

lm-sys / FastChat Public

<> Code Issues 877 Pull requests 151 Actions Security and quality Ins

⚠ This commit does not belong to any branch on this repository, and may belong to a fork outside of the repository.

Commit c9e84b8

kaisfree and claude committed 2 weeks ago

fix: wrap remaining blocking calls with asyncio.to_thread to prevent DoS

Co-Authored-By: Claude Opus 4.6 <noreply@anthropic.com>

1 parent [587d5cf](#) commit c9e84b8

3 files changed +3 -3 lines changed

↑ Top ⚙

Filter files...

- fastchat/serve
 - base_model_worker.py
 - huggingface_api_worker.py
 - multi_model_worker.py

3 files changed +3 -3 lines changed

Search within code ⚙

fastchat/serve/base_model_worker.py

```

@@ -215,7 +215,7 @@ async def api_generate(request: Request):
215 215     async def api_get_embeddings(request: Request):
216 216         params = await request.json()
217 217         await acquire_worker_semaphore()
218 -     embedding = worker.get_embeddings(params)
218 +     embedding = await asyncio.to_thread(worker.get_embeddings, params)
219 219     release_worker_semaphore()

```

```
220 220     return JsonResponse(content=embedding)
221 221
```



fastchat/serve/huggingface_api_worker.py



```
@@ -233,7 +233,7 @@ async def api_generate(request: Request):
233 233     params = await request.json()
234 234     worker = worker_map[params["model"]]
235 235     await acquire_worker_semaphore(worker)
236 -     output = worker.generate_gate(params)
+ 236 +     output = await asyncio.to_thread(worker.generate_gate, params)
237 237     release_worker_semaphore(worker)
238 238     return JsonResponse(output)
239 239
```



fastchat/serve/multi_model_worker.py



```
@@ -109,7 +109,7 @@ async def api_generate(request: Request):
109 109     params = await request.json()
110 110     await acquire_worker_semaphore()
111 111     worker = worker_map[params["model"]]
112 -     output = worker.generate_gate(params)
+ 112 +     output = await asyncio.to_thread(worker.generate_gate, params)
113 113     release_worker_semaphore()
114 114     return JsonResponse(output)
115 115
```



Comments 0



Please [sign in](#) to comment.