

## 多维影视剧元数据智能标注要求

Requirements for Intelligent Annotation of Multidimensional Film/TV Metadata

（征求意见稿）

XXXX-XX-XX 发布

XXXX-XX-XX 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 基本要求 ..... 1

    4.1 一般原则 ..... 1

    4.2 人工与机器协同 ..... 1

    4.3 结果导向与技术中立 ..... 2

5 元数据分类要求 ..... 2

    5.2 深度元数据 ..... 2

    5.3 周边元数据 ..... 3

    5.4 扩展元数据 ..... 4

6 智能标注要求 ..... 4

    6.1 智能标注模式 ..... 4

    6.2 标注工具与环境 ..... 4

    6.3 标注过程与质量控制 ..... 4

    6.4 数据管理与可追溯性 ..... 4

## 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专利的责任。

本文件由辽宁省数据局提出并归口。

本文件起草单位：集镁科技集团有限公司、中国联合网络通信有限公司沈阳市分公司、沈阳市市场监管事务服务中心（沈阳标准化研究院）。

本文件主要起草人：苏振宇、夏云鹏、姜楠、刘春龙、文博识、王磊、勾颖、黄尹、张洋、耿建伟、赵龙、张华刚。

本文件发布实施后，任何单位和个人如有问题和意见建议，均可通过来电、来函等方式进行反馈，有关单位将及时答复并认真处理，根据实际情况依法进行评估及复审。

归口管理部门和联系电话：辽宁省数据局（辽宁省沈阳市皇姑区崇山中路109号），联系电话：024-86916223。

标准起草单位和联系电话：集镁科技集团有限公司（沈阳市浑南区创新一路5号），联系电话：024-31335400

# 多维影视剧元数据智能标注要求

## 1 范围

本文件规定了多维影视剧元数据智能标注的要求，面向影视剧及其周边内容在人工与机器协同条件下的元数据生成、标注与管理活动，主要包括元数据的分类、各类元数据字段要求，以及适用于智能标注应用的相关管理要求。

本文件适用于电影、电视剧、网络剧、综艺、纪录片、动画片等各类影视剧及其周边信息在元数据层面的标注与管理。本文件不涉及具体算法模型或实现方案，也不涉及影视内容制作、内容审查或播控许可等业务。

## 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 42755—2023 人工智能 面向机器学习的数据标注规程
- GY/T 261—2012 广播电视数字版权管理元数据规范
- ISO/IEC 9834 信息技术标识和编码规范

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**元数据 metadata**

描述数据的数据，是对内容元素及其关系、形式、相关使用规则、义务和其他事项的结构化描述，可以嵌入内容元素或与之相关联。

[来源：GY/T 261—2012，3.1]

### 3.2

**数据标注 data labeling**

给数据样本指定目标变量和赋值的过程。

[来源：GB/T 42755—2023，3.1]

## 4 基本要求

### 4.1 一般原则

本文件用于规范多维影视剧元数据在智能标注场景下的标注与管理活动。

### 4.2 人工与机器协同

多维影视剧元数据智能标注应采用人工与机器协同的方式开展。

### 4.3 结果导向与技术中立

本文件仅对元数据标注结果及相关管理要求提出规定，不限定具体算法模型、工具链或实现方案。

## 5 元数据分类要求

### 5.1.1 作品标识类

作品标识类元数据用于唯一标识和基本识别影视剧作品，应能够支持作品溯源、区分和跨系统识别：

- 1) 源名称：由片方或版权方提供的原始作品名称，用于溯源；
- 2) 中文名称：影视剧在中国境内使用的正式名称；
- 3) 英文名称：影视剧的英文名称（如有）；
- 4) 别名：影视剧在不同地区或平台使用的译名、曾用名或宣传名称等，可为多值；
- 5) 唯一标识符：系统内部用于唯一标识影视剧作品的编码，应符合唯一性要求；
- 6) 平台标识符：公共数据库、行业平台或业务平台中用于标识影视剧作品的内容标识符（如豆瓣、IMDb 等，仅为示例），用于跨平台关联；平台类型和范围可根据实际业务需要扩展，示例不构成限定。

### 5.1.2 演职人员类

演职人员类元数据用于描述参与影视剧创作和制作的主要人员信息：

- 1) 导演：影视剧作品的导演信息，可为多个；
- 2) 编剧：影视剧作品的编剧信息，包括原创、改编或联合编剧等，可为多个；
- 3) 制片人：影视剧作品的主要制片人信息；
- 4) 演员：出演影视剧主要角色的演员信息；
- 5) 配音演员：为作品主要角色提供配音的演员信息；
- 6) 摄影、灯光、剪辑、美术、服装、化妆、音乐等制作人员：根据作品实际情况记录主要制作工种人员信息。

### 5.1.3 作品属性类

作品属性类元数据用于描述影视剧作品的类型、制作与播出属性等基本特征：

- 1) 主类型：影视剧作品所属的主要类型，如爱情、犯罪、动作、科幻、灾难、纪录片等；
- 2) 亚类型：影视剧作品所属的细分类型，如西部、公路、青春、童话等；
- 3) 制片国家/地区：影视剧作品制作时所属的国家或地区，可为多值；
- 4) 语言：影视剧作品的主要对白语言，可为多语种，包括方言；
- 5) 上映日期：影视剧作品的首映日期或平台上线日期等；
- 6) 上映地区：影视剧作品进行院线放映、影展放映或上线发布的主要国家、地区或平台；
- 7) 片长：影视剧作品的时长信息，单位为分钟；
- 8) 集数：影视剧作品的总集数信息，适用于电视剧、网络剧、动画剧集等分集播出的作品；
- 9) 剧情简介：对影视剧作品主要情节、核心冲突和整体基调的简要描述；
- 10) 分级信息：影视剧作品在不同国家或地区的官方分级信息；
- 11) 出品方：影视剧作品的主要出品或联合出品单位。

## 5.2 深度元数据

深度元数据用于描述影视剧作品在内容层面的语义特征和叙事属性,以支持对作品内容的结构化表达。

### 5.2.1 故事背景类

故事背景类元数据用于描述影视剧作品中故事发生的时间和空间背景:

- 1) 故事发生时间:作品中故事发生的公元纪年、年代、朝代、时期、季节、节日等;
- 2) 故事发生地点:作品中故事主要发生的国家、城市、地区、自然环境、人文环境或地标等;
- 3) 故事发生期限:作品中故事情节所覆盖的时间长度,如“一夜之间”“数日”“多年”等。

### 5.2.2 人物设定类

人物设定类元数据用于描述影视剧作品中主要人物的基本设定信息:

- 1) 主人公设定:主人公在作品中的身份、职业、年龄、家庭角色等基本设定;
- 2) 主人公关系设定:作品中主要人物之间呈现的关系类型,如亲属关系、师生关系、合作关系等。

### 5.2.3 题材与主题类

题材与主题类元数据用于概括影视剧作品在内容表达上的核心关注点:

- 1) 题材:作品所围绕的主要情节内容或事件类型,如爱情故事、警匪对抗、体育竞技等;
- 2) 主题:作品在整体叙事中所表达的核心思想或价值取向。

### 5.2.4 风格与情绪类

风格与情绪类元数据用于描述影视剧作品在整体表现形式和观感上的特征:

- 1) 作品风格:从整体氛围、叙事节奏或表现方式等方面对作品风格进行概括;
- 2) 观众情绪:概括作品在整体观感上可能引发的情绪体验。

## 5.3 周边元数据

### 5.3.1 宣传推广物料类

宣传推广物料类元数据用于描述与影视剧作品相关的宣传和推广信息:

- 1) 拍摄地点:影视剧作品拍摄的主要城市、区域或场景信息;
- 2) 花絮:与影视剧作品制作过程相关的幕后资料或花絮信息;
- 3) 看点:对影视剧作品中具有吸引力内容或特征的概括性描述;
- 4) 经典台词:影视剧作品中被观众广泛认知的代表性台词;
- 5) 海报:以视觉化设计为核心的影视剧宣传物料;
- 6) 剧照:从影视剧拍摄素材或正片中截取的静态画面;
- 7) 预告片:影视剧作品上映或播出前制作的宣传短片;
- 8) 原著:影视剧作品创作所依据的原始文本作品;
- 9) 原声音乐:为影视剧作品创作或选配的音乐作品集合,包括主题曲、插曲、片尾曲或配乐等。

### 5.3.2 官方信息渠道类

官方信息渠道类元数据用于记录影视剧作品相关的官方信息发布和访问渠道:

- 1) 官网地址:由影视剧作品发行方或相关单位设立的官方网站地址;
- 2) 流媒体播放地址:影视剧作品在流媒体平台上的播放访问路径。

### 5.3.3 口碑与数据类

口碑与数据类元数据用于记录影视剧作品在公开渠道中的评价和统计信息：

- 1) 评分信息：主流影视剧社区平台上的评分数值和评分人数（如豆瓣、IMDb 等，仅为示例）；
- 2) 点赞数量：影视剧作品在公开渠道中获得的点赞数量；
- 3) 评论数量：主流影视剧社区平台上的评论数量（如豆瓣、IMDb 等，仅为示例）；
- 4) 评论内容：具有代表性的专业评论或获奖评语等；
- 5) 票房数据：影视剧作品在公开渠道披露的票房收入信息。

### 5.3.4 荣誉与榜单类

荣誉与榜单类元数据用于记录影视剧作品获得的奖项和榜单信息：

- 1) 获奖情况：影视剧作品在国内外主要颁奖活动中获得的奖项或荣誉；
- 2) 榜单排名：影视剧作品入选的榜单及其对应排名信息。

## 5.4 扩展元数据

在实际应用中，元数据内容可由使用单位根据自身业务需求或行业特定需求进行定义，但应保持字段命名规范、结构清晰、取值可控，并与本标准规定的总体元数据体系保持一致性。扩展元数据应符合以下要求：

- 1) 不得破坏现有字段体系结构；
- 2) 字段命名应符合统一的命名约定，不得与已有字段冲突；
- 3) 字段应具备明确的数据类型、含义、取值范围及约束要求；
- 4) 扩展字段应可被独立识别，并与标准字段区分；
- 5) 扩展字段的启用、禁用或修改过程应记录相应的版本信息；
- 6) 新增或删除扩展字段时，应记录修改时间和版本号；
- 7) 扩展字段应与标准字段采用统一的数据交换格式。

## 6 智能标注要求

### 6.1 智能标注模式

多维影视剧元数据的智能标注应在人工参与的前提下，结合机器辅助方式开展。智能标注过程中，应根据元数据类型和业务需求，合理分配人工标注与机器处理的任务范围。

### 6.2 标注工具与环境

智能标注过程中所使用的标注工具和运行环境，应能够支持元数据字段的结构化录入、编辑和校验，并满足数据一致性和完整性的基本要求。

### 6.3 标注过程与质量控制

智能标注过程应围绕元数据标注结果开展必要的校验和复核，以保证标注内容的准确性和一致性。

当采用机器辅助方式生成或推荐标注结果时，应通过人工确认、抽样复核或其他方式对标注结果进行校验。

### 6.4 数据管理与可追溯性



智能标注过程中形成的元数据，应具备必要的版本记录和变更记录能力，以支持标注结果的追溯和核查。

当元数据发生修改、更新或替换时，应能够区分不同版本的标注结果，并保留必要的历史信息。

---