MR2020: Coding for METOC

Module 2: Querying ChatGPT/Teaming with Generative AI

What is ChatGPT? How does it work? (*According to ChatGPT)*

ChatGPT Overview:
•**AI Model:** Advanced conversational AI by OpenAI, most recently based on GPT-4.
•**Purpose:** Engages in human-like conversations, answers questions, and assists with tasks.

How ChatGPT Works:
1.Training:
1. **Data Collection:** Trained on diverse text data.
2. **Pre-training:** Learns language patterns via unsupervised learning.
3. **Fine-tuning:** Refined with supervised learning for specific tasks.

2.Architecture:
1. **Transformer Model:** Uses attention mechanisms for context understanding.
2. **Tokenization:** Breaks text into tokens, processes via neural networks.

3.Response Generation:
1. **Input Processing:** Converts input text to embeddings.
2. **Contextual Understanding:** Maintains conversation context.
3. **Output:** Generates and converts tokens back to text.

Key Features:
•Natural language understanding
•Contextual awareness
•Broad knowledge base
•Versatility across applications

Applications:
•Customer support
•Educational tools
•Content creation
•Programming assistance

A chatbot can be a useful tool for building code from scratch, finding errors, or suggesting ways to make code faster. However, you must know how to interact with it.

Generative AI can't read your mind! Thoughtfully articulate your inquiries and be concise and exact!

A simple example...

```
import numpy as np

A = np.ones(4,4)
A += 2
print(A)
```

Try running this! It will give you an error. Then enter the following to ChatGPT.

I have the following code:

import numpy as np
A = np.ones(4,4)
A += 2
print(A)
Why do I get this error? TypeError: Cannot interpret '4' as a data type

ChatGPT Response

The error occurs because the np.ones function is being used incorrectly. The correct way to create a 4x4 array of ones is to pass the shape as a tuple. Here's the corrected code:

```
import numpy as np

A = np.ones((4, 4)) # Correct way to define a 4x4 array of ones
A += 2 # This adds 2 to each element of the array
print(A)
```

A more detailed explanation then follows.

Question Specificity

General Question: "How do I read a file in Python?"

- *What's the problem with this inquiry?*

1) There are many different types of file formats used in earth science (e.g., NetCDF, HDF, text, CSV, zarr, grib, other binary).

2) There are sometimes multiple ways to load a file (e.g., read using native Python commands, use module specific to format, xarray, pandas, etc.).

Specific Question: "How can I read a CSV file in Python using pandas?"

- This question specifies the file type and the software used to load the data.

NOTE: Software changes over time so function names also change. This means you may need to mention the specific versions of modules/libraries you are using. While this is often not necessary, if ChatGPT produces code that calls a function that does not exist, then the problem may be associated with different functions available if different versions of a module.

Example: "How do I read a CSV file using pandas version 1.2.4 and Python 3.8?"

Contextualize

Without Context: "Why is my code not working?"

- A vague and open-ended question as worded above will prompt GPT to essentially look through your entire code for anything that may be wrong. You will often get potential solutions that don't have anything to do with the specific error you have encountered.

With Context: "Here's my Python code to read a CSV file using pandas, but it throws an error 'FileNotFoundError'. What could be wrong?"

- In this example, you have specified what you are trying to do and the specific error that has occurred. This gives GPT more information to work with, so it can more effectivecly focus on the problem.

If possible, include pieces of code.

In the examples on the previous slides, including code will help GPT identify the cause of problems. Otherwise, GPT might generate some boilerplate code and walk you through what it does; however, that code may not address what you are trying to do.

```
import pandas as pd

df = pd.read_csv('data.csv')
print(df.head())
```

Describe errors or unexpected/unwanted behavior

When receiving an error message: "I'm getting a 'TypeError: unsupported operand type(s) for +: 'int' and 'str'' when I run my code."

Don't say: "I'm getting an error when running my code."

Very frequently, our code will run without an error, but it produces unexpected behavior. Perhaps we're trying to make plot and the code generates a plot, but it doesn't look correct. Such errors are known as **logical errors**. These can be insidious and difficult to fix errors! Python will do exactly what you tell it to do without complaining, but you have nonetheless done something that gives you an undesirable result.

Examples of unexpected behavior inquiries:

"My function is supposed to return a sorted list, but it returns an unsorted list instead."

"The 'mean' column in dataframe 'df' was supposed to find the average value of 'Tsquare' in a 5 by 5 grid cell area around the coordinates shown in the 'x' and 'y' columns. However, all I get are NaNs."

Can you find the logical error?

```
def calculate_average(numbers):
    total = 0
    for num in numbers:
        total = num
    average = total / len(numbers)
    return average

# Test the function
numbers = [10, 20, 30, 40, 50]
print("Average:", calculate_average(numbers))
```

Run the code. The average printed out is not the actual average. Can you tell why by just looking at the code? If not, can you formulate a query for ChatGPT to tell you what is wrong?

Break down questions

If you have a complex issue, break it down into smaller, manageable parts. This also helps you evaluate new code piece by piece to ensure it is not producing errors or undesirable behaviors. *Small chunks of code are easier to debug than large blocks of new code!*

Complex: "How do I build a web scraper using BeautifulSoup in Python and store the data in an SQLite database?"

Broken Down:

"How do I scrape data from a webpage using BeautifulSoup in Python?"
Check to see that the code generated by ChatGPT works.

Then ask, "How do I store scraped data into a SQLite database [INSERT VARIABLE NAME] using SQLAlchemy in Python?" You might insert the variable name for the SQLite database from the code generated in response to the first question.

In general, avoid using ambiguous terms and ensure your questions are easy to understand. Be as exact as possible. AI does not easily understand nuance!