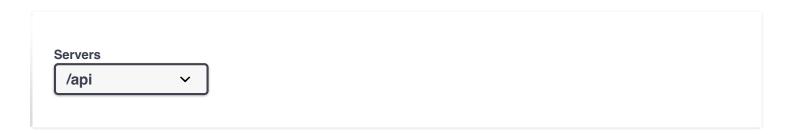


/api/openapi.json



## POST /deploy/deploy Deploy Model Endpoint / Deploy the specified model to the target device.

## Parameters:

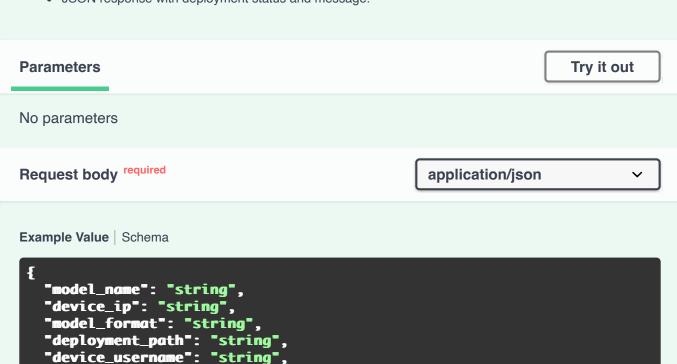
• model\_name: Name of the model to deploy.

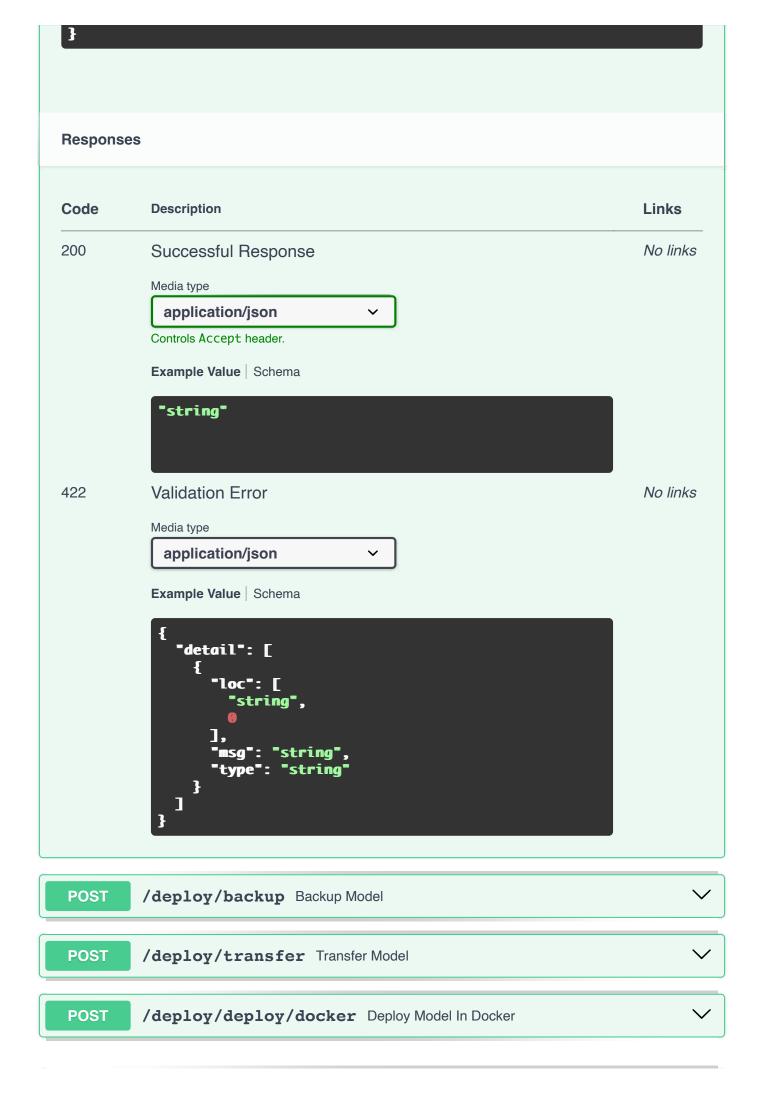
"is\_docker": false

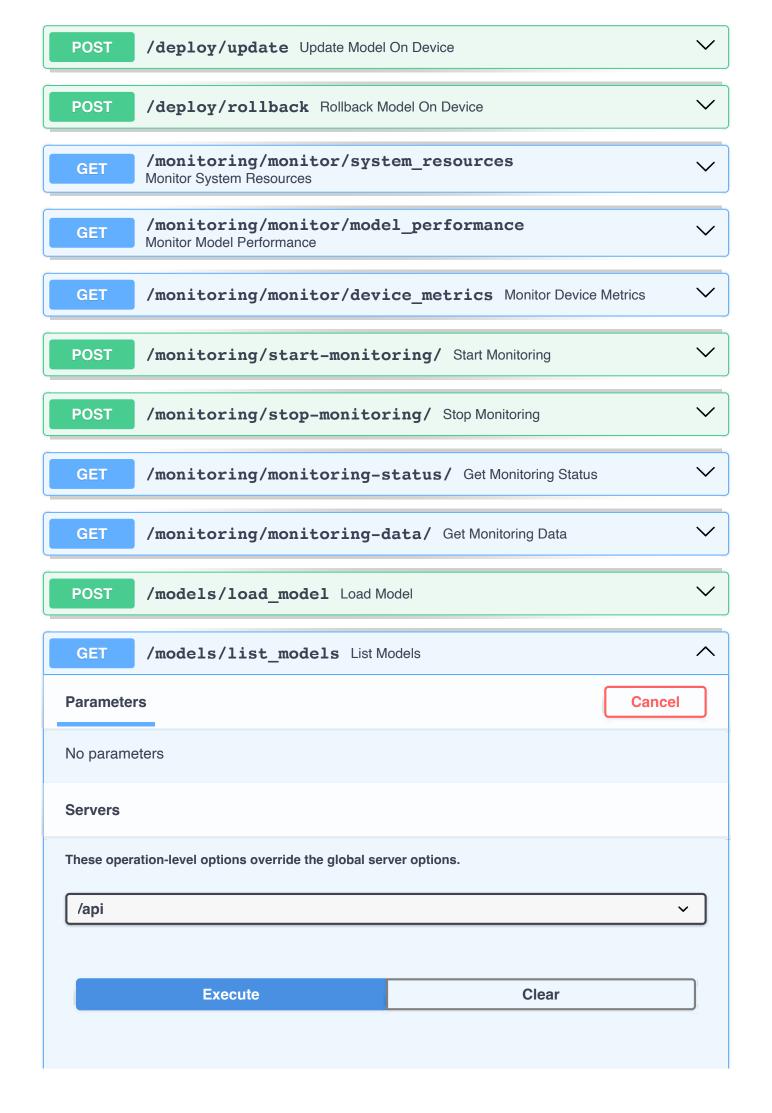
- device\_ip: IP address of the device where the model should be deployed.
- model\_format: The format of the model (e.g., 'onnx', 'tensorflow').
- deployment\_path: Path on the device where the model should be deployed.
- device\_username: Username for SSH login to the device.
- is\_docker: Whether the model should be deployed as a Docker container (default: False).

## Returns:

• JSON response with deployment status and message.







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

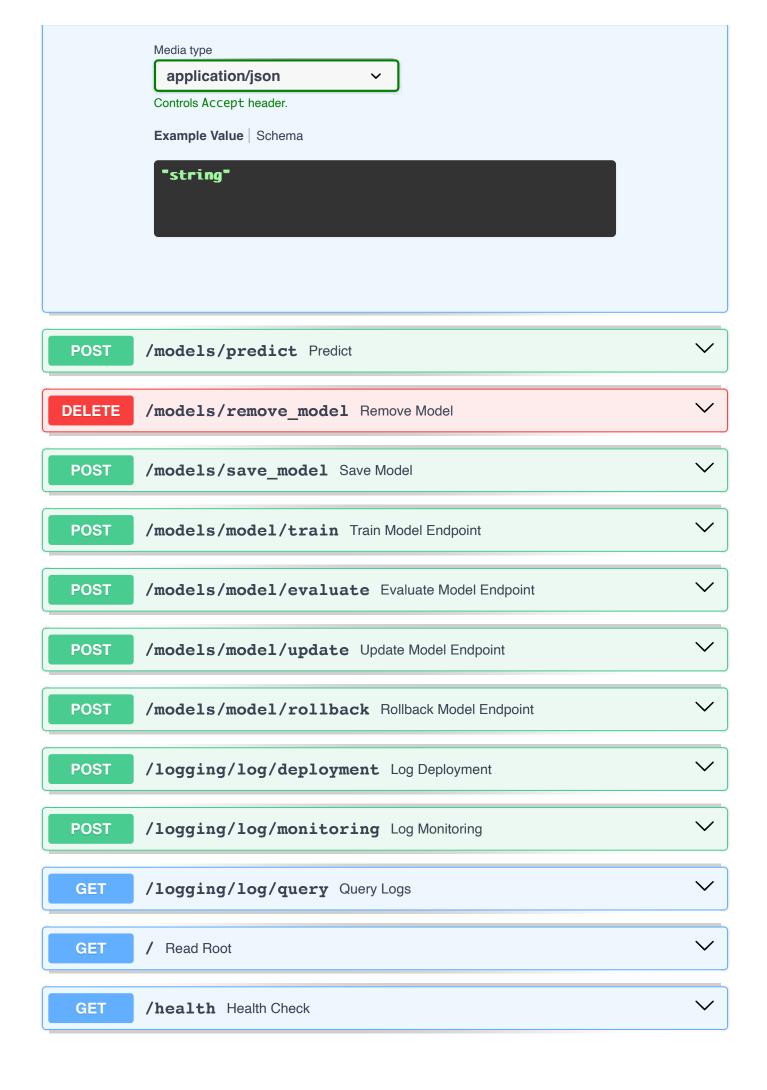
```
"models": [
      "model_name": "assembly_line_optimization",
      "framework_type": "lightgbm",
      "model_path": "/models/assembly_line_optimization.txt",
      "version": "2.0.0",
      "description": "LightGBM model for optimizing assembly
line throughput and efficiency.",
      "performance_metrics": {
        "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
      }
                                                     Download
      "model_name": "bert_sentiment_analysis",
      "framework type": "oppx"
```

**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



```
Schemas
   BackupModelRequest > Expand all object
   DeployDockerModelRequest >
   DeployModelRequest >
                                   object
   EvaluateModelRequest >
                                     object
   HTTPValidationError >
   LogEventRequest > Expand all
                                object
   MonitoringEventRequest >
                                      object
   TrainModelRequest >
                                 object
   TransferModelRequest >
   ValidationError >
                             object
   RollbackModelRequest >
   UpdateModelRequest >
                                   object
   RollbackModelRequest >
                                     object
   UpdateModelRequest >
                                   object
   LoadModelRequest > Expand all object
```

PredictRequest > Expand all object

RemoveModelRequest > Expand all object

SaveModelRequest > Expand all object