

# 徐嘉瑞

手机：(+86) 18115168409 · 邮箱：xujiarui98@foxmail.com

性别：男 · 籍贯：江苏南京 · 目标岗位：算法工程师

知乎 · in 领英 · XavierSpacy · CSDN



## 自我总结

- 就读于悉尼大学数据科学硕士（研究导向），机器学习专项，熟悉常见的机器学习、深度学习算法；
- 熟悉 Python、PyTorch 框架和 HuggingFace 生态，具备 Linux 开发的经验，参与过 Docker 企业化部署；
- 具备较强的编程和动手能力，主动学习和探索开源代码中的技术细节，并积极参与开源项目贡献等；
- 具备较强的学习能力和信息检索能力，能够快速地上手陌生领域，并掌握和应用该领域的基本知识；
- 能够阅读前沿文献、跟进研究进展，具备良好的沟通能力和团队协作能力，善于总结和独立思考。

## 专业技能

- 编程语言/数据库: Python, MySQL
- 科学计算: NumPy, Pandas, SciPy
- 机器学习框架: SciKit Learn, PyTorch, HuggingFace
- 检索增强: LangChain, Sentence-Transformers

## 教育背景

- 悉尼大学** 2023.02 - 2024.11
- 数据科学硕士 计算机学院 **WAM: 82.1 (优)** 专项: 机器学习
  - 获得 2024 年工程学部冬季假期研究实习奖学金，参与跨模态医疗图像表征学习的研究工作
- 南京工业大学** 2017.09 - 2021.06
- 工业工程学士 经济与管理学院 **GPA: 3.78 / 4.0** 专项: 物流与供应链管理

## 实习经历

- FunPlus** 2023.12-2024.02
- 技术中台 - AI 技术 - AIGC 算法实习生
- 产品演示:** 追踪业务相关的前沿模型并快速搭建产品演示，探索算法落地过程中的可行性与局限性。
  - 技术研究:** 参与 RAG 技术调研，跟踪和对比最新的嵌入模型、大型语言模型、检索增强框架等技术。
  - 模型部署:** 部署 ChatGLM、MiniCPM、Qwen 等多轮对话模型，利用 vLLM 加速推理，并外挂知识库。
  - 模块研发:** 参与检索增强中不同组件的研发，包括查询重写、文档重排等，并将其整合为通用框架。

## 研究经历

- Transformer 模型视角下安卓生态系统演变的研究** | 框架: PyTorch 2023.08 - 2024.06
- 对比了 7 种常见主题模型及 2 个嵌入变种在百万级应用描述上的连贯性、多样性、相似性等性能；
  - 引入了类别条件总结模块，分配主题词至每个应用类别，并在此基础上提出了 TopMEDAL 分析框架；
  - 从共同词定性与定量、词云可视化、大语言模型推理等多角度分析，得出应用市场的演变趋势。

## 项目经历

- 文本-图像多模态多标签分类问题** | 框架: PyTorch 2024.04 - 2024.06
- 选择 CLIP、BLIP 作为骨干网络，探究对比学习预训练模型在图文多模态下游分类任务上的表现；
  - 利用 CLIP 模型的图文特征在相同隐空间的特性，采用路由机制为不同模态动态分配权重并求和；
  - 通过 BLIP 模型文本编码器的交叉注意力机制，将可学习标记或完整的嵌入作为后续模块的输入；
  - 实现了开源代码库中与 BLIP 原论文相比的缺失功能，并提交更改，最终顺利合并至 transformers 库。

## 其他

- 机器学习:** 从零实现感知机、非负矩阵分解实现、手写字符识别、噪声标签图像分类、TabTransformer；
- 语言:** 普通话(母语)、全国大学英语 CET-4 及 CET-6 (541)、托福 97/120 (写作 27/30, 阅读 28/30)；
- 知乎图文创作:** 单篇最高阅读近 70 万、最高点赞量近 1 万、最高收藏量达 3 千，曾获知乎日报收录。