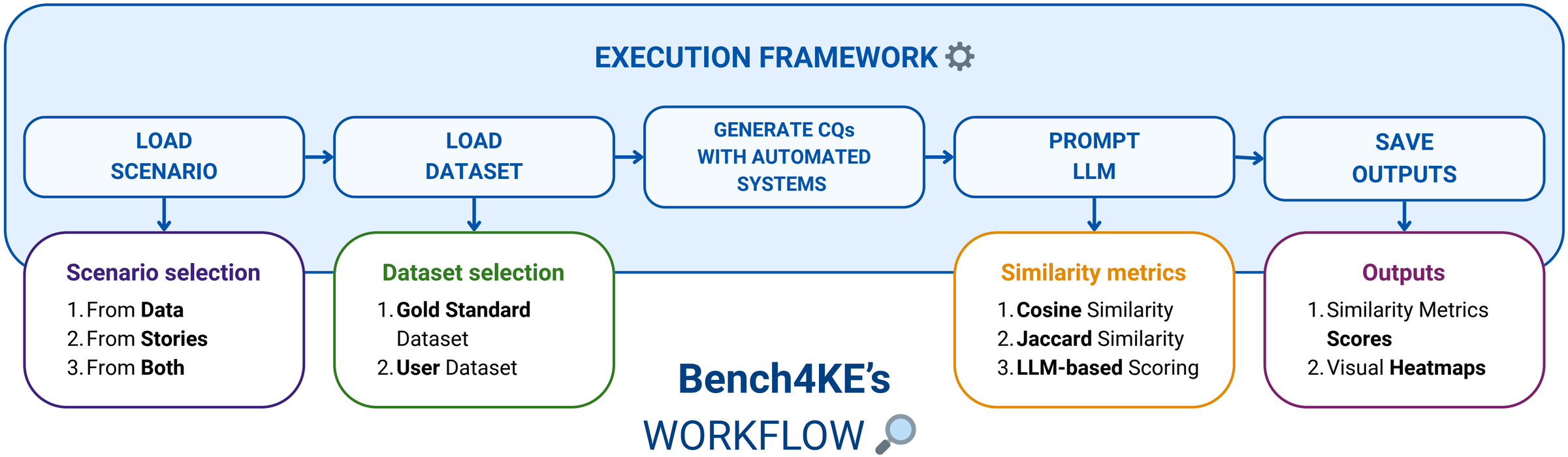
Bench4KE

An Overview



Bench4KE's WORKFLOW 🔎

Bench4KE

The API

- **CURRENT STATUS**: Not yet deployed online
- **FUTURE AVAILABILITY**: Will be deployed alongside paper submission

Bench4KE

Your API: Requirements

- Must be running at an external service URL
- Must accept a CSV dataframe as input
- Must be able to open URLs (including raw GitHub link for datasets)
- Must return the same dataframe with an added “generated” column

- ***If you’d like, we will provide you a template for GPT prompts to use as a “mock API”***

Bench4KE

Local Setup: Installation Requirements

- **Dependencies**: fastapi uvicorn pandas scikit-learn openai pytest
- **OpenAI API key** (ours are available if you need them)
- **Server launch**: uvicorn app.main:app --reload --host 127.0.0.1 --port 8000

Bench4KE

Local Setup: Endpoint Call

- curl or Python’s requests
- curl -X POST "http://127.0.0.1:8000/validate/" \
- -F "use_default_dataset=true" \
- -F "external_service_url=http://127.0.0.1:8001/newapi" \
- -F "api_key=key_for_external_service" \
- -F "validation_mode=all" \
- -F "output_folder=heatmaps" \
- -F "model=gpt-4" \

-F "save_results=true"

Bench4KE

Local Setup: Form Parameters

- use_default_dataset: Must be "true" to use our benchmark dataset
- external_service_url: URL of your CQ-generation service
- api_key: Any API key if needed by your external service
- validation_mode: One of: all | llm | cosine | jaccard
- output_folder: Local folder where heatmaps will be saved
- model: OpenAI model for LLM analysis (e.g. "gpt-4")
- save_results: "true" to dump a CSV of results into RESULTS_DIR

Bench4KE

Local Setup: Results Structure


```
{
  "message": "Processing complete",
  "results_saved_to": "results/validation_results_1714562823.csv",
  "validation_results": [
    {
      "Gold Standard": "What is the project about?",
      "Generated": "Tell me about the project?",
      "LLM Analysis": "1. The pairs with highest similarity are…\n2. Missing essential CQs…",
      "Average Cosine Similarity": 0.85,
      "Max Cosine Similarity": 1.00,
      "Average Jaccard Similarity": 0.67,
      "Cosine Heatmap": "heatmaps/cosine_heatmap_123456789.png",
      "Jaccard Heatmap": "heatmaps/jaccard_heatmap_987654321.png"
    },
    … more rows …
  ]
}
```


Bench4KE

Local Setup: Viewing Results

- **CSV File**: Open results/validation_results_<timestamp>.csv for a complete tabular view
- **Heatmaps**: Images referenced in "Cosine Heatmap" and "Jaccard Heatmap" are PNG files viewable with any image viewer

Bench4KE