

Servers

/api

default

POST

/deploy/deploy Deploy Model Endpoint

Deploy the specified model to the target device.

Parameters:

- **model_name**: Name of the model to deploy.
- **device_ip**: IP address of the device where the model should be deployed.
- **model_format**: The format of the model (e.g., 'onnx', 'tensorflow').
- **deployment_path**: Path on the device where the model should be deployed.
- **device_username**: Username for SSH login to the device.
- **is_docker**: Whether the model should be deployed as a Docker container (default: False).

Returns:

- JSON response with deployment status and message.

Parameters

Try it out

No parameters

Request body **required**

application/json

Example Value | Schema

```
{
  "model_name": "string",
  "device_ip": "string",
  "model_format": "string",
  "deployment_path": "string",
  "device_username": "string",
  "is_docker": false
}
```

}

Responses

Code	Description	Links
200	<div>Successful Response</div> <div>Media type</div> <div>application/json</div> <div>Controls Accept header.</div> <div>Example Value Schema</div> <div>"string"</div>	No links
422	<div>Validation Error</div> <div>Media type</div> <div>application/json</div> <div>Example Value Schema</div> <div>{ "detail": [{ "loc": ["string",], "msg": "string", "type": "string" }] }</div>	No links

POST

/deploy/backup Backup Model



POST

/deploy/transfer Transfer Model



POST

/deploy/deploy/docker Deploy Model In Docker



POST

/deploy/update

Update Model On Device

▼

POST

/deploy/rollback

Rollback Model On Device

▼

GET

/monitoring/monitor/system_resources

Monitor System Resources

▼

GET

/monitoring/monitor/model_performance

Monitor Model Performance

▼

GET

/monitoring/monitor/device_metrics

Monitor Device Metrics

▼

POST

/monitoring/start-monitoring/

Start Monitoring

▼

POST

/monitoring/stop-monitoring/

Stop Monitoring

▼

GET

/monitoring/monitoring-status/

Get Monitoring Status

▼

GET

/monitoring/monitoring-data/

Get Monitoring Data

▼

POST

/models/load_model

Load Model

▼

GET

/models/list_models

List Models

⬆

Parameters

Cancel

No parameters

Servers

These operation-level options override the global server options.

/api

▼

Execute

Clear

Responses

Curl

```
curl -X 'GET' \
  'http://localhost/api/models/list_models' \
  -H 'accept: application/json'
```



Request URL

```
http://localhost/api/models/list_models
```

Server response

Code	Details
------	---------

200

Response body

```
{
  "models": [
    {
      "model_name": "assembly_line_optimization",
      "framework_type": "lightgbm",
      "model_path": "/models/assembly_line_optimization.txt",
      "version": "2.0.0",
      "description": "LightGBM model for optimizing assembly line throughput and efficiency.",
      "performance_metrics": {
        "rmse": 0.8,
        "mae": 0.6,
        "r2_score": 0.95
      },
      "created_at": "2023-03-05T11:20:00Z",
      "updated_at": "2023-03-12T12:50:00Z",
      "additional_params": {
        "num_leaves": 31,
        "learning_rate": 0.1,
        "n_estimators": 200
      }
    },
    {
      "model_name": "bert_sentiment_analysis",
      "framework_type": "nnpx"
```



Download

Response headers

```
connection: keep-alive
content-length: 8234
content-type: application/json
date: Wed, 30 Oct 2024 21:55:35 GMT
server: nginx/1.27.2
```

Responses

Code	Description	Links
200	Successful Response	No links

Media type

application/json ▼

Controls Accept header.

Example Value | Schema

```
"string"
```

POST

/models/predict Predict



DELETE

/models/remove_model Remove Model



POST

/models/save_model Save Model



POST

/models/model/train Train Model Endpoint



POST

/models/model/evaluate Evaluate Model Endpoint



POST

/models/model/update Update Model Endpoint



POST

/models/model/rollback Rollback Model Endpoint



POST

/logging/log/deployment Log Deployment



POST

/logging/log/monitoring Log Monitoring



GET

/logging/log/query Query Logs



GET

/ Read Root



GET

/health Health Check





BackupModelRequest > Expand all object

DeployDockerModelRequest > Expand all object

DeployModelRequest > Expand all object

EvaluateModelRequest > Expand all object

HTTPValidationError > Expand all object

LogEventRequest > Expand all object

MonitoringEventRequest > Expand all object

TrainModelRequest > Expand all object

TransferModelRequest > Expand all object

ValidationError > Expand all object

RollbackModelRequest > Expand all object

UpdateModelRequest > Expand all object

RollbackModelRequest > Expand all object

UpdateModelRequest > Expand all object

LoadModelRequest > Expand all object

PredictRequest > Expand all object

RemoveModelRequest > Expand all object

SaveModelRequest > Expand all object