

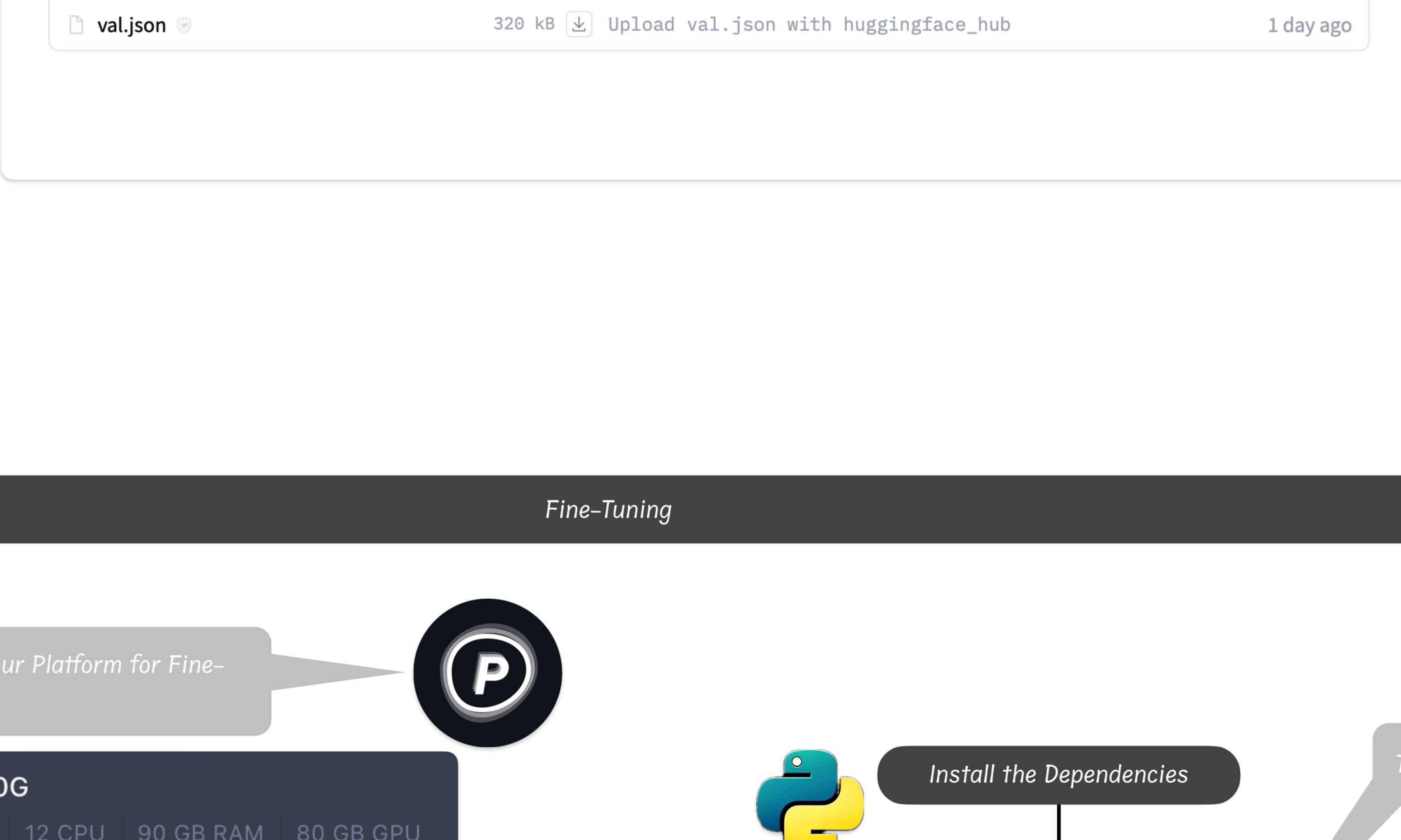
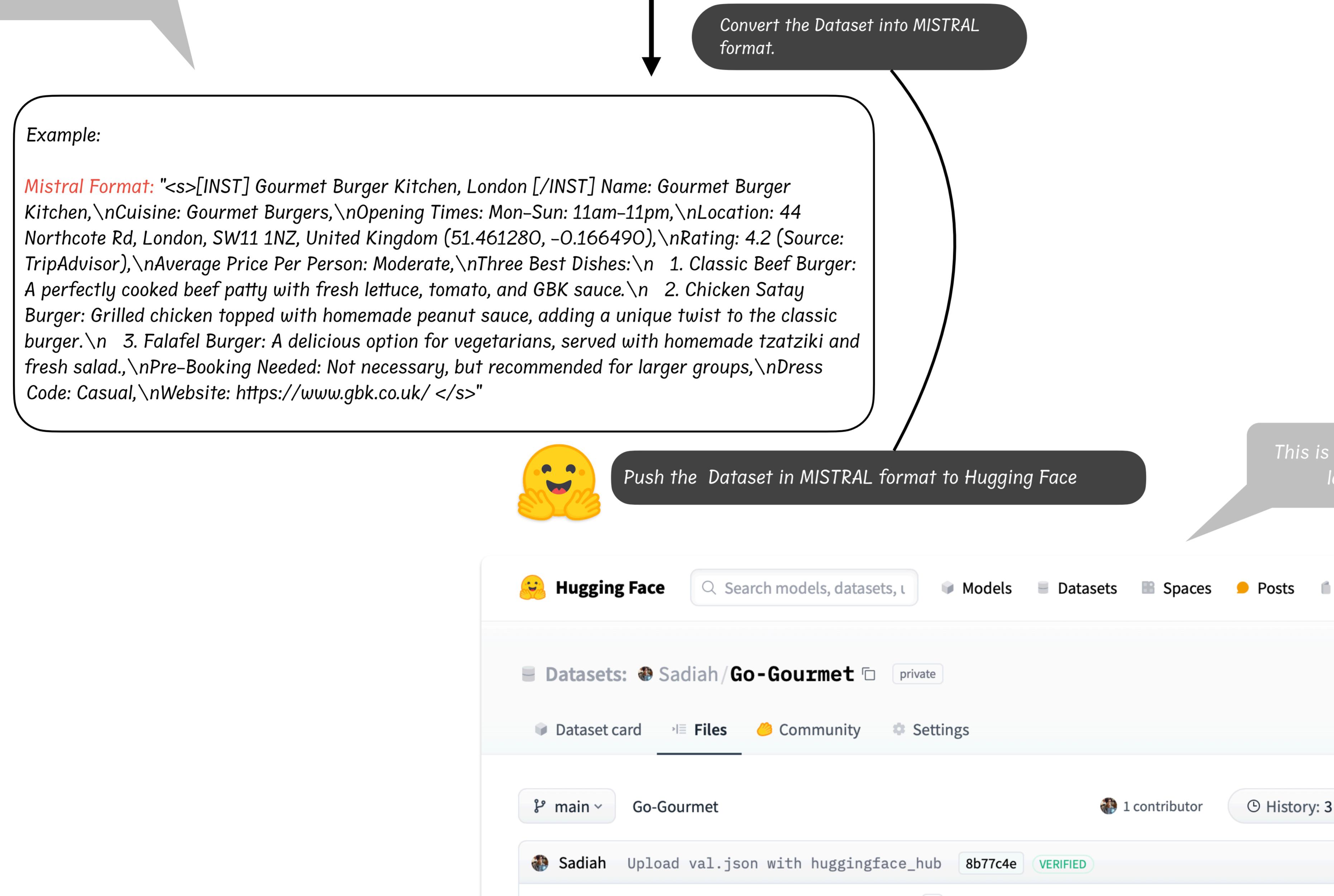
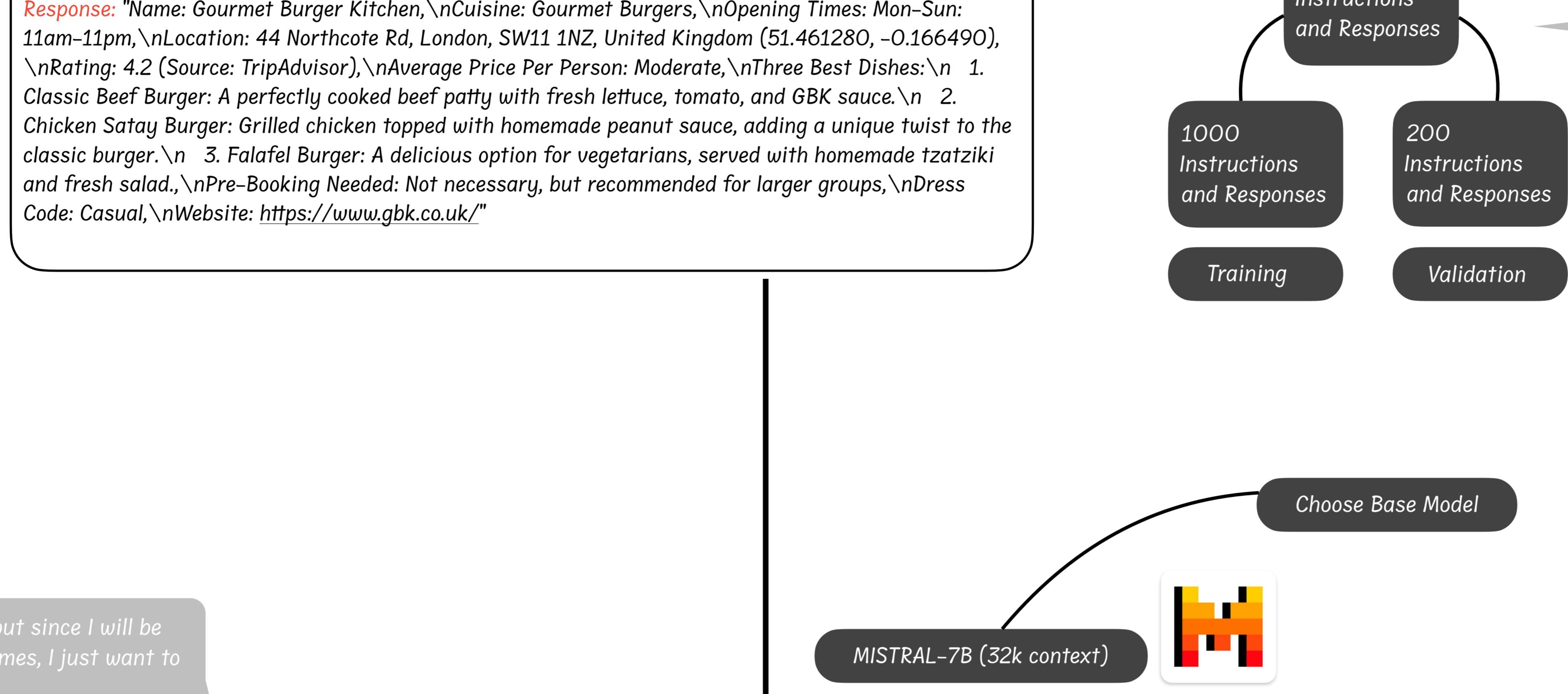


Go-Gourmet!

Fine-Tuning for uniform structure

Copyright © 2024 by @sadiyahzahoor.
All rights reserved.

Data-Preparation



Fine-Tuning

Choose your Platform for Fine-Tuning.



A100-80G
\$3.18/hr 12 CPU – 90 GB RAM – 80 GB GPU

Install the Dependencies



This is your Hugging Face Token

Pass on the credentials



Load the Base-Transformer model and its Tokeniser from Hugging Face with Quantization Configuration

mistral/MISTRAL-7B-Instruct-v0.2

load_in_4bit,...
bb_4bit_quant_type: ...
bnb_4bit_compute_dtype: ...
bnb_4bit_use_double_quant: ...
....add more

```
from transformers import AutoModelForCausalLM, AutoTokenizer
```

```
# Load the base model
```

```
model = AutoModelForCausalLM.from_pretrained(BASE_MODEL, trust_remote_code=True, device_map="auto")
```

```
# Load the entire model on the GPU
```

```
model.config.use_cache = False # Disable caching of model outputs
```

```
model.config.pretraining_tp = 1 # Tensor parallelism settings
```

```
# Load tokenizer
```

```
tokenizer = AutoTokenizer.from_pretrained(BASE_MODEL, trust_remote_code=True)
```

```
tokenizer.padding_side = "right" # Sequences are left indented
```

```
Tokens have not been saved to git credential helper. Pass add_to_git_credential=True if you want to set the git credential as well.
```

```
Token has a valid (permission: write).
```

```
Your token has been saved to ./root/.cache/huggingface/token
```

```
Login successful
```

```
config.json: 100% 596/696 [00:00:00, 41.0KB/s]
```

```
model.safetensors.index.json: 100% 25.4/25.1k [00:00:00, 1.97MB/s]
```

```
Downloading shards: 0% 2/3 [00:41<00:20, 20.07s/R]
```

```
model-00001-of-00003.safetensors: 100% 4.94/4.94G [00:24<00:00, 320MB/s]
```

```
model-00002-of-00003.safetensors: 100% 5.00/5.00G [00:16<00:00, 305MB/s]
```

```
model-00003-of-00003.safetensors: 78% 3.53/4.54G [00:53<00:31, 32.4MB/s]
```

Load PEFT configuration for training

PEFT used: QLoRA

Load the Go-Gourmet Dataset from Hugging Face

Add Training Configuration

Batch Size: ...
Learning Rate: ...
Optimizer: ...
Epochs: ...
.....add more

Start training loop

Time taken- 31 min (approx)

```
[46/469 2:03:36, Epoch 14/15]
```