日日是Oracle APEX

Oracle APEXを使った作業をしていて、気の付いたところを忘れないようにメモをとります。

2024年4月25日木曜日

Ollama + command-r-plus:104b-q2_KのローカルLLMでファンクション呼び出しを試す

以前の記事「OpenAI Chat Completions APIのツール呼び出しを使ってPL/SQLファンクションを呼び出す」で作成したAPEXアプリケーションを使って、Ollama + command-r-plus:104b-q2_Kでファンクション呼び出しを試してみます。

Ollama自体はファンクション呼び出しをサポートしていないはずなのですが、プロンプトを調整すればそれっぽいことができそう、とのことで試してみました。以下の記事を参考にしています。

Function calling using Ollama models

https://www.deskriders.dev/posts/1702742595-function-calling-ollama-models/

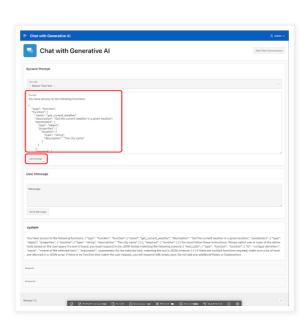
結論から先に言うと、プロンプトの記述だけでは厳しいかな、と感じました。以下、行った作業を 記述します。

APEXアプリケーション側の設定として、置換文字列の**G_API_ENDPOINT**は http://host.docker.internal:11434/v1/chat/completions、**G_MODEL_NAME**にはcommand-r-plus:104b-q2_Kを設定しています。

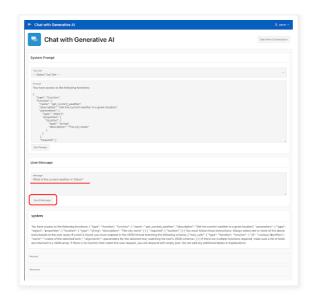
ollamaをローカルで実行し、APEXアプリケーションにアクセスします。

Promptとして以下を与え、**Set Prompt**をクリックします。ツール定義の部分は**OpenAI**のtoolの定義をそのまま埋め込んでいます。また出力の形式指定も**OpenAI**の仕様で返すように要求しています。

```
}
         },
         "required": [
                "location"
          ]
      }
}
You must follow these instructions:
Always select one or more of the above tools based on the user query
If a tool is found, you must respond in the JSON format matching the following schema:
{
   "tool_calls": {
        "type": "function",
        "function": {
            "id": "<unique identifier>",
            "name": "<name of the selected tool>",
            "arguments": <parameters for the selected tool, matching the tool's JSON schema>
      }
   }
}
If there are multiple functions required, make sure a list of tools are returned in a JSON arra
If there is no function that match the user request, you will respond with empty json.
Do not add any additional Notes or Explanations
system_prompt_for_ollama_fc.txt hosted with ♥ by GitHub
                                                                                         view raw
```



Message としてWhat is the current weather in Tokyo?を入力し、Send Messageをクリックします。



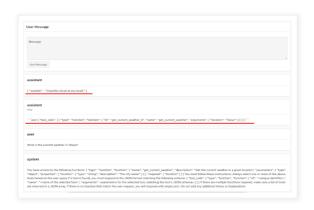
プロンプトが想定どおり解釈されると、roleがassistantのメッセージとして以下のようなJSONのレスポンスが返されます。

```
```json { "tool_calls": [ { "type": "function", "function": { "id": "get_current_weather_0", "name": "get_current_weather", "arguments": { "location": "Tokyo" } } } ] }```
```

パッケージUTL\_OPENAI\_CHAT\_APIには、OpenAI以外のエンドポイントから上記ようなJSONが返されたときは、ファンクション呼び出しと解釈してPL/SQLファンクションを呼び出すようにコーディングしています。

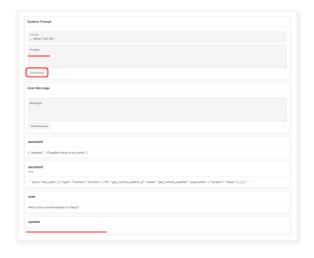
そのため、ファンクション**get\_current\_weather**が引数**{ "location": "Tokyo"}**で呼び出されます。ファンクション呼び出しのレスポンスは(**OpenAI**以外ではroleがtoolのメッセージに対応していないため)、roleをassistantにしてチャット履歴に追記します。

{ "weather" : "Clear(No cloud at any level)" }

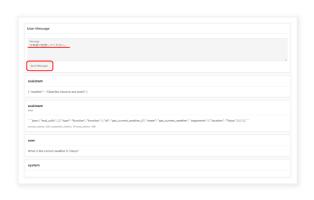


JSONで天候の回答は得ていますが、これから追加でメッセージを送信するにはsystemプロンプトからファンクション呼び出しの設定を除く必要があります。

**Prompt**を空にして**Set Prompt**をクリックします。 ロールがsystemのメッセージが無くなります。

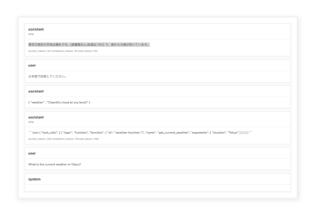


Messageに日本語で回答してください。と入力し、Send Messageをクリックします。



以下の回答が返されました。

東京の現在の天気は晴れです。(高層雲なし)気温は 13°C で、南からの風が吹いています。



東京の天気について聞いているということは理解されていて、天気が晴れで雲がないところまでは 合っていますが、気温や南からの風という情報は適当です。

systemプロンプトの記述によりOpenAIのtool\_callsのJSONが返されるのは少々驚きましたが、これから先には進めないように感じました。

今回の記事は以上です。何かの参考になれば幸いです。

**ベ** ホーム

# ウェブ バージョンを表示

#### 自己紹介

# Yuji N.

日本オラクル株式会社に勤務していて、Oracle APEXのGroundbreaker Advocateを拝命しました。 こちらの記事につきましては、免責事項の参照をお願いいたします。

## 詳細プロフィールを表示

Powered by Blogger.