**GenA11yHelper**

Smarter Prompt Testing for GenAI – Deployed with Terraform and AWS

**## Project Overview**

“GenA11yHelper” is a general-purpose tool that helps fine-tune different LLM system prompts through realtime user feedback on the responses, and automatically redeploy a base model with a promoted prompt. It makes it easier to improve GenAI systems by putting user feedback at the heart of the process.

**## How It Works**

* **Frontend Interface:** A Streamlit + LangChain app inside a Docker container allows users to input data and test prompts.
* **LLM Querying:** Prompts are processed via the OpenAI API to generate responses.
* **User Rating:** Each response is rated (e.g., thumbs up/down, or in our case a 1-5 slider), simulating real user feedback.
* **Prompt Versioning:** Prompt templates are versioned and stored in an AWS S3 bucket.
* **Feedback Logging:** Ratings are logged (evaluated via Weights & Biases).
* **DevOps Automation:** GitHub Actions CI/CD is used to promote the best-performing prompts in a GitHub repo (`prompt-templates/`) via cronjob.

**## Technologies Used**

* **Terraform** – Infrastructure as Code (IaC)
* **AWS EC2** – Hosts the containerized app
* **AWS S3** – Stores versioned prompt templates
* **Streamlit + LangChain** – Web app frontend
* **Docker** – Container environment
* **OpenAI API** – LLM querying
* **GitHub Actions** – Prompt promotion via CI/CD
* **Weights & Biases**  – Feedback evaluation script (visualization possible, TODO)

**## AWS Infrastructure (via Terraform)**

* Sets up a secure S3 bucket to store prompt versions
* Launches a small Ubuntu server (EC2) using the free tier
* Installs Docker on the server to run the app
* Opens ports for the app (8501) and SSH access (22)
* Creates SSH keys for secure login
* Displays the server's public IP and login details

**## Repository Structure**/terraform/ → Infrastructure code

flowchart.png → System architecture

prompt-templates/ → (Planned) Prompt storage

README.md → This file

**## Notes**Hackathon: DevOps for GenAI 2025

Track 5: Prompt Testing & Experimentation

AWS cloud environment hosting credits provided by the organizers

Project built within an 8-hour hackathon window