Fair AI Recruitment: Integrating Feedback Mechanisms for Ethical Hiring

DATA417-Group Project

Abstract

With the development of AI technology, its integration into various aspects of our lives has become increasingly prevalent, including in employment systems. One notable example is Amazon's AI recruitment recommendation system, which can not only help companies screen resumes more efficiently but also eliminate human emotional bias factors, making the recruitment process fairer and more reasonable. However, this system has exhibited significant ethical issues, such as gender bias and lack of transparency. Our project addresses the lack of transparency concerns by integrating a feedback mechanism into the AI recruitment process.

Our enhanced system not only matches candidates' resumes with job descriptions and calculates a similarity score but also provides constructive feedback to candidates. During the implementation, we built upon a publicly available algorithm as a foundation and used Python for its robust data analysis and machine learning capabilities. Our word-based algorithm processes resumes and job descriptions, removing common stop words to focus on key terms and skills. It then calculates the cosine similarity score and identifies missing words, generating feedback for candidates to help them improve their qualifications for future job opportunities.

We also integrated an API Large Language Model (LLM) to analyze the semantics of the resumes and job descriptions, providing deeper insights and more accurate feedback. For data management, we propose a consent-based strategy that ensures candidates have control over their personal information when they upload their resumes, aligning with ethical standards and fostering trust and transparency. This approach gives the agency to candidates, allowing them to decide how their data is used.

Both models provide valuable feedback and address the missing transparency in these AI recruitment systems. However, we encountered challenges with the LLM model's consistency, as it provided different results in every run, highlighting the need for well-crafted prompts and further refinement of the model. And for word-based algorithm is simply comparing keywords often misses nuances in meaning. Future work will focus on integrating more advanced technologies and refining the algorithms to better handle context and semantics, ensuring more consistent and accurate feedback. By introducing a feedback mechanism, we address transparency and candidate empowerment, ultimately contributing to a fairer and more effective recruitment process.

Keywords: Amazon's AI, recruitment, feedback mechanism, recommendation, LLM, Python, ethical issues, transparency, fairness, candidate empowerment.

Background

At the beginning of our project, after brainstorming our Project idea, we decided to address Amazon's AI recruitment system, which is used for selecting candidates based on their resumes in the initial phase of the recruitment process. When we decided that our data science algorithm idea would focus on Amazon's AI recruitment system, our group engaged in the spider game to deepen our understanding of the system. This basic diagram of our spider game session is presented in Figure 1.

Figure1: Spider Game Diagram on Amazon's AI Recruitment System

Artificial Intelligence (AI) is rapidly expanding globally, significantly impacting various sectors with both positive and negative outcomes. In 2014, Amazon developed an AI-driven resume rating algorithm to address the challenges of increasing recruitment times and costs for mid-to-high-level positions. This algorithm, which rates resumes from 1 to 5 stars, was designed to streamline the hiring process by comparing new applicant resumes to those of successful hires

from the past decade. Utilizing techniques in natural language processing (NLP) and machine learning (ML), the tool aims to identify top candidates by detecting key traits that align with previously successful profiles, thereby enhancing the efficiency and effectiveness of Amazon's recruitment strategy. (Alfons Kodiyan, 2019)

However, this AI recruitment system has been producing unfair results, with more men being selected than women, leading to gender bias, which is a significant ethical issue (*Amazon Ditched AI Recruiting Tool That Favored Men for Technical Jobs | Amazon | The Guardian*, n.d.).

Ethical issue

An ethical issue is a situation where a moral conflict arises, challenging the principles or standards of a society or workplace. Initially, we aimed to resolve this gender bias observed in Amazon's AI recruitment system. However, our focus then shifted to another ethical concern: Since the AI system rejects candidates based on its algorithm during the first phase, applicants never reach a human recruiter. This lack of transparency leaves them unsure of the reasons for rejection, potentially leading to demotivation. We believe this practice is unfair and raises ethical concerns regarding candidate experience and potential bias within the AI itself.

The ethical issue we are addressing in this project focuses on the lack of transparency in AI-driven resume screening. This opaqueness not only demotivates applicants but also hinders their ability to learn and improve for future applications. Additionally, concerns arise regarding potential bias within the AI itself. If trained on historical data that reflects existing inequalities. Addressing these concerns is important for ensuring fairness and equality in Amazon's AI-driven recruitment processes.

Introduction

According to the 2022 Candidate Experience Report, 60% of job candidates believe there is a significant need for companies to improve their recruitment experiences. Furthermore, 70% of candidates express a desire for feedback, highlighting a widespread demand for more transparent and communicative hiring practices. Interestingly, 61% of candidates also indicated they would be more inclined to reapply for future roles at a company from which they had previously received feedback. (2022 Candidate Experience Report, n.d.) These statistics underscore a critical opportunity: providing feedback is not just a valued practice, but it could also be an effective strategy to minimize biases in the recruitment process. Additionally, feedback helps candidates become more optimistic and view the process as fair, which can boost their confidence for future opportunities. By adopting a feedback-oriented approach, companies can enhance candidate engagement, foster goodwill, and potentially reduce the perpetuation of biases, thereby creating a more equitable recruitment environment. Therefore, we decided that improving an algorithm to provide feedback to candidates is a sound approach.

For companies and managers, the AI recruitment system gives them the list of applicants they want to interview, but for candidates, the recommendation system does not help them. For this reason, we want to change the system to make the recommendation system give some feedback and suggestions to the candidate. For example, if the candidate does not meet the recruitment requirement of "mastering Python language" in the resume, the system will give feedback that he needs to learn Python language to get the probability of getting a job. Rather than being screened by traditional AI recruitment systems without any useful improvements. Our group thinks this is very necessary.

Our team graded candidates mainly by comparing the similarity between their resumes and the company's job requirements. The higher the rating, the better the chance of getting an offer. At the same time, the recommendation system will also re-generate "missing words" feedback to the user for the weak part of the job requirement but the candidate's resume, reminding the user to improve and learn what skills to better meet the requirements of the position or industry. After giving guidance, the candidate can perform better in the next interview, thus increasing the success rate. For a company listing, the recommendation system generates an 'excess words' indicating the candidate has a skill not mentioned in the job description. This section can be opened to the company to select the best candidate for the position.

Methodology

Due to the proprietary nature of Amazon's recruitment algorithm, direct access to its code was not possible. However, we identified similar algorithms available in the public domain and found one on GitHub that closely aligned with our approach. This publicly available algorithm served as a foundation for our project algorithm. (*Rohitbakoliya/Sieve.Ai: Designed for Recruiters, Our AI-Powered Platform Can Filter out Top Resumes of the Stack*, n.d.)

For the implementation, we chose Python as our programming language due to its robust libraries for data analysis and machine learning tasks. To coordinate our efforts and ensure that all team members were aligned and could contribute effectively, we used a shared GitHub repository. (*MattTheRealYoung/DATA417_Group_Project: This Repo Is for DATA417 Group Project of Team DSC*, n.d.)

New Algorithm

This new algorithm has 2 parts: A word comparison and a LLM (Large Language Model) comparison (Figure2). Both methods start by using Textract to scrape text from a job description (written in text in our prototype) and a resume (a PDF in our prototype).
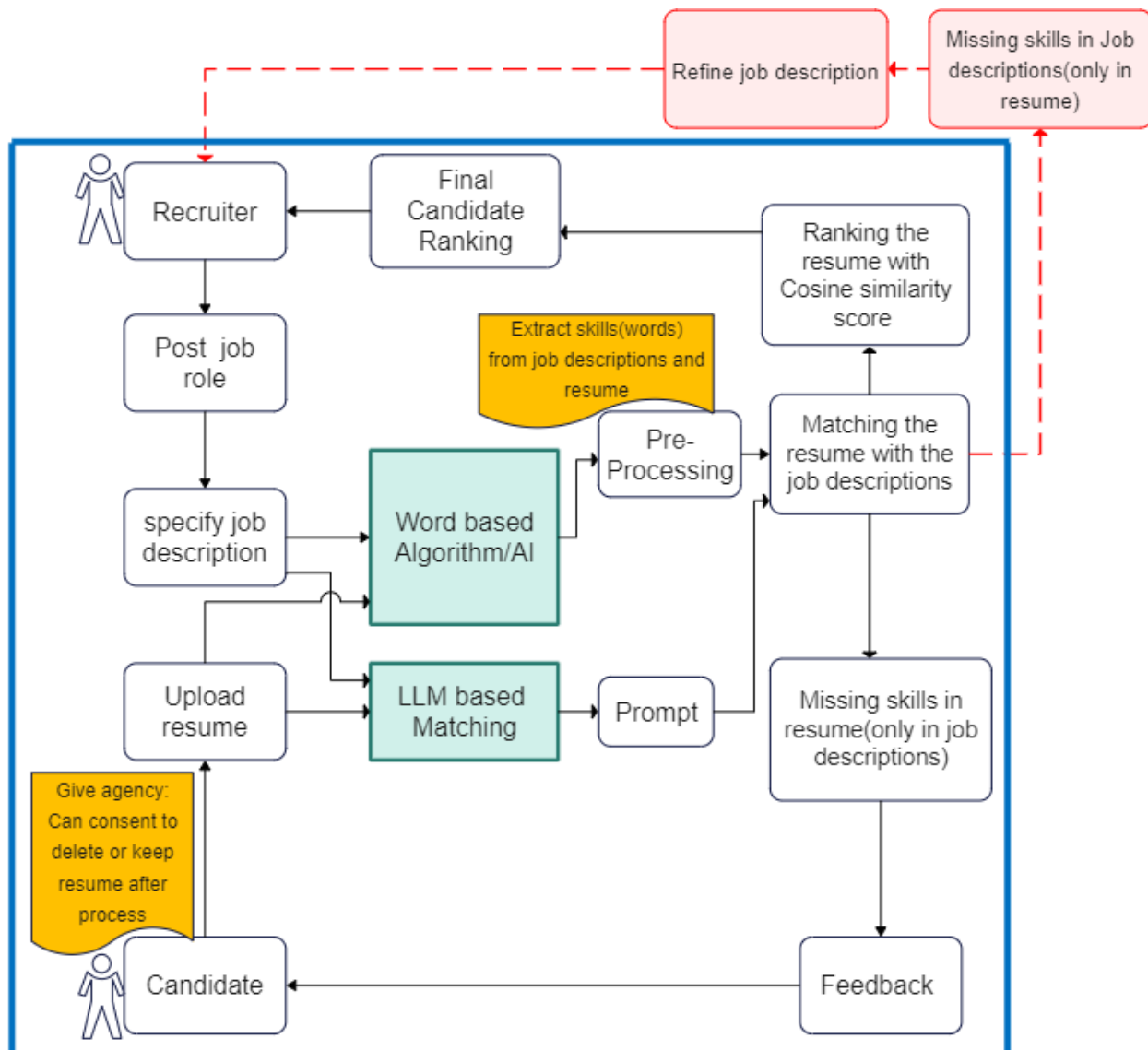
Figure2: System Structure

Word based matching feedback

The word comparison removes all punctuation and common words and turns the rest of the words into tokens. The words in both inputs are compared to the words in the job description with cosine similarity to give a percent match score. The words that are only in the job description are returned as missing words to let the candidate know which words are missing from their resume that can be added (if they have those skills) or work on in real life to be more appealing to the job. Words that are only in the resume can be returned to refine the job description if needed, but this would be a different use case than originally made for. The red arrows in the system diagram (Figure2) indicate potential future enhancements that are currently out of this project's scope, such as identifying missing skills in job descriptions that are only in resumes and refining job descriptions.

LLM (Large Language Model) based feedback

Leverage Langchain framework (*LangChain*, n.d.)*,* we connect to OpenAI LLM (Large Language Model) to analyze the resume and the job description. Compared to cosine similarity, semantics are considered in LLM, which will give us more insight about the feedback. However, since we don't understand how the comparison is processed, we receive different answers each time we ask for the comparison. Meanwhile, we can find unspecified requirements in our job description. We mentioned Bachelor of Engineering, Bachelor of data science and Bachelor of mathematics are needed, but we don't clarify do we need all these degrees or have one of them, thus LLM will get confused about process the bachelor's degrees as "or" or "and" relationships. To improve the accuracy of LLM results, we should detail our job description and improve our prompt to clarify our requirements either.

In LLM interaction code, first we use langchain Api to load the resume pdf and jobtype txt file for job description (Figure3).

```
/Users/mattyang/Documents/workspace/pythonProject/.venv/bin/python /Users/mattyang/Documents/workspace/pythonProject/Llm_explore.p
This is resume : PUBLICATION LISTJournal Articles1.He Zhang, Zhirong Yang, and Erkki Oja. Improving Cluster Analysis By Co-initial
This is job description : Data scientist

Needs:
machine learning algorithms experience

Bachelor of Engineering
Bachelor of Data Science
Bachelor of Mathematics

Must be fluent in English, Writing and Speaking

PhD preferred
```

Figure3: LLM interaction code (jobtype.txt file and resume pdf)

Next, use job descriptions and resumes to form our prompt to enquire OpenAI API, also add what we'd like LLM to help us with. Append the requirements "Compare the job description and resume, give a score of similarity between 0 and 100 percentages, List the matching skills, experiences, and qualifications. Identify any gaps or areas where the resume does not meet the job description requirements".

Data Management

As resumes contain personal data, handling candidates' resumes ethically and responsibly is very important. We can use a strategy that involves asking candidates for their consent regarding the future use of their personal information as they upload their resumes (Figure2). They can be given options to allow the company to retain their resumes for consideration in future job openings, consent to their resumes being used to train and improve the company's AI models or request the deletion of their resumes after the current recruitment process concludes. This approach ensures that candidates have control over their personal data, aligning with ethical standards and regulatory

requirements. By providing these options, we respect candidates' privacy and autonomy, fostering trust and transparency in the recruitment process. Additionally, for those who consent to their data being retained, it allows the company to build a more robust dataset for future hiring and model training, thereby improving the accuracy and fairness of the AI recruitment system. This consent-based data management strategy not only addresses ethical concerns but also enhances the overall efficiency and effectiveness of the recruitment process.

Results and Discussion

The feedback from word-based algorithm is limited as it can only tell you if the words match. It does not take into account the meaning of sentences. For example: if the job description said, "Needs to be fluent in English." and the resume said "Cannot speak English. Fluent in Java." The key words "Fluent" and "English" are present in both despite not being a match at all.

An example of a data science job description compared to a data scientist's resume can be seen in Figure 4. This same job description with a graphic designer's resume can be seen in Figure 5. The word "addedword" was added and compared to the data scientist's resume which gave the results in Figure 6.

```
Missing Words: ['mathematics', 'fluent']


 Similarity Score:
[[1.          0.93541435]
 [0.93541435 1.          ]]
Your Resume matches about 93.54% of the job
```

Figure 4: Example of a data science job description compared to a data scientist's resume.

```
Missing Words: ['data', 'scientist', 'machine', 'learning', 'algorithms', 'engineering', 'mathematics', 'fluent',
'english', 'writing', 'speaking', 'phd']


 Similarity Score:
[[1.  0.5]
 [0.5 1. ]]
Your Resume matches about 50.0% of the job
```

Figure 5: Example of a data science job description compared to a graphic designer's resume.

```
Missing Words: ['mathematics', 'fluent', 'addedword']


 Similarity Score:
[[1.          0.90748521]
 [0.90748521 1.          ]]
Your Resume matches about 90.75% of the job
```

Figure 6: Example of a data science job description compared to a data scientist's resume with a nonsense word added to the job description "addedword".

After the process of LLM, we will receive the result of comparison (Figure7).

```
The AI responsed: Score of similarity: 60%

Matching skills, experiences, and qualifications:
- Experience with machine learning algorithms
- PhD in a related field
- Fluent in English

Areas where the resume does not meet the job description requirements:
- The resume does not specifically mention a Bachelor's degree in Engineering, Data Science, or Mathematics, although it does list
- The resume does not explicitly mention experience in data science as a job role.

Process finished with exit code 0
```

Figure7: LLM result of comparison

We can put the LLM response and words cosine similarity response of previous algorithm together, looking for common ideas on both comparisons, and the different ideas could be used to improve our algorithms and future study.

The LLM aims to solve the lack of meaning transference by having an AI analyze the resume. It does this by taking the words from each text source and asking Open AI to compare them. The prompt provides both a percent similarity and feedback to note what is missing from the resume. The problem is the lack of absolute feedback, the AI model may not comment on something, or misinterpret what is passed through to it. Another massive issue is consistency. If you pass the same input multiple times, then the score and feedback will fluctuate (*Microsoft Azure OpenAI vs. OpenAI APIs: A Comparison*, n.d.). We have observed that the performance of the LLM fluctuates significantly with changes in the prompt, even when the task and input data remain constant (Thomas et al., 2023). All the general problems with AI are applicable here. However, it will not make the same mistake with the "Needs to be fluent in English." and "Cannot speak English. Fluent in Java." problem. It is very useful as recruiter AI systems will probably act similarly and thus can make your resume look better from an AI perspective, which is what we are trying to solve.

By using both these systems in tandem, the user can get both concrete and consistent feedback of missing words and more general AI based feedback which provides the missing transparency from these AI recruitment systems. The user will be able to iteratively improve their resume with these tools and have a higher chance of getting to the interview where they will have a higher chance of getting the job.

By integrating a feedback mechanism into the AI recruitment system, we can address the ethical issues related to transparency and candidate empowerment. Candidates who received feedback can get satisfaction from the recruitment process, as they gained a clearer understanding of the skills required for their desired positions. This transparency not only mitigates feelings of frustration and demotivation often associated with automated rejections but also empowers candidates to improve and tailor their skills to better align with job requirements in the future. Incorporating both methods in this feedback mechanism could enhance the effectiveness and fairness of AI-driven recruitment processes.

Conclusion

While both word-based feedback and LLM-based feedback provide much-needed transparency to the AI recruitment systems, there is still significant room for improvement. Future work could focus on refining the algorithms to better handle context and semantics, ensuring more consistent and accurate feedback.

By introducing a feedback mechanism for applicants into the AI recruitment recommendation system, our team has greatly improved the fairness and transparency of the process. This design shows that the recruitment recommendation system respects the time and effort of job seekers, and we hope that through this feedback, they can engage in targeted learning and improve their skills.

Major achievements:

1. A level playing field: The resumes of candidates recommended by the AI recruitment system can largely remove the bias of judging the selection, and all candidates are evaluated according to the same job criteria. All these ensure a more harmonious and fair playing field for candidates in job hunting and help job seekers to present themselves in the best state.

2. Improve self-knowledge: The increased feedback mechanism can help applicants better understand their strengths and weaknesses in the position. Through feedback, candidates can improve their resume, enhance their professional abilities, learn skills that may be overlooked, and make them more competitive in future job hunting.

3. Clear direction: Through the feedback mechanism, job seekers can clearly understand the reasons for their failure and whether the company is suitable for them. Feedback allows job seekers to see if they're a good match for the company's culture and skills, allowing them to focus on their other job applications.

4. Increase confidence: Because of the objective and positive feedback, even if the candidate is not selected, it can increase the information of the candidate and motivate them to continue to look for the next job.

On the other hand, it is essential to acknowledge the extensive role AI plays in hiring today. Candidates need to be made aware of the use of AI's services in hiring by employers. From

scanning and evaluating resumes to scoring candidates or conducting interviews, AI assists in various stages of the hiring process. However, we cannot ignore the fact that human involvement remains a vital component. Human judgment adds a layer of fairness and understanding that AI alone cannot achieve. (Burrell & Mcandrew, 2023)

Difficulties & Limitations

Limitations in Keyword Comparison and LLM Consistency: Simply comparing keywords may not be accurate, as it can miss nuances in meaning—for example, "fluent in English" versus "not fluent in English but can speak". Additionally, a significant challenge we observed when using the LLM API method was that the LLM model provided different results in every run, making it difficult to interpret and rely on the feedback. This inconsistency poses a problem for providing stable and trustworthy feedback to candidates. We understood that we need to provide a well-crafted prompt to achieve better and more consistent outcomes from the API.

Technical limitations: In the process of initial resume reading, we failed to utilize advanced AI technology, which may lead to certain errors in the reading of candidates' resume information. We realize that more advanced technologies need to be integrated in the actual input process to further improve the efficiency and accuracy of the recruitment recommendation system.

Quality of feedback: Feedback that is helpful to the candidate should be specific, actionable, and constructive. Providing vague or overly general feedback is not helpful. Therefore, we need to continuously optimize and ensure the timeliness and effectiveness of feedback.

Future work

Technology integration and optimization: It would be good add other LLM technology and more advanced technical solutions to the actual investment to strengthen the system's processing capacity for natural language, to enhance the recommendation system's information reading technology for job seekers' resumes. The multiple opinions will give the user a better idea of how different AI models will interpret their resume. And enhance the algorithms to better understand the context and nuances of job descriptions and resumes, reducing misinterpretations.

Quality of delivery: The feedback system should be more professional and timelier, which will be more helpful to the applicant. For practical input, consider providing telephone or face-to-face meetings to provide detailed, targeted, and interactive feedback to applicants. A nice graphical user interface will make this product much more professional and will increase ease of use.

Continuous monitoring and improvement: To cope with the future processing of large numbers of job seekers' resumes, we should continuously update and monitor algorithms and data to ensure that all steps in the recruitment process are fair and transparent, creating greater efficiency and value for both job seekers and employers.

Acknowledgements

References

2022 Candidate Experience Report. (n.d.).

Alfons Kodiyan, A. (2019). *An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool*.

Amazon ditched AI recruiting tool that favored men for technical jobs | Amazon | The Guardian. (n.d.). Retrieved May 13, 2024, from https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

Burrell, D. N., & Mcandrew, I. (2023). Exploring the Ethical Dynamics of the Use of Artificial Intelligence (AI) in Hiring in Healthcare Organizations. *Land Forces Academy Review*, *28*(4), 309–321. https://doi.org/10.2478/raft-2023-0037

LangChain. (n.d.). Retrieved May 29, 2024, from https://www.langchain.com/

MattTheRealYoung/DATA417_Group_Project: This repo is for DATA417 group project of team DSC. (n.d.). Retrieved May 14, 2024, from https://github.com/MattTheRealYoung/DATA417_Group_Project

Microsoft Azure OpenAI vs. OpenAI APIs: A Comparison. (n.d.). Retrieved May 29, 2024, from https://www.willowtreeapps.com/craft/openai-or-azure-openai-can-models-be-more-deterministic-depending-on-api

rohitbakoliya/sieve.ai: Designed for recruiters, Our AI-powered platform can filter out top resumes of the stack. (n.d.). Retrieved May 14, 2024, from https://github.com/rohitbakoliya/sieve.ai

Thomas, P., Spielman, S., Craswell, N., & Mitra, B. (2023). *Large language models can accurately predict searcher preferences*. http://arxiv.org/abs/2309.10621

Appendix
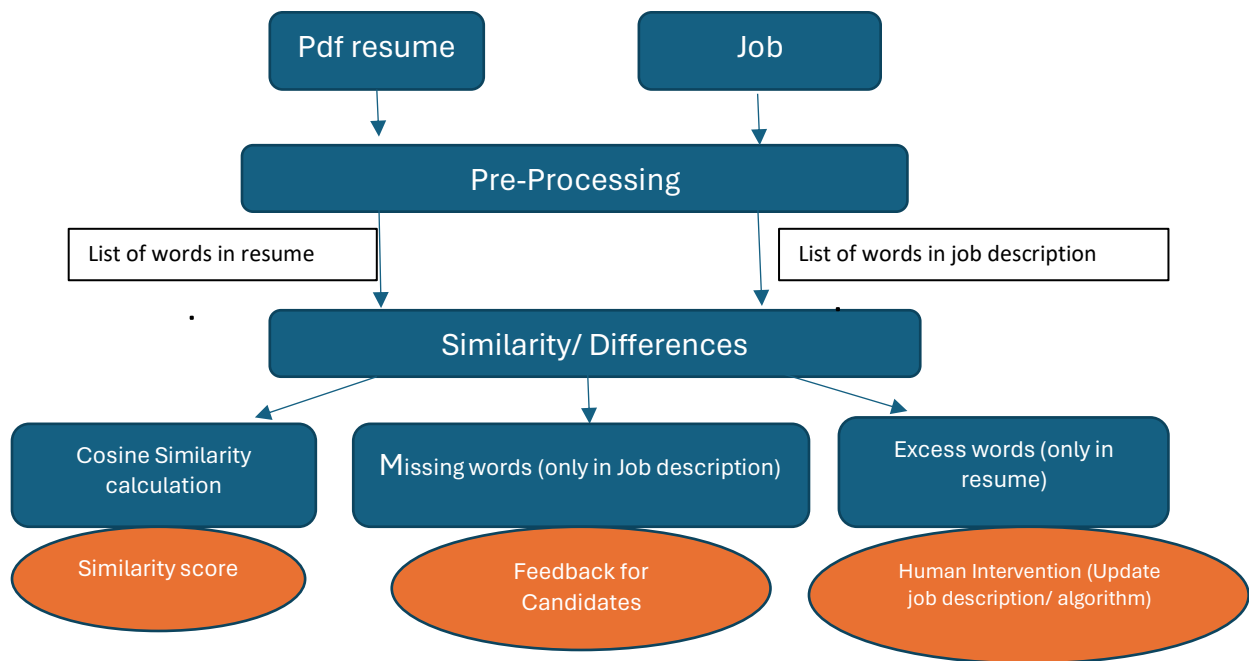
System Structure with Word based and LLM Integration