LLM Project Report: AI Job Application Assistant

ApplyPro.ai

Group 8: Jialin You, Xinghua Peng, Hamza Shahab Shafqat, Vishanth Hari Raj Balasubramanian

Collab Notebook

1. Introduction

Project Overview:

The **ApplyPro.ai** project is designed to streamline and automate the job application process for job seekers. The tool leverages a **Large Language Model (LLM)** to match resumes with relevant job postings and generates personalized cover letters and cold emails, significantly reducing the manual effort involved in job searching. The project focuses on four main components:

- **Problem Identification**: Job seekers face inefficiencies in searching for relevant roles and tailoring applications for each position.
- **Solution Proposal**: We propose an LLM-powered system that uses semantic embeddings to match resumes to job postings and generate customized cover letters.
- **Technical Implementation**: The solution incorporates various technologies, including resume parsing, embeddings for job matching, and LLMs like GPT for generating application materials.
- **Live Demo**: A user interface allows candidates to upload resumes, view job matches, and generate tailored cover letters and emails, showcasing the full workflow.

Business Problem Identification:

Job seekers often encounter several challenges while applying for positions:

- **Manual and Time-Consuming Searches**: Candidates spend significant time searching across multiple job boards, often resulting in missed opportunities.
- **Generic Applications**: Tailoring cover letters and emails for each job application can be overwhelming and often results in generic, ineffective communication.
- **Mismatch Between Job Descriptions and Applicant Profiles**: Keyword-based search algorithms may fail to match applicants with roles that align with their skill sets.

ApplyPro.ai addresses these inefficiencies by developing an AI-powered system that automates job search, application customization, and communication generation.

2. Business Problem Justification

Importance of the Problem:

Job searching is one of the most crucial yet time-consuming activities for job seekers. According to **Jobvite**, **65% of job seekers** report stress during the search process due to time constraints and the complexity of navigating different platforms. Additionally, **75% of resumes** submitted to job boards never reach human eyes because of mismatched parsing and inefficient algorithms (Forbes). The

recruitment inefficiencies are not only a burden for applicants but also result in **millions of dollars in lost revenue for businesses** due to long hiring processes and unfilled positions (McKinsey).

Failing to address these challenges can cause businesses to suffer from longer hiring timelines, increased recruitment costs, and poor candidate experiences, while job seekers continue to face inefficiencies that result in wasted time and fewer opportunities for employment.

Existing Solutions:

Current solutions include:

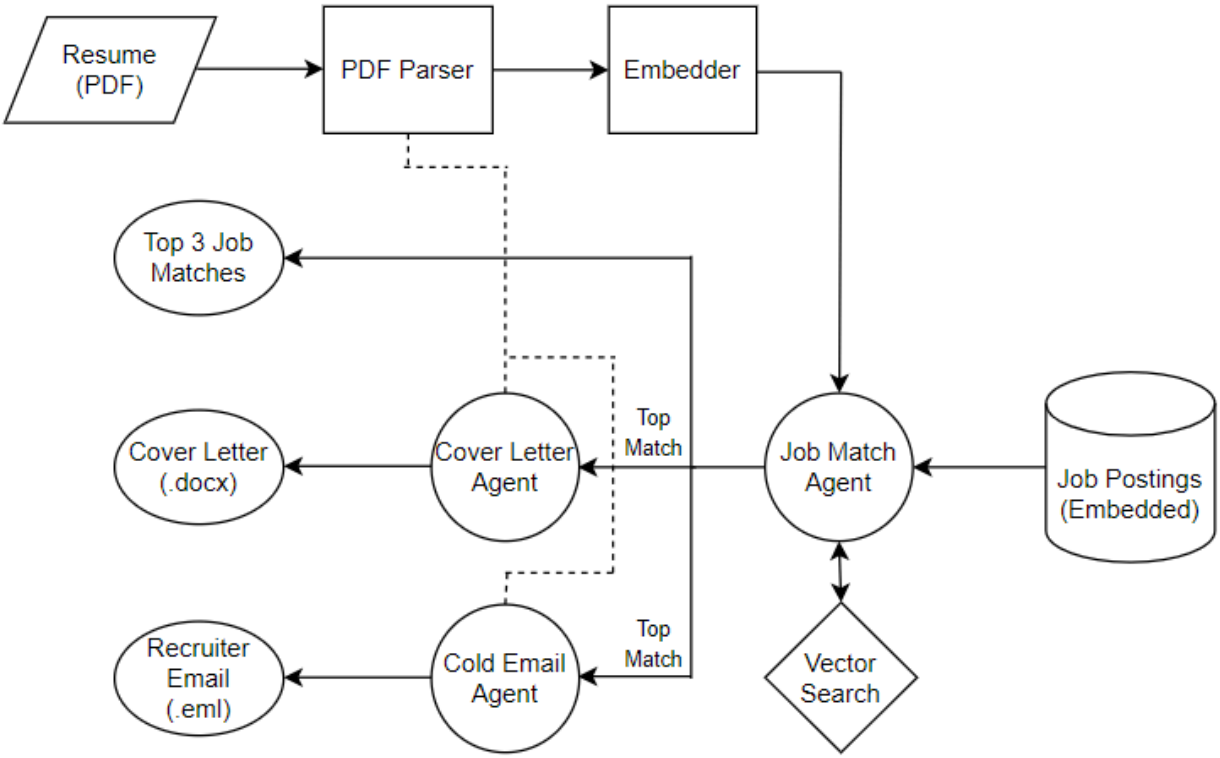
1. **Job Aggregators**: Platforms like LinkedIn, Indeed, and Workday provide job matching, but many rely on keyword-based search methods, which are prone to inaccuracy. These systems struggle to deeply understand candidate resumes and job descriptions, often leading to irrelevant results.
2. **Resume Builders**: Some services provide resume and cover letter templates but do not offer personalized generation or job-matching capabilities.
3. **Using ChatGPT for Job Applications**: Some users may manually input their resume, job descriptions, and company information into **ChatGPT** to generate personalized cover letters and emails. However, this approach has significant limitations:
    - **Potential Data Leaks**: Manually entering sensitive information into third-party platforms raises concerns about data privacy and security, as these platforms may not be designed for handling confidential user data such as personal resumes or proprietary job descriptions.
    - **Lack of Scalability**: While effective for one-off applications, using ChatGPT for multiple job applications becomes time-consuming and impractical. The process requires manual inputs for every job, which does not scale effectively when users need to apply for many jobs at once.

Limitations: These existing methods focus on one aspect of the process but lack end-to-end automation. They fail to provide semantic job matching that goes beyond keyword filtering, limiting the effectiveness of job matches and communication. Additionally, using tools like ChatGPT for manual generation introduces privacy risks and scalability issues, making it an impractical long-term solution for job seekers with numerous applications.

3. Solution Proposal

Proposed Solution Using LLM:

ApplyPro.ai provides a simple and easy-to-use solution to the job search problem. The process to use the application is described in the flowchart below:



The user uploads his resume in PDF format, which is then parsed to extract the text. This text is then embedded using a Large Language Model to get the numerical embedding vector for the resume. This will then be compared against a database of job postings that have also been embedded using the same embedder model before being stored. This matching is done through semantic vector search to identify the top three jobs that are most aligned with your profile as expressed in the resume. This task is performed by an agent that then outputs the details of the top three job matches. Based on the top job match and using the context from the applicant's resume, two separate agents generate the cover letter for the job and a cold email to the recruiter for the job application, respectively. The cover letter will be generated in .docx format and the email will be generated in .eml format that will also have the resume attached to the email.

Key Features: This application utilizes three key technologies:

1) LLM summarizer to summarize the details of the job posting before they are stored in the database

2) Embedder model (Nomic-Sentence Transformers) to transform both job postings and the resume to numerical embedding vectors for comparison.

3) LLM powered AI agents (ChatGPT 3.5 in this case) that autonomously perform tasks such as job matching and displaying top three matches, generating a cover letter in Word file for the top match, as well as generating an email message to the recruiter for the job application.

Impact: The ApplyPro.ai solution automates much of the job application process and has significant impact in terms of increasing the odds for successful recruitment. It not only picks out the jobs that are the best fit for you, but also saves you the time and effort of having to write a cover letter and application email - which, when considering the number of applications one has to send out in order to get a job, saves a significant amount of time that can be put to more productive use like interview preparation.

Addressing Ethical Implications and Challenges:

- **Bias in Job Matching Algorithms**
  - AI systems can unintentionally favor certain candidates, disadvantaging those from non-traditional backgrounds. For example, candidates with fewer years of experience or unconventional career paths may receive fewer relevant matches.
  - Mitigation Strategy: ApplyPro.ai will conduct regular audits to detect bias, ensuring fairer recommendations across diverse backgrounds. Techniques like re-weighting underrepresented profiles will also help reduce bias, making the system more inclusive.
- **Privacy and Data Security Risks**
  - Handling personal data like resumes poses significant privacy risks. Any improper handling of sensitive user information could lead to data breaches.
  - Mitigation Strategy: The system will use encryption to secure user data and comply with data protection laws like GDPR. ApplyPro.ai will also provide users with clear privacy controls, including data deletion options.
- **Transparency and Accountability in AI Decisions**
  - Users may not understand how job matches are generated or why certain recommendations are made. The AI model's decisions may appear difficult to interpret. This can create distrust if users don't feel confident in the system's recommendations.
  - Mitigation Strategy: ApplyPro.ai will implement explainable AI techniques, giving users insights into why certain jobs are recommended. This transparency will help build trust and clarify AI decision-making.
- **Hallucinations or Incorrect Content Generation**
  - The GPT-based models generating cover letters or emails could hallucinate (generate incorrect information), such as adding irrelevant content or fabricating job-related facts. This could lead to inaccurate or unprofessional communication.
  - Mitigation Strategy: ApplyPro.ai will fine-tune models to reduce hallucinations and provide users the ability to review and edit AI-generated content before submission (keeping the parameter like temperature parameter low), ensuring accuracy and professionalism.
- **Scalability**
  - Handling increasing user numbers and job postings may challenge the platform's capacity
  - Mitigation Strategy: ApplyPro.ai will use scalable cloud infrastructure, distributed databases, and optimized vector search engines fast, real-time performance even under heavy loads.
- **User Adoption**
  - Gaining user trust and adoption can be difficult, especially with concerns over AI-generated content and privacy.

    - Mitigation Strategy: ApplyPro.ai will focus on a user-friendly design, transparent data policies, and partnerships with career services, empowering users to confidently use the platform and boosting adoption.
- **Ethical Use of AI for Job Applications**
    - Using AI to automate cover letters and emails raises questions about authenticity. Recruiters may feel misled if the application appears to be handcrafted when it was AI-generated. This can create ethical concerns about fairness and transparency in the hiring process.
    - Mitigation Strategy: ApplyPro.ai will include a feature that informs users to disclose AI-generated content, ensuring transparency in communication. Additionally, users will be encouraged to review and personalize the AI-generated cover letters and emails, striking a balance between efficiency and authenticity. This approach promotes fairness while leveraging AI's capabilities.

4. Technical Implementation

Model and Algorithm Choices:

- **Embedding Model**: The system utilizes *SentenceTransformer (nomic-ai/nomic-embed-text-v1)* for generating text embeddings, which is optimized for semantic similarity tasks. This choice ensures that both job descriptions and resumes can be efficiently compared.
- **Summarization Model**: A transformer-based *facebook/bart-large-cnn* model is used for summarizing long job descriptions for job postings, allowing the system to condense relevant information.
- **Language Generation**: **GPT-3.5-turbo** is selected for generating coherent and contextually accurate cover letters and cold emails. This model provides high-quality natural language generation, making it ideal for producing professional and customized text.

Tokenization & Embedding:

- **Dynamic Tokenization**: Text is dynamically chunked into manageable segments, ensuring that summarization and embedding tasks are efficient.
- **Vector Search**: Embeddings generated from job descriptions and resumes are stored in a vector space, allowing for fast and accurate retrieval of the top matches using cosine similarity.

Step-by-Step Technical Workflow:

1. **Data Preprocessing**:
    - **Resume Parsing**: function of this is to extract and cleans text from PDFs, handling various layouts and formats. It ensures smooth integration with subsequent tasks.
    - **Summarization**: This function divides long job descriptions into smaller chunks, applies dynamic summarization using **BART**, and then recombines the summaries. This approach ensures that only the most relevant information is retained, facilitating more accurate matching.

2. **Embedding & Vector Search**:
    - **Embedding Generation**: The solution uses *SentenceTransformer* to convert summarized job descriptions into embeddings. These embeddings capture the semantic essence of the texts.
    - **Top Job Match Identification**: The **search_agent** utilizes the vector space to retrieve the top 3 job matches by comparing the resume embedding against stored job postings. This is achieved through efficient cosine similarity calculations.
3. **Text Generation for Applications**:
    - **Cover Letter Generation**: The **generate_cover_letter** function uses the GPT-3.5-turbo model to create personalized cover letters based on the top job match. The model incorporates details from the resume and job description to generate contextually relevant and professional text.
    - **Cold Email Drafting**: Similarly, the **generate_cold_email** function produces a polite email template that can be sent to recruiters, helping users initiate communication based on the job match.
4. **Custom Classes & Agents**:
    - **Agent Design**: The code introduces agents like **search_agent, cover_letter_agent, and cold_email_agent**, each tasked with a specific goal. This modular design allows for easy extension and testing of individual components.
    - **Utility Functions**: Functions for saving cover letters as DOCX files and cold emails as EML files streamline the user experience by providing readily usable outputs.
5. **Libraries & Technologies**:
    - **Transformers**: Utilized for summarization and language generation tasks, leveraging the *facebook/bart-large-cnn and GPT 3.5* models.
    - **SentenceTransformer & Cosine Similarity**: Used for efficient embedding generation and similarity search, enabling quick identification of job matches.

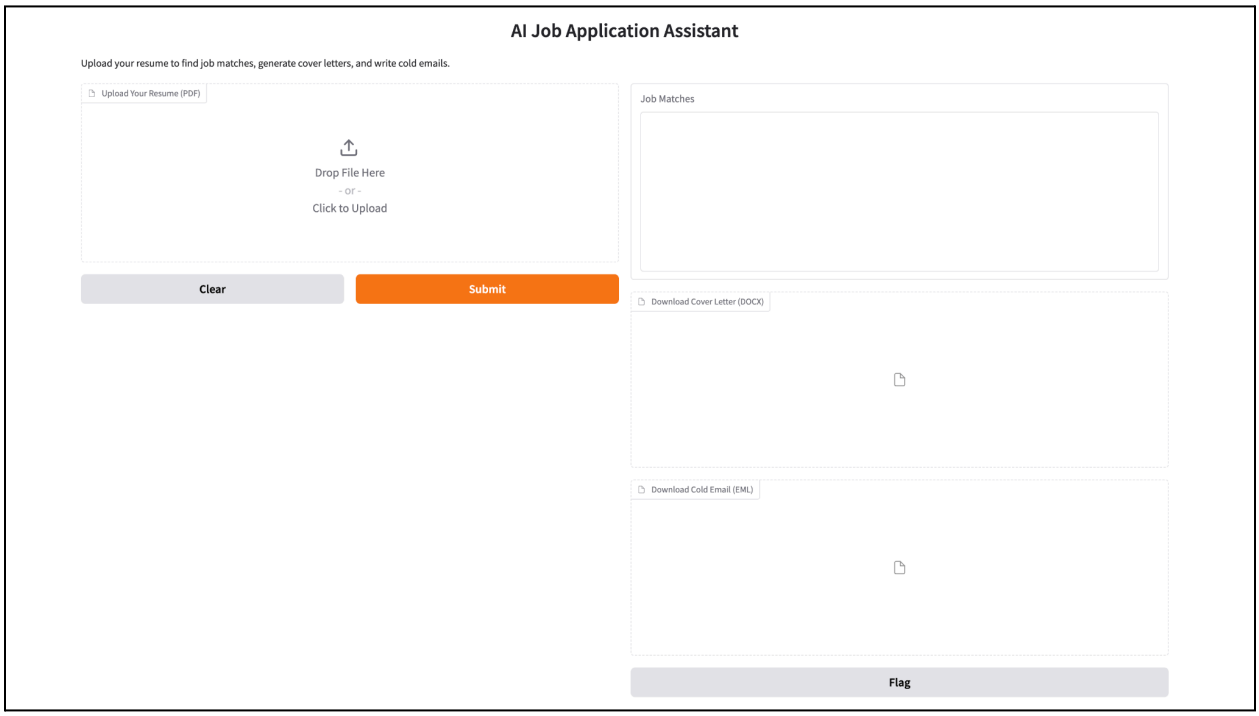
User-Friendly Design:

1. **User Interface & Experience**:
    - The solution is designed with an intuitive, straightforward interface that guides users through each step of the job application process. Users can easily upload their resume, receive job recommendations, and generate cover letters and emails with minimal manual input.
    - **Accessibility**: Users can upload resumes in multiple formats (e.g., PDF, DOCX), making the system flexible and widely usable.
    - **Automated Document Generation**: Once job matches are identified, users can immediately download generated cover letters and emails, saving time and ensuring professionalism.

5. Demo Walkthrough

Live Demo Features:

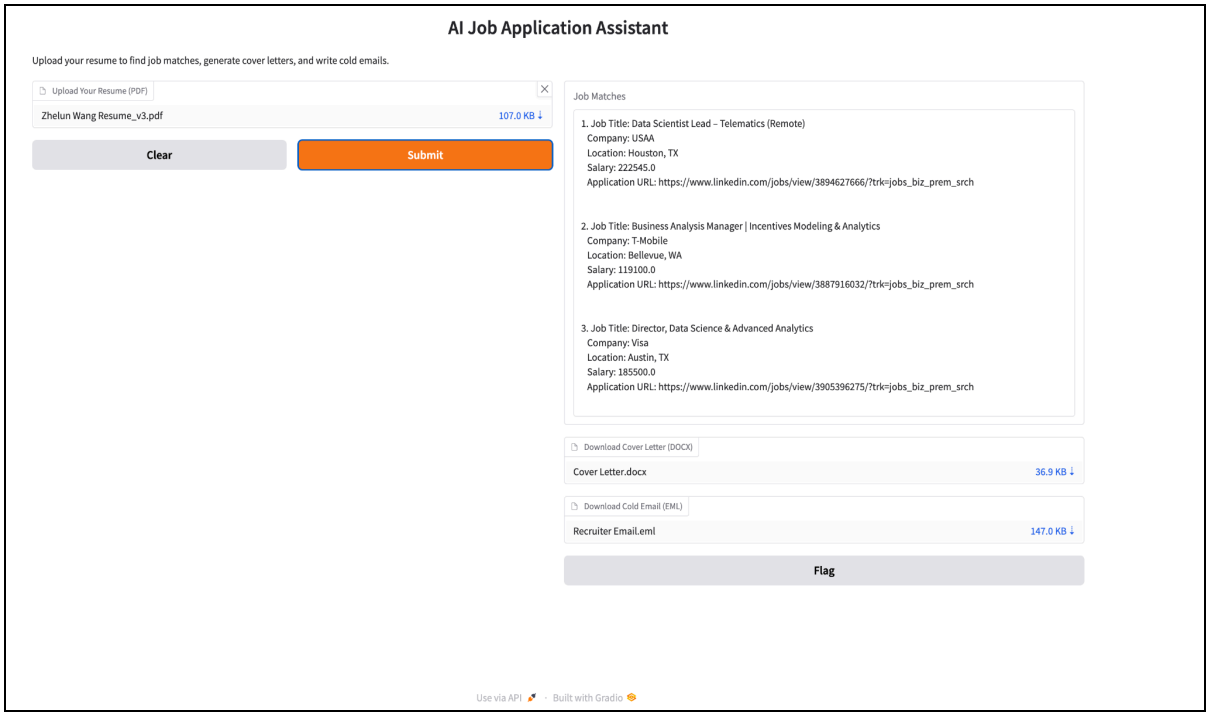
The home page of the application is as follows:



The entire process is just based on two steps:

1. The user uploads their resume on the left side pane and clicks the submit button.
2. The application takes a few seconds to process the resume, match against job postings and generate the outputs: top job matches, cover letter and application email. The user then downloads the cover letter and email.

The output of the application is shown below:



We can see that the three top job postings for you are displayed along with the key information such as job title, company name, location, salary and the application URL for you to apply. Also, the cover letter file and the application email file are also generated. These three outputs automate a great part of the job search and application process. You can now readily apply to the best jobs for you, attach a ready made cover letter, and just send the email draft to the recruiter.

6. Reflection on Learning

Learning Journey:

Working on **ApplyPro.ai** was a transformative experience for the entire team. Throughout the project, we collectively gained valuable technical and professional skills. One of the most important technical lessons was the ability to effectively utilize **Large Language Models (LLMs)** for natural language generation and job-matching tasks. Our team became proficient in implementing **semantic embeddings**, which enabled a more context-aware approach to resume-to-job matching, significantly improving the quality of the matches.

On the professional side, this project provided us with deep insights into how to solve real-world business problems. Understanding the pain points experienced by both job seekers and employers allowed us to translate these challenges into a technical solution that could add meaningful value. Balancing the technical aspects of the project with user experience and data privacy concerns was a key takeaway for the team.

Challenges and Solutions

The project presented several challenges, and overcoming them helped the team to learn and adapt throughout the development process:

1. **Matching Accuracy**: Our initial resume-to-job matching system relied on keyword matching, which often returned irrelevant results. By integrating **cosine similarity** with **semantic embeddings**, we were able to significantly enhance the accuracy and relevance of job recommendations. This improved our ability to match resumes to jobs based on context rather than simple keyword filtering. However, a key area that can be improved is the summarization of the job descriptions for the job postings using BART. Given the time and compute, ideally we would like to fine tune BART on job posting summarization data to accurately retain the core features of the job postings.
2. **Generating Personalized Content**: One of the biggest challenges was crafting cover letters and cold emails that felt personalized and relevant for each job application. Early versions of the content generated by the LLM felt too generic. To address this, we did some prompt engineering for the **GPT models** to produce more personalized and professional cover letters and emails based on the job descriptions and resume data.
3. **Data Security**: Ensuring the secure handling of sensitive data such as resumes and job descriptions is a critical challenge. A future improvement can be to use an in-house model to handle this data, or remove personally identifiable data before passing it through OpenAI.

Through these challenges, the team learned the importance of **iterative development** and continuously refining solutions to meet user needs and technical requirements.

7. Conclusion

Summary of Findings:

The **ApplyPro.ai** project successfully addressed key inefficiencies in the job application process by automating several time-consuming tasks. Our tool streamlined the process of job searching, resume matching, and personalized cover letter generation using **Large Language Models (LLMs)** and **semantic search** technologies. By integrating AI into the workflow, we significantly reduced the time and effort required by job seekers while improving the accuracy of job matches.

Our LLM-driven approach not only automates job matching but also ensures personalized and professional communication for job applications. This drastically improves the chances of applicants securing interviews, providing value to both job seekers and employers.

Next Steps

Looking ahead, there are several areas where **ApplyPro.ai** can be further optimized:

1. **Customization Options**: While the tool currently generates personalized cover letters and emails, future iterations could allow users to manually edit and customize content further before submission. Also, we can add options for job preferences such as salary, location etc. to incorporate into the search process to make job matching more customized.
2. **Real-Time Integration with Job Boards**: Connecting the system to live job boards will allow real-time job matching, ensuring users get the most up-to-date job recommendations. This will further enhance the relevance of the job postings.

3. **Fine-Tune BART Summarizer:** We can fine-tune the summarizer on data for our specific use-case in order to effectively summarize job posting descriptions and retain the core information important to the matching process in order to make the matches more accurate.
4. **Feedback and Analytics**: Adding an analytics feature to track the performance of job applications (e.g., response rates, interview invites) would provide users with valuable insights on how to improve their applications.

By focusing on these improvements, **ApplyPro.ai** has the potential to become an essential tool for job seekers, talent acquisition platforms, and career service providers, bringing significant efficiency and personalization to the job application process.