

## 教育背景

- **卡内基梅隆大学** 匹兹堡, 宾夕法尼亚, 美国  
化学工程硕士学位  
核心课程:  
机器学习                      深度学习                      计算机系统                      过程系统建模  
2019 年 8 月 - 2020 年 12 月
- **天津大学** 天津  
应用化学 (工) 学士  
2015 年 9 月 - 2019 年 7 月

## 工作经历

- **中国融通集团信息技术有限公司** 北京, 中国  
数据产品专家  
2024 年 3 月 - 至今
  - **数据中台建设与核心能力构建:**
    1. 参与数据中台建设, 负责数据资产整合、数据治理流程建立及数据仓库架构设计, 提升公司整体数据管理水平。
    2. 扩展并优化数据中台系统, 完成包括电商、司库等业务系统的数据接入, 并进行服务器及系统升级, 提升数据处理效率和系统性能。
    3. 支撑租赁业务数字化转型, 设计建设租赁业务决策支持系统, 完成关键数据初始化和质量提升工作, 为系统上线及业务数据纳管奠定基础。
    4. 赋能电商业务数据应用, 通过数据分析挖掘电商平台数据价值, 构建客户画像及偏好模型, 识别业务风险点, 并实现电商周报自动化, 为业务决策提供有力的数据支持。
- **京东零售数据中台** 北京, 中国  
内容智能研发组算法工程师  
2021 年 5 月 - 2024 年 3 月
  - **项目 1: 大型语言模型赋能电商业务场景:**
    1. 自主训练大型语言模型 (LLM), 提升模型在电商领域的专业性, 使其更精准理解电商领域的文本数据。
    2. 将训练后的 LLM 应用于智能客服和文案生成等电商业务场景: 在智能客服方面, 利用 LLM 自动回答客户商品咨询, 提高客服效率; 在文案生成方面, 使用 LLM 生成类似小红书风格的商品推荐文案, 提升营销效果。
    3. 利用 ChatGPT 辅助评估自训 LLM 的性能, 确保模型质量和应用效果。
  - **项目 2: 商品图、文特征挖掘与匹配:**
    1. 利用图、文多模态特征提取模型, 融合商品文本和图像信息, 自动生成商品的品类、材质、款式等特征标签, 实现商品多维度特征的深度挖掘。
    2. 开发商品特征标签分类模型, 有效过滤不准确或不合理的标签, 提升商品特征标签的质量。
    3. 应用商品特征匹配算法, 实现商品与特征关键词的精准匹配, 为商品推荐和搜索系统提供高质量数据支撑。
  - **项目 3: 基于深度学习的客户需求识别系统:**
    1. 主导开发客户情感分析系统自动分析商品评论, 识别并标记出客户表达的负面情绪, 例如不满、抱怨等。
    2. 通过改进算法和优化数据处理流程, 将负面情绪识别的准确率从 92% 提升至 98%, 大幅提高了系统识别客户真实情感的能力。
    3. 对情感分析系统进行性能优化, 使其处理评论速度提升了 3 倍, 显著提高了分析效率, 能够更快地处理海量客户评论数据。
    4. 应用用户观点分析技术, 对客户评论中的观点进行自动归类 and 匹配, 帮助企业更系统地了解客户对商品和服务的具体需求和偏好。

## 研究经验

- **LEGO-ABSA (详情见出版部分) 论文链接** 北京, 中国  
京东算法工程师  
2022 年 2 月 - 2022 年 7 月
  - 开发细粒度情感分析 (Aspect-Based Sentiment Analysis, ABSA) 的新框架, 旨在提升电商场景下客户评论的深层语义理解, 目标是更精确地分析客户评论中针对商品特定属性 (如质量、设计、价格) 的情感倾向。
  - 提出了一个统一的生成式框架, 通过设计可灵活调整的“任务提示 (Prompt)”来控制分析任务类型, 实现用同一模型高效解决多种细粒度情感分析子任务。例如, 一个任务提示可能被设计用于提取评论中提及的商品属性, 而另一个任务提示则用于判别用户对这些属性的情感极性。
  - 采用创新性的“乐高积木”式的模型构建方法, 先在简单任务上训练基础模型处理特定类型任务, 再通过组合不同的任务提示来处理更复杂的分析场景。
- **商业分析中的机器学习** CMU, 匹兹堡  
研究助理, Mi Zhou 教授  
2020 年 7 月 - 2020 年 12 月
  - 研究外貌吸引力与网络主播收入的关系, 使用计算机视觉技术分析主播面部图像, 量化面部特征 (包括使用 Face++ API 获取的“美貌分数”、视觉年龄、性别等), 并进行关联性分析。

- 开发基于深度学习的在线课程满意度预测模型，利用自然语言处理 (NLP) 技术分析线上课程字幕，捕捉课程内容的深层语义信息，并预测学生对课程的满意度。提出的模型在预测准确性上优于传统机器学习方法，RMSE 误差从 0.27 提升至 0.14。

- **气象能见度预测**

CMU, 匹兹堡

由 Hamish Gordon 教授指导的硕士研究项目

2020 年 1 月–2020 年 12 月

- 从多个气象数据网站收集了超过 30 万条历史天气记录，包含能见度、湿度、温度等多种气象参数，并使用数据处理工具对数据进行清洗和预处理。
- 应用机器学习模型预测天气能见度。结果显示，该模型预测雾天的准确率 (约 95%) 显著高于传统物理模型。

## 论文发表 & 专利

---

- \*Tianhao Gao, Jun Fang, \***Hanyu Liu** *et al.*, LEGO-ABSA: A Prompt-based Task Assemblable Unified Generative Framework for Multi-task Aspect-based Sentiment Analysis (\**Co-first author*, COLING 2022) ([论文链接](#))
- **刘涵宇**; 刘超; 包勇军, 样本处理方法、装置、模型更新方法、介质及电子设备, 专利号: CN118113862A
- **刘涵宇**; 刘朋樟; 刘超; 包勇军, 标签融合方法、装置、电子设备以及存储介质, 专利号: CN117807287A
- **刘涵宇**; 包勇军; 刘朋樟; 刘超, 文本处理方法、装置、电子设备及存储介质, 专利号: CN116610942A
- 方俊; 汪力; **刘涵宇**; 高天昊; 刘超; 刘朋樟; 包勇军, 模型训练方法、信息生成方法、装置、设备和介质, 专利号: CN119004091A
- 刘志远; 赵建博; **刘涵宇**, 一种图像语义理解方法、装置、设备及存储介质, 专利号: CN114429566A

## 技能

---

- **编程语言 & 工具**: Python, Bash, C++, SQL, JAVA, HTML/CSS, Docker, GIT, MySQL, L<sup>A</sup>T<sub>E</sub>X
- **深度学习框架 & 网络架构**: PyTorch, HuggingFace, CNN, RNN, LSTM, Transformers, ViT, GAN, Diffusion Model