



Repository Explorerã•«æ?³å®?è³å?•ã??æ??ã?ã?ã?ã?ã?ã?

Description

ã¼ã¼• èªæ?ã®ä½?å°ã®å?ã?Repository Explorerã§ã?ã??

ã•?ã??ã•-ã,?é??ãSè`?ã?•ã??ã•|ã•?ã??Alã?`ã?¼ã?,ã?§ã?³ã??ã•«ç?,ã½?ã•?ã??ã??ã•®ã•§ã?•ZIKUUã•
®ã?½ã??ã??ã??ã?|ã?§ã?¢ã?ªã?•ã?ã??ã?ªã??è?•ã?•ã•|èª?œ?»ã??ã?•ã•|ã•ã??ã??ã??ã?¼ã?«ã??

ä?·ä?¹ä??ä? ä·?å®?æ?·ä·?ä·?å¼ä?·ç§·ä»¥å¤?ä·®äººä·?ä?·ä?¹ä??ä? ä??é·?ç?·ä?·æ?jå¼µä?·ä??ä?·
 ä??ä?·ä?·ä·®æ£ä·ªä·©ä?·èj?ä·?é??ä·«äººé??ä??è£?å·©ä·?ä??ä·®ä·?ç?·®ç??ä·Sä½ä??ä??ä?·ä·ä·ä·
 ¾ä·?ä??

é??ç?°è? ã•šã•?ã??çš•è?°è°«ã•«ã•ã•£ã•|ã??ã?•ZIKUUã•®ã?•ã?¹ã??ã? ã•ã•ã•?ã•šã•
«è•?æ?¶ã??é ¼ã??ã•«æ??æ•|ã•šã•ã•è??è|•æ•|ã??è¶ã?•ã•|ã?•ã•¾ã?•ã?•ã??ã?•ä»?ã•®ã°ã•?ZIKUUã•
®ã?•ã?¹ã??ã? ã??æ?±ã•?ã•ã•ã•«ã•ã•ã•ã•?ã•?ã•?ã•?ã??ã?¹¼ã?«ã•?ã•ã•?ã•£ã•?ã??ã•|ã??ã•ã??ã•
?ã??ã•ã??ã•šã•?ã??

[illegible]

ã?½ã??ã??ã?|ã?§ã?¢é??ç?°ã®ç¾4å´ã•§ã•-ã?ã•ã?ã?ã?µ?é"?ã??ã?ã?|ã?ã?ã?ã°ã•-ã?ã?ã?ã?ã?

ã•ã??ã•ã•ã•ã•«æ?³å®?ã•?ã•??ã•??è³åå?•ã•??æ??ã•?ã•|èªæ?ã•??ç¶?ã•?ã•|ã•?ã•ã•¼ã•?ã•??

ä¼?ä°ä?•LLM Brokerä?ä?ä?½ä??ä?ä?!ä?Sä?¢ä-ä?LLMå?¼ä?ä?°ä?ä??è;?ä?æ??ä?
 @ä°é??æ?ç?ä??ä?ä?ä-ä??ä??æ?½è±jâ?ä?ä?ä?ä?LLMä??ä?ä?ä?ä?ä?¼äç¶?æ©?ä?Sä?
 ?ä??ç¾æ??ç?¹ä?Sä-llama.cppä?µä?¼ä?ä?¼ä?Ollamaä?«ä¾äç?ä?ä?ä?ä?¾ä?ä??

ã·?ã·®ã??ã·?ã·ªè³å?·ã??æ??ã·?ã·|ã·¿ã·¾ã·?ã·?ã·??

LLM Brokerã•®chatã?jã?½ã??ã??ã??æ?jã¼µã•?ã•jã??ã?ã?ã•?ãã??ã?½ã??ã?½ã??è½ã? ã•
?ã??æ?¹æ³?ã??çªªã•?ã•jã??

llama.cppã??Ollamaä»¥å¤?ã•®LLMă??ã?ã?ã?¤ã??ã?¼ã??ã½¿ã•?ã•?ã?ã•©ã•?ã•?ã??æ??ã??ã•
¤ã?ã??ã•è?_ã?ã•?ã?ã??è"?ã?ã•|ã?ã•¾ã?ã??

ã•ã•®å??ç?ã•?ã•?ã??ã??

é•©å?ãªå??ç?ã?è¿?ã•£ã|ã•ã|ã•ã??ã??ã?ã«æ?ã•ã¾ã?ã??

Answer

Summary

LLM Broker の chat メソッドを拡張してプロバイダーを追加するには、バックエンドクラスを実装し、profiles.yaml にプロファイルを登録する。

Details

Repository Explorer:backend/app/llm_broker_client.py lines 4-44 に示される LLMBrokerClient の generate_text は、設定された profile を /v1/chat に送る。LLM Broker の実装は backend/app/main.py で FastAPI ルートを定義し、app.backends.base.LLMBackend を継承したクラスをインポートする。実際のチャット処理は backend/app/backends/llamacpp.py と backend/app/backends/ollama.py の chat メソッドで行われている。したがって、新しいプロファイルを追加するには、次の手順を行う。

1. backend/app/backends/ 直下に新しいクラスを作成。例：

```
class MyProviderBackend(LLMBackend):  
    async def chat(self, *, profile: ProfileConfig, messages: list[ChatMessage], temperature: float | None):  
        ...` で Ollama や LlamaCpp と同様に HTTP POST を実装。
```
2. backend/profiles.yaml に新しい profile を追加。例：

```
```yaml  
my_provider:
 backend: myprovider
 base_url: http://my.provider:11434
 model: my-model
 temperature: 0.5
 timeout_sec: 120
...`
```
3. backend/app/main.py の import 文に `from app.backends.myprovider import MyProviderBackend` を追加し、FastAPI ルートで `MyProviderBackend` を使用する。
4. LLMBrokerClient から呼び出す際は、環境変数 `LLM_PROFILE=my_provider` を設定するか、`LLM_BROKER_PROFILE` を変更して `profile` 名を指定する。  
この手順で、LLM Broker の `chat` メソッドは新しいプロバイダーを選択して処理できるようになる。

### Unknowns

- LLM Broker の FastAPI ルートで実際にどのようにバックエンドクラスを選択しているかは、コード内のルーティング定義が不足しているが

### Next Actions

1. 確認する: Search for FastAPI route definitions in backend/app/main.py that map profile.backend to backend class instances。
2. 確認する: Inspect backend/app/backends/ directory for any factory or registry that selects backend implementation based on profile.back

å??ç?é?`å??ã??èª¸;ã??ã?ã?•œ? å½¢\$ã?ã?ã??ã?®ã?ã?ã?ã?ã??

```
Repository Explorer:backend/app/llm_broker_client.py lines 4-44
?LLMBrokerClient
? generate_text
? profile
? HTTP POST
? /v1/chat
```

```
LLM Broker @è£?~ backend/app/main.py S FastAPI ?«?%??â??â@?ç%â•
?â?•app.backends.base.LLMBackend ??ç¶?æ?¿â?•â?•â?~â?@â?¹â??â?ââ?³â?•â?%â??â
•â?•!â?•â??â??â@?é??â?@â?•â?£â??â??â?!ç?•â•
```

```
backend/app/backends/llamacpp.py ~ backend/app/backends/ollama.py ~
® chat ~|~%~?~??~·$è|~?·~??~·|~·?~??~?
~·?~·?~·?~·£~·|~?·æ°~·?~·?~??~?~·?~?~?~?~?~?~?~?~?~?~?
®æ??é ?~??è|~?~?~?
```

1. backend/app/backends/ ç?´ä, ?ã•«æ?°ã•?ã•?ã?¬ã?©ã?¹ã?²ã½?æ?•ã??

```
class MyProviderBackend(LLMBackend):
 async def chat(self, *, profile: ProfileConfig, messages: list[ChatMessage]
 ...`
 Ollama
 LlamaCpp
 HTTP POST
 @?£?`
```

2. backend/profiles.yaml    profile   

```
ä%?i¼?
 ``yaml
my_provider:
 backend: myprovider
 base_url: http://my.provider:11434
 model: my-model
 temperature: 0.5
 timeout_sec: 120
 ``
```

```
3. backend/app/main.py
from fastapi import FastAPI
from app.backends.myprovider import MyProviderBackend

app = FastAPI()
app.include_router(MyProviderBackend.router)
```

```
4. LLMBrokerClient ă•?ă??ă?¼ă•³ă?°ă•?é??ă•-ă?•
ç?°ăç?ăș?æ?° `LLM_PROFILE=my_provider` ă??è`ă@?ă•?ă??ă•?ă?•
`LLM_BROKER_PROFILE` ă??ăș?æ?`ă•?ă•| profile ă••ă??æ??ă@?ă•?ă??ă??
```

ã·?ã·®æ??é ?ã·\$ã·?LLM Broker ã·® chat ã·|ã·½ã·?ã·?ã·-æ°ã·?ã·?ã·?ã·?ã·?  
ã·?ã·?ã·?ã·?ã·?é· æ??ã·?ã·|ã·|ç·?ã·\$ã··ã·?ã·?ã·?ã·«ã·ªã·?ã·?

ç?«é•¢ã«ã¬ã?Unknownsã¬ã¬ã?ã?é ç?®ã¬ã¬ã¬ã?Unknownsã¬ã¬ã¬æ¹æ? ä,è¶³ã¬  
 §æ?è¬¬ã¬¬æ¬¬è¬¬ã¬¬§ã¬ã¬ã¬ç?¹ã¬«ã¬ã¬ã¬æ?ã¬¬ã¬¬ã¬¬4ã¬¬ã¬¬Next Actionsã¬ã¬ã¬  
 è½½ã? èªæ?»ã¬®æ¬æj?ã¬æ?ã¬ã¬ã¬ã¬4ã¬ã¬?

æ??çµ?ç??ã.«ã¯ã?ã?ã?®ã??ç?ã??èµ·ç?¹ã.«ã?·æ·±å ?ã??ã?·ã?·ã??ã?·æ¹æ? å¯³æ±|ã??ãº?ã?ã·  
 ?ã??ã·ã?ã·£ã?æ?·ä½?ã??ç¹ºã??è¿?ã?·ã·|ã?·ã·ã·®é?ç?²ã??ã??ã?·ã·ã·«ã·ã?ã?·æ?·ã·  
 ?ã·¾ã?ã??ã?ã·ã?ã?ã?ã?ã?·èªæ?»ã·®å±¥æã·?æ®?ã·£ã·|ã?·å ?ãºã?ã·ã·©ã·®ã??ã?·ã·«èªã·¹ã?ã·  
 ?ã?·ã?·ã?ã??ã?·ä»ã·®äººã?·èªæ?»ã?ã??é??ã·®å?è??ã·«ã·ªã??ã?ã??ã?·§ã?ã??

## Category

1. æ?¥ã? ã•®æ'»å??

## Tags

1. AI
2. LLM
3. Repository Explorer
4. ä.é??â??â.?

5. å±±æ¢“ç??

**Date Created**  
2026å¹7æ??13æ?¥  
**Author**  
kazuo-tsubaki

default watermark