

## 昆明理工大学毕业设计（论文）

# 开 题 报 告

# 题目：基于多模态深度学习的社交媒体虚假新闻检测系统

## 研究

学 院： 信息工程与自动化学院

专 业: 数据科学与大数据技术

学生姓名： 蒋学琛

指导教师： 李亚

日期: 2025 年 3 月

## 毕业设计（论文）的主要内容：

### 一、课题研究的背景及意义

社交媒体的发展提升了信息传播效率，但也加剧了虚假新闻的传播风险，影响社会舆论和稳定。传统单模态检测方法难以充分识别虚假新闻的复杂特征，推动了多模态深度学习方法的研究。

本研究旨在设计并实现基于先进的多模态深度学习技术的虚假新闻自动检测系统，通过精确的多模态信息融合与深层次特征提取，显著提高虚假新闻的识别准确性与泛化性能。本研究成果不仅在理论层面为多模态信息融合技术提供新的应用范例，也在实际应用层面能够为社交媒体平台、新闻机构以及监管部门提供高效、智能的虚假信息治理工具，推动信息治理体系的智能化转型。

### 二、课题研究内容

本课题的主要研究内容具体包括以下几个方面：

1. 数据集构建与数据预处理：从权威公开的社交媒体虚假新闻数据库中，系统性地提取具有代表性的多模态数据，包括新闻标题、正文文本及相关配图。对数据集进行高质量的清洗与标准化处理，包括缺失值填充、异常值检测与处理以及多模态数据的有效标注，为模型训练提供可靠的数据支持。

2. 多模态深度学习模型的设计与优化：研究与构建基于 Transformer 网络与卷积神经网络（CNN）的多模态融合深度学习模型，深入探索多模态数据特征融合的策略与方法。考虑实际情况中模型训练资源有限或难度较大的情况，可采用调用现有成熟的多模态大模型 API，或部署 Janus 等开源多模态模型进行系统实现。

3. 后端开发与接口设计：使用 Python 语言的 Django 框架构建系统的后端服务，提供包括用户注册、登录及新闻上传等一系列交互接口。设计高效、稳定的模型推理接口，支持用户上传文本及图像数据后，系统能够及时反馈虚假新闻识别结果。利用 MySQL 数据库进行高效的数据管理与存储，实现用户信息管理与检测结果的高效查询。

4. 前端界面设计与开发：基于 Vue.js 框架开发用户友好且交互性强的前端界面，实现新闻上传、用户认证与检测历史管理功能。提供历史检测结果的直观查询与展示功能，使用 Echarts 或类似的数据可视化工具，构建虚假新闻检测数据的分析界面，辅助用户深入理解历史检测趋势及模态特征。

本课题通过结合深度学习、多模态数据分析与现代 Web 技术，致力于开发高效且具有理论创新的虚假新闻检测系统，为领域内的后续研究与实际应用提供理论基础和技术借鉴。

**设计（论文）的技术路线及预期目标：**

**一、技术线路**

从公开的权威多模态数据集中系统性地采集包括新闻文本及相关配图的数据后进行数据清洗、预处理与标注，确保数据质量和一致性，以满足高精度模型训练的需求。

设计并实现基于 Transformer 和卷积神经网络（CNN）的多模态融合模型，深入探索文本与图像信息融合机制。当模型训练的复杂度较高时，采用现有成熟的多模态大模型（如 Janus、GPT）进行优化与性能提升。

基于 Django 框架搭建稳定高效的服务后端，实现用户管理、数据上传接口及模型调用接口。使用 MySQL 数据库实现数据的可靠存储及高效查询。使用 Vue.js 开发交互性强的前端界面，支持用户上传与历史检测结果的管理功能。利用 Echarts 等数据可视化技术，直观展示检测历史数据及分析结果。。

**二、预期目标**

完成本设计后系统将实现以下功能：

提供自动化且高准确性的多模态虚假新闻识别，系统能有效区分真实与虚假的新闻信息。

实现便捷易用的 Web 交互平台，用户可以方便地上传新闻文本与图像进行虚假新闻检测，系统能及时返回准确的检测结果。

用户可轻松管理和查询历史检测记录，并通过数据可视化工具直观分析虚假新闻检测的趋势与特征。

**课题进度计划**

**2025 年 1 月 1 日～2025 年 3 月 21 日：**确定研究课题目标，广泛调研国内外相关文献，明确研究内容与创新点，完成并提交开题报告。

**2025 年 3 月 22 日～2025 年 4 月 24 日：**系统性地完成数据集的采集、清洗、标注及标准化预处理，确保数据质量。设计初步的多模态深度学习模型架构，完成初步模型搭建与初始训练。完成并提交中期检查报告。

**2025 年 4 月 25 日～2025 年 5 月 4 日：**根据导师对中期报告的审阅反馈，深入调整和优化研究策略。完成中期检查。

**2025 年 5 月 6 日～2025 年 5 月 15 日：**深入开展模型优化与系统集成，准备论文查重所需的数据与环境。完成查重系统相关数据准备工作。

**2025 年 5 月 16 日~2025 年 5 月 29 日：**进行首次论文查重，根据结果修改和完善论文内容，提交最终修改稿。

**2025 年 5 月 30 日~2025 年 6 月 8 日：**完成论文的指导教师与评阅教师评审，安排毕业答辩相关事宜，确定答辩组成员。

**完成课题所需条件及落实措施：**

**数据资源：**获取来自公开渠道的权威多模态虚假新闻数据集，包含文本、图像等多种信息形式，以确保数据的多样性与真实性。

**软件条件：**搭建基于 Python 的研究环境，包括深度学习框架（如 PyTorch、TensorFlow）、Django 后端服务平台、Vue.js 前端框架以及 MySQL 数据库管理系统。

**硬件条件：**配备足够计算能力的 GPU 服务器或云端计算资源，以支持深度学习模型的训练和部署需求。

**学术指导与资源支持：**导师定期指导与深入交流，确保研究目标明确且进展有序，充分利用学术资源，包括图书馆电子资源、学术论文数据库，实时跟踪并掌握领域最新研究动态。

**参考文献、资料：**

[1] 朱枫, 张廷辉, 李鹏, 等. 基于多模态自适应融合的短视频虚假新闻检测[J]. 计算机科学, 2024, 51(11):39-46. DOI:10.11896/jsj.kx.240700062.

[2] 李基. 基于深度学习的多模态虚假新闻检测[D]. 四川:西南交通大学, 2022.

[3] 陈志涛, 刁建忠, 张宇辉. 基于知识图谱与 BERT 的多模态虚假新闻检测模型[J]. 信息与电脑, 2024, 36(4):220-222. DOI:10.3969/j.issn.1003-9767.2024.04.070.

[4] 李佳豪. 基于深度学习的多模态虚假新闻检测研究[D]. 江西:江西理工大学, 2024.

指导教师意见：

指导教师（签字）：

年      月      日

学院毕业设计（论文）工作领导小组意见：

组长（签字）：

年      月      日