


Agentic Workflow - Colab Implementation

Tracker:  Agentic Workflow - Colab Implementation

AHT:

- **Assertions:** ~60 mins/row
- **Actions:** ~60 mins/row
- **Blocked (API/Scenario):** ~15 mins/row
- **1st Reviews:** ~60 mins

Expected Rows to be moved in “**Ready for Review**”:

- Assertions: 6-8 rows/ day
- Actions: 6-8 rows/ day
- Anything less than 6/ day would need a valid justification



Objective:



The goal of this initiative is to operationalize the **Eval Orchestration workflow** by leveraging the **Autogen prompt-based system** to automatically generate **Actions** and **Final Assertions** based on model responses and capabilities. Unlike previous iterations where the Autogen code was provided as prefilled templates, we now use **Autogen prompts embedded within the shared Eval Orchestration Colab notebooks** to generate outputs dynamically. These outputs are handled within two distinct Eval Orchestration Colab notebooks:

- One dedicated to **Action generation**
- One dedicated to **Final Assertion generation**

Each Colab will take a different input structure and have slightly different usage patterns, as outlined in the documentation. With this updated agentic workflow, the process of writing code for Final Assertions or Actions is largely automated. The **Scenario team** will now prefill the **DB Initialization** and **Initial Assertion** sections, and the **Colab Implementation team** will focus solely on the generation of **Actions** and **Final Assertions**.


Relevant Colabs and Sample Inputs:

ID	Step	Eval Orchestration Colab	Sample Inputs
1	Final Assertion	 [Internal]_ - _Final_As...	 [Sample Input - Final Assertio...

2	Action	 [Internal]_ - _Actions...	 [Sample Input - Action] - Ag...
---	--------	---	--

Note: Please do create a copy of above colabs before utilizing them.


Video Reference:

 pxk-bkro-bpn (2025-05-13 18:16 GMT+1)

Workflow:

Building on the split workflow approach introduced on **March 31**, we will continue using separate workflows for **Action** and **Final Assertion** generation. These workflows will follow slightly different templates tailored to their respective outputs.


How to Use the Eval Orchestration Workflow for Final Assertion Generation:

- The Scenario team will provide input Colabs with the **DB Initialization** and **Initial Assertion** already completed.
- This also means we do the 1st pass from our end on the correctness of **DB Initialization** and **Initial Assertion**.
- These pre-filled sections should **not be modified** by the Colab Implementation team. If any changes are absolutely necessary, they must be logged in a designated tracker (a column for this purpose will be added).
- Example input:  [Sample Input - Final Assertion] - Agent-769_edge_2-Merged.ipynb

Execution Process:

- Run the Final Assertion Eval Orchestration Colab using the provided input.
- After each execution, the interface will prompt you to:
 - Indicate whether you are satisfied with the generated output.
 - Provide **feedback notes**, which will be used to refine the model's response in the next iteration.
- Each sample allows for **up to 20 retries** to achieve satisfactory results.
- **Note:** All existing guidelines around **API documentation adherence**, **avoidance of hardcoding**, and other standards still apply. The only change is in the **automated generation process** of the code.

How to Use the Eval Orchestration Workflow for Action Generation:


- Once the Final Assertions are completed by the Assertion team, the Action team will receive a Colab template that includes the following:
 - **Initial Assertion (pre-filled)**
 - **Final Assertion (pre-filled)**
- Your task as the Action Generation team is solely to generate the Action based on the context provided.
- Example Input:  [Sample Input - Action] - Agent-769_edge_2-Merged.ipynb

Execution Process:

- Run the Action Eval Orchestration Colab using the provided input.
- We wouldn't need human feedback in this loop, instead observe the model response to understand the correctness and model loss.
- Each sample allows for **up to 20 retries** to achieve satisfactory results.
- **Note:** All existing guidelines around **API documentation adherence**, **avoidance of hardcoding**, and other standards still apply. The only change is in the **automated generation process** of the code.

Note: Samples that require **Inspection in Final Assertion** cannot be generated through this Agentic Workflow and will need to be handled separately.

Mandatory attachments:

- Copy the output of the eval orchestration execution for the action step, save it in a txt file and attach it to column AJ (Eval orchestration execution output [txt file]).
- The txt files must be placed in this folder:  Eval_orchestration_txt_files