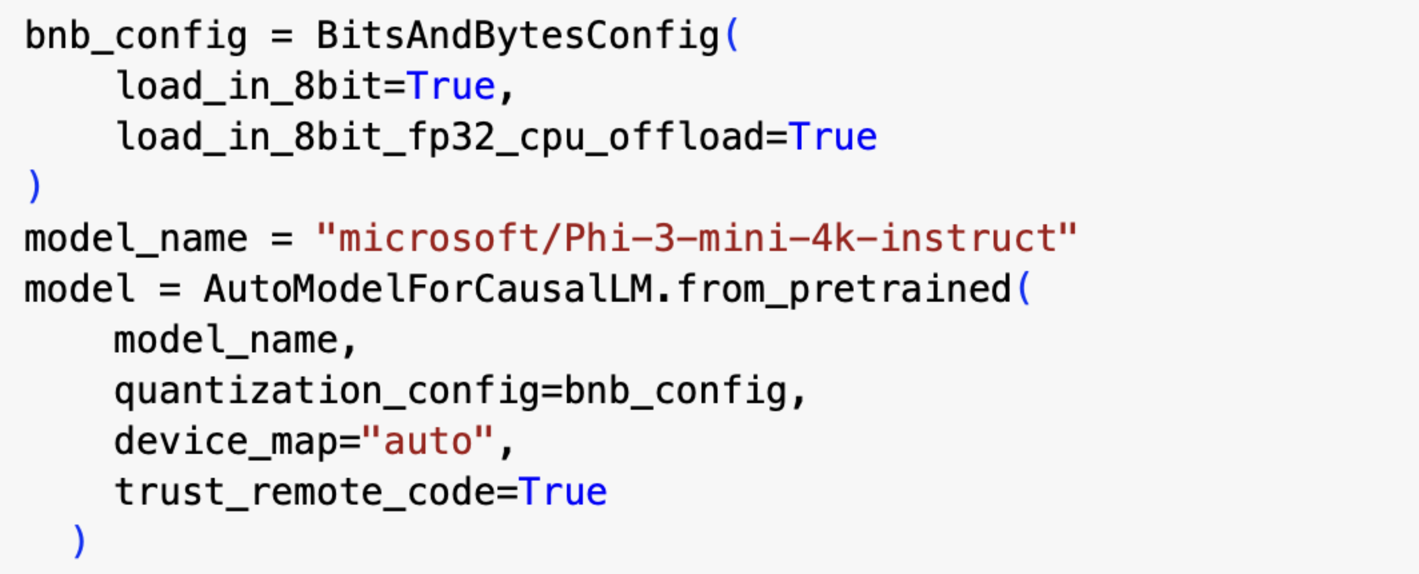

```
bnb_config = BitsAndBytesConfig(
    load_in_8bit=True,
    load_in_8bit_fp32_cpu_offload=True
)
model_name = "microsoft/Phi-3-mini-4k-instruct"
model = AutoModelForCausalLM.from_pretrained(
    model_name,
    quantization_config=bnb_config,
    device_map="auto",
    trust_remote_code=True
  )
```