

CSAI 422: Lab Assignment 4

Building Conversational Agents with Tool Use and Reasoning Techniques

This repository contains the implementation of a conversational agent that uses tools to gather information and solve problems through different reasoning techniques.

Setup Instructions

Prerequisites

- Python 3.7 or higher
- OpenAI API key
- WeatherAPI key

Installation

1. Clone this repository:
2. `git clone <your-repository-url>`
3. `cd <repository-directory>`
4. Install the required packages:
5. `pip install openai requests python-dotenv`
6. Create a `.env` file in the root directory with the following variables:
7. `API_KEY=your_openai_api_key`
8. `BASE_URL=https://api.openai.com/v1`
9. `LLM_MODEL=gpt-4-turbo-preview`
10. `WEATHER_API_KEY=your_weather_api_key`
11. Run the application:
12. `python conversational_agent.py`

Implementation Overview

The conversational agent is implemented with three different reasoning strategies:

1. **Basic Tool Calling:** A simple agent that can use weather-related tools to provide information.
2. **Chain of Thought (CoT):** An enhanced agent that uses step-by-step reasoning to solve more complex problems.
3. **ReAct (Reasoning and Acting):** An advanced agent that explicitly follows the Thought-Action-Observation pattern to solve multi-step problems.

Tool Implementation

The agent has access to the following tools:

- `get_current_weather`: Retrieves current weather data for a location
- `get_weather_forecast`: Retrieves a weather forecast for a location for a specified number of days
- `calculator`: Performs mathematical calculations
- `web_search`: Simulates web search for information (simplified implementation)

Agent Types

Each agent type uses a different system message to guide its behavior:

1. **Basic Agent**: Simple weather assistant with no specific reasoning instructions
2. **Chain of Thought Agent**: Follows a step-by-step approach to break down problems
3. **ReAct Agent**: Explicitly uses the Thought-Action-Observation pattern

Comparative Evaluation

A bonus feature allows users to evaluate the same query across all three agent types and rate their responses on a scale of 1-5. Results are saved to a CSV file for further analysis.

Example Conversations

Basic Agent

You: What's the weather like in Paris today?

Weather Assistant: I checked the current weather in Paris for you. Right now, it's 14.0°C (57.2°F) with Partly cloudy conditions. The humidity is at 82% and there's a light wind at 9.4 km/h.

You: Will it rain in London tomorrow?

Weather Assistant: I checked the weather forecast for London. Tomorrow, you can expect cloudy conditions with a 71% chance of rain. Temperatures will range between 8.8°C (minimum) and 14.1°C (maximum). I'd recommend having an umbrella handy!

Chain of Thought Agent

You: Should I wear a jacket if I'm visiting New York and Chicago next week?

Weather Assistant: To answer your question about whether you should wear a jacket for your visit to New York and Chicago next week, I'll need to check the weather forecasts for both cities.

First, let me get the weather forecast for New York for the next week:

Looking at the forecast for New York, the temperatures will range from lows of around 4.2°C to highs of 13.4°C over the next few days, with conditions varying from "Sunny" to "Overcast".

Now, let me check the forecast for Chicago:

The Chicago forecast shows temperatures ranging from lows of -1.1°C to highs of 9.8°C, with conditions including "Partly cloudy" and "Cloudy".

Analyzing this information:

- Both cities will have relatively cool temperatures
- New York will have lows in the 4°C range and highs up to 13°C
- Chicago will be colder, with temperatures potentially dropping below freezing at night
- Both cities will have variable conditions

Based on these forecasts, I would definitely recommend wearing a jacket for your trip. Chicago in particular will be quite cold, especially in the evenings and early mornings. A medium to heavy jacket would be appropriate, and you might want to consider layers for New York where daytime temperatures could be more moderate.

ReAct Agent

You: What's the temperature difference between Tokyo and Buenos Aires today, and which city should I choose for a beach day?

Weather Assistant: Thought: I need to find the current temperatures in both Tokyo and Buenos Aires, calculate the difference, and then determine which city would be better for a beach day based on temperature and conditions.

Action: First, I'll check the current weather in Tokyo.

I've checked the current weather in Tokyo. The temperature is 11.0°C (51.8°F) with Clear conditions. Humidity is at 36% with wind at 9.4 km/h.

Action: Now I'll check the current weather in Buenos Aires.

I've checked the current weather in Buenos Aires. The temperature is 24.0°C (75.2°F) with Sunny conditions. Humidity is at 48% with wind at 9.0 km/h.

Action: Now I'll calculate the temperature difference between these cities.

The temperature difference between Tokyo and Buenos Aires is 13.0°C (24°F).

Thought: Now I need to determine which city would be better for a beach day. This depends primarily on the temperature, but also on weather conditions, wind, etc.

Final Answer: The temperature difference between Tokyo and Buenos Aires today is 13.0°C (24°F), with Buenos Aires being significantly warmer at 24.0°C (75.2°F) compared to Tokyo's 11.0°C (51.8°F).

For a beach day, Buenos Aires would definitely be the better choice. At 24.0°C with sunny conditions, it offers ideal beach weather. Tokyo's temperature of 11.0°C is too cold for comfortable beach activities, even though the sky is clear. The warm, sunny conditions in Buenos Aires make it much more suitable for spending time at the beach today.

Analysis of Reasoning Strategies

Effectiveness of Different Strategies

1. **Basic Agent:**
 - Strengths: Simple, direct responses; efficient for straightforward queries
 - Limitations: Lacks transparency in reasoning; struggles with complex or multi-part questions
2. **Chain of Thought Agent:**
 - Strengths: Provides more detailed explanations; better at handling numerical comparisons and calculations; shows work clearly
 - Limitations: Responses can be longer; sometimes provides unnecessary details
3. **ReAct Agent:**
 - Strengths: Most transparent reasoning process; best at handling multi-step problems; clear structure helps track problem-solving
 - Limitations: Verbose responses; explicit thought process can seem mechanical

Impact on Response Quality

The reasoning strategies significantly affected response quality across different types of queries:

1. **Simple information retrieval** (e.g., "What's the weather in Paris?"): All agents performed similarly, with the Basic Agent being most concise.
2. **Comparative questions** (e.g., temperature differences): Chain of Thought and ReAct agents provided more reliable answers by showing their calculations.
3. **Decision-making questions** (e.g., should I wear a jacket?): ReAct and Chain of Thought agents provided better-justified recommendations by explicitly considering relevant factors.
4. **Multi-step problems**: ReAct agent clearly outperformed others by breaking down complex queries into manageable steps and explicitly tracking progress.

User Preference

Based on the comparative evaluation results, users generally preferred:

- Basic Agent for simple weather queries
- Chain of Thought for explanatory content
- ReAct for complex problem-solving tasks

Challenges and Solutions

API Integration Challenges

Challenge: Managing the asynchronous nature of tool calls and ensuring the model waits for tool responses before completing its answer.

Solution: Implemented a structured message processing system that handles tool calls, captures their responses, and feeds them back to the model for final response generation.

Tool Call Precision

Challenge: Ensuring the model correctly formats tool calls and selects appropriate tools.

Solution: Provided clear tool definitions with detailed parameter descriptions and implemented error handling for malformed tool calls.

Reasoning Strategy Implementation

Challenge: Getting the model to consistently follow the specified reasoning pattern, especially for the ReAct approach.

Solution: Created detailed system prompts with examples of the expected reasoning pattern and consistent formatting.

Weather API Limitations

Challenge: The free tier of WeatherAPI has query limits.

Solution: Implemented caching for frequently requested locations and ensured efficient API usage.

Evaluation Methodology

Challenge: Developing an objective method to compare agent performance.

Solution: Created the comparative evaluation feature that presents responses side-by-side and collects user ratings, storing results for analysis.

Conclusion

This implementation demonstrates how different reasoning strategies can significantly impact the quality and transparency of AI assistant responses. The ReAct approach proved particularly effective for complex, multi-step problems, while the Chain of Thought approach provided a

good balance of reasoning transparency and conciseness. The basic approach remains valuable for simple queries where efficiency is prioritized.

The project successfully demonstrates how tool use combined with explicit reasoning can create more capable AI assistants that can not only retrieve information but also process it meaningfully to address user needs.