Reddit Persona Gen

Reddit Persona Generator using LLMs

1) **praw==7.7.1** - Reddit API Wrapper - Used to **fetch posts and comments** of a Reddit user. Used in: `fetch_user_content()`

2) `google-generativeai==0.5.4` - Gemini SDK - Connects to **Gemini LLMs** via API. - i don't need `transformers` here if only using Gemini,

3) **spacy==3.7.4** - NLP pipeline - Like NLTK but more modern. Used for **entity recognition**, sentence segmentation, etc. Can help extract age, location, etc., from Reddit text

4) scikit-learn==1.5.0 - ML utilities - Includes clustering, TF-IDF, etc. Could be used for **text classification** or clustering Reddit users in bulk.

5) python-dotenv==1.0.1 - Load .env files - Loads API keys from `.env` files into `os.environ`. Currently replaced by manual credentials, but can be restored easily.

6) transformers==4.41.2 - Hugging Face model hub - like **FLAN-T5** locally. Needed if i want to use a **non-Gemini LLM**.

7) torch==2.3.0 - Required by `transformers` to run models locally. Handles tensor operations, model inference, etc

8) streamlit==1.35.0 - Used to build the web UI

LLM Comparison - FlanT5 vs Gemini flash 1.5

FLAN-T5 Output:

Analysis:

- The output from `google/flan-t5-base` was **minimal and unclear**.

- It lacked structure, context understanding, and failed to generate a persona.

- FLAN-T5, while lightweight and usable offline, is **not optimized** for complex tasks like **persona generation** without significant fine-tuning or prompt engineering.

- It's better suited for simpler tasks like translation, summarization, or classification.

Gemini Flash 1.5 Output Example:

Output:
A detailed persona with attributes:

- **Username**

- **Age**

- **Occupation**

- **Location**

- **Interests**

- **Personality**

- **Cited evidence** from multiple posts/comments

Analysis:

- Gemini Flash 1.5 provided a **comprehensive, human-like persona profile**.

- It correctly inferred subtle personality traits, occupation clues, interests, and supported claims with Reddit comment evidence.

- The model showed strong reasoning, empathy, and creativity, making it **ideal for NLP tasks like persona generation**, even when content is partially removed or vague.

Offline Capability - [Flan-T5] can Run Locally

[Gemini - Flash - 1.5] can Needs API / cloud access

Streamlit

🔍 Reddit User Persona Generator (LLM-powered)

Enter a Reddit profile URL below to generate a user persona using Gemini AI.

Reddit Profile URL

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

Generate Persona

Persona generated successfully!

📄 Download Persona

🟣 User Persona

Reddit User Persona:

* **Username:** /u/NYC_NightOwl_33
* **Age:** 33
* **Occupation:** Likely works in a professional field, possibly requiring a degree. The H1B visa posts suggest he might work in a tech field or a similar industry requiring specialized skills.
* **Interests:** Nightlife (specifically bars with a mature atmosphere), video games (lack of late-game content is a complaint), city life, history (implied by reflecting on generations of New Yorkers), and possibly farming/resource management games (based on comment about granaries).
* **Location:** New York City (explicitly stated)
* **Personality:** Introspective, somewhat cynical, but ultimately empathetic. He observes and analyzes his surroundings, reflecting on his own place within them. He's also a bit of a gamer and seems to enjoy discussing societal issues and engaging in some lighthearted banter.

Evidence from Posts and Comments:

Thank You