

win 本地部署ChatGLM3-6B微调及推理过程

weixin_40078854 已于 2024-04-01 12:22:22 修改

1. 操作系统及显卡信息

设备规格		复制	^
设备名称	liujingmao		
处理器	13th Gen Intel(R) Core(TM) i9-13900K 3.00 GHz		
机带 RAM	128 GB (128 GB 可用)		
设备 ID	D409E937-867A-41A1-8F52-EF3E32AAFB37		
产品 ID	00330-54040-69813-AAOEM		
系统类型	64 位操作系统, 基于 x64 的处理器		
笔和触控	没有可用于此显示器的笔或触控输入		
相关链接 域或工作组 系统保护 高级系统设置			
Windows 规格		复制	^
版本	Windows 11 专业版		
版本	23H2		
安装日期	2024/3/10		
操作系统版本	22631.3296		
体验	Windows Feature Experience Pack 1000.22687.1000.0		
Microsoft 服务协议			
Microsoft 软件许可条款			

CSDN @weixin_40078854

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

PS C:\Users\16014\11m> nvidia-smi.exe

Mon Apr 1 11:37:36 2024

NVIDIA-SMI 551.86			Driver Version: 551.86		CUDA Version: 12.4			
GPU	Name	TCC/WDDM	Bus-Id	Disp.A	Volatile	Uncorr.	ECC	
	Fan	Temp		Perf	Pwr:Usage/Cap	Memory-Usage	GPU-Util	Compute M.
0	NVIDIA GeForce RTX 3090	WDDM	00000000:01:00.0	On	32% 43C P8 29W / 350W	14097MiB / 24576MiB	15%	N/A Default N/A
Processes:								
GPU	GI	CI	PID	Type	Process name	GPU Memory Usage		
ID	ID							
0	N/A	N/A	1124	C+G	C:\Windows\System32\dwm.exe	N/A		
0	N/A	N/A	5824	C+G	..._8wekyb3d8bbwe\WindowsTerminal.exe	N/A		
0	N/A	N/A	5892	C+G	...les\GIGABYTE\Control Center\GCC.exe	N/A		
0	N/A	N/A	6192	C+G	...siveControlPanel\SystemSettings.exe	N/A		
0	N/A	N/A	7416	C+G	C:\Windows\explorer.exe	N/A		
0	N/A	N/A	8968	C+G	...nt.CBS_cw5n1h2txyewy\SearchHost.exe	N/A		
0	N/A	N/A	9000	C+G	...2txyewy\StartMenuExperienceHost.exe	N/A		
0	N/A	N/A	12676	C+G	...ekyb3d8bbwe\PhoneExperienceHost.exe	N/A		
0	N/A	N/A	13532	C+G	...GeForce Experience\NVIDIA Share.exe	N/A		
0	N/A	N/A	13728	C+G	...GeForce Experience\NVIDIA Share.exe	N/A		
0	N/A	N/A	14736	C+G	...CBS_cw5n1h2txyewy\TextInputHost.exe	N/A		
0	N/A	N/A	16712	C+G	... (x86)\letsvpn\app-3.4.0\LetsPRO.exe	N/A		
0	N/A	N/A	16952	C+G	...cal\Microsoft\OneDrive\OneDrive.exe	N/A		
0	N/A	N/A	18200	C+G	...5n1h2txyewy\ShellExperienceHost.exe	N/A		
0	N/A	N/A	19168	C+G	...oogle\Chrome\Application\chrome.exe	N/A		
0	N/A	N/A	19788	C+G	...9\extracted\runtime\WeChatAppEx.exe	N/A		
0	N/A	N/A	19964	C+G	...Programs\Microsoft VS Code\Code.exe	N/A		
0	N/A	N/A	22832	C+G	...crosoft\Edge\Application\msedge.exe	N/A		
0	N/A	N/A	32856	C+G	...wekyb3d8bbwe\XboxGameBarWidgets.exe	N/A		
0	N/A	N/A	57464	C+G	...s\ByteDance\douyin\3.4.0\douyin.exe	N/A		
0	N/A	N/A	61216	C	...anaconda3\envs\chatglm3\python.exe	N/A		

PS C:\Users\16014\11m>

CSDN@weixin_40078854

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

2. ChatGLM3 下载

https://blog.csdn.net/weixin_40078854/article/details/137226677?csdn_share_tail=%7B%22type%22%3A%22blog%22%2C%22rType%22%3A%22article%22%2C%22rId%22%3A%22137226677%22%2C%22source%22%3A%22weixin_40078854%22%7D

2.1 下载源代码

GitHub - THUDM/ChatGLM3: ChatGLM3 series: Open Bilingual Chat LLMs | 开源双语对话语言模型

2.2 下载模型

魔搭社区

git lfs install git clone 魔搭社区

3. 开始微调

3.1 参考官网[ChatGLM3/finetune_demo/README.md at main · THUDM/ChatGLM3 · GitHub](#)

3.2 微调过程

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

C:\Users\16014\.cache\huggingface\modules\transformers_modules\THUDM\chatglm3-6b\a5ba5501eb873d40d48bd0983bd2a8dd006bb838\modeling_chatglm.py:231: UserWarning: 1Torch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)

```
context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,
{'loss': 4.7867, 'grad_norm': 2.4454832077026367, 'learning_rate': 4.983333333333336e-05, 'epoch': 0.0}
{'loss': 4.6426, 'grad_norm': 2.757484197616577, 'learning_rate': 4.966666666666667e-05, 'epoch': 0.0}
{'loss': 4.2527, 'grad_norm': 3.266258716583252, 'learning_rate': 4.950000000000004e-05, 'epoch': 0.0}
{'loss': 4.2002, 'grad_norm': 3.29502010345459, 'learning_rate': 4.933333333333334e-05, 'epoch': 0.0}
{'loss': 4.0861, 'grad_norm': 2.647218942642212, 'learning_rate': 4.916666666666665e-05, 'epoch': 0.0}
{'loss': 3.7217, 'grad_norm': 3.1927905082702637, 'learning_rate': 4.9e-05, 'epoch': 0.0}
{'loss': 3.9412, 'grad_norm': 2.941864490509033, 'learning_rate': 4.883333333333334e-05, 'epoch': 0.0}
{'loss': 3.8096, 'grad_norm': 2.7546098232269287, 'learning_rate': 4.866666666666667e-05, 'epoch': 0.0}
{'loss': 3.7455, 'grad_norm': 2.9791178703308105, 'learning_rate': 4.85e-05, 'epoch': 0.0}
{'loss': 3.6756, 'grad_norm': 3.7696664333343506, 'learning_rate': 4.833333333333334e-05, 'epoch': 0.0}
{'loss': 3.8232, 'grad_norm': 3.2472739219665527, 'learning_rate': 4.8166666666666674e-05, 'epoch': 0.0}
{'loss': 3.6283, 'grad_norm': 3.664025068283081, 'learning_rate': 4.8e-05, 'epoch': 0.0}
{'loss': 3.5375, 'grad_norm': 3.887953519821167, 'learning_rate': 4.783333333333335e-05, 'epoch': 0.0}
{'loss': 3.6832, 'grad_norm': 3.584289073944092, 'learning_rate': 4.766666666666667e-05, 'epoch': 0.0}
{'loss': 3.6945, 'grad_norm': 4.085845470428467, 'learning_rate': 4.75e-05, 'epoch': 0.01}
{'loss': 3.5723, 'grad_norm': 3.9218673706054688, 'learning_rate': 4.733333333333336e-05, 'epoch': 0.01}
{'loss': 3.4973, 'grad_norm': 4.102350234985352, 'learning_rate': 4.716666666666667e-05, 'epoch': 0.01}
{'loss': 3.6479, 'grad_norm': 4.339003562927246, 'learning_rate': 4.7e-05, 'epoch': 0.01}
{'loss': 3.5801, 'grad_norm': 4.34035587310791, 'learning_rate': 4.683333333333334e-05, 'epoch': 0.01}
{'loss': 3.5586, 'grad_norm': 4.213550567626953, 'learning_rate': 4.666666666666667e-05, 'epoch': 0.01}
{'loss': 3.5918, 'grad_norm': 4.60715913772583, 'learning_rate': 4.6500000000000005e-05, 'epoch': 0.01}
{'loss': 3.5363, 'grad_norm': 5.015280723571777, 'learning_rate': 4.633333333333333e-05, 'epoch': 0.01}
{'loss': 3.5158, 'grad_norm': 5.237696170806885, 'learning_rate': 4.616666666666666e-05, 'epoch': 0.01}
{'loss': 3.4984, 'grad_norm': 4.997608184814453, 'learning_rate': 4.60000000000001e-05, 'epoch': 0.01}
{'loss': 3.5619, 'grad_norm': 5.232872009277344, 'learning_rate': 4.583333333333334e-05, 'epoch': 0.01}
{'loss': 3.5852, 'grad_norm': 5.7681708335876465, 'learning_rate': 4.566666666666667e-05, 'epoch': 0.01}
{'loss': 3.5881, 'grad_norm': 4.761264324188232, 'learning_rate': 4.55e-05, 'epoch': 0.01}
{'loss': 3.5068, 'grad_norm': 5.207887649536133, 'learning_rate': 4.533333333333335e-05, 'epoch': 0.01}
{'loss': 3.5859, 'grad_norm': 4.76151180267334, 'learning_rate': 4.516666666666667e-05, 'epoch': 0.01}
{'loss': 3.6053, 'grad_norm': 5.018894195556641, 'learning_rate': 4.5e-05, 'epoch': 0.01}
{'loss': 3.5223, 'grad_norm': 4.6577019691467285, 'learning_rate': 4.48333333333333e-05, 'epoch': 0.01}
{'loss': 3.5529, 'grad_norm': 5.834929466247559, 'learning_rate': 4.466666666666667e-05, 'epoch': 0.01}
{'loss': 3.6383, 'grad_norm': 4.845691680908203, 'learning_rate': 4.450000000000004e-05, 'epoch': 0.01}
{'loss': 3.4523, 'grad_norm': 5.5173234939575195, 'learning_rate': 4.433333333333334e-05, 'epoch': 0.01}
{'loss': 3.5078, 'grad_norm': 5.6048688888549805, 'learning_rate': 4.416666666666665e-05, 'epoch': 0.01}
{'loss': 3.5271, 'grad_norm': 5.077373504638672, 'learning_rate': 4.400000000000006e-05, 'epoch': 0.01}
{'loss': 3.5322, 'grad_norm': 5.08348274230957, 'learning_rate': 4.383333333333334e-05, 'epoch': 0.01}
{'loss': 3.3389, 'grad_norm': 5.091756343841553, 'learning_rate': 4.366666666666666e-05, 'epoch': 0.01}
{'loss': 3.6494, 'grad_norm': 5.9848952293396, 'learning_rate': 4.35e-05, 'epoch': 0.01}
{'loss': 3.4234, 'grad_norm': 5.4784447914123535, 'learning_rate': 4.333333333333334e-05, 'epoch': 0.01}
{'loss': 3.4371, 'grad_norm': 5.165726184844971, 'learning_rate': 4.316666666666667e-05, 'epoch': 0.01}
{'loss': 3.5904, 'grad_norm': 5.5213165283203125, 'learning_rate': 4.3e-05, 'epoch': 0.01}
```

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

```
{"loss": 3.4523, "grad_norm": 5.5173234939575195, "learning_rate": 4.433333333333334e-05, "epoch": 0.01}
{"loss": 3.5078, "grad_norm": 5.6048688888549805, "learning_rate": 4.416666666666665e-05, "epoch": 0.01}
{"loss": 3.5271, "grad_norm": 5.077373504638672, "learning_rate": 4.4000000000000006e-05, "epoch": 0.01}
{"loss": 3.5322, "grad_norm": 5.08348274230957, "learning_rate": 4.383333333333334e-05, "epoch": 0.01}
{"loss": 3.3389, "grad_norm": 5.091756343841553, "learning_rate": 4.366666666666666e-05, "epoch": 0.01}
{"loss": 3.6494, "grad_norm": 5.9848952293396, "learning_rate": 4.35e-05, "epoch": 0.01}
{"loss": 3.4234, "grad_norm": 5.478447914123535, "learning_rate": 4.333333333333334e-05, "epoch": 0.01}
{"loss": 3.4371, "grad_norm": 5.165726184844971, "learning_rate": 4.316666666666667e-05, "epoch": 0.01}
{"loss": 3.5904, "grad_norm": 5.5213165283203125, "learning_rate": 4.3e-05, "epoch": 0.01}
{"loss": 3.5932, "grad_norm": 6.598393440246582, "learning_rate": 4.283333333333335e-05, "epoch": 0.02}
{"loss": 3.54, "grad_norm": 5.622533321380615, "learning_rate": 4.266666666666667e-05, "epoch": 0.02}
{"loss": 3.4025, "grad_norm": 5.838737964630127, "learning_rate": 4.25e-05, "epoch": 0.02}
{"loss": 3.5566, "grad_norm": 6.134575366973877, "learning_rate": 4.23333333333334e-05, "epoch": 0.02}
{"loss": 3.5949, "grad_norm": 5.317560195922852, "learning_rate": 4.216666666666667e-05, "epoch": 0.02}
{"loss": 3.5307, "grad_norm": 5.620532989501953, "learning_rate": 4.2e-05, "epoch": 0.02}
{"loss": 3.4166, "grad_norm": 6.305963516235352, "learning_rate": 4.18333333333334e-05, "epoch": 0.02}
{"loss": 3.4125, "grad_norm": 7.557779788970947, "learning_rate": 4.166666666666667e-05, "epoch": 0.02}
17% | 500/3000 [05:37<26:16, 1.59it/s]*
```

**** Running Evaluation ****

```
Num examples = 50
Batch size = 16
```

CSDN @weixin_40078854

```
warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
{'eval_rouge-1': 30.745222, 'eval_rouge-2': 6.208084, 'eval_rouge-l': 22.782102, 'eval_bleu-4': 0.031044060594805578, 'eval_runtime': 93.4
974, 'eval_samples_per_second': 0.535, 'eval_steps_per_second': 0.043, 'epoch': 0.03}
{"loss": 3.4, "grad_norm": 7.237664222717285, "learning_rate": 3.316666666666667e-05, "epoch": 0.04}
{"loss": 3.4111, "grad_norm": 8.1635160446167, "learning_rate": 3.3e-05, "epoch": 0.04}
{"loss": 3.5139, "grad_norm": 6.7397308349609375, "learning_rate": 3.283333333333333e-05, "epoch": 0.04}
{"loss": 3.4123, "grad_norm": 7.087181568145752, "learning_rate": 3.266666666666667e-05, "epoch": 0.04}
{"loss": 3.2348, "grad_norm": 6.67784309387207, "learning_rate": 3.250000000000004e-05, "epoch": 0.04}
{"loss": 3.4186, "grad_norm": 7.084853172302246, "learning_rate": 3.233333333333333e-05, "epoch": 0.04}
{"loss": 3.4066, "grad_norm": 6.5051589012146, "learning_rate": 3.216666666666665e-05, "epoch": 0.04}
{"loss": 3.4105, "grad_norm": 6.650455951690674, "learning_rate": 3.2000000000000005e-05, "epoch": 0.04}
{"loss": 3.4115, "grad_norm": 7.152836322784424, "learning_rate": 3.183333333333334e-05, "epoch": 0.04}
{"loss": 3.5135, "grad_norm": 8.30115032196045, "learning_rate": 3.166666666666666e-05, "epoch": 0.04}
37% | 1109/3000 [14:38<17:38, 1.79it/s]
```

CSDN @weixin_40078854

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

```
{'loss': 3.5164, 'grad_norm': 6.449600696563721, 'learning_rate': 2.9666666666666672e-05, 'epoch': 0.04}
{'loss': 3.4018, 'grad_norm': 7.199850559234619, 'learning_rate': 2.95e-05, 'epoch': 0.04}
{'loss': 3.3107, 'grad_norm': 7.313725471496582, 'learning_rate': 2.933333333333336e-05, 'epoch': 0.04}
{'loss': 3.5068, 'grad_norm': 6.89715576171875, 'learning_rate': 2.916666666666667e-05, 'epoch': 0.04}
{'loss': 3.4146, 'grad_norm': 7.483893871307373, 'learning_rate': 2.9e-05, 'epoch': 0.04}
{'loss': 3.4, 'grad_norm': 7.177615165710449, 'learning_rate': 2.883333333333334e-05, 'epoch': 0.04}
{'loss': 3.5, 'grad_norm': 7.866086959838867, 'learning_rate': 2.8666666666666668e-05, 'epoch': 0.04}
{'loss': 3.5146, 'grad_norm': 8.68930721282959, 'learning_rate': 2.849999999999998e-05, 'epoch': 0.05}
{'loss': 3.4529, 'grad_norm': 7.792913913726807, 'learning_rate': 2.833333333333335e-05, 'epoch': 0.05}
{'loss': 3.416, 'grad_norm': 7.438651084899902, 'learning_rate': 2.8166666666666667e-05, 'epoch': 0.05}
{'loss': 3.3881, 'grad_norm': 7.13234281539917, 'learning_rate': 2.8000000000000003e-05, 'epoch': 0.05}
{'loss': 3.5561, 'grad_norm': 7.559863090515137, 'learning_rate': 2.783333333333333e-05, 'epoch': 0.05}
{'loss': 3.3529, 'grad_norm': 6.504519939422607, 'learning_rate': 2.7666666666666667e-05, 'epoch': 0.05}
{'loss': 3.5186, 'grad_norm': 7.67847204208374, 'learning_rate': 2.7500000000000004e-05, 'epoch': 0.05}
{'loss': 3.2891, 'grad_norm': 8.304348945617676, 'learning_rate': 2.733333333333333e-05, 'epoch': 0.05}
{'loss': 3.4344, 'grad_norm': 8.050036430358887, 'learning_rate': 2.7166666666666667e-05, 'epoch': 0.05}
{'loss': 3.5484, 'grad_norm': 7.238221168518066, 'learning_rate': 2.7000000000000002e-05, 'epoch': 0.05}
{'loss': 3.2625, 'grad_norm': 7.541359901428223, 'learning_rate': 2.683333333333333e-05, 'epoch': 0.05}
{'loss': 3.3986, 'grad_norm': 7.625448703765869, 'learning_rate': 2.6666666666666667e-05, 'epoch': 0.05}
{'loss': 3.3598, 'grad_norm': 8.393092155456543, 'learning_rate': 2.6500000000000004e-05, 'epoch': 0.05}
{'loss': 3.4721, 'grad_norm': 6.66404914855957, 'learning_rate': 2.633333333333333e-05, 'epoch': 0.05}
{'loss': 3.3332, 'grad_norm': 7.385687828063965, 'learning_rate': 2.6166666666666668e-05, 'epoch': 0.05}
{'loss': 3.2078, 'grad_norm': 6.915552139282227, 'learning_rate': 2.6000000000000002e-05, 'epoch': 0.05}
{'loss': 3.5498, 'grad_norm': 8.550110816955566, 'learning_rate': 2.583333333333336e-05, 'epoch': 0.05}
{'loss': 3.4027, 'grad_norm': 7.597967147827148, 'learning_rate': 2.566666666666666e-05, 'epoch': 0.05}
{'loss': 3.2166, 'grad_norm': 6.888904571533203, 'learning_rate': 2.5500000000000003e-05, 'epoch': 0.05}
{'loss': 3.4207, 'grad_norm': 7.880514621734619, 'learning_rate': 2.533333333333337e-05, 'epoch': 0.05}
49% |
```

CSDN@weixin_40078854
| 1482/3

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

```
管理员: Anaconda Prompt - p_ X + v
warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
warnings.warn(
{'eval_rouge-1': 32.082118, 'eval_rouge-2': 7.108884000000001, 'eval_rouge-l': 24.508034000000002, 'eval_bleu-4': 0.03280893935904207, 'ev
al_runtime': 97.1848, 'eval_samples_per_second': 0.514, 'eval_steps_per_second': 0.041, 'epoch': 0.09}
{'loss': 3.5129, 'grad_norm': 8.475122451782227, 'learning_rate': 8.16666666666668e-06, 'epoch': 0.09}
{'loss': 3.208, 'grad_norm': 8.285528182983398, 'learning_rate': 8.000000000000001e-06, 'epoch': 0.09}
{'loss': 3.4789, 'grad_norm': 7.606888771057129, 'learning_rate': 7.83333333333333e-06, 'epoch': 0.09}
{'loss': 3.3074, 'grad_norm': 9.601532936096191, 'learning_rate': 7.66666666666667e-06, 'epoch': 0.09}
{'loss': 3.1473, 'grad_norm': 7.468560218811035, 'learning_rate': 7.5e-06, 'epoch': 0.09}
{'loss': 3.251, 'grad_norm': 9.788201332092285, 'learning_rate': 7.33333333333334e-06, 'epoch': 0.09}
{'loss': 3.4271, 'grad_norm': 8.8259916305542, 'learning_rate': 7.16666666666667e-06, 'epoch': 0.09}
{'loss': 3.343, 'grad_norm': 8.214253425598145, 'learning_rate': 7.000000000000001e-06, 'epoch': 0.09}
{'loss': 3.3937, 'grad_norm': 8.670844078063965, 'learning_rate': 6.83333333333333e-06, 'epoch': 0.09}
{'loss': 3.3666, 'grad_norm': 7.838050365447998, 'learning_rate': 6.66666666666667e-06, 'epoch': 0.09}
{'loss': 3.4539, 'grad_norm': 8.194501876831055, 'learning_rate': 6.500000000000004e-06, 'epoch': 0.09}
{'loss': 3.3596, 'grad_norm': 8.18748664855957, 'learning_rate': 6.33333333333334e-06, 'epoch': 0.09}
{'loss': 3.3451, 'grad_norm': 8.989133834838867, 'learning_rate': 6.16666666666667e-06, 'epoch': 0.09}
{'loss': 3.2414, 'grad_norm': 8.749055862426758, 'learning_rate': 6e-06, 'epoch': 0.09}
{'loss': 3.3816, 'grad_norm': 7.964601993560791, 'learning_rate': 5.83333333333334e-06, 'epoch': 0.09}
{'loss': 3.2213, 'grad_norm': 7.849711894989014, 'learning_rate': 5.66666666666667e-06, 'epoch': 0.09}
{'loss': 3.3883, 'grad_norm': 8.309954643249512, 'learning_rate': 5.500000000000001e-06, 'epoch': 0.09}
{'loss': 3.3514, 'grad_norm': 8.940631866455078, 'learning_rate': 5.33333333333334e-06, 'epoch': 0.09}
{'loss': 3.2797, 'grad_norm': 9.925182342529297, 'learning_rate': 5.16666666666667e-06, 'epoch': 0.09}
{'loss': 3.4555, 'grad_norm': 8.14529037475586, 'learning_rate': 5e-06, 'epoch': 0.09}
{'loss': 3.2467, 'grad_norm': 6.839783191680908, 'learning_rate': 4.83333333333333e-06, 'epoch': 0.09}
{'loss': 3.3314, 'grad_norm': 9.924310684204102, 'learning_rate': 4.66666666666667e-06, 'epoch': 0.09}
{'loss': 3.343, 'grad_norm': 9.065497398376465, 'learning_rate': 4.5e-06, 'epoch': 0.1}
{'loss': 3.4131, 'grad_norm': 7.702908992767334, 'learning_rate': 4.33333333333334e-06, 'epoch': 0.1}
{'loss': 3.3291, 'grad_norm': 8.846440315246582, 'learning_rate': 4.16666666666667e-06, 'epoch': 0.1}
{'loss': 3.4951, 'grad_norm': 9.373692512512207, 'learning_rate': 4.000000000000001e-06, 'epoch': 0.1}
{'loss': 3.3871, 'grad_norm': 8.122434616088867, 'learning_rate': 3.83333333333334e-06, 'epoch': 0.1}
{'loss': 3.2883, 'grad_norm': 7.910025119781494, 'learning_rate': 3.66666666666667e-06, 'epoch': 0.1}
{'loss': 3.3949, 'grad_norm': 8.316025733947754, 'learning_rate': 3.500000000000004e-06, 'epoch': 0.1}
{'loss': 3.3994, 'grad_norm': 8.11379623413086, 'learning_rate': 3.33333333333333e-06, 'epoch': 0.1}
{'loss': 3.3684, 'grad_norm': 10.268864631652832, 'learning_rate': 3.16666666666667e-06, 'epoch': 0.1}
{'loss': 3.257, 'grad_norm': 8.653985023498535, 'learning_rate': 3e-06, 'epoch': 0.1}
{'loss': 3.2625, 'grad_norm': 8.049074172973633, 'learning_rate': 2.83333333333335e-06, 'epoch': 0.1}
{'loss': 3.4355, 'grad_norm': 8.686062812805176, 'learning_rate': 2.66666666666667e-06, 'epoch': 0.1}
{'loss': 3.4514, 'grad_norm': 7.830814838409424, 'learning_rate': 2.5e-06, 'epoch': 0.1}
{'loss': 3.2201, 'grad_norm': 8.983247756958008, 'learning_rate': 2.33333333333336e-06, 'epoch': 0.1}
{'loss': 3.359, 'grad_norm': 8.781380653381348, 'learning_rate': 2.16666666666667e-06, 'epoch': 0.1}
{'loss': 3.4941, 'grad_norm': 8.599340438842773, 'learning_rate': 2.000000000000003e-06, 'epoch': 0.1}
{'loss': 3.5113, 'grad_norm': 9.224966049194336, 'learning_rate': 1.83333333333335e-06, 'epoch': 0.1}
{'loss': 3.3488, 'grad_norm': 9.043194770812988, 'learning_rate': 1.66666666666667e-06, 'epoch': 0.1}
97% | 2904/3000 [36:28<00:58, 1.63it/s] | 2904/3000 [36:28<00:58, 1.63it/s]
https://blog.csdn.net/weixin_40078854/article/details/137226677?csdn_share_tail=%7B"type"%3A"blog"%2C"rType"%3A"article"%2C"rId"%3A"137226677"%2C"source"%3A"weixin_40078854"%7D
7/14
```

3.2 微调结果

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

管理员: Anaconda Prompt

```
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but can
not be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
```

3.4 微调结果与使用未微调的模型结果对比

```
of a model on the Hub.

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:\Users\newchatglm36n\ChatGLM3\finetune_demo\output\checkpoint-2000 --prompt 类型#裤*版型#宽松*风格#性感*图案#线条*裤型#阔腿裤
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
Loading checkpoint shards: 100%|██████████| 7/7 [00:04<00:00,  1.59it/s]
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
C:\Users\16014\.cache\huggingface\modules\transformers_modules\THUDM\chatglm3-6b\ba5ba5501eb873d40d48bd0983bd2a8dd006bb838\modeling_chatglm.py:226: UserWarning: 1Torch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,
宽松的阔腿裤，可以很好的修饰腿型，同时还能遮盖下半身的赘肉，将你的身材完美的展现出来，让你的身材更加性感迷人。裤身的线条设计，可以很好的修饰你的身材曲线，让你更加迷人。
.

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:\Users\chatglm6b\model\chatglm3-6b --prompt 类型#裤*版型#宽松*风格#性感*图案#线条*裤型#阔腿裤
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
Loading checkpoint shards: 100%|██████████| 7/7 [00:04<00:00,  1.69it/s]
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
C:\Users\16014\.cache\huggingface\modules\transformers_modules\chatglm3-6b\modeling_chatglm.py:226: UserWarning: 1Torch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,
这款宽松性感阔腿裤的设计线条流畅，图案独特，展现出女性魅力。它采用了高质量的面料，穿着舒适，同时具有较好的透气性。阔腿裤的版型设计使腿部线条更加修长，展现出修长的身材比例。总之，这款裤子无论是搭配什么衣服，都能让您在人群中脱颖而出，成为众人瞩目的焦点。

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:\Users\newchatglm36n\ChatGLM3\finetune_demo\output\checkpoint-2000 --prompt 类型#裤*版型#宽松*风格#性感*图案#线条*裤型#阔腿裤
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
    warnings.warn(
Loading checkpoint shards: 100%|██████████| 7/7 [00:04<00:00,  1.66it/s]
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
C:\Users\16014\.cache\huggingface\modules\transformers_modules\chatglm3-6b\modeling_chatglm.py:226: UserWarning: 1Torch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,
这款阔腿裤采用宽松的版型设计，穿着舒适不束缚，可以轻松驾驭。宽松的裤腿设计，修饰腿型，穿出大长腿的视觉效果。侧边开叉设计，凸显出性感的设计感，增添几分魅力。

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>
```

内容来源: csdn.net

作者昵称: weixin_40078854

CSDN @weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677作者主页: https://blog.csdn.net/weixin_40078854

```
'https://huggingface.co/C:/Users/newchatglm36n/ChatGLM3/finetune_demo/output/checkpoint-2800/None' for available files.

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:/Users/newchatglm36n/ChatGLM3/finetune_demo/output/checkpoint-2800/ --prompt 类型#裤#版型#宽松#风格#性感*图案#线条*裤型#阔腿裤
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
  warnings.warn(
Loading checkpoint shards: 100%|██████████| 7/7 [00:04<00:00,  1.55it/s]
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
C:\Users\16014\.cache\huggingface\modules\transformers_modules\THUDM\chatglm3-6b\aa5ba5501eb873d40d48bd0983bd2a8dd006bb838\modeling_chatglm.py:226: UserWarning: lTorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)
  context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,
这条阔腿裤的宽松设计，可以很好的修饰你的身形，轻松遮盖小腹赘肉，打造修身美腿的视觉效果。同时，裤身的线条设计，将女性的曲线美展现得淋漓尽致，尽显女性的性感魅力。

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py output/checkpoint-2800 --prompt 类型#裤#版型#宽松#风格#性感*图案#线条*裤型#阔腿裤
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
  warnings.warn(
Loading checkpoint shards: 100%|██████████| 7/7 [00:04<00:00,  1.56it/s]
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
C:\Users\16014\.cache\huggingface\modules\transformers_modules\THUDM\chatglm3-6b\aa5ba5501eb873d40d48bd0983bd2a8dd006bb838\modeling_chatglm.py:226: UserWarning: lTorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)
  context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,
阔腿裤是近年来非常流行的裤型，它的设计流畅，线条感强，搭配上高腰的设计，更具有时尚感，能够更好的展现女性身材曲线，同时宽松的版型，能够很好的遮盖腿部赘肉，打造纤细修长的腿部曲线，展现女性性感魅力。

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:/Users\chatglm6b\model\chatglm3-6b --prompt 类型#裤#版型#宽松#风格#性感*图案#线条*裤型#阔腿裤
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'
  warnings.warn(
Loading checkpoint shards: 100%|██████████| 7/7 [00:04<00:00,  1.64it/s]
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
C:\Users\16014\.cache\huggingface\modules\transformers_modules\chatglm3-6b\modeling_chatglm.py:226: UserWarning: lTorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)
  context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,
您好！看起来您提供的信息有些混乱，不过我猜您可能在询问关于宽松性感阔腿裤的问题。下面我将根据您的要求逐一进行解答。

1. 类型: #裤
  这里的#裤表示您可能在寻找一种类型的裤子，如直筒裤、修身裤、休闲裤等。

2. 版型: 宽松
  宽松版型的裤子通常具有较大的尺寸，穿着舒适，适合休闲或正式场合。

3. 风格: 性感
  性感风格通常体现在设计、面料或颜色上，使衣物更具吸引力，让人眼前一亮。您在提问中并未明确提及裤子的风格，如果您希望穿着更性感的裤子，可以考虑选择一条设计独特、颜色诱人的裤子。

4. 图案: 线条
  线条图案是指裤子上有一些明显的线条、花纹或几何图案。这种图案可以使裤子更具视觉冲击力，增加时尚感。您可以在挑选裤子时关注是否有线条图案，根据个人喜好来选择。

5. 裤型: 阔腿裤
  阔腿裤是指腿部宽松、向下延伸的裤子。它具有强大的包容性，适合不同身材的人穿着，阔腿裤可以搭配各种上衣，具有较强的时尚度。

根据以上分析，如果您想寻找一条宽松性感的阔腿裤，可以考虑在挑选时关注线条图案、颜色和材质等方面，以便找到符合您心意的裤子。希望我的回答对您有所帮助！

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>
```

CSDN @weixin_40078854

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

```
C:\Users\16014\cache\huggingface\modules\transformers_modules\THUDM\chatglm3-6b\5ba5501eb873d40d48bd0983bd2a8dd006bb838\modeling_chatglm.py:226: UserWarning: ITorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)  
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,  
这条阔腿裤的宽松设计，可以很好的修饰你的身形，轻松遮盖小腹赘肉，打造修身美腿的视觉效果。同时，裤子的线条设计，将女性的曲线美展现得淋漓尽致，尽显女性的性感魅力。  
(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py output\checkpoint-2000 --prompt 类型#裤*版型#宽松*风格#性感*图案#线条*裤型#阔腿裤  
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'  
    warnings.warn(  
Loading checkpoint shards: 100% | 7/7 [00:04<00:00, 1.56it/s]  
Setting eos_token is not supported, use the default one.  
Setting pad_token is not supported, use the default one.  
Setting unk_token is not supported, use the default one.  
C:\Users\16014\cache\huggingface\modules\transformers_modules\THUDM\chatglm3-6b\5ba5501eb873d40d48bd0983bd2a8dd006bb838\modeling_chatglm.py:226: UserWarning: ITorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)  
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,  
阔腿裤是近年来非常流行的裤型，它的设计流畅，线条感强，搭配上高腰的设计，更具有时尚感，能够更好的展现女性身材曲线，同时宽松的版型，能够很好的遮盖腿部赘肉，打造纤细修长的腿部曲线，展现女性魅力。  
(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:\Users\chatglm6b\model\chatglm3-6b --prompt 类型#裤*版型#宽松*风格#性感*图案#线条*裤型#阔腿裤  
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'  
    warnings.warn(  
Loading checkpoint shards: 100% | 7/7 [00:04<00:00, 1.64it/s]  
Setting eos_token is not supported, use the default one.  
Setting pad_token is not supported, use the default one.  
Setting unk_token is not supported, use the default one.  
C:\Users\16014\cache\huggingface\modules\transformers_modules\chatglm3-6b\modeling_chatglm.py:226: UserWarning: ITorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)  
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,  
您好！看起来您提供的信息有些混乱，不过我猜您可能在询问关于宽松性感阔腿裤的问题。下面我将根据您的要求逐一进行解答。
```

1. 类型: #裤
这里的#裤表示您可能在寻找一种类型的裤子，如直筒裤、修身裤、休闲裤等。

2. 版型: 宽松
宽松版型的裤子通常具有较大的尺寸，穿着舒适，适合休闲或正式场合。

3. 风格: 性感
性感风格通常体现在设计、面料或颜色上，使衣物更具吸引力，让人眼前一亮。您在提问中并未明确提及裤子的风格，如果您希望穿着更性感的裤子，可以考虑选择一条设计独特、颜色诱人的裤子。

4. 图案: 线条
线条图案是指裤子上有一些明显的线条、花纹或几何图案。这种图案可以使裤子更具视觉冲击力，增加时尚感。您可以在挑选裤子时关注是否有线条图案，根据个人喜好来选择。

5. 裤型: 阔腿裤
阔腿裤是指腿部宽松、向下延伸的裤子。它具有强大的包容性，适合不同身材的人穿着。阔腿裤可以搭配各种上衣，具有较强的时尚度。

根据以上分析，如果您想寻找一条宽松性感的阔腿裤，可以考虑在挑选时关注线条图案、颜色和材质等方面，以便找到符合您心意的裤子。希望我的回答对您有所帮助！

```
(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:\Users\newchatglm36n\ChatGLM3\finetune_demo\output\checkpoint-2000 --prompt 类型#裤*版型#宽松*风格#性感*图案#线条*裤型#阔腿裤  
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'  
    warnings.warn(  
Loading checkpoint shards: 100% | 7/7 [00:04<00:00, 1.56it/s]  
Setting eos_token is not supported, use the default one.  
Setting pad_token is not supported, use the default one.  
Setting unk_token is not supported, use the default one.  
C:\Users\16014\cache\huggingface\modules\transformers_modules\THUDM\chatglm3-6b\5ba5501eb873d40d48bd0983bd2a8dd006bb838\modeling_chatglm.py:226: UserWarning: ITorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)  
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,  
裤身设计采用宽松的版型，轻松遮肉又显瘦，舒适性好。腰带设计，提升腰线，更显高挑。裤身侧边采用走线装饰，提升层次感。宽松的阔腿裤型，搭配舒适的面料，穿着舒适又舒适。裤身上方采用小背心搭配，性感又迷人。  
(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>python inference_hf.py C:\Users\chatglm6b\model\chatglm3-6b --prompt 类型#裤*版型#宽松*风格#性感*图案#线条*裤型#阔腿裤  
C:\ProgramData\anaconda3\envs\chatglm3\lib\site-packages\huggingface_hub\utils\_runtime.py:184: UserWarning: Pydantic is installed but cannot be imported. Please check your installation. 'huggingface_hub' will default to not using Pydantic. Error message: '{e}'  
    warnings.warn(  
Loading checkpoint shards: 100% | 7/7 [00:04<00:00, 1.66it/s]  
Setting eos_token is not supported, use the default one.  
Setting pad_token is not supported, use the default one.  
Setting unk_token is not supported, use the default one.  
C:\Users\16014\cache\huggingface\modules\transformers_modules\chatglm3-6b\modeling_chatglm.py:226: UserWarning: ITorch was not compiled with flash attention. (Triggered internally at ..\aten\src\ATen\native\transformers\cuda\sdp_utils.cpp:263.)  
    context_layer = torch.nn.functional.scaled_dot_product_attention(query_layer, key_layer, value_layer,  
这条宽松性感阔腿裤拥有优美的线条设计，展现出迷人的风格。它的图案简洁而时尚，让人一眼就能感受到它的精致。裤型宽松舒适，非常适合在休闲场合穿着。阔腿裤的设计让您的双腿显得更加修长，让您在人群中脱颖而出。此外，这条裤子采用了高质量的面料，保证了它的耐用性和舒适度。无论是与朋友一起逛街，还是独自享受一个安静的夜晚，这条裤子都能给您带来完美的体验。
```

(chatglm3) C:\Users\newchatglm36n\ChatGLM3\finetune_demo>

CSDN @weixin_40078854

4. 部署到gradio_demo, 使用webui来调试和调用

4.1 使用未微调的模型调试结果

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

ChatGLM3-6B 使用未微调模型版本

The screenshot shows a conversational interface for the ChatGLM3-6B model. The history consists of two pairs of messages:

- User:** 类型#裤颜色#黑色风格#简约*图案#条纹
- Model:** #裤##黑色##简约##条纹

Below the history, there is an input field labeled "Input.." and a "Submit" button. To the right, there is a panel titled "Clear History" containing three sliders for configuration:

- Maximum length: 8192
- Top P: 0.8
- Temperature: 0.6

4.2 使用未微调的模型调试结果

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854

ChatGLM3-6B 使用微调后模型版本



This screenshot shows the CSDN AI writing tool interface. It includes an input field labeled "Input...", a "Submit" button, and a sidebar with model configuration options: "Clear History", "Maximum length" (set to 8192), "Top P" (set to 0.8, with a note "截图(Alt + A)" next to it), and "Temperature" (set to 0.6). The bottom right corner displays the text "CSDN @weixin_40078854".

文章知识点与官方知识档案匹配，可进一步学习相关知识

Python入门技能树 > 首页 > 概览 407036 人正在系统学习中

内容来源: csdn.net

作者昵称: weixin_40078854

原文链接: https://blog.csdn.net/weixin_40078854/article/details/137226677

作者主页: https://blog.csdn.net/weixin_40078854