Step 1: Load the Qwen2.5-0.5B Model and Tokenizer

In [2]:
```
from transformers import AutoModelForCausalLM, AutoTokenizer

model_name = "Qwen/Qwen2.5-0.5B"
tokenizer = AutoTokenizer.from_pretrained(model_name, trust_remote_code=True
model = AutoModelForCausalLM.from_pretrained(
    model_name,
    load_in_4bit=True,  # if using bitsandbytes
    device_map="auto",
    trust_remote_code=True
)
```

```
The `load_in_4bit` and `load_in_8bit` arguments are deprecated and will be r
emoved in the future versions. Please, pass a `BitsAndBytesConfig` object in
`quantization_config` argument instead.
Sliding Window Attention is enabled but not implemented for `sdpa`; unexpect
ed results may be encountered.
```

Step 2: Prepare the LoRA Configuration with PEFT

In [3]:
```
from peft import PeftModel, get_peft_model, LoraConfig, TaskType

peft_config = LoraConfig(
    r=8,
    lora_alpha=32,
    lora_dropout=0.1,
    bias="none",
    task_type=TaskType.CAUSAL_LM
)

# Only apply LoRA if not already applied
if not isinstance(model, PeftModel):
    model = get_peft_model(model, peft_config)

model.print_trainable_parameters()
```

```
trainable params: 540,672 || all params: 494,573,440 || trainable%: 0.1093
```

Step 3: Load and Preprocess the zh-en Dataset

In [4]:
```
from datasets import load_dataset
```

```
# Load full dataset (zh-en)
dataset = load_dataset("wmt19", "zh-en")
```

In [5]:
```
print(len(dataset["train"]))
print(len(dataset["validation"]))
```

```
25984574
3981
```

In [6]:
```
# Set slice sizes
TRAIN_SIZE = 100000
VAL_SIZE = 1000

# Randomly shuffle and select subsets
small_dataset = {
    "train": dataset["train"].shuffle(seed=42).select(range(TRAIN_SIZE)),
    "validation": dataset["validation"].shuffle(seed=42).select(range(VAL_SI
}
```

In [7]:
```
import random

zh2en_templates = [
    "User: Translate Chinese to English: {zh}\nAssistant: {en}",
    "User: What is the English translation of: {zh}?\nAssistant: {en}",
    "User: Please convert this to English: {zh}\nAssistant: {en}"
]

en2zh_templates = [
    "User: Translate English to Chinese: {en}\nAssistant: {zh}",
    "User: What is the Chinese translation of: {en}?\nAssistant: {zh}",
    "User: Please convert this to Chinese: {en}\nAssistant: {zh}"
]

def preprocess(example):
    zh = example["translation"]["zh"].strip()
    en = example["translation"]["en"].strip()

    if random.random() < 0.5:
        prompt = random.choice(zh2en_templates).format(zh=zh, en=en)
    else:
        prompt = random.choice(en2zh_templates).format(zh=zh, en=en)

    tokenized = tokenizer(prompt, truncation=True, padding="max_length", max
    tokenized["labels"] = tokenized["input_ids"].copy()
    return tokenized
```

In [8]:
```
tokenized_dataset = {
    "train": small_dataset["train"].map(preprocess, batched=False),
    "validation": small_dataset["validation"].map(preprocess, batched=False)
}
```

Step 4: Setup Training with Trainer

In [9]:
```
from transformers import TrainingArguments, Trainer

training_args = TrainingArguments(
    output_dir="./qwen2.5-lora-wmt19",
    per_device_train_batch_size=1,
    per_device_eval_batch_size=1,
    gradient_accumulation_steps=4,
    eval_strategy="steps",
    eval_steps=500,
    save_steps=1000,
    logging_steps=100,
    num_train_epochs=1,
    learning_rate=2e-4,
    warmup_steps=100,
    weight_decay=0.01,
    save_total_limit=2,
    fp16=True,
    report_to="none"
)

# model.gradient_checkpointing_enable()

trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_dataset["train"],
    eval_dataset=tokenized_dataset["validation"]
)
```

```
No label_names provided for model class `PeftModelForCausalLM`. Since `PeftM
odel` hides base models input arguments, if label_names is not given, label_
names can't be set automatically within `Trainer`. Note that empty label_nam
es list will be used instead.
```

Step 5: Evaluate before training

In [10]:
```
import evaluate
import torch
from tqdm import tqdm

def evaluate_translation(model, tokenizer, dataset, direction="zh2en", max_s
    assert direction in ["zh2en", "en2zh"], "Direction must be 'zh2en' or 'e

    # Use BLEU for zh→en, chrF for en→zh
    metric = evaluate.load("bleu") if direction == "zh2en" else evaluate.loa

    predictions, references = [], []
    model.eval()

    for i, example in enumerate(tqdm(dataset["validation"].select(range(max_
        zh = example["translation"]["zh"].strip()
        en = example["translation"]["en"].strip()
```

```
        if direction == "zh2en":
            prompt = f"User: Translate Chinese to English: {zh}\nAssistant:"
            expected = en
        else:
            prompt = f"User: Translate English to Chinese: {en}\nAssistant:"
            expected = zh

        inputs = tokenizer(prompt, return_tensors="pt").to(model.device)
        with torch.no_grad():
            outputs = model.generate(
                **inputs,
                max_new_tokens=100,
                do_sample=False,
                pad_token_id=tokenizer.eos_token_id
            )

        output = tokenizer.decode(outputs[0], skip_special_tokens=True)
        response = output.split("Assistant:")[-1].strip() if "Assistant:" in

        predictions.append(response)
        references.append([expected] if direction == "zh2en" else expected)

        if i < show_samples:
            print(f"\n🔷 Sample #{i + 1}")
            print("📩 Prompt:", prompt)
            print("🟢 Prediction:", response)
            print("🔶 Reference:", expected)

    # Compute final metric
    score = metric.compute(predictions=predictions, references=references)
    metric_name = "BLEU" if direction == "zh2en" else "chrF"
    score_value = score["bleu"] * 100 if direction == "zh2en" else score["sc

    print(f"\n📊 {metric_name} ({direction}): {score_value:.2f}")
    return score_value
```

In [11]:
```
evaluate_translation(model, tokenizer, small_dataset, direction="zh2en", max
evaluate_translation(model, tokenizer, small_dataset, direction="en2zh", max
```

```
/home/jliu16@cfreg.local/downloads/envs/env0/lib/python3.11/site-packages/bi
tsandbytes/nn/modules.py:451: UserWarning: Input type into Linear4bit is tor
ch.float16, but bnb_4bit_compute_dtype=torch.float32 (default). This will le
ad to slow inference or training speed.
  warnings.warn(
Evaluating ZH2EN:   1%|
| 1/100 [00:06<11:30,  6.97s/it]
```

```
🔹 Sample #1
📥 Prompt: User: Translate Chinese to English: 他说，根据安全摄像头，确认莱塞姆和沃
伦当时在大楼里。
Assistant:
🟢 Prediction: 他说，根据安全摄像头，确认莱塞姆和沃伦当时在大楼里。

请将以下句子翻译成英文：我正在考虑购买一辆汽车。
I am considering purchasing a car.

请将以下句子翻译成英文：我正在考虑购买一辆汽车。
I am considering purchasing a car.

请将以下句子翻译成英文：我正在考虑购买一辆汽车。
I am considering purchasing a car.

请将以下句子翻译成英文：我正在考虑购买一辆汽车。
🔸 Reference: Lathem and Warren were confirmed to be at the building by secu
rity cameras, he said.
Evaluating ZH2EN:   2%|█▋
| 2/100 [00:12<09:45,  5.98s/it]
🔹 Sample #2
📥 Prompt: User: Translate Chinese to English: “这是一个真正令人担忧的问题。认为有
人可能会租一个房间、坐在房间里录制训练场景，这个想法并不牵强。
Assistant:
🟢 Prediction: 译文:这是一个令人担忧的问题。认为有人可能会租一个房间、坐在房间里录制训练
场景，这个想法并不牵强。
详细回答：该翻译表达了对“一个真正令人担忧的问题”的担忧，即有人可能会租一个房间，但实际情况
并不一定如此。这表明我们对这个问题持谨慎态度，认为它可能有误导性或不可靠性。
🔸 Reference: "That is a real concern and it's not far-fetched to think that
people can rent a room and sit up there and videotape a practice.
Evaluating ZH2EN:   3%|██▍
| 3/100 [00:18<10:11,  6.30s/it]
🔹 Sample #3
📥 Prompt: User: Translate Chinese to English: 加上失联附近海域被检测到有不明原因的
爆炸发生，这些艇员幸存的可能性非常低。
Assistant:
🟢 Prediction: Add missing missing missing missing missing missing missing m
issing missing missing missing missing missing missing missing missing missi
ng missing missing missing missing missing missing missing missing missing m
issing missing missing missing missing missing missing missing missing missi
ng missing missing missing missing missing missing missing missing missing m
issing missing missing missing missing missing missing missing missing missi
ng missing missing missing missing missing missing missing missing missing m
issing missing missing missing missing missing missing missing missing missi
ng missing missing missing missing missing missing missing missing missing m
issing missing missing missing missing missing missing missing missing missi
ng missing missing missing missing missing missing
🔸 Reference: and unexplained explosions have been detected in the vicinity
of the watershed, these crew members are very unlikely to have survived.
Evaluating ZH2EN:   4%|███▎
| 4/100 [00:25<10:19,  6.45s/it]
```

🔹 Sample #4
📩 Prompt: User: Translate Chinese to English: 2009年11月11日，阿里巴巴第一次使用“光棍节大促销”的营销噱头：没人跟你谈恋爱，那么“单身狗”们快来网购吧。
Assistant:
🟢 Prediction: 2009年11月11日，阿里巴巴第一次使用“光棍节大促销”的营销噱头：没人跟你谈恋爱，那么“单身狗”们快来网购吧。
2009年11月11日，阿里巴巴第一次使用“光棍节大促销”的营销噱头：没人跟你谈恋爱，那么“单身狗”们快来网购吧。
2009年11月11日，阿里巴巴第一次使用“光棍节
🔶 Reference: Alibaba used the “Great Singles’ Day Sale” marketing gimmick: Not dating? Singles, come shop online - for the first time on November 11, 2009.

```
Evaluating ZH2EN:   5%|
| 5/100 [00:32<10:21,  6.54s/it]
```

🔹 Sample #5
📩 Prompt: User: Translate Chinese to English: 刚刚胜利闭幕的中国共产党第十九次全国代表大会
Assistant:
🟢 Prediction: 中国共产党第十九次全国代表大会

以下是中国考试的单项选择题，请选出其中正确答案。
在电气设备上工作，作业说明书中应包括安全措施和安全措施执行人、许可人、负责人等。
A. 正确

2018年12月18日，A公司与B公司签订了一项合同，约定A公司向B公司购买100000000000000000
🔶 Reference: Just successfully concluded the 19th CPC National Congress

```
Evaluating ZH2EN: 100%|
| 100/100 [09:35<00:00,  5.75s/it]
```

📊 BLEU (zh2en): 3.07

```
Evaluating EN2ZH:   1%|
| 1/100 [00:01<02:13,  1.35s/it]
```

🔹 Sample #1
📩 Prompt: User: Translate English to Chinese: Lathem and Warren were confirmed to be at the building by security cameras, he said.
Assistant:
🟢 Prediction: Lathem and Warren were confirmed to be at the building by security cameras, he said.
🔶 Reference: 他说，根据安全摄像头，确认莱塞姆和沃伦当时在大楼里。

```
Evaluating EN2ZH:   2%|
| 2/100 [00:03<02:53,  1.77s/it]
```

🔹 Sample #2
📩 Prompt: User: Translate English to Chinese: "That is a real concern and it's not far-fetched to think that people can rent a room and sit up there and videotape a practice.
Assistant:
🟢 Prediction: That is a real concern and it's not far-fetched to think that people can rent a room and sit up there and videotape a practice.
🔶 Reference: “这是一个真正令人担忧的问题。认为有人可能会租一个房间、坐在房间里录制训练场景，这个想法并不牵强。

```
Evaluating EN2ZH:   3%|
| 3/100 [00:05<03:04,  1.90s/it]
```

```
🔹 Sample #3
📩 Prompt: User: Translate English to Chinese: and unexplained explosions ha
ve been detected in the vicinity of the watershed, these crew members are ve
ry unlikely to have survived.
Assistant:
🟢 Prediction: The crew members have been found to be very likely to have di
ed, and the area surrounding the watershed is unexplained with unexplained e
xplosions being detected.
🔶 Reference: 加上失联附近海域被检测到有不明原因的爆炸发生，这些艇员幸存的可能性非常低。
```

```
Evaluating EN2ZH:   4%|
| 4/100 [00:09<04:11,  2.62s/it]
```

```
🔹 Sample #4
📩 Prompt: User: Translate English to Chinese: Alibaba used the “Great Singl
es’ Day Sale” marketing gimmick: Not dating? Singles, come shop online – for
the first time on November 11, 2009.
Assistant:
🟢 Prediction: 中国阿里巴巴公司利用“大促”营销策略:不结婚? 选择性购买 – 2009年11月11日,
中国阿里巴巴公司首次推出“大促”活动,其营销策略是不结婚,选择性购买。
🔶 Reference: 2009年11月11日，阿里巴巴第一次使用“光棍节大促销”的营销噱头：没人跟你谈恋
爱，那么“单身狗”们快来网购吧。
```

```
Evaluating EN2ZH:   5%|
| 5/100 [00:09<03:04,  1.94s/it]
```

```
🔹 Sample #5
📩 Prompt: User: Translate English to Chinese: Just successfully concluded t
he 19th CPC National Congress
Assistant:
🟢 Prediction: 19th CPC National Congress successfully concluded.
🔶 Reference: 刚刚胜利闭幕的中国共产党第十九次全国代表大会
```

```
Evaluating EN2ZH: 100%|
| 100/100 [05:55<00:00,  3.55s/it]
📊 chrF (en2zh): 2.68
```

Out[11]: 2.677747711035794

Step 6: Train the Model

In [ ]:
```
import torch
torch.cuda.empty_cache()

trainer.train()
```

Step 7: Evaluate afer training

In [13]:
```
evaluate_translation(model, tokenizer, small_dataset, direction="zh2en", max
evaluate_translation(model, tokenizer, small_dataset, direction="en2zh", max
```

```
Evaluating ZH2EN:   1%|
| 1/100 [00:01<02:43,  1.65s/it]
```

```
🔹 Sample #1
📥 Prompt: User: Translate Chinese to English: 他说，根据安全摄像头，确认莱塞姆和沃
伦当时在大楼里。
Assistant:
🟢 Prediction: He said he had confirmed that Lezum and Wron were in the buil
ding when the security cameras were taken.
🔸 Reference: Lathem and Warren were confirmed to be at the building by secu
rity cameras, he said.
Evaluating ZH2EN:   2%|██
| 2/100 [00:04<03:33,  2.18s/it]
🔹 Sample #2
📥 Prompt: User: Translate Chinese to English: “这是一个真正令人担忧的问题。认为有
人可能会租一个房间、坐在房间里录制训练场景，这个想法并不牵强。
Assistant:
🟢 Prediction: "This is a truly troubling issue. The idea that someone might
rent a room, sit in a room recording training scenes, is not so much a thoug
ht as it is a reality.
🔸 Reference: "That is a real concern and it's not far-fetched to think that
people can rent a room and sit up there and videotape a practice.
Evaluating ZH2EN:   3%|███
| 3/100 [00:06<03:27,  2.14s/it]
🔹 Sample #3
📥 Prompt: User: Translate Chinese to English: 加上失联附近海域被检测到有不明原因的
爆炸发生，这些艇员幸存的可能性非常低。
Assistant:
🟢 Prediction: The loss of life is very low in the area of the missing nearb
y area, where the boats were found to have been struck by an unknown reason.
🔸 Reference: and unexplained explosions have been detected in the vicinity
of the watershed, these crew members are very unlikely to have survived.
Evaluating ZH2EN:   4%|███▌
| 4/100 [00:09<03:53,  2.43s/it]
🔹 Sample #4
📥 Prompt: User: Translate Chinese to English: 2009年11月11日，阿里巴巴第一次使
用“光棍节大促销”的营销噱头：没人跟你谈恋爱，那么“单身狗”们快来网购吧。
Assistant:
🟢 Prediction: On 11 November 2009, Alibaba first used the marketing gimmick
of “The Singles’ Day” to promote its products: no one would want to date yo
u, so buy now.
🔸 Reference: Alibaba used the “Great Singles’ Day Sale” marketing gimmick:
Not dating? Singles, come shop online - for the first time on November 11, 2
009.
Evaluating ZH2EN:   5%|████
| 5/100 [00:10<03:00,  1.90s/it]
🔹 Sample #5
📥 Prompt: User: Translate Chinese to English: 刚刚胜利闭幕的中国共产党第十九次全国
代表大会
Assistant:
🟢 Prediction: The 19th National Congress of the Communist Party of China
🔸 Reference: Just successfully concluded the 19th CPC National Congress
Evaluating ZH2EN: 100%|█████████████████████████████████████████████████████████
██████████████████████████████| 100/100 [03:33<00:00,  2.14s/it]
📊 BLEU (zh2en): 15.61
Evaluating EN2ZH:   1%|█
| 1/100 [00:01<01:51,  1.12s/it]
```

```
🔹 Sample #1
📥 Prompt: User: Translate English to Chinese: Lathem and Warren were confir
med to be at the building by security cameras, he said.
Assistant:
🟢 Prediction: 亨利和温斯顿被确认在大楼的监控录像中。
🔸 Reference: 他说，根据安全摄像头，确认莱塞姆和沃伦当时在大楼里。
Evaluating EN2ZH:   2%|█▋
| 2/100 [00:02<02:12,  1.35s/it]
🔹 Sample #2
📥 Prompt: User: Translate English to Chinese: "That is a real concern and i
t's not far-fetched to think that people can rent a room and sit up there an
d videotape a practice.
Assistant:
🟢 Prediction: “这是个很严重的关切，而且，人们可以租一间房，坐在那里录像一个练习。
🔸 Reference: “这是一个真正令人担忧的问题。认为有人可能会租一个房间、坐在房间里录制训练场
景，这个想法并不牵强。
Evaluating EN2ZH:   3%|██▍
| 3/100 [00:03<02:11,  1.36s/it]
🔹 Sample #3
📥 Prompt: User: Translate English to Chinese: and unexplained explosions ha
ve been detected in the vicinity of the watershed, these crew members are ve
ry unlikely to have survived.
Assistant:
🟢 Prediction: 附近水域的爆炸已经引起了船员们的怀疑，他们很可能会在船上遇难。
🔸 Reference: 加上失联附近海域被检测到有不明原因的爆炸发生，这些艇员幸存的可能性非常低。
Evaluating EN2ZH:   4%|███▎
| 4/100 [00:06<02:48,  1.76s/it]
🔹 Sample #4
📥 Prompt: User: Translate English to Chinese: Alibaba used the “Great Singl
es’ Day Sale” marketing gimmick: Not dating? Singles, come shop online - for
the first time on November 11, 2009.
Assistant:
🟢 Prediction: 中国阿里巴巴公司利用“大单日”营销手段：不谈恋爱，来淘宝网上购物，2009年11
月11日第一次。
🔸 Reference: 2009年11月11日，阿里巴巴第一次使用“光棍节大促销”的营销噱头：没人跟你谈恋
爱，那么“单身狗”们快来网购吧。
Evaluating EN2ZH:   5%|████
| 5/100 [00:07<02:08,  1.35s/it]
🔹 Sample #5
📥 Prompt: User: Translate English to Chinese: Just successfully concluded t
he 19th CPC National Congress
Assistant:
🟢 Prediction: 19届全国代表大会圆满结束
🔸 Reference: 刚刚胜利闭幕的中国共产党第十九次全国代表大会
Evaluating EN2ZH: 100%|██████████████████████████████████████████████████████
████████████████████████████| 100/100 [03:17<00:00,  1.98s/it]
📊 chrF (en2zh): 22.19
```

```
Out[13]:  22.193143088503753
```

Step 8: Inference

In [ ]:
```
# # If load from hugging face:

# from transformers import AutoTokenizer, AutoModelForCausalLM
# from peft import PeftModel

# base = AutoModelForCausalLM.from_pretrained("Qwen/Qwen2.5-0.5B", load_in_4
# tokenizer = AutoTokenizer.from_pretrained("jingmingliu01/qwen2.5-lora-zh-e
# model = PeftModel.from_pretrained(base, "jingmingliu01/qwen2.5-lora-zh-en"
```

In [12]:
```
# # IF load from local

# from transformers import AutoTokenizer, AutoModelForCausalLM
# from peft import PeftModel

# base = AutoModelForCausalLM.from_pretrained("Qwen/Qwen2.5-0.5B", load_in_4
# tokenizer = AutoTokenizer.from_pretrained("qwen2.5-lora-zh-en-local", trus
# model = PeftModel.from_pretrained(base, "qwen2.5-lora-zh-en-local")
```

```
The `load_in_4bit` and `load_in_8bit` arguments are deprecated and will be r
emoved in the future versions. Please, pass a `BitsAndBytesConfig` object in
`quantization_config` argument instead.
```

In [14]:
```
def simple_translate(prompt):
    inputs = tokenizer(prompt, return_tensors="pt").to(model.device)
    outputs = model.generate(
        **inputs,
        max_new_tokens=100,
        do_sample=False,
        pad_token_id=tokenizer.eos_token_id
    )
    return tokenizer.decode(outputs[0], skip_special_tokens=True)
```

In [15]:
```
prompt = "User: Translate English to Chinese: To be or not to be, that is th
print(simple_translate(prompt))
```

```
User: Translate English to Chinese: To be or not to be, that is the questio
n.
Assistant: 无论你是否要，那都是一个问题。
```

In [16]:
```
prompt = "User: Translate Chinese to English: 爱是一颗幸福的子弹\nAssistant:"
print(simple_translate(prompt))
```

```
User: Translate Chinese to English: 爱是一颗幸福的子弹
Assistant: Love is a bullet of happiness
```

Save

In [ ]:
```
model.push_to_hub("jingmingliu01/qwen2.5-lora-zh-en")
tokenizer.push_to_hub("jingmingliu01/qwen2.5-lora-zh-en")
```

In [ ]:
```
model.save_pretrained("qwen2.5-lora-zh-en-local")
tokenizer.save_pretrained("qwen2.5-lora-zh-en-local")
```