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.

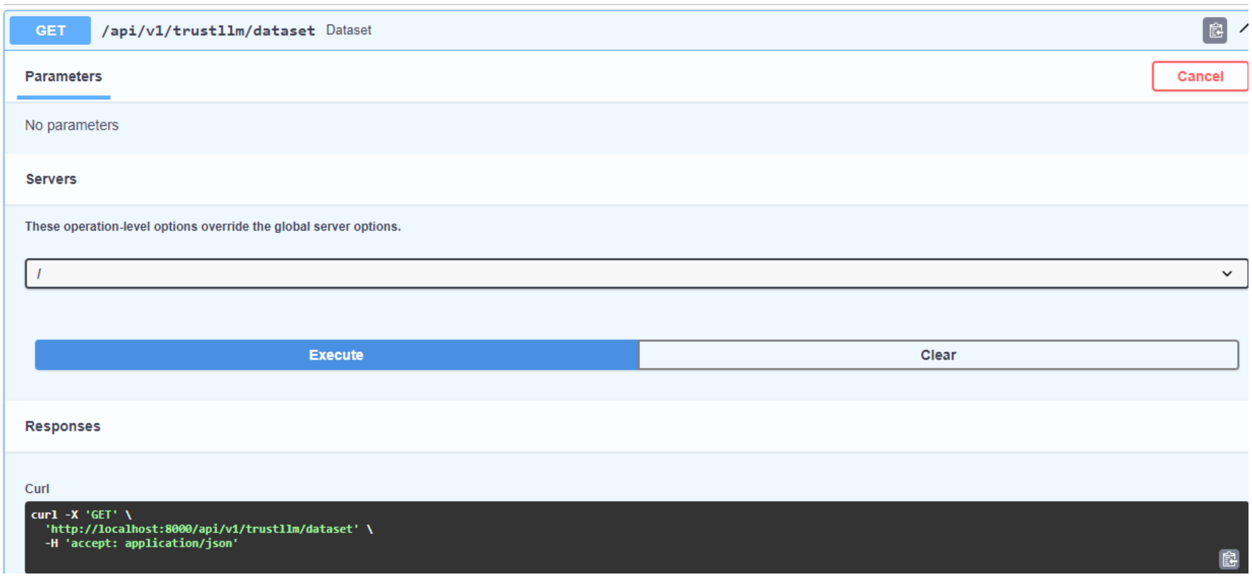
Swagger url: http://localhost:8000/api/v1/trustllm/docs

Steps for generating responses:

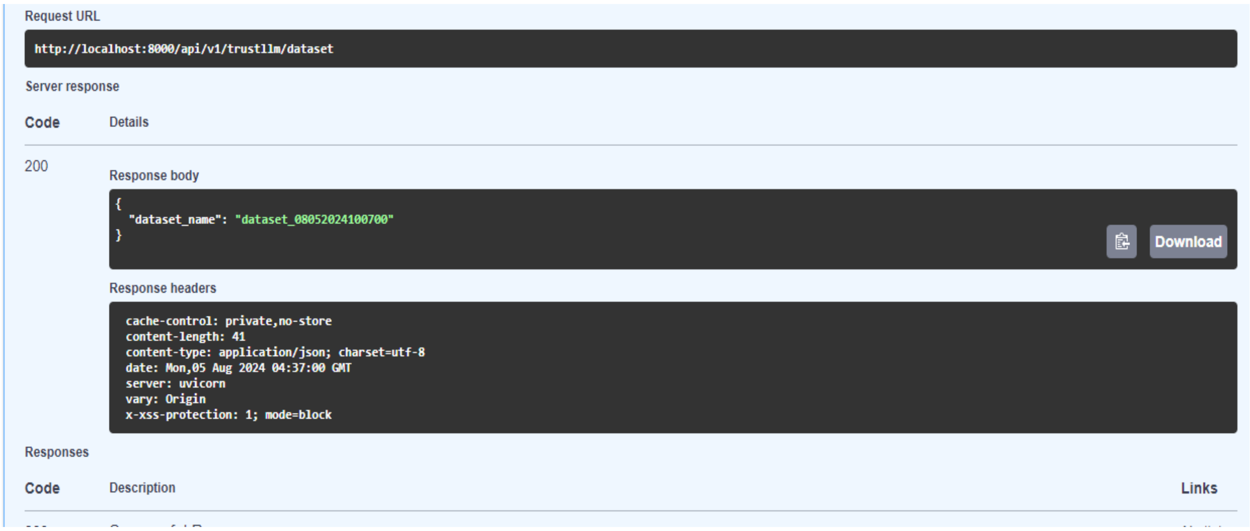
Create a dataset:

1) /api/v1/trustllm/dataset.

2)Copy the dataset name generated.



```
curl -X 'GET' \
  'http://localhost:8000/api/v1/trustllm/dataset' \
  -H 'accept: application/json'
```



```
http://localhost:8000/api/v1/trustllm/dataset
```

```
{
  "dataset_name": "dataset_08052024100700"
}
```

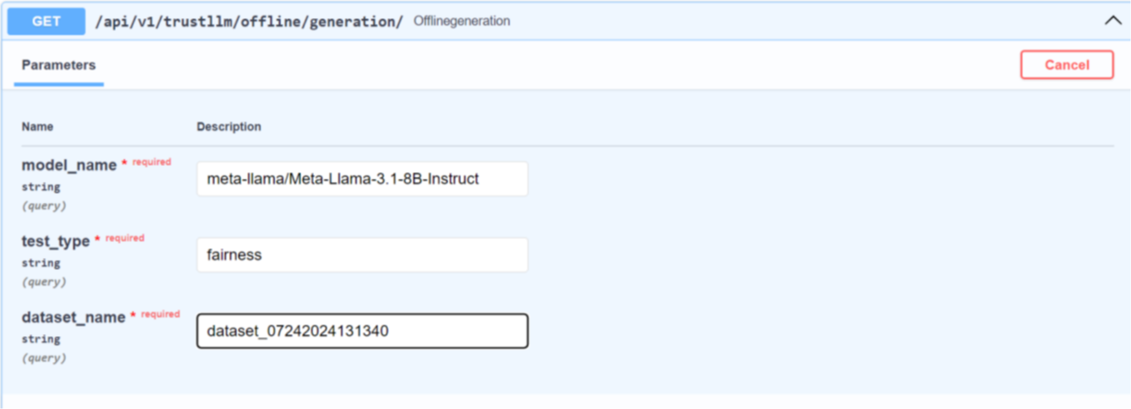
```
cache-control: private,no-store
content-length: 41
content-type: application/json; charset=utf-8
date: Mon,05 Aug 2024 04:37:00 GMT
server: uvicorn
vary: Origin
x-xss-protection: 1; mode=block
```

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.

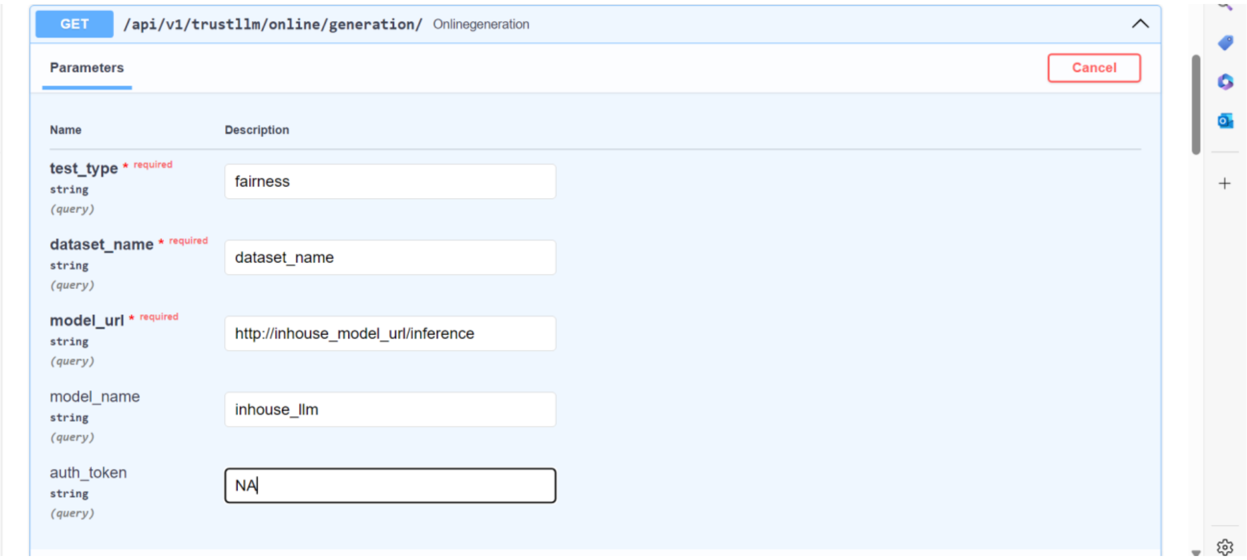


Name	Description
model_name * required string *(query)*	meta-llama/Meta-Llama-3.1-8B-Instruct
test_type * required string *(query)*	fairness
dataset_name * required string *(query)*	dataset_07242024131340

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.



/api/v1/trustllm/online/generation/ Onlinegeneration

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

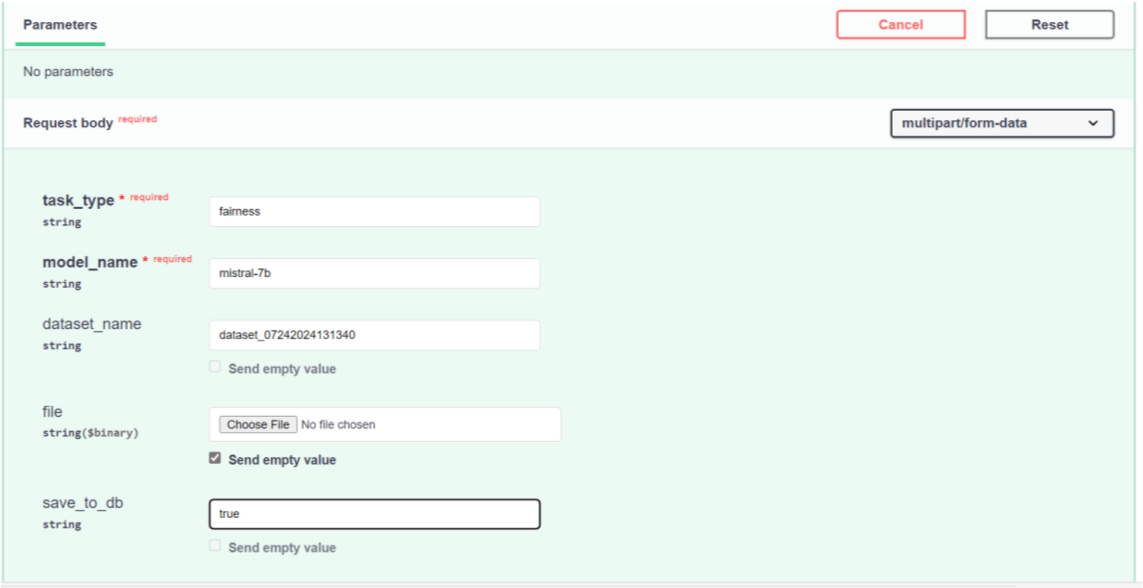
Note: If we have different input payload or response structure, we need to change it inside the code.

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 enterd 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.



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.