

ā•ā??ā??ā?©ç?`ā•ā!āpšè!æ`jè`?è^aā?ā?ā?«ā??æ?±ā•ā?ā?•ā•®ā?μā?³ā??ā?«ā??ā?-
 ā?°ā?©ā? ā??æ?_ā?ā•ā•è©|è|é?`è^aā?ā•ā?ā•ā?4ā•ā??

main.rs - LLM-Inference-API-in-Rust - VSCode

ファイル 編集 選択 表示 移動 実行 ターミナル ヘルプ

main.rs x Cargo.toml

language_model_server > src > main.rs > infer

```
1 use hyper::service::{make_service_fn, service_fn};
2 use hyper::{Body, Request, Response, Server};
3 use std::net::SocketAddr;
4 use serde::{Deserialize, Serialize};
5 use std::convert::Infallible;
6 use io::Write;
7 use path::PathBuf;
8
9 #[derive(Debug, Deserialize)]
10 struct ChatRequest {
11     prompt: String,
12 }
13
14 #[derive(Debug, Serialize)]
15 struct ChatResponse {
16     response: String,
17 }
```

問題 出力 デバッグ コンソール ターミナル ポート

```
Running 'target/debug/lln_handler'
Server listening on port 8083...
^C
(streamlit) ktsubaki@ubuntu2204-i7:~/workspace/LLM-Inference-API-in-Rust/language_model_server$ cargo run
Compiling lln_handler v0.1.0 (/home/ktsubaki/workspace/LLM-Inference-API-in-Rust/language_model_server)
Finished dev [unoptimized + debuginfo] target(s) in 6.65s
Running 'target/debug/lln_handler'
Server listening on port 8083...
Error in inference: FileDoesNotExist { path: "/home/ktsubaki/workspace/LLM-Inference-API-in-Rust/language_model_server/open_llama_3b-f16.bin" }
^C
(streamlit) ktsubaki@ubuntu2204-i7:~/workspace/LLM-Inference-API-in-Rust/language_model_server$ pwd
/home/ktsubaki/workspace/LLM-Inference-API-in-Rust/language_model_server
(streamlit) ktsubaki@ubuntu2204-i7:~/workspace/LLM-Inference-API-in-Rust/language_model_server$ cargo run
Compiling lln_handler v0.1.0 (/home/ktsubaki/workspace/LLM-Inference-API-in-Rust/language_model_server)
Finished dev [unoptimized + debuginfo] target(s) in 6.66s
Running 'target/debug/lln_handler'
Server listening on port 8083...
Error in inference: InvalidMagic { path: "/home/ktsubaki/workspace/LLM-Inference-API-in-Rust/open_llama_3b-f16.bin", magic: 4034b50 (PK) }
^C
(streamlit) ktsubaki@ubuntu2204-i7:~/workspace/LLM-Inference-API-in-Rust/language_model_server$
```

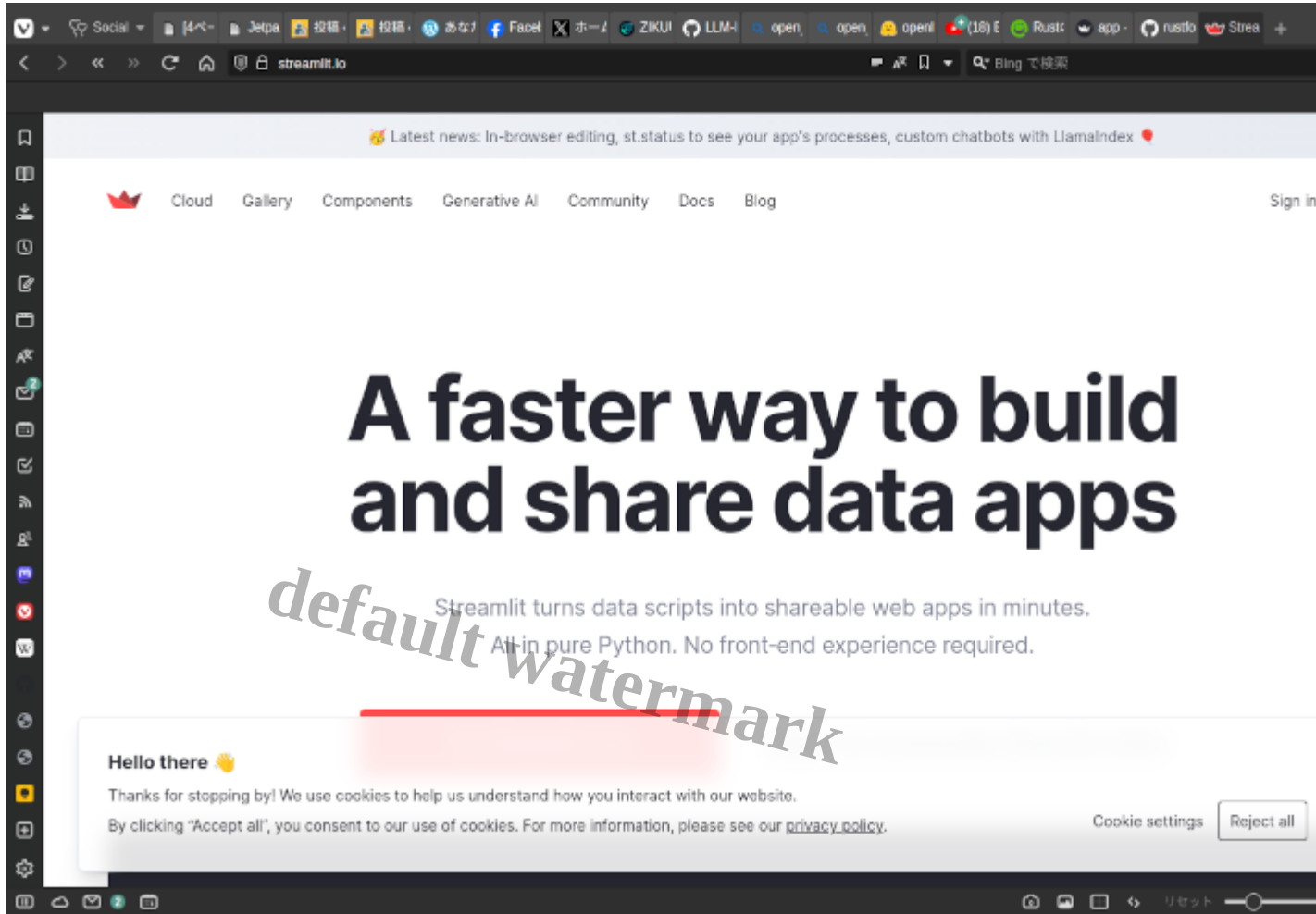
Rust [language_model_server] Starting rust-analyzer

行 17、列 73 スペース:4 UTF-8 LF

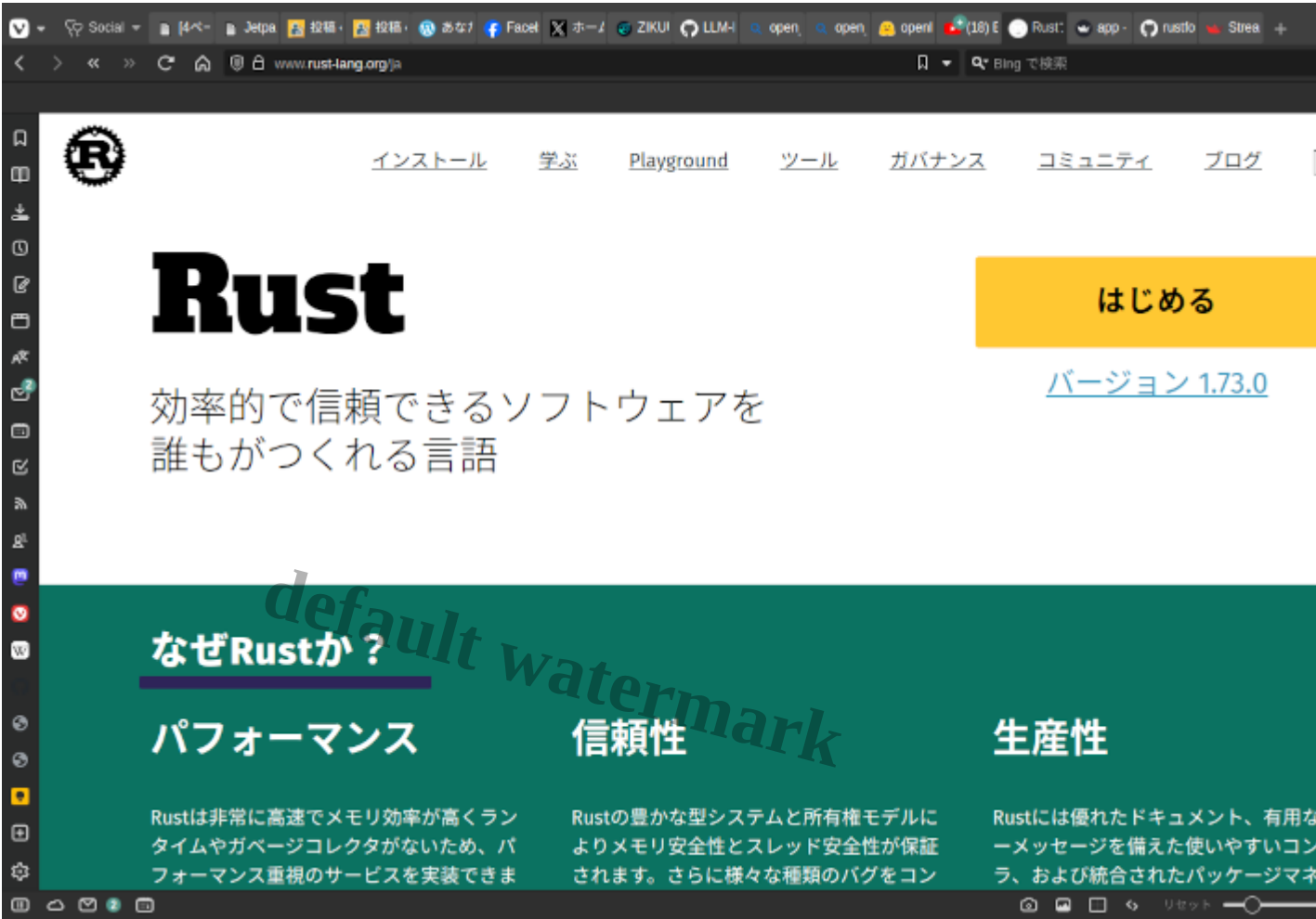
ã?ã?©ã?¼ã•®é?£ç||?ã•šã•ªã•?ã•ªã•?¿?ã•«é?²ã•¿ã•¾ã•?ã•??ã•?ã½?ã•¨ã•?ã•??ã•??é•?ã•?ã•?ã•?ã•¨ã•?ã•?
?ã•¾ã•?ã•??

Rustã•Sæ?_ã•?ã??ã•?è`?èª?ã?¢ã??ã?«ã??ä½¿ã•£ã|æ??ç« ç??æ?•ã??ã•?ã??APIã?µã?¼ã?•
ã?¼ã??ã½?ã•£ã|ã•?ã¾ã•?ã°ã•?ã??ã?ã³ã??ã?ã³ã??ã•®Webã?¢ã??ã?ªã?±ã?¼ã?ã?Sã?³ã•?ã??ã
•ã•®APIã??å?¼ã•³å?ã•?ã•|æ??æ?_ç??æ?•ã?ã•Sã•ã¾ã•?ã??

ã??ã?ã³ã??ã?“ã³ã??ã•ç°jã?ã«Ulã??ã½?ã??ã??Streamlitã•ã•©ã??ã½¿ã•?ã•è?¬ã?ã•?ã??ã•
?ã??ã•¾ã•?ã??ã??



ä•?ä??ä•?ä??ä•®ä?½ä??ä??ä?|ä?§ä?¢ä?¨ä?³ä?_ä??ä?¢ä??ç?®æ??ä•?ä•ªä??ä?•Cä?•C++ä?•Javaä•ªä
 •©ä•®æ?ä?ä•?ä??ä?³ä?³ä??ä?ªä?«ä?²ä•®è¨èª?ä•ä•?ä•§ä•ä•ä?•Pythonä??JavaScript (TypeScript)ä•
 ªä•©ä•®ä?ªä?³ä?¿ä?¼ä??ä?ªä?¿ä?¼ä??è¨èª?ä??æ?¥è??ä•«ä?©ç?¨è?ä?ä?¢?ä?ä?|ä?ä??Rustäªä
 ©ä?²ä?|ä•¶ä•®ä•²ä¿é?ä?²ä•ä•æ?ä?ä?ä?¼ä?ä??ç?¹ä•«æ?§è?½ä¨ä¿¿ä?¼æ?§ä•«ä?ªä??ä?Rustä•
 ¯ä»?ä¾?æ¨?æ°è¨èª?ä«äªä??ä•æ?ä•£ä•|ä?ä?ä?¼ä?ä??



Category

1. 言語

Tags

1. AI
2. Python
3. Rust
4. Streamlit

Date Created

2023年11月4日

Author

kazuo-tsubaki