价时,如何建立等价评价条件或最小等价评价条件,以夯实跨模态、跨场景下评价结果可比较性的理论基础;第二,当前案例聚焦于时序数据的完整性维度,其余模态(如图像、文本、晶体结构等)以及其他维度(如准确性、一致性、洞察力等)的评价指标设计与体系构建有待逐一展开,尤其是面向多模态数据集联合评价时,模态间的质量维度尚未

得到明确关注。上述方向的持续研究，将有望进一步完善 AI-Ready 多模态数据集质量评价体系，为数据质量治理与 AI 可信评测提供更加坚实的理论支撑与方法工具。

[AI工具声明]

本论文在撰写过程中未使用任何人工智能辅助工具（包括但不限于大语言模型），全部内容均由作者团队独立完成。

参考文献：

[1] Polat C, Kurban H, Serpedin E, et al. Beyond atomic geometry representations in materials science: a human-in-the-loop multimodal framework[C]// Proceedings of the 13th International Conference on Learning Representations (ICLR). Singapore, 2025.

[2] Dreger M, Malek K, Eikerling M. Large language models for knowledge graph extraction from tables in materials science [J]. Digital Discovery, 2025, 4(5): 1221-1231.

[3] Shahzad K, Mardare A I, Hassel A W. Accelerating materials discovery: combinatorial synthesis, high-throughput characterization, and computational advances[J]. Science and Technology of Advanced Materials: Methods, 2024, 4 (1): 2314561.

[4] Liu Y, Yang Z W, Zou X X, et al. A general framework to govern machine learning oriented materials data quality [J]. Materials Science and Engineering: R: Reports, 2025, 162: 100821.

[5] Liu Y, Yang Z W, Zou X X, et al. Data quantity governance for machine learning in materials science[J]. National Science Review, 2023, 10(7): nwad125.

[6] Liu Y, Ma S C, Yang Z W, et al. Domain knowledge-assisted materials data

anomaly detection towards constructing high-performance machine learning models[J]. Journal of Materiomics, 2025, 11(6): 101066.

[7] Liu Y, Zou X X, Ma S C, et al. Feature selection method reducing correlations among features by embedding domain knowledge[J]. Acta Materialia, 2022, 238: 118195.

[8] Liu, Y, Wang S Y, Yang Z W, et al. Auto-MatRegressor: liberating machine learning alchemists[J]. Science bulletin 2023

[9] Liu Y, Liu D H, Ge X Y, et al. A high-quality dataset construction method for text mining in materials science[J]. Acta Physica Sinica, 2023, 72(7): 070701.

[10] Liu Y, Liu, Zuo W, et al. A closed-loop framework for material text annotation correction and validation based on large language models[C]//Proceedings of the International Joint Conference on Neural Networks (IJCNN). 2026.

[11] Liu Y, Yu Z Y, Liu Z T, et al. Improving disentanglement in variational auto-encoders via feature imbalance informed dimension weighting[J]. Knowledge-Based Systems, 2024, 263: 110261.

[12] Yu Z Y, Yang Z W, Zhang W, et al. Focus on latent: toward a meaningful and interpretable latent representation for diffusion models[J]. Information Processing & Management, 2026, 63(7), 104854.

[13] Zhan J F, Wang L, Gao W L, et al. Evaluatology: The science and engineering of evaluation[J]. BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 2024, 4(1): 100162.

[14] Bernardo M C, Jo o V S, Marcelo C S, et al. Optimizing deep neural networks for nuclear power plant temperature estimation: A study on feature importance

- and outlier detection[J]. Progress in Nuclear Energy, 2025, 191: 106039.
- [15] Li K M, Persaud D, Choudhary K, et al. Exploiting redundancy in large materials datasets for efficient machine learning with less data[J]. Nature Communications, 2023, 14(1): 7845.
- [16] Xu P, Ji X, Li M, et al. Small data machine learning in materials science[J]. npj Computational Materials, 2023, 9(1): 85.
- [17] Khan A, et al. Leveraging generative adversarial networks to create realistic scanning transmission electron microscopy images[J]. npj Computational Materials, 2023, 9(1): 85.
- [18] Liu N, Jafarzadeh S, Lattimer B Y, et al. Harnessing large language models for data-scarce learning of polymer properties[J]. Nature Computational Science, 2025, 5: 245–254.
- [19] Wang R Y, Strong D M. Beyond accuracy: what data quality means to data consumers[J]. Journal of Management Information Systems, 1996, 12(4): 5–33.
- [20] Luo C X, Xiong H X, Ye Y Z, et al. TDQE: a quality evaluation method for text data in deep learning[J]. Big Data Research, 2025, 11(6): 95–107.
- [21] 崔丁洁. 面向语义的文本质量评价研究[D]. 哈尔滨: 哈尔滨工业大学, 2022.
- Cui D J. Research on Semantic-Oriented Text Quality Evaluation[D]. Harbin: Harbin Institute of Technology, 2022.
- [22] 谢沛文. 面向知识图谱的质量评估技术研究[D]. 南京: 东南大学, 2022.
- Xie P W. Research on Quality Assessment Technology for Knowledge Graphs[D]. Nanjing: Southeast University, 2022.
- [23] 曹玉东, 刘海燕, 贾旭, 等. 基于深度学习的图像质量评价方法综述[J]. 计算机工程与应用, 2021, 57(23): 27–36.
- CAO Y D, LIU H Y, JIA X, et al. Review of Image Quality Assessment Methods Based on Deep Learning[J]. Computer Engineering and Applications, 2021, 57(23): 27–36.
- [24] 宋洪涛, 于江生, 韩启龙. 工业多元时序数据质量评估方法[J]. 计算机应用, 2024, 44(6): 1743–1750.
- Song H T, Yu J S, Han Q L. Quality Assessment Method for Industrial Multivariate Time Series Data[J]. Journal of Computer Applications, 2024, 44(6): 1743–1750.
- [25] 刘小戈. 多维传感器时序数据的质量评估及分类研究[D]. 哈尔滨: 哈尔滨工程大学, 2023.
- Liu X G. Research on Quality Assessment and Classification of Multidimensional Sensor Time Series Data[D]. Harbin: Harbin Engineering University, 2023.
- [26] 尤祺, 袁堂晓, 汪惠芬. 基于神经网络的工业时序数据质量管理方法[J]. 机械制造与自动化, 2022, 51(3): 96–99, 112.
- You Q, Yuan T X, Wang H F. Industrial Time Series Data Quality Management Method Based on Neural Network[J]. Machine Building & Automation, 2022, 51(3): 96–99, 112.
- [27] Alampara N, Schilling-Wilhelmi M, Jablonka K. Lessons from the trenches on evaluating machine-learning systems in materials science[J]. Computational Materials Science, 2025, 248: 113452.
- [28] 忤丹丹, 梅健, 李明, 等. 工程建设环境时序数据质量评价模型研究[J]. 环境科学与管理, 2024, 49(12): 168–173.
- Wu D D, Mei J, Li M, et al. Research on quality evaluation model of time-series data for engineering construction environment[J]. Environmental Science and Management, 2024, 49(12): 168–173.
- [29] 刘晋瑞, 刘雪鹏, 刘速, 等. 面向炼化生产运行的高质量多源时序数据集构建方法研究[J]. 化工学报, 2026, 1–14.

Liu J R, Liu X P, Liu S, et al. Research on construction method of high-quality multi-source time-series dataset for refining and chemical production operation[J]. CIESC Journal, 2026, 1-14.

[30] Zhang Y, Tang Q, Zhang Y, et al. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning[J]. Nature Communications, 2020, 11(1): 1706.

[31] Battery Archive. Battery Archive [EB/OL]. <https://www.batteryarchive.org/>, 2021

[32] NASA Ames Research Center. NASA Prognostics Center of Excellence Data Set Repository [EB/OL]. <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>, 2007

[33] Deng Z, Xu L, Liu H, et al. Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles[J]. Applied Energy, 2023, 339: 120954.

[34] Zhu J, Wang Y, Huang Y, et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation[J]. Nature Communications, 2022, 13(1): 2261.

[35] Center for Advanced Life Cycle Engineering (CALCE). CALCE Battery Research Group [EB/OL]. <https://web.calce.umd.edu/batteries/>, 2011

[36] Lu J, Xiong R, Tian J, et al. Battery Degradation Datasets (Two Types of Lithium-ion Batteries) [EB/OL]. Data Mendeley, V1, 2023. doi: 10.17632/v8k6bsr6tf.1.

[37] Birkl C, Howey D. Oxford Battery Degradation Dataset 1 [EB/OL]. University of Oxford, 2017. doi: 10.5287/ora-pb591qk7p.

[38] Ma G, Xu S, Jiang B, et al. Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning[J]. Energy & Environmental Science, 2022, 15(10): 4083-4094.

[39] Severson K A, Attia P M, Jin N, et al. Data-driven prediction of battery cycle life before capacity degradation[J]. Nature Energy, 2019, 4(5): 383-391.

[40] Xu-Ancha. Datasets [EB/OL]. <https://www.xu-ancha.com/datasets>.

[41] Materials Cloud Archive. Materials Cloud Archive: Record ed9q1-h0m33 [EB/OL]. <https://archive.materialscloud.org/records/ed9q1-h0m33>.

作者简介



左维 (2002-), 男, 上海大学计算机工程与科学学院博士研究生, 主要研究方向为多模态数据质量治理。

录用日期: XXXX-XX-XX
通信作者: 刘悦, yueliu@shu.edu.cn
基金项目: 国家自然科学基金资助项目(No. 92270124, 52073169和61936001)
Foundation Items: The National Natural Science Foundation of China (No. 92270124, 52073169 & 61936001)