Enhancing the Efficiency and Quality of Internship Matching: An AI-Driven Approach to Student-Company Alignment

Written by Kaj Hansteen Izora
Aug 23, 2024

Introduction

The process of matching students from underrepresented communities with high-quality internships is a crucial component of fostering diversity and inclusion in the workforce. However, traditional matching methods, which rely heavily on manual processes, have proven to be time-consuming, resource-intensive, and challenging to scale. As The Script program continues to expand, serving an increasing number of students and companies, the limitations of these methods become more pronounced, necessitating a more efficient and scalable solution.

This paper explores the implementation of an AI-driven matching pipeline designed to enhance the efficiency and quality of student-company pairings. By integrating advanced techniques such as keyword extraction, alignment scoring, and iterative filtering, this approach aims to streamline the matching process while maintaining high standards of match quality. The new system not only reduces the time required for manual reviews but also improves the precision of matches, ensuring that students are paired with internships that align closely with their skills, interests, and career goals.

Through a detailed examination of the current matching process, the challenges it presents, and the improvements introduced by the AI-driven pipeline, this paper outlines the steps taken to optimize both the speed and accuracy of internship matching. It also addresses potential areas for further refinement, ensuring that as The Script grows, it can continue to meet its mission of providing meaningful opportunities to underserved college students.

Part 1: Overview of Current Matching Pipeline

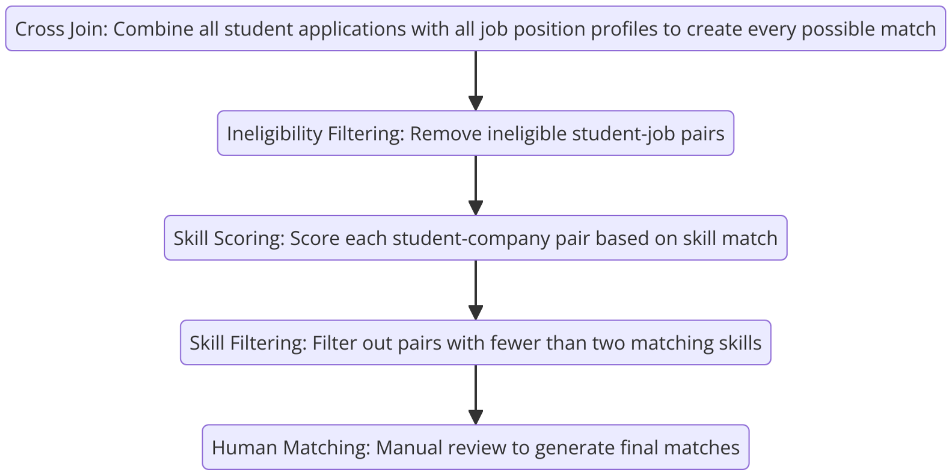
The current matching system consists of two primary components: an automated matching process and a manual review process.

1. **Automated Matching**: Matches are generated based on predefined options in the application forms. Students and companies select from provided fields, such as eligibility requirements and skills. This structured data is used to create initial matches between students and companies using a weighted scoring algorithm.

2. **Manual Review**: After the automated matching, a manual process is conducted. Human reviewers examine both the structured data (skills and requirements selected in the forms) and unstructured data

(such as student essays, resumes, job descriptions, and free text fields within the application) to refine and select the strongest matches.

A simplified overview of this matching pipeline is illustrated in the flowchart below.



Key Issues with the Current Matching Pipeline

I will focus on three primary issues with the current pipeline, ordered from most to least significant: efficiency of the matching process, user experience for companies and students filling out forms, and the quality of the matches.

1. Efficiency of the Matching Process

The most critical issue with the current pipeline is its inefficiency. Completing the full matching process—providing each company with three strong matches—can take a team of three people up to three weeks. This timeline is particularly problematic given that it involves a relatively small number of companies (around 100). Ideally, the matching process is streamlined as follows:

1. **Select a company.**
2. **Generate 15 matches** based on automated match scores.
3. **Review the job description and summary.**
4. **Read the resumes, essays, and summaries** of the matched students.
5. **Select the top three candidates** for the company.
6. **Repeat** the process for each company.

This ideal scenario already takes a significant amount of time. However, several factors can slow down this ideal process even further, leading to significant inefficiencies:

- **No Matches Found**: In some cases, the automated system fails to generate any matches for a company, often due to the company's niche requirements or insufficient skills selected during the form-filling process. This forces human matchers to:
    - Manually search for close matches.
    - Create spreadsheets to track these potential candidates.

    - Review the company's job description and student profiles.
    - Identify three eligible candidates manually.
- This manual intervention can be time-consuming and drastically slows down the process.
- **Most Students Already Matched**: As the process progresses, the pool of available students shrinks because many have already been matched to companies. In such cases, human matchers must:
    - Attempt to generate new matches but find that most students are fully matched.
    - Continuously generate more matches until eligible candidates are found.
    - Review the company's job description and student profiles.
    - Identify three eligible candidates from the remaining pool.
- This situation further slows down the process as it becomes increasingly difficult to find unmatched students who fit the company's requirements.

Improving the efficiency of the matching process is crucial for saving both time and resources. More importantly, The Script's current process cannot scale effectively as the program grows. As The Script expands to match thousands of companies with tens of thousands of students, the current labor-intensive process will become a bottleneck, limiting the program's ability to scale.
To fulfill its mission of providing high-quality internships to all underserved college students, The Script must streamline and accelerate the matching process. This will ensure that the program can grow and serve a larger number of students without being constrained by the current operational limitations.

2. User Experience for Companies and Students

The primary goal of The Script is to place as many students from underrepresented communities into high-quality internships as possible. A significant barrier to achieving this is the potential drop-off from both students and companies due to the application process. Some may not complete the form or may decide not to sign up at all because of the perceived complexity.
There needs to be a careful balance between simplifying the application process to encourage participation and collecting enough information to make effective matches. This balance is especially crucial for companies since maximizing their participation increases the number of available internship spots for students. While a more detailed application form may provide better data for matching, it can also discourage both students and companies from enrolling if the process is perceived as too burdensome.

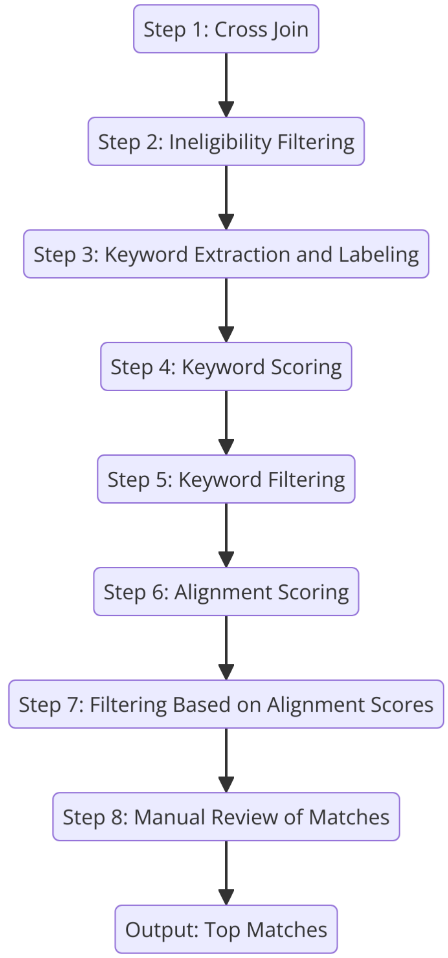
3. Quality of Matches

Although the current matching process has a high satisfaction rate—over 90% based on survey data—it is still important to address potential areas where match quality could be improved.
One area of concern is the somewhat arbitrary order in which students are matched with companies. If a student is identified as a strong match early in the process, they may be paired with a company even if they might be a better fit for another company that comes up later. Once a student has reached their maximum number of matches, they become ineligible for further consideration, even if later matches might be more suitable. This approach, while preventing overmatching, can lead to less-than-ideal pairings for both students and companies, potentially reducing the overall quality of the matches

---

Part 2: Overview of the AI Matching Pipeline

The AI matching pipeline utilizes a multi-step process to filter and score matches between students and companies. The following section provides a brief overview of the process, which includes various stages from input to output, along with a focus on how keyword extraction and alignment scoring contribute to the final matches. A simplified overview of this matching pipeline is illustrated in the flowchart below.



Input
The process begins with a cross-joined table that includes all student applications paired with all position profiles, creating a comprehensive dataset for matching.

Step 1: Cross Join
All student applications are joined with all company position profiles to create an initial dataset. This ensures that every possible student-company pair is considered for matching.

Step 2: Ineligibility Filtering
Ineligible students are filtered out based on criteria such as age, year in college, and location. This step removes any matches that do not meet the basic eligibility requirements set by the companies, reducing the dataset to relevant matches.

Step 3: Keyword Extraction and Labeling

In this stage, keywords are extracted and labeled from both student applications and company position profiles. The goal is to identify relevant keywords that can be used to match students with companies effectively. Keywords are categorized and labeled to add more detail, which enhances their utility in the matching process.

Step 4: Keyword Scoring
Using the extracted and labeled keywords, a scoring system is applied to match students and companies. The scores help determine the strength of the match based on how well the keywords align between the student and the company.

Step 5: Keyword Filtering
Keywords are used to filter out student-company pairs that do not meet a certain threshold of matching criteria. This step further refines the dataset by eliminating weaker matches.

Step 6: Alignment Scoring
An alignment score is generated for the remaining student-company matches. This score provides a more nuanced evaluation of how well the student fits the company's needs and culture.

Step 7: Filtering Based on Alignment Scores
Matches are filtered based on their alignment scores to ensure that only the top matches for each company are retained.

Step 8: Manual Review of Matches
After the pipeline outputs the top matches, a human will review top candidates, and approve 3 strong matches for each company.

Output
The final output is a set of unique top matches for every company. These matches are based on the alignment scores generated in the previous steps, ensuring that each company receives the best possible candidates from the student pool.

Part 3: Detailed Exploration of the AI Matching Pipeline

This section will go into detail on each aspect of the AI matching Pipeline. Throughout this explanation, examples will be drawn from the 2024 Script cohort, which includes:

- **103 position profiles**
- **394 student profiles**
- **40,582 initial matches after cross join**

Pipeline Stage 1: Ineligibility Filtering

Ineligibility filtering is consistent with the current matching pipeline. This process removes ineligible matches based on specific criteria such as location, age, and year in college. After applying these filters, the dataset is reduced to 27,697 rows.

Pipeline Stage 2: Keyword Extraction and Labeling

The next stage in the pipeline involves utilizing AI to extract and label keywords. The aim is to produce a formatted list of keywords that can effectively match students with companies. Keyword labeling adds context and detail, enhancing the precision of the matching process.

- **Extraction Process:** AI is used to read and analyze profiles, extracting and labeling keywords related to technical skills, education, industry of interest, soft skills, and company values/culture/mission.
- **Data Sources:** All data sources are from unstructured data
    - **Students:** Resumes, essays, and majors/minors.
    - **Companies:** Job descriptions, skill requirements/preferences, and position summaries.

<u>Keyword Extraction Instructions:</u>

An engineered prompt is provided to the AI that details what keywords should be extracted, and what labels to add to them. Here is an overview of the instructions provided to the AI:

Task: Extract and label keywords from both student and company profiles, focusing on technical skills, education, industries of interest, soft skills, and values/motivations/mission.

Extraction Criteria:

1. **Technical Skills:**
    - Identify and extract technical skills from student resumes, essays, and education summaries, as well as from company position descriptions and skill/requirements summaries.
2. **Education:**
    - Extract relevant educational background details from student profiles, such as majors, minors, and other academic qualifications.
    - For companies, include education-related keywords mentioned in job requirements, such as specific degrees or fields of study.
3. **Industries of Interest:**
    - For students: Extract and infer industry-related keywords based on the applicant's academic background, career goals, and experience.
    - For companies: Identify industry keywords from the position title, description, and requirements.
    - Consider the context to add inferred industries if not explicitly stated.
4. **Soft Skills:**
    - Identify and extract soft skills from both student and company profiles.
5. **Values, Motivations, and Mission:**
    - Extract keywords representing the values, motivations, and mission of the student or company.

Labeling Criteria:

1. **Technical Skills:**
    - **Relevance/Necessity Score:**

        - **High:** Critical requirement (for companies) or directly related to the applicant's career goals and experience (for students).
        - **Medium:** Preferred but not essential (for companies) or somewhat related, with limited experience or interest (for students).
        - **Low:** Nice to have but not necessary (for companies) or not directly related or mentioned with little context (for students).
    - **Skill Level Score:**
        - **Advanced:** High proficiency and extensive experience.
        - **Intermediate:** Moderate proficiency and applied knowledge.
        - **Beginner:** Basic knowledge or limited experience.
2. **Education:**
    - No labeling required; only extract keywords.
3. **Industries of Interest:**
    - **Industry of Interest Score:**
        - **Primary:** Directly aligned with the applicant's or company's main focus.
        - **Secondary:** Closely related, complementary fields.
        - **Tertiary:** Less related or peripheral industries.
4. **Soft Skills:**
    - **Skill Level Score:**
        - **Advanced:** High proficiency and extensive experience.
        - **Intermediate:** Moderate proficiency and some experience.
        - **Beginner:** Basic understanding or minimal experience.
5. **Values, Motivations, and Mission:**
    - No labeling required; only extract keywords.

The full extraction and labeling instructions can be found here:
Extraction and Labeling Instructions
A sample of keywords extracted across each category for students and companies can be found here:
Examples of Company Keywords
Examples of Student Keywords

Reasoning Behind Keyword Extraction and Labeling:

The inclusion of keyword extraction and labeling in the AI matching pipeline serves several critical purposes:

1. **Reducing Matches Before Alignment Scoring:** Before directly comparing student and company profiles for alignment, it is essential to reduce the number of potential matches. Keyword extraction allows us to do this efficiently by analyzing each student and company profile individually, rather than comparing every student to every company. This approach significantly reduces the number of API calls required—from approximately 28,000 to just 900—making the process more cost-effective.
2. **Enhancing Nuanced Matching:** The current automated matching process depends on skills selected from application forms, but these selections often lack the nuance required for more precise matches. A significant advantage of AI-powered keyword extraction and labeling is the

AI's ability to not only extract keywords but also label them based on the context within the document. This adds a deeper level of detail to the skills, industries, and values discussed in student and position profiles.

- **Overselection by Students:** Students frequently select a broad range of skills, some of which may not align with their actual career interests. For instance, in the current system, if a student lists both "Videography" and "Data Visualization" as skills, they would typically be given equal weight in the matching process. However, with AI, the system would analyze the student's data, recognize that the student is more focused on industries related to data visualization, and label "Videography" as having low relevance and "Data Visualization" as having high relevance. This contextual understanding enables more accurate and meaningful matches.
- **Underselection by Companies:** Conversely, companies might underselect skills, particularly if they operate in niche industries or do not thoroughly complete the application form. Keyword extraction ensures that all relevant skills are considered during matching, even if they were not explicitly selected by the company.

3. **No Need for Frequent Application Form Updates:** The dynamic nature of job markets means that relevant skills can change rapidly, requiring constant updates to the skills listed on application forms. With keyword extraction, there is no need for these frequent updates. The AI is capable of identifying and extracting new or specialized skills directly from job descriptions, ensuring that emerging fields are incorporated into the matching process without the need for manual intervention. This capability has been effectively demonstrated with sample resumes and job descriptions from cutting-edge industries such as AI Prompt Engineering, VR/AR Development, Human-Computer Interaction, and Quantum Algorithm Design, where the AI successfully extracted all pertinent keywords related to these fields.
4. **Simplifying the Application Process:** Keyword extraction allows for a simplified application process. Instead of requiring applicants and companies to select from predefined skills, they can use free text to provide detailed job descriptions and resumes. The AI will then extract the necessary keywords, making the process more user-friendly while still ensuring effective matches. Additionally, combining free text with predefined skill selections enables the AI to label and analyze both structured and unstructured data, enriching the structured data with added nuance and organizing unstructured data in a way that makes it more actionable for matching purposes.

Challenges and Areas for Improvement:

1. **Prompt Engineering:** The prompts used for keyword extraction were developed quickly and could benefit from significant refinement. The categories of keywords being extracted and the labels assigned to them should be carefully reviewed and improved. A thorough analysis of the extraction results is necessary to optimize the prompts and achieve the best possible outcomes.
2. **Source of Keyword Extraction:** Currently, the AI matching pipeline relies solely on unstructured data from position and student profiles for keyword extraction. Although this approach has yielded high-quality matches, the accuracy and relevance of the extracted keywords are closely tied to the quality of the information provided. To further enhance the effectiveness of the matching process, the application process could be improved by offering more opportunities

for companies and students to explicitly list relevant skills. This would allow the AI to extract and utilize more precise and comprehensive keywords, thereby increasing the overall quality of the matches generated.

Pipeline Stage 3: Keyword Scoring and Weighting

After keywords are extracted, a scoring mechanism similar to the current method, but with more detailed scoring criteria is used. The scoring process uses different weights for each category, reflecting the relative importance of each dimension in determining a strong match between a student and a job posting. The weighted scoring helps prioritize critical factors like technical skills and industry alignment while still considering soft skills and values as supplementary factors.

Overview of Scoring Criteria:

- **Technical Skills Matching and Scoring:**
  - The technical skills from student profiles are compared against those required by job postings.
  - **Weights:**
    - **Relevance Score:** Skills with a High relevance score are given more weight, with up to **20 points** for a High-High match.
    - **Skill Level Score:** Additional points are awarded based on the alignment of skill levels, with up to **4 extra points** for an Advanced-Advanced match.
  - The final score for technical skills is a combination of the relevance and skill level scores, with higher alignment yielding a higher total score. This process ensures that the matching focuses on the most pertinent skills.
- **Industry Matching and Scoring:**
  - Industry keywords are compared between student profiles and job postings, with the matching process heavily weighted based on the industry labels.
  - **Weights:**
    - **Primary-Primary Match: 40 points**
    - **Primary-Secondary Match: 30 points**
    - **Primary-Tertiary Match: 5 points**
  - This weighting emphasizes strong alignment in industry interests, ensuring that students are matched with job opportunities that best fit their career aspirations.
- **Soft Skills Matching and Scoring:**
  - Soft skills are compared based on their skill levels, with the scoring process emphasizing alignment between the student's soft skills and the job's requirements.
  - **Weights:**
    - **Advanced-Advanced Match: 3 points**
    - **Advanced-Intermediate Match: 2 points**
    - **Advanced-Beginner Match: 1 point**
  - While soft skills are given lower overall weight compared to technical skills and industry alignment, they are still crucial in differentiating candidates who are otherwise technically qualified.

- **Values Matching and Scoring:**
  - Values are matched through exact comparisons, with a fixed score awarded for each match.
  - **Weights:**
    - Each exact match between a student's values and a job posting's values earns **3 points**.
  - Although values carry a lower weight in the overall matching process, they contribute to ensuring that candidates align with the company's culture and mission.

Final Output
The pipeline aggregates the scores from these four categories into a total score (Extraction_total_points). This score represents the overall alignment between student profiles and job postings, weighted according to the importance of each category:

- **Technical Skills:** Heavily weighted due to their direct impact on job performance.
- **Industry Alignment:** Also heavily weighted to ensure that students are matched with jobs in relevant industries.
- **Soft Skills and Values:** Given lower weights but still factored into the total score to refine the matching process.

The resulting DataFrame, with these weighted scores, will be used in the next stage of the pipeline to select the best matches. This ensures that the most aligned candidates are paired with the most suitable opportunities, optimizing the matching process for both students and employers. By weighting the scores, the pipeline prioritizes the most critical factors, leading to higher-quality matches.

Examples of Keyword scores for Student-Company matches can be found here:
Examples of Keyword scores for Student-Company matches

Reasoning Behind Keyword Scoring:
The keyword scoring process is designed to enhance the accuracy and quality of student-company matches by leveraging labeled keywords across various categories. By assigning relevance and skill level scores to each keyword, the system prioritizes matches where a student's skills are not only aligned with a company's needs but also relevant to the student's industry of interest. This approach ensures that the highest-scoring matches are those where the student's skills are directly applicable to the job's specific requirements and aligned with their career goals. For instance, if a student lists skills that are irrelevant to their industry of interest, these skills will contribute fewer points in the matching process, ensuring that the final matches reflect true alignment between the student's aspirations and the job's needs.

Challenges and Areas for Improvement:
The current comparison and scoring scripts were developed quickly and, while functional, require further refinement. One of the main challenges lies in the use of fuzzy matching, which, while helpful in capturing related skills, can sometimes produce unintended results, such as matching unrelated skills or overemphasizing certain skills. Additionally, the weighting of specific skills needs further adjustment. For example, a student listing general skills like "Microsoft Office" may receive fewer points compared to a

student who lists more specific tools like "Microsoft SharePoint" or "Microsoft Word." This discrepancy can sometimes lead to overemphasizing less critical skills or underrepresenting more relevant ones. To improve the accuracy and effectiveness of the scoring, these weights should be adjusted based on further analysis and testing to identify what produces the best matching results.

Pipeline Stage 4: Keyword Filtering:

After scoring keywords, a filtering process is utilized to reduce the number of matches each student and company has. Because of the more detailed scoring mechanism, we are able to improve how this filtering is done, ensuring that students and companies receive both high quality matches, and are not over or under matched.

Description of the Filtering Process for Students and Companies

Overview: This stage of the pipeline focuses on filtering and selecting the best matches between students and companies based on the keyword scores generated in the previous stage. The filtering process ensures that only the most aligned student-company pairs, as indicated by the highest scores, are retained for further consideration. The goal is to match each student with up to 5 companies, ensuring that each match is meaningful and aligned with both the student's and company's needs.

Step-by-Step Filtering Process:

1. **Initial Sorting:**
    - The DataFrame containing student and company matches is sorted by company and then by the keyword alignment score (Extraction_total_points) in descending order. This sorting ensures that the top matches (those with the highest scores) are prioritized for each company.
2. **Iterative Matching Rounds:**
    - The matching process is performed in rounds. In each round, the top match for each company (the student with the highest alignment score) is selected. This ensures that the strongest matches are considered first.
3. **Processing Top Matches:**
    - For each company, the top student match is identified and collected into a list. This list represents the best possible matches for that round.
4. **Tracking and Limiting Student Matches:**
    - The code tracks the matches for each student across all rounds using a dictionary. After each round, it ensures that each student retains only their top 5 matches based on the alignment score. This prevents any student from being matched with more than 5 companies, ensuring a fair distribution of matches.
5. **Updating the Final Matches:**
    - The selected matches from each round are added to the final list of matches, ensuring that only the top 5 matches for each student are kept. This list accumulates the best matches across all rounds.
6. **Removing Fully Matched Students:**

   - Students who have reached their maximum of 5 matches are removed from the pool, meaning they are not considered in subsequent rounds. This step ensures that the remaining students have a fair chance to be matched with companies.
7. **Loop Continuation:**
   - The matching process continues in rounds until no students remain in the pool, meaning all students have either been matched to their limit or no suitable matches are left.
8. **Final Output:**
   - The final output is a DataFrame containing all the matches, with each student being matched with up to 5 companies. This filtered list of matches is then ready for further analysis or narrowing down to select the top candidates for each job posting.

This filtering process effectively narrows down the list of potential matches to ensure that only the most relevant and strongest student-company pairs are retained. By focusing on the highest keyword alignment scores and limiting each student to a maximum of 5 matches, the process ensures a balanced and optimized matching outcome that benefits both students and companies.

The 2024 Matches after keyword filtering can be found here:
2024 Matches after keyword filtering

Reasoning Behind Keyword Filtering in the Updated Matching Pipeline

The introduction of keyword filtering in the updated matching pipeline addresses several critical issues inherent in the current system, enhancing the efficiency, fairness, and quality of the matching process. This process is essential for optimizing the match between students and companies, particularly as The Script scales to serve more participants.

1. Match Reduction for Efficiency:
The keyword filtering process significantly reduces the number of potential matches, streamlining the next steps in the pipeline. For instance, in the 2024 cohort, filtering reduced the number of matches from 27,700 to just 993. This drastic reduction minimizes the workload for human matchers in subsequent stages and speeds up the overall process. By narrowing down the pool to only the most relevant candidates, the system ensures that resources are focused on the most promising matches, thereby improving the efficiency of the pipeline.

2. Fair Distribution of Matches:
The current system struggles with both overmatching and undermatching, where students and companies either receive too many or too few matches. This imbalance can result in suboptimal outcomes, with some companies finding it challenging to identify suitable candidates and some not being matched at all.
The updated filtering system mitigates these issues by enforcing a more balanced distribution of matches. By prioritizing the most relevant and aligned matches, the system ensures that every company receives a fair number of high-quality candidates.

The distribution of company and student matches for 2024 after filtering using the current pipeline vs the AI pipeline can be found in the links below

i. Distribution Of Company Matches - Current Pipeline
ii. Distribution of Student Matches - Current Pipeline
iii. Distribution Of Company Matches - Post AI Pipeline
iv. Distribution Of Company Matches - Post AI Pipeline

Looking at the tables, you can see that the AI pipeline distributes matches significantly more evenly across students and companies when compared to the current pipeline. The match distributions using the current pipeline ranges from 0 to 294 for companies and 1 to 72 for students. With the AI pipeline, it ranges from 2 to 7 for companies and 1 to 2 for students.

3. Ensuring the Best Matches:

In the current pipeline, the order in which students are matched with companies can be somewhat arbitrary, leading to less-than-ideal pairings. For example, a student might be matched early on with a company that is not the best fit simply because they appeared first in the matching sequence. This randomness can compromise the overall quality of the matches.

The updated filtering system addresses this by ensuring that each student is matched with the companies that are most aligned with their skills, interests, and career goals. By considering all potential matches for a student and selecting only the top ones, the process eliminates the arbitrary nature of the current system. This approach ensures that the best possible matches are made, improving the satisfaction rates for both students and companies and enhancing the overall quality of the internships provided.

Challenges and Areas for Improvement:

One of the main challenges in the filtering process is determining the optimal cutoff for the number of matches a student can receive. If the cutoff is set too high, the most qualified candidates may end up being matched with multiple companies, which could lead to overmatching—where a few students dominate the available opportunities. On the other hand, setting the cutoff too low can result in a more equitable distribution of candidates across companies, but it might prevent some companies from being matched with the highest-quality candidates.

Striking the right balance between ensuring that companies have access to top-tier candidates and preventing the overmatching of those candidates is critical. This balance will require ongoing analysis of the matching results and careful adjustment of the cutoff criteria to optimize outcomes for both students and companies.

Pipeline Stage 5: Alignment Scoring

Alignment scoring is a process used to determine how well a student's profile matches a job opportunity. By leveraging AI, we compare the student's resume, essays, and skills with the job description to find the most compatible pairings. This helps ensure that students are matched with roles where they can excel, and companies receive candidates well-suited to their needs.

Key Areas of Alignment:

1. **Company and Industry Alignment:**

   - This assesses whether the student's interests and career goals align with the company's industry and mission. For example, if a student is passionate about sustainability and the company focuses on environmental conservation, it would be a strong match.
2. **Job Role and Responsibilities vs. Applicant Experience:**
   - Here, we evaluate how well the student's previous job roles and experiences align with the responsibilities of the job they're applying for. High alignment occurs if the student's experience directly matches the job's requirements.
3. **Education, Technical Skills, and Tools:**
   - This part examines whether the student has the necessary education, skills, and familiarity with the tools required for the job. For instance, proficiency in a specific programming language needed by the job would result in a high score.
4. **Values, Perks, Development Opportunities, and Company Culture Alignment:**
   - This evaluates how well the student's personal values and preferred work environment match the company's culture. A strong alignment is noted if both share similar values, such as a focus on work-life balance.

Overall Alignment Score: After evaluating these four areas, the scores are averaged to produce an overall alignment score. A higher score indicates a stronger match between the student and the job.

<u>Alignment Scoring Instructions Overview:</u>

Below is an overview of the engineered prompt that is given to the AI for alignment scoring.

Task: Your task is to evaluate how well internship applicants align with specific job descriptions. The alignment will be broken down into four categories, each assessing a different aspect of the match between the student profile and the job description.

Alignment Categories:

1. **Company and Industry Alignment:**
   - **Job Description:** Evaluate the industry focus, mission, and overall goals of the company as described in the job description.
   - **Applicant:** Assess the applicant's industry interests, career aspirations, and how they align with the company's focus.
   - **Summary of Alignment:** Determine how closely the applicant's interests align with the company's industry and mission.
2. **Job Role and Responsibilities vs. Applicant Experience:**
   - **Job Description:** Identify the key responsibilities and roles outlined in the job description.
   - **Applicant:** Review the applicant's relevant past roles, work experience, and how they align with the job's responsibilities.
   - **Summary of Alignment:** Evaluate the extent to which the applicant's experience matches the responsibilities of the job.
3. **Education, Technical Skills, and Tools:**

- **Job Description:** List the required and preferred educational qualifications, technical skills, and tools needed for the job.
- **Applicant:** Highlight the applicant's education, skills, and familiarity with the required tools.
- **Summary of Alignment:** Assess how well the applicant's educational background and skills match the job's requirements.

4. **Values, Perks, Development Opportunities, and Company Culture Alignment:**
   - **Job Description:** Examine the company's values, culture, perks, and development opportunities as described in the job description.
   - **Applicant:** Evaluate the applicant's values, what they seek in a company culture, and their interest in the perks and development opportunities offered.
   - **Summary of Alignment:** Analyze how well the applicant's values and cultural preferences align with those of the company.

Alignment Score Notes:

- For each category, assign an alignment score out of 10, reflecting how well the applicant matches the job description in that particular area.
- Calculate the overall alignment score by averaging the scores from each category.

Referencing Notes:

- When referencing specific information from the student profile or job description, indicate the line number from which the information is taken.
  - For student profiles, use the format: (SP line #).
  - For job descriptions, use the format: (JD line #).
- Example: If a student's resume mentions "SQL" on line 5 (SP line 5) and the job description mentions "SQL" on line 17 (JD line 17), you should reference these line numbers when discussing the alignment.

Conclusion:

- After assessing all categories, provide a brief conclusion summarizing the overall alignment between the applicant and the job. Highlight key strengths and areas where alignment is lacking to help gauge the applicant's fit for the role.

Full Alignment Scoring instructions can be found here:
[Alignment Scoring Instructions]
The 2024 matches after Alignment Scoring can be found here:
[Matches Post Alignment Scoring]

Reasoning Behind Keyword Filtering in the Updated Matching Pipeline:

Alignment scoring is the most detailed automated analysis in the matching process. It mimics the human evaluation of student-company pairs by providing scores that can be easily sorted and filtered. Additionally, it offers summaries explaining the rationale behind the scoring, enabling quick and informed decision-making.

Challenges and Areas for Improvement:

1. Prompt Engineering:
The prompt used for scoring can be significantly refined. Enhancing the specific criteria for matching, improving the details in the summaries, and refining how the AI references information can lead to more accurate and higher-quality matches. By improving prompt engineering, we can achieve better alignment between students and companies.

2. Costs:
Since each student-company pair is sent for scoring via an API, the costs associated with these calls can add up quickly. While alignment scoring is essential, it is also the most expensive part of the matching process. For context, each comparison currently costs about 0.03 cents. If we were to skip keyword filtering and compare all students with all companies directly (around 28000 rows), the cost could escalate to around $900. To manage costs effectively, it's crucial to reduce the number of pairs sent for alignment scoring. By using keyword matching and filtering, we narrow down the pool to about 1,000 student-company pairs, reducing the cost to approximately $30. As the program grows, these costs will continue to rise, making it vital to maintain efficient filtering before alignment scoring. Although this automated process is still far more cost-effective than manual matching, the expenses should be carefully monitored.

Pipeline Stage 6: Alignment Filtering

The last stage in our matching pipeline is alignment filtering. This step takes the rows that have passed keyword filtering and applies the newly generated alignment scores to narrow down the best matches. The goal is to produce four strong matches for each position profile, similar to the final stage in the current manual matching process.

How Alignment Filtering Works:

1. **Input and Process:**
    - The input to this stage includes all the rows that have passed keyword filtering, now enhanced with alignment scores.
    - We use the same filtering algorithm applied during keyword filtering, but instead of using keyword scores, we rely on alignment scores.
    - The process is iterative, matching students with the companies they align with best, across multiple rounds. In each round, students are limited to a maximum of two matches with companies.
2. **Selecting the Top Candidates:**

- After assigning matches based on alignment scores, we focus on selecting the top four candidates for each position.
- Up to this point, no students have been eliminated from the potential pool of candidates. However, if a student does not rank in the top four for any company, they are removed from the candidate pool. For the 2024 cohort, this resulted in 60 students being removed out of 397.
- After filtering, each company will have up to four strong matches, completing the automated matching process. The final step would involve a manual review to ensure that the candidates are indeed high-quality matches.

Example of a Match Generated by the AI Matching Pipeline:

Company: Holst Architecture
Position Name: Architectural Internship
Student: Maribel Zepeda

- **Company and Industry Alignment:**
    - **Job Description:** Holst Architecture is a Portland-based design studio that prioritizes the human experience of space in building designs and strives to express high environmental, social, and aesthetic ideals (JD line 6).
    - **Applicant:** Maribel aspires to design affordable housing for low-income families/minorities and emphasizes sustainable resources in her architectural goals (SP line 4).
    - **Summary of Alignment:** Maribel's focus on affordable housing and sustainability aligns with Holst's commitment to social and environmental ideals. However, her primary interest in affordable housing may diverge slightly from Holst's broader focus on design.
    - **Alignment Score:** 7.0/10
- **Job Role and Responsibilities vs. Applicant Experience:**
    - **Job Description:** Responsibilities include collaboration with the design team, participation in design charrettes, preparation of presentation materials, and creation of working drawings (JD lines 8-11).
    - **Applicant:** Maribel has experience in 3D design, product design for retail environments, and hands-on construction involvement through her work with Eduardo's Construction (SP lines 3-4 and 7).
    - **Summary of Alignment:** Maribel's experience in design and construction tasks reflects the collaborative and creative responsibilities of the internship, though her direct experience with specific tasks (like design charrettes) may be limited.
    - **Alignment Score:** 8.0/10
- **Education, Skills, and Tools:**
    - **Job Description:** Strong candidates should be current architecture students with skills in Revit, Sketchup, Microsoft Office, and Adobe Suite (JD lines 14-16).
    - **Applicant:** Maribel is currently enrolled in an architecture program and is proficient in Revit, Sketchup, and Microsoft Office; she also has familiarity with Adobe software (SP line 5 and 6).

    - **Summary of Alignment:** Maribel meets all the educational requirements and possesses critical technical skills needed for the role, demonstrating a strong alignment.
    - **Alignment Score:** 9.0/10
- **Values, Perks, Development Opportunities, and Company Culture Alignment:**
    - **Job Description:** Holst Architecture values a supportive environment for learning, diversity, and mentorship (JD lines 1-4).
    - **Applicant:** Maribel seeks an environment that promotes continuous learning and diversity, valuing mentors and support (SP line 3).
    - **Summary of Alignment:** Maribel's emphasis on a mentor-driven, inclusive environment aligns well with Holst's company culture and values.
    - **Alignment Score:** 8.0/10
- **Overall Alignment Score:** 8.0/10
- **Conclusion:** Maribel Zepeda demonstrates a strong alignment with the internship position at Holst Architecture, particularly through her educational background and technical skills. While there is significant alignment in values and company culture, her primary focus on affordable housing may not entirely align with the company's broader mission. However, her hands-on experience and passion for sustainable and socially equitable architecture present her as a compelling candidate for the internship.

The final results of our AI pipeline can be found below. They show all matches made by the AI pipeline, the top 4 matches for each company, and the number 1 match for each company.

- Table with 2024 matches post AI pipeline
- Table with Top 4 Matches for each Company in 2024 post AI pipeline
- Table with Top Match of each Company in 2024

Reasoning Behind Alignment Filtering:

- **Speed of Matching:** Alignment filtering is the final step in a fully automated matching process, capable of providing strong matches in a fraction of the time it traditionally took. The full pipeline only took 15 minutes to run. After alignment filtering, a manual audit is conducted to ensure the top candidates are qualified. During testing, manually auditing 100 Student-Company matches across 25 companies took around 3 hours, with all matches found to be qualified candidates. If the work is spread out across 3 people, this approach would reduce the time needed to make strong matches from three weeks to less then one day.
- **Quality of Matches:** In a manual audit, all matches were considered strong, with an overall alignment score of 8 or above typically indicating a strong match. Across all position profiles in 2024, only six number-one candidates scored below an 8, suggesting that the matching process produced a qualified candidate for 94% of companies.

Challenges and Areas for Improvement:

- **Unsuccessful Matches:** While most candidates were qualified for their positions, some companies, particularly those that are harder to match, had top candidates with alignment scores ranging from 3-7. These would be considered unsuccessful matches, and a manual audit confirmed they were not ideal fits. Improving the AI matching pipeline could help reduce these instances.

- **Student Elimination:** If a student doesn't make it into the top four for any company, they are eliminated from the candidate pool. This could be unfair to students who are qualified but were not matched earlier in the pipeline. To ensure fairness, it might be better to avoid eliminating any students until after the manual audit, which would involve reviewing more candidates but ensuring no student is unfairly removed. For the 2024 cohort, this would mean performing a manual audit on 600 students across 100 companies as opposed to 400 students across 100 companies.
- **Optimizing the Filtering and Matching Algorithm:** Currently, each student is limited to two strong matches, meaning they are only paired with the companies where their alignment scores are the highest. For instance, if a student has alignment scores of 9, 8.5, and 8.3 with three different companies, they will be matched with the companies where they scored 9 and 8.5, leaving the company with the 8.3 score without that match. Allowing students to be matched with more companies could provide better candidates for more companies, but it may also result in students being overmatched. Striking the right balance between offering more matches and avoiding overmatching is crucial for improving the overall matching process.

Summary of the AI Matching Pipeline Process

Overview of Improvements:

The AI matching pipeline introduces several enhancements to the traditional student-company matching process. By utilizing a multi-step approach that includes keyword extraction, scoring, and alignment filtering, this system significantly improves the efficiency, fairness, and quality of the matches. Here's a summary of the key improvements and areas for potential enhancement:

1. Efficiency:
- Automation: The process automates what was previously a time-consuming manual task, reducing the time required to generate strong matches from up to three weeks to potentially just one day.
- Match Reduction: The pipeline reduces the number of potential matches early on, from tens of thousands down to a manageable number (e.g., from 27,700 to 993 in the 2024 cohort). This reduction helps streamline the manual review process and focuses resources on the most promising candidates.

2. Enhanced Matching Accuracy:
- Keyword Extraction and Scoring: The use of AI to extract and label keywords from both student profiles and company job descriptions adds a layer of nuance and detail to the matching process that goes beyond simple selection from predefined lists.
- Weighted Scoring: By assigning relevance and skill level scores to keywords, the system prioritizes matches that align closely with both the student's career interests and the company's needs. This helps ensure that the strongest, most relevant matches are selected.

3. Fairness and Balanced Distribution:

- Controlled Matching: The pipeline limits the number of matches a student can receive, preventing overmatching (where a few students dominate the opportunities) and ensuring a more equitable distribution of candidates across companies.
- Optimized for Scale: As The Script program grows, this automated approach will be essential for managing a larger volume of students and companies without sacrificing match quality.

4. Alignment Filtering:
- Final Match Selection: The alignment filtering process ensures that only the top candidates for each position are retained, providing companies with a curated list of strong matches.
- Speed of Matching: This step finalizes the automated process, allowing for a quick yet thorough selection of the best candidates, followed by a manual audit to verify match quality.

Areas for Improvement:

1. Prompt Engineering:
- The prompts used for keyword extraction and alignment scoring could be refined to improve the accuracy and relevance of the extracted information. Better prompts could lead to higher-quality matches and more precise alignment scores.

2. Cost Management:
- The cost of API calls for alignment scoring is a significant factor, particularly as the program scales. While the process is more cost-effective than manual matching, careful management is necessary to keep expenses under control.

3. Unsuccessful Matches:
- Some companies, especially those with niche or hard-to-match positions, still receive candidates with lower alignment scores (3-7), which may not be ideal. Improving the AI pipeline to better handle these cases could further enhance match quality.

4. Student Elimination Concerns:
- The current process eliminates students who do not rank in the top four for any company. This could potentially exclude qualified students. To address this, the pipeline could be adjusted to ensure that no students are removed until after a manual review, thus ensuring that all qualified candidates are considered.

5. Optimization of Filtering and Matching:
- The balance between limiting student matches and ensuring that companies receive strong candidates is critical. Allowing more matches per student might benefit companies but could lead to overmatching. Finding the right balance will be essential for the continued improvement of the matching process.

Note on the Complexity of the AI Pipeline

The AI matching pipeline, while promising, requires significant research and evaluation to transition from a Minimum Viable Product (MVP) to a fully integrated tool within The Script's ecosystem. This involves:

- **Maintaining Code:** It's essential to have someone who fully understands the pipeline's code to manage maintenance, updates, and bug fixes effectively.

- **Refining the AI's Capabilities:** To ensure the AI delivers high-quality matches, ongoing discussions are needed to refine the prompts, scoring criteria, and overall evaluation metrics. This ensures that the AI aligns closely with The Script's evolving needs.
- **Iterative Improvements:** Regular assessments of the pipeline's output are crucial. These evaluations will help identify areas for improvement, allowing for iterative enhancements that increase the accuracy and efficiency of the matching process.
- **Research and Development:** Continued investment in research is essential to explore new technologies, optimize processing speed, and enhance the overall quality of matches. This includes staying updated on advances in AI and machine learning to keep the tool at the forefront of innovation.

While there are initial costs and complexities involved, the AI pipeline offers significant long-term advantages over the current system:

- **Scalability:** Unlike the current manual process, the AI pipeline is designed to handle increasing volumes of data and complexity as The Script grows, without a proportional increase in costs or time.
- **Efficiency:** The pipeline can process and analyze data at a speed and scale far beyond manual capabilities, ensuring faster and more accurate matches.
- **Future-Proofing:** Investing in the development of the AI pipeline now ensures that The Script remains competitive and capable of expanding its services in the future.

Overall, the AI matching pipeline offers significant improvements in terms of efficiency, accuracy, and fairness in matching students with companies. However, there are still areas that can be optimized, particularly in prompt engineering, cost management, research and evaluation of output, and ensuring that all qualified candidates are fairly considered. Continuous refinement and monitoring will be necessary to ensure that the pipeline remains effective as The Script program continues to grow.

Part 4: Other Stuff That is Important to Think about

API Pricing Breakdown for AI Matching Pipeline

The AI matching pipeline relies on the OpenAI API for tasks like keyword extraction, labeling, and alignment scoring. Each API call incurs a cost calculated based on the number of tokens processed. Here's a detailed explanation of how these costs break down, along with an expanded explanation of the math involved.

Token Costs

- Token Definition: Tokens are the units of text that AI models use to process and generate language. Typically, about 4 characters equal 1 token.
- Cost per Token:
  - Input Tokens: $5 per million tokens.
  - Output Tokens: $15 per million tokens.

For tasks like keyword extraction and alignment scoring, the cost is approximately $0.03 per row processed. This cost is influenced by both the number of tokens sent to the model (input tokens) and the number of tokens generated by the model (output tokens).

Keyword Extraction and Labeling

Keyword extraction is divided into four categories:
1. Student Technical Skills and Education
2. Student Industry, Soft Skills, and Values
3. Company Technical Skills
4. Company Industry, Soft Skills, and Values

For each student, two API calls are made—one for technical skills and education, and another for industry, soft skills, and values. Similarly, two API calls are made for each company. The reasoning for this is because breaking down tasks performed by AI into smaller chunks results in better output. However, breaking down the task too much (e.g having an API call for every category) significantly increases price. This approach minimizes API calls while maintaining quality results.

- Formula: Cost=(2×Number of Students+2×Number of Companies)×0.03
- Example Calculation:
  - For 400 students and 100 companies:
    - Number of API Calls: (2×400)+(2×100)=800+200=1000
    - Total Cost: 1000×0.03=$30

You could further optimize results by breaking down each category into separate API calls. However, this increases the number of calls, thereby increasing the cost:

- 5 student categories: Technical skills, Education, Soft skills, Values, and Industry of interest
- 4 company categories: Technical skills, Soft skills, Values, and Industry
- Expanded Formula: Cost=(5×Number of Students+4×Number of Companies)×0.03
- Example Calculation:
  - For 400 students and 100 companies:
    - Number of API Calls: (5×400)+(4×100)=2000+400=2400
    - Total Cost: 2400×0.03=$72

Model Choice:
- GPT-4o vs. GPT-4o-mini: While the GPT-4o-mini model is 33x cheaper, it is less effective at keyword extraction. The balance between cost and output quality should be evaluated, but sticking with GPT-4o for keyword extraction is recommended.

Alignment Scoring

Process Overview: The alignment scoring process calculates how well student-company matches align, based on extracted keywords and other factors. The cost is proportional to the number of rows being processed.

Cost Calculation:

- Formula: Cost=Number of Matches×0.03
- Example Calculation:
  - For 1000 matches:
    - Total Cost: 1000×0.03=$30

Optimization Strategy:

- Without filtering, processing all possible matches (e.g., 28,000 rows) would be prohibitively expensive, costing about $900. Keyword filtering can significantly reduce the number of rows needing alignment scoring, lowering costs. For instance, filtering 2024 matches reduced 28,000 matches to 900, dropping the cost to $27.

Model Choice:

- GPT-4o-mini for Alignment: The GPT-4o-mini model performs relatively well in alignment scoring. Although results from GPT-4o are more nuanced, the cost difference is significant. Using GPT-4o-mini could reduce the price from $30 per 1000 rows to less than $1, making it a viable option for larger-scale processing. I suggest evaluating the outputs from both models. If GPT-4o-mini proves effective for alignment scoring, you could process more rows without incurring significant costs, potentially improving the overall results. For instance, processing all matches for 2024 could cost only $30 using GPT-4o-mini, allowing you to rely less on keyword filtering.

Additional Note on Pricing: Batch Queuing:

Utilizing batch queuing for API requests can significantly reduce costs by allowing large batches of tokens to be processed at a discounted rate. For instance, you can submit a batch of 2,000,000 tokens to be queued and processed within 24 hours. The advantage of this approach is that the cost per API request is reduced by half when using batch queuing. This makes it an ideal option for processes like The Script internships, where all matches are made after the application process is complete. By opting for batch queuing, the organization can achieve substantial cost savings while maintaining the quality and efficiency of the matching process. For products like the associate module, using batch queuing may not be feasible because it requires real-time matching

---

Data Security with AI and API Calls

When integrating AI into the matching pipeline, data security becomes a crucial concern. The use of APIs to send and receive potentially sensitive data, such as student profiles and job descriptions, introduces several security risks that must be carefully managed. This section discusses possible security vulnerabilities, such as prompt injection attacks, and outlines mitigation techniques to protect the integrity and confidentiality of the data.

Potential Security Risks

1. **Prompt Injection Attacks:**
   - A prompt injection attack occurs when malicious input is crafted in such a way that it manipulates the behavior of the AI model. For example, an attacker might introduce special characters or misleading text into a job description or student profile to alter the AI's response or extract unintended information.
   - If successful, such an attack could lead to incorrect matchings, exposure of sensitive information, or manipulation of the data for malicious purposes.
   - *Example:* An attacker could embed commands in a job description that instruct the AI to ignore certain matching criteria, leading to unfair advantages or skewed results.
2. **Data Integrity Risks:**
   - Data integrity involves ensuring that the data remains accurate and consistent throughout the process. During multiple API calls, there's a risk that data could be altered either intentionally or accidentally.
   - If data is tampered with, it could lead to incorrect alignments and matches, thereby compromising the quality of the outcomes.
   - *Example:* An error or attack that alters the alignment scores between a student and a company could result in suboptimal or completely mismatched pairs.

Mitigation Techniques

1. **Isolating API Calls on Separate Threads:**
   - One effective mitigation technique is to isolate each API call on a separate thread, even if it increases operational costs. Doing so makes it impossible for any prompt injection or error to affect other API calls. The current implementation keeps all API calls on separate threads, eliminating any risk of prompt injection attacks. This approach enhances the security of the system by reducing the risk of cross-contamination between different data sets. If an attacker manages to exploit one API call, they won't be able to access or affect other data processed in parallel.
   - *Drawback:* The main drawback is the increased cost and complexity of the system, as running multiple isolated threads requires more computational resources.
2. **Input Validation and Sanitization:**
   - To prevent prompt injection attacks, it is crucial to thoroughly validate and sanitize all input data before it is processed by the AI model. This process involves removing or escaping special characters and potentially malicious code or prompts that could be used to manipulate the model's behavior. Additionally, an extra layer of AI can be employed to validate inputs, ensuring that no prompt injections are present in the data sent to the API. This proactive approach helps maintain the integrity and security of the AI-driven matching process.
3. **Human Validation of Matches**
   - The AI matching pipeline is designed to enhance, not replace, human involvement in the matching process. Its primary purpose is to significantly accelerate the identification of strong matches between candidates and job positions. However, human validation

remains crucial. After the AI generates matches, a human reviewer should audit each top match by carefully reviewing all resumes, essays, and job descriptions to ensure that the AI's selections are accurate and aligned with the intended criteria. This collaborative approach combines the efficiency of AI with the critical judgment of human expertise, ensuring the highest quality matches.

Handling Rate Limits: Challenges and Solutions

When working with AI models via API calls, particularly for large-scale operations like keyword extraction, handling rate limits effectively is crucial to ensure smooth and uninterrupted processing. Rate limits are imposed by API providers, such as OpenAI, to prevent abuse and ensure fair usage of their resources. These limits typically restrict the number of requests per minute (RPM) and the number of tokens that can be processed per minute (tokens per minute, or TPM).

Different Rate Limits by API Tier and Model

OpenAI offers different API tiers, each with its own rate limits. The API tiers are linked to the amount paid and time since the first payment:

- **Free Tier:**
  - Available to users in allowed geographies.
  - $100/month usage cap.
- **Tier 1:**
  - $5 paid and $100/month usage cap.
- **Tier 2:**
  - $50 paid and 7+ days since the first successful payment.
  - $500/month usage cap.
- **Tier 3:**
  - $100 paid and 7+ days since the first successful payment.
  - $1,000/month usage cap.
- **Tier 4:**
  - $250 paid and 14+ days since the first successful payment.
  - $5,000/month usage cap.
- **Tier 5:**
  - $1,000 paid and 30+ days since the first successful payment.
  - $50,000/month usage cap.

Additionally, the model being used, such as gpt-4o-mini or gpt-4o, influences these limits. Below is a summary of typical rate limits encountered:

GPT-4o-mini:

- **Tier 1:**
  - RPM (Requests Per Minute): 500
  - TPM (Tokens Per Minute): 200,000

    - Batch Queue Limit: 2,000,000
- **Tier 2:**
    - RPM: 5,000
    - TPM: 2,000,000
    - Batch Queue Limit: 20,000,000
- **Tier 3:**
    - RPM: 5,000
    - TPM: 4,000,000
    - Batch Queue Limit: 40,000,000
- **Tier 4:**
    - RPM: 10,000
    - TPM: 10,000,000
    - Batch Queue Limit: 1,000,000,000
- **Tier 5:**
    - RPM: 30,000
    - TPM: 150,000,000
    - Batch Queue Limit: 15,000,000,000

GPT-4o:

- **Tier 1:**
    - RPM: 500
    - TPM: 30,000
    - Batch Queue Limit: 90,000
- **Tier 2:**
    - RPM: 5,000
    - TPM: 450,000
    - Batch Queue Limit: 1,350,000
- **Tier 3:**
    - RPM: 5,000
    - TPM: 800,000
    - Batch Queue Limit: 50,000,000
- **Tier 4:**
    - RPM: 10,000
    - TPM: 2,000,000
    - Batch Queue Limit: 200,000,000
- **Tier 5:**
    - RPM: 10,000
    - TPM: 30,000,000
    - Batch Queue Limit: 5,000,000,000

Challenges Faced with Rate Limit Errors

In practice, managing rate limits involves several challenges:

1. **Unexpected Rate Limit Errors:**
   - Despite configuring the system according to the documented rate limits, I encountered rate limit errors more frequently than anticipated. For instance, even at a higher tier, rate limiting at 200 RPM was encountered, which was lower than the expected limit. This could be due to how requests are sent in bursts rather than being evenly distributed, leading to temporary spikes in request rates, though even after addressing this, I continued to get rate limit errors. Further investigation on improving throughput for API calls may be necessary.
2. **Handling Rate Limit Exceeded Errors:**
   - Rate limit exceeded errors require handling within the code to ensure the process can continue without crashing. This involves implementing retry mechanisms that wait for a specified period before attempting the request again. For example, after encountering a rate limit error, the system can wait for 5 to 10 seconds before retrying, allowing the request rate to fall back within acceptable limits.
3. **Impact on Processing Time:**
   - When rate limits are hit frequently, it can significantly increase the total processing time.

<u>**Mitigation Strategies**</u>

To mitigate these issues and ensure smooth processing, several strategies were implemented:

1. **Semaphore for Request and Token Limits:**
   - Semaphores are used to control the number of concurrent API calls (rate_limit) and the number of tokens processed (token_limit). These semaphores are periodically replenished to maintain the balance between requests and avoid exceeding the rate limits.
2. **Thread Management:**
   - Threads are used to handle multiple requests concurrently while respecting the rate limits. A separate thread monitors the number of requests and tokens, replenishing the semaphores as needed to ensure continuous processing without hitting the rate limits.
3. **Retry Mechanisms:**
   - When a rate limit is exceeded, the code includes a retry mechanism that waits for a few seconds before trying again. This helps prevent the process from failing and allows it to recover gracefully from temporary rate limit exceedances.
4. **Load Distribution:**
   - By distributing the load across multiple threads and using a batch processing approach, the system can handle large-scale operations efficiently, minimizing the risk of hitting rate limits.

Currently, the code can process around 150 rows per minute when operating at Tier 2 or higher, which allows for the processing of approximately 2,000 rows for keyword extraction and alignment scoring in under 15 minutes. This was the number of rows required for processing during the 2024 matching process. However, as the scale of The Script product expands, this rate of processing may need to be revisited. While Tier 2 rate limits theoretically support processing up to 1,000 rows per minute, increasing throughput beyond 150 rows per minute proved challenging. For the time being, this rate suffices given the current volume of company-student matches. However, if the participant numbers scale to thousands or tens of thousands, further optimization of throughput will be essential to meet the increased demand efficiently.

The Future of AI and Machine Learning Driven Matching

Investing time and resources into refining and integrating the AI matching pipeline within The Script product has the potential to unlock significant advancements in how AI and machine learning are utilized in the organization. Once the pipeline's results are validated, it could pave the way for training machine learning models that predict student-company fit with greater speed and accuracy than the current system. When I first began this project, challenges in data collection prevented clear insights into why certain matches were strong, making it difficult for machine learning models to replicate or enhance the process. However, validated data from the AI pipeline could allow for the generation of synthetic data from past matches, leading to a robust dataset capable of training more sophisticated models.

The journey to develop an MVP (Minimum Viable Product) for an AI-based matching algorithm was filled with challenges—technical, systemic, and ethical. It involved significant trial and error, with 60-70% of my time spent in a 'Research and Development' phase before finding a viable path forward. This experience highlights an important lesson for The Contingent: integrating AI into organizational products requires substantial resources dedicated to exploration and experimentation.

It's crucial to recognize that not every idea will work as expected. Some paths will prove unfeasible due to data structure issues, practical limitations, or ethical concerns. The key takeaway is to embrace a mindset of rapid experimentation and iteration. If an approach fails, it's essential to pivot quickly and explore other options. The landscape of AI integration is still in its infancy, with no established blueprint for success. Therefore, you must be prepared as an organization for a process that involves navigating numerous dead ends before discovering what works best. The quickly evolving nature of how AI will be integrated into society means that utilizing it to create effective solutions will require persistence, adaptability, and a willingness to learn from failures. By continuing to explore and refine AI-driven processes, The Script and The Contingent can position themselves at the forefront of social innovation, ultimately achieving their mission of creating scalable social good.