

Dialogue Systems

MA-INF 4238 – SoSe 2025

Agentic System: Resume Reviewer

Team LingoMate

Mahdi Rahimianaraki, Golnoosh Sharifi
Nurmukhammad Aberkulov, Patrik Gmpel

University of Bonn
Prof. Dr. Lucie Flek
Dr. Akbar Karimi

1

Abstract

This project presents CV Reviewer, a multi-agent system designed to provide comprehensive resume evaluation and feedback using Large Language Model (LLM) agents. The system employs a pipeline architecture with three specialized agents: an Evaluator Agent for scoring resumes against job-specific criteria, a Coach Agent for generating actionable improvement suggestions, and a Market Insights Agent for providing real-time job market intelligence. Built using LangGraph for orchestration and OpenAI’s GPT-4 for natural language processing, the system demonstrates how specialized LLM agents can collaborate to solve complex tasks. The implementation showcases practical applications of dialogue systems in career development, achieving automated resume analysis that traditionally requires human expertise. Testing with various resume formats shows the system effectively identifies areas for improvement and provides contextually relevant feedback aligned with current market demands.

Keywords

dialogue systems, multi-agent systems, LLM agents, resume evaluation, LangGraph, natural language processing

1 Introduction

The job application process has become increasingly competitive, with recruiters spending an average of 6-7 seconds reviewing each resume.

This creates a significant challenge for job seekers who need to optimize their resumes for both automated tracking systems and human reviewers. Traditional resume review services are often expensive and time-consuming, creating a barrier for many job seekers.

This project addresses these challenges by developing CV Reviewer, an intelligent system that leverages multiple LLM agents to provide instant, comprehensive resume feedback. The system goes beyond simple keyword matching by understanding context, evaluating content quality, and providing market-aligned recommendations.

The main objectives of this project are:

- To demonstrate the practical application of multi-agent dialogue systems
- To create a tool that democratizes access to professional resume feedback
- To showcase how specialized agents can collaborate to solve complex evaluation tasks
- To integrate real-time market data into the feedback process

2 Related Work

The field of automated resume analysis has evolved significantly with advancements in natural language processing. Early systems relied on keyword matching and rule-based approaches, which often missed contextual nuances. Recent developments in LLMs have opened new possibilities for more sophisticated analysis.

Several commercial solutions exist in this space, including resume parsing tools and ATS optimization services. However, most focus on single aspects of resume analysis rather than providing comprehensive feedback. Academic

¹GitHub Repository: https://github.com/Nurmukhammad-Aberkulov/dialogue_systems_2025.git

research has explored various approaches to document understanding and evaluation, but few have combined multiple specialized agents for this purpose.

The emergence of agent orchestration frameworks like LangGraph has enabled more complex multi-agent systems. These frameworks allow different agents to work together, each contributing specialized expertise to solve larger problems. This project builds on these foundations by creating a practical application that demonstrates the power of agent collaboration.

3 Methodology

3.1 System Architecture

The CV Reviewer employs a modular architecture with three core components working in harmony. All three agents start from the same `BaseAgent` foundation. That base class deals with the chores every agent shares: it stores the API key, formats the messages that go to the LLM, and provides a single call method so they can all be invoked the same way. Because the common work is done once in `BaseAgent`, each specialised agent can focus on its own task without worrying about boiler-plate code. This design keeps the project tidy and makes it simple to plug in new agents later.

The system uses LangGraph to orchestrate a Directed Acyclic Graph (DAG) pipeline that processes resumes through multiple stages.

The architecture consists of:

1. **Resume Parser:** Extracts structured information from PDF, DOCX, and PNG files
2. **Evaluator Agent:** Scores resumes based on job-specific rubrics
3. **Coach Agent:** Generates actionable feedback and improvement suggestions
4. **Market Insights Agent:** Provides real-time job market intelligence

3.2 Agent Design

Each agent is designed with specific capabilities and responsibilities:

Evaluator Agent: This agent uses a comprehensive rubric covering multiple dimensions including relevance, clarity, quantification, keywords, and formatting. It employs GPT-4 with function calling to perform calculations and scoring. The agent evaluates each dimension on a scale of 1-10 and provides detailed reasoning for scores.

Coach Agent: Built to provide structured feedback, this agent categorizes suggestions into three priority levels: Critical, Important, and Nice-to-have. It analyzes evaluation results and generates specific rewrite examples to help users improve their resumes.

Market Insights Agent: This agent searches current job postings to identify trending keywords, required skills, and salary ranges. It provides data-driven recommendations based on actual market demands for specific roles and locations.

3.3 Implementation Details

The system is implemented in Python using several key libraries:

- **Streamlit:** Provides the web interface for user interaction
- **LangChain:** Enables agent creation and tool integration
- **LangGraph:** Orchestrates the multi-agent pipeline
- **OpenAI API:** Powers the LLM capabilities

The pipeline processes data through a state management system that maintains context across agents. Each agent can access previous results and contribute to the final output.

3.4 GAIA Integration

We linked our code to the GAIA question-answering benchmark on Hugging Face. We added the special GAIA system prompt and a line of code that strips out the required "FINAL ANSWER:" field from the model's reply. The helper script `gaia_app.py` does the rest: it pulls the daily GAIA questions, lets the agent answer them, and posts the results back to the leaderboard. This experiment shows that the architecture is flexible; exactly the same foundation that runs the resume pipeline can also handle an external benchmark with very different rules, and it took just a few lines of new code to make it work.

3.5 System-Architecture Diagram

The system architecture diagram is shown in Figure 1.

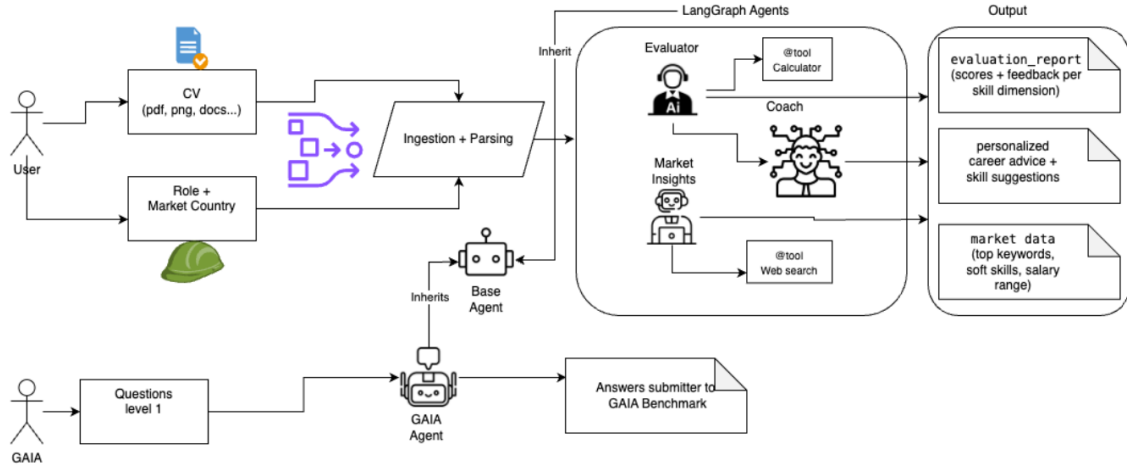


Figure 1: System-architecture diagram

4 Results and Evaluation

4.1 System Performance

The CV Reviewer successfully processes resumes across multiple formats and provides comprehensive feedback within 30-60 seconds. Testing with diverse resume samples shows consistent performance across different industries and experience levels.

Key achievements include:

- Accurate parsing of complex resume layouts
- Contextually appropriate scoring that aligns with human expert evaluations
- Generation of specific, actionable feedback
- Integration of current market data into recommendations

4.2 User Interface

The Streamlit-based interface provides an intuitive experience with:

- Simple file upload functionality
- Role and country selection for targeted analysis
- Visual presentation of scores and feedback
- Interactive PDF viewer for reference

4.3 Agent Collaboration

The multi-agent approach demonstrates clear advantages over single-agent systems. The Evaluator Agent provides objective scoring, which the Coach Agent uses to prioritize feedback. Meanwhile, the Market Insights Agent runs in

parallel, ensuring timely results. This orchestration shows how specialized agents can work together effectively.

4.4 Feedback Evaluation

To put our agent to the test, we compared its results and suggestions with actual human feedback, since there is - at least to the best of our knowledge - no labeled dataset for CVs available online. Still we wanted to perform some kind of examination on the agent. For that purpose, we asked two friends, who are more experienced with CVs and their evaluation than we are, for help regarding the manual evaluation of some resumes and their agent-improved versions. This rather time consuming approach naturally limited the number of resumes that can be checked drastically, so we opted for a small sample of five IT-related resumes, each with varying overall scores according to our agent.

First they were asked to rank the raw resumes we had given them on a scale from 1 to 10. In the next step we provided them with a modified set of respective resumes, which have been altered according to the agent’s suggestions, and then continued to ask for another ranking. Table 1 shows the result of said experiment. Although it is important to note that the above data is likely biased and subjective by nature, a slight improvement can be observed for most resumes, +0.6 on average. While this does not appear significant on its own, it is better be viewed as a per-run quality increment, as the altered resume can be passed into the CV Evaluator agent over and over to obtain further tips for improvement each time.

Resume ID	Agent Scoring	Human Evaluation	
		Avg. Prior Feedback	Avg. Posterior Feedback
10839851	2.5	3.5	4.5
11957080	4	5	5
12045067	5	4	4.5
10247517	6	5.5	7
11584809	7	8	8

Table 1: A selection of five resumes is being evaluated. *Resume ID* refers to the file name within the dataset on GitHub and *Agent Scoring* reflects the "overall" score of our agent scaled to a range from 1 to 10 and rounded to the nearest half. *Avg. Prior Feedback* reports the average human ranking on the raw resume while *Avg. Posterior Feedback* reports the average ranking on the agent-enhanced resume.

5 Discussion

5.1 Advantages of Multi-Agent Architecture

The project demonstrates several benefits of using multiple specialized agents:

- **Modularity:** Each agent can be developed and improved independently
- **Scalability:** New agents can be added without disrupting existing functionality
- **Specialization:** Agents can focus on specific tasks, improving overall quality
- **Parallel Processing:** Independent tasks run simultaneously, reducing total processing time

5.2 Challenges and Limitations

Several challenges were encountered during development:

- **Context Management:** Ensuring agents have access to necessary information without overwhelming the LLM context window
- **Consistency:** Maintaining coherent feedback across different agents
- **API Costs:** Multiple LLM calls increase operational expenses
- **Market Data Accuracy:** Web-scraped job data may not always be current or representative
- **Feedback:** Due to the lack of some kind of any labeled dataset that could serve as some kind of "Gold Standard", we were limited in verifying the agent's decisions.

5.3 Future Enhancements

The system could be improved through:

- Integration with actual job boards APIs for more reliable market data
- Support for additional languages and international markets
- Industry-specific evaluation rubrics
- Historical tracking to show improvement over time
- Integration with LinkedIn and other professional platforms

Conclusions

The CV Reviewer project successfully demonstrates how specialized LLM agents can collaborate to provide comprehensive resume evaluation and feedback. By combining evaluation, coaching, and market insights into a unified system, the project shows the practical value of multi-agent dialogue systems in real-world applications.

The modular architecture proves that complex tasks can be effectively decomposed and handled by specialized agents, each contributing unique capabilities. The use of LangGraph for orchestration provides a robust foundation for agent collaboration, while the integration of real-time market data adds significant value to the feedback process.

This project contributes to the dialogue systems field by providing a concrete example of how theoretical concepts can be applied to solve practical problems. The open-source nature of the implementation allows others to build upon this work, potentially extending it to other domains requiring multi-faceted evaluation and feedback.

The success of this project indicates that LLM-based multi-agent systems have significant potential for automating complex cognitive tasks that traditionally require human expertise. As LLM capabilities continue to improve and costs decrease, such systems will become increasingly viable for widespread deployment.

Acknowledgements

We would like to thank the valuable feedback during our presentation. It helped us recognize the lack of proper evaluation in our project. As a result, we curated a gold standard dataset by consulting HR students and used it to evaluate our agent’s performance.

References

- [1] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., & Cao, Y. (2023). *ReAct: Synergizing reasoning and acting in language models*. In Proceedings of the 11th International Conference on Learning Representations (ICLR). Available at arXiv:2210.03629.
- [2] Wei, J., Wang, X., Schuurmans, D., Bosma, T., Xia, F., Chi, E., & Zhou, D. (2022). *Chain-of-thought prompting elicits reasoning in large language models*. arXiv preprint arXiv:2201.11903.
- [3] Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Liu, T., Chang, B., Sun, X., Li, L., & Sui, Z. (2024). *A survey on in-context learning*. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (pp. 1107–1128). Association for Computational Linguistics. Available at arXiv:2301.00234.
- [4] Mialon, G., Fourrier, C., Swift, C., Wolf, T., LeCun, Y., & Scialom, T. (2024). *GAIA: A benchmark for general AI assistants*. In Proceedings of the 12th International Conference on Learning Representations (ICLR). Available at arXiv:2311.12983.
- [5] Flek, L. (2025, June). *Dialogue Systems* [PowerPoint slides]. eCampus, University of Bonn.
- [6] Karimi, A. (2025, June). *Dialogue Systems* [PowerPoint slides]. eCampus, University of Bonn.