PromptChainer: Chaining Large Language Model Prompts through Visual Programming

Sherry Tongshuang Wu*, Ellen Jiang*, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, Carrie Cai
wtshuang@cs.washington.edu, { ellenj, donsbach, jeffgray, alemolinata, michaelterry, cjcai }@google.com

TL;DR: Explore LLM chain authoring for realistic prototyping.

Motivation and Problem

Large Language Models enables prototyping with AI.

Realistic apps are too complex to prototype with a single LLM run,

But possible with a chain of multiple LLM calls, each for one sub-task.

How to support users in authoring their own LLM chains?

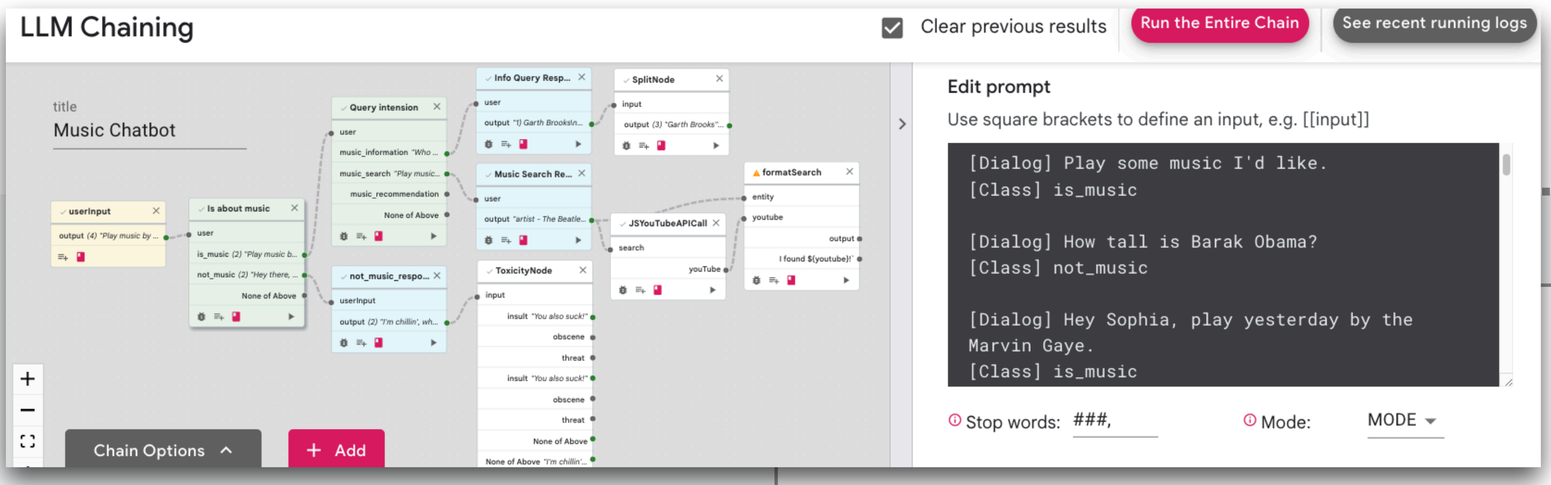
Contribution

Summarize challenges in chain authoring.

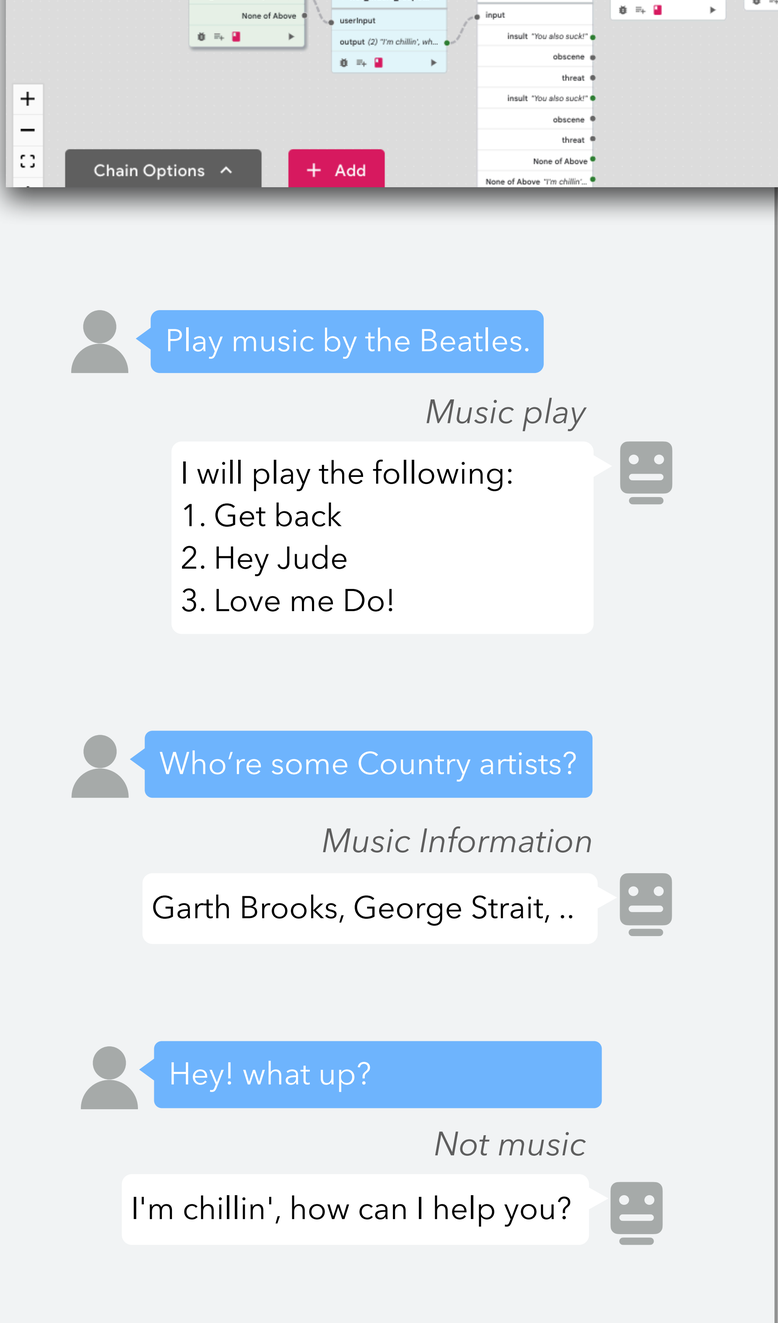
PromptChainer: Interface, visually program chains.

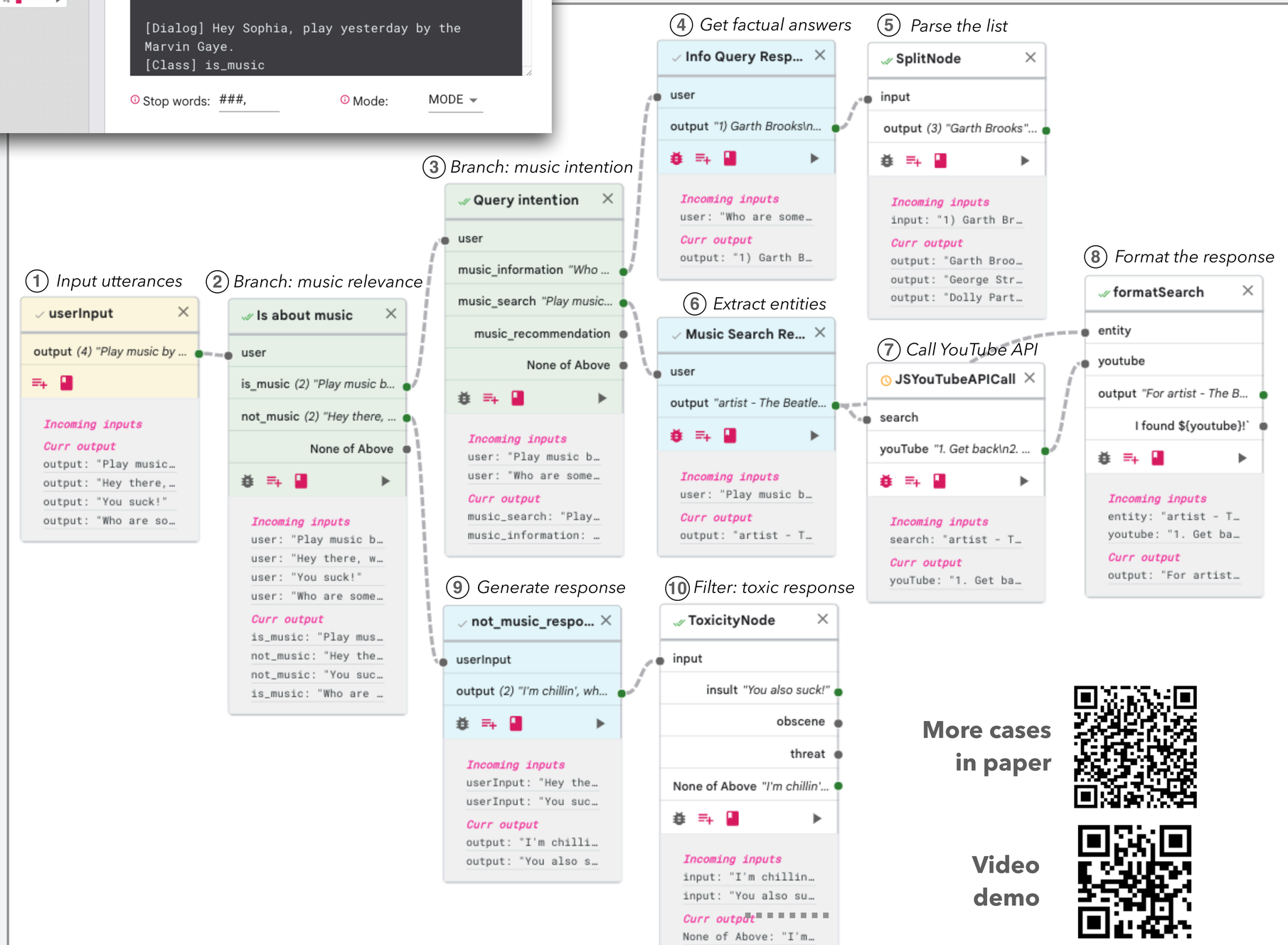
Case studies: Summarize patterns in chain authoring.

Future work: scalability, low-fi chain prototyping.



PromptChainer Interface & Walkthrough example: music chatbot





More cases in paper

Video demo

Chaining challenges & PromptChainer Solutions

LLMs are hard to chain.

Q: Handle LLMs' arbitrary str outputs?

A: Provide scaffolding nodes.

LLMs chains are hard to design.

Q: What sub-tasks are feasible?

A: Example galleries for classifier, etc.

LLMs chains are hard to debug.

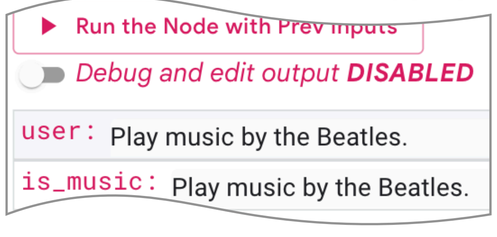
Q: How to handle cascading errors?

A: Multi-level, interactive debugging

Evaluation	⑩ Toxicity classifier
Processing	⑤ Split by number
API Call	⑥ Call YouTube API

Unit testing

Breakpoint Debugging



Case study & open questions

4 designers / developers, self-proposed diverse tasks & chains

Showed multiple chaining objective:

Address LLM limitations;

Build extensible prototypes (e.g. add additional nodes);

Used different chaining construction strategies:

Sequentially implement each step;

Sketch out placeholder nodes first before filling them in.

Needed "in-context" debugging:

Refine prompts w.r.t interactions between LLM steps.

Discovered Additional challenges:

Coherence b/w interdependent sub-tasks?

Low-fi prototyping: refine chain structures, vs. local prompts?