

PROMPT DESIGN DOC – Jinnie

Prompt Objective: *The goal is to generate short, friendly, and accurate answers based solely on the content of the current web page, mimicking a helpful human assistant without hallucinating beyond the input context.*

The LLM model used: **DeepSeek/deepseek-r1-0528-qwen3-8b**

When the user input the question, the extension scrapes the entire website text. Then the entire text along with the guiding instruction is inputted as prompt to the LLM

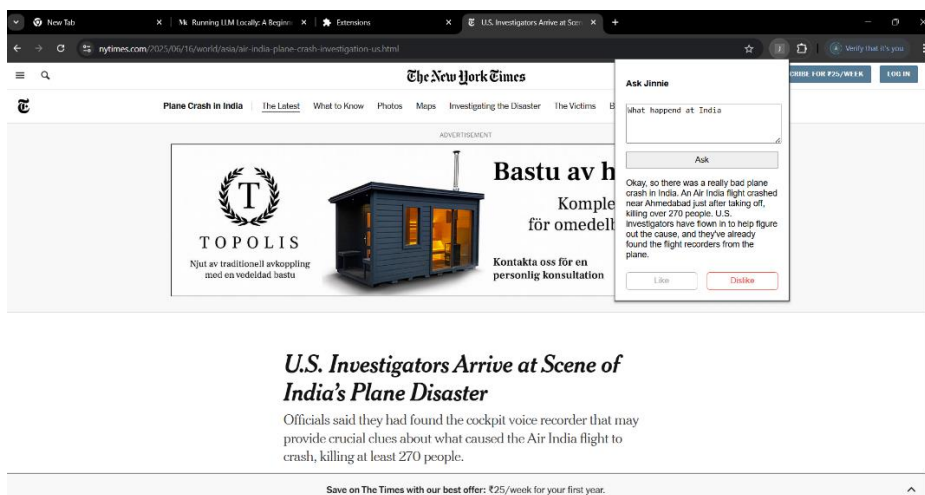
Prompt: “Based only on the content below, answer this question in a short and clear way like you're talking to a friend. Keep it to just 3 to 4 sentences”

Code snippet

```
prompt = f"Based only on the content below, answer this question in a short and clear way like you're talking to a friend. Keep it to just 3 to 4 sentences.\n\nContent:\n{content}\n\nQuestion: {question}"
```

This prompt will reduce verbosity, control tone and avoid hallucination

eg:



A screenshot of a Medium article titled "Running LLM Locally: A Beginner's Guide to Using Ollama" by Arun Patidar. The article has 81 likes and 5 comments. A chatbot interface is overlaid on the right side of the article, showing a conversation with "Ask Jinnie". The chatbot has asked "How many claps and reactions this article got?" and received the response "This article received "81" claps and got "5" reactions." The chatbot also has buttons for "Like" and "Dislike". The article's header image shows a laptop, a server, and a tablet with the text "RUNNING LLMs LOCALLY WITH OLLAMA".

New TabRunning LLM Locally: A BeginnExtensionsU.S. Investigators Arrive at Scemedium.com/@arunpatidar26/run-llm-locally-ollama-6ba296747505

MediumSearch

Running LLM Locally: A Beginner's Guide to Using Ollama


Arun Patidar

Follow

3 min read · Oct 28, 2024

B1

5



RUNNING

LLMS

LOCALLY

WITH

OLLAMA

Ask Jinnie

Who wrote this article?

Ask

This article was written by Arun Patidar.

LikeDislike

Sign up

Sign in