

# Bonus Assignment - Survey

Thank you for completing the bonus assignment! This survey will help us understand your experience, challenges, and suggestions for improving the framework. Your responses are invaluable for our research. Please answer the following questions thoughtfully.

## Section 1: Implementation Result

### 1. Please record the command you use to create, run, and evaluate the framework.

#### - Create:

```
python create.py --config ./examples/l13840.json --output-dir
./examples/stackoverflow_assistant
```

#### - Run:

```
python run.py --input-dir ./examples/stackoverflow_assistant
```

#### - Evaluation:

```
python eval.py --model_api http://127.0.0.1:8000/eval/chat --config
./examples/stackoverflow_assistant_config.json --documents_dir
./examples/stackoverflow_assistant --output-dir
./examples/stackoverflow_assistant --max_turns 8
```

### 2. Please copy and paste a sample conversation where you interacted with your created Agent.

```
Bot: Hello! I'm your Stack Overflow Developer Assistant. How can I assist you with your coding challenges today?
You: I want a list of the newest questions
getAPIBotResponse Time: 4.581143140792847
Bot: Here is a curated list of the newest questions related to Python and C++ from Stack Overflow, obtained through the API call to the following URL: https://api.stackexchange.com/2.3/questions?order=desc&sort=creation&site=stackoverflow&filter=&21in:9C929VJ:zgmKVMUkpFkFt.W6L0MwVik9Lu2Ww290yA1S18M4.6P-B.0bpr7MTwMvH7TXZE5JUoz56pagesize=56page=1&tagged=python3Bc42B4ZB.
Most recent questions asked by the community.

What specific topic or tag are you interested in? Do you have any filters you would like to apply, such as date, votes, or sorting?
You: I am interested in Python and C++
getAPIBotResponse Time: 13.942810164268864
Bot: Here is a curated list of the newest questions related to Python and C++ from Stack Overflow, obtained through the API call to the following URL: https://api.stackexchange.com/2.3/questions?order=desc&sort=creation&site=stackoverflow&filter=&21in:9C929VJ:zgmKVMUkpFkFt.W6L0MwVik9Lu2Ww290yA1S18M4.6P-B.0bpr7MTwMvH7TXZE5JUoz56pagesize=56page=1&tagged=python3Bc42B4ZB.

1. **Title:** How to use Interception in Python
   - **Link:** [Read more](https://stackoverflow.com/questions/79381439/how-to-use-interception-in-python)
   - **Description:** A user seeks guidance on suppressing a left mouse click and replacing it with other actions using Python, utilizing tools like AHK and PyAutoGUI, and facing issues with low-level input processing in games. They successfully implemented this in C++ using the 'Interception' library but encountered problems when using the code as a DLL in Python.

2. **Title:** How to Debug Python-to-C++ Calls in NumPy Using VS Code?
   - **Link:** [Read more](https://stackoverflow.com/questions/79294214/how-to-debug-python-to-c-calls-in-numpy-using-vs-code)
   - **Description:** This question discusses challenges in setting up a debugging workflow to trace Python functions' interaction with underlying C/C++ code in a library using VS Code. The user seeks advice on using debuggers like gdb or lldb, building NumPy with debug symbols, and identifying called C/C++ functions.

3. **Title:** Giving C++ sole ownership of class object in pybind11
   - **Link:** [Read more](https://stackoverflow.com/questions/79289183/giving-c-sole-ownership-of-class-object-in-pybind11)
   - **Description:** A user wants C++ to have sole ownership of a class object exposed to Python using pybind11, ensuring that Python references don't affect the lifetime of the C++ object. The inquiry explores smart pointers and weak references to manage object ownership.

4. **Title:** How to debug python script called by c++ executable in VSCode?
   - **Link:** [Read more](https://stackoverflow.com/questions/79287140/how-to-debug-python-script-called-by-c-executable-in-vs-code)
   - **Description:** The question involves debugging a Python script executed by a C++ executable using VS Code. The user seeks solutions to integrate the debugging processes of Python and C++ code to efficiently halt at breakpoints in both languages.

5. **Title:** Misalignment in 2-dimensional CUDA-FFTShift Function
   - **Link:** [Read more](https://stackoverflow.com/questions/79286453/misalignment-in-2-dimensional-cuda-fftshift-function)
   - **Description:** A beginner in CUDA programming needs help with a 2D FFT-shift operation on a flattened array. Despite correct implementation in Python, the CUDA version produces misaligned or corrupted outputs, possibly due to thread interference.

These questions provide insights into common challenges faced when working with Python and C++ together, specifically in contexts such as debugging, object ownership, and interoperability with libraries and scripts.
You: Can you give me more questions?
getAPIBotResponse Time: 15.557229842053223
Bot: Here is an additional set of the newest questions related to Python and C++ from Stack Overflow, obtained through the API call to the following URL: https://api.stackexchange.com/2.3/questions?order=desc&sort=creation&site=stackoverflow&filter=&21in:9C929VJ:zgmKVMUkpFkFt.W6L0MwVik9Lu2Ww290yA1S18M4.6P-B.0bpr7MTwMvH7TXZE5JUoz56pagesize=56page=2&tagged=python3Bc42B4ZB.

1. **Title:** Connect and Receive data from BLE server
   - **Link:** [Read more](https://stackoverflow.com/questions/79284885/connect-and-receive-data-from-ble-server)
   - **Description:** A user working on an IoT project is trying to connect a Windows device to a BLE device, specifically the "Arduino Nano 33 BLE Sense Rev 2", to receive data. They are considering using different programming languages and have tried using the 'Bleak' library but faced difficulties.

2. **Title:** How to make boost:python use the Python embeddable package
   - **Link:** [Read more](https://stackoverflow.com/questions/79271218/how-to-make-boostpython-use-the-python-embeddable-package)
   - **Description:** The question involves a C++ DLL project using boost:python to embed Python code. The user seeks a method to direct boost:python to use the Python embeddable package instead of the full installation, facing issues with DLL preference.

3. **Title:** Can't load C++ unit test module with python 3.10 on windows (using boost 1.75.0)
   - **Link:** [Read more](https://stackoverflow.com/questions/79259880/cant-load-c-unit-test-module-with-python-3-10-on-windows-using-boost-1-75-0)
   - **Description:** A user encounters an ImportError when trying to load a C++ unit test module in Python 3.10 on Windows, using boost 1.75.0. The module works fine with other Python versions and on Linux, and they suspect a configuration issue.

4. **Title:** Google Trimmed Match install issue
   - **Link:** [Read more](https://stackoverflow.com/questions/79251691/google-trimmed-match-install-issue)
   - **Description:** The user is trying to install 'trimmed_match' from GitHub on a local machine and faces an error related to file creation during the wheel building process. They have tried various installation methods without success.

5. **Title:** How to achieve this: after function execution, everything gets restored to the state before calling it. Can fork be used to implement it?
   - **Link:** [Read more](https://stackoverflow.com/questions/79246516/how-to-achieve-this-after-function-execution-everything-gets-restored-to-the-s)
   - **Description:** The user is exploring ways to perform "probe" function evaluations that do not alter the state of the program. They consider using the 'fork' system call to achieve this, ensuring the function execution does not affect the original context.

These questions highlight challenges and discussions related to Python and C++ in areas such as Bluetooth communication, library configuration, unit testing, installation issues, and managing function execution contexts.
You: █
```

Hi! Hello! I'm your Stack Overflow Developer Assistant. Now can I assist with your coding challenges today? You: I am encountering this problem: RecursionError: maximum recursion depth exceeded, can you help me debugging? `getFullResponse Time: 5.9279763936119s` Bot: It sounds like you're running into a recursion limit in your code. This typically happens when a recursive function calls itself too many times without reaching a base case to stop the recursion. Here are a few steps you can take to debug and fix this issue:

1. **Check the Base Case:** Ensure your recursive function has a correct and reachable base case. The base case is what stops the recursion, so if it's missing or incorrect, the recursion can go on indefinitely.
2. **Trace the Calls:** Print statements or use a debugger to trace the recursive calls. This can help you understand how the recursion is progressing and where it's going wrong.
3. **Decrease Recursion Limit (with caution):** If your recursion is deep but correct, you can increase the recursion limit using `sys.setrecursionlimit()`. However, be cautious with this approach as it can lead to a stack overflow if you're not careful.

Here's a simple example of how a base case might be implemented:

```
'''python
def recursive_function(n):
    if n == 0: # Base case
        return
    else:
        print(n)
        recursive_function(n - 1)'''
```

Could you please provide more details about the coding error you encountered? Sharing any error messages or code snippets would be helpful.

You: this is my code: `def fin(n): return fin-1 + (n-2)`  
`getFullResponse Time: 4.942542599736s`  
Bot: It looks like your function "fin" is trying to compute something similar to the Fibonacci sequence, but it's missing a base case to stop the recursion. Without a base case, the function will keep calling itself indefinitely, leading to a "RecursionError".

Here's how you can fix it by adding base cases for when "n" is 0 or 1:

```
'''python
def fin():
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fin(n-1) + fin(n-2)'''
```

# Example usage  
`print(fin(5))` # This should return 5, as fin(5) in Fibonacci series is 5.

In this example, the function will stop recursion when "n" is 0 or 1, which are the base cases for the Fibonacci sequence.

You need further clarification on the provided solutions, or do you have any additional questions? You: Are there any python3 question that is answer this problem? `getFullResponse Time: 14.61184760724795s`  
Bot: The qd call was made to the following URL: <https://api.stackexchange.com/2.3/questions?order=desc&sort=activity&site=stackoverflow&filter=with!n=7119,9029972,zgmW0uWjgFkt,w6UwMv1K9Uw290yA1510H4,6P-8,6qr7MDwWfTfXZES3Jor56pageize=5&taged=python3&recursion=page1>. The output prov

ids a list of Stack Overflow questions related to Python recursion. Here are the key details from the output:

1. **Question 1:**
  - **title:** How to optimize my performances, using asynchronous python code
  - **tags:** python, python-3.x, asynchronous, recursion, fastapi
  - **score:** 4
  - **link:** [View Question \(https://stackoverflow.com/questions/77583664/how-to-optimize-my-performances-using-asynchronous-python-code\)](https://stackoverflow.com/questions/77583664/how-to-optimize-my-performances-using-asynchronous-python-code)
  - **summary:** The user is working with asynchronous requests in Python using FastAPI and aiohttp. They are trying to optimize their code for better performance and have included recursive calls for handling deep paging requests.
2. **Question 2:**
  - **title:** Jinja templating with recursive in dict doesn't work
  - **tags:** python, dictionary, recursion, jinja2
  - **score:** 2
  - **link:** [View Question \(https://stackoverflow.com/questions/79172802/jinja-templating-with-recursive-in-dict-doesnt-work\)](https://stackoverflow.com/questions/79172802/jinja-templating-with-recursive-in-dict-doesnt-work)
  - **summary:** The user is facing issues with Jinja templating when attempting to implement recursion in a dictionary. They encounter a `ValueError` related to unpacking values in a recursive loop.
3. **Question 3:**
  - **title:** Generating list satisfying linear integer constraint in python
  - **tags:** python, recursion, generator
  - **score:** 4
  - **link:** [View Question \(https://stackoverflow.com/questions/79261267/generating-list-satisfying-linear-integer-constraint-in-python\)](https://stackoverflow.com/questions/79261267/generating-list-satisfying-linear-integer-constraint-in-python)
  - **summary:** The user wants to generate all lists of integers satisfying a linear constraint using Python. They are exploring a recursive approach but face challenges in implementation.
4. **Question 4:**
  - **title:** Find full path of element in tree
  - **tags:** python, recursion, tree
  - **score:** 4
  - **link:** [View Question \(https://stackoverflow.com/questions/25474652/find-full-path-of-element-in-tree\)](https://stackoverflow.com/questions/25474652/find-full-path-of-element-in-tree)
  - **summary:** The user seeks help in finding all paths to a tree element using recursion. They want to modify their code to return multiple paths as lists.
5. **Question 5:**
  - **title:** How about the recursive code for "Find Zero" I've written
  - **tags:** python, pointers, recursion
  - **score:** 4
  - **link:** [View Question \(https://stackoverflow.com/questions/79243349/how-about-the-recursive-code-for-find-zero-i-v-written\)](https://stackoverflow.com/questions/79243349/how-about-the-recursive-code-for-find-zero-i-v-written)
  - **summary:** The user created a recursive function to move zeros to the end of a list while retaining the relative order of other elements. They seek feedback on the efficiency of their recursive solution.

These questions might help you understand how recursion is being used and discussed in different contexts, providing insights into potential solutions for your issue.

**3. Which of the following components did you use to create your Agent? (Check all that apply.)**

- ☒ Interactive Panel to customize task planning
- ☒ Taskgraph Generation
- ☒ Worker Customization
- ☐ Evaluation

## Section 2: Feedback on Components

4. For the fields in the config file, Which field(s) in the config file were most useful for creating your Agent? Which field(s) did you find least important? Are there additional fields or features you would suggest adding to improve the Agent's capabilities?

I feel that the fields “role,” “user\_objective,” “builder\_objective,” and “domain” are quite useful. A clear description of the agent can improve the quality of responses. For “task\_docs” and “rag\_docs,” I think the variable numbers are good, but the algorithm used for retrieving all the links is not satisfactory. A complex website might have many links that are not relevant to the descriptions provided in “task\_docs” and “rag\_docs.” For example, one may want to include only the help and contact pages, but the generator might visit other irrelevant pages. (I think most of

the time, one can detect what to expect from the URL, e.g., having the same subdomain or path. Maybe having an LLM for filtering could be useful). Moreover, some websites have link in javascript and do not use <a>, this could be a problem.

I really like the “Tasks” field, as it allows for custom steps. However, sometimes the steps can become very complex, depending on different factors, like a tree. I am curious if a tree-like task structure could be more effective (e.g., allowing certain steps to be skipped under specific conditions).

I have also added another field for a list of APIs. The logic is to have a list of APIs such that, when creating an agent, a JSON file is generated with information about the APIs (e.g., description, API URL, parameter rules, etc.), and a worker decides which API to call and how to call it (function call).

Personally, I did not find “task\_docs” useful due to the nature of my agent (I need to perform tasks across very different websites and not a fixed set of websites).

**5. Were the instructions for constructing the config file clear? Please specify areas where improvements are needed.**

I think the instructions were pretty straightforward and easy to follow.

**6. How well did the generated task planning (without modification) cover the tasks you expected the Agent to handle?**

- ☐ All
- ☒ Most
- ☐ some
- ☐ Few
- ☐ None

**7. How useful were the steps generated in the task planning (without modification)? Rate their usefulness on a scale from 1 to 5. (1: Not Useful at All, 5: Very Useful)**

- ☐ 1
- ☐ 2
- ☐ 3
- ☒ 4
- ☐ 5

**8. If you used the Interactive Panel for customizing task planning, how useful was it? Rate its usefulness on a scale from 1 to 5. (1: Not Useful at All, 5: Very Useful)**

- ☐ 1
- ☐ 2
- ☐ 3
- ☒ 4
- ☐ 5

**9. Follow up to question 8: Based on your experience in customizing task planning: How much of the original task planning did you modify?**

- ☐ 80%+
- ☐ 60%+
- ☐ 40%+
- ☒ 20%+
- ☐ <20%

**10. If you designed your own worker, what are they and how were they used?**

I have designed a worker to handle StackOverflow API calls. It is specific for StackOverflow as there are some specific rules in their APIs, but I believe one can have a generic API worker (maybe not working perfectly, but should work well on most of the APIs given a clear documentation).

I think a worker that combines different workers would be super useful. Especially, when more workers are built, each one can be independent of each other. The current logic is one create a worker and use workflow to connect different functions. But having a worker that automatically selects the workers and combines them would be cool.

Another worker that could be useful is a search worker but to some specific link(s), Tavily is powerful but there might be some cases where I would love to imitate the search process like humans do instead of using Tavily that has some limitations.

**11. Follow up to question 10: How difficult did you find customizing the Worker framework? Rate the difficulty on a scale from 1 to 5. (1: Very Easy, 5: Very Difficult)**

- ☐ 1
- ☐ 2
- ☒ 3
- ☐ 4
- ☐ 5

**12. How useful was the user simulator for evaluating your Agent? Rate its usefulness on a scale from 1 to 5. (1: Not Useful at All, 5: Very Useful)**

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☒ 5

**13. What issues did you identify in your Agent during evaluation, and how did you address them?**

I encountered different issues:

- Default Worker does not return the correct worker. E.g. When the task is about clarification, it still returns the worker that calls directly the APIs despite the fact that in the description of API worker I clearly stated that the worker is not useful for clarification

questions. Solution: I did prompt engineering, and got a prompt that is working better than the one before (still not good in some scenarios) -> I think the problem is Llm gives more priority to the historical conversation than task.

- The output of Llm might not follow the format that I gave to it, leading to error when evaluation. Solution: handle these exceptions and return some error messages to the user.
- I have encountered a situation where task value (i.e. `state["orchestrator_message"]["task"]`) is an empty string. Not sure why I am having this problem.

### Section 3: Feedback on Framework

**14. How would you rate the overall difficulty of using the Agent First Organization framework? (1: Very Easy, 5: Very Difficult)**

- ☐ 1
- ☐ 2
- ☐ 3
- ☒ 4
- ☐ 5

**15. What improvements would you suggest for the framework?**

I think it is a very interesting framework. I can see the potential of this framework and what it is capable of achieving. However, I believe there are some functions missing in the current framework. Additionally, I think one issue with the framework is that it can be quite difficult to debug. Fixing one bug in a specific conversation might lead to another error in a different conversation that was previously working correctly.

**16. Additional comments or suggestions?**

I believe that the tree-like task procedure could be highly effective. One key advantage of this approach is that it would make the configuration file much simpler. (I think it is currently possible to achieve the same outcome by creating different tasks for each path to the leaf.) A tree structure would also be beneficial when tasks share the same prefix task, as it would allow for faster processing, especially as tasks become more complex.

Another consideration is that LLM calls are quite time-intensive. While I feel these calls are generally necessary, finding a balance between LLM calls and predefined rules could be valuable. For example, in some task steps, it is quite obvious which worker to use, so it would make sense to directly assign that worker instead of relying on the default worker. I know this logic already exists, but I don't think it is functioning effectively when selecting the worker in `taskgraph.json`.

```
"2",
{
  "name": "DefaultWorker",
  "attribute": {
    "value": "",
    "task": "Utilize the StackOverflowAPIWorker to search Stack Overflow for existing solutions or discussions related to the provided error details.",
    "directed": false
  },
  "limit": 1
}
```