

Finetune LLMs with GRPO

This notebook shows how to finetune an LLM with GRPO, using the `trl` library.

It's by [Ben Burtenshaw](#) and [Maxime Labonne](#).

This is a minimal example. For a complete example, refer to the GRPO chapter in the [course](#).

Install dependencies

```
In [1]: !pip install -qqq datasets==3.2.0 transformers==4.47.1 trl==0.14.0 peft==0.14.0 accelerate==1.2.1 bitsandbytes==0.40.1
!pip install -qqq flash-attn --no-build-isolation --progress-bar off
```

```
In [16]: import torch
print(torch.version.cuda)

12.1
```

Load Dataset

```
In [1]: from huggingface_hub import notebook_login
notebook_login()

VBox(children=(HTML(value='<center> <img\nsrc=https://huggingface.co/front/assets/huggingface_logo-noborder.svg...'
```

```
In [2]: from dotenv import load_dotenv
import os
import wandb

# Load environment variables from .env
load_dotenv()

# W&B reads your key from environment automatically
wandb.login() # no key needed in notebook
```

```
wandb: (1) Create a W&B account
wandb: (2) Use an existing W&B account
wandb: (3) Don't visualize my results
wandb: Enter your choice:wandb: You chose 'Use an existing W&B account'
wandb: Logging into https://api.wandb.ai. (Learn how to deploy a W&B server locally: https://wandb.me/wandb-server)
wandb: Find your API key here: https://wandb.ai/authorize?ref=models
wandb: Paste an API key from your profile and hit enter:wandb: No netrc file found, creating one.
wandb: Appending key for api.wandb.ai to your netrc file: /root/.netrc
wandb: Currently logged in as: beshoyarnest01 (beshoyarnest01-minia-university) to https://api.wandb.ai. Use `wandb login --relogin` to force relogin
```

```
Out[2]: True
```

```
In [5]: import torch
from datasets import load_dataset
from peft import LoraConfig, get_peft_model
from transformers import AutoModelForCausalLM, AutoTokenizer
from trl import GRPOConfig, GRPOTrainer
```

```
In [6]: # Load dataset
dataset = load_dataset("mlabonne/smoltdlr")
print(dataset)
```

```
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:104: UserWarning:
Error while fetching `HF_TOKEN` secret value from your vault: 'Requesting secret HF_TOKEN timed out. Secrets can only be fetched when running from the Colab UI.'
You are not authenticated with the Hugging Face Hub in this notebook.
If the error persists, please let us know by opening an issue on GitHub (https://github.com/huggingface/huggingface_hub/issues/new).
warnings.warn(
```

```
README.md: 0%|          | 0.00/981 [00:00<?, ?B/s]
data/train-00000-of-00001.parquet: 0%|          | 0.00/1.44M [00:00<?, ?B/s]
data/validation-00000-of-00001.parquet: 0%|          | 0.00/151k [00:00<?, ?B/s]
data/test-00000-of-00001.parquet: 0%|          | 0.00/151k [00:00<?, ?B/s]
Generating train split: 0%|          | 0/2000 [00:00<?, ? examples/s]
Generating validation split: 0%|          | 0/200 [00:00<?, ? examples/s]
Generating test split: 0%|          | 0/200 [00:00<?, ? examples/s]
```

```
DatasetDict({
  train: Dataset({
    features: ['prompt', 'completion'],
    num_rows: 2000
  })
  validation: Dataset({
    features: ['prompt', 'completion'],
    num_rows: 200
  })
  test: Dataset({
    features: ['prompt', 'completion'],
    num_rows: 200
  })
})
```

Load Model

```
In [7]: # Load model with fallback if flash_attn is not available
model_id = "HuggingFaceTB/SmolLM-135M-Instruct"
try:
    model = AutoModelForCausalLM.from_pretrained(
        model_id,
        torch_dtype="auto",
        device_map="auto",
        attn_implementation="flash_attention_2",
    )
except ImportError:
    model = AutoModelForCausalLM.from_pretrained(
        model_id,
        torch_dtype="auto",
        device_map="auto",
    )
tokenizer = AutoTokenizer.from_pretrained(model_id)

config.json: 0%|          | 0.00/723 [00:00<?, ?B/s]
`torch_dtype` is deprecated! Use `dtype` instead!
model.safetensors: 0%|          | 0.00/269M [00:00<?, ?B/s]
generation_config.json: 0%|          | 0.00/156 [00:00<?, ?B/s]
tokenizer_config.json: 0.00B [00:00, ?B/s]
vocab.json: 0.00B [00:00, ?B/s]
merges.txt: 0.00B [00:00, ?B/s]
tokenizer.json: 0.00B [00:00, ?B/s]
special_tokens_map.json: 0%|          | 0.00/565 [00:00<?, ?B/s]
```

LoRA Config

```
In [8]: # Load LoRA
lora_config = LoraConfig(
    task_type="CAUSAL_LM",
    r=16,
    lora_alpha=32,
    target_modules="all-linear",
)
```

```
In [9]: model = get_peft_model(model, lora_config)
print(model.print_trainable_parameters())
```

```
trainable params: 4,884,480 || all params: 139,399,488 || trainable%: 3.5039
None
```

Define Reward Function

```
In [10]: # Reward function
def reward_len(completions, **kwargs):
    return [-abs(50 - len(completion)) for completion in completions]
```

Define Training Arguments

```
In [23]: # Training arguments
training_args = GRPOConfig(
    output_dir="GRPO",
    learning_rate=2e-5,
    per_device_train_batch_size=8,
    gradient_accumulation_steps=2,
    max_prompt_length=512,
    max_completion_length=96,
    num_generations=8,
    optim="adamw_torch",
```

```

num_train_epochs=1,
bf16=True,
report_to=["wandb"],
remove_unused_columns=False,
logging_steps=1,
)

```

<string>:196: FutureWarning: The `max_prompt_length` argument is deprecated and will be removed in version 0.28.0. You should instead filter your dataset before training to ensure that prompts do not exceed your desired length.

Initialize the trainer

```

In [24]: # Trainer
trainer = GRPOTrainer(
    model=model,
    reward_funcs=[reward_len],
    args=training_args,
    train_dataset=dataset["train"],
)

```

The model is already on multiple devices. Skipping the move to device specified in `args`.

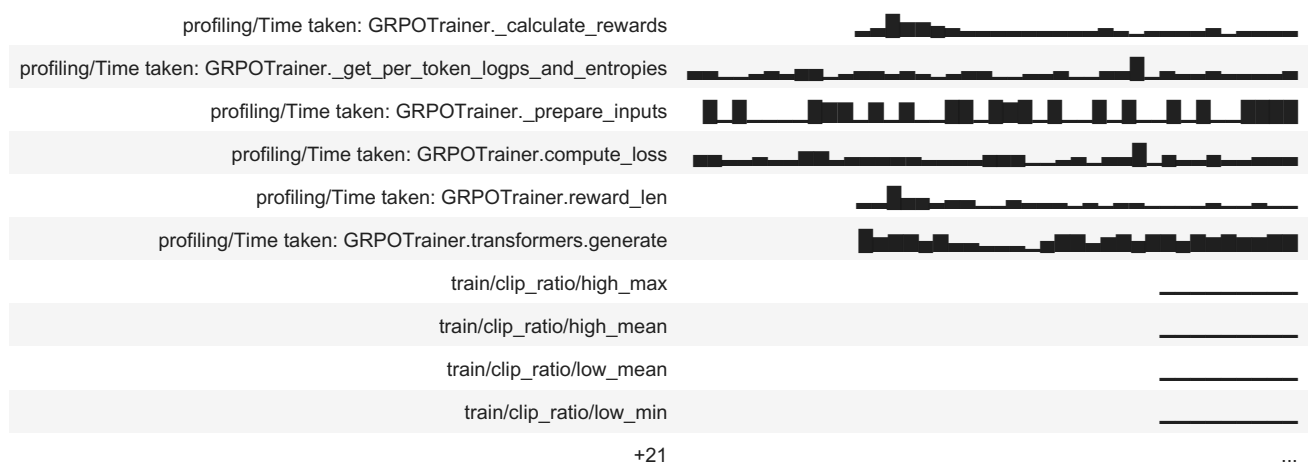
```

In [25]: wandb.init(project="GRPO")

```

Finishing previous runs because reinit is set to 'default'.

Run history:



Run summary:

profiling/Time taken: GRPOTrainer._calculate_rewards	0.00079
profiling/Time taken: GRPOTrainer._get_per_token_logps_and_entropies	0.33982
profiling/Time taken: GRPOTrainer._prepare_inputs	8.95228
profiling/Time taken: GRPOTrainer.compute_loss	0.49418
profiling/Time taken: GRPOTrainer.reward_len	8e-05
profiling/Time taken: GRPOTrainer.transformers.generate	8.91702
train/clip_ratio/high_max	0
train/clip_ratio/high_mean	0
train/clip_ratio/low_mean	0
train/clip_ratio/low_min	0
+21	...

View run **worthy-feather-5** at: <https://wandb.ai/beshoyarnest01-minia-university/GRPO/runs/nyec78pm>

View project at: <https://wandb.ai/beshoyarnest01-minia-university/GRPO>

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

Find logs at: `./wandb/run-20260122_101306-nyec78pm/logs`

Tracking run with wandb version 0.24.0

Run data is saved locally in `/content/wandb/run-20260122_101928-mxx8ehzz`

Syncing run **distinctive-planet-6** to [Weights & Biases \(docs\)](#)

View project at <https://wandb.ai/beshoyarnest01-minia-university/GRPO>

View run at <https://wandb.ai/beshoyarnest01-minia-university/GRPO/runs/mxx8ehzz>

Out[25]: <wandb.sdk.wandb_run.Run object at 0x7aab79015f70>

```
In [26]: # Train model
trainer.train()
```

[1000/1000 1:49:09, Epoch 1/1]

Step	Training Loss
------	---------------

1	0.156700
---	----------

2	0.186400
---	----------

3	0.191700
---	----------

4	0.314500
---	----------

5	0.088000
---	----------

6	0.284900
---	----------

7	0.272900
---	----------

8	0.235400
---	----------

9	0.230800
---	----------

10	0.161600
----	----------

11	0.162500
----	----------

12	0.324300
----	----------

13	0.089700
----	----------

14	0.178900
----	----------

15	0.216100
----	----------

16	0.164700
----	----------

17	0.135400
----	----------

18	0.155700
----	----------

19	0.086600
----	----------

20	0.245300
----	----------

21	0.204800
----	----------

22	0.338700
----	----------

23	0.133100
----	----------

24	0.222200
----	----------

25	0.229900
----	----------

26	0.282300
----	----------

27	0.146300
----	----------

28	0.296900
----	----------

29	0.319200
----	----------

30	0.294600
----	----------

31	0.293800
----	----------

32	-0.000000
----	-----------

33	0.294400
----	----------

34	0.104000
----	----------

35	0.244000
----	----------

36	0.293800
----	----------

37	0.151800
----	----------

38	0.189100
----	----------

39	0.065500
----	----------

40	0.277600
----	----------

41	0.053000
----	----------

42	0.070400
43	0.090400
44	0.080100
45	0.247200
46	0.195500
47	0.102000
48	0.347600
49	0.290400
50	0.158600
51	0.421200
52	0.297300
53	0.158200
54	0.256800
55	0.286900
56	0.232800
57	0.189600
58	0.269900
59	0.288000
60	0.119000
61	0.153900
62	0.412700
63	0.327400
64	0.228600
65	0.361400
66	0.328100
67	0.266900
68	0.364300
69	0.375100
70	0.398000
71	0.324100
72	0.280100
73	0.372600
74	0.431700
75	0.450900
76	0.409000
77	0.279100
78	0.237100
79	0.455300
80	0.404800
81	0.410600
82	0.318000
83	0.492400
84	0.258200
85	0.534200
86	0.372500
87	0.386300
88	0.460700
89	0.369900
90	0.369000
91	0.564800
92	0.463800

93	0.482100
94	0.597900
95	0.378500
96	0.646700
97	0.497000
98	0.403900
99	0.619100
100	0.377100
101	0.553900
102	0.479900
103	0.649800
104	0.535500
105	0.666100
106	0.662800
107	0.520900
108	0.487500
109	0.434700
110	0.576200
111	0.392000
112	0.552900
113	0.517700
114	0.680400
115	0.415200
116	0.551500
117	0.434100
118	0.577200
119	0.450500
120	0.398500
121	0.220200
122	0.516600
123	0.513300
124	0.400100
125	0.622500
126	0.244500
127	0.267100
128	0.431800
129	0.422500
130	0.712400
131	0.327100
132	0.408300
133	0.448000
134	0.341200
135	0.377900
136	0.562900
137	0.584100
138	0.413000
139	0.491600
140	0.765800
141	0.318100
142	0.540400
143	0.408000

144	0.222700
145	0.314500
146	0.567700
147	0.637700
148	0.370600
149	0.342900
150	0.360800
151	0.382200
152	0.243900
153	0.246300
154	0.309700
155	0.484500
156	0.638700
157	0.356300
158	0.834400
159	0.489200
160	-0.160100
161	0.336300
162	0.337700
163	0.489600
164	0.405000
165	0.109000
166	0.193000
167	0.198100
168	-0.073200
169	0.363600
170	0.273500
171	0.125600
172	0.535300
173	0.146400
174	0.095100
175	0.269000
176	0.303600
177	0.089600
178	-0.171700
179	0.254000
180	0.063300
181	0.781400
182	0.176900
183	0.403400
184	0.933300
185	0.208700
186	-0.170800
187	-0.044600
188	0.521400
189	-0.178500
190	0.151200
191	-0.269200
192	0.219600
193	0.353400

194	0.036100
195	0.410900
196	0.079300
197	0.203500
198	0.116400
199	0.190000
200	-0.346700
201	0.009100
202	0.058600
203	0.181000
204	0.078400
205	0.092500
206	-0.019200
207	-0.043100
208	0.761800
209	0.468700
210	0.173600
211	0.248500
212	0.191700
213	-0.140200
214	0.042500
215	0.057800
216	0.130000
217	0.022600
218	0.325800
219	-0.202800
220	0.105800
221	0.175300
222	0.169300
223	0.132100
224	0.026000
225	-0.361000
226	-0.213100
227	-0.111000
228	0.083700
229	-0.023300
230	-0.156600
231	0.749400
232	-0.138400
233	0.117100
234	0.048500
235	-0.131100
236	-0.038000
237	0.119600
238	-0.021500
239	0.056300
240	0.242700
241	0.003300
242	-0.023200
243	-0.163900
244	0.293100

245	-0.209400
246	0.147000
247	0.128200
248	-0.140700
249	0.713300
250	-0.159300
251	0.149900
252	-0.180600
253	0.021900
254	0.052600
255	-0.001600
256	0.041100
257	0.079200
258	-0.022500
259	-0.054400
260	-0.112900
261	-0.052500
262	0.092300
263	-0.059300
264	-0.013300
265	0.077700
266	0.093400
267	-0.045300
268	0.242000
269	0.161600
270	0.819000
271	-0.300300
272	0.147400
273	0.449900
274	0.219100
275	-0.233500
276	-0.010100
277	0.129700
278	0.021500
279	0.689400
280	-0.106400
281	0.171500
282	0.214800
283	0.225500
284	0.324600
285	0.133100
286	-0.006100
287	-0.051400
288	0.878000
289	0.215800
290	0.169400
291	-0.203300
292	0.084000
293	0.134800
294	0.226900
295	0.195800

296	0.171800
297	0.338200
298	-0.019100
299	-0.113900
300	-0.026200
301	-0.188300
302	0.334500
303	0.064600
304	0.388100
305	0.532700
306	0.371200
307	0.012700
308	-0.051200
309	0.045900
310	-0.010600
311	-0.200600
312	0.165700
313	-0.219100
314	-0.024400
315	-0.127600
316	-0.071400
317	0.057500
318	0.011500
319	-0.194300
320	-0.090100
321	0.085900
322	0.199100
323	0.003500
324	0.050200
325	0.033500
326	-0.112100
327	0.027000
328	0.051600
329	0.191700
330	-0.088300
331	0.096300
332	0.138700
333	-0.113000
334	0.204300
335	0.166700
336	0.043400
337	0.049700
338	0.018500
339	-0.020600
340	0.148900
341	0.200100
342	0.169800
343	0.131600
344	0.078300
345	0.434800
346	0.001700

347	0.229800
348	0.144700
349	-0.042600
350	0.006800
351	0.118100
352	0.172900
353	0.021100
354	0.367900
355	0.105200
356	0.345900
357	0.029600
358	0.204300
359	0.779200
360	-0.146400
361	-0.035000
362	-0.033300
363	0.061500
364	-0.070700
365	0.013300
366	0.114000
367	-0.008700
368	0.120100
369	0.119300
370	0.252100
371	0.231000
372	-0.009500
373	0.180000
374	0.134600
375	0.266900
376	0.728500
377	0.523700
378	0.295900
379	0.495000
380	0.028500
381	0.115100
382	0.088100
383	0.034300
384	0.182200
385	0.126000
386	0.081300
387	-0.007200
388	0.073000
389	0.026500
390	-0.121000
391	-0.004800
392	-0.048800
393	-0.018400
394	-0.118800
395	0.681200
396	-0.082400
397	-0.055700

398	-0.180400
399	-0.073600
400	0.185400
401	-0.119900
402	0.134000
403	0.080000
404	-0.079400
405	-0.028900
406	-0.054200
407	0.094300
408	-0.119500
409	0.026100
410	0.066500
411	0.081800
412	0.263100
413	0.109000
414	0.028300
415	0.225700
416	-0.077200
417	-0.075000
418	-0.044000
419	-0.063000
420	0.035300
421	-0.036600
422	0.723100
423	0.127100
424	0.050100
425	-0.040300
426	0.365100
427	0.126000
428	0.115800
429	0.151000
430	0.251000
431	0.070900
432	0.048700
433	0.013600
434	-0.071800
435	0.155300
436	0.574000
437	-0.011800
438	-0.034600
439	0.020400
440	0.203400
441	-0.097700
442	0.077700
443	0.141400
444	0.046200
445	-0.077100
446	0.251300
447	-0.032200
448	0.110300

449	0.044900
450	0.173300
451	0.018500
452	-0.003800
453	0.113200
454	0.359400
455	0.664100
456	0.020600
457	-0.052900
458	0.286900
459	0.142700
460	0.156800
461	0.177100
462	0.701300
463	-0.029800
464	0.024100
465	0.060500
466	0.048100
467	-0.103500
468	-0.170400
469	0.716500
470	0.132600
471	0.135200
472	0.037800
473	-0.028800
474	-0.034900
475	0.096500
476	0.024100
477	0.232800
478	0.098100
479	0.396300
480	0.065200
481	0.015100
482	0.004500
483	-0.028000
484	-0.000700
485	-0.025000
486	0.151600
487	0.254900
488	0.067500
489	-0.066100
490	-0.146500
491	0.114000
492	0.786100
493	0.033600
494	-0.005100
495	-0.137800
496	-0.034300
497	0.140000
498	-0.104100
499	-0.033100

500	-0.036300
501	0.147600
502	0.352900
503	0.113300
504	-0.143000
505	0.272600
506	-0.022800
507	-0.162000
508	0.022400
509	-0.006900
510	-0.010300
511	-0.090600
512	0.238900
513	-0.013600
514	-0.082800
515	-0.037100
516	-0.093300
517	0.003900
518	0.037400
519	0.226700
520	-0.049100
521	0.134100
522	0.186500
523	0.095400
524	0.589800
525	-0.072400
526	0.703500
527	0.142300
528	-0.148200
529	0.108800
530	0.139900
531	0.069100
532	-0.075300
533	0.463500
534	0.145200
535	0.034800
536	0.205000
537	-0.005700
538	0.130000
539	0.243000
540	0.035700
541	0.435800
542	0.138000
543	0.055600
544	-0.205600
545	-0.055000
546	0.090700
547	-0.017500
548	0.013100
549	-0.033300

550	0.164200
551	0.096900
552	0.150700
553	0.014800
554	-0.106700
555	0.080600
556	0.063000
557	0.213500
558	0.708000
559	0.069100
560	0.510300
561	0.124300
562	0.031900
563	0.122400
564	0.019400
565	0.189100
566	-0.055100
567	-0.141300
568	0.053300
569	0.175100
570	0.034800
571	0.228000
572	0.010200
573	0.090900
574	0.326400
575	0.002000
576	0.042000
577	-0.008800
578	-0.057900
579	-0.163800
580	0.070500
581	0.220400
582	0.264000
583	0.019300
584	-0.092600
585	-0.037700
586	-0.051900
587	0.258500
588	-0.125000
589	0.007000
590	-0.176300
591	-0.018600
592	-0.005000
593	0.043200
594	0.043300
595	0.727800
596	-0.037100
597	0.156300
598	0.071000
599	0.668100
600	0.056400

601	0.083700
602	-0.135700
603	0.699300
604	0.073600
605	0.016800
606	0.174700
607	0.093100
608	-0.002300
609	0.042600
610	0.087200
611	0.132700
612	-0.040700
613	0.195000
614	0.038900
615	0.079400
616	0.124700
617	-0.184600
618	0.113600
619	0.102500
620	-0.139700
621	0.068800
622	0.005200
623	0.029100
624	-0.127900
625	0.057000
626	-0.036300
627	-0.023100
628	0.050800
629	-0.046300
630	0.028800
631	-0.133900
632	0.064400
633	0.058000
634	0.013000
635	-0.095400
636	0.027500
637	0.168200
638	0.208700
639	-0.038600
640	-0.010800
641	0.039400
642	-0.027400
643	0.009100
644	0.059700
645	-0.081600
646	-0.033900
647	-0.117900
648	0.117100
649	0.208200
650	0.211800
651	-0.054900

652	-0.130700
653	0.163100
654	-0.022100
655	0.032400
656	0.233200
657	0.146300
658	0.092400
659	0.185200
660	0.048900
661	0.132900
662	0.006700
663	0.272200
664	-0.010600
665	0.162700
666	0.362700
667	-0.107600
668	-0.009500
669	-0.020600
670	-0.008100
671	-0.105900
672	-0.016900
673	0.140400
674	0.103100
675	-0.046100
676	-0.026700
677	0.158900
678	0.018100
679	-0.080900
680	-0.093200
681	0.107700
682	0.640100
683	0.338700
684	0.149900
685	0.126400
686	0.220000
687	0.130900
688	-0.001400
689	0.040300
690	-0.022500
691	0.044900
692	-0.078100
693	-0.072800
694	0.156900
695	0.389900
696	0.129800
697	0.133000
698	-0.013700
699	0.073500
700	0.116300
701	-0.039900
702	0.004800

703	0.004100
704	-0.052300
705	0.000200
706	0.035700
707	-0.105300
708	0.337000
709	-0.023800
710	0.022600
711	-0.037100
712	-0.054700
713	0.144400
714	0.660700
715	0.094900
716	0.087700
717	0.148600
718	0.241100
719	-0.042900
720	0.045800
721	-0.036100
722	0.075300
723	-0.102100
724	0.100000
725	0.112100
726	-0.004000
727	0.167500
728	0.132100
729	0.091200
730	-0.201900
731	0.097500
732	-0.006400
733	-0.116700
734	-0.087500
735	0.044100
736	0.131800
737	-0.136400
738	-0.029100
739	-0.038800
740	0.060000
741	0.171200
742	0.118100
743	0.045400
744	0.120800
745	-0.123700
746	-0.138700
747	0.081600
748	-0.133000
749	0.278000
750	-0.055900
751	0.024300
752	-0.100700
753	0.128500

754	0.748100
755	0.037500
756	0.039300
757	0.111000
758	-0.050200
759	-0.023500
760	0.167200
761	0.138500
762	0.037100
763	0.147500
764	0.609000
765	0.203100
766	-0.099000
767	-0.018000
768	0.072800
769	0.072000
770	0.658600
771	-0.018200
772	0.581500
773	-0.175000
774	0.068900
775	0.013900
776	0.049500
777	0.066800
778	0.129200
779	-0.078800
780	0.066200
781	-0.027300
782	0.114200
783	0.114000
784	0.196800
785	0.069700
786	-0.003200
787	0.069500
788	0.629000
789	-0.009500
790	0.045200
791	0.152800
792	0.026100
793	0.388500
794	0.079700
795	0.093600
796	-0.051100
797	0.165200
798	0.036800
799	-0.035400
800	-0.008800
801	-0.040900
802	-0.078800
803	0.072300
804	-0.098800

805	0.180200
806	0.209800
807	-0.019300
808	-0.022500
809	0.061400
810	0.015600
811	0.060900
812	0.127500
813	0.150100
814	-0.104700
815	-0.110200
816	0.200100
817	0.067000
818	-0.016300
819	-0.100300
820	0.056700
821	-0.023900
822	-0.055800
823	0.270100
824	0.057500
825	0.125500
826	0.050500
827	0.059500
828	0.117600
829	0.004000
830	0.117500
831	-0.008300
832	0.019400
833	0.195100
834	0.085600
835	-0.049800
836	0.187700
837	0.098700
838	0.136800
839	-0.077000
840	0.711900
841	0.047800
842	0.224300
843	-0.002900
844	0.079900
845	0.062300
846	-0.010600
847	0.103700
848	0.075500
849	0.045100
850	0.018500
851	-0.046800
852	0.072400
853	0.039400
854	0.135100
855	-0.112800

856	-0.026900
857	-0.295200
858	-0.012700
859	0.150400
860	0.007500
861	-0.030500
862	-0.008800
863	0.038900
864	0.091700
865	-0.062300
866	-0.101600
867	0.106400
868	-0.170700
869	0.105100
870	-0.074100
871	0.020500
872	-0.106300
873	-0.025900
874	0.038000
875	-0.034000
876	0.075700
877	0.173200
878	-0.084900
879	0.105700
880	0.008900
881	0.066500
882	-0.064400
883	0.044900
884	0.064500
885	0.066700
886	0.114200
887	0.061000
888	-0.175800
889	-0.014200
890	0.492400
891	-0.105300
892	0.110200
893	0.042300
894	-0.104400
895	0.007100
896	0.098400
897	0.038700
898	0.061300
899	-0.056500
900	0.102100
901	0.045800
902	0.026500
903	0.072200
904	-0.026900
905	-0.045600

906	0.114600
907	0.075500
908	0.192400
909	0.066800
910	0.009500
911	-0.032100
912	0.079000
913	0.089600
914	0.088200
915	0.014000
916	0.125900
917	-0.016800
918	0.051000
919	-0.034700
920	0.045000
921	0.045100
922	0.116100
923	-0.056100
924	-0.060800
925	0.032400
926	0.141600
927	0.089800
928	-0.023100
929	-0.077100
930	0.027600
931	0.134400
932	0.180000
933	0.016600
934	0.067800
935	-0.005500
936	0.053400
937	-0.015600
938	0.051500
939	0.028100
940	0.068400
941	-0.008300
942	0.136000
943	0.103500
944	0.064700
945	-0.013700
946	0.183900
947	0.057600
948	0.014400
949	0.013400
950	-0.041200
951	0.085400
952	0.010000
953	0.040500
954	0.046200
955	0.011100
956	-0.138300

957	-0.094400
958	0.000900
959	-0.071700
960	0.052400
961	0.166200
962	0.014600
963	0.068200
964	0.070300
965	0.213800
966	-0.017300
967	0.132400
968	-0.012100
969	0.003300
970	0.087000
971	0.103900
972	-0.037100
973	-0.052500
974	-0.085100
975	-0.106900
976	-0.023800
977	0.006600
978	-0.019600
979	-0.014000
980	0.123400
981	-0.068200
982	0.112100
983	0.076600
984	-0.015800
985	0.532200
986	0.060900
987	0.094300
988	0.037100
989	0.086000
990	-0.070500
991	-0.170500
992	-0.023000
993	0.031400
994	0.129700
995	0.146000
996	0.027900
997	0.144500
998	0.045700
999	-0.000800
1000	0.069800

```
Out[26]: TrainOutput(global_step=1000, training_loss=0.12347275123046711, metrics={'train_runtime': 6562.9425, 'train_samples_per_second': 0.305, 'train_steps_per_second': 0.152, 'total_flos': 0.0, 'train_loss': 0.12347275123046711})
```

Push Model to Hub

```
In [33]: # Save model
merged_model = trainer.model.merge_and_unload()
merged_model.push_to_hub( "fine_tuned_SmolGRP0-135M_using_GRP0", private=False, tags=["GRP0"])
```

```
README.md: 0.00B [00:00, ?B/s]
Processing Files (0 / 0)      : |          | 0.00B / 0.00B
New Data Upload              : |          | 0.00B / 0.00B
...90ggj6m/model.safetensors: 9%|9      | 25.1MB / 269MB
```

No files have been modified since last commit. Skipping to prevent empty commit.

WARNING:huggingface_hub.hf_api:No files have been modified since last commit. Skipping to prevent empty commit.

```
Out[33]: CommitInfo(commit_url='https://huggingface.co/7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO/commit/1d8a123b78172cf6dc9b4169ed1a0326f0bece5b', commit_message='Upload LlamaForCausalLM', commit_description='', oid='1d8a123b78172cf6dc9b4169ed1a0326f0bece5b', pr_url=None, repo_url=RepoUrl('https://huggingface.co/7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO', endpoint='https://huggingface.co', repo_type='model', repo_id='7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO'), pr_revision=None, pr_num=None)
```

```
In [38]: tokenizer.save_pretrained("fine_tuned_SmolGRPO-135M_using_GRPO")
```

```
Out[38]: ('fine_tuned_SmolGRPO-135M_using_GRPO/tokenizer_config.json',
'fine_tuned_SmolGRPO-135M_using_GRPO/special_tokens_map.json',
'fine_tuned_SmolGRPO-135M_using_GRPO/chat_template.jinja',
'fine_tuned_SmolGRPO-135M_using_GRPO/vocab.json',
'fine_tuned_SmolGRPO-135M_using_GRPO/merges.txt',
'fine_tuned_SmolGRPO-135M_using_GRPO/added_tokens.json',
'fine_tuned_SmolGRPO-135M_using_GRPO/tokenizer.json')
```

```
In [39]: tokenizer.push_to_hub("fine_tuned_SmolGRPO-135M_using_GRPO", private=False, tags=["GRPO"])
```

```
Out[39]: CommitInfo(commit_url='https://huggingface.co/7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO/commit/9b01868f9efde21e3a61c1da0d517cc89f9bdb52', commit_message='Upload tokenizer', commit_description='', oid='9b01868f9efde21e3a61c1da0d517cc89f9bdb52', pr_url=None, repo_url=RepoUrl('https://huggingface.co/7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO', endpoint='https://huggingface.co', repo_type='model', repo_id='7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO'), pr_revision=None, pr_num=None)
```

Generate Text

```
In [40]: prompt = """
# A long document about the Cat
```

The cat (*Felis catus*), also referred to as the domestic cat or house cat, is a small domesticated carnivorous mammal. It is the only domesticated species of the family Felidae. Advances in archaeology and genetics have shown that the domestication of the cat occurred in the Near East around 7500 BC. It is commonly kept as a pet and farm cat, but also ranges freely as a feral cat avoiding human contact. It is valued by humans for companionship and its ability to kill vermin. Its retractable claws are adapted to killing small prey species such as mice and rats. It has a strong, flexible body, quick reflexes, and sharp teeth, and its night vision and sense of smell are well developed. It is a social species, but a solitary hunter and a crepuscular predator. Cat communication includes vocalizations—including meowing, purring, trilling, hissing, growling, and grunting—as well as body language. It can hear sounds too faint or too high in frequency for human ears, such as those made by small mammals. It secretes and perceives pheromones.

```
"""

messages = [
    {"role": "user", "content": prompt},
]
```

```
In [44]: # Generate text
from transformers import pipeline
from transformers import AutoModelForCausalLM, AutoTokenizer

model = AutoModelForCausalLM.from_pretrained("7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO")
tokenizer = AutoTokenizer.from_pretrained("7beshoyarnest/fine_tuned_SmolGRPO-135M_using_GRPO")

generator = pipeline("text-generation", model=model, tokenizer=tokenizer)

## Or use the model and tokenizer we defined earlier
# generator = pipeline("text-generation", model=model, tokenizer=tokenizer)

generate_kwargs = {
    "max_new_tokens": 256,
    "do_sample": True,
    "temperature": 0.5,
    "min_p": 0.1,
}

generated_text = generator(messages, **generate_kwargs)

# Clean dictionary format
output_dict = {
    "user": messages,
    "assistant": generated_text[0]["generated_text"]
}

# Pretty-print
```



```
import json
print(json.dumps(output_dict, indent=4, ensure_ascii=False))
# print(generated_text)
```

Device set to use cuda:0

```
{
  "user": [
    {
      "role": "user",
      "content": "\n# A long document about the Cat\n\nThe cat (Felis catus), also referred to as the domestic cat or house cat, is a small \ndomesticated carnivorous mammal. It is the only domesticated species of the family Felidae.\nAdvances in archaeology and genetics have shown that the domestication of the cat occurred\nin the Near East around 7500 BC. It is commonly kept as a pet and farm cat, but also ranges\nfreely as a feral cat avoiding human contact. It is valued by humans for companionship and\nits ability to kill vermin. Its retractable claws are adapted to killing small prey species\nsuch as mice and rats. It has a strong, flexible body, quick reflexes, and sharp teeth,\nand its night vision and sense of smell are well developed. It is a social species,\nbut a solitary hunter and a crepuscular predator. Cat communication includes\nvocalizations—including meowing, purring, trilling, hissing, growling, and grunting—as\nwell as body language. It can hear sounds too faint or too high in frequency for human ears,\nsuch as those made by small mammals. It secretes and perceives pheromones.\n\n"
    }
  ],
  "assistant": [
    {
      "role": "user",
      "content": "\n# A long document about the Cat\n\nThe cat (Felis catus), also referred to as the domestic cat or house cat, is a small \ndomesticated carnivorous mammal. It is the only domesticated species of the family Felidae.\nAdvances in archaeology and genetics have shown that the domestication of the cat occurred\nin the Near East around 7500 BC. It is commonly kept as a pet and farm cat, but also ranges\nfreely as a feral cat avoiding human contact. It is valued by humans for companionship and\nits ability to kill vermin. Its retractable claws are adapted to killing small prey species\nsuch as mice and rats. It has a strong, flexible body, quick reflexes, and sharp teeth,\nand its night vision and sense of smell are well developed. It is a social species,\nbut a solitary hunter and a crepuscular predator. Cat communication includes\nvocalizations—including meowing, purring, trilling, hissing, growling, and grunting—as\nwell as body language. It can hear sounds too faint or too high in frequency for human ears,\nsuch as those made by small mammals. It secretes and perceives pheromones.\n\n"
    },
    {
      "role": "assistant",
      "content": "The cat! A beloved and fascinating creature that has captivated humans for centuries. Here's a comprehensive overview of the cat:\n\n**Physical Characteristics:**\n\n* **Body:** The cat's body is a sleek, sleek cat with a slender body, short legs, and a short tail.\n* **Head:** The cat's head is small and rounded, with a pointed snout and a small, expressive eye.\n* **Eyes:** The cat's eyes are large and round, with vertical pupils that are almost black. They are capable of seeing in low light conditions.\n* **Whiskers:** The cat's whiskers are long and sticky, used for detecting prey and communicating with other cats.\n* **Fur:** The cat's fur is soft, smooth, and often has a distinctive pattern of stripes or spots.\n* **Eyes:** The cat's eyes are large and round, with vertical pupils that are almost black. They are capable of seeing in low light conditions.\n* **Tail:** The cat's tail is long and muscular, used for balance and display.\n\n**Behavior:**\n\nThe cat was domesticated in the Near East around 7500 BC, during the Neolithic"
    }
  ]
}
```