

## Generation:

First before evaluation of model, its required to generate the response from model. Here model can be offline or online, depending on this, one should select the api endpoint for generation.

Swagger url: <http://localhost:8000/api/v1/trustllm/docs>

## Steps for generating responses:

### Create a dataset:

- 1) /api/v1/trustllm/dataset.
- 2) Copy the dataset name generated.

The screenshot displays the Swagger UI interface for the API endpoint `/api/v1/trustllm/dataset`. The interface is divided into several sections:

- GET /api/v1/trustllm/dataset Dataset**: The top section showing the method and endpoint.
- Parameters**: A section indicating "No parameters" with a "Cancel" button.
- Servers**: A section showing the server URL `/` with a dropdown menu and "Execute" and "Clear" buttons.
- Responses**: A section showing the response details for a 200 status code.

The response details include:

- Code**: 200
- Response body**: A JSON object `{ "dataset_name": "dataset_08052024100700" }` with a "Download" button.
- Response headers**: A list of headers including `cache-control: private, no-store`, `content-length: 41`, `content-type: application/json; charset=utf-8`, `date: Mon, 05 Aug 2024 04:37:00 GMT`, `server: uvicorn`, `vary: Origin`, and `x-xss-protection: 1; mode=block`.

At the bottom, there is a "Responses" table with columns for "Code", "Description", and "Links".

## Generation For offline models (Huggingface models).

To evaluate the models present on huggingface

/api/v1/trustllm/offline/generation/

- **model\_name**: Enter the model name. It should be same as the one present on huggingface.
- **test\_type**: Enter any one of the following:
  - [privacy, fairness, ethics, safety, truthfulness]
- **dataset\_name**: Enter the dataset that have generated in first step.

The screenshot shows a web interface for the API endpoint `/api/v1/trustllm/offline/generation/`. The interface has a header bar with the method `GET` and the endpoint path. Below the header is a section titled "Parameters" with a "Cancel" button. The parameters are listed in a table with columns "Name" and "Description".

Name	Description
<b>model_name</b> * required string (query)	meta-llama/Meta-Llama-3.1-8B-Instruct
<b>test_type</b> * required string (query)	fairness
<b>dataset_name</b> * required string (query)	dataset_07242024131340

## Generation for online models:

/api/v1/trustllm/online/generation/

- **test\_type**: Enter any one of the following:
  - [privacy, fairness, ethics, safety, truthfulness]
- **dataset\_name**: Enter the dataset that have generated in first step.
- **model\_url**: Enter the url of the hosted model
- **model\_name**: Add model name, it can be anything.
- **auth\_token**: Add auth token if required.

Name	Description
<b>test_type</b> * required string (query)	fairness
<b>dataset_name</b> * required string (query)	dataset_name
<b>model_url</b> * required string (query)	http://inhouse_model_url/inference
model_name string (query)	inhouse_llm
auth_token string (query)	NA

Note: If we have different input payload or response structure, we need to change it inside the code.

### Evaluation of the responses (GPU required)

/api/v1/trustllm/evaluation

- **task\_type:** Enter any one of the following:  
[privacy, fairness, ethics, safety, truthfulness]
- **model\_name:** enter the model\_name.
- **dataset\_name:** Enter the dataset that we have generated in first step. Don't upload file if already entered dataset\_name.
- **file:** One can upload the generated dataset file as well in zip format. Don't enter dataset\_name if already uploaded the generated file.
- **save\_to\_db:** true if wanted to save in a mongodb or false.

**Response will be a Json format.**

Parameters

Cancel

Reset

No parameters

Request body required

multipart/form-data

task\_type required

string

fairness

model\_name required

string

mistral-7b

dataset\_name

string

dataset\_07242024131340

☐ Send empty value

file

string(\$binary)

Choose File

No file chosen

☒ Send empty value

save\_to\_db

string

true

☐ Send empty value

#### FAQ:

Where to get the generated dataset?

Ans: Go to endpoint `/api/v1/trustllm/get/generatedDataset`, present in **utils** and add a `dataset_name` generated.

Or one can go to `src/generated_results` to get the `generated_results` as well.