Style Guide - AI Fashion Assistant

NVIDIA + LangChain Generative AI Contest entry by Karim Lalani

Presented at Austin LangChain Meetup on July 10th 2024

Socials

LinkedIn
@-karim-lalani/

Github
@lalanikarim

Medium
@klcoder

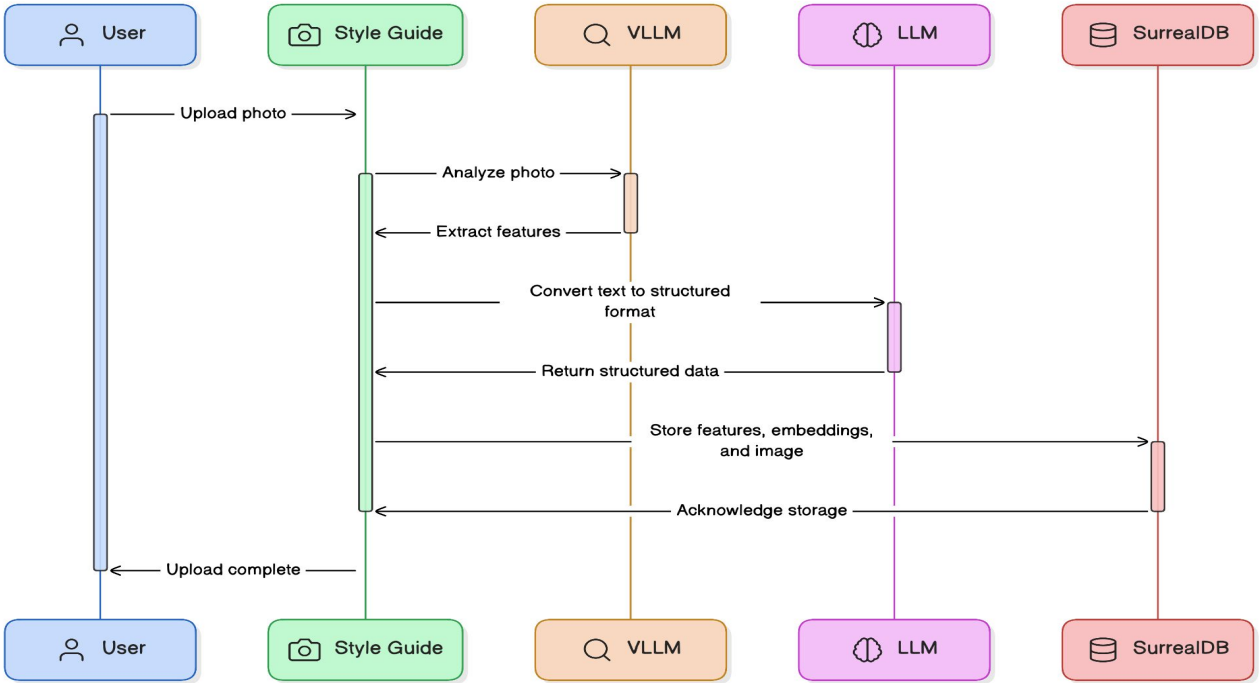
Project Overview

- Voice enabled AI assistant
- Lets you talk to your wardrobe
- Analyzes uploaded photos
- Identifies the kind of clothes, color, style, etc.
- Stores attributes along with photo in a vector database
- Suggests outfits based on user’s query
  - Show me casual wear
  - Show me formal wear
  - Show me blue shirts
  - I am going on a vacation to the Mediterranean. Suggest some outfits to pack.

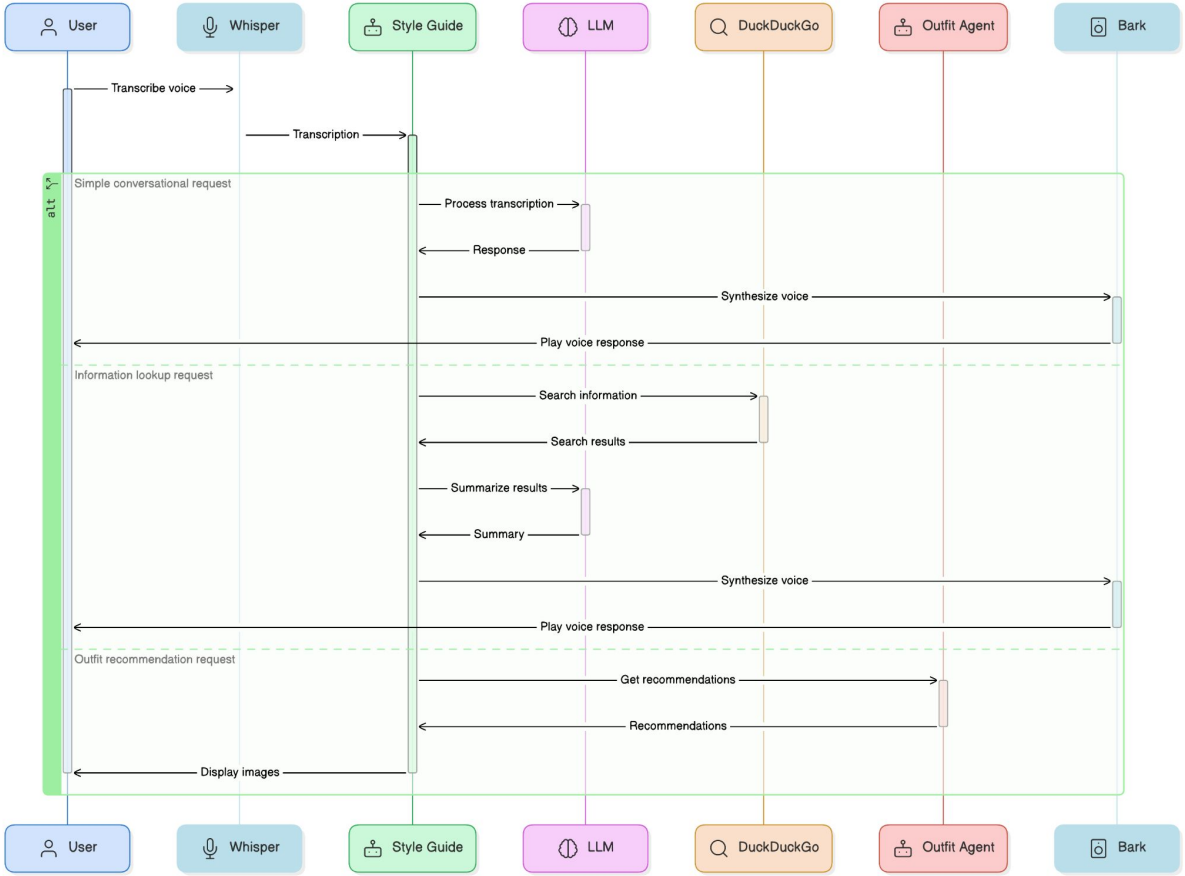
Technologies

- Speech to text is using OpenAI's open source Whisper mini model.
- Text to Audio is using Suno's open source Bark small model.
- WebRTC scaffolding from WebRTC AI Voice Chat project.
- Hosted AI Inference NVIDIA Inference Microservice (NIM):
  - Text-to-text LLM meta/llama3-8b-instruct
  - Image-to-text VLLM microsoft/phi-3-vision-128k-instruct
  - Retrieval Embeddings nvidia/embed-qa-4
- Document and Vector store database SurrealDB
- Web Search using DuckDuckGo
- GenAI framework LangChain
- GenAI API calls monitoring and tracing LangSmith from LangChain

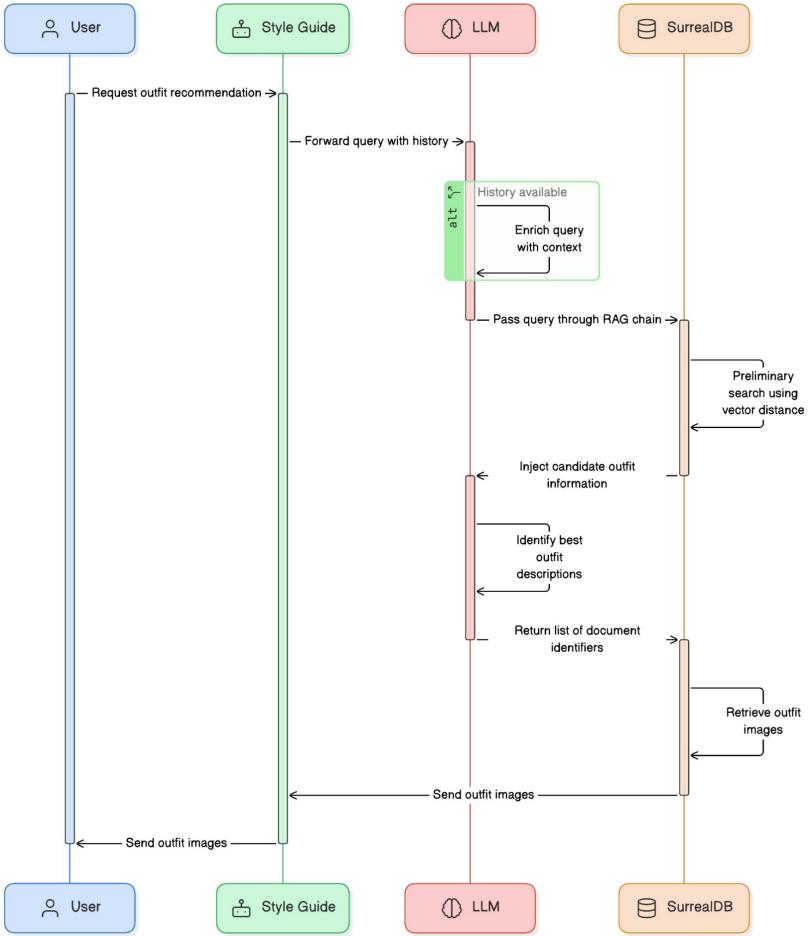
Photo upload workflow



AI Agent Workflow



Outfit Recommendation Workflow



Challenge - ChatNVIDIA lacks tool calling support

ChatNVIDIA model within **langchain-nvidia-ai-endpoints** package didn't support tool calling at the time of contest

Resolution

Wrote a wrapper inspired by **OllamaFunctions** to add tool calling capabilities to ChatNVIDIA model

```
nvidia_functions = create_tool_calling_model(ChatNVIDIA, "chat_nvidia_functions")
```

https://github.com/lalanikarim/style-guide-ai-assistant/blob/main/tool_calling_model.py

PR #23193 has been submitted to LangChain with a Python mixin version of the wrapper

Challenge - phi-3-vision’s poor JSON output quality

phi-3-vision-128k-instruct model didn’t generate json output which was needed to extract features as structured output

Resolution

Deferred structured output parsing to **llama3-8b-instruct** which handled tool calling and structured output parsing reliably

```
outfit_output_llm = _llm.with_structured_output(Outfit, include_raw=False)
nv_fc_chain = RunnableLambda(get_image_prompt)
              | _vllm
              | RunnableLambda(get_message_content)
              | outfit_output_llm
```

Demo

Style Guide - AI Assistant

Questions?

Resources

Github - https://github.com/lalanikarim/style-guide-ai-assistant/

NVIDIA NIM Playground - https://build.nvidia.com/explore/discover

Surreal DB - https://surrealdb.com/

LangChain - https://www.langchain.com/

Most Awesome LangChain Resource in Central TX - https://www.meetup.com/austin-langchain-ai-group/

Thank you!