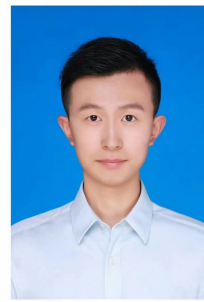


# 徐嘉瑞

手机：(+61) 0412696563 · 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 ·  XavierSpycy · CSDN



## 自我总结

- 就读于悉尼大学数据科学硕士（研究方向），机器学习专项，熟悉常见的机器学习、深度学习算法；
- 研究方向聚焦于大语言模型应用，有主题建模、模型部署、检索增强、加速推理等的经验；
- 活跃于开源社区，善于通过开源代码学习和探索各种工程技术细节，并积极参与对开源项目的贡献；
- 熟悉 Linux 开发环境，熟悉 PyTorch 框架和 HuggingFace 平台，参与过 Docker 和 Shell 企业化部署。

## 专业技能

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

## 教育背景

悉尼大学			2023.02 - 2024.06
• 数据科学硕士	计算机学院	WAM: 82.1	专项: 机器学习
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
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，外挂知识库，并利用 vLLM 加速推理。
- 模块研发: 参与检索增强中不同组件的研发，包括查询路由、文档重排等，并将其整合为通用框架。
- 团队协作: 协助团队开发多模态搜索工具，独立负责大语言模型文档检索增强部分的开发与部署。

## 项目经历

基于 NumPy 的机器学习框架开发   框架: NumPy	2023.08 - 2023.11
--------------------------------	-------------------

- 多层感知机
  - 实现了多种激活函数、隐藏层、优化器、学习率调度器等组件，融合早停法、小批量训练等技巧。
  - 整合了可用于回归或分类任务的损失函数，在经典数据集上可以高效地达到令人满意的性能。
- 非负矩阵分解
  - 实现了多种不同的有效且高效的非负矩阵分解算法，并完善了便于研究者创新的简洁开发框架。
  - 设计了可与用户交互的算法体验界面，并通过在人脸修复任务上的性能，展示低秩矩阵的鲁棒性。
  - 引入了多进程技术，可并行处理多组实验，相较于顺序处理总时间缩短至原有的 30%~40%。

用于文本-图像分类的多模态融合   框架: PyTorch	2023.04 - 2023.07
-------------------------------	-------------------

- 特征提取: 部署用于文本特征的蒸馏版本的 BERT 模型和用于图像特征的 DenseNet 等卷积神经网络。
- 特征融合: 实现了自注意力和交叉注意力机制，有效地加权融合了提取自多模态源的特征。
- 项目成果: 通过个人独立完成以 87.5% 的 F1 得分取得了在 Kaggle 竞赛中前 22% (31/142) 的排名。

## 其他

- 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)
- 自媒体运营: 知乎个人账号运营图文内容创作，达 10 万阅读量的回答共计 6 篇，单篇回答阅读量最高达 70 万，单篇点赞量最高近 1 万，单篇收藏量最高达 3 千。