王泽宇

19722751024 | admin@9998k.cn | AI 算法岗
https://github.com/moyitech

教育经历

太原理工大学 计算机科学与技术学院 (大数据学院) 计算机科学与技术专业【本科】 2022.09-2026.06

GPA: 4.02 / 5.0 **参与组织**：太原理工大学 云顶书院 人工智能方向负责人
课程成绩：机器学习 91，操作系统 97，高等数学 94，数据库管理及应用 90，线性代数 98，大学物理 95，概率论与数理统计 94，电子设计 98，大学英语 96，数据结构与算法 86，计算机数值方法 87，JAVA 语言程序设计 86。
荣誉奖励：2022-2023 校优秀本科生一等奖学金，2023-2024 校优秀本科生二等奖学金。
荣誉称号：学业成绩优秀称号、对外交流优秀称号、科技竞赛、学术研究优秀称号。

专业技能

深度学习：掌握经典深度学习算法，熟练使用 PyTorch 复现经典网络并训练模型
LLM：掌握主流的量化算法的应用（awq 等），模型本地部署 (vllm、llama.cpp)、LoRA 微调，RAG 开发经验，AI-Agent 智能体开发经验，主导 Agent 垂类体 AI 项目。
开源经历：self-llm(13k+ star) 核心贡献者，MetaGPT PR，Mutual-AI 核心贡献者
其他：Docker，实验室算力服务器运维，使用 CentOS 运维个人博客及 Python 项目。

项目经历

论文：Multi-Stage Multimodal Distillation for Audio-Visual Speaker Tracking 2024.05 −2024.12

发表会议：2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP，**CCF-B**)
DOI：10.1109/ICASSP49660.2025.10888838
承担工作：学生二作。使用 SiamFC、stGCF 搭建音视频多模态网络，三阶段深缩框架，实现像素级 (3 pixel) 的说话人追踪精度。
音视频说话人追踪在人机交互、智能监控、具身智能领域具有重要作用，本课题使用两个单模态的教师网络在三个不同阶段对音视频模态的学组网络进行蒸馏，多模态的学生网络使用对称交叉注意力进行融合。使用 AV16.3 数据集进行实验，seq1、2、3 训练，seq8、11、12 测试，平均绝对误差 3.02。

开源项目：self-llm 2024.08 −2025.03

影响力：13.7k star **担任角色**：核心贡献者
项目链接：https://github.com/datawhalechina/self-llm
项目围绕开源大模型在 Linux 平台的环境配置、部署应用与算法优化，面向国内初学者提供了一套系统化、全流程的实践教程。针对 LLAMA、ChatGLM、InternLM 等主流开源 LLM，详细阐述了环境镜像管理、模型部署流程及集成 LangChain 框架的工程化应用实践；同时实现了包括分布式全量微调、LoRA、P-tuning 等高效微调方法，加强了开源大模型的实际落地与推广应用。项目成功入选 2024 年 Google 开发者大会优秀开源算法项目案例。

学校项目：太原理工大学 AI 辅导员 2025.02 −2025.03

担任角色：发起人、主要开发者
介绍链接：https://mp.weixin.qq.com/s/It5YYVXDfUXFA7APnv7kkg
项目内容：
知识库收集：1. 利用爬虫技术，对学校官网的通知通告等信息进行采集；2. 使用 MySQL+Milvus 数据库，实现管理员对 PDF、CSV 知识库的增删改查。
检索增强生成：1. 将 BM25 搜索和 Text Embedding 搜索分别封装成 Tools；2. 使用 Doubao-1.5-lite 模型对聊天内容进行提取，并对两个检索起分别进行 Function Call；3. 使用 DeepSeek-v3 对整合上下文进行生成。4. 在进行 function call 搜索的同时，对用户信息进行有害性辨别，分类为：校园信息问答、提示词注入、无关信息等，对于提示词注入和无关信息的情况进行中断并直接返回预设文本。

竞赛项目：AI 面试官 2023.12

担任角色：队长、主要开发者
竞赛介绍：阿里云天池 Agent Builder 挑战赛 应用技术奖
项目内容：该项目发起于 AI Agent 刚兴起的 23 年末，使用 Multi-Agent 技术实现了 AI 面试工具流：简历评估、RAG 抽取题目、问答评估、最终打分。题库结合爬虫从网络中获取 AI 相关的面试题，并使用 GPT 生成部分题目，结合 faiss 构建题库搜索 tool。

竞赛项目：Home Agent 2024.9

担任角色：队长、主要开发者
竞赛介绍：全国大学生智能照明与智能穿戴创新创业大赛 全国特等奖
项目内容：二次开发 Blockly 图形化编程工具使其支持小米智能家居 API 的接入。使用 LoRA 微调的 Qwen2 将用户输入自然语言转换为 JSON 表示的积木块，微调数据集使用 GPT-4o 进行 few-shot 构建。

开源项目：Mutual-AI 2023.04 –2023.08

担任角色：核心贡献者 (算法、运维)
项目链接：https://github.com/YinHan-Zhang/Mutual-AI **线上地址**：http://ai.9998k.cn/
背景：于太原理工大学云顶书院人工智能方向学习期间，出于学习目的，训练 CV、NLP 等方向的的“Hello World”项目，并接入前后端分布式部署在 5 台服务器中供线上交互。**贡献**：使用 CNN 训练手写数字识别模型；使用微博评论数据集微调 Bert 实现情感二分类任务；模型与 FastAPI 对接；多服务器部署运维。

实习经历

SentimenTrader(美国上市) Agent 算法工程师 2024.10 –2025.04

检索引擎优化：项目使用 RAG 对站内文章、YouTube 视频链接能非开放领域的内容进行检索增强生成。检索前进行 Function Call 从用户上下文中分别提取语义信息和关键词信息；使用 bge-large 和 bm25 对数据进行双路检索；检索后使用 bge-rerank 对召回的数据进行重排序；用 LLM 结合 Query 对召回内容进行摘要总结。
Agent 开发：结合文章、YouTube 视频知识库、回测工具操作实现 AutoAgent。其中回测工具首先对用户意图进行分析并分析信息完整性，提示模型结合召回信息进行动态迭代的 Web 操作，完成用户目标的股票回测。

华创智慧（北京）科技有限公司 大模型算法工程师 2024.06 –2024.09

大模型的本地部署：使用 vllm 部署、awq 量化，显存降低 2/3，output token 速度提升一倍
Agent 开发：基于大语言模型的心理教育大模型，通过 CBT 方法对用户行为进行分析，为用户进行心理疏导，帮助用户改善情绪调节，以及针对解决当前问题的个人应对策略的发展。

上海鲸科信息科技有限公司 AIGC 产品研发 2023.12 –2024.06

在实习期间，主要开发了两款基于 GPT-4 的 AIGC 应用：**1. 定制化翻译**：使用 RAG（Retrieval-Augmented Generation）架构思想，Milvus 向量数据库及特定领域国际化词汇库，实现对复杂上下文对话的精准翻译输出，结果以结构化的 JSON 形式呈现；**2. 客服辅助 QA**：通过切分整合文档资源，运用 Agentic-RAG 实时检索匹配相关文本片段，自动生成针对性回答并附带源文链接，有效赋能客服团队快速、准确解答用户疑问。

科技竞赛

全国大学生数学建模竞赛 全国二等奖 2023.09

中国移动梧桐杯大数据竞赛 全国三等奖 2023.11

阿里云天池 AgentBuilder 挑战赛 英伟达技术奖 2023.12

中国大学生智能照明和智能穿戴创新创业大赛 全国特等奖 2024.09

国际青年人工智能大赛 国际一等奖 2024.11

全球校园人工智能算法精英大赛 全国二等奖 2024.11

全国大学生职业规划大赛 太原理工大学 Top1 2024.12