

生成式人工智能智能体平台在大学生就业指导中的应用

龚翔, 袁小刚, 高旭昊, 常有康, 马亚燕
甘肃政法大学网络空间安全学院, 甘肃 兰州 730070
DOI: 10.61369/TACS.2025100041

摘 要 : 本文探讨了生成式人工智能在大学生就业指导中的应用, 聚焦于构建基于开源大模型的本地化智能体平台。研究通过模型量化压缩与本地部署, 实现了低成本、轻量化的技术方案, 并开发了涵盖职业画像、岗位推荐、智能问答与政策解读的核心功能模块, 创新性地引入 “AI+ 人工” 协同服务机制以提升服务质量。平台在甘肃某高校的实证实验显示, 其在响应速度、数据安全等方面表现基本达标, 显著提升了岗位匹配与申请转化效率, 但在高峰并发性能、跨专业建议精准度及对基础薄弱学生的帮扶效果等方面仍存在不足。研究表明, 生成式 AI 在标准化就业指导服务中具有效率优势, 但区域化落地需在技术轻量化、数据生态构建与服务模式创新之间寻求平衡。未来研究应关注轻量化架构、动态区域数据库建设与 “政校企” 协同运营, 以推动智能就业指导体系在区域层面的规模化应用与持续优化。

关 键 词 : 智能体; 本地化部署; 就业指导

The Application of Generative Artificial Intelligence Agent Platform in College Students' Career Guidance

Gong Xiang, Yuan Xiaogang, Gao Xuhao, Chang Youkang, Ma Yayan
School of Cyberspace Security, Gansu Political Science and Law University, Lanzhou, Gansu 730070

Abstract : This paper investigates the application of generative artificial intelligence in university career guidance, focusing on the development of a localized intelligent agent platform based on an open-source large language model. The study achieves a cost-effective, lightweight technical solution through model quantization, compression, and local deployment, while developing core functional modules covering career profiling, position matching, intelligent Q&A, and policy interpretation. An innovative "AI + human" collaborative service mechanism is introduced to enhance service quality. Empirical experiments conducted at a university in Gansu Province demonstrate that the platform meets baseline requirements in areas such as response speed and data security, and significantly improves job matching accuracy and application conversion rates. However, limitations remain in peak concurrency performance, accuracy of cross-disciplinary recommendations, and support effectiveness for students with weaker foundational skills. The research indicates that generative AI holds efficiency advantages in standardized career guidance services, but its regional implementation requires a balance among technological lightweight design, data ecosystem development, and service model innovation. Future studies should focus on lightweight architectures, dynamic regional database construction, and "government-industry-university" collaborative operations to promote the scalable application and continuous optimization of intelligent career guidance systems at the regional level.

Keywords : intelligent agent; localized deployment; career guidance

引言

(一) 研究背景及意义

我国高校毕业生规模持续扩大, 2025年预计突破1200万人, 就业市场面临 “供大于求” 与 “人岗错配” 的结构性矛盾。传统就业指导模式存在覆盖有限、信息滞后、精准度不足等问题。在技术层面, 生成式人工智能在自然语言处理、个性化内容生成等方面取得突破, 为构建精准、高效的就业指导平台提供了可能^[1]。因此, 探索基于生成式 AI 的就业指导平台, 成为应对当前就业服务痛点的重要途径。

本研究旨在通过构建动态职业画像、个性化岗位推荐、智能简历优化等功能, 帮助毕业生明晰自身优势与短板, 实现从 “被动求职” 向 “主动匹配” 转变。研究不仅有助于优化毕业生就业体验与成功率, 也为高校就业服务体系升级提供实践与理论支持。

项目信息: 本项目得到甘肃省教育科学 “十四五” 规划2025年度 “大学生职业规划与就业指导” 专项课题《AI 大模型驱动的可信大学生职业规划与就业指导智能体平台构建研究》(项目号: GS[2025]GHBZX0059) 的支持。

（二）研究目标与内容

本研究基于开源大模型 DeepSeek V1.0，围绕低成本本地化部署、就业场景精准训练与区域化服务落地展开。目标包括实现模型在高校本地环境中的轻量化稳定运行，完成就业场景适配训练，并通过在甘肃高校的应用验证，形成可推广的本地化就业服务解决方案。

研究内容围绕研究目标展开，主要包括三个层面：首先，基于现有服务器通过模型量化与压缩技术完成 DeepSeek V1.0 的本地化部署，构建隐私保护与成本监控体系，并建立多维度训练语料库，采用增量预训练与指令微调策略提升模型在就业场景中的适配能力；其次，开发动态职业画像、区域化岗位推荐、求职辅助及政策解读等核心功能，建立“AI+ 人工”协同服务机制，实现复杂需求的智能转接；最后，在甘肃某高校开展对照实验以评估平台效能，根据反馈持续优化模型与功能，完成本地化适配改进，并为后续推广提供实证依据。本研究聚焦于模型部署优化与场景化应用，不涉及底层算法创新，旨在为区域高校提供一套可落地实施的智能就业指导解决方案。

一、相关理论与研究现状

（一）DeepSeek V1.0 本地化部署智能体

指将 DeepSeek V1.0 大模型从云端迁移至高校本地算力环境，通过技术优化构建的、专用于大学生就业指导的智能化服务载体^[2]。本研究的核心技术载体采用“量化压缩 + 分布式算力调度”方案，与网络空间安全学院个性化学习试点采用的 ChatGLM2-6B 模型相比，DeepSeek V1.0 在文本生成的逻辑性与结构化输出上更具优势，且原生支持多轮对话记忆功能，无需额外开发会话管理模块，更符合职业规划咨询的交互场景。

本研究在模型量化策略上借鉴了前者的 INT8 量化经验，同时针对就业场景文本较短、关键词密集的特点，通过 llama.cpp 框架适配 CPU-GPU 混合推理优化了量化精度损失补偿方案，将模型显存占用从原生 16GB 降至 4GB 以下。

（二）国内研究现状

国内研究呈现“学术先行、应用跟进”的特点。学术方面，已有研究量化了大模型对就业市场的影响，揭示了产业集聚效应，并探索了轻量化模型在职业教育领域的应用^[3]。实践层面，部分高校引入 AI 职业评测工具，生成了基于需求模型的动态职业画像；甘肃省等地也试点推行了“AI 推荐 + 人工核验”的就业服务模式。但普遍存在数据更新滞后、算法透明度不足、跨行业建议精准性待提升等问题。整体来看，国内研究侧重于工具开发，对数据隐私、算法伦理及区域适配性等问题的系统性探讨尚显不足^[4]。

（三）国外研究现状

国际研究体现出“技术深化与伦理先行”的双轨特征。技术应用上，欧美企业已利用 AI 实现动态岗位匹配和构建职业培训平台，但面临算法黑箱、文化技能差异等适配性挑战^[5]。政策与理论层面，欧盟《人工智能法案》等举措致力于为可信 AI 建立法律框架。国外的研究与实践多基于发达经济体的中心化数据平台，在算法透明度与伦理规范方面具有前瞻性，但其高成本模式与特定市场环境，导致其方案对中国西部等新兴市场的适配性与推广性存在局限^[6]。

二、智能体平台设计与部署

（一）平台设计理念

以“可信赋能、精准适配、协同高效”为核心，立足于大学

生职业规划全流程与区域产业发展实际，构建技术合规与服务实用相统一的智能就业指导生态。

严格遵循隐私保护法规，对原始数据进行脱敏处理，采用泛化与端到端加密技术，全面防范数据泄露风险。通过推荐依据可视化，清晰展示岗位匹配的核心维度，增强决策可解释性，破解算法黑箱问题。采用接口冗余设计与操作日志审计机制，确保服务连续性与行为可追溯，筑牢平台可信基础。

（二）技术路线与架构

1. 技术路线

平台设计与实施遵循从需求分析到应用优化的闭环流程。首先，基于毕业生数据、求职痛点、企业需求及区域政策，完成了核心功能界定与需求文档制定。随后进入技术验证阶段，依托 RTX 4090 服务器进行 DeepSeek V1.0 模型接口的适配测试，并通过性能压力测试验证了“大模型 + 规则引擎”混合架构的可行性，同时评估了数据脱敏与匹配算法的安全性及效果^[7]。在平台开发阶段，依照分层架构完成了应用层、算法层与数据层的功能实现，并集成了数据加密、权限控制等安全机制，通过容器化技术完成了模块集成与系统联调^[8]。最后进入应用优化阶段，通过功能与性能测试保障平台可用性，结合用户反馈进行迭代改进，并最终完成了小范围试点验证。

2. 技术架构

本研究构建的生成式 AI 就业指导智能体平台采用多层架构设计，自上而下分为应用层、算法层、数据层与安全层，各层职责清晰、协同联动，确保系统稳定性、扩展性与安全性的统一。

在应用层，平台通过 Web 端与小程序双端部署，为用户提供职业画像生成、智能问答、岗位匹配、规划咨询及数据可视化等核心功能，满足从自我认知到决策参考的全流程就业指导需求^[9]。算法层作为智能核心，采用“大模型 API+ 规则引擎”的混合架构，依托 DeepSeek V1.0 实现自然语言处理与内容生成，同时结合 Drools 规则引擎对输出进行合规性与精准性校验，并运用协同过滤与知识图谱相结合的混合推荐算法提升匹配效果，配套建立多维评估体系以持续优化服务。数据层负责为平台运行提供支撑，对脱敏后的毕业生数据进行 ETL 清洗与集成，采用 MySQL 与 MongoDB 混合存储架构管理结构与非结构化数据，并通过定期同步机制保障数据时效性。安全层贯穿整体设计，综合运用信息脱敏、SSL 传输加密、基于角色的访问控制（RBAC）以及操作日志审计等多种技术与措施，全面保障数据安全与系统可信^[10]。

（三）平台部署过程

平台部署过程遵循“环境搭建－模型对接－数据预处理－系统集成”的流程。硬件基于 RTX 4090 服务器进行资源优化配置，软件环境选用 Ubuntu 22.04 系统，并部署 Python 3.9、PyTorch 2.0、Nginx、MySQL、MongoDB 及 Redis 等必要组件。模型对接方面，完成了 DeepSeek V1.0 的本地部署，通过 FastAPI 封装推理接口，并引入 Redis 缓存与超时重试机制以提升服务稳定性与成本效率；同时采用检索增强生成（RAG）技术优化输出准确性，并实施“增量更新＋轻量化微调”策略保障信息时效性。数据预处理阶段对原始毕业生数据集进行了脱敏、清洗与标准化处理，随后完成数据库初始化与分类导入，并构建了职业画像标签体系。最终，通过 Docker 容器化技术将各模块集成部署，利用 Docker Compose 进行编排管理，并配置 Nginx 反向代理与 HTTPS 确保访问安全，通过 Apifox 实现接口监控与系统运维，保障平台整体稳定运行。

三、平台实证实验与结果分析

（一）实验目的

本实验旨在多维度验证平台的性能与应用价值。具体目的包括：评估平台对学生能力特征与职业倾向的刻画准确性；衡量其“人岗匹配”效果；考察学生用户对平台建议的实际采纳率和满意度；测试平台的并发、响应与资源占用等性能指标，验证其稳定性。

（二）实验对象

实验对象为甘肃政法大学网络空间安全学院与人工智能学院 2026 届的 800 名毕业生。样本覆盖网络安全、人工智能等核心专业，包含本科与硕士两个学历层次，学历结构与实际生源保持一致。同时，样本兼顾了拥有职业认证、实习经历的学生与暂无相关经验的学生，以确保能够全面反映不同能力背景与就业准备状态毕业生的真实情况，从而保障实验结果的代表性与可靠性。

（三）实验结果统计

以甘肃政法大学网络空间安全学院、人工智能学院 2026 届 800 名毕业生为样本，分为使用 DeepSeek V1.0 本地化智能体平台的 400 人实验组与采用传统就业指导模式的 400 人对照组，以下为针对性统计的实验结果：

1. 平台基础功能效能统计

平台在核心功能的响应速度上表现基本达标，职业画像与岗位推荐的平均响应时间为 2.8 秒，符合预设 ≤ 3 秒的目标；但在秋招高峰期面对超过 200 人的并发访问时，响应延迟上升至 5.7 秒，并有 4.1% 的学生出现了请求超时。数据处理效率方面，平台对多源数据融合的准确率为 81%，略低于 85% 的预期，其中人工智能专业学生的非标准化实践项目数据融合准确率仅为 74%。在安全合规性上，实验期间未发生数据泄露事件，合规率达到 100%。

五大核心模块平均可用率达到 92.4%，满足预设 90% 的目标。各模块可用性表现分化明显：数据可视化模块最高，达 98%；而职业画像模块因学生输入信息格式不统一，可用性最低，仅为 84%。在内容准确率方面，平台输出内容经就业指导教师核

验，平均准确率为 71%，未达到预设 80% 的要求。

2. 核心应用效果指标统计

在匹配准确率方面，实验组获得的有效岗位匹配占比为 50%，较对照组的 32% 提升 18 个百分点，但未达到预设的“提升 30%”的目标。跨专业转型建议的准确率从 34% 提升至 46%，印证了 RAG 技术在专业适配中的有效性，但受限于区域知识库覆盖，小众专业的适配效果仍有提升空间。在岗位申请转化率方面，实验组基于平台推荐获得面试邀请的比例为 42%，显著高于对照组的 16%，提升 26 个百分点。但持有职业资格证书的学生转化率仅达 31%，而无技能认证的学生仅为 15%，表明平台对基础薄弱学生的帮扶效果仍显不足。

四、总结与展望

本文研究与实践表明，生成式 AI 在标准化就业指导服务（如简历优化、政策解读）中效率与便捷性优势显著，但受成本、技术适配及区域数据生态等因素制约，平台在岗位匹配精准度、区域契合度及用户信任度等核心指标上未达预期，最终未能实现规模化应用。研究发现，要开发成功的区域化智能就业平台，需在技术、数据、服务与资源间取得动态平衡，并解决技术轻量化与服务精准化、通用方案与区域特色适配两大核心矛盾。

未来可从三个方面深化研究：技术上，探索采用远程 API 与联邦学习等轻量化架构，在控制成本的同时提升服务稳定性与算法透明度；数据上，推动与地方政府、产业园区合作，构建覆盖区域重点产业的动态需求数据库，实现数据实时更新与融合；模式上，探索“政校企”协同的可持续运营模式，整合多方资源与政策支持，推动平台从高校试点向区域就业服务体系规模化延伸。通过技术、数据与模式的三维优化，旨在为高校就业智能化转型提供更具实操性的解决方案，助力缓解区域就业结构性矛盾。

参考文献

- [1] 王冀扬, 杨卓然, 杨波, 等. 智能体在银行业的应用探索[J]. 金融纵横, 2025, (04): 69-74.
- [2] 谢鑫, 赵正, 崔永辉. 基于多专家智能体的智能化课程评价模型设计与应用[J]. 航海教育研究, 2025, 42(1): 54-60, 68.
- [3] 王文晟, 谭宁, 黄凯, 等. 基于大模型的具身智能系统综述[J]. 自动化学报, 2025, 51(1): 1-19.
- [4] 张帆, 谭跃刚. 生成式预训练模型机器人及其潜力与挑战[J]. China Mechanical Engineering, 2024, 35(7).
- [5] Pang X, Li Z, Chen J, et al. Ai2apps: A visual ide for building llm-based ai agent applications[J]. arXiv preprint arXiv:2404.04902, 2024.
- [6] Rao H, Zhao Y, Hou X, et al. Software Engineering for Large Language Models: Research Status, Challenges and the Road Ahead[J]. arXiv preprint arXiv:2506.23762, 2025.
- [7] ZHAI Jie, Li Yanhao, Meng Tianxin. Exploration and practice of personalized computer laboratory teaching based on decision trees and large models[J]. Experimental Technology and Management, 2023, 40(12): 8-15.
- [8] Zhang K, Yao W, Liu Z, et al. Diversity empowers intelligence: Integrating expertise of software engineering agents[J]. arXiv preprint arXiv:2408.07060, 2024.
- [9] 卢宇, 余京蕾, 陈鹏鹤. 基于大模型的教学智能体构建与应用研究[J]. 中国电化教育, 2024(7): 99-108.
- [10] Zhang B, Luo P. Or-llm-agent: Automating modeling and solving of operations research optimization problem with reasoning large language model[J]. arXiv preprint arXiv:2503.10009, 2025.