Leveraging Tool-Calling to enhance LLM Capabilities

Student

Moritz Schiesser

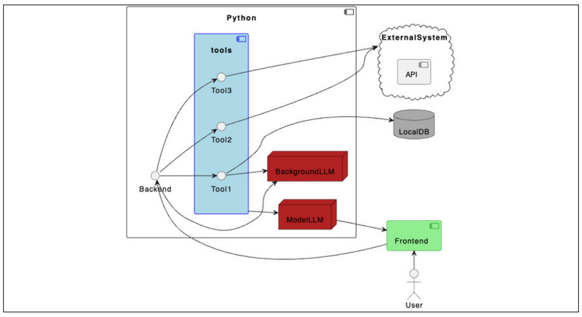
Introduction: Recent advancements in generative natural language processing (NLP) models have lead companies and academic research institutions alike to develop and deploy increasingly sophisticated models for a variety of applications.
These models have shown to be capable of generating grammatically correct and coherent human-like text.
Since these models not only learn to produce grammatically correct text but also learn to recall and reproduce information from their training data, they are often unable to talk about topics that are not present in their training data.
While they are able to generalize well to unseen data, when it comes to factual real-world and real-time information, they often fall short.
This can be bypassed by providing the model information about a topic by extending the prompt with this information.
Further, LLMs cannot directly interact with other systems.
Tool-calling can be used to take action as a consequence of the user's input, be it providing information to the model, or taking action on the user's behalf.

This project explores how tool-calling can be leveraged to enhance the capabilities of large language models.

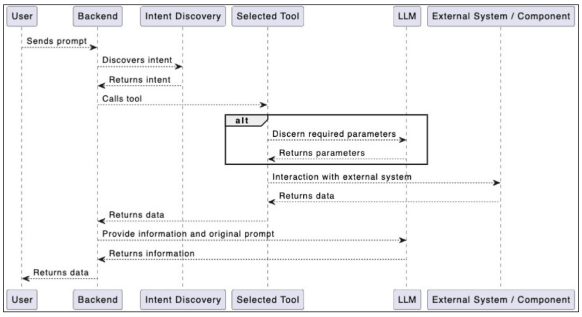
Approach / Technology: LLMs do not have access to real-time information, which makes it difficult for them to generate up-to-date information, neither can they interact with other systems directly. Tool-calling can be used to enhance the capabilities of LLMs.
A tool, as referred to in this context, is a fragment of code that interacts with other systems and provides relevant information to the LLM. For example, a tool can be used to fetch real-time information from a database or an API and provide this information to the model. To decide, which tool should be invoked, an LLM is presented with the available tools and their descriptions, alongside the user prompt. Based on the selection, required information is extracted by querying the LLM again, which is used to invoke the tool. The result of the tool is fed back to the LLM, alongside the original user prompt, to generate the final output. This project explores and illustrates this by building a chatbot system that covers parts of the \`themovieDB\` API.

Result: The sample application successfully demonstrates how tool-calling can be leveraged to enhance the capabilities of various models.
The approach of using generative AI to select which programmatic tool to call, and then using the output of that tool to help the LLM generate the final output, is shown to be viable and effective.
While the selection of the generative text model has a great impact on the reliability and performance of the application, even comparatively small models can be used to great effect.

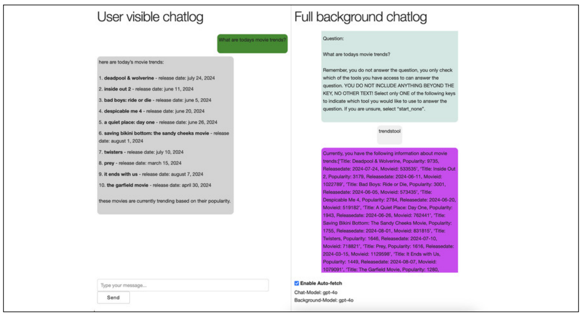
An overview over the components of this approach, and how they interact with each other
Own presentment



Sequence diagram of tool-calling
Own presentment



The UI of the sample application with the typical start of a conversation between user and system
Own presentment



User visible chatlog

What are todays movie trends?

here are today's movie trends:

1. **deadpool & wolverine** - release date: july 24, 2024
2. **inside out 2** - release date: june 11, 2024
3. **bad boys: ride or die** - release date: june 5, 2024
4. **despicable me 4** - release date: june 20, 2024
5. **a quiet place: day one** - release date: june 26, 2024
6. **saving bikini bottom: the sandy cheeks movie** - release date: august 1, 2024
7. **twisters** - release date: july 10, 2024
8. **prey** - release date: march 15, 2024
9. **it ends with us** - release date: august 7, 2024
10. **the garfield movie** - release date: april 30, 2024

these movies are currently trending based on their popularity.

Full background chatlog

Question:

What are todays movie trends?

Remember, you do not answer the question, you only check which of the tools you have access to can answer the question. YOU DO NOT INCLUDE ANYTHING BEYOND THE KEY, NO OTHER TEXT! Select only ONE of the following keys to indicate which tool you would like to use to answer the question. If you are unsure, select "start_none".

trendstool

Currently, you have the following information about movie trends:['Title: Deadpool & Wolverine, Popularity: 9735, Releasedate: 2024-07-24, Movieid: 533535', 'Title: Inside Out 2, Popularity: 3179, Releasedate: 2024-06-11, Movieid: 1022789', 'Title: Bad Boys: Ride or Die, Popularity: 3001, Releasedate: 2024-06-05, Movieid: 573435', 'Title: Despicable Me 4, Popularity: 2784, Releasedate: 2024-06-20, Movieid: 519182', 'Title: A Quiet Place: Day One, Popularity: 1943, Releasedate: 2024-06-26, Movieid: 762441', 'Title: Saving Bikini Bottom: The Sandy Cheeks Movie, Popularity: 1755, Releasedate: 2024-08-01, Movieid: 831815', 'Title: Twisters, Popularity: 1646, Releasedate: 2024-07-10, Movieid: 718821', 'Title: Prey, Popularity: 1616, Releasedate: 2024-03-15, Movieid: 1129598', 'Title: It Ends with Us, Popularity: 1449, Releasedate: 2024-08-07, Movieid: 1079091', 'Title: The Garfield Movie, Popularity: 1280,

☑ **Enable Auto-fetch**
Chat-Model: gpt-4o
Background-Model: gpt-4o

Advisor
Prof. Dr. Mitra Purandare

Subject Area
Data Science