

```
In [1]: import torch
        from datasets import load_dataset
        from transformers import AutoModelForCausalLM, AutoTokenizer, TrainingArguments
        from trl import DPOTrainer
        import wandb
```

WARNING: BNB_CUDA_VERSION=124 environment variable detected; loading libbitsandbytes_cuda124.dll.
This can be used to load a bitsandbytes version that is different from the PyTorch CUDA version.
If this was unintended set the BNB_CUDA_VERSION variable to an empty string: export BNB_CUDA_VERSION=
If you use the manual override make sure the right libcudart.so is in your LD_LIBRARY_PATH
For example by adding the following to your .bashrc: export LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:<path_to_cuda_dir/lib64

```
In [2]: # --- Device Configuration ---
        device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
        print(f"Using device: {device}")
```

Using device: cuda

```
In [3]: # --- Model and Tokenizer Setup ---
        model_name_or_path = "gpt2"
        ignore_bias_buffers = False

        model = AutoModelForCausalLM.from_pretrained(model_name_or_path)
        if ignore_bias_buffers:
            model._ddp_params_and_buffers_to_ignore = [name for name, buffer in model.named_buffers() if buffer.dtype == torch.bool]

        model_ref = AutoModelForCausalLM.from_pretrained(model_name_or_path)
        tokenizer = AutoTokenizer.from_pretrained(model_name_or_path)
        if tokenizer.pad_token is None:
            tokenizer.pad_token = tokenizer.eos_token
```

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\transformers\tokenization_utils_base.py:1617: FutureWarning: `clean_up_tokenization_spaces` was not set. It will be set to `True` by default. This behavior will be deprecated in transformers v4.45, and will be then set to `False` by default. For more details check this issue: <https://github.com/huggingface/transformers/issues/31884>
warnings.warn(

```
In [4]: # --- Dataset Loading and Prompt Extraction ---
        def create_prompt(sample):
            return {"prompt": sample["input"]}

        train_dataset = load_dataset("argilla/distilabel-intel-orca-dpo-pairs", split="train[:90%]")
        eval_dataset = load_dataset("argilla/distilabel-intel-orca-dpo-pairs", split="train[90%:]")

        # Extract the prompt
        train_dataset = train_dataset.map(create_prompt)
        eval_dataset = eval_dataset.map(create_prompt)
```

```
Map:   0%|          | 0/11573 [00:00<?, ? examples/s]
Map:   0%|          | 0/1286 [00:00<?, ? examples/s]
```

```
In [5]: # --- Hyperparameters ---
        learning_rate = 0.0000538
        per_device_train_batch_size = 8
        gradient_accumulation_steps = 4
        max_length = 512
        max_prompt_length = 256
        max_target_length = 256
        max_steps = 500
        report_to = "wandb"
        gradient_checkpointing = True
        beta = 0.2
```

```
In [6]: # --- Weights & Biases Setup ---
        wandb.init(  

```

```

project="dpo-training",
config={
    "model_name": model_name_or_path,
    "dataset": "argilla/distilabel-intel-orca-dpo-pairs",
    "learning_rate": learning_rate,
    "batch_size": per_device_train_batch_size,
    "gradient_accumulation_steps": gradient_accumulation_steps,
    "max_length": max_length,
    "max_prompt_length": max_prompt_length,
    "max_target_length": max_target_length,
    "max_steps": max_steps,
    "beta": beta,
    "gradient_checkpointing": gradient_checkpointing,
    "optimizer": "adamw_torch",
},
)

# --- Training Arguments ---
training_args = TrainingArguments(
    per_device_train_batch_size=per_device_train_batch_size,
    max_steps=max_steps,
    remove_unused_columns=False,
    gradient_accumulation_steps=gradient_accumulation_steps,
    learning_rate=learning_rate,
    evaluation_strategy="steps",
    logging_first_step=True,
    logging_steps=5,
    eval_steps=50,
    output_dir="./nlp-a5",
    optim="adamw_torch",
    warmup_steps=50,
    report_to=report_to,
    bf16=True,
    gradient_checkpointing=gradient_checkpointing,
    save_strategy="steps",
    save_steps=50,
)

```

wandb: Using wandb-core as the SDK backend. Please refer to <https://wandb.me/wandb-core> for more information.

wandb: Currently logged in as: sila-nmht (sila-nmht-asian-institute-of-technology) to <https://api.wandb.ai>. Use `wandb login --relogin` to force relogin

Tracking run with wandb version 0.19.7

Run data is saved locally in c:\Users\silan\Desktop\A5\wandb\run-20250302_162206-b5jpbbnp

Syncing run **spring-music-1** to Weights & Biases (docs)

View project at <https://wandb.ai/sila-nmht-asian-institute-of-technology/dpo-training>

View run at <https://wandb.ai/sila-nmht-asian-institute-of-technology/dpo-training/runs/b5jpbbnp>

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\transformers\training_args.py:1545: FutureWarning: `evaluation_strategy` is deprecated and will be removed in version 4.46 of 🤗 Transformers. Use `eval_strategy` instead
 warnings.warn(

```

In [7]: # --- DPOTrainer Initialization ---
dpo_trainer = DPOTrainer(
    model,
    model_ref,
    args=training_args,
    beta=beta,
    train_dataset=train_dataset,
    eval_dataset=eval_dataset,
    tokenizer=tokenizer,
    max_length=max_length,
    max_target_length=max_target_length,
    max_prompt_length=max_prompt_length,
    generate_during_eval=True,
)

```

```
Map:      0%|          | 0/11573 [00:00<?, ? examples/s]
Token indices sequence length is longer than the specified maximum sequence length for this model (1105 > 1024). Running this sequence through the model will result in indexing errors
Map:      0%|          | 0/1286 [00:00<?, ? examples/s]
max_steps is given, it will override any value given in num_train_epochs
```

```
In [8]: # --- Training ---
        dpo_trainer.train()
```

```
wandb: WARNING The `run_name` is currently set to the same value as `TrainingArguments.output_dir`. If this was not intended, please specify a different run name by setting the `TrainingArguments.run_name`
parameter.
      0%|          | 0/500 [00:00<?, ?it/s]
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\transformers\models\gpt2\modeling_gpt2.py:545: UserWarning: 1Torch was not compiled with flash attention. (Triggered internally at C:\actions-runner\_work
\pytorch\pytorch\builder\windows\pytorch\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:555.)
  attn_output = torch.nn.functional.scaled_dot_product_attention(
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
Could not estimate the number of tokens of the input, floating-point operations will not be computed
{'loss': 0.6931, 'grad_norm': 86.32217407226562, 'learning_rate': 1.076e-06, 'rewards/chosen': 0.0, 'rewards/rejected': 0.0, 'rewards/accuracies': 0.0, 'rewards/margins': 0.0, 'logps/rejected': -429.841064
453125, 'logps/chosen': -380.14141845703125, 'logits/rejected': -101.80658721923828, 'logits/chosen': -101.73731994628906, 'epoch': 0.0}
{'loss': 0.6917, 'grad_norm': 82.48517608642578, 'learning_rate': 5.38e-06, 'rewards/chosen': -0.655906617641449, 'rewards/rejected': -1.1794593334197998, 'rewards/accuracies': 0.515625, 'rewards/margins':
0.5235527753829956, 'logps/rejected': -413.0698547363281, 'logps/chosen': -359.45428466796875, 'logits/rejected': -104.19410705566406, 'logits/chosen': -102.56261444091797, 'epoch': 0.01}
{'loss': 0.4781, 'grad_norm': 51.92380142211914, 'learning_rate': 1.076e-05, 'rewards/chosen': -0.745120644569397, 'rewards/rejected': -2.0767862796783447, 'rewards/accuracies': 0.8125, 'rewards/margins':
1.3316656351089478, 'logps/rejected': -422.7705078125, 'logps/chosen': -330.71630859375, 'logits/rejected': -104.4666976928711, 'logits/chosen': -103.62039947509766, 'epoch': 0.03}
{'loss': 0.5277, 'grad_norm': 46.469303131103516, 'learning_rate': 1.6139999999999998e-05, 'rewards/chosen': -0.696223795413971, 'rewards/rejected': -2.133615016937256, 'rewards/accuracies': 0.800000011920
929, 'rewards/margins': 1.4373910427093506, 'logps/rejected': -432.337158203125, 'logps/chosen': -382.0255126953125, 'logits/rejected': -103.82987213134766, 'logits/chosen': -102.584228515625, 'epoch': 0.0
4}
{'loss': 0.385, 'grad_norm': 66.70952606201172, 'learning_rate': 2.152e-05, 'rewards/chosen': -0.7804359793663025, 'rewards/rejected': -2.7457351684570312, 'rewards/accuracies': 0.84375, 'rewards/margins':
1.9652992486953735, 'logps/rejected': -407.06201171875, 'logps/chosen': -355.43988037109375, 'logits/rejected': -102.24687194824219, 'logits/chosen': -101.85172271728516, 'epoch': 0.06}
{'loss': 0.7909, 'grad_norm': 64.98677825927734, 'learning_rate': 2.69e-05, 'rewards/chosen': -0.947333812713623, 'rewards/rejected': -2.540565252304077, 'rewards/accuracies': 0.7749999761581421, 'rewards/
margins': 1.5932317972183228, 'logps/rejected': -419.9918518066406, 'logps/chosen': -333.4473571777344, 'logits/rejected': -103.3801498413086, 'logits/chosen': -101.32579040527344, 'epoch': 0.07}
{'loss': 0.9666, 'grad_norm': 94.71665954589844, 'learning_rate': 3.2279999999999996e-05, 'rewards/chosen': -1.1336230039596558, 'rewards/rejected': -2.738132953643799, 'rewards/accuracies': 0.78125, 'rewa
rds/margins': 1.604509711265564, 'logps/rejected': -418.307861328125, 'logps/chosen': -351.4735107421875, 'logits/rejected': -102.21714782714844, 'logits/chosen': -101.72389221191406, 'epoch': 0.08}
{'loss': 1.4839, 'grad_norm': 83.86888885498047, 'learning_rate': 3.7659999999999995e-05, 'rewards/chosen': -1.5797709226608276, 'rewards/rejected': -2.902073860168457, 'rewards/accuracies': 0.731249988079
071, 'rewards/margins': 1.3223025798797607, 'logps/rejected': -393.5597839355469, 'logps/chosen': -364.721435546875, 'logits/rejected': -101.4345932006836, 'logits/chosen': -100.98875427246094, 'epoch': 0.
1}
{'loss': 0.8129, 'grad_norm': 53.89474868774414, 'learning_rate': 4.304e-05, 'rewards/chosen': 0.34803858399391174, 'rewards/rejected': -1.3461812734603882, 'rewards/accuracies': 0.762499988079071, 'reward
s/margins': 1.694219946861267, 'logps/rejected': -398.78253173828125, 'logps/chosen': -349.1788330078125, 'logits/rejected': -100.0361557006836, 'logits/chosen': -99.02516174316406, 'epoch': 0.11}
{'loss': 0.5471, 'grad_norm': 45.81199264526367, 'learning_rate': 4.842e-05, 'rewards/chosen': 0.4214906692504883, 'rewards/rejected': -2.0227017402648926, 'rewards/accuracies': 0.7875000238418579, 'reward
s/margins': 2.444192409515381, 'logps/rejected': -421.54150390625, 'logps/chosen': -342.899658203125, 'logits/rejected': -98.65272521972656, 'logits/chosen': -98.01823425292969, 'epoch': 0.12}
{'loss': 0.6454, 'grad_norm': 73.66812133789062, 'learning_rate': 5.38e-05, 'rewards/chosen': 0.3888837695121765, 'rewards/rejected': -1.6650291681289673, 'rewards/accuracies': 0.8062499761581421, 'reward
s/margins': 2.05391263961792, 'logps/rejected': -420.1739807128906, 'logps/chosen': -389.46099853515625, 'logits/rejected': -97.13334655761719, 'logits/chosen': -95.84834289550781, 'epoch': 0.14}
`use_cache=True` is incompatible with gradient checkpointing. Setting `use_cache=False`...
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002470D24FD90>, 'epoch': 0.14}
      0%|          | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.7701175808906555, 'eval_runtime': 118.3686, 'eval_samples_per_second': 10.864, 'eval_steps_per_second': 1.36, 'eval_rewards/chosen': 0.46673712134361267, 'eval_rewards/rejected': -1.3878153
562545776, 'eval_rewards/accuracies': 0.7590579390525818, 'eval_rewards/margins': 1.8545525074005127, 'eval_logps/rejected': -406.8403015136719, 'eval_logps/chosen': -339.9384765625, 'eval_logits/rejecte
d': -95.71627044677734, 'eval_logits/chosen': -95.33931732177734, 'epoch': 0.14}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
```

```
{'loss': 0.558, 'grad_norm': 79.04922485351562, 'learning_rate': 5.320222222222222e-05, 'rewards/chosen': 0.4374000132083893, 'rewards/rejected': -1.672490119934082, 'rewards/accuracies': 0.8062499761581421, 'rewards/margins': 2.1098902225494385, 'logps/rejected': -417.58819580078125, 'logps/chosen': -351.3348388671875, 'logits/rejected': -96.4607162475586, 'logits/chosen': -95.52983856201172, 'epoch': 0.15}
{'loss': 0.8263, 'grad_norm': 47.55983352661133, 'learning_rate': 5.260444444444444e-05, 'rewards/chosen': -1.0790084600448608, 'rewards/rejected': -3.867455244064331, 'rewards/accuracies': 0.762499988079071, 'rewards/margins': 2.7884469032287598, 'logps/rejected': -433.04083251953125, 'logps/chosen': -375.5039978027344, 'logits/rejected': -96.74488830566406, 'logits/chosen': -96.39244079589844, 'epoch': 0.17}
{'loss': 0.6953, 'grad_norm': 54.888671875, 'learning_rate': 5.2006666666666665e-05, 'rewards/chosen': 0.06507492810487747, 'rewards/rejected': -2.1249938011169434, 'rewards/accuracies': 0.78125, 'rewards/margins': 2.1900687217712402, 'logps/rejected': -408.31561279296875, 'logps/chosen': -342.8404541015625, 'logits/rejected': -95.37672424316406, 'logits/chosen': -95.01136779785156, 'epoch': 0.18}
{'loss': 0.6132, 'grad_norm': 84.78803253173828, 'learning_rate': 5.140888888888889e-05, 'rewards/chosen': 1.1877286434173584, 'rewards/rejected': -1.399277925491333, 'rewards/accuracies': 0.800000011920929, 'rewards/margins': 2.5870065689086914, 'logps/rejected': -400.93487548828125, 'logps/chosen': -331.5931091308594, 'logits/rejected': -93.27552795410156, 'logits/chosen': -93.60957336425781, 'epoch': 0.19}
{'loss': 1.0169, 'grad_norm': 49.93352127075195, 'learning_rate': 5.081111111111111e-05, 'rewards/chosen': -0.07504285871982574, 'rewards/rejected': -2.3349924087524414, 'rewards/accuracies': 0.8062499761581421, 'rewards/margins': 2.2599496841430664, 'logps/rejected': -403.9119873046875, 'logps/chosen': -335.65911865234375, 'logits/rejected': -91.37318420410156, 'logits/chosen': -91.47337341308594, 'epoch': 0.21}
{'loss': 0.6825, 'grad_norm': 52.41972732543945, 'learning_rate': 5.0213333333333336e-05, 'rewards/chosen': -0.767310619354248, 'rewards/rejected': -4.384817600250244, 'rewards/accuracies': 0.8374999761581421, 'rewards/margins': 3.617506742477417, 'logps/rejected': -418.5055236816406, 'logps/chosen': -336.71527099609375, 'logits/rejected': -90.05725860595703, 'logits/chosen': -90.01679229736328, 'epoch': 0.22}
{'loss': 1.7861, 'grad_norm': 117.04061126708984, 'learning_rate': 4.961555555555556e-05, 'rewards/chosen': -1.8621944189071655, 'rewards/rejected': -3.74086332321167, 'rewards/accuracies': 0.800000011920929, 'rewards/margins': 1.8786687850952148, 'logps/rejected': -420.4522399902344, 'logps/chosen': -386.14910888671875, 'logits/rejected': -90.28736114501953, 'logits/chosen': -89.07003784179688, 'epoch': 0.23}
{'loss': 0.9483, 'grad_norm': 111.1344985961914, 'learning_rate': 4.901777777777778e-05, 'rewards/chosen': -0.37004974484443665, 'rewards/rejected': -2.7511322498321533, 'rewards/accuracies': 0.8062499761581421, 'rewards/margins': 2.381082773208618, 'logps/rejected': -430.1302185058594, 'logps/chosen': -369.4415588378906, 'logits/rejected': -88.33872985839844, 'logits/chosen': -87.91231536865234, 'epoch': 0.25}
{'loss': 0.8893, 'grad_norm': 51.4976921081543, 'learning_rate': 4.842e-05, 'rewards/chosen': 0.4231366515159607, 'rewards/rejected': -2.0979807376861572, 'rewards/accuracies': 0.7437499761581421, 'rewards/margins': 2.5211174488067627, 'logps/rejected': -401.6153259277344, 'logps/chosen': -343.760009765625, 'logits/rejected': -87.45030212402344, 'logits/chosen': -86.86721801757812, 'epoch': 0.26}
{'loss': 0.7265, 'grad_norm': 50.081626892089844, 'learning_rate': 4.782222222222222e-05, 'rewards/chosen': -0.06916000694036484, 'rewards/rejected': -2.681837558746338, 'rewards/accuracies': 0.8374999761581421, 'rewards/margins': 2.6126773357391357, 'logps/rejected': -409.5958251953125, 'logps/chosen': -359.68389892578125, 'logits/rejected': -86.97504425048828, 'logits/chosen': -85.43632507324219, 'epoch': 0.28}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002470BA3E550>, 'epoch': 0.28}
0%|          | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.7531054019927979, 'eval_runtime': 117.5481, 'eval_samples_per_second': 10.94, 'eval_steps_per_second': 1.37, 'eval_rewards/chosen': 0.27912962436676025, 'eval_rewards/rejected': -2.1547889709472656, 'eval_rewards/accuracies': 0.7776915431022644, 'eval_rewards/margins': 2.4339182376861572, 'eval_logps/rejected': -410.67523193359375, 'eval_logps/chosen': -340.8764953613281, 'eval_logits/rejected': -85.44556427001953, 'eval_logits/chosen': -85.26910400390625, 'epoch': 0.28}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
{'loss': 0.8066, 'grad_norm': 41.73305892944336, 'learning_rate': 4.722444444444444e-05, 'rewards/chosen': 0.4517017900943756, 'rewards/rejected': -1.5693342685699463, 'rewards/accuracies': 0.7875000238418579, 'rewards/margins': 2.021036148071289, 'logps/rejected': -420.3628845214844, 'logps/chosen': -355.36126708984375, 'logits/rejected': -86.13858795166016, 'logits/chosen': -85.57471466064453, 'epoch': 0.29}
{'loss': 0.7271, 'grad_norm': 67.39576721191406, 'learning_rate': 4.6626666666666665e-05, 'rewards/chosen': 0.5972003936767578, 'rewards/rejected': -1.8452813625335693, 'rewards/accuracies': 0.800000011920929, 'rewards/margins': 2.442481756210327, 'logps/rejected': -399.28125, 'logps/chosen': -358.2326354980469, 'logits/rejected': -85.44587707519531, 'logits/chosen': -84.99433898925781, 'epoch': 0.3}
{'loss': 0.6502, 'grad_norm': 41.41155242919922, 'learning_rate': 4.602888888888889e-05, 'rewards/chosen': 0.0006014645332470536, 'rewards/rejected': -2.6945464611053467, 'rewards/accuracies': 0.793749988079071, 'rewards/margins': 2.695147752761841, 'logps/rejected': -399.70941162109375, 'logps/chosen': -348.1581115722656, 'logits/rejected': -85.673828125, 'logits/chosen': -85.75601196289062, 'epoch': 0.32}
{'loss': 0.7564, 'grad_norm': 49.70991897583008, 'learning_rate': 4.543111111111111e-05, 'rewards/chosen': 0.5910513401031494, 'rewards/rejected': -2.240988254547119, 'rewards/accuracies': 0.78125, 'rewards/margins': 2.8320398330688477, 'logps/rejected': -445.3321838378906, 'logps/chosen': -377.35931396484375, 'logits/rejected': -84.32178497314453, 'logits/chosen': -84.53287506103516, 'epoch': 0.33}
{'loss': 0.9219, 'grad_norm': 49.939979553222656, 'learning_rate': 4.483333333333334e-05, 'rewards/chosen': 0.9032796621322632, 'rewards/rejected': -1.4474371671676636, 'rewards/accuracies': 0.75, 'rewards/margins': 2.350717067718506, 'logps/rejected': -392.47528076171875, 'logps/chosen': -338.84942626953125, 'logits/rejected': -82.7908706665039, 'logits/chosen': -83.20585632324219, 'epoch': 0.35}
{'loss': 0.6825, 'grad_norm': 43.489234924316406, 'learning_rate': 4.423555555555555e-05, 'rewards/chosen': 0.7985666394233704, 'rewards/rejected': -1.5062321424484253, 'rewards/accuracies': 0.7562500238418579, 'rewards/margins': 2.3047986030578613, 'logps/rejected': -407.49896240234375, 'logps/chosen': -349.68365478515625, 'logits/rejected': -81.65785217285156, 'logits/chosen': -82.04661560058594, 'epoch': 0.36}
{'loss': 0.788, 'grad_norm': 44.61458206176758, 'learning_rate': 4.363777777777778e-05, 'rewards/chosen': -0.6791790127754211, 'rewards/rejected': -3.2409214973449707, 'rewards/accuracies': 0.762499988079071, 'rewards/margins': 2.5617427825927734, 'logps/rejected': -400.8290710449219, 'logps/chosen': -343.03173828125, 'logits/rejected': -81.57640838623047, 'logits/chosen': -82.28054809570312, 'epoch': 0.37}
{'loss': 0.6962, 'grad_norm': 48.99653625488281, 'learning_rate': 4.304e-05, 'rewards/chosen': -0.26395973563194275, 'rewards/rejected': -2.7384941577911377, 'rewards/accuracies': 0.78125, 'rewards/margins': 2.474534749984741, 'logps/rejected': -421.305908203125, 'logps/chosen': -373.29522705078125, 'logits/rejected': -81.13308715820312, 'logits/chosen': -81.98467254638672, 'epoch': 0.39}
{'loss': 0.6925, 'grad_norm': 70.9882583618164, 'learning_rate': 4.244222222222222e-05, 'rewards/chosen': 0.5938063263893127, 'rewards/rejected': -2.5369081497192383, 'rewards/accuracies': 0.824999988079071, 'rewards/margins': 3.130714178085327, 'logps/rejected': -409.7278747558594, 'logps/chosen': -364.85125732421875, 'logits/rejected': -79.89900970458984, 'logits/chosen': -80.49259185791016, 'epoch': 0.4}
{'loss': 0.5317, 'grad_norm': 27.644611358642578, 'learning_rate': 4.184444444444444e-05, 'rewards/chosen': 0.38659635186195374, 'rewards/rejected': -2.2567315101623535, 'rewards/accuracies': 0.800000011920929, 'rewards/margins': 2.6433279514312744, 'logps/rejected': -392.61029052734375, 'logps/chosen': -326.4050598144531, 'logits/rejected': -80.21643829345703, 'logits/chosen': -80.45060729980469, 'epoch': 0.41}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002474D7F9490>, 'epoch': 0.41}
0%|          | 0/161 [00:00<?, ?it/s]
```



```
{'eval_loss': 0.716432329925537, 'eval_runtime': 117.5167, 'eval_samples_per_second': 10.943, 'eval_steps_per_second': 1.37, 'eval_rewards/chosen': 0.04009328410029411, 'eval_rewards/rejected': -2.6230392456054688, 'eval_rewards/accuracies': 0.7743271589279175, 'eval_rewards/margins': 2.663132667541504, 'eval_logps/rejected': -413.0164489746094, 'eval_logps/chosen': -342.0716857910156, 'eval_logits/rejected': -77.79004669189453, 'eval_logits/chosen': -78.4781494140625, 'epoch': 0.41}

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]

{'loss': 0.6313, 'grad_norm': 61.368038177490234, 'learning_rate': 4.124666666666667e-05, 'rewards/chosen': -0.0895061120390892, 'rewards/rejected': -3.2197909355163574, 'rewards/accuracies': 0.768750011920929, 'rewards/margins': 3.1302852630615234, 'logps/rejected': -424.0433044433594, 'logps/chosen': -370.56231689453125, 'logits/rejected': -77.89362335205078, 'logits/chosen': -78.12007904052734, 'epoch': 0.43}
{'loss': 0.9444, 'grad_norm': 36.428863525390625, 'learning_rate': 4.064888888888889e-05, 'rewards/chosen': 0.01371699571609497, 'rewards/rejected': -2.3674216270446777, 'rewards/accuracies': 0.7875000238418579, 'rewards/margins': 2.381138324737549, 'logps/rejected': -397.24969482421875, 'logps/chosen': -346.69927978515625, 'logits/rejected': -76.88948822021484, 'logits/chosen': -77.67308044433594, 'epoch': 0.44}
{'loss': 0.6902, 'grad_norm': 41.582950592041016, 'learning_rate': 4.005111111111111e-05, 'rewards/chosen': 0.8930956721305847, 'rewards/rejected': -1.7185869216918945, 'rewards/accuracies': 0.7749999761581421, 'rewards/margins': 2.611682415008545, 'logps/rejected': -391.7558898925781, 'logps/chosen': -332.7657470703125, 'logits/rejected': -76.559326171875, 'logits/chosen': -76.60616302490234, 'epoch': 0.46}
{'loss': 0.8164, 'grad_norm': 32.99758529663086, 'learning_rate': 3.945333333333333e-05, 'rewards/chosen': 0.66659015417099, 'rewards/rejected': -1.5941461324691772, 'rewards/accuracies': 0.731249988079071, 'rewards/margins': 2.2607362270355225, 'logps/rejected': -371.60015869140625, 'logps/chosen': -317.7023010253906, 'logits/rejected': -74.71438598632812, 'logits/chosen': -75.0600357055664, 'epoch': 0.47}
{'loss': 0.7898, 'grad_norm': 44.49337387084961, 'learning_rate': 3.885555555555555e-05, 'rewards/chosen': -0.1970115453004837, 'rewards/rejected': -3.1468515396118164, 'rewards/accuracies': 0.8125, 'rewards/margins': 2.9498403072357178, 'logps/rejected': -443.7132873535156, 'logps/chosen': -349.10968017578125, 'logits/rejected': -73.07229614257812, 'logits/chosen': -74.90153503417969, 'epoch': 0.48}
{'loss': 0.6749, 'grad_norm': 44.76820373535156, 'learning_rate': 3.825777777777777e-05, 'rewards/chosen': -0.3300102651119232, 'rewards/rejected': -3.2335567474365234, 'rewards/accuracies': 0.7875000238418579, 'rewards/margins': 2.9035465717315674, 'logps/rejected': -423.0830078125, 'logps/chosen': -365.2192077636719, 'logits/rejected': -75.81597900390625, 'logits/chosen': -76.04116821289062, 'epoch': 0.5}
{'loss': 0.8762, 'grad_norm': 43.37041473388672, 'learning_rate': 3.7659999999999995e-05, 'rewards/chosen': 0.4160152077674866, 'rewards/rejected': -2.0366153717041016, 'rewards/accuracies': 0.7437499761581421, 'rewards/margins': 2.4526305198669434, 'logps/rejected': -413.62701416015625, 'logps/chosen': -374.9130859375, 'logits/rejected': -74.63163757324219, 'logits/chosen': -75.06140899658203, 'epoch': 0.51}
{'loss': 0.6281, 'grad_norm': 42.40433883666992, 'learning_rate': 3.706222222222222e-05, 'rewards/chosen': 0.18032363057136536, 'rewards/rejected': -2.136779308319092, 'rewards/accuracies': 0.7562500238418579, 'rewards/margins': 2.317103147506714, 'logps/rejected': -390.0362243652344, 'logps/chosen': -327.7127990722656, 'logits/rejected': -75.75682067871094, 'logits/chosen': -76.13587188720703, 'epoch': 0.53}
{'loss': 0.5412, 'grad_norm': 20.47016716003418, 'learning_rate': 3.646444444444444e-05, 'rewards/chosen': -0.11494920402765274, 'rewards/rejected': -3.1286449432373047, 'rewards/accuracies': 0.7749999761581421, 'rewards/margins': 3.013695478439331, 'logps/rejected': -395.9088134765625, 'logps/chosen': -320.53106689453125, 'logits/rejected': -76.11039733886719, 'logits/chosen': -76.77498626708984, 'epoch': 0.54}
{'loss': 0.8947, 'grad_norm': 55.328277587890625, 'learning_rate': 3.586666666666667e-05, 'rewards/chosen': -0.5571750998497009, 'rewards/rejected': -3.331686019897461, 'rewards/accuracies': 0.800000011920929, 'rewards/margins': 2.7745108604431152, 'logps/rejected': -403.8236083984375, 'logps/chosen': -356.4224548339844, 'logits/rejected': -76.06084442138672, 'logits/chosen': -76.36619567871094, 'epoch': 0.55}

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002474D819970>, 'epoch': 0.55}
 0%|          | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.7222762703895569, 'eval_runtime': 121.0032, 'eval_samples_per_second': 10.628, 'eval_steps_per_second': 1.331, 'eval_rewards/chosen': -0.032697293907403946, 'eval_rewards/rejected': -3.1584994792938232, 'eval_rewards/accuracies': 0.796066164970398, 'eval_rewards/margins': 3.1258020401000977, 'eval_logps/rejected': -415.6938171386719, 'eval_logps/chosen': -342.43560791015625, 'eval_logits/rejected': -73.72225952148438, 'eval_logits/chosen': -74.38448333740234, 'epoch': 0.55}

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
```

```
{'loss': 0.5457, 'grad_norm': 42.19619369506836, 'learning_rate': 3.526888888888889e-05, 'rewards/chosen': 0.3945646286010742, 'rewards/rejected': -3.658681869506836, 'rewards/accuracies': 0.8187500238418579, 'rewards/margins': 4.05324649810791, 'logps/rejected': -466.9610900878906, 'logps/chosen': -357.0932922363281, 'logits/rejected': -72.65667724609375, 'logits/chosen': -73.61896514892578, 'epoch': 0.57}
{'loss': 0.7351, 'grad_norm': 44.63496780395508, 'learning_rate': 3.4671111111111116e-05, 'rewards/chosen': 1.0736498832702637, 'rewards/rejected': -1.8460795879364014, 'rewards/accuracies': 0.7749999761581421, 'rewards/margins': 2.919729232788086, 'logps/rejected': -423.306884765625, 'logps/chosen': -362.35284423828125, 'logits/rejected': -70.54705810546875, 'logits/chosen': -70.61494445800781, 'epoch': 0.58}
{'loss': 0.826, 'grad_norm': 63.51594161987305, 'learning_rate': 3.407333333333333e-05, 'rewards/chosen': 1.1699292659759521, 'rewards/rejected': -1.4635512828826904, 'rewards/accuracies': 0.7749999761581421, 'rewards/margins': 2.6334807872772217, 'logps/rejected': -432.220458984375, 'logps/chosen': -371.22259521484375, 'logits/rejected': -68.8424301147461, 'logits/chosen': -69.8603286743164, 'epoch': 0.59}
{'loss': 0.862, 'grad_norm': 33.00798416137695, 'learning_rate': 3.347555555555555e-05, 'rewards/chosen': 0.6025430560112, 'rewards/rejected': -1.7918682098388672, 'rewards/accuracies': 0.78125, 'rewards/margins': 2.394411325454712, 'logps/rejected': -386.277099609375, 'logps/chosen': -316.5558166503906, 'logits/rejected': -71.87672424316406, 'logits/chosen': -72.52440643310547, 'epoch': 0.61}
{'loss': 0.6404, 'grad_norm': 43.626895904541016, 'learning_rate': 3.287777777777778e-05, 'rewards/chosen': 0.7662680149078369, 'rewards/rejected': -1.6726696491241455, 'rewards/accuracies': 0.7437499761581421, 'rewards/margins': 2.4389376640319824, 'logps/rejected': -423.9674377441406, 'logps/chosen': -339.8011474609375, 'logits/rejected': -69.69376373291016, 'logits/chosen': -71.6377944946289, 'epoch': 0.62}
{'loss': 0.9638, 'grad_norm': 65.55481719970703, 'learning_rate': 3.2279999999999996e-05, 'rewards/chosen': 0.3657369613647461, 'rewards/rejected': -1.8621606826782227, 'rewards/accuracies': 0.7124999761581421, 'rewards/margins': 2.227898120880127, 'logps/rejected': -406.73431396484375, 'logps/chosen': -378.902099609375, 'logits/rejected': -68.96754455566406, 'logits/chosen': -69.4041519165039, 'epoch': 0.64}
{'loss': 0.5815, 'grad_norm': 47.545780181884766, 'learning_rate': 3.1682222222222224e-05, 'rewards/chosen': 1.316070556640625, 'rewards/rejected': -1.4784131050109863, 'rewards/accuracies': 0.793749988079071, 'rewards/margins': 2.7944839000701904, 'logps/rejected': -409.5979919433594, 'logps/chosen': -385.15472412109375, 'logits/rejected': -67.60984802246094, 'logits/chosen': -68.28776550292969, 'epoch': 0.65}
{'loss': 0.5846, 'grad_norm': 34.15302658081055, 'learning_rate': 3.1084444444444444e-05, 'rewards/chosen': 1.0559507608413696, 'rewards/rejected': -1.4341124296188354, 'rewards/accuracies': 0.8125, 'rewards/margins': 2.490063190460205, 'logps/rejected': -385.544677734375, 'logps/chosen': -335.39129638671875, 'logits/rejected': -68.02262115478516, 'logits/chosen': -68.2630386352539, 'epoch': 0.66}
{'loss': 0.9241, 'grad_norm': 32.87361145019531, 'learning_rate': 3.0486666666666667e-05, 'rewards/chosen': 0.27785366773605347, 'rewards/rejected': -1.7676045894622803, 'rewards/accuracies': 0.7562500238418579, 'rewards/margins': 2.0454583168029785, 'logps/rejected': -412.60931396484375, 'logps/chosen': -351.5341796875, 'logits/rejected': -67.31730651855469, 'logits/chosen': -68.31941986083984, 'epoch': 0.68}
{'loss': 0.6882, 'grad_norm': 24.867473602294922, 'learning_rate': 2.988888888888889e-05, 'rewards/chosen': 0.531781792640686, 'rewards/rejected': -2.517037868499756, 'rewards/accuracies': 0.8187500238418579, 'rewards/margins': 3.0488193035125732, 'logps/rejected': -433.2159118652344, 'logps/chosen': -337.3003234863281, 'logits/rejected': -66.43065643310547, 'logits/chosen': -67.45755004882812, 'epoch': 0.69}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002474D80AC70>, 'epoch': 0.69}
0%|          | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.6676872372627258, 'eval_runtime': 121.6993, 'eval_samples_per_second': 10.567, 'eval_steps_per_second': 1.323, 'eval_rewards/chosen': 0.6186218857765198, 'eval_rewards/rejected': -2.0402002334594727, 'eval_rewards/accuracies': 0.7903726696968079, 'eval_rewards/margins': 2.6588220596313477, 'eval_logps/rejected': -410.102294921875, 'eval_logps/chosen': -339.17901611328125, 'eval_logits/rejected': -66.41825103759766, 'eval_logits/chosen': -67.22669219970703, 'epoch': 0.69}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
{'loss': 0.7138, 'grad_norm': 52.590721130371094, 'learning_rate': 2.9291111111111107e-05, 'rewards/chosen': 0.6930635571479797, 'rewards/rejected': -2.0636179447174072, 'rewards/accuracies': 0.8062499761581421, 'rewards/margins': 2.756681442260742, 'logps/rejected': -430.04669189453125, 'logps/chosen': -366.04327392578125, 'logits/rejected': -66.62266540527344, 'logits/chosen': -67.40010833740234, 'epoch': 0.7}
{'loss': 0.4698, 'grad_norm': 29.189342498779297, 'learning_rate': 2.8693333333333332e-05, 'rewards/chosen': 0.6795655488967896, 'rewards/rejected': -2.4594383239746094, 'rewards/accuracies': 0.84375, 'rewards/margins': 3.1390042304992676, 'logps/rejected': -394.29034423828125, 'logps/chosen': -321.86761474609375, 'logits/rejected': -67.69721984863281, 'logits/chosen': -68.16939544677734, 'epoch': 0.72}
{'loss': 1.2202, 'grad_norm': 43.41884994506836, 'learning_rate': 2.8095555555555557e-05, 'rewards/chosen': 0.10626578330993652, 'rewards/rejected': -2.0635104179382324, 'rewards/accuracies': 0.762499988079071, 'rewards/margins': 2.169776439666748, 'logps/rejected': -437.454833984375, 'logps/chosen': -396.3990173339844, 'logits/rejected': -64.95469665527344, 'logits/chosen': -65.3371810913086, 'epoch': 0.73}
{'loss': 0.688, 'grad_norm': 54.823150634765625, 'learning_rate': 2.7497777777777775e-05, 'rewards/chosen': 0.6668751239776611, 'rewards/rejected': -1.925405740737915, 'rewards/accuracies': 0.8062499761581421, 'rewards/margins': 2.592280864715576, 'logps/rejected': -384.44769287109375, 'logps/chosen': -328.5667419433594, 'logits/rejected': -64.68936157226562, 'logits/chosen': -65.6404037475586, 'epoch': 0.75}
{'loss': 0.876, 'grad_norm': 25.46416473388672, 'learning_rate': 2.69e-05, 'rewards/chosen': 0.7416917681694031, 'rewards/rejected': -1.2324570417404175, 'rewards/accuracies': 0.7875000238418579, 'rewards/margins': 1.9741485118865967, 'logps/rejected': -387.96417236328125, 'logps/chosen': -337.8142395019531, 'logits/rejected': -66.47028350830078, 'logits/chosen': -66.3119888305664, 'epoch': 0.76}
{'loss': 0.6601, 'grad_norm': 39.12128448486328, 'learning_rate': 2.630222222222222e-05, 'rewards/chosen': 1.2589921951293945, 'rewards/rejected': -1.1641098260879517, 'rewards/accuracies': 0.762499988079071, 'rewards/margins': 2.4231019020080566, 'logps/rejected': -414.93707275390625, 'logps/chosen': -343.75286865234375, 'logits/rejected': -66.33631134033203, 'logits/chosen': -67.12257385253906, 'epoch': 0.77}
{'loss': 0.8265, 'grad_norm': 45.296104431152344, 'learning_rate': 2.5704444444444446e-05, 'rewards/chosen': 1.0840585231781006, 'rewards/rejected': -0.956757664680481, 'rewards/accuracies': 0.793749988079071, 'rewards/margins': 2.040816068649292, 'logps/rejected': -379.8754577636719, 'logps/chosen': -305.32147216796875, 'logits/rejected': -68.24330139160156, 'logits/chosen': -68.14559936523438, 'epoch': 0.79}
{'loss': 0.5377, 'grad_norm': 44.5572395324707, 'learning_rate': 2.5106666666666668e-05, 'rewards/chosen': 0.13226111233234406, 'rewards/rejected': -2.8885281085968018, 'rewards/accuracies': 0.831250011920929, 'rewards/margins': 3.020789384841919, 'logps/rejected': -423.3558654785156, 'logps/chosen': -337.54180908203125, 'logits/rejected': -67.06298828125, 'logits/chosen': -68.6774673461914, 'epoch': 0.8}
{'loss': 0.5555, 'grad_norm': 55.0949821472168, 'learning_rate': 2.450888888888889e-05, 'rewards/chosen': 0.19217965006828308, 'rewards/rejected': -3.7852272987365723, 'rewards/accuracies': 0.84375, 'rewards/margins': 3.9774067401885986, 'logps/rejected': -462.6588439941406, 'logps/chosen': -334.9466552734375, 'logits/rejected': -67.49727630615234, 'logits/chosen': -69.19988250732422, 'epoch': 0.82}
{'loss': 0.4596, 'grad_norm': 39.8045654296875, 'learning_rate': 2.391111111111111e-05, 'rewards/chosen': 0.30405426025390625, 'rewards/rejected': -3.699218273162842, 'rewards/accuracies': 0.8374999761581421, 'rewards/margins': 4.003273010253906, 'logps/rejected': -432.2752990722656, 'logps/chosen': -373.6387939453125, 'logits/rejected': -65.89004516601562, 'logits/chosen': -67.20271301269531, 'epoch': 0.83}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002474D7F0910>, 'epoch': 0.83}
0%|          | 0/161 [00:00<?, ?it/s]
```

```
{ 'eval_loss': 0.6199276447296143, 'eval_runtime': 115.807, 'eval_samples_per_second': 11.105, 'eval_steps_per_second': 1.39, 'eval_rewards/chosen': 0.5863352417945862, 'eval_rewards/rejected': -2.493701934814453, 'eval_rewards/accuracies': 0.8115941286087036, 'eval_rewards/margins': 3.0800371170043945, 'eval_logps/rejected': -412.36981201171875, 'eval_logps/chosen': -339.3404846191406, 'eval_logits/rejected': -66.51512908935547, 'eval_logits/chosen': -67.28250885009766, 'epoch': 0.83}

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]

{'loss': 0.5988, 'grad_norm': 44.31158447265625, 'learning_rate': 2.331333333333333e-05, 'rewards/chosen': 0.581853449344635, 'rewards/rejected': -1.9521478414535522, 'rewards/accuracies': 0.800000011920929, 'rewards/margins': 2.534001111984253, 'logps/rejected': -392.5242614746094, 'logps/chosen': -335.1625061035156, 'logits/rejected': -66.18138122558594, 'logits/chosen': -66.18134307861328, 'epoch': 0.84}
{'loss': 0.5134, 'grad_norm': 32.24290084838867, 'learning_rate': 2.2715555555555554e-05, 'rewards/chosen': 1.1869428157806396, 'rewards/rejected': -1.7295684814453125, 'rewards/accuracies': 0.8125, 'rewards/margins': 2.9165115356445312, 'logps/rejected': -374.95172119140625, 'logps/chosen': -303.47381591796875, 'logits/rejected': -67.24435424804688, 'logits/chosen': -67.9585952758789, 'epoch': 0.86}
{'loss': 0.4492, 'grad_norm': 41.668556213378906, 'learning_rate': 2.2117777777777776e-05, 'rewards/chosen': 1.2061660289764404, 'rewards/rejected': -2.1104722023010254, 'rewards/accuracies': 0.83749999761581421, 'rewards/margins': 3.316638231277466, 'logps/rejected': -399.46197509765625, 'logps/chosen': -340.9095153808594, 'logits/rejected': -65.80945587158203, 'logits/chosen': -66.53375244140625, 'epoch': 0.87}
{'loss': 0.5723, 'grad_norm': 21.580129623413086, 'learning_rate': 2.152e-05, 'rewards/chosen': 1.0073246955871582, 'rewards/rejected': -2.201137065887451, 'rewards/accuracies': 0.793749988079071, 'rewards/margins': 3.2084617614746094, 'logps/rejected': -415.3838806152344, 'logps/chosen': -319.04046630859375, 'logits/rejected': -66.15020751953125, 'logits/chosen': -67.0469970703125, 'epoch': 0.88}
{'loss': 0.5127, 'grad_norm': 29.153423309326172, 'learning_rate': 2.092222222222222e-05, 'rewards/chosen': 1.2220170497894287, 'rewards/rejected': -2.398667573928833, 'rewards/accuracies': 0.84375, 'rewards/margins': 3.6206843852996826, 'logps/rejected': -398.83544921875, 'logps/chosen': -347.98675537109375, 'logits/rejected': -66.57186126708984, 'logits/chosen': -66.97994232177734, 'epoch': 0.9}
{'loss': 0.5312, 'grad_norm': 43.84840774536133, 'learning_rate': 2.0324444444444444e-05, 'rewards/chosen': 1.4688034057617188, 'rewards/rejected': -1.6887305974960327, 'rewards/accuracies': 0.8187500238418579, 'rewards/margins': 3.157534122467041, 'logps/rejected': -434.2273864746094, 'logps/chosen': -347.01348876953125, 'logits/rejected': -65.7993392944336, 'logits/chosen': -66.83378601074219, 'epoch': 0.91}
{'loss': 0.7506, 'grad_norm': 47.407249450683594, 'learning_rate': 1.9726666666666665e-05, 'rewards/chosen': 1.1566593647003174, 'rewards/rejected': -1.271723985671997, 'rewards/accuracies': 0.8187500238418579, 'rewards/margins': 2.4283833503723145, 'logps/rejected': -402.28375244140625, 'logps/chosen': -362.5733947753906, 'logits/rejected': -65.85303497314453, 'logits/chosen': -66.76921081542969, 'epoch': 0.93}
{'loss': 0.8776, 'grad_norm': 41.36626434326172, 'learning_rate': 1.9128888888888889e-05, 'rewards/chosen': 0.836691677570343, 'rewards/rejected': -1.5871068239212036, 'rewards/accuracies': 0.8062499761581421, 'rewards/margins': 2.4237983226776123, 'logps/rejected': -390.8427734375, 'logps/chosen': -337.61285400390625, 'logits/rejected': -67.0473861694336, 'logits/chosen': -67.52189636230469, 'epoch': 0.94}
{'loss': 0.6146, 'grad_norm': 30.994197845458984, 'learning_rate': 1.8531111111111112e-05, 'rewards/chosen': 0.6284633874893188, 'rewards/rejected': -2.1303293704986572, 'rewards/accuracies': 0.7875000238418579, 'rewards/margins': 2.7587926387786865, 'logps/rejected': -394.22418212890625, 'logps/chosen': -347.57647705078125, 'logits/rejected': -67.00053405761719, 'logits/chosen': -67.45765686035156, 'epoch': 0.95}
{'loss': 0.6719, 'grad_norm': 49.60293197631836, 'learning_rate': 1.7933333333333333e-05, 'rewards/chosen': 0.9248493909835815, 'rewards/rejected': -1.4640328884124756, 'rewards/accuracies': 0.75, 'rewards/margins': 2.3888823986053467, 'logps/rejected': -392.7113342285156, 'logps/chosen': -329.45001220703125, 'logits/rejected': -64.57225036621094, 'logits/chosen': -66.01065826416016, 'epoch': 0.97}

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(

{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002474D7EA730>, 'epoch': 0.97}
 0%|          | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.6213732957839966, 'eval_runtime': 117.5062, 'eval_samples_per_second': 10.944, 'eval_steps_per_second': 1.37, 'eval_rewards/chosen': 1.101764440536499, 'eval_rewards/rejected': -1.4390374422073364, 'eval_rewards/accuracies': 0.7841615080833435, 'eval_rewards/margins': 2.540802001953125, 'eval_logps/rejected': -407.0964660644531, 'eval_logps/chosen': -336.7633056640625, 'eval_logits/rejected': -64.94148254394531, 'eval_logits/chosen': -65.81298828125, 'epoch': 0.97}

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
```



```
{'loss': 0.4819, 'grad_norm': 30.804285049438477, 'learning_rate': 1.7335555555555558e-05, 'rewards/chosen': 1.484919786453247, 'rewards/rejected': -1.1313515901565552, 'rewards/accuracies': 0.8187500238418579, 'rewards/margins': 2.6162712574005127, 'logps/rejected': -415.90447998046875, 'logps/chosen': -360.48321533203125, 'logits/rejected': -62.61079025268555, 'logits/chosen': -63.80034255981445, 'epoch': 0.98}
{'loss': 0.874, 'grad_norm': 38.822994232177734, 'learning_rate': 1.6737777777777776e-05, 'rewards/chosen': 0.8699784278869629, 'rewards/rejected': -1.6794168949127197, 'rewards/accuracies': 0.768750011920929, 'rewards/margins': 2.5493953227996826, 'logps/rejected': -416.244140625, 'logps/chosen': -341.0328674316406, 'logits/rejected': -64.50392150878906, 'logits/chosen': -65.36022186279297, 'epoch': 1.0}
{'loss': 0.9483, 'grad_norm': 9.058503150939941, 'learning_rate': 1.6139999999999998e-05, 'rewards/chosen': 0.955286026009766, 'rewards/rejected': -2.8314433097839355, 'rewards/accuracies': 0.908749938011694, 'rewards/margins': 3.7867298126220703, 'logps/rejected': -427.146728515625, 'logps/chosen': -327.68560791015625, 'logits/rejected': -62.80608367919922, 'logits/chosen': -64.36961364746094, 'epoch': 1.01}
{'loss': 0.0904, 'grad_norm': 16.072532653808594, 'learning_rate': 1.554222222222222e-05, 'rewards/chosen': 1.9007396697998047, 'rewards/rejected': -3.3445019721984863, 'rewards/accuracies': 0.9750000238418579, 'rewards/margins': 5.245241641998291, 'logps/rejected': -406.4423522949219, 'logps/chosen': -343.7262268066406, 'logits/rejected': -60.882476806640625, 'logits/chosen': -62.21611404418945, 'epoch': 1.02}
{'loss': 0.2638, 'grad_norm': 19.947404861450195, 'learning_rate': 1.4944444444444444e-05, 'rewards/chosen': 2.4832510948181152, 'rewards/rejected': -3.2184066772460938, 'rewards/accuracies': 0.949999988079071, 'rewards/margins': 5.701657772064209, 'logps/rejected': -427.52197265625, 'logps/chosen': -391.9784240722656, 'logits/rejected': -61.08870315551758, 'logits/chosen': -60.74915313720703, 'epoch': 1.04}
{'loss': 0.0845, 'grad_norm': 53.526275634765625, 'learning_rate': 1.4346666666666666e-05, 'rewards/chosen': 1.8216516971588135, 'rewards/rejected': -3.6578879356384277, 'rewards/accuracies': 0.96875, 'rewards/margins': 5.47953987121582, 'logps/rejected': -395.97344970703125, 'logps/chosen': -302.2493896484375, 'logits/rejected': -63.5843391418457, 'logits/chosen': -65.01161193847656, 'epoch': 1.05}
{'loss': 0.0635, 'grad_norm': 7.371969699859619, 'learning_rate': 1.3748888888888888e-05, 'rewards/chosen': 2.2166666984558105, 'rewards/rejected': -4.150076389312744, 'rewards/accuracies': 0.987500011920929, 'rewards/margins': 6.3667426109313965, 'logps/rejected': -420.4534606933594, 'logps/chosen': -342.7113342285156, 'logits/rejected': -61.518348693847656, 'logits/chosen': -63.04655075073242, 'epoch': 1.06}
{'loss': 0.1123, 'grad_norm': 13.665971755981445, 'learning_rate': 1.3151111111111111e-05, 'rewards/chosen': 1.868035912513733, 'rewards/rejected': -3.8165340423583984, 'rewards/accuracies': 0.96875, 'rewards/margins': 5.684570789337158, 'logps/rejected': -403.8265686035156, 'logps/chosen': -335.9253845214844, 'logits/rejected': -63.952598571777344, 'logits/chosen': -65.65969848632812, 'epoch': 1.08}
{'loss': 0.0647, 'grad_norm': 6.713380336761475, 'learning_rate': 1.2553333333333334e-05, 'rewards/chosen': 1.633339524269104, 'rewards/rejected': -3.799732208251953, 'rewards/accuracies': 0.981249988079071, 'rewards/margins': 5.433071613311768, 'logps/rejected': -423.66448974609375, 'logps/chosen': -330.0780029296875, 'logits/rejected': -64.35719299316406, 'logits/chosen': -65.02449035644531, 'epoch': 1.09}
{'loss': 0.119, 'grad_norm': 17.787675857543945, 'learning_rate': 1.1955555555555556e-05, 'rewards/chosen': 1.9392048120498657, 'rewards/rejected': -4.776007652282715, 'rewards/accuracies': 0.9750000238418579, 'rewards/margins': 6.715212345123291, 'logps/rejected': -449.86053466796875, 'logps/chosen': -363.79644775390625, 'logits/rejected': -65.53288269042969, 'logits/chosen': -65.39407348632812, 'epoch': 1.11}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x0000024784B3CFA0>, 'epoch': 1.11}
0% | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.644171953201294, 'eval_runtime': 117.6817, 'eval_samples_per_second': 10.928, 'eval_steps_per_second': 1.368, 'eval_rewards/chosen': 0.4069330096244812, 'eval_rewards/rejected': -2.869373321533203, 'eval_rewards/accuracies': 0.8281574249267578, 'eval_rewards/margins': 3.276306629180908, 'eval_logps/rejected': -414.2481689453125, 'eval_logps/chosen': -340.2374572753906, 'eval_logits/rejected': -64.66106414794922, 'eval_logits/chosen': -65.4554443359375, 'epoch': 1.11}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
{'loss': 0.1039, 'grad_norm': 3.904911994934082, 'learning_rate': 1.1357777777777777e-05, 'rewards/chosen': 1.911709189414978, 'rewards/rejected': -4.101197242736816, 'rewards/accuracies': 0.9750000238418579, 'rewards/margins': 6.012906551361084, 'logps/rejected': -411.5382385253906, 'logps/chosen': -335.3758850097656, 'logits/rejected': -64.78614044189453, 'logits/chosen': -65.04131317138672, 'epoch': 1.12}
{'loss': 0.0957, 'grad_norm': 6.655702590942383, 'learning_rate': 1.076e-05, 'rewards/chosen': 1.8591140508651733, 'rewards/rejected': -3.2132465839385986, 'rewards/accuracies': 0.9750000238418579, 'rewards/margins': 5.072360992431641, 'logps/rejected': -390.045166015625, 'logps/chosen': -317.8333740234375, 'logits/rejected': -63.70588302612305, 'logits/chosen': -65.53626251220703, 'epoch': 1.13}
{'loss': 0.1196, 'grad_norm': 10.346579551696777, 'learning_rate': 1.0162222222222222e-05, 'rewards/chosen': 1.9247852563858032, 'rewards/rejected': -3.6962714195251465, 'rewards/accuracies': 0.949999988079071, 'rewards/margins': 5.621057033538818, 'logps/rejected': -442.06646728515625, 'logps/chosen': -354.2452087402344, 'logits/rejected': -65.61262512207031, 'logits/chosen': -65.46220397949219, 'epoch': 1.15}
{'loss': 0.0603, 'grad_norm': 7.645101547241211, 'learning_rate': 9.564444444444445e-06, 'rewards/chosen': 2.158538818359375, 'rewards/rejected': -3.908759355545044, 'rewards/accuracies': 0.96875, 'rewards/margins': 6.06729793548584, 'logps/rejected': -418.2589416503906, 'logps/chosen': -330.3628845214844, 'logits/rejected': -66.86517333984375, 'logits/chosen': -67.24533081054688, 'epoch': 1.16}
{'loss': 0.1123, 'grad_norm': 9.361799240112305, 'learning_rate': 8.966666666666667e-06, 'rewards/chosen': 1.8783397674560547, 'rewards/rejected': -3.651076555252075, 'rewards/accuracies': 0.96875, 'rewards/margins': 5.529416084289551, 'logps/rejected': -393.2381896972656, 'logps/chosen': -335.4851989746094, 'logits/rejected': -67.01258850097656, 'logits/chosen': -66.44461822509766, 'epoch': 1.17}
{'loss': 0.1506, 'grad_norm': 12.935763359069824, 'learning_rate': 8.368888888888888e-06, 'rewards/chosen': 1.7521289587020874, 'rewards/rejected': -3.634990692138672, 'rewards/accuracies': 0.949999988079071, 'rewards/margins': 5.387120246887207, 'logps/rejected': -386.4699401855469, 'logps/chosen': -324.6338806152344, 'logits/rejected': -66.38020324707031, 'logits/chosen': -66.79615783691406, 'epoch': 1.19}
{'loss': 0.5986, 'grad_norm': 9.284168243408203, 'learning_rate': 7.771111111111111e-06, 'rewards/chosen': 1.9469410181045532, 'rewards/rejected': -3.7310092449188232, 'rewards/accuracies': 0.96875, 'rewards/margins': 5.677950859069824, 'logps/rejected': -411.90435791015625, 'logps/chosen': -364.34222412109375, 'logits/rejected': -66.46623229980469, 'logits/chosen': -66.51963806152344, 'epoch': 1.2}
{'loss': 0.1027, 'grad_norm': 20.745498657226562, 'learning_rate': 7.173333333333333e-06, 'rewards/chosen': 2.3401854038238525, 'rewards/rejected': -3.038395404815674, 'rewards/accuracies': 0.9624999761581421, 'rewards/margins': 5.3785810470581055, 'logps/rejected': -406.11322021484375, 'logps/chosen': -338.1979675292969, 'logits/rejected': -66.12078094482422, 'logits/chosen': -66.09214782714844, 'epoch': 1.22}
{'loss': 0.1437, 'grad_norm': 22.323652267456055, 'learning_rate': 6.575555555555555e-06, 'rewards/chosen': 2.4637179374694824, 'rewards/rejected': -3.6244139671325684, 'rewards/accuracies': 0.956250011920929, 'rewards/margins': 6.088131904602051, 'logps/rejected': -410.036865234375, 'logps/chosen': -298.9726867675781, 'logits/rejected': -66.15055084228516, 'logits/chosen': -67.33503723144531, 'epoch': 1.23}
{'loss': 0.1427, 'grad_norm': 4.509947776794434, 'learning_rate': 5.977777777777778e-06, 'rewards/chosen': 2.6652581691741943, 'rewards/rejected': -3.020510196685791, 'rewards/accuracies': 0.9624999761581421, 'rewards/margins': 5.685769081115723, 'logps/rejected': -399.15313720703125, 'logps/chosen': -329.51861572265625, 'logits/rejected': -67.18165588378906, 'logits/chosen': -67.73406982421875, 'epoch': 1.24}
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x000002474D7F9D90>, 'epoch': 1.24}
```



```
0%|          | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.6730199456214905, 'eval_runtime': 115.8401, 'eval_samples_per_second': 11.102, 'eval_steps_per_second': 1.39, 'eval_rewards/chosen': 1.1133185625076294, 'eval_rewards/rejected': -1.9897171258926392, 'eval_rewards/accuracies': 0.8131469488143921, 'eval_rewards/margins': 3.1030356884002686, 'eval_logps/rejected': -409.849853515625, 'eval_logps/chosen': -336.70556640625, 'eval_logits/rejected': -65.83477783203125, 'eval_logits/chosen': -66.72872161865234, 'epoch': 1.24}

c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:295: FutureWarning: `torch.cpu.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cpu', args...)` instead.
  with torch.enable_grad(), device_autocast_ctx, torch.cpu.amp.autocast(**ctx.cpu_autocast_kwargs): # type: ignore[attr-defined]
{'loss': 0.1846, 'grad_norm': 46.18772888183594, 'learning_rate': 5.38e-06, 'rewards/chosen': 2.432614326477051, 'rewards/rejected': -3.799328565597534, 'rewards/accuracies': 0.949999988079071, 'rewards/margins': 6.231942653656006, 'logps/rejected': -419.77239990234375, 'logps/chosen': -336.6206970214844, 'logits/rejected': -66.97927856445312, 'logits/chosen': -66.82938385009766, 'epoch': 1.26}
{'loss': 0.1087, 'grad_norm': 8.702348709106445, 'learning_rate': 4.782222222222226e-06, 'rewards/chosen': 2.0061357021331787, 'rewards/rejected': -3.8021392822265625, 'rewards/accuracies': 0.96875, 'rewards/margins': 5.80827522277832, 'logps/rejected': -417.48211669921875, 'logps/chosen': -344.32440185546875, 'logits/rejected': -66.13191223144531, 'logits/chosen': -66.78289031982422, 'epoch': 1.27}
{'loss': 0.0955, 'grad_norm': 8.919349670410156, 'learning_rate': 4.184444444444444e-06, 'rewards/chosen': 2.3867013454437256, 'rewards/rejected': -3.760941743850708, 'rewards/accuracies': 0.96875, 'rewards/margins': 6.147643089294434, 'logps/rejected': -435.34796142578125, 'logps/chosen': -372.0423583984375, 'logits/rejected': -66.47483825683594, 'logits/chosen': -67.8721694946289, 'epoch': 1.29}
{'loss': 0.1056, 'grad_norm': 19.481855392456055, 'learning_rate': 3.5866666666666665e-06, 'rewards/chosen': 2.1070446968078613, 'rewards/rejected': -3.533687114715576, 'rewards/accuracies': 0.9624999761581421, 'rewards/margins': 5.640732288360596, 'logps/rejected': -391.87481689453125, 'logps/chosen': -305.5421447753906, 'logits/rejected': -66.63484191894531, 'logits/chosen': -67.28287506103516, 'epoch': 1.3}
{'loss': 0.1037, 'grad_norm': 23.20883560180664, 'learning_rate': 2.988888888888889e-06, 'rewards/chosen': 2.509688377380371, 'rewards/rejected': -3.807267665863037, 'rewards/accuracies': 0.96875, 'rewards/margins': 6.316956043243408, 'logps/rejected': -425.3314514160156, 'logps/chosen': -325.8824768066406, 'logits/rejected': -64.90129852294922, 'logits/chosen': -66.94100952148438, 'epoch': 1.31}
{'loss': 0.0931, 'grad_norm': 28.788654327392578, 'learning_rate': 2.3911111111111113e-06, 'rewards/chosen': 2.178412675857544, 'rewards/rejected': -3.834437608718872, 'rewards/accuracies': 0.96875, 'rewards/margins': 6.012849807739258, 'logps/rejected': -422.89215087890625, 'logps/chosen': -313.00494384765625, 'logits/rejected': -67.72181701660156, 'logits/chosen': -68.94268798828125, 'epoch': 1.33}
{'loss': 0.1153, 'grad_norm': 9.82227611541748, 'learning_rate': 1.7933333333333332e-06, 'rewards/chosen': 2.7431321144104004, 'rewards/rejected': -3.94225811958313, 'rewards/accuracies': 0.9624999761581421, 'rewards/margins': 6.685390472412109, 'logps/rejected': -427.3273010253906, 'logps/chosen': -371.89990234375, 'logits/rejected': -67.59329223632812, 'logits/chosen': -67.47795867919922, 'epoch': 1.34}
{'loss': 0.1147, 'grad_norm': 8.28013801574707, 'learning_rate': 1.1955555555555556e-06, 'rewards/chosen': 2.537787437438965, 'rewards/rejected': -3.677978992462158, 'rewards/accuracies': 0.956250011920929, 'rewards/margins': 6.215766906738281, 'logps/rejected': -399.5307922363281, 'logps/chosen': -319.8611755371094, 'logits/rejected': -66.5009994506836, 'logits/chosen': -67.45075988769531, 'epoch': 1.35}
{'loss': 0.0852, 'grad_norm': 12.39801025390625, 'learning_rate': 5.977777777777778e-07, 'rewards/chosen': 2.178774118423462, 'rewards/rejected': -3.451228380203247, 'rewards/accuracies': 0.9624999761581421, 'rewards/margins': 5.630002975463867, 'logps/rejected': -410.73822021484375, 'logps/chosen': -362.87518310546875, 'logits/rejected': -66.65019989013672, 'logits/chosen': -66.91692352294922, 'epoch': 1.37}
{'loss': 0.1022, 'grad_norm': 22.168128967285156, 'learning_rate': 0.0, 'rewards/chosen': 2.5318286418914795, 'rewards/rejected': -3.725522994995117, 'rewards/accuracies': 0.96875, 'rewards/margins': 6.257351875305176, 'logps/rejected': -435.3894958496094, 'logps/chosen': -361.06976318359375, 'logits/rejected': -66.29678344726562, 'logits/chosen': -67.5572509765625, 'epoch': 1.38}

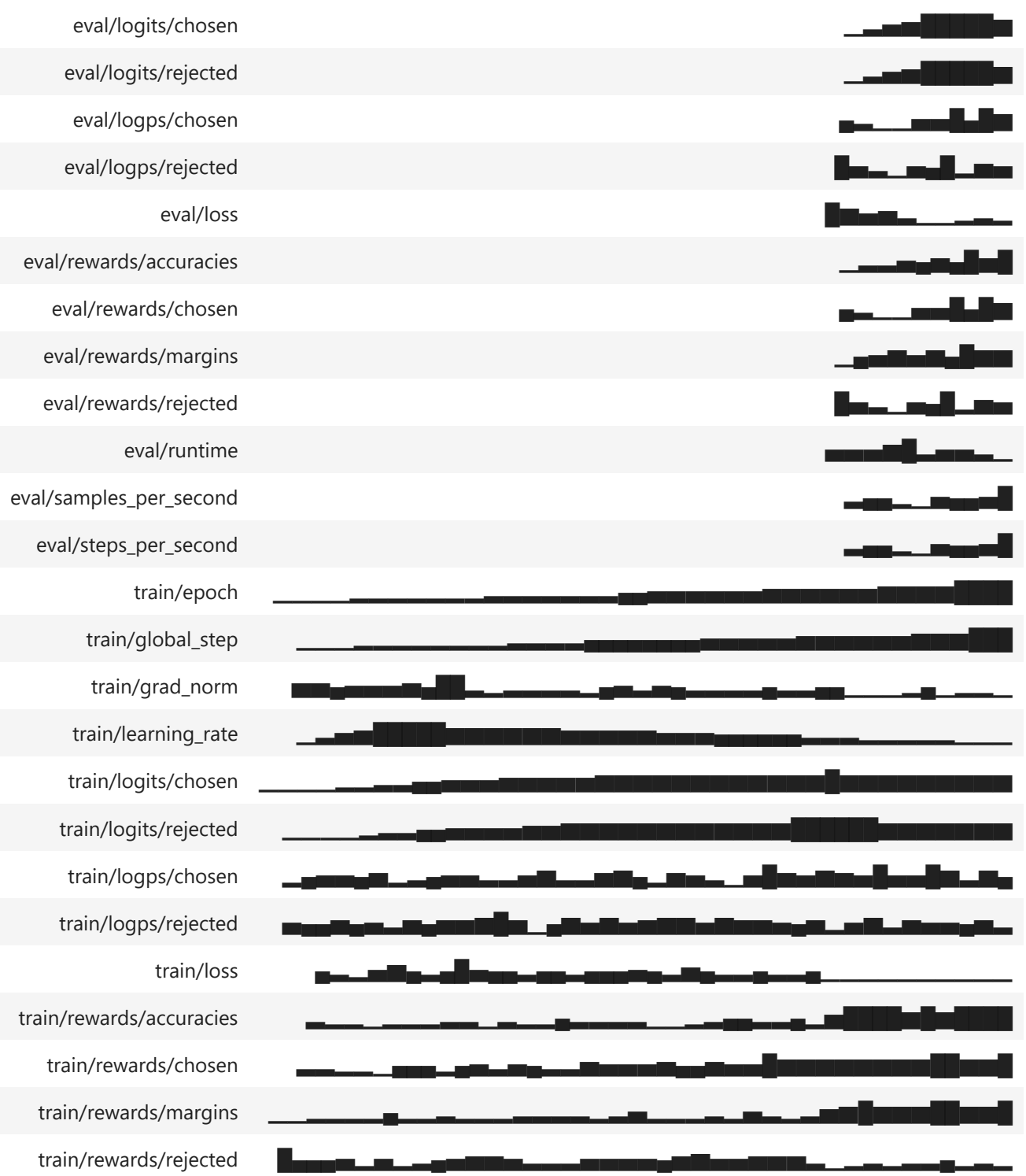
c:\Users\silan\Desktop\A5\.venv\lib\site-packages\torch\utils\checkpoint.py:92: UserWarning: None of the inputs have requires_grad=True. Gradients will be None
  warnings.warn(
{'game_log': <wandb.sdk.data_types.table.Table object at 0x00002474D813F70>, 'epoch': 1.38}
0%|          | 0/161 [00:00<?, ?it/s]
{'eval_loss': 0.6408655047416687, 'eval_runtime': 112.62, 'eval_samples_per_second': 11.419, 'eval_steps_per_second': 1.43, 'eval_rewards/chosen': 0.9778485894203186, 'eval_rewards/rejected': -2.149104356765747, 'eval_rewards/accuracies': 0.8234990239143372, 'eval_rewards/margins': 3.126952648162842, 'eval_logps/rejected': -410.6468505859375, 'eval_logps/chosen': -337.3829040527344, 'eval_logits/rejected': -66.98162841796875, 'eval_logits/chosen': -67.84807586669922, 'epoch': 1.38}
{'train_runtime': 4061.3209, 'train_samples_per_second': 3.94, 'train_steps_per_second': 0.123, 'train_loss': 0.5710406829714775, 'epoch': 1.38}

Out[8]: TrainOutput(global_step=500, training_loss=0.5710406829714775, metrics={'train_runtime': 4061.3209, 'train_samples_per_second': 3.94, 'train_steps_per_second': 0.123, 'total_flos': 0.0, 'train_loss': 0.5710406829714775, 'epoch': 1.38217000691085})

In [9]: # --- Saving ---
dpo_trainer.save_model("./trained_model_orca_dpo_wandb")

In [10]: # --- Finish wandb run ---
wandb.finish()
```

Run history:



Run summary:

eval/logits/chosen	-67.84808
eval/logits/rejected	-66.98163
eval/logps/chosen	-337.3829
eval/logps/rejected	-410.64685

eval/loss	0.64087
eval/rewards/accuracies	0.8235
eval/rewards/chosen	0.97785
eval/rewards/margins	3.12695
eval/rewards/rejected	-2.1491
eval/runtime	112.62
eval/samples_per_second	11.419
eval/steps_per_second	1.43
total_flos	0
train/epoch	1.38217
train/global_step	500
train/grad_norm	22.16813
train/learning_rate	0
train/logits/chosen	-67.55725
train/logits/rejected	-66.29678
train/logps/chosen	-361.06976
train/logps/rejected	-435.3895
train/loss	0.1022
train/rewards/accuracies	0.96875
train/rewards/chosen	2.53183
train/rewards/margins	6.25735
train/rewards/rejected	-3.72552
train_loss	0.57104
train_runtime	4061.3209
train_samples_per_second	3.94
train_steps_per_second	0.123

View run **spring-music-1** at: <https://wandb.ai/sila-nmht-asian-institute-of-technology/dpo-training/runs/b5jpbbnp>

View project at: <https://wandb.ai/sila-nmht-asian-institute-of-technology/dpo-training>

Synced 5 W&B file(s), 10 media file(s), 20 artifact file(s) and 0 other file(s)

Find logs at: `.\wandb\run-20250302_162206-b5jpbbnp\logs`

In []:

```
from huggingface_hub import login

login()
dpo_trainer.push_to_hub()
```

Upload 2 LFS files: 0%| | 0/2 [00:00<?, ?it/s]
model.safetensors: 0%| | 0.00/498M [00:00<?, ?B/s]
training_args.bin: 0%| | 0.00/5.18k [00:00<?, ?B/s]


```
Out[ ]: CommitInfo(commit_url='https://huggingface.co/silanm/nlp-a5/commit/9c2eccfc0a7c23c6f8e54e323523296a810202a6', commit_message='End of training', commit_description='', oid='9c2eccfc0a7c23c6f8e54e323523296a810202a6', pr_url=None, repo_url=RepoUrl('https://huggingface.co/silanm/nlp-a5', endpoint='https://huggingface.co', repo_type='model', repo_id='silanm/nlp-a5'), pr_revision=None, pr_num=None)
```