

## Introduction

This report discusses the Python script designed to fetch paginated data from an API, match response texts to their sources using the **difflib** library, and display the results using Streamlit. The script includes logging for error handling and rate limiting to prevent hitting API rate limits.

## Script Breakdown

The script is organized into several functions, each with a specific responsibility:

### 1. **fetch\_page\_data** Function

- Purpose: Fetches data from a specific page of the API.
- Input: **session** (requests session), **base\_url** (API endpoint), **page** (page number).
- Output: Returns a list of data items from the specified page or an empty list if an error occurs

### 2. **fetch\_all\_data** Function

- Purpose: Fetches all paginated data from the API by iterating through pages until no data is returned.
- Output: Returns a list containing all the fetched data.

### 3. **match\_response\_to\_sources** Function

- Purpose: Matches a response text to its sources using **difflib** to find similar contexts.
- Input: **response\_text** (response text to match), **sources** (list of source items).
- Output: Returns a list of citations where each citation is a dictionary with the source ID and link.

### 4. **process\_item** Function

- Purpose: Processes a single data item by matching the response to its sources.
- Input: **item** (data item containing response and sources).
- Output: Returns a dictionary with matched citations.

## 5. `process_data` Function

- Purpose: Processes a list of data items.
- Input: **data** (list of data items).
- Output: Returns a list of processed data items.

## 6. `Main` Function

- Purpose: Main function to fetch, process data and display results using Streamlit.
- Workflow: Fetches data when a button is clicked, processes the data, and displays the results.

## 7. Script Execution

- The script is executed by calling the **main** function if the script is run as the main module.

## Conclusion

This script efficiently fetches and processes data from a paginated API. The use of logging ensures that any errors are reported, and the rate limiting prevents API overuse. The **difflib** library provides a simple yet effective way to match response texts to their sources. The integration with Streamlit makes the process interactive and user-friendly, allowing users to fetch and view data with a button click.