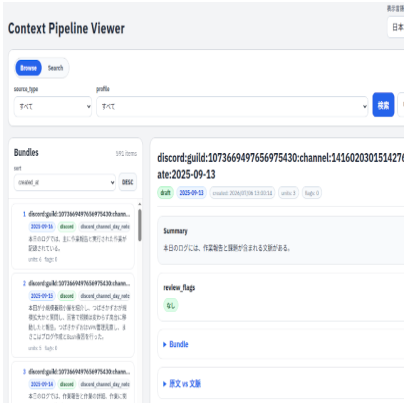


CPS â?? LLM Brokerã•«å¯¾å¿?ã?ã•?ã•¦Regression Test

Description

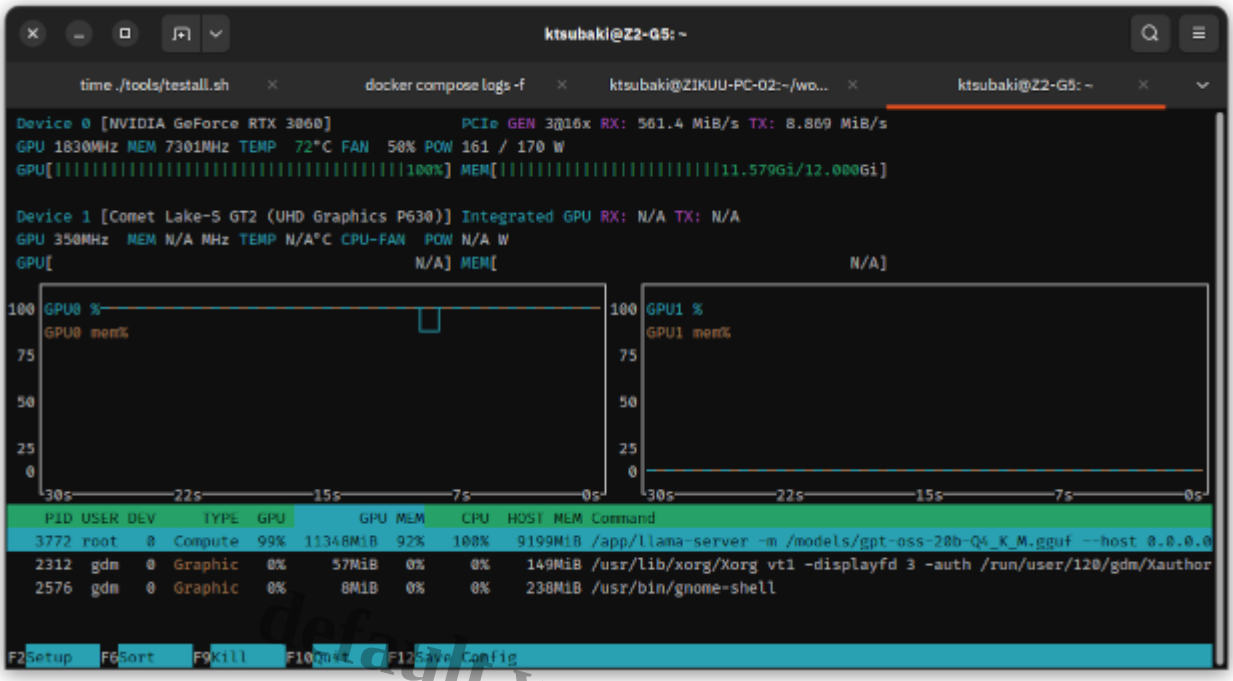
default watermark

ä¸?æ?¨æ?¥ä½?ã•£ã•?**LLM Broker**ã??ä½¿ç?¨ã•?ã??ã??ã•?ã•«CPSã??æ?¹é? ã•?ã•¾ã•?ã•?ã??

Ollama + gpt-oss:20bã•?ã??ã?•llama.cpp + gpt-oss-20b-Q4_K_M.ggufã•¸ã•®ç§»è¡?ã•§ã•?ã??

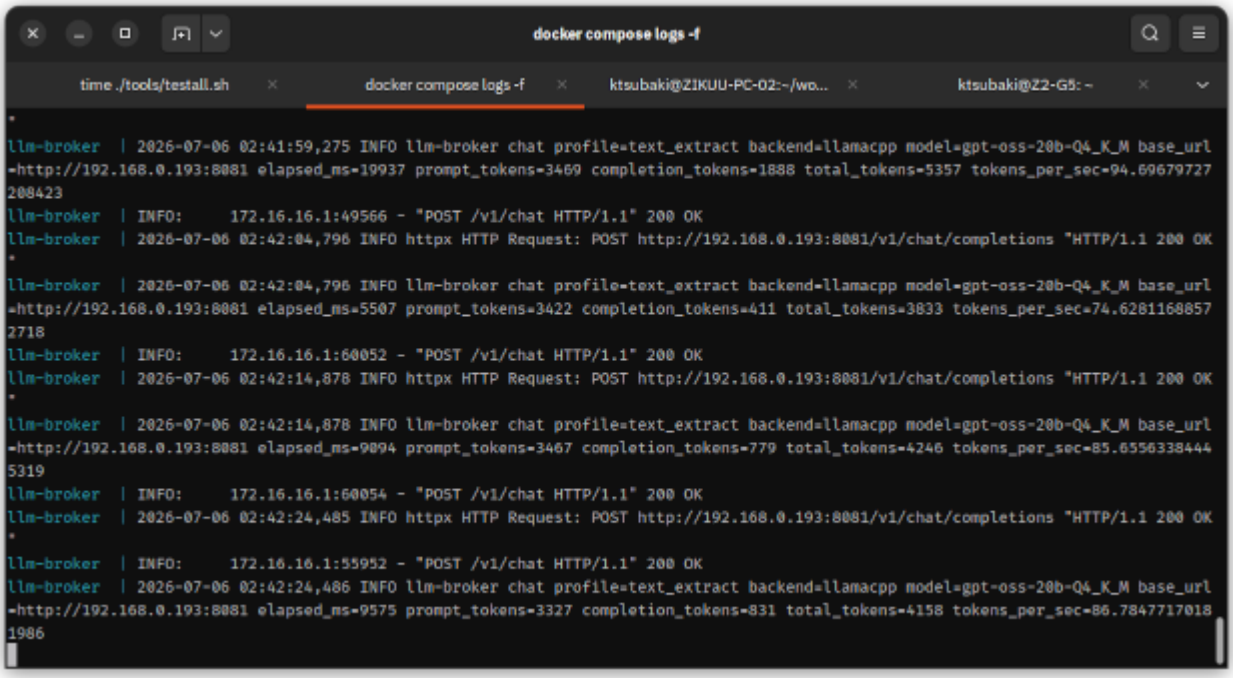
é??ç?ºç?¨ã•¨ã•?ã•¦ã?•Xeon W-1250P (6C/12T)ã•¨RTX 3060ã•?è¼?ã•£ã•?ã?¯ã?¼ã?¯ã?¹ã??ã?¼ã?·ã?§ã?³æ§llama.cpp ã?µã?¼ã?•ã?¼ã??èµ°ã??ã•?ã•¦ã•?ã•¾ã•?ã??æ?¬ç?ªã•¯ Core i5 13400 + RTX 2000 Adaã??ä½¿ã?ã•¾ã•?ã??ï¼?å¤?å??æ?¬ç?ªã•®æ?¹ã•?å°?ã•?é??ã•?ï¼?

GPUè² è•·ã??ç?£è¦?ã•?ã?•GPUã•®æ¸©åº¦ã??ã??ã?¡ã?³ã•®å??è»¢ã??è¦?ã•¾ã•?ã?? ã??ã?«ã??ã?¯ã?¼æ§1æ??é??44å??å?¦ç•?ã??ã•?ç¶?ã•?ã•¦ã?•GPUæ¸©åº¦ã•¯æ??å¤§72åº¦ã?• ã??ã?¡ã?³å??è»¢ã•¯50%ã•?ã??ã•?ã•«å•?ã•¾ã??ã•¾ã•?ã•?ã?? 12GBã•®VRAMã•§ã•¯ã?®ã?ªã?®ã?ªæ§ã•?ã•?ã?•å®?å®?ã•?ã•¦å??ä½?ã•?ã•¾ã•?ã•?ã??



default watermark

LLM Brokerã•®APIå?¼ã•³å?ºã•?ã??ç?£è¦?ã??
ã•?ã•?ã•§ã•¯ã??ã?¼ã?¯ã?³ç??æ?é??åº¦ã??è¦?ã•¾ã•?ã??
70 ã?? 100 tok/secã••ã??ã•?ã•§æ?¨ç§»ã•?ã•¦ã•?ã•¾ã•?ã•?ã??



ã??ã?¹ã??ã•?å®?èµ°ã•?ã•?ã•®ã•§ã?•æ??å¾?ã•«ã?ã?°ã•®è§£æ?•ã??ã•?ã•¾ã•?ã??

cbes_log_summary

file	segments	stored	skipped	errors	units	re
logs/store/discord_2025-09-12.json	1	1	0	0	6	0
logs/store/discord_2025-09-13.json	1	1	0	0	3	0
logs/store/discord_2025-09-14.json	1	1	0	0	4	0
logs/store/discord_2025-09-15.json	1	1	0	0	5	0
logs/store/discord_2025-09-16.json	1	1	0	0	6	0
logs/store/discord_general_2025-09-12.json	8	7	0	1	11	0
logs/store/discord_general_2025-09-13.json	3	3	0	0	9	0
logs/store/discord_general_2025-09-14.json	9	9	0	0	19	0
logs/store/discord_general_2025-09-15.json	5	5	0	0	11	0
logs/store/discord_general_2025-09-16.json	10	10	0	0	21	0
logs/store/news_aljazeera_world.json	0	5	0	0	5	0
logs/store/news_aljazeera_world_general.json	0	5	0	0	9	0
logs/store/news_bbc_business.json	0	5	0	0	5	0
logs/store/news_bbc_business_general.json	0	5	0	0	7	0
logs/store/news_bbc_world.json	0	5	0	0	5	0
logs/store/news_bbc_world_general.json	0	5	0	0	12	0
logs/store/news_guardian_business.json	0	5	0	0	5	0
logs/store/news_guardian_business_general.json	0	4	0	1	20	0
logs/store/news_guardian_technology.json	0	5	0	0	6	0
logs/store/news_guardian_technology_general.json	0	5	0	0	29	0
logs/store/news_guardian_world.json	0	5	0	0	6	0
logs/store/news_guardian_world_general.json	0	5	0	0	19	0
logs/store/news_nhk_world.json	0	5	0	0	6	0
logs/store/news_nhk_world_general.json	0	5	0	0	12	0
logs/store/news_nyt_business.json	0	5	0	0	6	0
logs/store/news_nyt_business_general.json	0	4	0	1	8	0
logs/store/news_nyt_home.json	0	5	0	0	5	0
logs/store/news_nyt_home_general.json	0	5	0	0	9	0

default watermark

Ollamaã•?ã??llama.cppã•«LLM Runnerã??å¤?æ?´ã•?ã•¦ã?•GGUFå½¢å¼•ã•®ã?¢ã??ã?«ã•«å¤?ã?•ã•£ã•¦ã?•è?¥å¹²ã?•LLMã•®ã?¬ã?¹ã?•ã?³ã?¹ã•«å¤?å??ã•?ã??ã•¾ã•?ã??å°?ã•?ã?¨ã?©ã?¼ã??å¢?ã•?ã•¾ã•?ã?? ã•?ã??ã•?ã??ã?•ä¸?ã•¤ã•?ã•¤ã?¨ã?©ã?¼ã??æ½°ã•?ã•¦ã•?ã••ã•¾ã•?ã•?ã?•å±±å ´ã•¯è¶?ã?ã•?æ??ã•?ã•?ã•?ã•¾ã•?ã??

Category

1. æ?¥ã? ã•®æ´»å??

Tags

1. AI
2. CPS
3. llama.cpp
4. LLM
5. ä¸?é??å??å¸?
6. å±±æ¢¨ç??
7. æ??è??æ•?

default watermark

Date Created
2026å¹´7æ??6æ?¥
Author
kazuo-tsubaki