```
/Users/rohithkattamuri/PycharmProjects/pythonProject1
/.venv/bin/python /Users/rohithkattamuri/
PycharmProjects/pythonProject1/sample.py
/Users/rohithkattamuri/PycharmProjects/pythonProject1
/.venv/lib/python3.9/site-packages/urllib3/__init__.py
:35: NotOpenSSLWarning: urllib3 v2 only supports
OpenSSL 1.1.1+, currently the 'ssl' module is compiled
 with 'LibreSSL 2.8.3'. See: https://github.com/
urllib3/urllib3/issues/3020
  warnings.warn(
Loading checkpoint shards: 100%|████████████| 3/3 [00:16
<00:00,  5.44s/it]
Some parameters are on the meta device because they
were offloaded to the disk.
MistralForCausalLM(
  (model): MistralModel(
    (embed_tokens): Embedding(32768, 4096)
    (layers): ModuleList(
      (0-31): 32 x MistralDecoderLayer(
        (self_attn): MistralSdpaAttention(
          (q_proj): Linear(in_features=4096,
out_features=4096, bias=False)
          (k_proj): Linear(in_features=4096,
out_features=1024, bias=False)
          (v_proj): Linear(in_features=4096,
out_features=1024, bias=False)
          (o_proj): Linear(in_features=4096,
out_features=4096, bias=False)
          (rotary_emb): MistralRotaryEmbedding()
        )
        (mlp): MistralMLP(
          (gate_proj): Linear(in_features=4096,
out_features=14336, bias=False)
          (up_proj): Linear(in_features=4096,
out_features=14336, bias=False)
          (down_proj): Linear(in_features=14336,
out_features=4096, bias=False)
          (act_fn): SiLU()
        )
        (input_layernorm): MistralRMSNorm((4096,), eps
=1e-05)
```

```
        (post_attention_layernorm): MistralRMSNorm((
4096,), eps=1e-05)
      )
    )
    (norm): MistralRMSNorm((4096,), eps=1e-05)
  )
  (lm_head): Linear(in_features=4096, out_features=
32768, bias=False)
) MistralConfig {
  "_attn_implementation_autoset": true,
  "_name_or_path": "mistralai/Mistral-7B-Instruct-v0.3
",
  "architectures": [
    "MistralForCausalLM"
  ],
  "attention_dropout": 0.0,
  "bos_token_id": 1,
  "eos_token_id": 2,
  "head_dim": 128,
  "hidden_act": "silu",
  "hidden_size": 4096,
  "initializer_range": 0.02,
  "intermediate_size": 14336,
  "max_position_embeddings": 32768,
  "model_type": "mistral",
  "num_attention_heads": 32,
  "num_hidden_layers": 32,
  "num_key_value_heads": 8,
  "rms_norm_eps": 1e-05,
  "rope_theta": 1000000.0,
  "sliding_window": null,
  "tie_word_embeddings": false,
  "torch_dtype": "float16",
  "transformers_version": "4.46.3",
  "use_cache": true,
  "vocab_size": 32768
}


Asking to truncate to max_length but no maximum length
 is provided and the model has no predefined maximum
length. Default to no truncation.
```

```
huggingface/tokenizers: The current process just got
forked, after parallelism has already been used.
Disabling parallelism to avoid deadlocks...
To disable this warning, you can either:
    - Avoid using `tokenizers` before the fork if
possible
    - Explicitly set the environment variable
TOKENIZERS_PARALLELISM=(true | false)
xxxxxxxxxxxxxTraceback (most recent call last):
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/sample.py", line 304, in <module>
    generated_text = generate_text(input_text)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/sample.py", line 251, in generate_text
    generated_tokens = model.generate(
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/utils/_contextlib.py", line 116, in decorate_context
    return func(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
transformers/generation/utils.py", line 2215, in
generate
    result = self._sample(
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
transformers/generation/utils.py", line 3206, in
_sample
    outputs = self(**model_inputs, return_dict=True)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1736, in
_wrapped_call_impl
    return self._call_impl(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1747, in _call_impl
    return forward_call(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
accelerate/hooks.py", line 170, in new_forward
```

```
    output = module._old_forward(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
transformers/models/mistral/modeling_mistral.py", line
 1065, in forward
    outputs = self.model(
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1736, in
_wrapped_call_impl
    return self._call_impl(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1747, in _call_impl
    return forward_call(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
transformers/models/mistral/modeling_mistral.py", line
 791, in forward
    layer_outputs = decoder_layer(
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1736, in
_wrapped_call_impl
    return self._call_impl(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1747, in _call_impl
    return forward_call(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
accelerate/hooks.py", line 170, in new_forward
    output = module._old_forward(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
transformers/models/mistral/modeling_mistral.py", line
 528, in forward
    hidden_states, self_attn_weights,
present_key_value = self.self_attn(
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
```

```
/nn/modules/module.py", line 1736, in
_wrapped_call_impl
    return self._call_impl(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1747, in _call_impl
    return forward_call(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
accelerate/hooks.py", line 170, in new_forward
    output = module._old_forward(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
transformers/models/mistral/modeling_mistral.py", line
 422, in forward
    query_states = self.q_proj(hidden_states)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1736, in
_wrapped_call_impl
    return self._call_impl(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/nn/modules/module.py", line 1747, in _call_impl
    return forward_call(*args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
accelerate/hooks.py", line 165, in new_forward
    args, kwargs = module._hf_hook.pre_forward(module
, *args, **kwargs)
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
accelerate/hooks.py", line 336, in pre_forward
    value = self.weights_map[name]
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
accelerate/utils/offload.py", line 118, in __getitem__
    return self.dataset[f"{self.prefix}{key}"]
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/
accelerate/utils/offload.py", line 170, in __getitem__
```

```
    with safe_open(weight_info["safetensors_file"],
framework="pt", device=device) as f:
  File "/Users/rohithkattamuri/PycharmProjects/
pythonProject1/.venv/lib/python3.9/site-packages/torch
/storage.py", line 404, in untyped
    def untyped(self):
KeyboardInterrupt

Process finished with exit code 130 (interrupted by
signal 2:SIGINT)
```