## **Reddit User Persona Generator — Project Summary and Workflow**

This project is designed to automatically generate detailed user personas based on the public Reddit activity of any given Reddit username. The tool scrapes a user’s posts and comments and feeds this content into a local Large Language Model (LLM), which then constructs a comprehensive profile including demographics, behavioral patterns, motivations, frustrations, and more. The final persona is saved as a plain .txt file and includes citations for transparency.

### **1. Objective**

The main objective of this script is to:

* Accept a Reddit username as input
* Scrape public posts and comments
* Analyze the content using a locally hosted language model
* Generate a rich, human-like persona based on behavioral cues
* Save the output as a citation-backed persona in a .txt file

This approach mimics how UX researchers might profile users for research, but it automates the entire pipeline using modern NLP capabilities.

### **2. Tools and Technologies Used**

* **Python** – The primary programming language used for automation and orchestration
* **PRAW** – A Python Reddit API Wrapper to access Reddit data
* **Ollama** – A framework to run large language models locally
* **Local LLM (e.g., Phi-2 or LLaMA3)** – Used to analyze the content and generate personas
* **dotenv** – To manage API keys securely
* **Requests** – To interact with the local LLM server

**3. Workflow Breakdown**

#### **Step 1: Setup and Initialization**

* API credentials for Reddit (client ID, secret, user agent) are stored securely in a .env file.
* The script uses the PRAW library to authenticate with Reddit using these credentials.
* A local Ollama server must be running with the desired LLM model (e.g., phi, llama3) already pulled and available.

#### **Step 2: Fetching Reddit Activity**

* The user specifies a Reddit username (e.g., kojied).
* The script fetches the latest 50 posts and 50 comments from the user’s public profile.
* Each post and comment is formatted with title, content, and URL.

#### **Step 3: Generating Persona**

* The Reddit activity is bundled into a single prompt.
* The prompt is designed to simulate the thought process of a UX researcher and is sent to the local LLM using a POST request via the Ollama API.
* The LLM returns a structured user persona based on language patterns, topics of interest, tone, and self-expression found in the Reddit content.

#### **Step 4: Saving Output**

* The generated persona is saved into a .txt file under the sample\_users/ directory, named after the Reddit username.
* Each trait in the persona is supported by a Reddit URL, ensuring transparency and traceability of the inference.

### **4. Features and Considerations**

* **Offline Capabilities**: The script works entirely offline after model and Reddit data are fetched, making it suitable for privacy-focused environments.
* **Explainability**: By citing source Reddit URLs for each personality trait, the output is verifiable.
* **Scalability**: The design is modular and can be extended to analyze multiple users in batch mode or integrate into web-based tools.

### **5. Ethical Considerations**

This tool operates solely on public Reddit data and is intended for educational and research use. It does not attempt to de-anonymize users or make any definitive claims about real-world identity. All personas generated are synthetic and probabilistic in nature.

### **6. Future Scope**

While this project focuses on a command-line interface, it has potential to be deployed as a public-facing web app where users can input a Reddit username and receive an instant persona profile. Authentication, caching, and user consent layers can also be integrated for a more production-ready experience.

### **7. Sample Users:**

<https://www.reddit.com/user/kojied/>

<https://www.reddit.com/user/Hungry-Move-6603/comments/>