

README.md

Ramavath Simhadri Naik

May 2025

1 Introduction

This project provides a REST API to scrape public GitHub profile data for a given username, summarize the developer's profile using OpenAI, and return structured data.

2 Tech Stack Used

- **Node.js** – JavaScript runtime environment
- **Express.js** – Web framework
- **Axios** – For making HTTP requests
- **Cheerio** – For HTML scraping
- **OpenAI SDK** – For GPT-based summarization
- **dotenv** – For managing environment variables
- **Nodemon** (dev only) – Auto-restart for development

3 Assumptions Made

- GitHub scraping is limited to public profile data (e.g., bio, repositories, followers, etc.).
- OpenAI summarization assumes the prompt formatting remains consistent.
- Usernames provided via the API are valid GitHub usernames.
- No pagination or rate-limiting handling for large GitHub profiles.
- The environment contains a valid `OPENAI_API_KEY`.

4 Setup Instructions

1. Clone the Repository

```
git clone <your-repo-url>
cd github-user-scraper
```

2. Install Dependencies

```
npm install
```

3. Add Environment Variables

Create a `.env` file in the root directory:

```
OPENAI_API_KEY=your_openai_api_key_here
PORT=3000
```

4. Run the Project

For development (with live-reload):

```
npm run dev
```

For production:

```
npm start
```

5 Running Locally and Hitting Endpoints

Base URL

`http://localhost:3000`

POST `/scrape/:username`

Scrapes and summarizes a GitHub user's profile.

Sample cURL Request

```
curl -X POST http://localhost:3000/scrape/octocat
```

Sample Postman Request

- Method: POST
- URL: `http://localhost:3000/scrape/octocat`
- Headers: None required
- Body: Empty

Sample Response

```
{
  "username": "octocat",
  "bio": "GitHub's official test account.",
  "publicRepos": ["repo1", "repo2", "..."],
  "summary": "Octocat is a GitHub test user with contributions in..."
}
```

6 License

This project is licensed under the ISC License.