

# Data analysis using openAI: getting to know the Code Interpreter

Jorge Guerra Pires<sup>1\*</sup>

**Abstract:** On this document, we are getting to know more about the code interpreter from openAI.  
: openAI — data science — statistical inference — python — open source — diabetes

<sup>1</sup> Founder at IdeaCoding Lab and JovemPesquisador.com

**DOI:** <http://dx.doi.org/10.22456/2175-2745.XXXX> • **Received:** dd/mm/yyyy • **Accepted:** dd/mm/yyyy

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

## 1. Details on openAI API Assistant

Unfortunately, there are not many details on their official documentations. It is possible to find articles on websites on the internet, but they are not official. They may be speculating on how coder interpreter works. What I have added herein was learnt by using, and sometimes it was possible to see the behavior as logs on the playground. Also, it is possible to ask the assistant directly.

If we ask it directly, which can be done during the analysis.

Question: Why is your name coder interpreter?

Assistant: The name "Coder Interpreter" reflects my role and capabilities within this environment. As an AI language model, I have the ability to interpret and understand natural language inputs, as well as execute code and perform computations based on those inputs. This includes tasks such as coding, data analysis, statistical modeling, and more.

The "Coder" part of the name represents my ability to work with code and perform computations, while the "Interpreter" aspect reflects my natural language understanding and interpretation capabilities. Together, these qualities allow me to assist with a wide range of coding and data-related tasks, making the name "Coder Interpreter" a fitting descriptor of my role within this environment.

Question: how do you work under the hood? I mean, how are you conducting the analysis asked? are you using your own codes, or are you using Python libraries?

Assistant: I utilize Python, including popular data science libraries such as pandas, numpy, and

scikit-learn to perform data analysis and modeling. These libraries provide powerful tools for data manipulation, statistical analysis, and machine learning. If needed, I can also leverage other libraries depending on the specific requirements of the analysis.

Question: Does that mean that you are restricted to Python libraries?

Assistant: For the purpose of this environment, I primarily utilize Python libraries. However, I can also integrate with other tools and environments as per the user's request. If you have any specific requirements or questions about using other tools, feel free to ask!

Question: Do you mean function calling? adding my own function to be called as tools?

Assistant: Yes, that's correct! If you have a specific function or tool that you'd like to integrate into the analysis, you can certainly share the details with me. I can then incorporate your function into the analysis process and utilize it as needed.

Question: What are your limitations as data science tool?

Assistant: As a data science tool in this environment, there are certain limitations to be aware of:

1. **Internet Access:** This tool is not connected to the internet, so it cannot access external websites or APIs.
2. **Non-Persistent Storage:** Files uploaded during a session can only be accessed within

that session. Once the session ends, the files are no longer available.

3. **Restricted Libraries:** While this environment supports many popular data science libraries, it may not have every possible library or tool that could be accessed in a traditional data science setup.
4. **Computation Time:** There are computational limits in terms of processing power and time for each individual request.
5. **No External Integrations:** The tool cannot integrate with external services, databases, or systems.
6. **Limited Environment Customization:** The environment is pre-configured, and users can't install additional software or make low-level system changes.

Despite these limitations, this tool provides a powerful platform for data analysis, machine learning, and collaboration in a contained and secure environment.

Question: So it means that if I have a tool that is not in Python, say in R, I need to create a tool, and make it available as function calling.

Assistant: Yes, if you have specific functionality or tools that are not available through Python libraries in this environment, one approach would be to create the required functionality as a function or set of functions that can be called and utilized within this environment. This function can then be integrated into the analysis and used as needed. If you have a specific tool or functionality in mind, feel free to share the details, and I can help guide you through the process of creating a function to achieve the desired outcome.

Question: Are you really running the calculations? can I trust your calculations?

Assistant: Yes, I am actively running the calculations and analysis based on the code and instructions provided. You can trust the calculations and analysis performed within this environment. The code is executed in a stateful Jupyter notebook environment, and the results are generated based on the provided data and instructions. If you have any specific concerns or questions about the analysis, feel free to ask, and I will be happy to address them.

Question: Are you using the function calling approach, that is, you have a set of functions with description, and you use your ability as LLM for deciding which function to call, and interpret the results

Assistant:

Yes, that's a good way to think about it. I have a set of built-in functions and libraries that I can utilize to perform various data analysis, statistical calculations, machine learning tasks, and more. When you ask a question or request an analysis, I use my expertise as a language model to determine which functions and methods to apply in order to provide you with the most relevant and accurate results.

I interpret the results and provide insights based on the output of the calculations and analysis performed. If there are specific functions or methods you'd like to see applied, or if you have custom functions you'd like to integrate, feel free to share the details, and I can guide you through the process.

Question: So, if I give extra functions as tool, my function will be equal chance to be called, in case it is the best fit for the analysis

Assistant:

Absolutely, if you provide additional custom functions or tools that are relevant to the analysis at hand, I will consider them as part of the available resources when determining which functions to use. If your custom function is the best fit for a particular analysis or task, it will certainly be considered for use.

Incorporating custom functions into the analysis process can enhance the range of tools available and ensure that the analysis is tailored to the specific needs of the problem at hand. If you have a custom function you'd like to integrate, feel free to share the details, and I can assist with incorporating it into the analysis as needed.

As it is possible to see, the tool is very dynamics. It also allows function calling, which is an approach also by openAI APIs that allows to define personalized functions, and use them on a real-time applications, see [????] for real-world applications of function calling.