

T/SAIAS

上海市人工智能行业协会团体标准

T/SAIAS 030—2025

人工智能音乐语料库建设导则

Artificial Intelligence—Construction Guidelines for Music Corpus

2025-07-24 发布

2025-07-25 实施

目次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 数据资源 1

 5.1 一般要求 1

 5.2 数据资源类型 2

6 语料生产 3

 6.1 数据采集 3

 6.2 数据清洗 3

 6.3 数据标注 3

 6.4 数据质检 4

 6.5 数据更新与维护 4

7 安全要求 5

 7.1 价值对齐 5

 7.2 数据安全 5

附录 A（资料性） 人工智能音乐语料库适用的典型应用场景 6

参考文献 8

前 言

本文件按照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》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这些专利的责任。

本文件由上海市人工智能行业协会提出并归口。

本文件起草单位：上海库帕思科技有限公司、上海音乐学院、上海人工智能行业协会、上海师范大学、东华大学、上海华东电信研究院、上海阶跃星辰智能科技有限公司、上海商汤智能科技有限公司、上海信投智能科技股份有限公司、上海文广演艺集团上海轻音乐团、上海音田文化传媒有限公司、上海智瑜数创信息科技有限公司、上海商汤智能科技有限公司、联通（上海）产业互联网有限公司、中国电信股份有限公司上海分公司、国创智造科技（上海）有限公司、中文在线集团股份有限公司。

本文件主要起草人：山栋明、黄海清、施佳樑、刘灏、钟俊浩、申林、胡茂密、郭汉杰、魏飞、沈滨、蔡岳均、常永波、王薇、田碧野、贺仁驹、孙丹诚、邓思文、张丹宁、陈紫娟、饶雪、孙雯、钱浅莹、路少卿、刘红、董德平、吴子龙、孙令晔、吴正阳、曹宇、杨闻博、曹艳平、于博、王娜、陈巧慧、孙雯、刘隽欢、宫健、赵春昊、胡力旗、王若茗、王凯靖、堵炜炜、虞祝豪、朱莉琴、杨文恺、郑更河、华静、宋佳琪、贺仁龙、郑茂宽、纪婷婷。

本文件首次制定。

首期执行单位：上海库帕思科技有限公司、上海音乐学院、上海人工智能行业协会、上海师范大学、东华大学、上海华东电信研究院、上海阶跃星辰智能科技有限公司、上海商汤智能科技有限公司、上海信投智能科技股份有限公司、上海文广演艺集团上海轻音乐团、上海音田文化传媒有限公司、上海智瑜数创信息科技有限公司、上海商汤智能科技有限公司、联通（上海）产业互联网有限公司、中国电信股份有限公司上海分公司、国创智造科技（上海）有限公司、中文在线集团股份有限公司。

本文件版权归上海市人工智能行业协会所有。未经许可，不得擅自复制、转载、抄袭、改编、汇编、翻译或将本文件用于其他任何商业目的。

人工智能音乐语料库建设导则

1 范围

本文件规定了人工智能音乐语料库的数据资源、语料生产与安全要求。
本文件适用于人工智能音乐语料库的研究、开发、维护、应用、管理等工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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 31777—2015 数字音乐元数据
GB/T 35273—2020 信息安全技术 个人信息安全规范
GB/T 35274—2023 信息安全技术 大数据服务安全能力要求
GB/T 42016—2022 信息安全技术 网络音视频服务数据安全要求
GB/T 45574—2025 数据安全技术 敏感个人信息处理安全要求
GB/T 45652—2025 网络安全技术 生成式人工智能预训练和优化训练数据安全规范
T/SAIAS 015—2024 语料库建设导则

3 术语和定义

T/SAIAS 015—2024界定的以及下列术语和定义适用于本文件。

3.1

音乐语料库 music corpus

系统收集、整理和存储与音乐相关的多模态电子数据库。

4 缩略语

下列缩略语适用于本文件。

API:应用程序编程接口 (Application Programming Interface)
BPM:拍子数 (Beat Per Minute)
DOC:文档文件格式 (Document File Format)
FLAC:无损音频压缩编码 (Free Lossless Audio Codec)
LRC:歌词文件格式 (Lyrics File Format)
MIDI:音乐数字接口 (Musical Instrument Digital Interface)
MP3:动态影像专家压缩标准音频层面3 (Moving Picture Experts Group Audio Layer III)
MusicXML:音乐扩展标记语言 (Music Extensible Markup Language)
SATB:混声四部合唱 (Soprano, Alto, Tenor, Bass)
SSA:女声三部合唱 (Soprano, Soprano, Alto)
TXT:文本文件格式 (Text File Format)
WAV:波形音频文件格式 (Waveform Audio File Format)

5 数据资源

5.1 一般要求

作为音乐语料库素材的数据资源应符合下列要求：

a) 数据模态应包含多模态数据，如音频、乐谱、歌词等；

- b) 多模态数据应保持时间同步，如音频与乐谱对齐（需精确到小节或拍点），歌词与音频时间轴对齐（需精确到字或音节）；
- c) 数据资源具有足够规模，覆盖不同音乐风格、时期和文化背景；
- d) 按地域、风格、速度、节拍、调式调性、织体、配器/乐器组合、演唱/演奏技法等维度设定数据比例，避免单一化；
- e) 确保数据采集和使用符合著作权法规定，采用“授权-采集-标注-使用”全流程合规管理，优先使用公共版权或授权数据；
- f) 数据质量应具有规范性、一致性、完整性、准确性，不含水印和不适宜内容；
- g) 数据偏见应平衡性别、地域、文化代表性（如避免西方古典音乐占比过高），时代代表性，以及流派内部的多样性。

5.2 数据资源类型

5.2.1 音频数据

音频数据以声波数字化形式存储的音乐内容载体，包含乐器演奏、人声演唱、环境音效等可听化信息。音频数据指标见表1。

表 1 音频数据的指标

序号	指标	说明
1	音频类型	见表 2
2	数据资源内容	音频文件及对应的精确元数据描述（包括但不限于来源、版权信息、表演者、录制环境等）
3	采样率	不小于 44.1kHz
4	通道数	单声道/立体声
5	量化精度	不低于 16 位
6	信噪比	≥48dB
7	文件格式	FLAC、WAV、MP3（VBR）
8	单一样本时长	30 秒以上

音频数据资源的分类说明见表2。

表 2 音频数据分类表

序号	类型	说明
1	器乐独奏录音	单一乐器演奏录音
2	器乐合奏录音	多乐器组合演奏录音
3	人声独唱录音	单一人声录音，需明确标注伴奏形式及伴奏是否包含在音频中，若包含伴奏，需有对应乐谱或伴奏音轨分离信息。
4	人声合唱录音	多人合唱录音，需明确标注伴奏形式及伴奏是否包含在音频中，若包含伴奏，需有对应乐谱或伴奏音轨分离信息，需标注声部构成（如 SATB, SSA 等）。
5	音效录音	环境噪音或自然音效录音
6	电子音乐	合成器生成的电子音乐
7	电子原声音乐	原声乐器或声音经电子音乐技术处理或与之相结合的声音
8	混合录音	混合以上两种或多种形式的录音（需依照上述分类标注混合形式）

5.2.2 乐谱数据

乐谱数据以符号、文字、数字或图形等形式，对音乐的音高、节奏、力度、速度、演奏方式等要素进行结构化记录和描述的数字化信息。乐谱数据指标见表3。

表 3 乐谱数据的指标

序号	指标	说明
1	文件格式	MusicXML、MIDI
2	内容完整性	包含音高、节奏、力度、速度、演奏记号、装饰音、踏板标记、歌词文本、段落标记、速度与表情术语版本等信息
3	版本标识	标注原版乐谱或改编版本信息、编者信息、出版信息、原始手稿/权威版本依据
4	同步关联信息	指明乐谱数据与对应音频文件、歌词文件的同步关系（如通过时间码、小节/拍号映射）

5.2.3 歌词数据

歌词数据以文本形式对音乐作品中歌曲的歌词内容进行结构化记录、存储和处理的数字化信息。歌词数据指标见表4。

表 4 歌词数据的指标

序号	指标	说明
1	语种	标注语种，支持方言及少数民族语言
2	编码	UTF-8
3	时间对齐	与音频时间轴同步
4	情感标签	标注每段歌词的情感倾向
5	文本格式	TXT、LRC、DOC

5.2.4 元数据

音乐语料库元数据应符合GB/T 31777-2015的规定。

6 语料生产

6.1 数据采集

采集方式包括但不限于流媒体API、用户上传、合作方共享等。数据采集应遵循版权合规原则，确保采集的数据来源合法，并获得必要的授权。应建立数据供应商资质审核制度，留存授权文件，禁止采集未授权的盗版数据。

6.2 数据清洗

数据清洗与预处理过程应包括去噪、去重、格式转换、标准化等步骤，以确保数据质量符合语料库建设要求。具体处理方式如下：

- a) 使用降噪算法去除环境噪声，多轨录音对齐，音频格式统一，音量标准化；
- b) 修正音符错位、调号错误等问题；
- c) 核对应时序信息及内容信息，文本内容规范化处理（统一大小写与拼写、去除特殊字符与标点等）；对照原唱音频，播放歌曲，核对应时间戳一致性，重点标记发音模糊或语速较快的段落；
- d) 基于音频指纹技术检测重复数据，确保语料库内数据唯一性≥99%。

6.3 数据标注

数据标注采用分层标注，分为基础信息层、音乐结构层、情感场景层，应符合下列要求：

- a) 基础信息层按照 GB/T 31777-2015 的规定进行结构化标注；
- b) 音乐结构层针对声波信号的声学特征或语义标签标注，宜包含以下内容：
 - 按照旋律走向进行划分与标注；
 - 速度、节拍、动态范围；
 - 调式调性、和弦进行、音色频域特性；
 - 按音乐结构划分段落；
 - 从旋律层、和声层、节奏层标注乐器类型与演奏方式。

- c) 情感场景层分段落标注音乐情感标签、适用场景，见附录 A。

6.4 数据质检

6.4.1 音频数据质检

音频数据质检从完整性和音质进行检查，应符合下列要求：

- a) 检查音频文件，不存在损坏或缺失片段。通过专业音频编辑软件对音频进行逐段播放检测，或使用音频分析工具计算音频文件时长、字节数等关键信息，与原始数据记录进行比对，若时长为 0 或字节数异常，则判定该音频文件存在完整性问题；
- b) 音频采样率、比特率等音质参数应符合标准。音频的采样率应不低于 44.1kHz，量化位数不低于 16bit，MP3 格式比特率不低于 128kbps。使用音频分析软件，提取音频音质参数，对于不符合标准的音频，标记为音质问题，进行重新采样或转换处理。

6.4.2 乐谱数据质检

乐谱数据质检从规范性、准确性、完整性、可行性和一致性进行检查，应符合下列要求：

- a) MusicXML/MIDI 文件解析完整且无报错，扫描文件不包含未授权水印或版权标记；
- b) 音符位置、休止符时长符合音乐逻辑；验证力度记号、装饰音、表情术语与音频数据同步；检查调性标记、节拍标记与元数据一致；
- c) 乐谱包含完整结构，无缺失小节，重复段落的起始/终止位置逻辑正确，合唱或合奏谱中各声部纵向对齐；
- d) 音符不超出乐器演奏范围，复杂段落的人工标注指法符合人体工程学；
- e) MIDI 文件解析的 BPM 值与音频元数据偏差 $\leq \pm 5\%$ ；乐谱与对应音频的强拍位置对齐误差 $\leq 50\text{ms}$ 。

6.4.3 歌词数据质检

歌词数据质检从准确性和完整性进行检查，应符合下列要求：

- a) 人工校对歌词内容，不存在错别字、语句不通顺等问题。利用自然语言处理技术，如文本相似度计算，将歌词与权威歌词库进行比对，识别歌词的准确性。对存在错误的歌词，进行手动修正；
- b) 检查歌词文件，不存在段落或句子缺失。通过分析歌词的结构和逻辑，判断歌词的完整性，从其他来源补充缺失部分。

6.4.4 元数据质检

元数据质检从准确性、完整性和一致性进行检查，应符合下列要求：

- a) 通过与权威音乐数据库进行数据比对，检查元数据的准确性。存在差异的数据，需进一步核实，例如通过搜索多个音乐平台的相关信息交叉验证；
- b) 检查所有元数据字段，不存在缺失情况。建立元数据字段完整性检查表，将缺失字段标记为完整性问题；
- c) 确保同一歌曲在不同元数据字段中信息一致。通过编写数据查询脚本，对元数据进行批量检查，发现不一致及时进行修正。

6.5 数据更新与维护

数据更新与维护应符合下列要求：

- a) 定期发布更新版本，新增数据占比 $\geq 5\%$ ，标注错误修正率 $\geq 95\%$ ；
- b) 建立“数据贡献者积分体系”，通过 API 开放数据检索权限，对优质贡献者给予版权收益分成；
- c) 采用“YYYY.MM.BUILD”版本号（如 2025.05.03 表示 2025 年 5 月第 3 次更新），记录数据集增量与变更日志；

- d) 音乐语料平台的维护需兼顾“数据质量”、“系统稳定”、“用户需求”三大目标，通过常态化监控、周期性优化、应急响应机制形成闭环管理。

7 安全要求

7.1 价值对齐

音乐语料的价值对齐应符合T/SAIAS 015-2024 第9章的规定。

7.2 数据安全

音乐语料库数据安全至少应执行以下要求：

- a) 涉及个人信息的，执行并符合 GB/T 35273-2020、GB/T 45574-2025 的规定；
- b) 用于提供大数据服务，应符合 GB/T 35274-2023 的规定；
- c) 用于提供生成式人工智能预训练和优化训练数据，执行并符合 GB/T 45652-2025 的规定；
- d) 用于提供网络音视频服务，应符合 GB/T 42016-2022 的规定；
- e) 应符合 T/SAIAS 015-2024 第 10 章的规定。

附录 A

(资料性)

人工智能音乐语料库适用的典型应用场景

人工智能音乐语料库可应用于音乐创作与智能编曲、音乐推荐与个性化服务、商业与品牌营销、音乐教育与学术研究、医疗与心理健康、版权管理和技术创新等场景，典型应用场景见下。

A.1 音乐创作与智能编曲**A.1.1 多模态辅助作曲**

利用音频、乐谱、歌词的多模态数据协同，通过学习语料库中不同音乐风格的旋律走向、和弦进行规律等，实现原创乐曲的生成。例如依据乐谱数据中的MusicXML格式信息，能自动创作出符合特定风格（如巴赫复调风格）的钢琴前奏曲；还能结合歌词情感标签（如“浪漫”）与音频旋律特征，生成匹配的流行歌曲主歌段落。

A.1.2 影视游戏动态配乐

根据场景需求从语料库提取音乐元素实时生成配乐。例如在游戏中，通过分析电子音乐语料库的合成器音色与节奏模式，生成随剧情推进的氛围音乐；影视剪辑软件可利用语料库中不同情绪的交响乐片段，实现一键匹配剧情的配乐生成。

A.2 音乐推荐与个性化服务**A.2.1 智能推荐系统**

基于用户行为数据，结合语料库的元数据标签（如BPM、调式调性等），实现精准的音乐推荐。例如音乐平台通过分析用户常听歌曲的频域特征，从语料库中筛选相似风格的音乐生成歌单；短视频平台依据语料库中标注的音乐结构特征，为用户推送同类型热门音乐。

A.2.2 场景化音乐生成

将环境数据与语料库元素相结合，实现实时编曲。例如智能家居系统根据用户设定的特定场景（如“瑜伽冥想”），从语料库提取相应频率的自然音效与旋律，生成定制化背景音；健身APP通过运动强度数据匹配语料库中对应节奏的鼓点，动态调整运动音乐的节奏密度。

A.3 商业与品牌营销**A.3.1 广告音效定制**

基于品牌调性，从语料库中提取合适的音乐元素，生成专属的广告音效。例如运动品牌广告通过分析语料库中高频打击乐器与电吉他音色，生成充满活力的广告配乐；奢侈品品牌利用语料库中的古典钢琴小品，制作门店背景音乐以强化品牌高端质感。

A.3.2 商业空间音乐管理

根据不同的消费场景需求，从语料库中选取合适的音乐进行播放，以影响消费者行为。例如连锁咖啡店在早餐时段播放语料库中特定BPM的爵士乐片段，营造轻松氛围；超市促销区调用语料库中快节奏电子音乐，加速顾客流动以提升购买效率。

A.4 音乐教育与学术研究**A.4.1 智能教学系统**

利用标注的乐谱与音频数据，实现交互式音乐学习。例如钢琴学习APP通过比对用户演奏音频与语料库中的专业演奏数据，实时标注错误并生成练习建议；音乐理论平台基于语料库中乐曲的结构标注，自动生成和声分析习题。

A.4.2 音乐学研究

通过对语料库中的大数据进行分析，探究音乐发展规律。例如学者对语料库中特定时期古典音乐的元数据进行统计，研究该时期音乐创作趋势；流行音乐机构分析热歌的歌词情感标签，预测特定主题在流媒体平台的传播潜力。

A.5 医疗与心理健康

A.5.1 音乐疗法定制

将语料库的声学特征与医学数据相结合，为患者定制音乐治疗方案。例如针对失眠患者，从语料库中筛选特定频率的音色与节奏，生成助眠音频；抑郁症干预系统通过分析语料库中特定调式、速度的音乐，生成渐进式情绪疏导播放列表。

A.5.2 神经反馈音乐

结合脑电数据与语料库，根据用户的心理状态实时调整音乐。例如冥想设备通过检测用户脑电波状态，在发现焦虑信号时，自动从语料库切换至标注为“宁静”的音乐；注意力训练系统根据脑波集中度数据，动态调整语料库中音乐的节奏复杂度，以维持用户专注度。

A.6 版权管理与技术创新

A.6.1 侵权检测系统

利用语料库的音频指纹与结构标注，实现对音乐作品的侵权检测。例如音乐平台通过特定算法比对上传歌曲与语料库的旋律片段，自动标记疑似侵权作品；版权机构基于语料库中乐谱的MIDI数据，检测改编作品是否超出授权范围的旋律修改。

A.6.2 虚拟内容生产

基于语料库学习，实现虚拟偶像创作等技术创新。例如虚拟歌手通过训练语料库中真人歌手的声线特征，生成原创歌曲的演唱音频；元宇宙演唱会利用语料库的乐器编排数据，实时渲染3D空间音效。

参 考 文 献

- [1]中国分类主题词表 (第二版)
 - [2]中国图书馆分类法
 - [3]中国档案分类法
 - [4] GB/T 31781—2015 数字音乐分类与代码
 - [5] WH/T 99.3—2023 非物质文化遗产数字化保护 数字资源采集和著录 第3部分:传统音乐
-