



Linux知識不要！手のひらサイズの
ローカルLLM チャットコンピュータ

LLMCardputer の つくりかた



@GOROman

甘

口



自己紹介



(省略)



LT おしながき (5分)

- LLMCardputer とは？
- M5 Cardputer とは？
- M5 Module LLM とは？
- つくりかた
- 苦労した点（罨・地雷）



209



150

 @GOROman

ローカルLLMを手のひらサイズで動かしてみよう！ M5 Cardputer + ModuleLLM

AI M5stack LLM Cardputer Module-LLM

♡ 209 🗨️ 👍 🙌 +3

最終更新日 2025年02月03日 投稿日 2025年01月26日 197553 views

はじめに

OpenAI の ChatGPT のような LLM(大規模言語モデル)、このLLMをローカルで動かすにはCUDAが動くGPUだったり、Apple SiliconなMac (しかもメモリ大量) だったりが必要でした。そんな中、近年では組み込み向けのマイコン用のエッジLLMモジュールなどが登場してきました。

この記事では、M5Stack社が発売しているクレジットカードサイズのキーボード付きマイコンボード「M5Stack Core2」にローカルLLMが動く「ModuleLLM」を紹介...



209



150

 @GOROman

ローカルLLMを手のひらサイズで動かしてみよう！ M5 Cardputer + ModuleLLM

AI M5stack LLM Cardputer Module-LLM

♡ 209 🗨️ 👍 🙌 +3

最終更新日 2025年02月03日 投稿日 2025年01月26日 197553 views

はじめに

OpenAI の ChatGPT のような LLM(大規模言語モデル)、このLLMで動かすにはCUDAが動くGPUだったり、Apple SiliconなMac (メモリ大量) だったりが必要でした。そんな中、近年では組み込み向けコン用のエッジLLMモジュールなどが登場してきました。

この記事では、M5Stack社が発売しているクレジットカードサイズのキーボード付きコンピュータ「M5 Cardputer」でローカルLLMが動く「ModuleLLM」を紹介...

puter + ModuleLLM

Module-LLM

2025年01月26日 197553 views



LLMCardputer とは？



LLMCardputerとは？

M5Stack社が出したLLM用モジュールを使って
ChatGPTのショボいやつがネットなしで
手のひらの上で動く。ローカルLLM！

一応日本語でチャットもできる。





LLMCardputerとは？

M5Stack社が出したLLM用モジュールを使って
ChatGPTのショボいやつがネットなしで
手のひらの上で動く。ローカルLLM！

一応日本語でチャットもできる。

実用度はない（←重要）





LLMCardputerとは？

M5Stack社が出したLLM用モジュールを使って
ChatGPTのショボいやつがネットなしで
手のひらの上で動く。ローカルLLM！

一応日本語でチャットもできる。

実用度はない（←重要） 浪漫はある（←重要）





Cardputer とは？



SKU 9277

M5Stack Cardputerキット (M5StampS3付属)

送料区分: 200

商品コード: M5STACK-K132

発売日: 2023-10-13

在庫数: 39

短いURL: [scci.to/9277](https://ssci.to/9277)

¥5,962 (税込)

数量

-	1	+
---	---	---

カートに追加



Module LLM

(LLM モジュール?)

とは?



Module LLM (LLMモジュール) とは？

AXERA (アレクサではない) 社のAX630Cを登載。

今話題のLLM (大規模言語モデル Large Language Model) が使えるモジュール。あんま大規模は無理。

モジュールと言いながら、実態はSBC (シングル・ボード・コンピュータ) なのでラズパイとかの方が近いLinux(ubuntu)がプリン。

LLMな理由は、AI向けの計算機能 (NPU) がチップ内にある。



Home

Products

Algorithms

Solutions

About Us



CN

Smart Driving Chips



Smart night vision Chips →

Smart Vision Chips

Light Smart Vision Chips

Coaxial HD TX

Coaxial HD RX

Analog-Digital Hybrid Chips

Smart night vision Chips

AX620A



AX620U



AX650A



AX650N



AX630C



AX620Q



Smart night vision Chips

Home

Products

Smart Driving Chips

Smart night vision Chips

Smart Vision Chips

Light Smart Vision Chips

Coaxial HD TX

Coaxial HD RX

Analog-Digital Hybrid Chips



F0.9 StarLight Camera

FHD Video



Motion detecti



Easy for installa



Remote view



Multi-user view



All Color XM



XMEYE



ICsee



6MP:AX620+SC450AI
8MP:AX620U+OS04A1



CN

AX650A



AX620Q





Module LLM (LLMモジュール) とは？

AXERA (アレクサではない) 社のAX630Cを登載。

今話題のLLM (大規模言語モデル Large Language Model) が使えるモジュール。あんま大規模は無理。

モジュールと言いながら、実態はSBC (シングル・ボード・コンピュータ) なのでラズパイとかの方が近いLinux(ubuntu)がプリン。

LLMな理由は、AI向けの計算機能 **(NPU)** がチップ内にある。





NPU

NPU (Neural Processing Unit) とは？

ニューラルネットワークを
プロセッシングする
ユニット

ニューラルネットワーク (Neural Network) は、人間の脳内にある神経細胞 (ニューロン) とその接続構造を模倣した機械学習モデルです。入力層、隠れ層 (中間層)、出力層の3つの層から構成され、各層のニューロンがシナプスに相当する「重み」を介して接続されています。 WIKIPEDIA

ニューラルネットワークは、入力データに対して重みを調整し、活性化関数を適用することで、複雑なパターンの学習や予測を行います。特に、隠れ層を多層に重ねたディープニューラルネットワーク (DNN) は、ディープラーニング (深層学習) として知られ、画像認識や音声認識、自然言語処理など、多様な分野で高い性能を発揮しています。 ビジネス+IT

ニューラルネットワークの学習には、誤差逆伝播法 (バックプロパゲーション) や確率的勾配降下法 (SGD) などの手法が用いられます。これらの手法により、モデルの予測誤差を最小化するように重みが調整されます。 楽天グループ株式会社

また、ニューラルネットワークにはさまざまな種類があり、畳み込みニューラルネットワーク (CNN) は画像処理に、再帰型ニューラルネットワーク (RNN) は時系列データや自然言語処理に適しています。 SKYGROUP

このように、ニューラルネットワークは多様なデータ ↓ パターンを学習し、高度な予測や分類を可能にする強力なツールとして、現代のAI技術の中核を担っています。

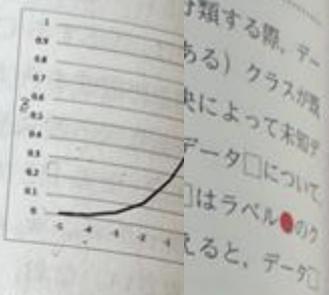
高校 情報Ⅱ

一般教養に？



1 シグモイド関数 表計算ソフトウェアを用いて、変数 x を入力とするシグモイド関数の値を計算し、関数を可視化することを考える。

	A	B	C	D	E	F	G	H	I	J	K	L
1	w_1	1										
2	w_0	0										
3	x	-5	-4	-3	-2	-1	0	1	2	3	4	5
4	$f(x)$											



セル B4 から L4 の行では、対応するセル B3 から L3 の値を入力とするシグモイド関数の値を計算する。そのため、数式を入力し、それをセル C4 から L4 までのセルにコピーして、セル C4 から L4 までのセルにコピーする。なお、ネイピア数 e を底とする $\text{EXP}()$ と表す時、シグモイド関数は $1/(1+\text{EXP}(-(w_1 \cdot x + w_0)))$ と表すことができる。 w_1, w_0 はそれぞれ変数 x と定数項の係数で、 w_1 の値はセル B1 に、 w_0 の値はセル B2 にそれぞれ入力とする。数式を入力したら、セル B3 から L4 までを選択してグラフ折れ線グラフを選択しグラフを可視化することができる。

分類する際、データがある) クラスが異なる) によって未知データについてラベルを付ける。データ

セル●のデータがよって、データと分類される。

2 ロジスティック回帰による分類 次の(1)~(3)の空欄に適切な語句を答えなさい。
 (1)は、入力を任意のクラスに割り当てる問題である。ロジスティック回帰は(1)の代表的な

2 ニューラルネットワーク

人間の神経細胞(ニューロン)は図8(右)に示すように、外部からの刺激やほかの神経細胞からの情報を受け取る樹状突起、情報を神経細胞の末端に伝達する軸索、情報をほかの神経細胞に伝達するシナプスからなる。ニューラルネットワークは、このような神経細胞間の情報伝達の仕組みをコンピュータでモデル化したものである。
 ニューラルネットワークの基本的な仕組みは、入力に対する重みの線形変換(線形変換)とその非線形変換であり、図8(左)に示すように、ロジスティック回帰モデルは入力の重み付き和とシグモイド関数による非線形変換という単純なニューラルネットワークと見ることができる。ここで、非線形変換を行う関数は、一般に活性化関数と呼ばれ、神経細胞間の情報伝達を模したものになっている。

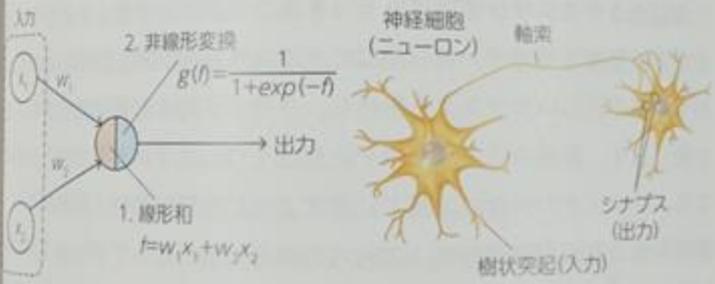


図8 ロジスティック回帰モデル(左)とニューロンの構造(右)

図9(左)に示すように、ニューラルネットワークは、入力、中間層、出力層から構成される。中間層では入力の線形変換と非線形変換の処理

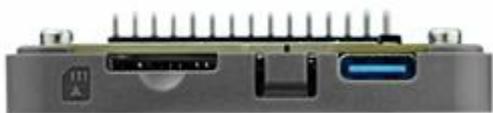


Module LLMのつかいかた

M5Core2などにブッ刺してスタックさせて使うのが本来の使い方。

基本はUART（シリアル通信）でLLM機能と通信する。

（TCPポート10001も使えるが今回は割愛）



STAMPS3 CARDPUTER CARD SIZE COMPUTER



M5STACK



MODULE LLM

KWS ASR LLM TTS



BUTTON
BOOT

UART
INTERFACE
AX630C
3.2TOPs@INT8

Tx	Rx	Core	DEFAULT
Core1	013	014	015
016	017	018	019
020	021	022	023
024	025	026	027
028	029	030	031
032	033	034	035
036	037	038	039
040	041	042	043
044	045	046	047
048	049	050	051

4GB
LPDDR4

RGB
x3

MIC

USB-OTG
HOST
DEVICE
UPGRADE

1W@SPK

32GB
eMMC5.1



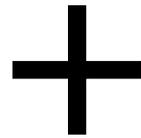
M5STACK

COMPATIBLE & NOT INCLUDE



フュー^ジジョン

!!
!!!!
!!





+

+





+

+





+

+





+



+



+

+





+

+





+

+





+

+





+

+





+

+





+

+





+

+





+

+





+

+



合体

!!!

フュージヨン

合体

!!!



+



合体

!!!



+



合体

!!!



+



合体

!!!



+



合体

!!!

合体

!!!

合体

!!!

合体

!!!

合体

!!!

合体

!!!

合体

!!!

合体

!!!

合体

!!!

合体

!!!

CARD
COMPUTER

Mic DataG46
ClkG43

ortance. It is also
the world's 4th-la
rgest city by popul
ation.

M5





LLMCardputer

1分でわかる
LLMCardputer
の
つくりかた

①分解

と

ケースの加工

削る

ニッパーでゴリゴリやりました。

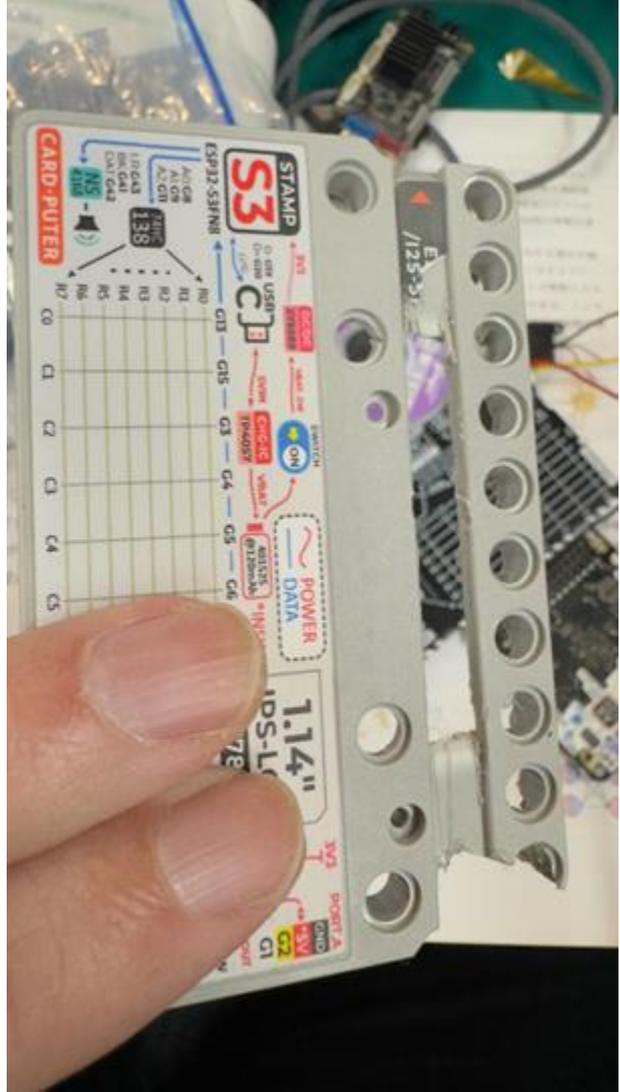
(超音波カッターがいらしい)



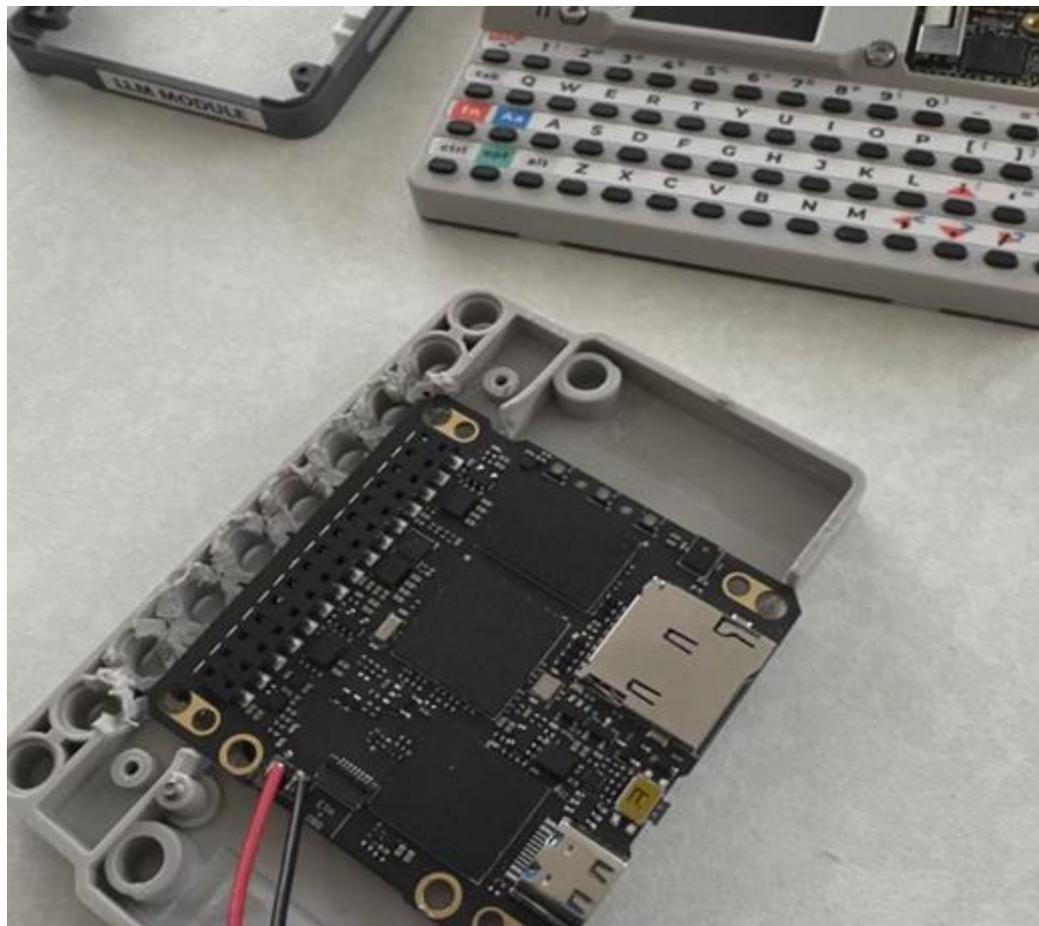
切る

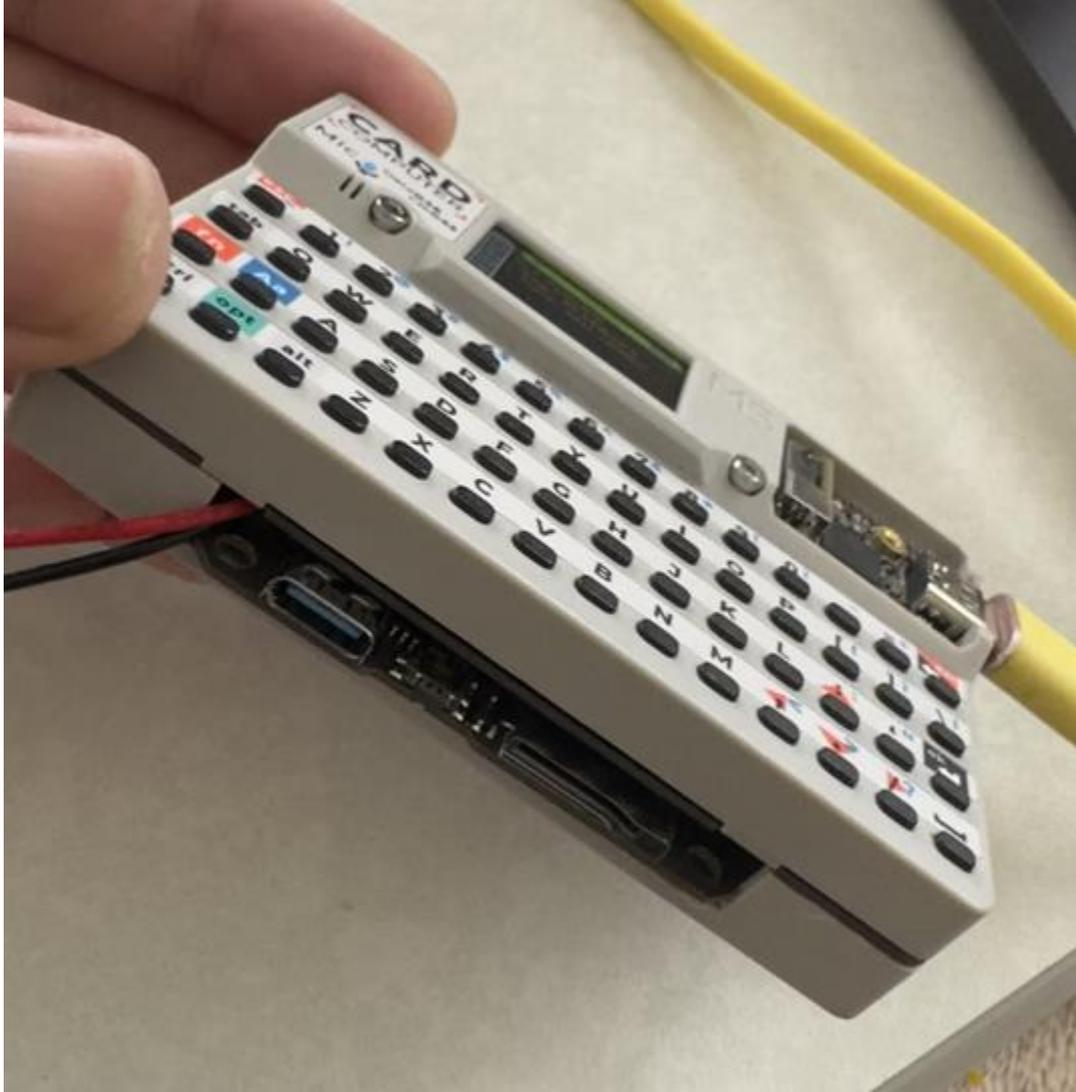
カッターでゴリゴリやりました。

(超音波カッターがいらしい)



あらかじめ分解しておいた
Module LLM基板を入れる





②ハンダづけ

ハンダづけ

Cardputer とつなぐ



Cardputer側(GROVE)

ModuleLLM側(M.BUS)

G(黒)

GND

5V(赤)

5V

G2(黄)

UART(Tx)

G1(白)

UART(Rx)

③ プログラム を 書く

Files

master + 🔍

Go to file t

- examples
 - Advanced
 - Basic
 - button
 - buzzer
 - display
 - ir_nec
 - keyboard
 - inputText
 - inputText.ino
 - multiPress
 - singlePress

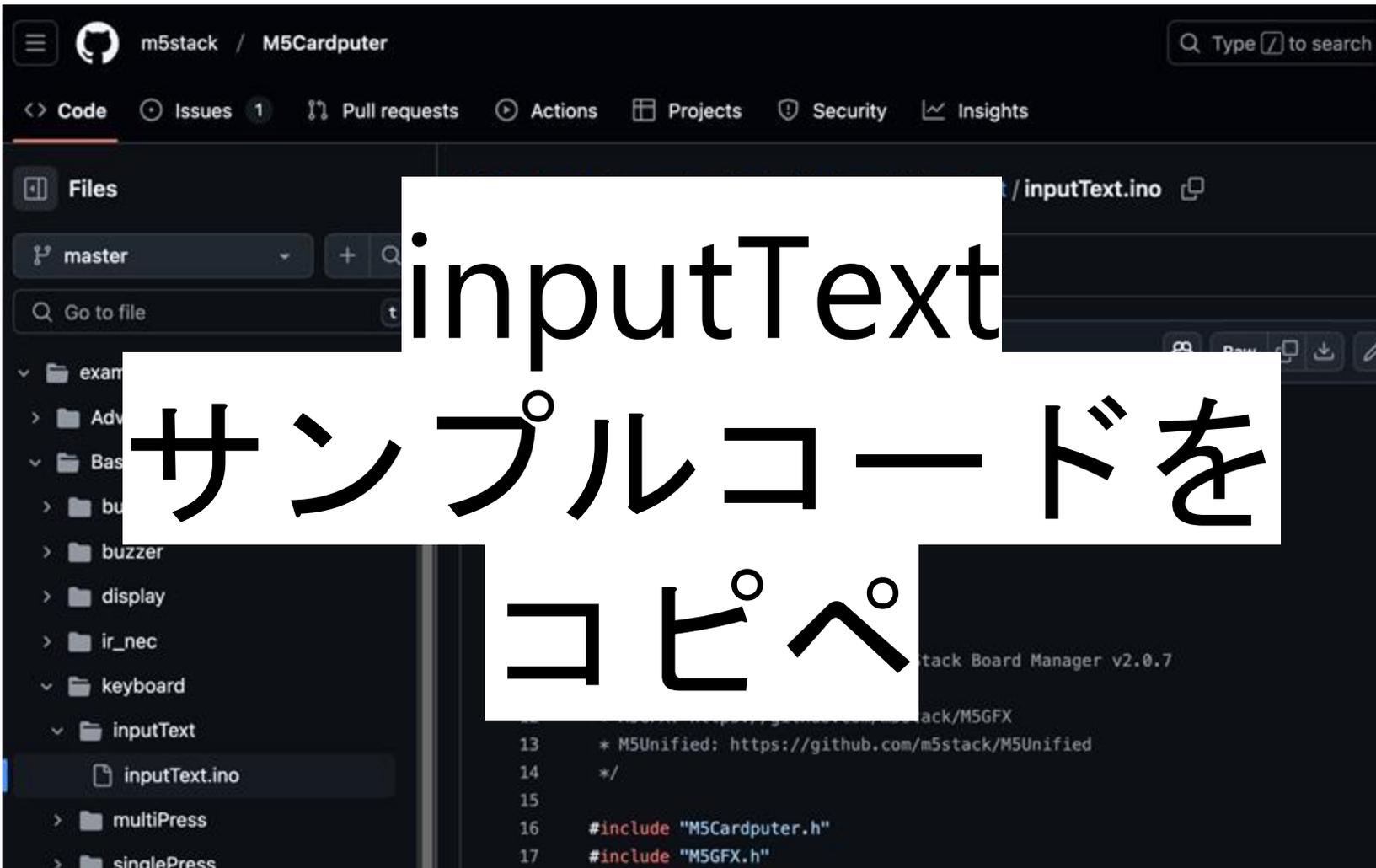
M5Cardputer / examples / Basic / keyboard / inputText / inputText.ino

Gitshaoxiang update lib and add usb keyboard demo

Code Blame 73 lines (62 loc) · 2.27 KB

Raw Copy Download Edit

```
1  /**
2   * @file inputText.ino
3   * @author SeanKwok (shaoxiang@m5stack.com)
4   * @brief M5Cardputer input text test
5   * @version 0.1
6   * @date 2023-10-13
7   *
8   *
9   * @Hardwares: M5Cardputer
10  * @Platform Version: Arduino M5Stack Board Manager v2.0.7
11  * @Dependent Library:
12  * MSGFX: https://github.com/m5stack/MSGFX
13  * M5Unified: https://github.com/m5stack/M5Unified
14  */
15
16  #include "M5Cardputer.h"
17  #include "MSGFX.h"
```



2つのサンプル
プログラムを

フュー^ジジョン

!!
!!!!
!!

甘口

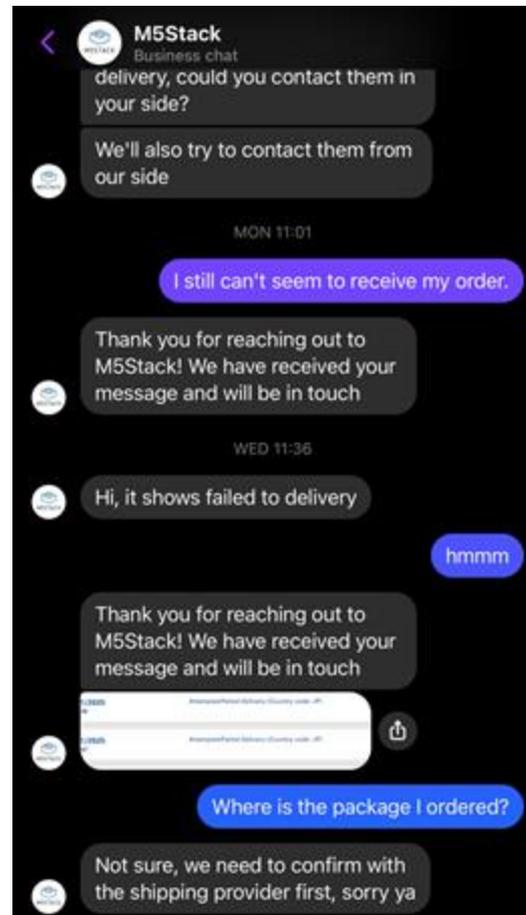


ね、簡単でしょ？

LLMCardputer開発

苦労した点（罨・地雷）

- ・ 去年M5 Storeで注文したModule LLMが**ロス**トして未だ届かない。スイッチサイエンスで買った。
- ・ IR LEDがモゲるとCardputerの画面がズれる（関係ない）
- ・ ファミリーベーシック風の音の再現（関係ない）
- ・ LLMの反応で null がやたら出る



ケースを
破壊したくない
あなたに . . .

3Dプリント済みケース

@Mu_Alexius000







Coming Soon....

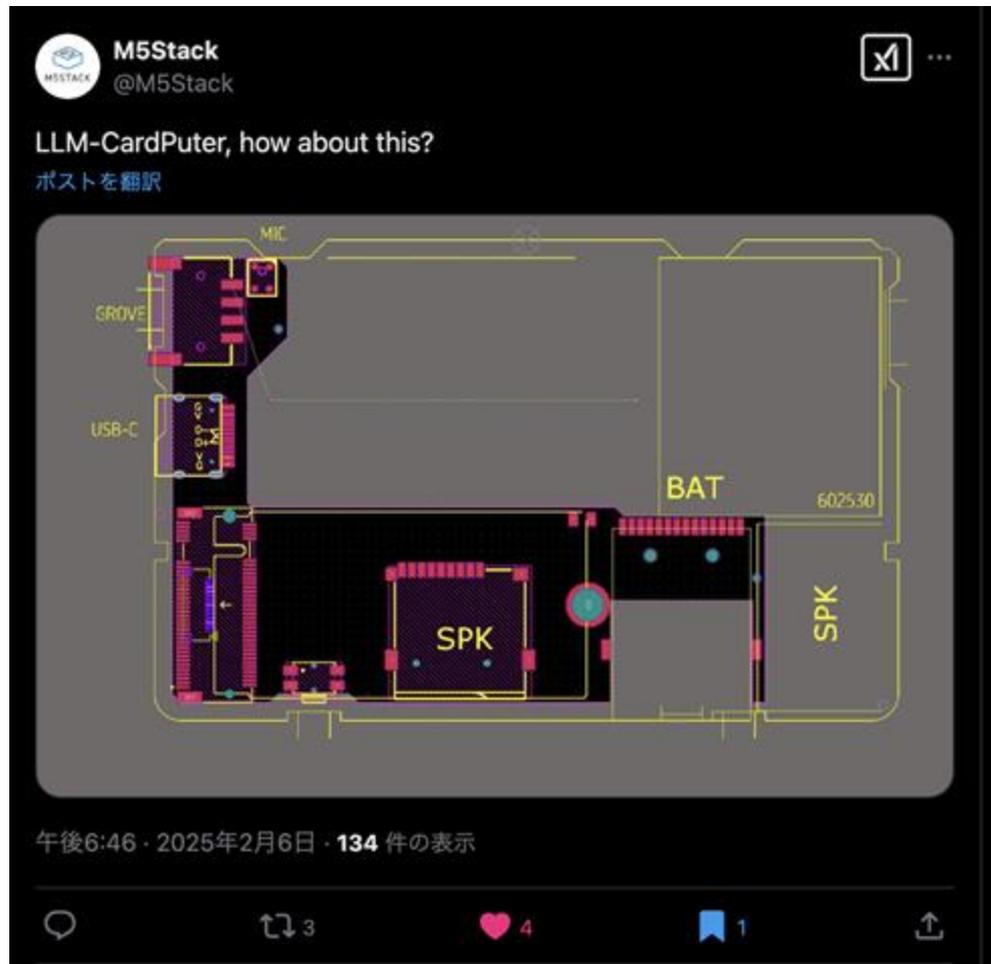


みんなも
つくってみてね

END

One More Thing

LLM-CardPuter 公式で 開発開始???



True

END