

# API DOCUMENTATION

## API ENDPOINT REPORT

### Introduction

This report provides a detailed analysis of the API endpoints implemented in the FastAPI-based application. The application processes user queries based on extracted text from resumes available in CSV and PDF formats. It then uses embeddings and a pre-trained language model to generate a comprehensive answer to the user's question.

### Endpoint Overview

The application consists of a single API endpoint that enables users to interact with the system:

| ATTRIBUTE     | DESCRIPTION   |
|---------------|---|
| HTTP METHOD   | POST  |
| ENDPOINT PATH | /query/   |
| PURPOSE       | Processes user queries and provides detailed answers based on information extracted from resumes. |
| INPUT FORMAT  | JSON  |
| OUTPUT FORMAT | JSON  |

### Request Details

#### Request Payload

The API expects a JSON object with the following structure:

| FIELD    | TYPE   | DESCRIPTION   |
|----------|--------|---|
| question | string | Represents the question that the user wants to be answered based on the resumes' content. |

## Response Details

The API returns a JSON object containing the generated answer. The structure is as follows:

| FIELD    | TYPE   | DESCRIPTION  |
|----------|--------|--|
| response | string | The generated answer based on relevant resume content. |

## Functionality

### How the Endpoint Works

#### 1. User Input Handling:

- The user sends a POST request to the /query/ endpoint with a JSON payload containing their question.

#### 2. Processing the Question:

- The handle\_query function receives the request and extracts the question.
- The extracted question is passed to the handle\_user\_input function, which utilizes various techniques to find relevant information from the resumes, including:
  - **FAISS Index:** Uses the pre-trained HuggingFaceEmbeddings model to search for the most similar texts within the extracted resume content.
  - **BM25 Algorithm:** Applies a BM25 ranking mechanism to retrieve the most relevant resume texts based on the user's query.

#### 3. Generating the Response:

- The generate\_response\_from\_inference\_api function integrates with the Hugging Face Inference API to generate a comprehensive answer based on the relevant resume content and user query.

### Integration with External API

- The API utilizes Hugging Face's Inference API (mistralai/Mistral-Nemo-Instruct-2407 model) to generate responses based on the preprocessed data from resumes.
- An environment variable (HUGGINGFACEHUB\_API\_TOKEN) is used to manage authentication with the Hugging Face Inference API.

## Error Handling and Edge Cases

- **Validation:** The endpoint relies on FastAPI's built-in validation to ensure that the question field is correctly provided as a string.

- **Missing Data:** If no relevant text is found for a given question, the API may still attempt to generate a response, but it might be less accurate.
- **External API Issues:** If there's a problem with the Hugging Face Inference API (e.g., invalid API token or connectivity issues), a generic error message will be returned, indicating that the response generation failed.

## Conclusion

The /query/ API endpoint is a robust and versatile feature of this application, capable of processing complex user questions and generating detailed answers based on extracted resume data. By integrating advanced text retrieval methods and leveraging a powerful language model, this endpoint serves as an efficient tool for querying extensive resume databases.

### Summary of the Endpoint:

- **Endpoint:** /query/
- **Method:** POST
- **Input:** JSON with question field.
- **Output:** JSON with response field containing the generated answer.
- **Technology Stack:** FastAPI, FAISS, BM25, Hugging Face Inference API

## SAMPLE QUERIES

Google Colab link for the queries:

[https://colab.research.google.com/drive/14Hrh-RyMpIN8\\_IcYfJMDvjvpzdScgW6v?usp=sharing](https://colab.research.google.com/drive/14Hrh-RyMpIN8_IcYfJMDvjvpzdScgW6v?usp=sharing)

**Query 1:** Summarise the first HR Resume

**Response :**

```
You: summarise the first hr resume
Bot: The first HR resume provided is from a candidate with the title "HR Specialist Horizontal Engineer". Here's a summary of their resume:

**HR Specialist Horizontal Engineer**
- **Professional Summary:** Passionate HR specialist with over 10 years of extensive experience in completing reference checks, preliminary phone interviews, job a
- **Work History:**
  - **HR Specialist Horizontal Engineer** (09/2007 - Present, Company Name)
    - Carried out all recruitment processes, including headhunting, reaching potential candidates, resume screening, interviewing, background checks, and closing v
    - Created, organized, updated, and maintained personnel files and the HR database.
    - Managed and coordinated employee training and development programs.
    - Provided administrative support for all HR functions, including filing, copying, general correspondence, and meeting preparations.
- **Education:**
  - Master of Business Administration (MBA) in Human Resources (2011, University Name)
  - Bachelor of Science (BSc) in Biotechnology (2009, Mount Carmel College)
- **Skills:** Word, Excel, PowerPoint, Outlook, oral and written communication, employee relations, project management, budget management, purchasing, time manager
- **Certifications:** Human Resource Business Partner, Employee Relations, Workforce Planning and Analytics, Training and Development (2011, Present University).
```

## Query 2: What is the Technical Stack of a random candidate?

### Response :

```
You: what are the technical stack of a random candidates
Bot: Based on the provided resumes, here is a random candidate's technical stack:

**Java Developer** (from resume 1)

- Java, JEE, Web Services, JSP, Servlet, Struts, JDBC, JMS, MDB, JMX, Spring (IoC, Spring MVC, Spring Security, Spring Remoting, Spring AMQP), Hibernate/MyBatis, M
```

## Query 3: Which programming languages are commonly known by candidates for software engineering roles?

### Response :

```
You: Which programming languages are commonly known by candidates for software engineering roles?
Bot: Based on the provided resumes, here are the programming languages commonly known by candidates for software engineering roles:

1. **Java**: This is the most commonly mentioned programming language, with 7 out of 8 candidates listing it as part of their technical stack. It is used for ent
2. **Python**: 4 out of 8 candidates mentioned Python, which is widely used for data analysis, machine learning, web development (Django, Flask), and automation
3. **JavaScript/TypeScript**: 3 out of 8 candidates mentioned JavaScript or TypeScript, which are essential for front-end web development and some back-end devel
4. **C/C++**: 2 out of 8 candidates mentioned C or C++, which are often used for system programming, game development, and embedded systems.
5. **SQL**: 5 out of 8 candidates mentioned SQL, which is crucial for database management and working with relational data.
6. **HTML/CSS**: 3 out of 8 candidates mentioned HTML or CSS, which are essential for web development and designing user interfaces.
7. **Other languages**: Some candidates also mentioned knowledge of other languages such as:
   - **Go**: 1 candidate
   - **Ruby**: 1 candidate
   - **Swift**: 1 candidate
   - **PHP**: 1 candidate
   - **R**: 1 candidate (for data analysis)
   - **Matlab**: 1 candidate (for scientific computing)
   - **Verilog**: 1 candidate (for hardware description language)
   - **Assembly**: 1 candidate (for low-level programming)
```

## Query 4: Which data visualization tools are candidates familiar with?

### Response :

```
You: Which data visualization tools are candidates familiar with?
Bot: Based on the provided resumes, here are the data visualization tools that candidates are familiar with:

1. **Microsoft Excel**: This is the most commonly mentioned tool, with 6 out of 8 candidates listing it. Excel is widely used for creating charts, graphs, and o
2. **Power BI**: 3 out of 8 candidates mentioned Power BI, which is a powerful data visualization tool by Microsoft that provides interactive visualizations and
3. **Tableau**: 2 out of 8 candidates mentioned Tableau, a popular data visualization tool that helps users to explore, analyze, and present data with interacti
4. **Qlikview**: 1 out of 8 candidates mentioned Qlikview, a data visualization tool that enables users to create interactive dashboards and reports to explore
5. **Minitab**: 1 out of 8 candidates mentioned Minitab, a statistical software suite that includes data visualization capabilities for creating charts, graphs,
6. **Matplotlib/Seaborn**: 1 out of 8 candidates mentioned Matplotlib or Seaborn, which are popular data visualization libraries in Python for creating static,
7. **D3.js**: 1 out of 8 candidates mentioned D3.js, a JavaScript library for data visualization in the web browser. D3.js helps users create dynamic and intera
8. **Power Query**: 1 out of 8 candidates mentioned Power Query, a powerful data connectivity and transformation add-in for Excel that enables users to extract,
9. **Power Pivot**: 1 out of 8 candidates mentioned Power Pivot, a data analysis and visualization tool that enables users to create interactive pivot tables an
10. **Google Data Studio**: 1 out of 8 candidates mentioned Google Data Studio, a web-based data visualization tool that enables users to create interactive das

These tools cover a range of data visualization options, from spreadsheet-based tools like Excel to specialized data visualization software like Tableau and Pow
```

## Query 5: How many years of experience do candidates usually have in machine learning?

### Response :

```
You: How many years of experience do candidates usually have in machine learning?
Bot: Based on the provided resumes, the experience level of candidates in machine learning varies widely. Here's a breakdown of the experience levels mentioned by t

1. **Less than 1 year of experience**:
   - 1 candidate (e.g., "Junior Data Scientist" with 0-1 years of experience in machine learning)

2. **1-3 years of experience**:
   - 3 candidates (e.g., "Data Scientist" with 1-3 years of experience in machine learning, "Machine Learning Engineer" with 2 years of experience, and "AI Speci

3. **4-6 years of experience**:
   - 2 candidates (e.g., "Senior Data Scientist" with 5 years of experience in machine learning and "Machine Learning Engineer" with 6 years of experience)

4. **7-10 years of experience**:
   - 1 candidate (e.g., "Principal Data Scientist" with 10 years of experience in machine learning)

5. **More than 10 years of experience**:
   - 1 candidate (e.g., "Director of Data Science" with 15 years of experience in machine learning)

The majority of candidates (5 out of 8) have 3 years or less of experience in machine learning, with only 1 candidate having more than 10 years of experience. Th
```