TEAM PROCRASTINATION

INTRODUCTION

In the ever-evolving talent acquisition landscape, HR teams face the monumental task of efficiently evaluating countless resumes and recommendations to identify the best candidates. The traditional manual review process is time-consuming and prone to human bias, inconsistencies, and oversight of crucial details. Here comes Satya, an innovative AI-powered system designed to revolutionize HR decision-making.

Satya's primary mission is to assist HR professionals in navigating the complex world of resumes, recommendation letters, and professional networks. By leveraging advanced AI techniques, Satya aims to uncover hidden patterns, verify credentials, and provide insightful analysis beyond surface-level information.

The core challenge lies in developing an AI system that can assess a candidate's qualifications based on their resume while evaluating the credibility of their professional network and recommendations. This includes detecting potential fraud or biases in hiring and ensuring that only qualified and genuine candidates are identified.

Through this project, we seek to explore the potential of AI in transforming HR practices, making the hiring process more efficient, accurate, and fair. The utilization of AI not only streamlines the recruitment process but also enhances the quality of hiring decisions, thereby contributing to the overall success of organizations.

METHODOLOGY AND MODELS

Our implementation of Satya leverages advanced AI techniques to analyse resumes and recommendations comprehensively. Here's a detailed overview of our approach:

1. LLaMA-based Models: We utilized LLaMA 3.2 (Large Language Model Meta AI) as the foundation for our AI analysis. This choice allows us to harness sophisticated natural language processing (NLP) capabilities, enabling the system to understand and interpret the nuances of human language within resumes and recommendations. LLaMA's architecture is particularly adept at contextual analysis, ensuring the model captures the significance of phrases and terminologies used in candidate submissions.

2. Vagueness Scoring: We developed a model that assigns a vagueness score to evaluate the specificity of experiences described in the CV. This scoring system identifies potentially inflated or ambiguous claims by analyzing the language used to describe work experiences. For example, vague phrases like "great potential" and "incredible enthusiasm" etc. may have a higher "vagueness score", while concrete metrics like "increased sales by 20%" would score lower in this regard. This helps HR professionals to pinpoint areas where candidates may exaggerate their qualifications.

3. Recommendation Trust Factor: We implemented a scoring system to assess the credibility and strength of recommendations provided by recommenders of candidates. This trust factor is determined by analyzing various elements, such as the specificity of the recommendation, the

language of the recommendation, the semantics of the word flow, etc. A higher value of it tells a favorable choice of candidates

4. Reciprocal Endorsement Detection: To enhance our ability to detect potential bias in recommendations, we created a mechanism for identifying and flagging instances of mutual endorsements. This feature recognizes patterns where professionals endorse each other reciprocally, which may suggest collusion or bias. By highlighting these endorsements, Satya empowers HR teams to evaluate the recommendations' authenticity critically.

5. CV Section Extraction: Automation is crucial to Satya's efficiency. We developed an automated process for extracting and categorizing different sections from resumes, such as education, work experience, skills, etc. This enables a more focused analysis of each component, allowing the system to assess each section for specific criteria, which can be pivotal in making informed hiring decisions.

6. Networking Analysis: The idea of our version of Satya's Networking Analysis is to create a graph representation of professional networks, illustrating relationships among candidates, recommenders, and work histories. Such a graph with differently weighted connections based on certain criteria (discussed later) allows HR personnel to identify key individuals and strongly connected people in the community. This allows for measuring the collaboration capabilities and soft skills of a candidate.

7. Timeline Analysis: Satya also performs a timeline analysis to evaluate the continuity and progression of a candidate's career. Large, unexplained gaps in the employment history are flagged as areas of concern. For instance, if a candidate has a gap of several years between two roles without any clear reason (e.g., education, sabbatical), Satya will lower the overall score for that candidate. This helps HR teams identify candidates whose career trajectory may need further investigation.

8. Comprehensive CV Scoring: The final scoring system integrates all these factors into a comprehensive CV assessment. By providing a holistic view of a candidate's qualifications, Satya helps HR professionals make more informed choices.

This multi-faceted approach allows Satya to provide a nuanced and comprehensive evaluation of each candidate, helping HR teams make better decisions in the hiring process.

PERFORMANCE METRICS AND FEATURES

The performance of Satya's CV analysis system is based on a multifaceted approach that delves deep into various aspects of each résumé. While most metrics are numeric, some are in the form of flags which will appear in the interactive dashboard to be brought to the attention of HR personnel. Our system's effectiveness is determined by its ability to calculate and integrate several key features:

1. Vagueness in Experience: This metric measures the specificity and clarity of described work experiences. A high vagueness score may indicate a candidate overstates their qualifications or lacks relevant experience. By analyzing language patterns, Satya helps identify potentially inflated or ambiguous claims that could misrepresent a candidate's capabilities.

2. Trust Score on Recommendations: This score evaluates the credibility and strength of the provided recommendations. By assessing the recommender's authority and the specificity of their endorsement, this metric contributes significantly to assessing the overall reliability of a candidate's profile. Higher trust scores correlate with stronger, more relevant endorsements, increasing the candidate's perceived value.

3. Recommendation Sentiments: This metric allows us to assess the feedback and insights provided in the possibly multiple recommendation letters provided by people who have worked with a candidate. For this metric, we assume that everything said is true, which is balanced out by the trust score.

4. Skill Trust Score: Through this metric, we put together the skills and experience of a candidate, seeing whether or not they make sense. Generally, candidates gain skills through experience which is an assumption to make, and nevertheless, must be flagged for deeper analysis by a recruiter.

5. Skill Relevance Score: This metric utilizes three inputs: candidate experience, candidate skills, and role expectations. We utilize large language models here as well to process and obtain a score, relying on the fact that state of the art LLMs have enough understanding of what a role is. The role expectations are defined by the HR personnel or may be obtained from the technical department of the company as per requirements. For the purpose of this problem statement, we used Few Shot Learning with instruct models to find the right skills for over 450 roles in the given dataset.

6. Recursive Endorsements: This feature detects patterns of mutual endorsements within professional networks, flagging potential biases or collusion in recommendations. The identification of reciprocal endorsements helps HR teams recognize when candidates may be engaging in practices that could distort the true picture of their professional relationships and competencies.

7. Employment Gaps: Satya penalizes candidates with significant, unexplained gaps in their employment history, which are almost certainly a red flag. The system's timeline analysis identifies such gaps, which could indicate potential issues with career progression or candidate reliability. These gaps contribute to lowering the overall candidate score.

8. Candidate Networking: We analyze the extent and quality of a candidate's professional connections to provide insights into their industry influence and reach. A solid professional network can indicate a candidate's active engagement within their field, suggesting potential fit and adaptability in various roles. Since the given dataset does not have company names available, we provide a proof of concept along with the code to analyze a network of connected individuals. Following are some parameters on which the Candidate Networking will be analysed :

    A. Degree Centrality: Measures direct connections; individuals with a high degree of centrality are influential, well-connected, and can access information quickly.
    B. Betweenness Centrality: Reflects how often someone bridges between others; high betweenness indicates control over information flow and cross-group collaboration.
    C. Closeness Centrality: Indicates proximity to others in the network; high closeness means faster access to people and resources, aiding efficient communication.

D. Clustering Coefficient: Measures how tightly-knit a person's connections are; high clustering suggests strong group relationships and potential team cohesion.

Our testing has shown that this multi-faceted approach leads to robust and insightful CV evaluations. Its utility lies in the ability to uncover hidden patterns and connections within different portions of an applicant's data.

ARCHITECTURE AND SOFTWARE DESIGN
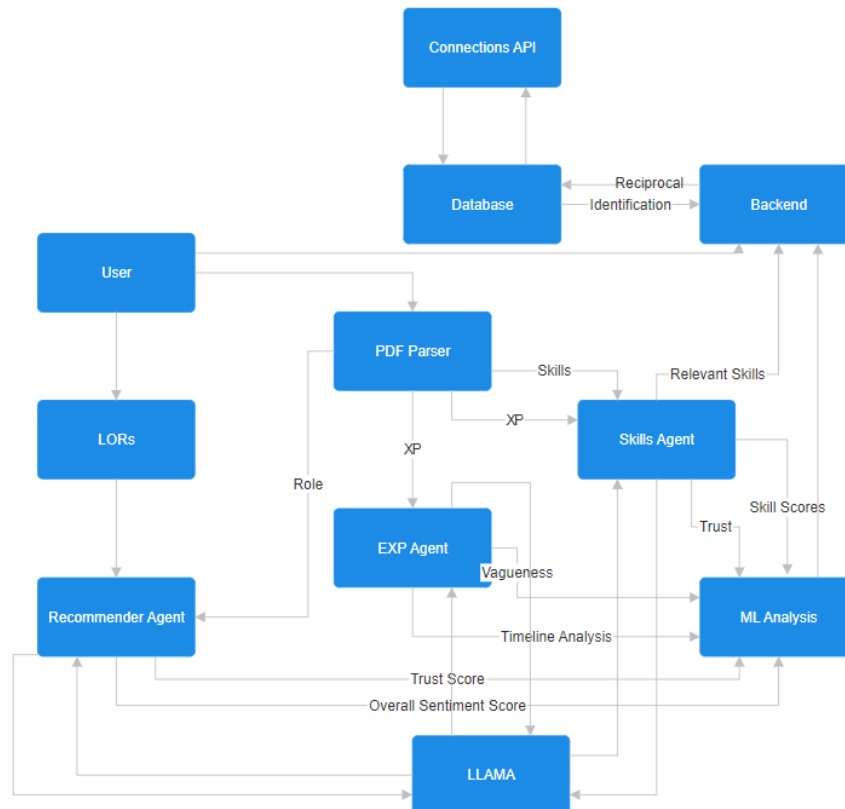
1. Aiming for a modular software design, we implemented an AI Agentic Framework. For each section of the CV, we add a different Agent for performing both candidate scoring and risk analysis. Each agent has its own positive and negative metric, as mentioned in the metrics section.
2. As with any good framework, we have divided the codebase in to three sections: Frontend, Backend, and Analysis. Analysis is done only at the time of the entry of a candidate. We have kept this separate to offload as much of the heavy lifting as possible, enabling parallel processing.
3. Since processing is a one-time expense, we maximize the use of large language models. We performed tests on various models of the LLaMa family, Mistral, and Palm. We performed some minor tests on closed source models as well, which, as expected, had much better results (we do

not share an accuracy measure, since most of it is open to interpretation. These results are true to the best of our team's understanding). This shows that while the current accuracy of the system isn't feasible at scale, one can easily use the GPT API (4o showed promising results).

4. The web development side of our framework involves three Models: Applicant, Role, and Recommendation. The attributes are trivial. We use a MERN stack for implementation of the interactive dashboard, along with a FastAPI implementation of the Analysis end, making the usage of LLaMa via HuggingFace transformers possible.

SCALABILITY AND OPTIMIZATION

Our proposed solution for Satya is designed with scalability and future adaptability. Here are the key aspects of our approach to ensuring long-term viability and performance:

1. Scalable Architecture: The system is built to handle increasing volumes of data without compromising performance. A modular design allows for easy integration of new features and analysis techniques. This scalability ensures that as organizations grow and hiring demands increase, Satya can maintain its efficiency and effectiveness.

2. Versatile Agentic Framework: The solution can be continuously updated to incorporate more sections of the applicant's CV. Currently, we look at the Experience, Skills, Recommendation Letters, Summary, and Education of the applicants. With our architecture, it is simple to add another agent and give more and more valuable insights as the system evolves.

3. Adaptability to Future Machine Learning Trends: As more powerful models come into picture, the accuracy of our system increases without any effort. Seeing the trend of tech giants fighting to lead the AI space, research in the area of LLMs is bound to continue, fueling the power of this system.

4. Time-Efficiency Trade-offs: Almost none of our processing is done on query time. This makes it faster to access data multiple times without loss in performance.

By focusing on these aspects, our solution is well-positioned to meet current HR needs and evolve and improve over time. Satya's scalability and optimization potential ensure its relevance and effectiveness in the dynamic landscape of AI-powered HR decision-making.

PROVISIONS AND FUTURE WORK

An AI-driven system for resume fraud detection and analysis is a highly comprehensive system having multiple considerations. While the system must be fair to applicants in terms of ranking them for various job applications, it has to be provided with a lot of data (such as hiring history of a company). There are a few aspects which could be covered in light of the current project:

1. Individual Networking: The idea is that we build a graph with weighted edge depicting the number of companies/organizations those individuals have had worked together, these past correlations can build a HR insight of them working as a team, and also can be looked upon as a connecting link between individuals within the professional circles, and also a bridge between company-to-

company contacts. This will help HR visualize how the career paths of the individuals evolve over time.

2. Sentence Inferences: A major portion of understanding the qualities of a candidate depends on the way they frame their sentences in the Resume. It is highly essential to correctly understand the context as well as understand the context correctly. For this, various NLP concepts can be utilized which can group semantically similar sentences based on contexts. A very basic aspect of this has been implemented in our project, where we identify verbs and relate them to usual traits, so that given a sentence, a particular trait may be inferred from it.

3. Domain-specific model training: With thousands and thousands of designations in various companies, it is highly difficult to categorize each designation. Various clustering can be made by identifying domain and role specific jobs rather than designation. Thus once the general domain is identified, the model be trained to learn the requisite skills and level of skills required to "survive" in the world of that domain. This helps to rank candidates better for a job application when the HR requests for such rankings.

CONCLUSION

Satya is an AI-driven system designed to enhance the hiring process by providing insightful, data-backed evaluations of resumes and recommendations. It utilizes advanced language models and innovative techniques like vagueness scoring, recommendation trust metrics, and professional network analysis to give HR teams a comprehensive view of each candidate's profile. With modular, scalable architecture and the potential for seamless integration of future AI advancements, Satya aims to streamline HR workflows, eliminate bias, and improve the quality of hiring decisions efficiently and accurately.

The multi-faceted approach to CV scoring and the system's ability to dig deeper into résumé data provides HR professionals with a powerful tool for making more informed hiring decisions. Our focus on scalability and optimization ensures that Satya can continue to deliver value as data volumes grow and new trends emerge in the job market.