ICS 35.240
L 70

团 体 标 准

T/BSIA 00X-2024

跨媒体虚假新闻识别系统要求

Cross-media fake news identification system requirements

（征求意见稿）

2024-xx-xx 发布　　2024-xx-xx 实施

北京软件和信息服务业协会　发布

目 次

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。

本文件由北京软件和信息服务业协会提出并归口。

本文件起草单位：×××

本文件主要起草人：×××

本文件为首次发布。

跨媒体虚假新闻识别系统要求

1 范围

本文件规范详细界定了跨媒体虚假新闻识别领域人工智能建模系统的系统架构、高级技术功能需求及精细化应用功能规范。

本文件适用于跨媒体环境下，针对虚假新闻识别任务，对人工智能技术建模系统及解决方案的数据预处理策略、高效算法设计原则、精细化模型训练流程、以及智能模型管理机制的全面要求，旨在为企业构建、优化、评估及验证跨媒体虚假新闻识别系统提供权威性指导与标准化依据。

2 规范性引用文件

下列文件对于本文的应用是必不可少的。凡是注日期版本的引用文件，仅注日期版本适用本文件。凡是不注日期的引用文件，最新版本(包括所有的修改版)适用本文件。

GB/T 18030-2005 信息技术 中文编码字符集

GB/T 20273-2019 信息安全技术 数据库管理系统安全技术要求

GB/T 41867-2022 信息技术 人工智能 术语

GB/T 42135-2022 智能制造 多模态数据融合技术要求

3 术语和定义

3.1

虚假新闻 false news

特指那些出于故意目的，旨在误导公众认知或歪曲客观事实的信息，其背后往往隐藏着推动特定政治立场、商业利益或社会舆论的动机。

3.2

跨媒体虚假新闻 cross-media false news

指跨越多种媒体形态（包括但不限于文字、图片、视频、音频等）传播，利用多模态信息融合技术增强欺骗性的虚假新闻内容，这些新闻信息在不同平台间相互印证或混淆视听，加大了识别难度。

3.3

特征提取 feature extraction

在跨媒体虚假新闻识别系统中，特指针对多模态数据（文字、图像、视频帧、音频片段等）进行深度分析和处理的过程，旨在从复杂的原始数据中高效提取出对于识别虚假新闻至关重要的特征信息，以支持机器学习模型更精准地理解和区分真实与虚假的新闻内容。

3.4

模型训练　model training

跨媒体虚假新闻识别系统中，采用特定领域的已知数据集对机器学习模型进行训练，通过精细调整模型参数，优化模型在识别多模态虚假新闻信息（包括文字、图片、视频、音频等）方面的准确性。

3.5

特征选择　feature selection

针对跨媒体数据的复杂性，从原始多模态数据中精心挑选出与虚假新闻识别最为相关且最具判别力的特征集合，旨在降低模型训练成本、提高计算效率，并显著提升模型在跨媒体环境下的识别性能。

3.6

模型评估　model evaluation

采用严格独立的跨媒体测试数据集，对训练好的机器学习模型进行全面而细致的评估。评估过程不仅关注模型的总体准确性和稳定性，还深入分析模型在不同媒体形态下的表现差异，以确保模型在复杂多变的跨媒体虚假新闻环境中具备高度的鲁棒性和适应性。

3.7

准确率　accuracy

在跨媒体虚假新闻识别任务中，模型准确地将虚假新闻样本与真实新闻样本区分开来的比例，反映了模型在复杂多模态数据环境下的整体判断性能。

3.8

召回率　recall

针对所有实际被标记为虚假新闻的样本，跨媒体虚假新闻识别模型成功识别并标记为正例（即虚假新闻）的比例，衡量了模型对虚假新闻的全面捕捉能力。

3.9

精确率　precision

在所有被跨媒体虚假新闻识别模型预测为虚假新闻的样本中，实际确实为虚假新闻的比例，评估了模型在预测虚假新闻时的准确性，避免了对真实新闻的误报。

3.10

F1 分数 F1 score

作为精确率与召回率的调和平均数，F1 分数是衡量跨媒体虚假新闻识别系统性能的关键指标，它能够综合评估系统在正确识别虚假新闻与避免误报非虚假新闻之间的平衡能力。

3.11

交叉验证 cross-validation

在跨媒体虚假新闻识别系统的评估中，交叉验证是一种高级且必要的技术。它通过将大规模、多样化的数据集有效划分为多个互不重叠的子集，并循环使用每个子集作为独立的测试集，剩余部分作为训练集，从而确保系统在不同数据分布上的泛化能力和稳定性，避免了对单一数据集的过拟合现象。

4 系统架构

4.1 系统设计

跨媒体虚假新闻识别系统架构由平台层和应用层构成（见图 1）。平台层作为基础，提供了数据的导入导出、预览探索、预处理、特征工程、算法选择、模型培训与评估、模型管理、模型市场以及工作流调度等核心功能。这些功能构成了系统运行的基石，为上层应用提供了强大的数据处理和模型管理能力。应用层建立在平台层之上，专注于虚假新闻识别的具体应用，包括事件识别、事件预警、智能发现与预警以及分析决策等核心应用功能。这一层直接面向用户，提供了直观的界面和智能的分析工具，帮助用户快速识别和响应虚假新闻事件。整个系统的设计旨在通过自动化和智能化的手段，提高虚假新闻识别的效率和准确性，同时降低操作的复杂性，确保用户能够轻松地管理和利用虚假新闻识别技术。

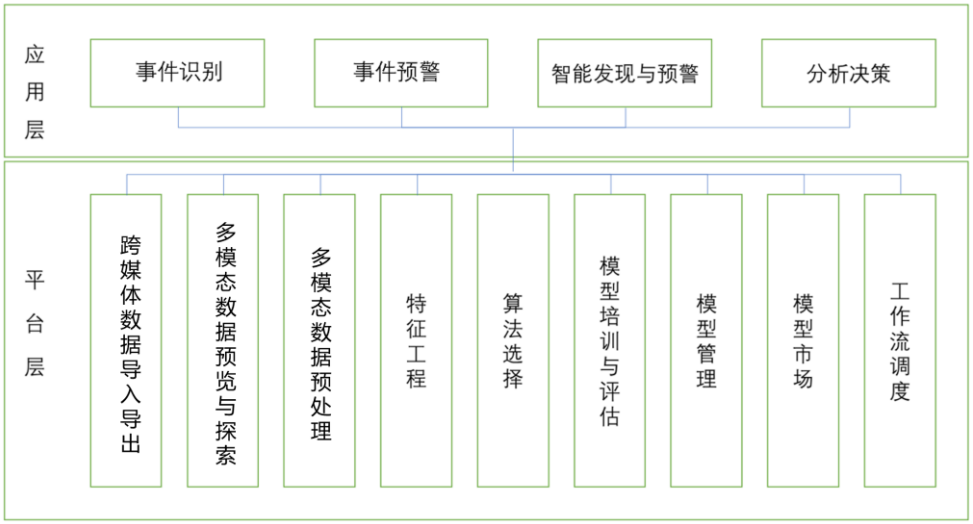

图 1 跨媒体虚假新闻识别系统架构

4.2 平台层核心功能

平台层核心功能有：

a) 数据导入导出：支持从多元化外部数据源自动导入数据，并灵活导出处理结果至各类系统或存储，保障数据流通的顺畅与高效；

b) 数据预览与探索：通过直观的用户界面，提供强大的数据可视化与分析工具，让用户能够轻松查看数据概况，深入探索数据间的关联与趋势；

c) 数据预处理：集成智能算法，自动执行数据清洗、标准化、去重等预处理步骤，显著提升数据处理效率与准确性，为后续分析奠定坚实基础；

d) 特征工程：深度剖析数据，构建针对性强、高效识别虚假新闻的特征体系；

e) 算法选择：集成多样化先进算法，灵活匹配各类虚假新闻识别任务，提升识别效能；

f) 模型训练与评估：遵循高标准流程训练模型，实施全面细致的评估，确保模型在准确性与稳定性上均达到最优；

g) 模型管理：实施严格的模型版本控制，持续监控模型性能，确保模型稳定可靠，支持快速迭代优化；

h) 模型市场：构建安全、高效的模型交易平台，鼓励用户共享高质量模型，促进跨领域知识共享与合作；

i) 工作流调度：实现全流程自动化调度，精准控制各阶段任务执行，提升系统处理效率与任务完成质量。

4.3 应用层核心应用功能

应用层核心功能有：

a) 事件识别：通过高级算法自动检测并识别虚假新闻事件，确保高效且准确。

b) 事件预警：建立实时监测系统，对潜在虚假新闻事件迅速发出预警，增强风险防范能力。

c) 智能发现与预警：融合多种智能算法，深度挖掘并即时预警虚假新闻，提升系统智能化水平。

d) 分析决策：提供基于大数据和AI的深度分析报告，辅助用户制定精准有效的虚假新闻应对策略。

5 技术功能要求

5.1 数据导入导出

5.1.1 数据导入

数据导入模块需全面兼容多样化跨媒体数据类型，包括但不限于文本、高清图像、多格式视频及音频文件，确保系统无缝对接各类数据源，提升数据处理效率与兼容性，进而增强虚假新闻识别的广泛适用性与深度分析能力。

5.1.2 数据导出

数据导出功能应灵活适配多种输出格式，支持将处理结果导出为标准化文本文件、直接对接主流数据库系统，或转换为可视化报告形式，便于用户进行深度分析、数据备份及跨平台共享，充分满足用户对数据处理结果的多样化需求。

5.1.3 数据样例

数据样例的展示和处理对于用户了解数据特征、质量以及建立模型的初步认识至关重要。在数据导入过程中，系统应该提供数据样例的预览功能，让用户能够快速浏览和了解数据的基本情况。通过数据样例，用户可以对数据的分布、特征和标签等信息有一个初步的认识，为后续的数据预处理和特征工程提供参考。

5.2 数据预览与探索

5.2.1 数据质量分析

数据质量分析是确保识别准确性的基础。通过对数据源的验证和检查，系统可以识别并处理可能存在的数据错误、缺失或异常，确保后续分析的可靠性和准确性。对于跨媒体数据

而言，这意味着需要特别关注文字、图片和视频等不同形式数据的质量问题，以确保跨媒体信息的一致性和可信度。支持对脏数据，数据缺失值、异常值等的检查。

5.2.2 数据统计分析

数据统计分析是对数据进行整体性的了解和把握。系统需要能够对数据集的规模、分布、频率等进行统计分析，以便识别出可能的模式和趋势。在跨媒体虚假新闻识别中，这种分析可以帮助确定不同类型媒体数据的比例和分布情况，为后续的特征提取和模型选择提供重要参考。支持查看数据的分布情况和统计学指标。支持图形化自定义统计分析数据。

5.2.3 数据特征分析

数据特征分析旨在精准挖掘跨媒体数据的核心特征。系统采用自然语言处理（NLP）技术提取文本的语义向量，并结合词频与 TF-IDF 算法识别关键信息。对于图片数据，系统运用深度学习模型分析颜色、纹理等视觉特征。在处理视频数据时，系统通过运动估计与对象跟踪技术捕捉帧间动态与对象行为模式，以专业手段揭露潜在的虚假信息。

5.2.4 复杂数据特征分析

针对复杂的跨媒体数据，系统深化分析层次，将语音情感识别技术融入视频语音分析中，同时结合图像内容验证与上下文语义理解，构建一个多维度的特征网络。系统利用机器学习模型整合各类特征，实现跨模态虚假信息的高精度综合判断，从而提升系统的专业性和识别准确性。

5.3 数据预处理

5.3.1 数据清洗

数据清洗环节强化了对跨媒体数据的专项处理，支持自定义清洗规则，精准识别并剔除来自不同媒体源的噪音数据、异常值及缺失项。采用先进的数据校验算法，确保数据完整性与准确性，为模型训练奠定坚实的数据基础，显著提升模型对虚假新闻识别的精确度和可信度。

5.3.2 数据变换

数据变换阶段专注于跨媒体数据的深度融合与标准化处理，支持自动将文字、图像、视频等多模态数据转换为统一格式，便于高效特征提取。同时，引入高级数据预处理技术，如自适应标准化、归一化及智能编码，增强数据一致性，优化模型训练效率与稳定性，提升跨媒体虚假新闻识别的专业性与准确性。

5.3.3 数据规约

数据规约聚焦于跨媒体数据的高效精简，采用先进的特征选择与降维策略，确保关键信息保留的同时，显著降低数据维度与复杂度。此过程专为虚假新闻识别优化，旨在提升模型处理效率与准确性，同时大幅降低计算负荷与存储需求。

5.3.4 自动化预处理

自动化预处理模块集成行业特定模板，实现数据清洗、变换、规约等流程的全自动化，减少人工干预，提升处理效率与结果一致性。通过智能算法动态调整预处理策略，确保预处理过程适应不同跨媒体数据的特性与需求。

5.3.5 预处理行业模板

行业模板可以根据不同媒体类型和特定领域的需求，定义标准化的预处理流程和参数设置，以确保数据预处理的一致性和可重复性。通过使用行业模板，可以降低虚假新闻识别技术的实施门槛，提高系统的易用性和适用性，同时也有助于促进技术的标准化和推广。

5.4 特征工程

5.4.1 特征提取流程

特征提取流程包括特征变换、特征重要性评估、特征选择、特征生成等。特征工程流程需要清晰定义，以确保在识别虚假新闻时能够有效地处理不同类型的数据。这个流程应当包括特征选择、特征转换和特征构建等步骤，以最大程度地挖掘数据中的信息。

5.4.2 特征工程自动化

特征工程自动化是提高效率和准确性的关键。系统需要具备自动化的特征工程功能，能够根据数据的特点和识别需求，自动选择和生成最相关的特征。通过自动化，可以降低人工干预的成本，并且能够更快速地响应不同数据形式和规模的变化。特征工程自动化包括自动多表扩展、自动特征变换、自动特征选择及自动特征生成等。

5.4.3 特征提取模板

针对跨媒体虚假新闻识别，系统应当提供一系列标准化的特征提取模板，涵盖文本、图像和视频等不同媒体类型的特征。这些模板应当经过充分验证和优化，以确保能够捕获到与虚假新闻相关的关键信息，从而提高识别准确性和效率。

5.5 算法选择

5.5.1 基础能力

支持针对文本、图像和视频等不同媒体类型的假新闻检测算法，算法参数可配置。

5.5.2 算法类型

集成但不限于以下算法类型：特征权重、预处理、机器学习等核心算法，NLP、CV、集成学习及深度学习前沿技术。

5.5.3 自定义算法

支持通过 Python 等实现自定义算法，支持用户自定义持续化扩展算子库。

5.5.4 实用工具库

提供文本分析工具用于对新闻文本进行分析和处理，包括自然语言处理（NLP）库、文本向量化技术、词嵌入模型等，以便于对新闻内容进行特征提取和相似度比较；提供多媒体分析工具用于对图片、视频等多媒体内容的新闻进行分析，包括图像处理库、视频分析技术等，以辅助识别虚假新闻中可能存在的图像或视频篡改；提供包含已标注的虚假新闻和真实新闻数据的数据集，用于模型训练和评估；提供用于评估虚假新闻识别模型性能的工具，包括常用的评估指标计算方法、交叉验证技术等；提供部署和集成工具用于将开发好的虚假新闻识别模型集成到实际应用中，包括模型优化和压缩技术、部署平台等。

5.5.5 算法样例库

提供章节 5.5.2 所列算法的使用样例。

5.6 模型训练与评估

5.6.1 训练过程

支持训练任务的精细控制，可以查看运行日志。训练过程中支持调试功能，可进行单步调试，断点调试。支持训练过程中间数据查看、导出。

5.6.2 资源共享

支持多个用户分组管理和共享计算资源。

5.6.3 资源管控

支持对物理资源进行虚拟化管控，可以动态进行资源的申请或释放。

5.6.4 复杂任务依赖

支持多任务之间图形化构建依赖，以构建复杂的模型训练任务及数据分析任务。

5.6.5 自动调参

支持自动调参和搜索网格，包括在给定命中率和覆盖率的要求下搜索参数输出结果，及在给定参数搜索最优结果。

5.6.6 自动建模

支持自动建模，自动选择算法及参数。

5.6.7 交叉验证

支持基于分层采样的按比例随机分配训练与测试集，确保数据分布的一致性，同时支持K 折交叉验证（K-fold Cross-Validation）以增强模型泛化能力。

5.6.8 评估指标

系统需支持全面的评估指标集，包括但不限于准确率、精确率、召回率、F1 分数以及混淆矩阵输出，以便深入分析模型性能。评估要求设定为：要求准确率大于 80%，召回率大于85%，F1 分数大于 80%。

5.6.9 评估样例库

提供所有评估算子样例。

5.7 模型管理

5.7.1 模型的版本管理

系统应提供强大的模型版本管理功能，包括历史、新建及外部导入模型保存和版本管理，支持模型架构、训练参数、性能指标等详细信息的全面查看，以及模型预测结果的深度分析与可视化展示。

5.7.2 模型导入导出

对训练好的模型进行有效的保存和加载，以便于在不同环境中进行部署和使用。支持导出 TensorFlow 的 SavedModel 或 PyTorch 的.pth 文件，提供相应的接口和方法来加载这些保存的模型，使其可以在生产环境中轻松调用和应用。

5.7.3 深度学习模型管理

支持深度学习模型导入导出和可视化查看。同时，需提供实验管理工具，支持模型训练、参数调整、性能评估及结果分析，确保实验阶段模型迭代与优化的高效进行。

5.8 模型市场

5.8.1 模型用户管理

支持高级管理员对普通用户进行精细化权限配置，包括项目访问、模型使用权限等的详细管理。

5.8.2 模型服务上架

支持经过严格审核的模型、算法、代码及自定义环境等资源在模型市场中发布，确保所有上架内容的质量及合规性。

5.8.3 模型服务上、下线

支持模型服务的动态上线与及时下线操作，同时提供全面的状态监控与日志记录功能，以保障服务的稳定性和可追踪性。

5.8.4 模型服务更新

支持动态滚动更新和智能灰度更新策略，包括流量权重的动态调整，以确保更新过程中服务的连续性和稳定性。

5.8.5 模型服务测试

实施全面的 API 功能和性能测试，包括压力测试和兼容性测试，确保模型服务在上线前的可靠性和兼容性

5.8.6 模型服务管理

提供灵活的自定义模型部署选项，自动配置 REST API 接口，支持自动扩展和智能负载均衡，同时提供详细的 API 管理界面，包括 API 的监控和日志分析功能，以优化模型服务的性能和可管理性。

5.8.7 模型服务监控

实现全方位的模型服务监控，包括实时监控模型服务内容、运行状态、实例详情与资源设置，以及精细化的调用情况与结果统计，支持预警机制和故障诊断，确保服务的高可用性和可维护性。

5.8.8 模型服务使用

提供标准化的 REST API 接口，支持灵活的参数配置与安全的数据传输，确保预测结果的准确性与及时性。同时，增强 API 的文档化和示例代码，以提升开发者的使用体验和集成效率。

5.9 工作流程调度

5.9.1 任务配置

包括定义识别目标、确定数据来源、标注标准等。首先，需明确识别目标，如识别文字、图片或视频中的虚假信息。其次，选择数据来源，涵盖社交媒体、新闻网站等多个渠道，以确保覆盖广泛的信息类型和来源。此外，标注标准至关重要，需要明确虚假信息的定义及标注方式，可能包括真实性、内容准确性等方面。

5.9.2 设计工作流

包括需求分析、数据收集与预处理、特征提取与选择、模型设计与训练以及评估与优化等环节。建立一个系统化的框架，确保从问题定义到技术实现的无缝衔接，以达到高效识别虚假新闻。

5.9.3 执行工作流

支持数据采集、模型训练、模型测试和结果分析等阶段。按照设计好的流程依次执行各项任务，确保数据的准确性和模型的稳定性，最终实现对跨媒体虚假新闻的准确识别。

5.9.4 工作流上、下线

建立标准化工作流管理流程，确保工作流上线时经过严格的质量控制与合规审查。提供灵活的下线机制，依据性能监控和用户反馈，动态调整工作流状态，保障系统运行的高效性和用户体验。

5.9.5 工作流导入导出

提供标准化的接口以兼容多种格式的工作流文件导入，包括但不限于 JSON、XML 等，增强系统的灵活性与互操作性。支持将工作流完整导出至本地，包括工作流定义、参数配置及依赖关系，确保数据的完整性和一致性。

5.9.6 工作流详情

支持查看单个工作流的每次执行时间、执行状态及其资源消耗情况，提升监控与分析能力。提供任务级日志追踪，允许查看工作流下单个任务的执行时间、状态、日志详情及异常信息，便于故障排查与性能优化。

6. 应用功能要求

6.1 事件识别

6.1.1 新闻上报

网格员在日常监控中，运用高级检测算法，对发现的疑似热点虚假新闻进行深度语义和

情感分析，智能分类并初步识别新闻主题。结合多模态信息融合技术，对系统识别出的虚假新闻进行综合评估与初步分析。

6.1.2 新闻评估

模型通过多维度特征提取，结合新闻的来源可靠性、内容一致性及历史相似案例分析，对新闻的真伪进行综合评估。优先上报置信度高且可疑度显著的新闻，确保评估的精准性和有效性。

6.1.3 新闻核查

舆情分析中心将疑似虚假新闻的核实任务派发给网格员，网格员利用专业信息检索工具，结合权威数据源，对新闻的真实性进行深度核实。结合模型分析结果，网格员将提供详细核实报告，包括核实方法、证据链及最终判断，确保核查过程的透明性和结论的可靠性。

6.1.4 新闻检索

利用深度跨媒体关联分析技术，对已被模型判别的虚假新闻进行全网追踪检索，覆盖社交网络、新闻网站、视频平台等多渠道，迅速阻断虚假信息的二次传播，并分析传播路径，以减少社会负面影响的扩散。

6.1.5 新闻溯源

采用高级数据挖掘与分析技术，追溯虚假新闻的首发源头，包括内容与发布账号。对源头账号的历史发布内容进行深度审查，识别潜在的虚假信息模式，防止同账号持续发布虚假新闻，减少社会影响。

6.1.6 事件结案

在确认虚假新闻相关事件检测完毕后，运用事件管理平台，记录关键信息，包括涉事账户、平台类型、传播途径及影响范围，形成详尽的事件报告。对事件进行系统存档与案例分析，为未来的事件处理提供数据支持与经验借鉴。

6.2 智能发现与预警

6.2.1 重点账号识别

运用大数据分析技术，精准识别频繁发布、传播虚假新闻的账号与平台，建立高风险账号黑名单，实施持续监控与智能预警。对列入黑名单的账号，实行实时内容监测与即时上报机制，确保虚假新闻的快速识别与阻断。

6.2.2 媒体平台分析

依托历史监管数据与 AI 智能分析，深入剖析各社交媒体平台，识别虚假新闻热点话题与内容模式。针对平台特定话题，部署智能实时监测系统，对潜在虚假新闻热点进行精准预警，及时通知网格员进行人工复核，提升预警的准确性和响应速度。

6.2.3 热点新闻预警

采用深度学习算法，对多平台新闻动态进行实时热度预测分析，提前识别潜在热点新闻，实施优先级监测与深度内容分析，有效预防虚假信息迅速扩散，减少社会负面影响。

6.2.4 社交群体发现

运用社交网络分析技术，结合重点账号行为模式挖掘，智能识别高关联度社交群体及虚假新闻传播节点。对关键账号执行精细化监测，结合内容智能识别技术，实现对社交群体中虚假新闻传播的精准防控。

6.3 分析决策

6.3.1 预测分析

运用跨媒体多模态分析技术，对各平台相似事件进行深度整合与模式识别，结合新闻模型的决策算法，精准评估热点事件的虚假概率。对高置信度的虚假事件实施强化监测与预警，确保重点监管的有效性。

6.3.2 问题发现

依托传播源头追溯技术与路径分析算法，深入剖析虚假新闻的生成机制与扩散模式，揭示其传播背后的驱动因素。为预警模型的优化与防控策略的制定提供数据支持与理论依据，增强系统预警的精准度与前瞻性。

6.3.3 关联分析

采用实体识别与链接分析技术，对虚假新闻中涉及的人物、地点、时间等实体进行精细关联分析。通过构建事件图谱，对关联新闻进行智能筛选与深度检测，有效识别潜在未检测虚假新闻，强化对类似虚假信息的源头遏制与传播阻断。