

# CSE 472: Social Media Mining

Amey Bhilegaonkar, Saurabh Zinjad  
*Ira Fulton School of Engineering*  
*Arizona State University*  
Tempe, Arizona  
abhilega@asu.edu, szinjad@asu.edu

## I. ABSTRACT

In a world driven by digital connectivity and an ever-evolving job market, the need for tailored and efficient resume creation has become paramount. This report outlines a groundbreaking project that leverages the power of Language Model APIs, primarily OpenAI, to streamline the resume creation process. The workflow involves parsing generic job descriptions, extracting crucial information from resumes, and utilizing advanced queries to tailor resumes based on specific job requirements. The generated resumes are then formatted using Latex templates, ensuring compatibility with Applicant Tracking Systems (ATS). This innovative approach not only enhances the efficiency of resume customization but also positions the applicant for success in the competitive job landscape.

## II. INTRODUCTION

In the age of digital interconnectedness, this research delves into the realm of social media mining, with a specific focus on the contentious Canada-India controversy as a case study. It explores the intricate web of social media networks, investigating data collection techniques, network construction, and large language model (LLM) classification to decipher diverse perspectives within this digital discourse. By leveraging Mastodon's REST APIs, we gather a wealth of data, construct Friendship Networks to visualize user relationships, and employ LLMs to classify users into pro, neutral, or anti-categories. This research contributes to our understanding of online discourse, network structures, and sentiment analysis, offering valuable insights into the dynamics of contemporary digital conversations. Our code can be found here: [GitHub Link](#)

## III. RELATED WORKS

In the expansive landscape of resume creation tools and technologies, several approaches have been explored to harness the power of Language Models (LLMs) for enhanced personalization. Our exploration of related works involved an in-depth analysis of existing projects and tools to understand the prevailing trends and methodologies. Notably, we observed that many existing solutions primarily focus on bolstering the strength of a resume using LLMs, emphasizing general language proficiency and coherence. While these tools undoubtedly add a layer of sophistication to the content, they

often fall short in tailoring resumes to specific job requirements. Our project stands out by going beyond mere linguistic enhancement, diving into the intricacies of job descriptions and dynamically customizing resumes to align precisely with the expectations of potential employers.

In our market exploration, it became evident that many tools focus on predefined templates and generic keyword enhancements. They lack the adaptability required to cater to the diversity of job roles and industries. In contrast, our approach involves parsing job descriptions in a generic manner, creating a standardized understanding that is not bound by hardcoded templates. This flexibility is a key differentiator, allowing our project to cater to a wide array of job categories without the need for manual intervention or template adjustments. The ability to tailor resumes according to the unique demands of each job sets our project apart, ensuring a level of personalization that goes beyond the capabilities of many existing tools.

One significant distinction lies in our parsing technique for job descriptions. Traditional approaches often involve hardcoding specific keywords or criteria for each job category, limiting their applicability. Our project, on the other hand, utilizes a dynamic parsing technique that extracts essential information without relying on predefined rules. This not only enhances the adaptability of the model to various job contexts but also eliminates the need for constant updates or adjustments as the job market evolves. The absence of hardcoded constraints allows our project to stay agile and responsive to the ever-changing landscape of employment, providing a level of versatility that is a departure from conventional resume creation tools.

## IV. MODEL DESCRIPTION

The success of this innovative resume creation project hinges on the adept utilization of advanced language models, with a primary focus on OpenAI APIs. OpenAI, a pioneer in natural language processing, provides a suite of powerful tools that form the backbone of the project's intelligence. The primary model employed is likely GPT (Generative Pre-trained Transformer), renowned for its ability to understand context, generate coherent text, and respond intelligently to complex queries. GPT's pre-trained nature allows it to grasp the intricacies of language, making it an ideal candidate for parsing generic job descriptions. The project harnesses GPT's contextual understanding to extract key phrases, industry-specific terminology, and essential skills from a variety of job

descriptions, creating a standardized representation that serves as the foundation for resume customization.

Furthermore, the model description delves into the specifics of how the project extracts information from resumes using OpenAI. This involves employing techniques to comprehend the content, identify relevant keywords, and map these to the standardized representation derived from the parsed job descriptions. The model intelligently recognizes patterns in the resumes, understanding the applicant's skills, experiences, and qualifications. This intricate process involves leveraging OpenAI's capabilities in information extraction, ensuring that the nuances of each resume are captured accurately.

The project's model description also encompasses the dynamic aspect of query formulation with OpenAI. Queries play a pivotal role in tailoring resumes to specific job requirements. The report elucidates how the project formulates queries that dynamically adapt to the job's unique characteristics, prompting the model to prioritize certain skills or experiences. This dynamic interaction with the model ensures a personalized and finely tuned output that aligns seamlessly with the expectations of potential employers. Overall, the model description provides a comprehensive overview of how OpenAI's sophisticated language models are harnessed at each stage of the project, from parsing generic job descriptions to the intricacies of resume customization.

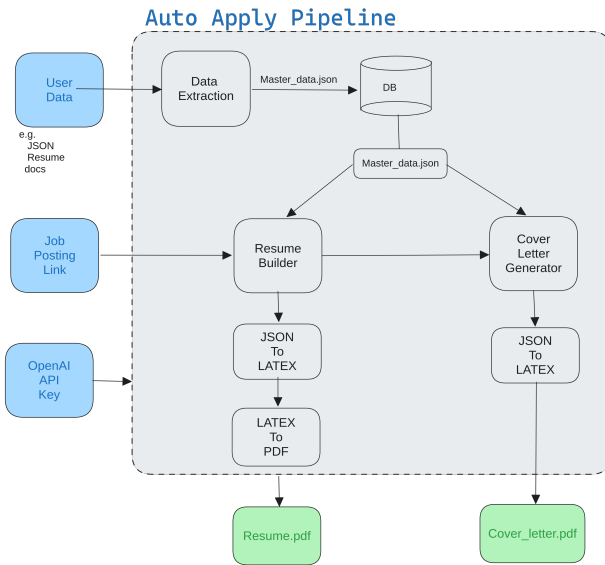


Fig. 1. AutoML Pipeline - Resume Generator

## V. EXPERIMENT

Our experimental phase was marked by a meticulous exploration of various parameters and methodologies, aimed at optimizing the efficacy of our resume customization pipeline.

- 1) **Experimenting with Different Prompts:** One key aspect of our experimentation involved fine-tuning the prompts used to engage with OpenAI's language models. We systematically tested an array of prompts to gauge

their impact on the quality and relevance of information extracted from both job descriptions and resumes. By dynamically altering the queries, we aimed to optimize the model's comprehension of context and specificity, ensuring that the tailored resumes align precisely with the expectations outlined in diverse job postings. This iterative process allowed us to identify the most effective prompts, contributing to the project's adaptability across a spectrum of job roles and industries.

- 2) **Different Techniques of Parsing Resumes:** Our experiments extended to the parsing of resumes, exploring various techniques to convert unstructured data into a structured format. We implemented different approaches, including natural language processing techniques and rule-based systems, to discern the nuances of diverse resume formats. This comprehensive exploration ensured that our pipeline could adeptly handle a variety of resume styles, accommodating variations in section organization, language usage, and formatting. The iterative testing and refinement of parsing techniques enabled us to develop a robust system capable of extracting essential information accurately and efficiently.
- 3) **Different Unified Formats - Finalized JSON at the Last:** The experimentation also delved into determining the most suitable unified format for storing parsed information before generating the final tailored resume. We considered various structures, including database schemas and intermediate formats. After a thorough evaluation, we settled on a finalized JSON format. This format proved to be versatile, facilitating seamless integration with subsequent processes such as template formatting using LaTeX. The adoption of JSON not only ensured data integrity but also streamlined the transition between different stages of the resume customization pipeline.
- 4) **Different Techniques of Parsing Job Links:** Parsing job links emerged as a critical component of our experimentation, involving the exploration of multiple techniques. We experimented with web scraping tools like BeautifulSoup and Selenium to extract relevant information from job listings. Additionally, we integrated OpenAI APIs into this process to enhance the comprehension of job descriptions. By systematically comparing these approaches, we aimed to identify the most efficient and accurate method for extracting key details from job postings. This multifaceted approach allowed us to create a parsing technique that is not only robust but also adaptable to the dynamic nature of job listings on various platforms.

These keywords and hashtags were carefully selected to cover a wide range of discussions and opinions related to the Canada-India controversy.

## VI. FUTURE WORKS

Anticipating the trajectory of our project, we are poised to introduce several features aimed at augmenting user ex-

perience and optimizing the functionality of our resume customization tool. One significant enhancement on the horizon is the incorporation of a Streamlit frontend, an interface designed for user-friendliness and interactivity. This feature will furnish users with an intuitive platform to seamlessly input job links, upload resumes, and observe the dynamic tailoring process unfold in real time. The user-centric design of the Streamlit front end aligns with our commitment to providing a visually compelling and accessible interface, ensuring that individuals, regardless of technical proficiency, can navigate the resume customization process effortlessly. This development reflects our dedication to democratizing the resume-tailoring experience, making it more widely accessible.

In parallel, our strategic roadmap includes the implementation of dockerized deployments, signifying a commitment to scalability and operational efficiency. Dockerization offers users the capability to deploy the resume customization pipeline across diverse computing environments without concerns regarding compatibility. This containerization not only simplifies deployment for individual users but also positions the project for seamless integration into larger organizational workflows. The forward-looking adoption of dockerized deployments ensures the adaptability of our project to the dynamic technological landscape.

Furthermore, we are poised to introduce an email-sending module into the project. This module will serve a dual purpose by disseminating tailored resumes directly to users and facilitating correspondence with company members affiliated with specific job postings. Users can anticipate the convenience of receiving their customized resumes and cover letters directly in their email inboxes, streamlining the application process. Simultaneously, the email module will enable the direct transmission of tailored documents to relevant company members, establishing a more immediate and personalized channel of communication between applicants and hiring teams. These upcoming features collectively underscore our commitment to a comprehensive, user-centric, and forward-looking approach in redefining the paradigm of resume customization tools.

## REFERENCES

- [1] openAI API documentation: [Link](#).