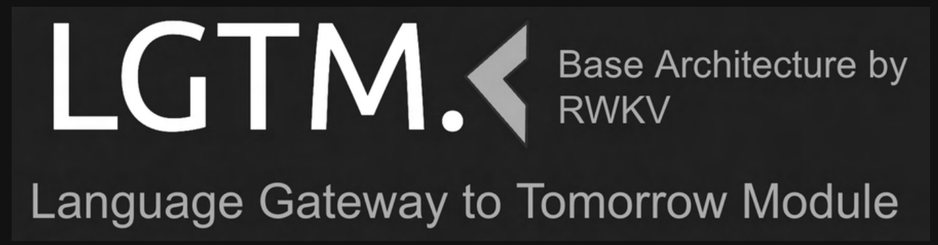
Language Gateway to Tomorrow Module

みんなが使えるＡＩを。


LGTM.
Base Architecture by
RWKV
Language Gateway to Tomorrow Module

About LGTM

LGTM開発統括よりご挨拶

OpenMOSE @_m0se_ ・XX月XX日

LLMが数多く誕生していく中でLGTMはどこまでも
小さくて賢い言語モデルを追求していきます。

どんなスマートフォン / PCでもAIが当たり前に
機能の一つとして備わっているので、小さくて賢
いモデルが必要になると考えております。

「身近なAI」を築き上げる日本の先駆者として
日々LGTMを磨き続けます。
詳しくはX(@_m0se_)をご覧ください。

❤ 10100

LGTMって何？

大規模言語モデルの一つ、RWKVを基にした言語モデルです。
「みんなが使えるAI」を実現するために、誰もが触りやすく、製品を作りやすい言語モデル LGTM を開発しました。

RWKV

LGTMのベースモデルで
多言語に対応している

LGTM. Base Architecture by RWKV
Language Gateway to Tomorrow Module

チューニングを容易に改良し
特化型言語モデルの開発など
優秀な小さいAIを開発可能

製品例
チャットボット
「くるるえーあい」

誰でも安く簡単に
オリジナルAIサービスの開発が可能に

お悩みの声

社内規則を社員に周知したい、、
社員同士の会話を増やしたい！

誰にも話せない愚痴を聞いてくれる
相手が欲しい

自分の代わりにコーディングしてくれる人
がいたら、、

AIを取り入れたいけどクラウドだと
情報漏洩が怖い、、

ChatGPTやLlaMAを使っても出来ない、
AIの活用方法が分からない、というお悩みの声をいただきます

LGTMの活用シーン

社内の情報屋 wikiボット

社内のあらゆる情報を学習させることで社内ルールや、社員の趣味や嬉しいことなどをいつでも聞いて相談ができ、コミュニケーションの促進だけでなく新卒社員の退職率軽減にも役立ちます。

コーディングAI

プログラマー、足りてますか？
御社のプログラミング規則、独自ライブラリ、これまでに制作したコードを学習させることで、御社独自のコーディングAIを作成することが可能です。

自分だけの理想の話し相手

自由に口調や性格をカスタマイズ可能なキャラクター性の高いAIを作ることができます。AI Tuberの制作や、介護現場での話し相手、また、あなただけの理想のパートナーを作ることも可能です。

LGTMは他にも様々な課題を解決することができます

比較表

	LGTM	GPT	LLaMA
導入レベル	初心者～上級者	中級～	上級者
生成スピード	◎（RTX3090にて7B 200Token/sec）	○	○
学習	クラウド/オンプレミスで簡単 最短２分で学習	限定的なファインチューン/ 基本はプロンプトチューニング	難しい
コスト	世界最安・月額固定でファインチューニング可能	APIによる従量課金	高額な機材が必要
ビジネス面	製品用特化モデルの作成にかかる時間が削減できるため、エンド製品を早くリリースできる	OpenAI社に依存したシステムを構築する必要がある	ファインチューニングに時間がかかるため、製品実用化までに時間がかかる

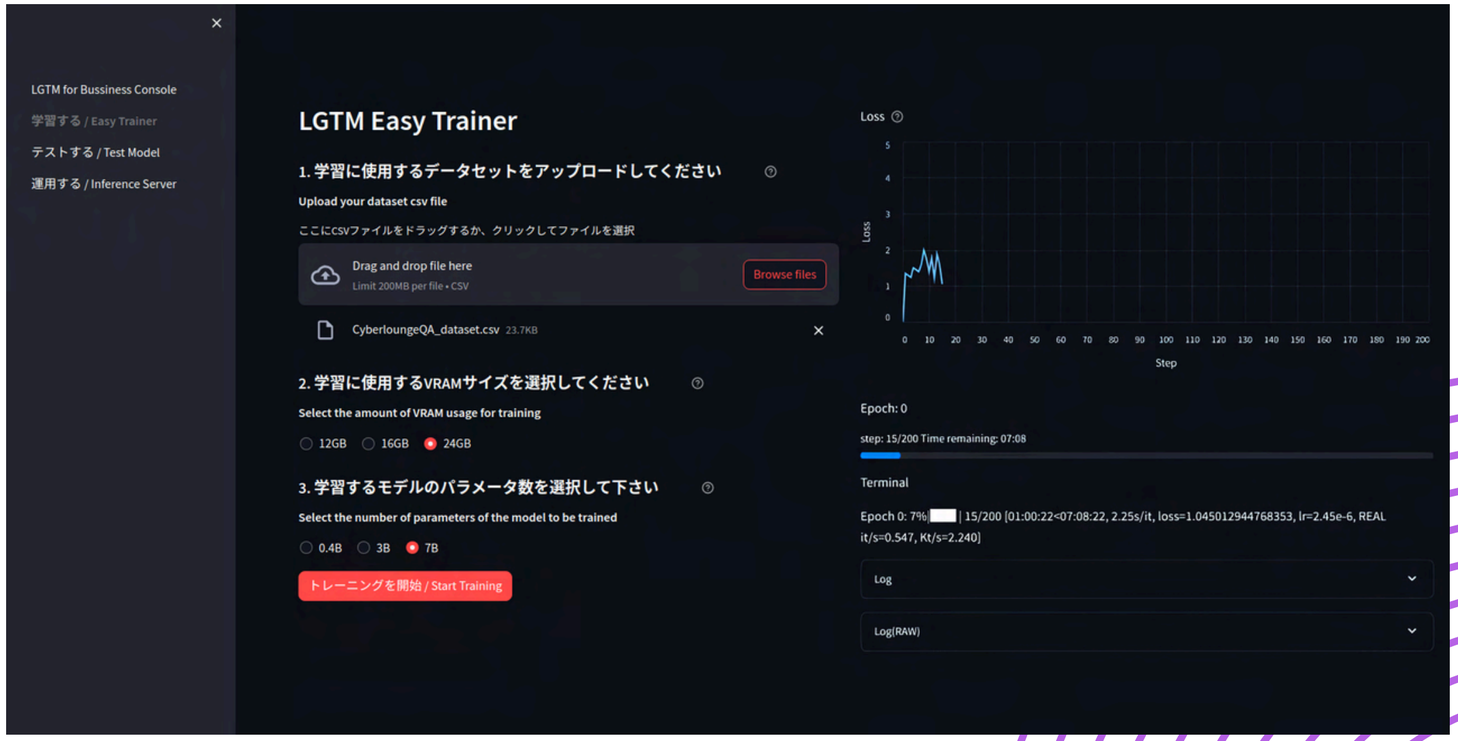
直感的に操作できる学習環境
LGTM Easy Trainer

「LGTM Easy Trainer」では、所定の形式のCSVファイルを読み込ませて、画面の指示に従っていくだけで、あなただけの特化言語モデルを作成できます。

直感的な操作

簡単なデータセット

圧倒的な学習スピード



LGTM. Base Architecture by RWKV
Language Gateway to Tomorrow Module

LGTM 技術者向け説明

小規模で高い日本語性能

LGTMでは、30億パラメータおよび70億パラメータのモデルを提供予定

少ない計算資源で高い効率と性能を追求した日本語の理解と生成に特化させたモデル

コンテキスト数無限処理

多数の言語モデルが採用しているトランスフォーマーアーキテクチャとは異なり、LGTMはRNNを使用

時系列データを扱うのに優れており、自然言語の流れをより連続的に捉えることが可能

RNN型アーキテクチャ

RNNの特性として、入力の長さに依存しないため、非常に長い文脈も理論上処理可能

長い文書や会話の理解に有利であり、コンテキストが非常に重要な翻訳や要約タスクにおいて強みを発揮

高速な推論

RNNアーキテクチャのメリットとして、RNNは過去の情報を隠れ状態として保持

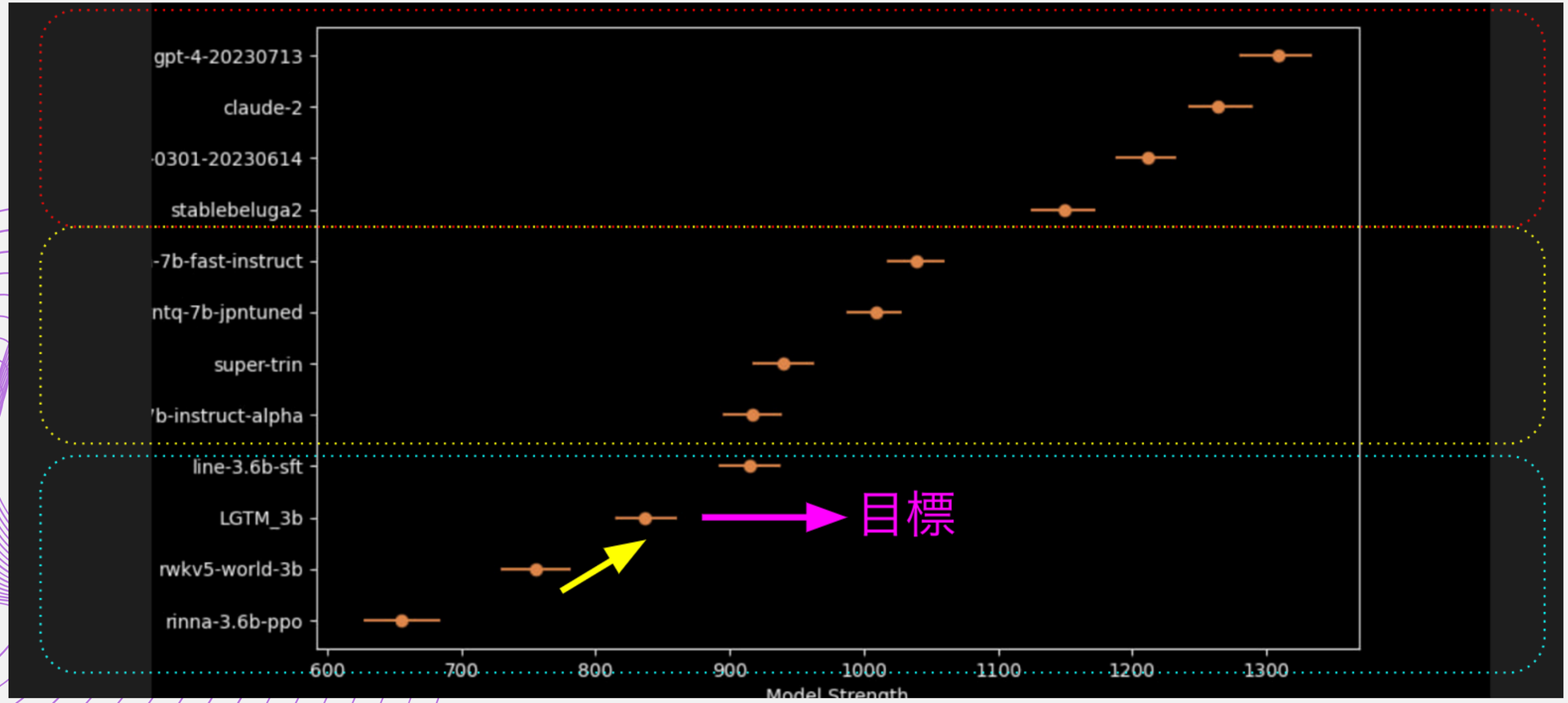
Transformer系と比べ再計算頻度を低減し、少ないGPUもしくは、CPUでの高速な推論が可能

日本語性能検証結果

LGTM 3bの日本語性能(RAKUDAベンチマーク)

検証日：2024年3月19日時点

Transformer系と
遜色がないレベルに到達



大規模
70b+領域

中規模
7b領域

小規模
3b領域

製品化までの流れ

ファインチューニング

1. 少量のターゲットデータで
   高度に最適化されたモデル調整
2. ユーザー固有の用途や業界に特化
3. 短時間でのデプロイメントが現実に

RAG

1. 関連情報の検索と内容生成が
   一体で行われる
2. データベースからの情報取得と
   自然で正確な文の生成

製品例:チャットボット

1. 自然な日本語で流暢に対話し、
   ユーザーからの質問に対して適切な回答
2. 幅広いトピックに対応した
   満足度の高い対話体験

技術資料（LGTM for Business）

システム稼働要件(推奨)

OS：Ubuntu 22.04 LTS
CPU：x86-64互換CPU
RAM：128GB以上
GPU：90TFlops(BF16)，VRAM 24GB以上
　　CUDA 12.3以上 もしくは ROCm 6.0以上
　　が動作するもの

配布形態

実行環境はDockerイメージで配布
（導入後は自動更新されます）
※ インターネットに接続できない環境への導入につきましてはお問い合わせください。

サーバ要件

・SSHでアクセスできる状態であること
・Webサーバーが動作し、クライアントとの送受信ができる環境であること

推論結果（テキスト生成）の利用方法

LGTM for Businessを導入したサーバから発行されるAPIエンドポイントとAPIキーをクライアント側で設定してください。
（詳細はオンラインダッシュボードからご確認ください。）

学習方法

オンラインダッシュボードの「LGTM Easy Trainer」から学習（ファインチューニング）が可能です。

技術資料（LGTMのモデル構造・特徴）

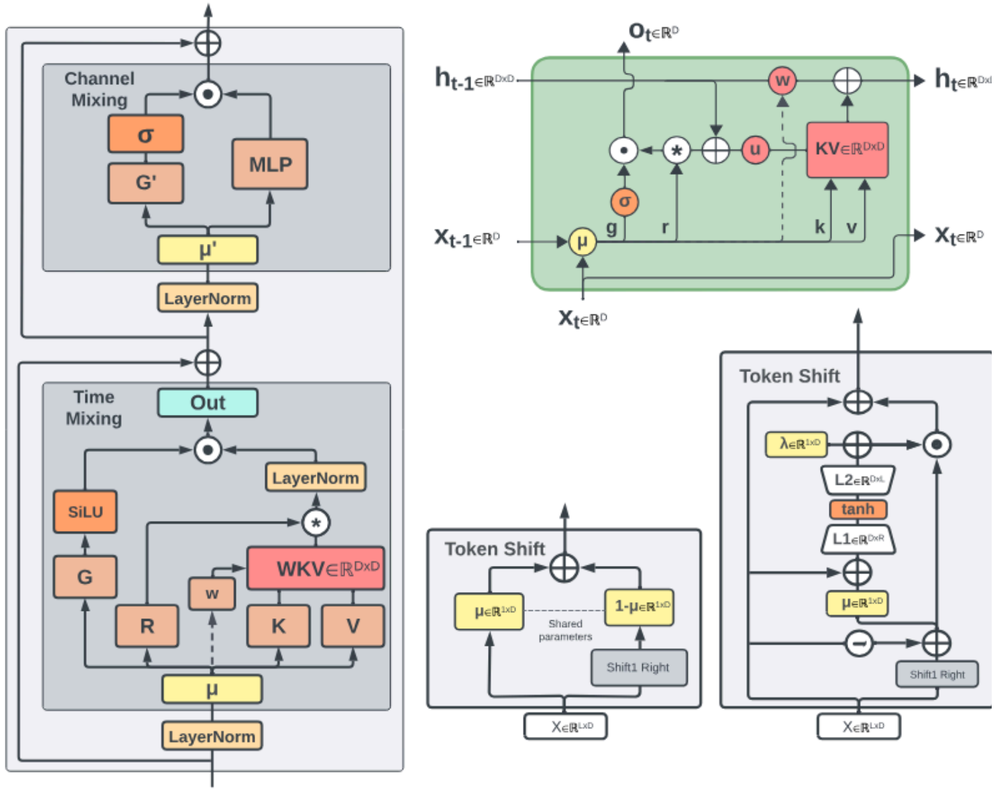
モデル構造

LGTMはRNN(回帰型ニューラルネットワーク)を使用したLLM(大規模言語モデル)です。

基本モデルアーキテクチャとして、RWKVという言語モデルを採用し、それを元に自社開発の学習手法とデータセットを用いてチューニングを行っています。

モデルの特徴

・実行およびトレーニング時の計算が少なく、大きなコンテキストサイズを持つTransformer系モデルと比べて、計算能力の要求が10分の1以下。
・計算負荷が任意のコンテキスト長に線形でスケールする。（Transformerは二乗にスケールする）
・ベースモデルが多言語でトレーニングされており、日本語だけでなく英語、中国語にも対応可。



図．モデル構造
Eagle and Finch: RWKV with Matrix-Valued States and Dynamic Recurrence (https://arxiv.org/abs/2404.05892)

技術資料（学習手法）

学習手法(Fine-Tuning)

「LGTM Easy Trainer」では State-Tuning という手法とRLHF（人間からのフィードバックによる強化学習）を組み合わせたものを採用しています。

その他の手法を使用してLGTMをファインチューニングする方法につきましてはお問い合わせください。

State-Tuning

State-Tuningとは、入力データセットに対して適切なpromptベクトルをバックプロパゲーションし、その時点のモデルのニューロン発火状態をニューラルネットワーク内の各層に埋め込む手法です。

RLHF

RLHFとは、機械学習モデル学習時に人間の価値基準を元に、趣向に沿う回答が出るように調整する強化学習の手法です。

本製品では弊社独自実装の学習手法によりこれを実現しています。

LGTM. Base Architecture by RWKV
Language Gateway to Tomorrow Module

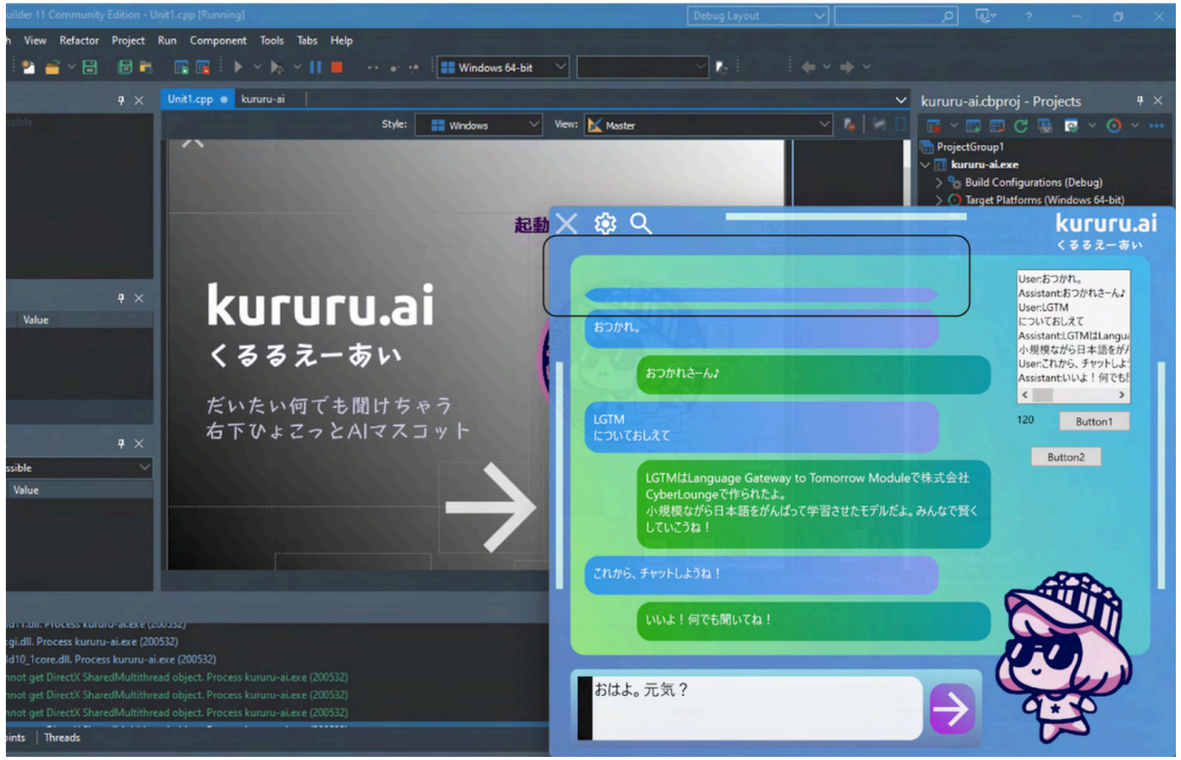
kururu.ai

AIデスクトップマスコット

LGTM. Base Architecture by RWKV
Language Gateway to Tomorrow Module

kururu.aiは、LGTMを使った無料AIチャットボットです。

ネット環境不要の Windowsアプリ



1. PCにインストールするだけオフラインで利用可能
2. 音声合成や天気予報などプラグインの追加が可能
3. プロンプトチューニングでキャラ設定ができる
4. クオリティを上げるならファインチューンで簡単学習

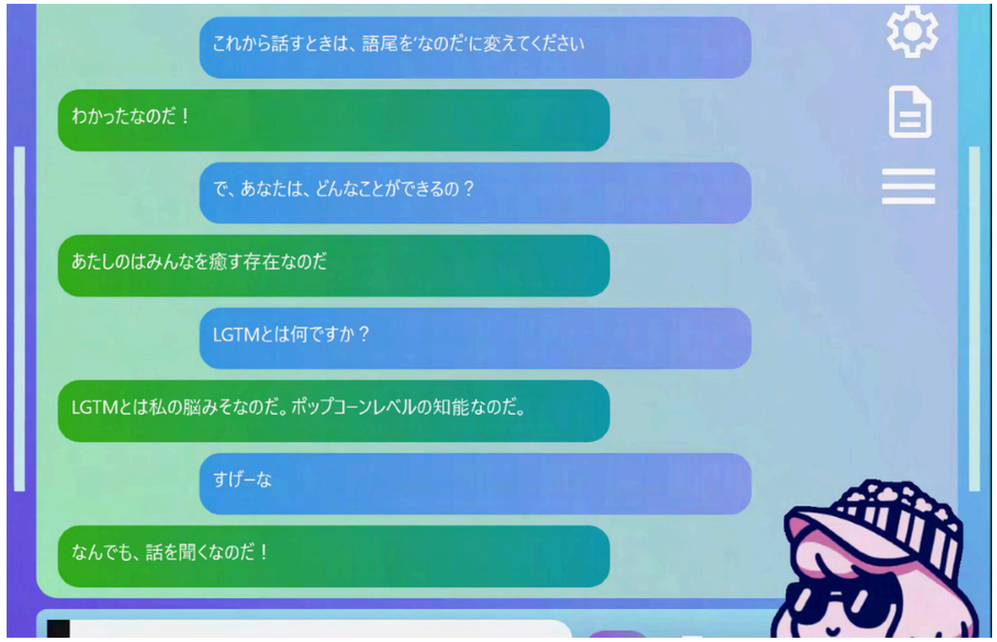
簡単にキャラ設定

初期設定でカスタマイズ

初期画面で設定するだけの簡単になりきり設定！

プロンプト(チャット)で指示

カスタマイズで足りないならチャットで直接指示してみよう！



LGTM. Base Architecture by RWKV
Language Gateway to Tomorrow Module

動作環境

- OS Windows 11
- CPU Intel Core i5 以上もしくは同等の互換CPU
- RAM 16GB以上
- SSD 20GB以上
- **GPU 不要（生成）**

PCのメモリで動くから
wi-fiががなくても動いちゃうぞ

アプリダウンロード後、初回起動時に
言語モデルファイルがダウンロードされます。

LGTM.

Base Architecture by RWKV

Language Gateway to Tomorrow Module

ロードマップ

LGTM. Base Architecture by RWKV
Language Gateway to Tomorrow Module

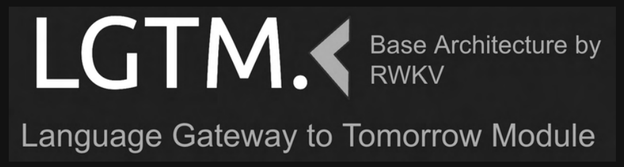
LGTM　ロードマップ

翻訳特化(7B)
2B開発(低スペックPCでも動く)
AIサービスフリマ
7B開発(ゲーミングPC)
議事録(14B)
0.7B(組込機器向け)
7B(ラズパイ)
14B開発(クラウド)

24年5月　24年8月　24年10月　25年2月　25年8月　25年9月　26年8月　27年12月

料金プラン

Ask to staff



LGTM.

Base Architecture by RWKV

Language Gateway to Tomorrow Module

皆さま、ご拝読
ありがとうございました

TEL：03-5795-0067
MAIL：info@cyberlounge.co.jp

Cyber Lounge