

Obsidianã•¨LLM Brokerã??ã•¤ã•ªã••

Description

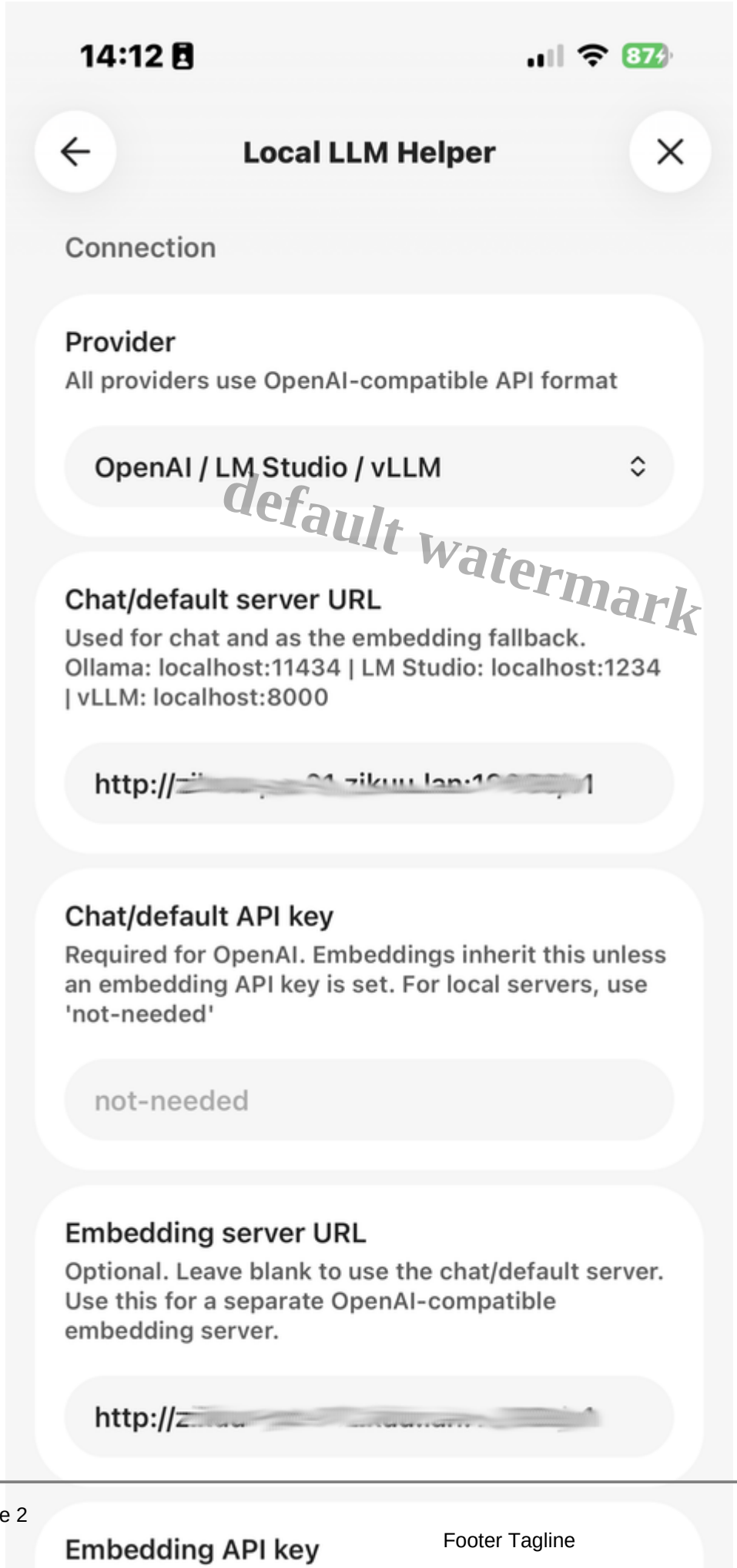
default watermark

**LLM Brokerã•«Obisidianã•®Local LLM Helperã•¨ã•?ã•?ã??ã?©ã?°ã?¤ã?³ã??ä½¿ã•£ã•¦æ?¥ç¶?ã•?ã•¦ã•¿ã•¾ã•?ã•?ã??

è¨å®?é ç?®ã•¯

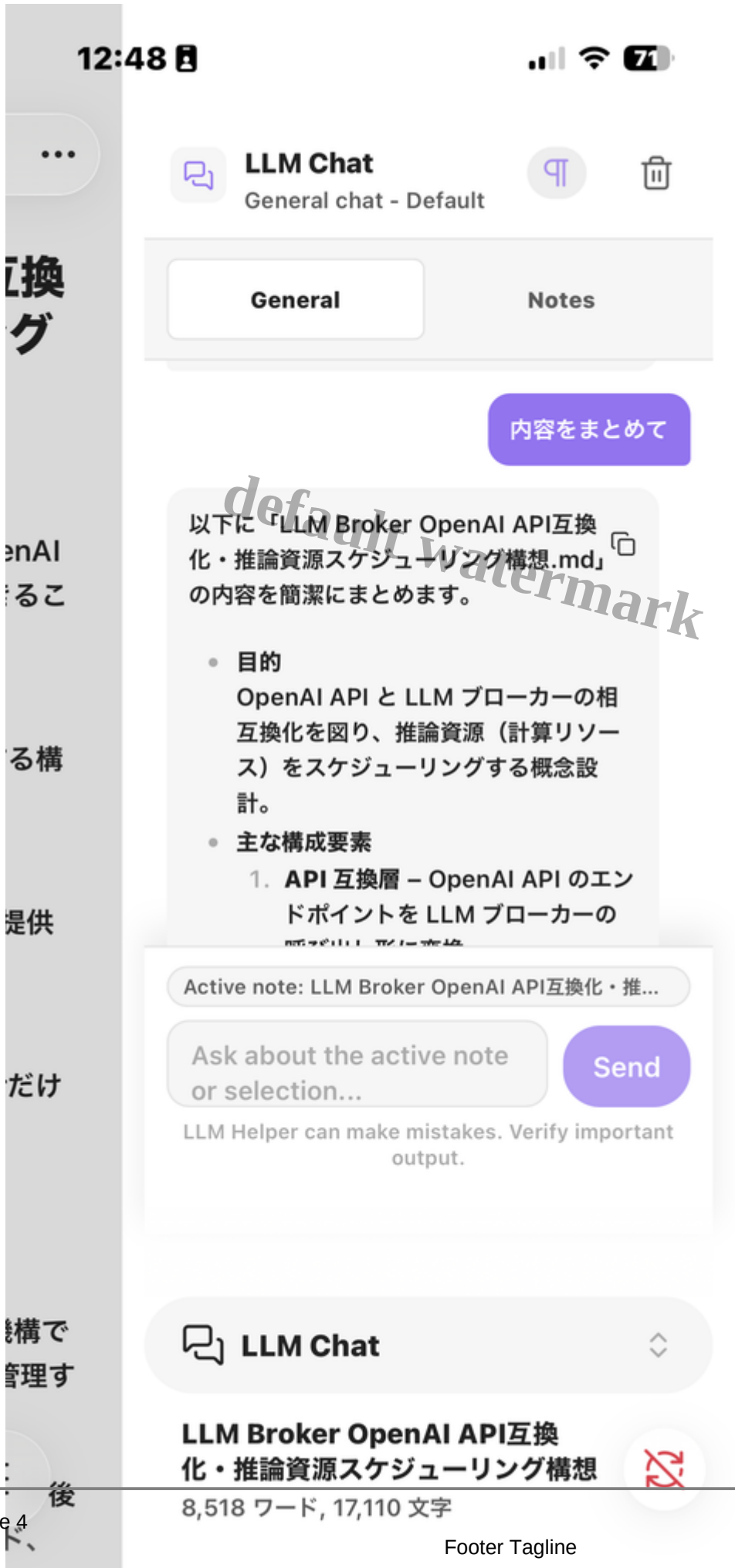
- Providerã•« OpenAI ã??é•¸ã•¶
- LLM Brokerã•®URLã??å ¥å??ã•?ã??
- ã?¢ã??ã?«ï¼?LLM Brokerã•®ä¸?ç??ã•§ã•¯ã??ã?ã??ã?£ã?¼ã?«ï¼?ã??é¸ã•¶

ã• ã•?ã•§ã•?ã??


14:12
87
Local LLM Helper
Connection
Provider
All providers use OpenAI-compatible API format
OpenAI / LM Studio / vLLM
default watermark
Chat/default server URL
Used for chat and as the embedding fallback.
Ollama: localhost:11434 | LM Studio: localhost:1234
| vLLM: localhost:8000
http://
Chat/default API key
Required for OpenAI. Embeddings inherit this unless
an embedding API key is set. For local servers, use
'not-needed'
not-needed
Embedding server URL
Optional. Leave blank to use the chat/default server.
Use this for a separate OpenAI-compatible
embedding server.
http://
Embedding API key

ã•?ã??ã•¯RAG Chatã??å®?è¡?ã?ã•?æ§?å•ã??

default watermark


12:48
LLM Chat
General chat - Default
General
Notes
内容をまとめて
以下に「LLM Broker OpenAI API互換化・推論資源スケジューリング構想.md」の内容を簡潔にまとめます。
目的
OpenAI API と LLM ブローカーの相互換化を図り、推論資源（計算リソース）をスケジューリングする概念設計。
主な構成要素
1. API 互換層 – OpenAI API のエンドポイントを LLM ブローカーの
default watermark
Active note: LLM Broker OpenAI API互換化・推...
Ask about the active note or selection...
Send
LLM Helper can make mistakes. Verify important output.
LLM Chat
LLM Broker OpenAI API互換化・推論資源スケジューリング構想
8,518 ワード, 17,110 文字

æ®?å¿µã•ªã•?ã??ã?•LLM Brokerã•®OpenAI äº?æ•?APIã•?å??ã?•è¾¼ã•¿ã?¢ã??ã?«ã??è¿?ã•?ã•¦ã•?ã•ªã•?ã•®ã•§ã?•å??ã?•è¾¼ã•¿ã•?ä½¿ã•?ã•¾ã•?ã??ã??

ã•?ã??ã•¯æ?©æ?¥ã«ä¿®æ£ã•?ã•¦ã•?ã•?ã•¨æ?•ã•?ã•¾ã•?ã??

ã•?ã??ã•¾ã•§ã•«ã?•

- Codex
- Open WebUI
- VS Codium (Continue)
- FreeCAD
- Obsidian (Local LLM Helper)
- ã•?ã•¹ã•¦ã•®ZIKUUã?¢ã??ã?ªã?±ã?¼ã?·ã?§ã?³

ã•¨ã•®æ?¥ç¶?ã??ç¢ºèª•ã•?ã•¾ã•?ã•?ã??

å°•ã•?ã•ªä¿®æ£ã•§ã•?ã•®Obsidianç?¨ã•®ã??ã?©ã?°ã?¤ã?³ã•«ã??å¯¾å¿?ã§ã••ã??ã•¯ã?ã•§ã•?ã??

default watermark

å½?å?•ã•¯ã?ã•?ã??ã?¯ã?¨ã?³ã??ã®LLMã??ã?ã?•ã?¤ã??ã?¼ã•®ä¸¸ç¶?ã??ã•?ã??ã• ã•?ã•®è??ã•?ã??ã?ã?¼ã?«ã?¼ã•¨ã•?ã•¦é??ç?ºã??å§?ã?•ã•¾ã•?ã•?ã?•æ§?ã? ã•ªã?¯ã?©ã?¤ã?¢ã?³ã??ã®æ?¥ç¶?ã??ç¢ºèª•ã•?ã•?ã??ã?•ã•?ã??ã•«é?²å??ã•?ã•¾ã•?ã??

ä°?å®?ã•?ã•¦ã•?ã??æ?¡å¼µã•¯ã?

1. ã?ªã?¯ã?¨ã?¹ã??ã??ã?ã?¥ã?¼ã?¤ã?³ã?°
2. è¤?æ?°GPUã•¸ã•®ã?ªã?¯ã?¨ã?¹ã??ã®æ?¯ã??å??ã?
3. ã??ã?«ã?•ã?¢ã?¼ã??ã?«å¯¾å¿?ï¼?GPU #1ã«ã??ã?ã?¹ã??æ?¨è«?ã?•GPU #2ã•«ç?»å?•å?¦ç•?ã•¨ã•?ã•?å½¢ï¼?
4. Observability APIã•®å®?è£ ï¼?å??ä½?ã•®è¿½è·¡ï¼?

ã•ªã•©ã•§ã•?ã??

1ã•«ã•¤ã•?ã•¦ã•¯ã?è¤?æ?°ç®?æ??ã•?ã??ã•®ã?•ã??ã?•ã?ªã?¯ã?¨ã?¹ã??ã?é• ã••ã•ªã•£ã•¦ã??ã•?ã•?ã•?ã??ç¢ºå®?ã•«å?¦ç•?ã•?ã•?ã•?ã•?ã??ã??
ZIKUUã•®ã?³ã?¢ã?µã?¼ã??ã?¹ã•¯ã?ã?•ã??ã?¯ã?°ã?©ã?³ã??ã§LLMã??ä½¿ç?¨ã•?ã??ç®?æ??ã•?å¤?ã•?ã•®ã•§ã•?ã??

2ã•«ã•¤ã•?ã•¦ã•¯ã?ä»?å¾?ã?•GPUã?µã?¼ã•?ã?¼ã??è¤?æ?°å•¸ä¸¦ã•¹ã•?é??ã•«ã?•ç?¶æ³•ã•«å¿?ã•?ã•¦é•©å??ã•ªGPUã??ä½¿ã•?ã•?ã?•ã??

3ã•«ã•¤ã•?ã•¦ã•¯ã?LLMè?ªä½?ã•?ã??ã?«ã?•ã?¢ã?¼ã??ã?«å¯¾å¿?ã§ã•ªã•?å´å•?ã•§ã??ã?•ã?¢ã??ã?ªã?±ã?¼ã?·ã?§ã?³ã•?ã??è¦?ã•?ã??ã??ã?«ã?•ã?¢ã?¼ã??ã?«ã•®ã??ã•?ã•«æ?¯ã??è??ã?•ã??ã??ã•?ã•«ã•?ã•?ã•?ã•?ã??ã??

4ã•¯ã?»ã?ã?¥ã?ªã??ã?£ã?¨ã??ã?¬ã?¼ã?µã??ã?ªã??ã?£ã?¼ã??æ?•ã•?ã•?ã??ã•?ã?•ã??
ã•?ã??ã•¯ã?ä»?ã•®ä¸»è¦•ã•ªZIKUUã?¢ã??ã?ªã?±ã?¼ã?·ã?§ã?³ã•«ã•¤ã•?ã•¦ã??ã?•å®?è£ ã??ä°?å®?ã•?ã•¦ã•?ã??æ©?è?½ã•§ã•?ã??

ã•?ã•§ã•«ã?•GPUã•?ä½?ã•?ã?•LLMã??ã?ã?•ã?¤ã??ã?¼ã•?ä½?ã•?ã?•ã?¢ã??ã?«ã•?ä½?ã•?ã??æ?•è?ã•?ã•?ã•«ã?•

- èª¬æ??ã•?ã•¦
- ç¿»è¨³ã•?ã•¦
- å??æ?•ã•?ã•¦

ã•®ã??ã•?ã•«ã?•å?¦ç•?ã•®æ?•å?³ï¼?ã??ã?ã??ã?£ã?¼ã?«ï¼?ã??ä½¿ã•£ã•¦ã?¢ã??ã?ªã?±ã?¼ã?·ã?§ã?³ã•?ã??æ?¨è«?æ©?è?½ã??å?¼ã•³å?ºã•?å?ºç?¤ã•«ã•ªã•£ã•¦ã•?ã•¾ã•?ã•?ã?•ã•«ã?•ã••ã?•Brokerã•¨ã•?ã•?å••å?•ã•§ã•¯å?ã•¾ã??ã•ªã•?è¦•æ¨¡ã•®ã?½ã??ã??ã?¦ã?§ã?¢ã•«ã•ªã??ã••ã•?ã•§ã•?ã??
æ?¨è«?è³?æº•ã?ªã?¼ã?±ã?¹ã??ã?¬ã?¼ã?¿ã?¼ã•®ã??ã•?ã•ªè¡¨ç•¾ã•?ã•µã•?ã?•ã•?ã•?ã•?ã??ã•?ã??ã•¾ã•?ã??ã??

Category

1. æ?¥ã? ã•®æ´»å?? default watermark

Tags

1. AI
2. LLM
3. LLM Broker
4. Obsidian
5. ä¸?é??å??å¸?
6. å±±æ¢¨ç??

Date Created
2026å¹´8æ??22æ?¥
Author
kazuo-tsubaki