

目次

 HOME	1.1
 概要	1.2
 セットアップ	1.3
 v1.2.0リリースノート	1.4
 v1.0から1.1へのアップデートガイド	1.5
 主な操作方法	2.1
 ファイルビューの操作	2.1.1
 ファイルをインデックスに登録する	2.1.2
 コンフィグファイルを作成する	2.1.3
 AIを使う	3.1
 AIに質問する	3.1.1
 文書ファイルについて質問する	3.1.2
 画像ファイルについて質問する	3.1.3

カスタムLLMの設定



文書ファイルを要約する

3.1.4



文書ファイルを翻訳する

3.1.5



文書ファイルを校正する

3.1.6



画像を検索する

3.1.7



画像を生成する

3.1.8



議事録を取る

3.1.9



音声入力を利用する場合

3.1.10



便利機能

4.1



ローカルLLMの設定

4.1.1



ベンチマーク機能

4.1.2



さらに詳しく

5.1



セキュアな環境構築

5.1.1

カスタムLLMの設定

	カスタムLLMの設定	5.1.1.1
	LMスタジオを使用したカスタムLLM構築	5.1.1.2
	Ollamaを使用したカスタムLLM構築	5.1.1.3
	FastChatを使用したカスタムLLM構築	5.1.1.4
	弊社推奨LLM	5.1.1.5
	RAGによる回答精度の向上	5.1.2
	エンベディングについて	5.1.3
	Azure OpenAI Service	5.1.4
	Gemini	5.1.5
	OpenAIについて	6.1
	APIキーの取得	6.1.1
	料金体系	6.1.2
	外部リンク	7.1

カスタムLLMの設定



評価用アプリDL

7.1.1



説明動画（Youtube）

7.1.2



Langage

8.1



大規模言語モデル(LLM)の導入による業務効率改善

ChatGPTを代表とする大規模言語モデル(LLM)の登場により、オフィスワークをAIで効率化する動きが加速しています。大規模言語モデルの業務活用が企業の競争力向上にも重視され、AIはDXの中心になりつつあります。「ailia DX Insight」は、大規模言語モデルをオフィスワークに導入させ、さまざまな業務を多角的にサポートします。

- 2024/12/18

アップデート

ailia DX Insight 1.2がリリースされました。アップデートに関する内容はこちらをご確認ください。

- 2024/12/18

アップデートガイド

ailia DX Insight 1.1から1.2へのアップデート方法はこちらをご確認ください。

- 2024/8/06

アップデートガイド

ailia DX Insight 1.0から1.1へのアップデートがまだの方はこちらからアップデートガイドをご確認ください。

概要

[セットアップ](#)

[主な操作方法](#)

[AIを使う](#)

[次のページへ >](#)

概要

aillia DX Insightは、ax株式会社および株式会社アクセルが提供する、企業のDXのためのAIスーパーAPIです。WindowsやmacOSにインストールするだけで誰でもAIを業務に活用可能です。大規模言語モデルに文章検索や画像検索、音声認識、企業内情報検索などaillia独自のAIをプラスすることで、AIを使った業務効率の改善が簡単に行えます。



機能

aillia DX insightでは下記の機能を使用可能です。各機能で使用されるモデルと実行環境は下記となります。

機能	クラウドで動作	端末上で!
文章検索 (RAG)	ChatGPT、Azure OpenAI、Gemini、Claude(* 準備中) (回答生成)	aillia SDK (インデックス作成、ベクトル化) カスタムLLM (回答生成)
画像生成	DALLE	StableDiffusion(* 準備中)
要約・アイデア出し	ChatGPT、Azure OpenAI、Gemini、Claude(* 準備中) (回答生成)	カスタムLLM
翻訳	ChatGPT、Azure OpenAI、Gemini、Claude(* 準備中) (回答生成)	aillia SDK、カスタムLLM
画像検索	-	aillia SDK
音声入力	-	aillia SDK
音声認識	Whisper Large V3 Turbo	aillia SDK
議事録	Whisper Large V3 Turbo	aillia SDK
OCR	Azure AI Vision(* 準備中)	aillia SDK

対応ファイルフォーマット

カスタムLLMの設定

ailia DX Insightがサポートしている形式は下記となります。

一般的なオフィス系ドキュメントフォーマット

```
pdf, doc, docx, txt, md, xlsx, csv, tex, css, html
```

プログラム系ドキュメントフォーマット

```
c, cpp, json, java, py, rb, php
```

js, swift, .kt, rust, dart, lua は設定 /インデックスから追加することができます。



画像フォーマット

```
'png', 'jpg', 'jpeg'
```

音声フォーマット

```
'wav'
```

ファイルの保存先

ailia DX Insightは下記のフォルダにデータベースやモデルファイルを保存します。このフォルダを削除することで、ailia DX Insightを初期化することができます。

```
Windows : c:/Users/[UserName]/Documents(*)/ailia DX Insight
```

```
macOS : ~/Documents(*)/ailia DX Insight
```

(*)Windowsでは「マイドキュメント」、Macでは「書類」と表示されることもあります。

[次のページへ >](#)

セットアップ

[ailia DX Insight 1.0から1.1へのアップデートがまだの方はこちらをご確認ください。](#)

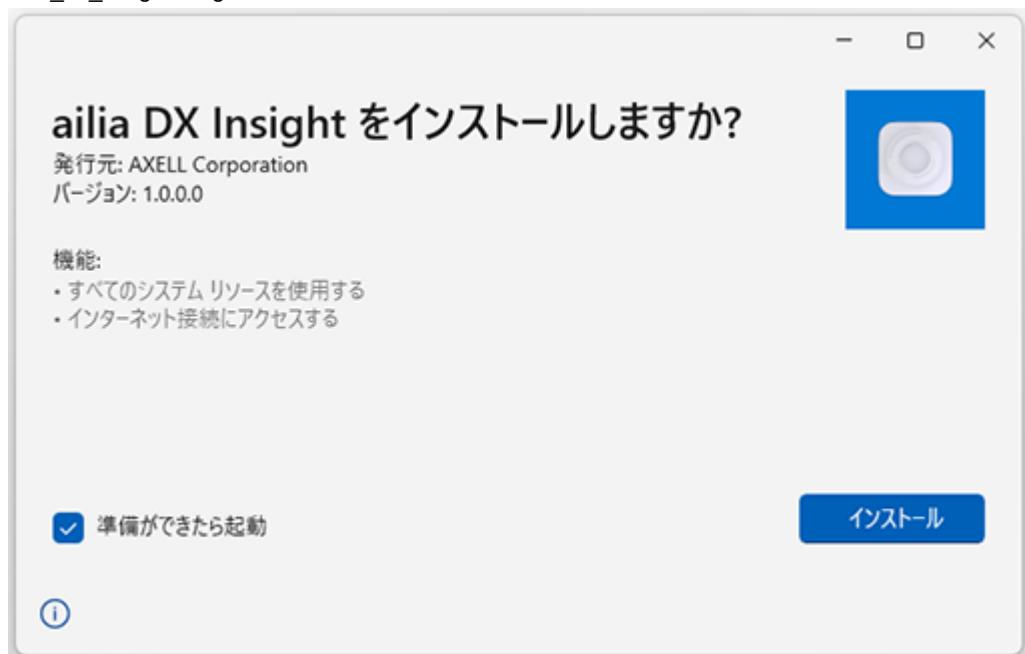
[ailia DX Insight 1.2のアップデート内容はこちらをご確認ください](#)

ダウンロード

ailia DX Insightを[ダウンロード](#)して、zipを解凍します。

インストーラの起動

Windowsの場合は、aillia_dx_insight.msixをダブルクリックしてインストーラを起動します。macOSの場合は、aillia_dx_insight.dmgを右クリックして開くで起動します。



ライセンスもしくはコンフィグファイルの選択

初回起動時に「ライセンスもしくはコンフィグファイルを選択してください」とウィンドウに表示されます。ライセンスファイルをフォルダから選択します。

カスタムLLMの設定

起動に必要なライセンスファイルはmacOSの場合は[HOME]/Library/SHALO、windowsの場合は[ROOT]/ProgramData/SHALOに格納されます。



コンフィグファイルを選択することでアプリケーションをカスタマイズすることができます。

OpenAIのAPIキーの取得

OpenAIのWEBページでアカウント登録を行い、[OpenAIのAPIキーを取得します](#)。OpenAIのAPIキーを入力しなくてもilia DX Insightは使用できますが、機能が制限されます。

セットアップ

カスタムLLMの設定

ライセンスもしくはコンフィグファイルを選択後、チュートリアル画面が表示されます。



チュートリアルの最終画面にて、OpenAI APIキーの設定が可能です。

OpenAI APIキーの設定

チュートリアルから設定する場合

1. チュートリアルの最後、「APIキーの設定」まで進みます。

カスタムLLMの設定

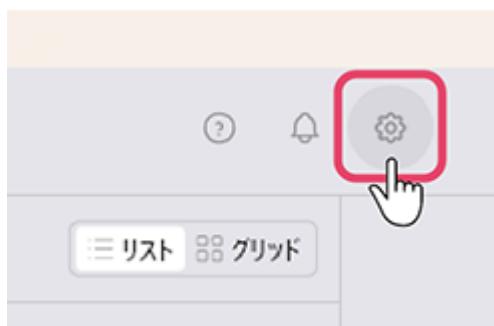
2. 「OpenAI APIキー」の下にあるテキストボックスに、"sk-"から始まるOpenAI APIキーを入力します。



3. 「完了」ボタンを押します。

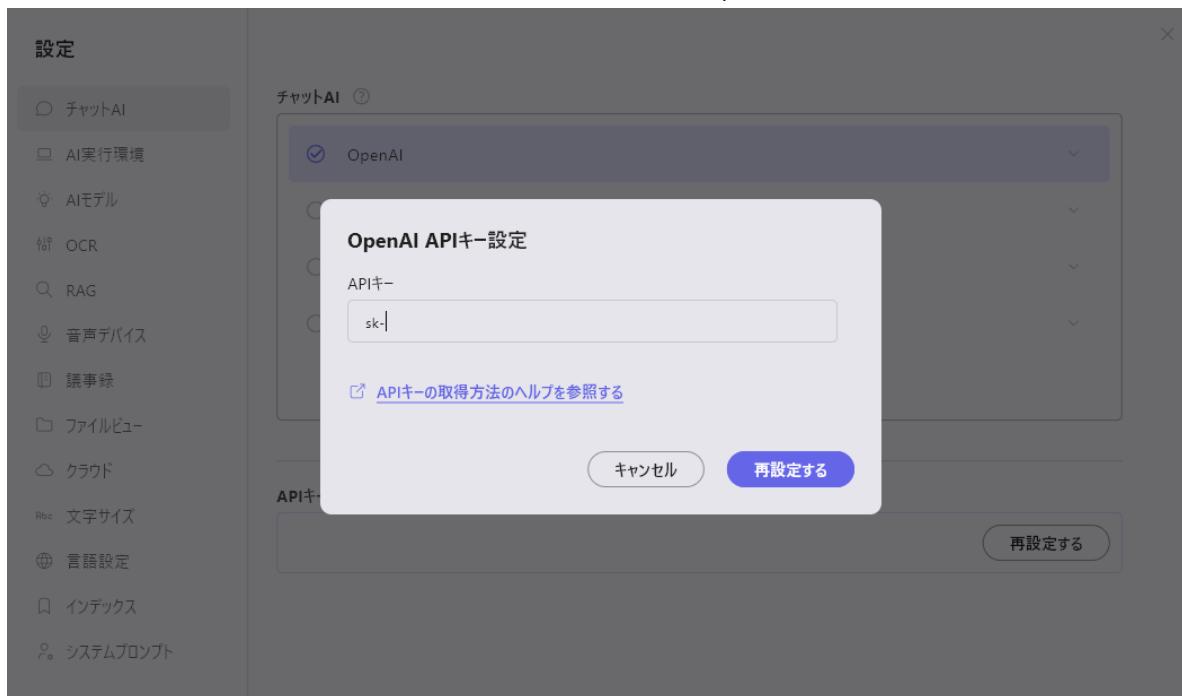
通常画面から設定する場合

4. 画面右上にある歯車アイコンを押します。



カスタムLLMの設定

- 表示されたダイアログ内のテキストボックスに、"sk-"から始まるOpenAIのAPIキーを入力します。



- 「OpenAI」をクリックして、使用するchatGPTのバージョンを選択し、「閉じる」ボタンを押します。

AIモデルのダウンロード

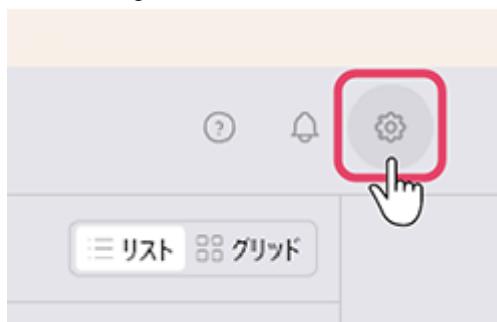
初回起動時のチュートリアルの終了後、AIモデルのダウンロードが開始されます。ダウンロードの進捗は左のサイドバー下部にて確認できます。



ダウンロード完了後、ailia DX Insightが使用可能となります。

OpenAIモデルの選択

- ailia DX insightの画面にて、右上の歯車アイコンをクリックして設定ウィンドウを表示させます。



カスタムLLMの設定

2. 「チャットAI」の項目の中にある「OpenAI」をクリックし、モデルを選択します。



GPTo3-miniの追加方法

カスタムLLMの設定

- 「チャットAI」の項目の中にある「OpenAI」をクリックし、「+追加」を選択します。



- 表示されるウィンドウ内で以下のように入力することでChatGPT o3-miniを使用することができます。



- モデル：o3-mini
- 最大トーケン長：200000
- 説明：必要に応じてメモとして使用

カスタムLLMの設定

OpenAIがAPIでアクセスできるようにしているモデルであれば、同様の方法でチャットAIに追加することで使用できます。

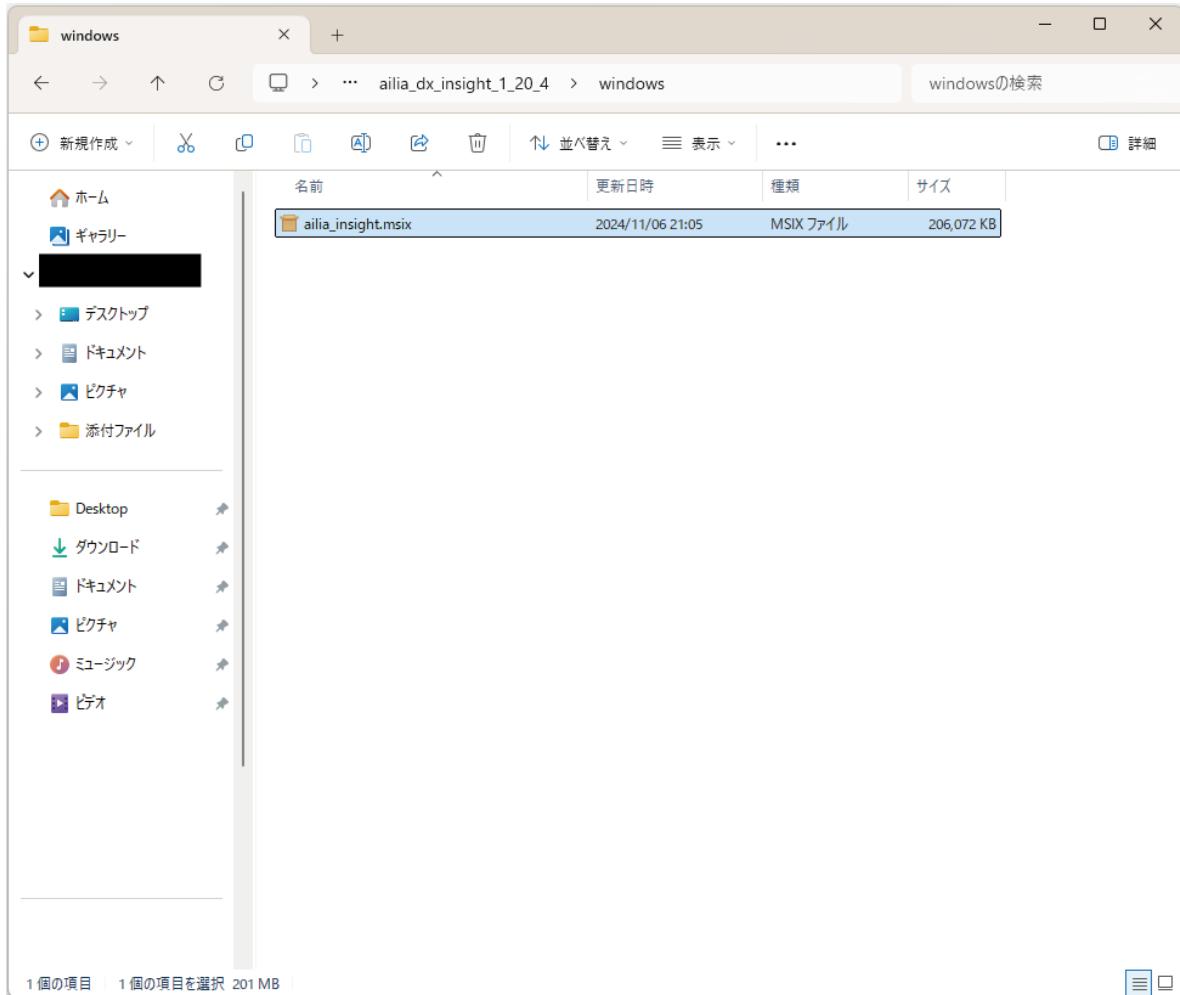
[次のページへ >](#)

バージョン1.2.0リリースノート

ailia DX Insight1.1から1.2へのアップデートガイド

ailia DX Insight1.1と1.2は電子証明書が同一になっています。

1. Windowsの場合はailia DX Insight1.1から1.2へアップデートは「ailia_insight.msix」をダブルクリックします。



カスタムLLMの設定

2. 「ailia DX Insightを更新しますか？」というウィンドウが表示されますので、「更新」を選択してください。



3. ailia DX Insight1.2がインストールされた後、アプリケーションが起動されます。

機能追加

チャット機能

- ローカルLLMに対応
- マークダウンの表示に対応
- コードをコピーするボタンを追加
- 送信時にモデル選択する機能を追加

ファイルビュー

- ファイルビューにクイックアクセス用のボタンを追加
- 初期ディレクトリ設定に対応

RAG

- 文章全体を渡してチャットするモードを追加 (DOCモード)
- 複数のファイルを選択してDOCモードを使用する機能を追加

議事録

- Whisper Large V3 Turboに対応

その他

- AI実行環境のベンチマーク機能の追加

OpenAIのモデル変更

カスタムLLMの設定

- GPT3.5の削除
- GPT-4o mini対応

UI

- 設定画面のコンフィグファイルの設定にライセンスファイルの有効期限を表示
- 校正と翻訳で変更前と変更後を表示
- 画像編集で画像全体を選択するボタンを追加

改善

- GPUを使用した場合のドキュメント登録の高速化
- エクセルのインポートにおいて、末尾の空白を削除
- 1行の中でテキストが分割されたPDFの翻訳の精度改善

不具合修正

- 翻訳のコマンドプロンプトを編集しようとすると校正のプロンプトという表示がある問題を修正
- チャット欄に大量の文章を入れるとオーバフローする問題を修正
- チャットヒストリーの左右の幅を広げにくい問題を修正
- Geminiでシステムプロンプトが適用されない問題を修正
- libllama.dylibがmacOS14向けにビルドされているためmacOS13で起動できない問題を修正

[次のページへ >](#)

ailia DX Insight 1.0から1.1へのアップデートガイド

ailia DX Insight 1.1では、アプリケーションの電子証明書が更新されています。Microsoftのインストーラであるmsixでは、電子証明書が更新された場合、別のアプリケーションとして認識されます。

(参考情報：<https://github.com/microsoft/msix-packaging/issues/365>)

そのため、ailia DX Insight 1.0から1.1にアップデートする場合、下記の手順に従い、旧バージョンをアンインストール後にインストールを行なってください。

windowsの場合

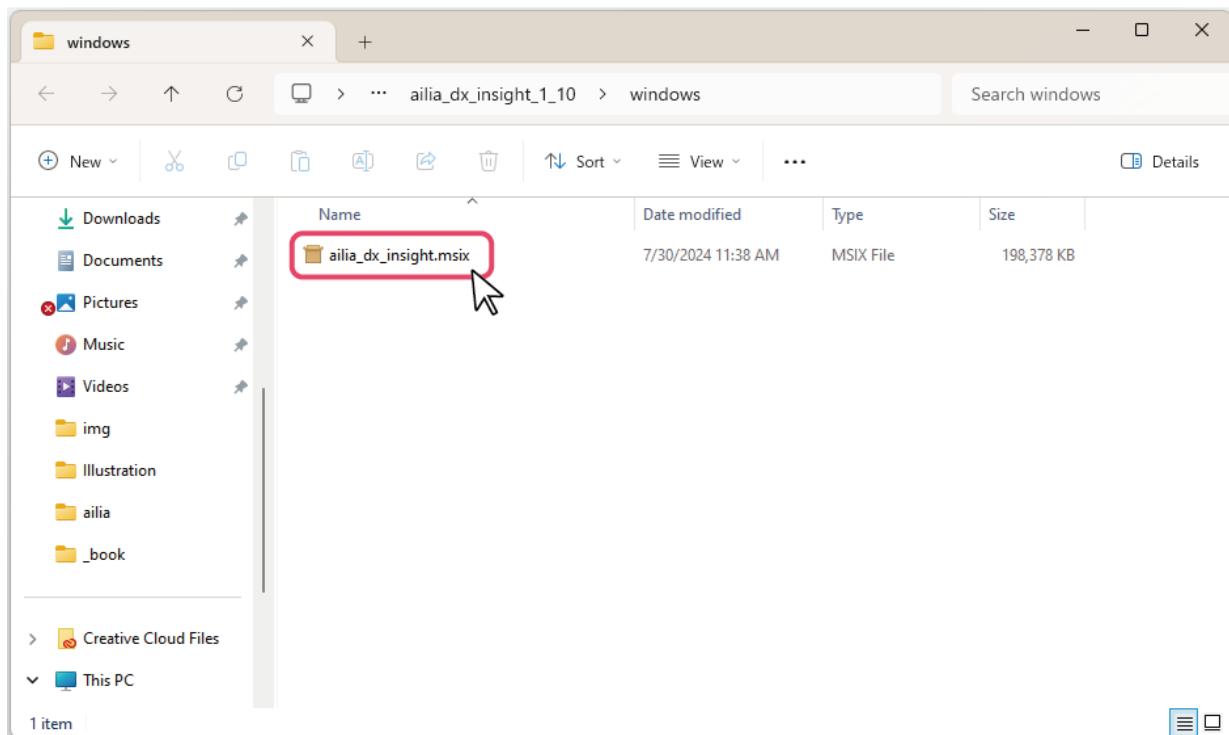
スタートメニューからailia_dx_insightを選択し、右クリックでアンインストールを行います。

(スタートメニューの中に見当たらない場合は「すべてのアプリ」の中からお探し下さい)



カスタムLLMの設定

ailia_dx_insight.msixをダブルクリックし、インストールを行います。



ユーザデータは c:/Users/[UserName]/Documents(*)/ailia DX Insight に保存されているため、保持されます。

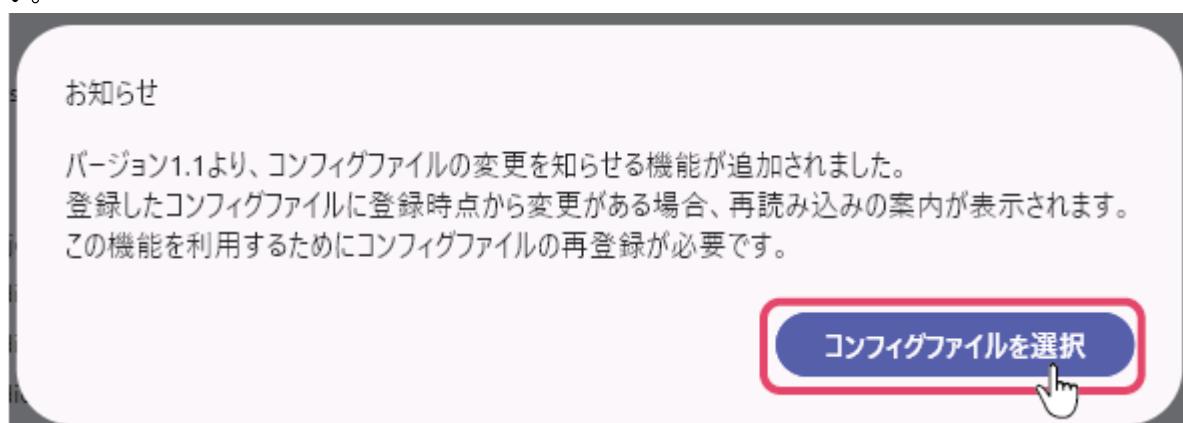
(*) 「マイドキュメント」と表示されることもあります。

Macの場合

ailia_dx_insight.dmgを開き、ailia_dx_insight.appをApplicationフォルダにドロップすることで上書きします。

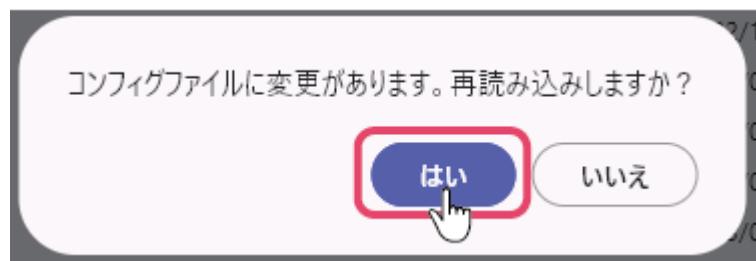
コンフィグファイルの更新

ailia DX Insight 1.0でコンフィグファイルを使用している場合、コンフィグファイルの更新が必要です。ailia DX Insight 1.1の初回起動時に、下記のダイアログが表示されるため、コンフィグファイルのパスを設定してください。



コンフィグファイルの登録時以降にファイルに変更があった場合、リロードするか確認する下記のダイアログが表示されます。「はい」を選択するとコンフィグファイルがリロードされ、変更が適用されます。

カスタムLLMの設定



また、設定画面→コンフィグファイルより、手動でのコンフィグファイルの登録が可能です。
同画面にてコンフィグファイルの登録解除も可能です。

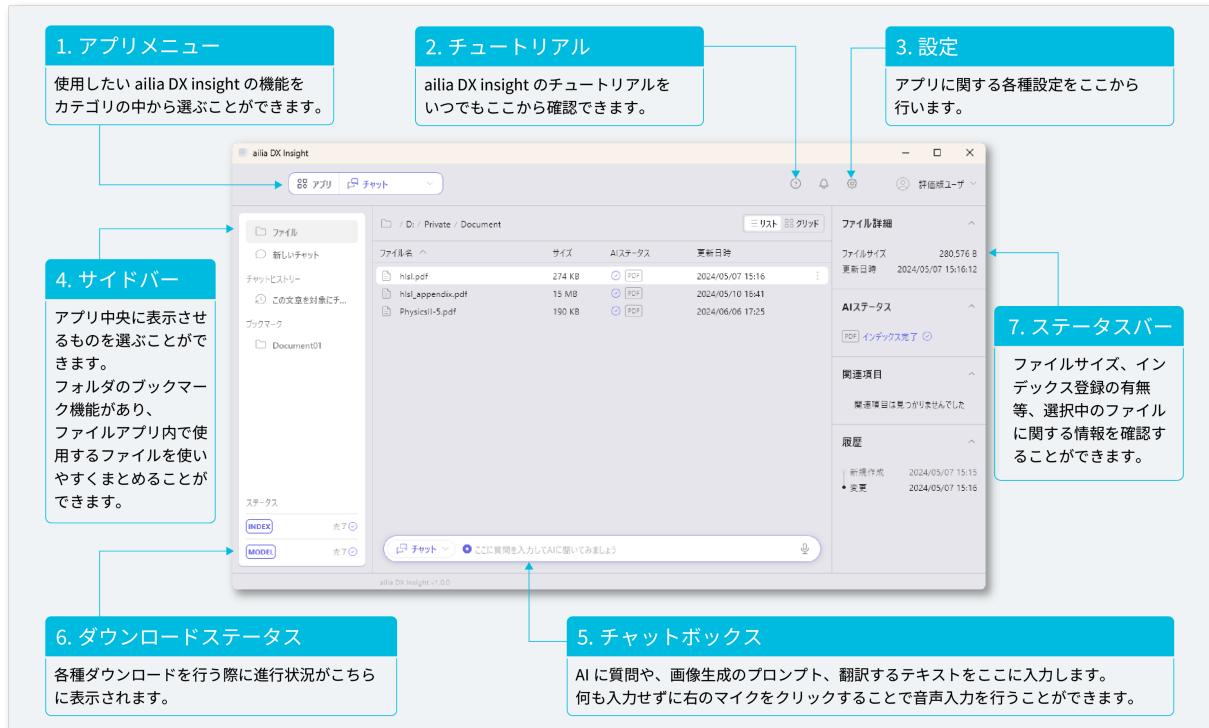


[次のページへ >](#)

主な操作方法

画面構成

aillia DX insightの主な画面構成



1. アプリメニュー

使用したいaillia DX insightの機能をカテゴリの中から選ぶことができます。

2. チュートリアル

aillia DX insightのチュートリアルをいつでもここから確認できます。

3. 設定

アプリに関する各種設定をここから行います。

4. サイドバー

アプリ中央に表示させるものを選ぶことができます。フォルダのブックマーク機能があり、ファイルアプリ内で使用するファイルを使いやすくまとめることができます。

5. チャットボックス

カスタムLLMの設定

AIに質問や、画像生成のプロンプト、翻訳するテキストをここに入力します。何も入力せずに右のマイクをクリックすることで音声入力を行うことができます。

6. ダウンロードステータス

各種ダウンロードを行う際に進行状況がこちらに表示されます。

7. ステータスバー

ファイルサイズ、インデックス登録の有無等、選択中のファイルに関する情報を確認することができます。

[次のページへ >](#)

ファイルビューの操作

ファイルの並び順を変える

ファイルビュー上部に並んでいる項目名をクリックすると、ファイルの並び順が変わります。



ファイル名	サイズ	AIステータス	更新日時
ailiaDXInsight_manual.pdf	5 MB	PDF	2024/07/10 19:56
ailiaDXInsight_manual_copy.pdf	2 MB	PDF	2024/07/18 20:22
DietAndHealth.pdf	8 MB	PDF	2024/07/29 18:12
DietAndHealth_ja.pdf	8 MB	PDF	2024/08/09 17:59
doc.txt	2 KB		2024/10/17 10:12

- ファイル名で並び替え
「ファイル名」をクリックすると、名前順(A-Z/Z-A)に並びます。
- ファイルサイズで並び替え
「サイズ」をクリックすると、サイズが大きい/小さい順に並びます。
- インデックス登録の有無で並び替え
「AIステータス」をクリックすると、インデックス登録済/未登録で並びます。
- 更新日時で並び替え
「更新日時」をクリックすると、ファイルの更新日時の新しい/古い順に並びます。

ディレクトリの移動

ファイルビューの項目の左上の記号から、ディレクトリ間を移動することができます。

また、カレントディレクトリの表示から、上の階層に直接移動することもできます。

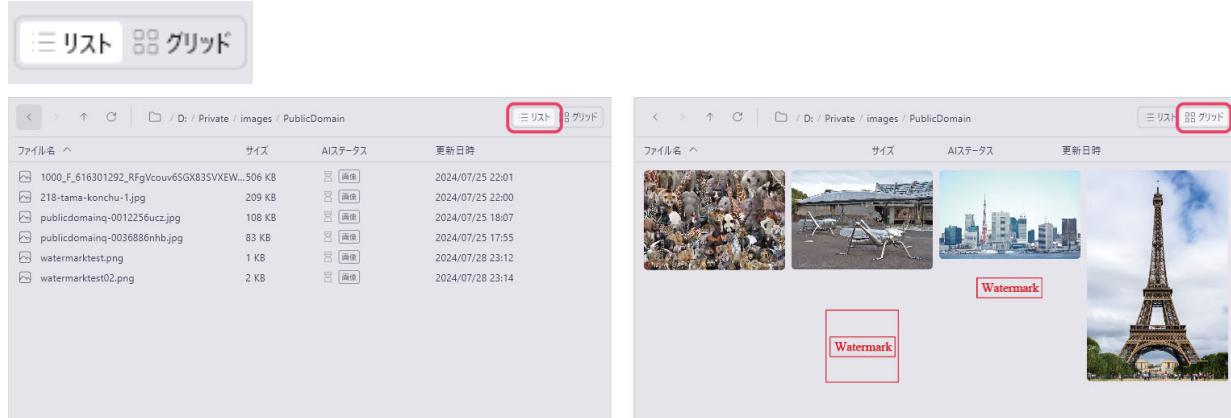


ファイル名	サイズ	AIステータス	更新日時
ailiaDXInsight_manual.pdf	5 MB	PDF	2024/07/10 19:56
ailiaDXInsight_manual_copy.pdf	2 MB	PDF	2024/07/18 20:22
DietAndHealth.pdf	8 MB	PDF	2024/07/29 18:12
DietAndHealth_ja.pdf	8 MB	PDF	2024/08/09 17:59
doc.txt	2 KB		2024/10/17 10:12

ファイルの表示方法を変える

カスタムLLMの設定

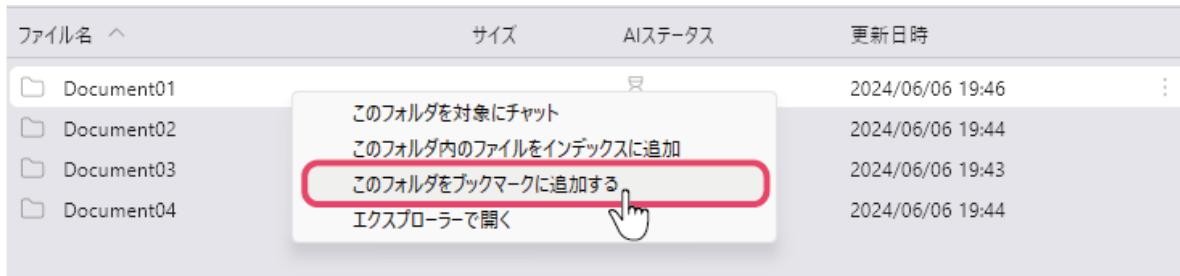
ファイルビューの右上にある表示モードボタンをクリックすると、ファイルの表示方法（リスト表示/グリッド表示）を変えることができます。グリッド表示では、画像ファイルのサムネイルがグリッド状に並びます。



フォルダをブックマークに追加する

頻繁に開くフォルダをブックマークに追加することができます。

1. ブックマークに追加したいフォルダにカーソルを合わせ右クリックし、「このフォルダをブックマークに追加する」を選択します。



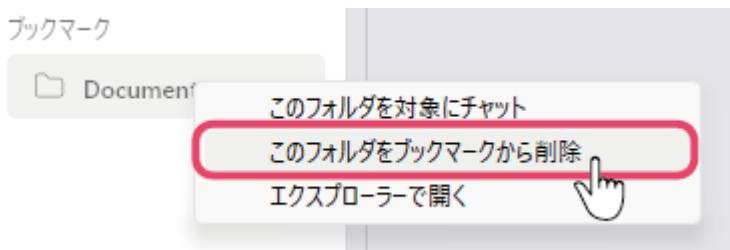
2. 左のサイドバーにあるブックマークにフォルダが追加されます。



フォルダをブックマークから削除する

この操作ではフォルダ本体の削除は行われません。

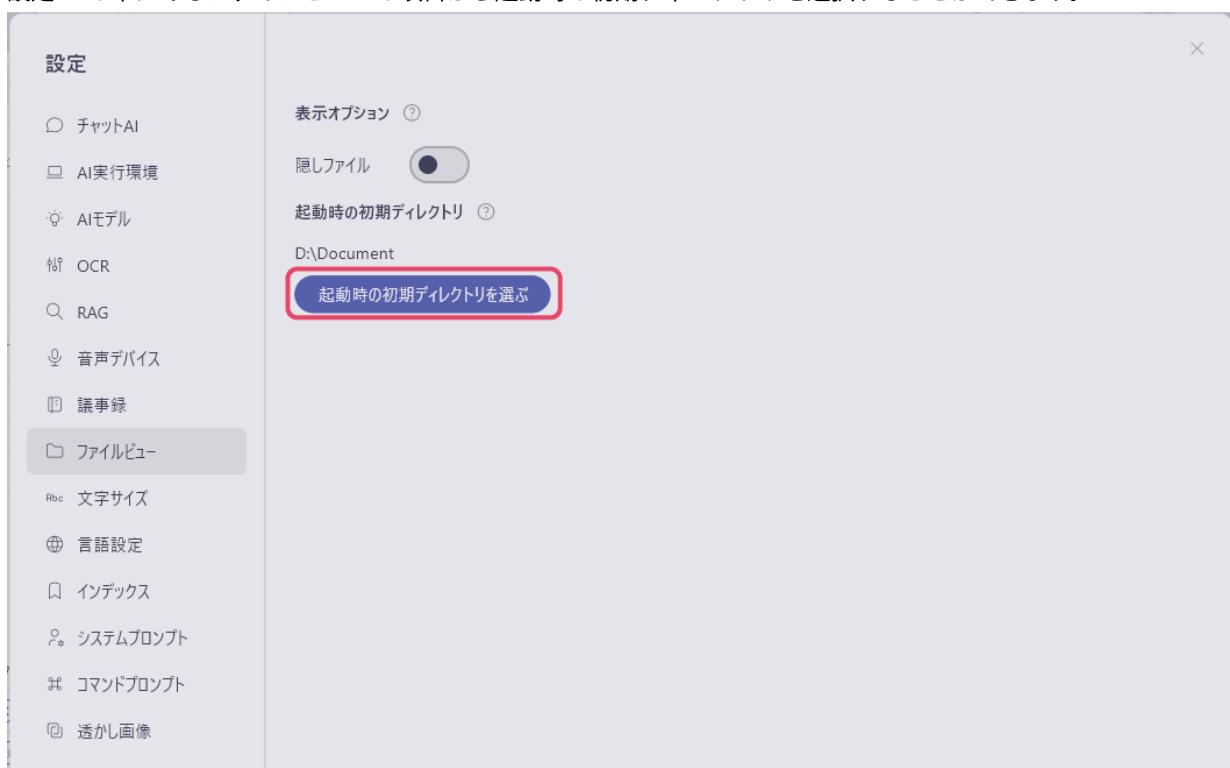
- 左サイドバーにあるブックマーク内、削除したいフォルダを右クリックし、「このフォルダをブックマークから削除」を選択します



- ブックマークからフォルダが削除されます。

初期ディレクトリを設定する

設定 の中にあるファイルビューの項目から起動時の初期ディレクトリを選択することができます。



[次のページへ >](#)

ファイルをインデックスに登録する

ファイルをインデックス登録することでAI機能を活用した各種機能をそのファイルに対して実行することが出来ます。

[文書ファイルについて質問する](#)

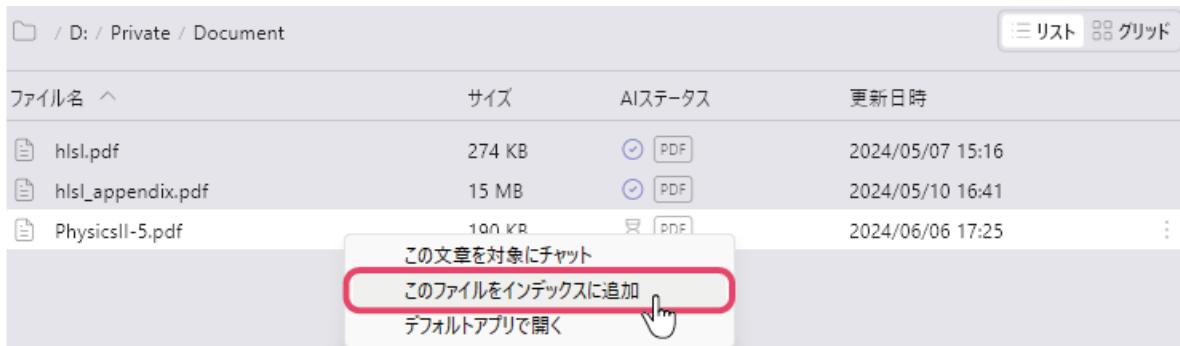
[文書ファイルを要約する](#)

[文書ファイルを翻訳する](#)

[画像を検索する](#)

ファイルをインデックス登録する

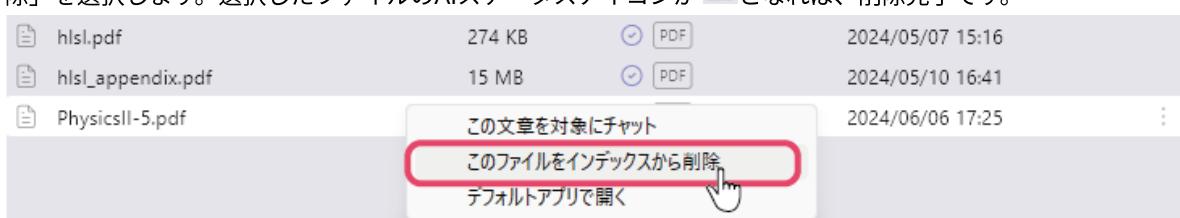
- 登録したいファイルを右クリックし、「このファイルをインデックスに追加」を選択します。選択したファイルのAIステータスアイコンが となれば、登録完了です。



- フォルダを右クリックし、「このフォルダ内のファイルをインデックスに追加」を選択すると、フォルダ内の全てのファイルをインデックス登録することができます。

ファイルをインデックスから削除する

- インデックスから削除したい場合、登録済のファイルを右クリックし、「このファイルをインデックスから削除」を選択します。選択したファイルのAIステータスアイコンが となれば、削除完了です。



[次のページへ >](#)

コンフィグファイルを作成する

コンフィグファイルを作成することによって利用規約の画像をアプリケーション起動時に表示させたり、企業ロゴを常時表示させることができます。

コンフィグファイルの作成例

コンフィグのファイルフォーマットはjsonで作成してください。

```
{  
  "apiKey" : {  
    "openAI": "YourOpenAIApiKey"  
  },  
  "licenseFile": "/path/to/license/file",  
  "terms": {  
    "image": "/path/to/terms/image"  
  },  
  "logo": "path/to/logo/image"  
}
```

"apiKey"

OpenAIのAPIキーをこちらに入力します。

"licenseFile"

ライセンスファイルへのパスをコンフィグファイルからの相対パスで指定します。

"terms" "logo"

画像ファイルへのパスをコンフィグファイルからの相対パスで指定します。

ライセンスファイル以外の項目は省略してもアプリの起動には影響はありません。

作成したコンフィグファイルは、社内の共有フォルダなどに配置してください。

設定や画像ファイルはローカルにコピーされるため、初期設定以外では共有フォルダへのアクセスは不要です。

コンフィグファイルの適用/ライセンスの適用

カスタムLLMの設定

ailia DX Insightの初回起動時に「ライセンスファイルもしくはコンフィグファイルを選択してください」というウインドウが表示されます。フォルダを指定して適用させてください。



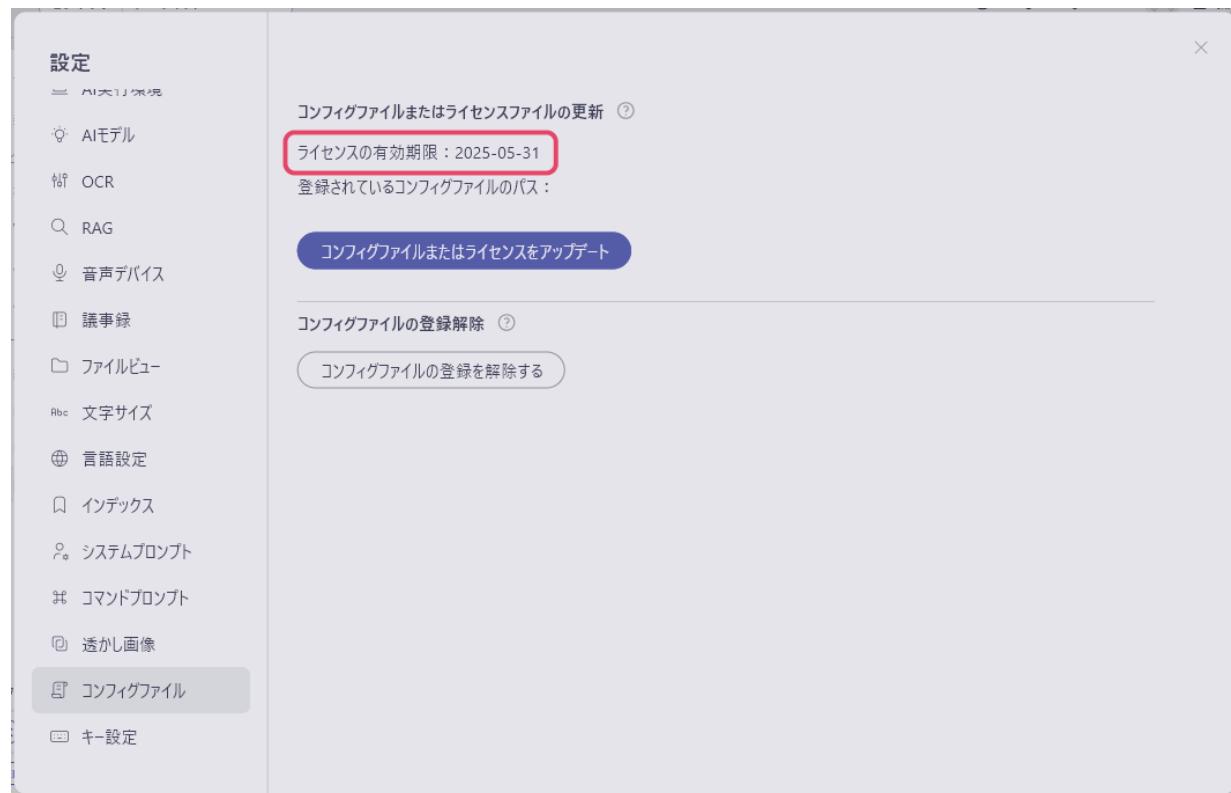
または、設定画面 の「コンフィグファイル」項目内にてライセンスの登録/更新をすることができます。
また、同項目内にて、コンフィグファイルの登録を解除することができます。



ライセンスファイルの有効期限の確認

カスタムLLMの設定

設定画面の「コンフィグファイル」項目内にてライセンスの有効期限を確認することができます。



次のページへ >

AIを使う

aillia DX insightで出来ること

[AIに質問する](#)

[文書ファイルについて質問する](#)

[画像ファイルについて質問する](#)

[文書ファイルを要約する](#)

[文書ファイルを翻訳する](#)

[文書ファイルを校正する](#)

[画像を検索する](#)

[画像を生成する](#)

[議事録を取る](#)

[次のページへ >](#)

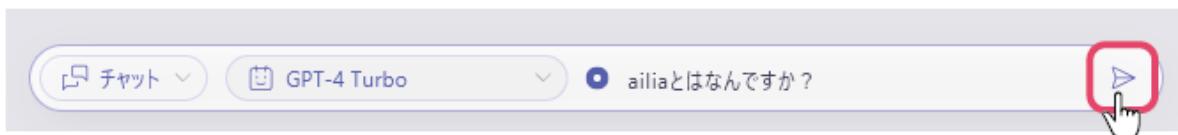
AIに質問する

AIにチャットで質問する

- 画面左上のアプリメニューを開き、一覧の中から「チャット」を選択します。（起動時のデフォルトではチャットの状態になっています）



- 画面中央下のチャットボックスに質問を入力し、送信ボタンをクリック（またはWindowsの場合はCtrl + Enter、macOSの場合はCmd + Enter）します。



- 送信前にチャットボックスの左のLLM名から、使用するLLMを選択・変更することができます。



- AIからの回答が表示されます。

- 音声入力を利用する場合

コードのコピー

チャットでAIにコードを生成させた場合、AIからの回答の中のコードに該当する部分が小窓で表示されます。小窓内の右上のコピーボタンをクリックすることで、対象のコードだけをコピーすることができます。

カスタムLLMの設定

- ・ サーバー側言語 (例 : Node.js, Python, Ruby)

HTML

HTMLは画面レイアウトのために使用します。`canvas`要素を使用してゲーム画面を描画します。

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <title>テトリスゲーム</title>
  <link rel="stylesheet" type="text/css" href="styles.css">
</head>
<body>
  <canvas id="tetris" width="300" height="600"></canvas>
  <script src="tetris.js"></script>
</body>
</html>
```



CSS

CSSを利用して、キャンバスのスタイルを定義します。

```
canvas {
  border: 4px solid #333;
  background-color: #111;
}
```



マークダウン形式で生成した場合も同様に小窓で表示され、右上のコピーボタンから対象のマークダウン記述のみをコピーすることができます。

● OpenAI GPT-4 Turboが回答を作成しました

下記は、指定されたコンテキストをマークダウン形式に変換したものです。



対応ファイルフォーマット
ailia DX Insightがサポートしている形式は下記となります。

一般的なオフィス系ドキュメントフォーマット

- pdf
- doc
- docx
- txt
- md
- xlsx
- csv
- tex
- css
- html

プログラム系ドキュメントフォーマット

- c
- cpp
- json
- java
- py
- rb
- php
- js
- swift
- .kt
- rust
- dart
- lua

注: js, swift, .kt, rust, dart, lua は設定/インデックスから追加することができます。

画像フォーマット

- 'png'

[次のページへ >](#)

[一覧に戻る](#)

文書ファイルについて質問する

文書ファイルをインデックスに登録すると、登録した文書に記載された情報についてAIに質問することができます。

ファイルをインデックス登録する方法については[こちら](#)をご確認ください。

文書ファイル内の情報について質問する

1. 画面左上のアプリメニューから「チャット」を選択します。



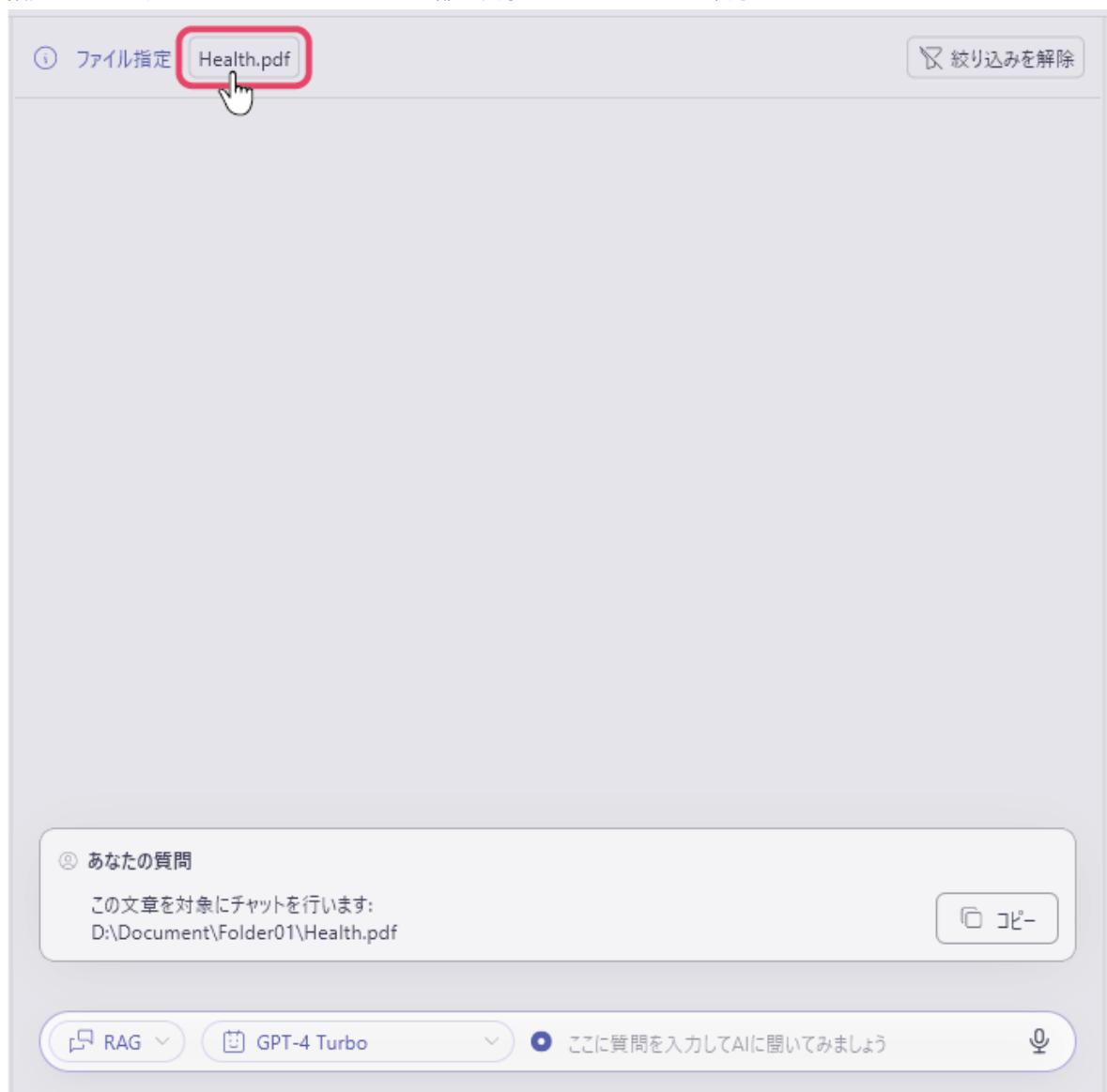
2. インデックス登録した質問したい文書ファイルを右クリックし、「この文章を対象にチャット」を選択します。

The screenshot shows a file list in a grid view. A context menu is open over a PDF file named 'Health.pdf'. The menu items are: 'この文章を対象にチャット (RAG)' (highlighted with a red box), 'この文書を質問 (DOC)', 'このファイルをインデックスから削除', and 'デフォルトアプリで開く'. The table below lists the files in the folder:

ファイル名	サイズ	AIステータス	更新日時
ailiaDXInsight_manual.pdf	5 MB	PDF	2024/07/10 19:56
ailiaDXInsight_manual_copy.pdf	2 MB	PDF	2024/07/18 20:22
DietAndHealth.pdf	8 MB	PDF	2024/07/29 18:12
DietAndHealth_ja.pdf	8 MB	PDF	2024/08/09 17:59
doc.txt	2 KB		2024/10/17 10:12
file.txt	960 B	✓	2024/10/24 17:13
Health.pdf	5.1 MB	✓ PDF	2024/07/29 19:00
hsls.pdf	この文書を質問 (DOC)	✓ PDF	2024/05/07 15:16
hsls_ap	このファイルをインデックスから削除	✓ PDF	2024/05/10 16:41
overview	デフォルトアプリで開く		2024/07/25 23:03
PhysicsII-5.pdf	190 KB	PDF	2024/06/06 17:25

カスタムLLMの設定

3. 指定されたファイル、またはフォルダが上部に表示されていることを確認します。



4. チャットボックスの左側の表示が「RAG」となるので、質問を入力し、送信します。



カスタムLLMの設定

5. AIが関連するファイルを調べた回答が表示されます。（検索はローカルAIで行われます）

① ファイル指定 Health.pdf

□ 紋り込みを解除

もし、食事と健康の関連に関する関心も増加しています。適切な食事は、単に病気の予防だけでなく、生活の質を向上させ、心身のウェルネスを促進する重要な要素です。本論文では、栄養学的視点から、食事が人間の健康に与える影響を多角的に分析します。具体的には、バランスの取れた食事がどのようにして体と心に良い影響を与えるか、そして不適切な食生活が引き起こす潜在的リスクについて探求します。2. 栄養素の基本とその機能 人間の体は、生きていく上で必要なエネルギーと栄養素を食事から取り入

p.1

● OpenAI GPT-4 Turboが回答を作成しました

この文章の大半の部分は、食事と健康についての複数の重要な側面を説明している各章の要点です。具体的には、以下のポイントが重要です：

- 「序論」で述べられた、適切な食事が人間のウェルネスを促進する重要な要素である点。
- 「栄養素の基本とその機能」章で詳細に説明されている、主要な栄養素（炭水化物、タンパク質、脂質、ビタミン・ミネラル）の役割。
- 「適切な食事の健康効果」章において、健康的な食生活が心血管疾患の予防、体重管理と糖尿病予防、精神健康の支持にどのように寄与するかの説明。
- 「食生活と公衆衛生」章では、不適切な食生活が公衆衛生に及ぼす悪影響と、健康的な食生活を推進するための公共キャンペーンの重要性が強調されています。
- 「個々の食生活選択と社会への影響」章で、個々人の食選択が社会全体の健康水準に与える影響と、それによる経済的及び環境的メリットが説明されています。
- 最後の「結論」章では、バランスの取れた食生活がもたらす健康効果と、それを支持するための個人と社会の役割がまとめられています。

これらの点が、文章全体の流れと主要なメッセージを理解するために特に重要です。

□ 保存 □ コピー

RAG ▾ ● ここに質問を入力してAIに聞いてみましょう

• 生成された回答を上にスクロールしていくと、AIが参照した文章が表示されます。

参照された文章をクリックすることで参照箇所を確認することができます。

① ファイル指定 Health.pdf

□ 紋り込みを解除

この文章の大半の部分は、食事と健康についての複数の重要な側面を説明している各章の要点です。具体的には、以下のポイントが重要です：

AIが関連するファイルを調べています...

Health.pdf
もし、食事と健康の関連に関する関心も増加しています。適切な食事は、単に病気の予防だけでなく、生活の質を向上させ、心身のウェルネスを促進する重要な要素です。本論文では、栄養学的視点から、食事が人間の健康に与える影響を多角的に分析します。具体的には、バランスの取れた食事がどのようにして体と心に良い影響を与えるか、そして不適切な食生活が引き起こす潜在的リスクについて探求します。2. 栄養素の基本とその機能 人間の体は、生きていく上で必要なエネルギーと栄養素を食事から取り入

p.1

● OpenAI GPT-4 Turboが回答を作成しました

この文章の大半の部分は、食事と健康についての複数の重要な側面を説明している各章の要点です。具体的には、以下のポイントが重要です：

- 「序論」で述べられた、適切な食事が人間のウェルネスを促進する重要な要素である点。
- 「栄養素の基本とその機能」章で詳細に説明されている、主要な栄養素（炭水化物、タンパク質、脂質、ビタミン・ミネラル）の役割。

RAG ▾ ● ここに質問を入力してAIに聞いてみましょう

● 食事と健康 栄養に基づいたライフスタイルの選択が長寿とウェルネスに及ぼす影響

1.序論

現代社会における健康志向の高まりとともに、食事と健康の関連に関する関心も増加しています。適切な食事は、単に病気の予防だけでなく、生活の質を向上させ、心身のウェルネスを促進する重要な要素です。本論文では、栄養学的視点から、食事が人間の健康に与える影響を多角的に分析します。具体的には、バランスの取れた食事がどのようにして体と心に良い影響を与えるか、そして不適切な食生活が引き起こす潜在的リスクについて探求します。2. 栄養素の基本とその機能

人間の体は、生きていく上で必要なエネルギーと栄養素を食事から取り入れる必要があります。主要な栄養素には以下のがあります：

炭水化物：エネルギーの主要な供給源です。全粒穀物、野菜、果物に豊富に含まれています。

タンパク質：筋肉や細胞の構築、修復に必要です。肉類、魚類、豆類、乳製品に多く含まれています。

カスタムLLMの設定

- 通常のチャットに戻る場合には「絞り込みを解除」をクリックします。

The screenshot shows a user interface for a custom LLM. At the top, there's a header with a file icon, the text '① ファイル指定' (File Specified), and the file name 'Health.pdf'. To the right of the file name is a button labeled '絞り込みを解除' (Remove Filter) with a red rectangular highlight and a hand cursor icon. Below the header is a large text area containing a document about nutrition and health. In the bottom right corner of this text area, it says 'p.1'. On the far right edge of the text area, there's a small circular icon with a question mark.

● OpenAI GPT-4 Turboが回答を作成しました

この文章の大変な部分は、食事と健康についての複数の重要な側面を説明している各章の要点です。具体的には、以下のポイントが重要です：

1. 「序論」で述べられた、適切な食事が人間のウェルネスを促進する重要な要素である点。
2. 「栄養素の基本とその機能」章で詳細に説明されている、主要な栄養素（炭水化物、タンパク質、脂質、ビタミン・ミネラル）の役割。
3. 「適切な食事の健康効果」章において、健康的な食生活が心血管疾患の予防、体重管理と糖尿病予防、精神健康の支持にどのように寄与するかの説明。
4. 「食生活と公衆衛生」章では、不適切な食生活が公衆衛生に及ぼす悪影響と、健康的な食生活を推進するための公共キャンペーンの重要性が強調されています。
5. 「個々の食生活選択と社会への影響」章で、個々人の食選択が社会全体の健康水準に与える影響と、それによる経済的及び環境的メリットが説明されています。
6. 最後の「総論」章では、バランスの取れた食生活がもたらす健康効果と、それを支持するための個人と社会の役割がまとめられています。

これらの点が、文章全体の流れと主要なメッセージを理解するために特に重要です。

● 保存

□ コピー

RAG ▾ ● ここに質問を入力してAIに聞いてみましょう

文書全体を渡して質問 (DOCモード)

ailia DX InsightではRAGを使用することによって回答性能を向上させています。[RAGについて](#)はこちらをご確認ください。

しかし、要件によってはテキストのすべての要素を洗い出したい場面があります。

その場合は、文章ファイルを右クリックして、「この文章全体を渡して質問 (DOC)」を選択してください。

カスタムLLMの設定

ファイル名 ^	サイズ	AIステータス	更新日時	操作
ailiaDXInsight_manual.pdf	5 MB	PDF	2024/07/10 19:56	
ailiaDXInsight_manual_copy.pdf	2 MB	PDF	2024/07/18 20:22	
DietAndHealth.pdf	8 MB	PDF	2024/07/29 18:12	
DietAndHealth_ja.pdf	8 MB	PDF	2024/08/09 17:59	
doc.txt	2 KB		2024/10/17 10:12	
file.txt	960 B	✓	2024/10/24 17:13	
Health.pdf	8 MB	✓ PDF	2024/07/29 19:00	⋮
hsl.pdf	この文章を対象にチャット (RAG)			
hsl_a...	この文章全体を渡して質問 (DOC)			
overv...	このノードをインテリジェンスから削除			
PhysicsII-1.pdf	190 KB	PDF	2024/06/06 17:25	
Rerank.ai	4 MB		2024/08/12 09:10	
Rerank.pdf	4 MB	PDF	2024/08/09 10:30	

DOCの場合、RAGによる情報検索を行わず、ドキュメント全体をLLMに投げます。コストはかかりますが、より高精度になります。なお、チャットを継続した場合も、その都度、ドキュメント全体のトークンのコストがかかります。

複数の文書ファイルについて質問する

関連の文書ファイルをフォルダにまとめておくことで、フォルダ内の複数文書ファイルに跨った内容の質問をすることができます。

画面左上のアプリメニューはチャットの状態にしておいてください。

1. インデックス登録済みのファイルが格納されたフォルダを右クリックし、「このフォルダを対象にチャット」を選択します。

操作	リスト	グリッド
/ D: / Private / Document		
リスト		
グリッド		
ファイル名 ^	サイズ	AIステータス
Folder01		
Folder02		
Folder03		

右側メニュー

- このフォルダを対象にチャット
- このフォルダ内のファイルをインデックスに追加
- このフォルダをブックマークに追加する
- エクスプローラーで開く

2. 指定されたフォルダが上部に表示されていることを確認します。

下部にはフォルダに含まれるファイルが一覧で表示されます。

カスタムLLMの設定

The screenshot shows the configuration interface for a custom LLM. At the top, there is a file selection bar with a red border around the 'ファイル指定' (File Selection) button and the path 'D:\Private\Document\Folder02\'. To the right of this is a checkbox labeled '絞り込みを解除' (Remove filter). Below this is a large empty area for document preview.

In the center, there is a red-bordered box containing a section titled 'あなたの質問' (Your Question). It lists several PDF files from the selected folder:

- D:\Private\Document\Folder02\ailiaDXInsight_manual.pdf
- D:\Private\Document\Folder02\ailiaDXInsight_manual_copy.pdf
- D:\Private\Document\Folder02\Health.pdf
- D:\Private\Document\Folder02\hisl.pdf
- D:\Private\Document\Folder02\hisl_appendix.pdf
- D:\Private\Document\Folder02\water_healing.pdf

To the right of the list is a black button with a white 'COPY' icon and the word 'コピー' (Copy).

At the bottom of the interface is a search bar with a dropdown menu set to 'RAG'. The input field contains the text 'ここに質問を入力してAIに聞いてみましょう' (Input your question and let AI answer it). To the right of the input field is a small counter '0' and a blue circular icon with a white question mark.

3. チャットボックスの左側の表示が「RAG」となるので、質問を入力し、送信します。

The screenshot shows the chat interface. At the top, there is a navigation bar with tabs: 'おすすめプロンプト' (Recommended Prompt), 'コンサルタント' (Consultant), '営業' (Sales), 'プロダクトマネージャー' (Product Manager), and 'システムエンジニア' (System Engineer). Below this is a red-bordered input field containing the text 'この中でストレスについて記述があるファイルはどれ？」 (Which file in this list contains a description of stress?). To the right of the input field is a large red-bordered send button with a white right-pointing arrow icon.

カスタムLLMの設定

4. AIが関連するファイルを調べた回答が表示されます。（検索はローカルAIで行われます）

① ファイル指定 D:\Private\Document\Folder02\

絞り込みを解除

視点から、食事が人間の健康に与える影響を多角的に分析します。具体的には、バランスの取れた食事がどのようにして体と心に良い影響を与えるか、そして不適切な食生活が引き起こす潜在的リスクについて探求します。2. 栄養素の基本とその機能 人間の体は、生きていく上で必要なエネルギーと栄養素を食事から取り入

water_healing.pdf

されています 実験デザイン 本研究では、被験者をランダムに2群に分け、一方には水の音を聞かせ、もう一方には静かな室内環境を提供しました。各セッションは30分間で、心理的及び生理的パラメーターをセッション前後で測定しました。0001 ### 測定指標 1. 心理的影響：ストレスレベル（自己評価スケール）、心理状態（心理的アセスメントツール） 2. 生理的影响：心拍数、血圧、呼吸率 ## 結果 ### 心理的結果 水の音を聞いた群では、ストレスレベルが有意に低下し、全体的な心理状態が改善しました。対照群と比較して、明らかな差異が観察されました ($p < 0.05$)。### 生理的結果 同様に、水の音を聞いた群では心拍

p.1

② OpenAI GPT-4 Turboが回答を作成しました

ストレスについて記述があるファイルは、「D:\Private\Document\Folder02\water_healing.pdf」です。このファイルの中で、「心理的影響：ストレスレベル（自己評価スケール）、心理状態（心理的アセスメントツール）」との記述があります。また、水の音を聞いた群ではストレスレベルが有意に低下したという研究結果が述べられています。

保存

コピー

RAG

ここに質問を入力してAIに聞いてみましょう



5. 生成された回答を上にスクロールしていくと、AIが参照したファイル、文章を確認することができます。

① ファイル指定 D:\Private\Document\Folder02\

絞り込みを解除

Health.pdf

食事と健康 単調な生活のライフスタイルの過度なストレスは、心身への影響を及ぼすことがあります。一方で、健康的な食事は、心身の健康を維持するのに役立ちます。特に、高野菜や高穀物の食事は、心身の健康を保つのに効果的です。しかし、高野菜や高穀物だけではなく、他の健康的な食事も重要です。例えば、魚や豆類などのタンパク質を多く含む食品は、心身の健康を保つのに役立つことがあります。また、低GIの食事は、血糖値の変動を最小限に抑え、心身の健康を保つのに役立つことがあります。これらの食事は、過度なエネルギー消費を防ぎ、心身の健康を保つのに役立つことがあります。

water_healing.pdf

されています 実験デザイン 本研究では、被験者をランダムに2群に分け、一方には水の音を聞かせ、もう一方には静かな室内環境を提供しました。各セッションは30分間で、心理的及び生理的パラメーターをセッション前後で測定しました。0001 ### 測定指標 1. 心理的影響：ストレスレベル（自己評価スケール）、心理状態（心理的アセスメントツール） 2. 生理的影響：心拍数、血圧、呼吸率 ## 結果 ### 心理的結果 水の音を聞いた群では、ストレスレベルが有意に低下し、心身の健康を保つのに役立つことがあります。対照群と比較して、明らかに心身の健康が改善されました ($p < 0.05$)。### 生理的結果 同様に、水の音を聞いた群では心拍

はじめに
自然界の音は古来より人々の心身に影響を与えてきたとされ、特に水の音はそのリラクゼーション効果が注目されています。本論文では、水の音が人間の心理的、生理的健康に与える影響について、科学的な観点から詳細に調査し、そのヒーリング効果に焦点を当てています。



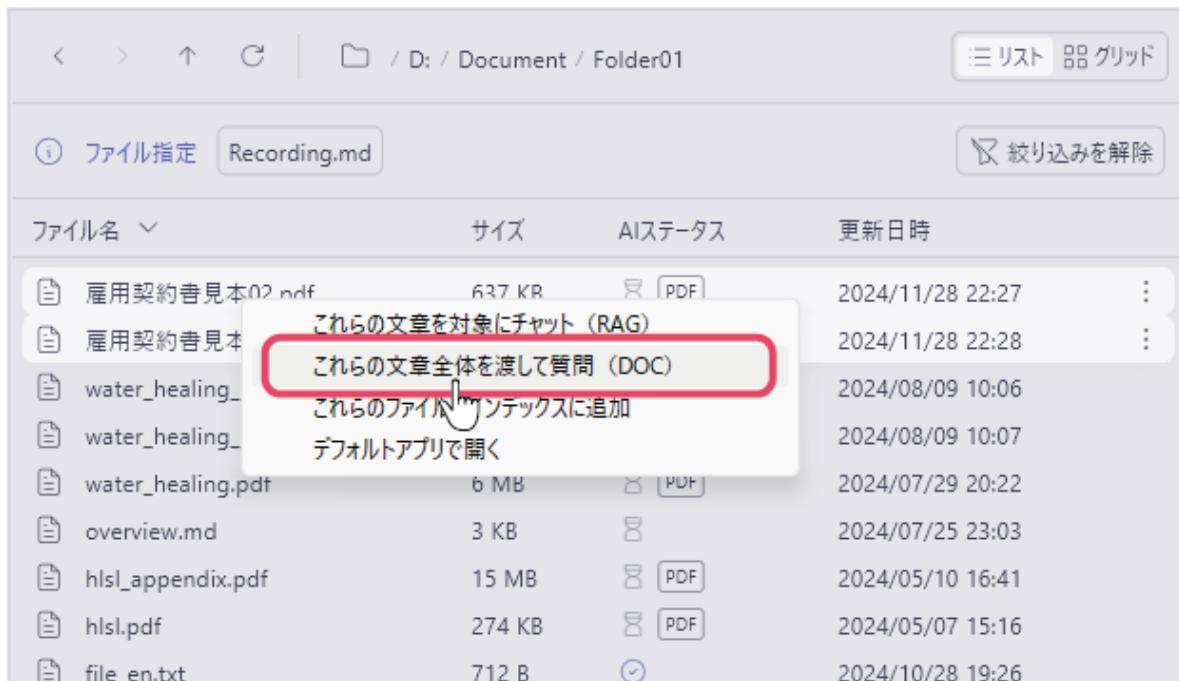
- 通常のチャットに戻る場合には「絞り込みを解除」をクリックします。

複数ファイルを比較して質問する

2つ以上のファイルを選択し、DOCモードで質問をすることでファイル同士をAIによって比較させることができます。

更新された契約書の変更箇所の確認や、文書ファイルの変更箇所の確認を素早く簡単に行うことができます。

1. チャットモードでShift（もしくはCtrl）押しながら2つ以上のファイルを選択した状態で右クリックメニューから「これらの文章全体を渡して質問（DOC）」を選択します。



2. チャットボックスに質問を入力し、送信します。

この時、チャットボックスの左側の表示が「DOC」となっていることを確認してください

あなたの質問

DocModeで使用するトークンを測定中です...

あなたの質問

これらの文章全体を対象に質問を行います (推定: 1641トークン) :

D:\Document\Folder01\雇用契約書見本02.pdf
D:\Document\Folder01\雇用契約書見本.pdf

2つのファイルの差分を教えて下さい

3. AIが選択されたファイルを参照し、回答を生成します。

● OpenAI GPT-4oが回答を作成しました

2つのファイル「雇用契約書見本02.pdf」と「雇用契約書見本.pdf」の主な差分は以下の通りです。

1. 契約期間の違い:

- 「雇用契約書見本02.pdf」では、契約期間が2025年4月1日から2026年3月31日までの1年間です。
- 「雇用契約書見本.pdf」では、契約期間が2024年4月1日から2025年3月31日までの1年間です。

2. 報酬（年俸）の違い:

- 「雇用契約書見本02.pdf」では、被雇用者の年俸は4,300,000円です。
- 「雇用契約書見本.pdf」では、被雇用者の年俸は3,600,000円です。

3. 年俸の支払日:

- 「雇用契約書見本02.pdf」では、年俸の支払日は毎月15日です。
- 「雇用契約書見本.pdf」では、年俸の支払日は毎月25日です。

その他の内容については基本的には同じです。契約書の条項や守秘義務の内容、契約の変更・解除の手続きなどは一致しています。

保存

コピー

DOC ▾

GPT-4o

▼



ここに質問を入力してAIに聞いてみましょう

0

RAG

ailia DX Insightは回答を生成する際、Retrieval-Augmented Generation (RAG) を使用することで出力結果の根拠を明確にし、事実に基づかない情報の生成を抑制しています。

RAGに関して、詳しくは[こちら](#)をご参照ください。

[次のページへ >](#)

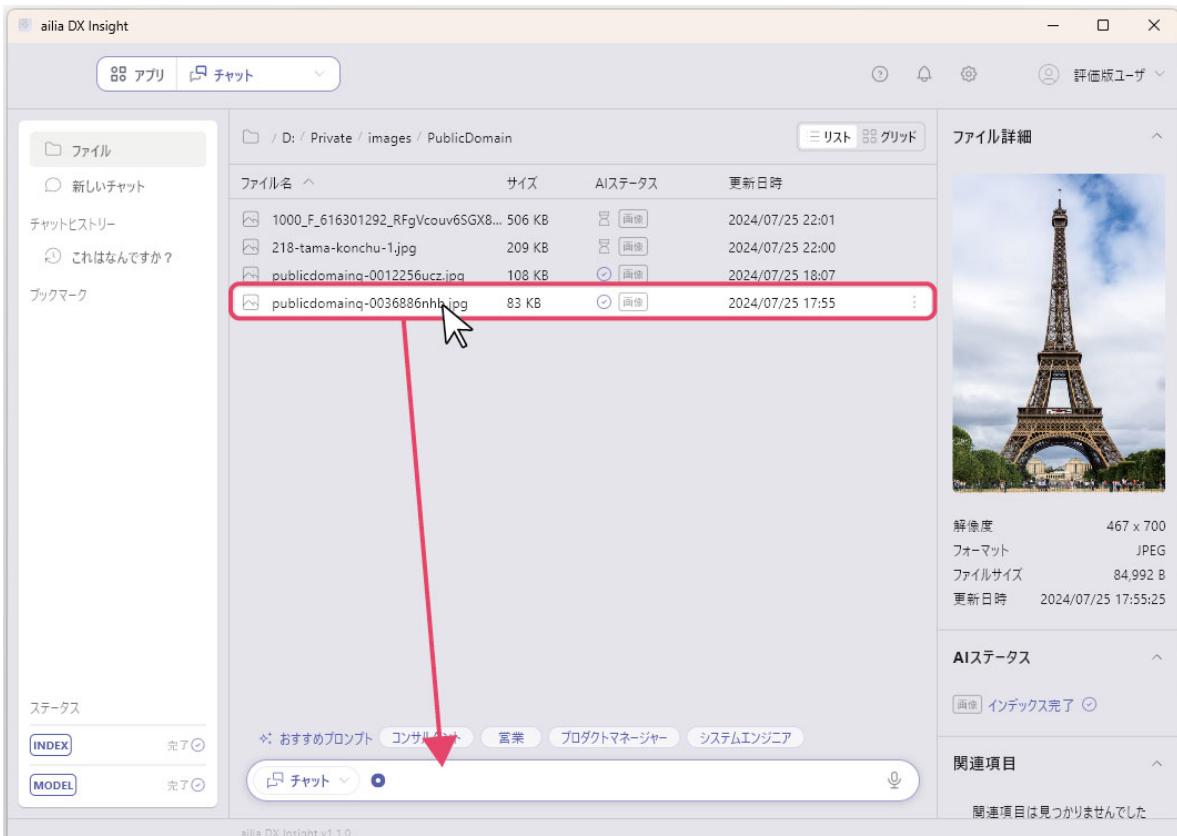
[一覧に戻る](#)

画像ファイルについて質問する

画像ファイルをインデックスに登録するとその画像について質問をすることができます。画面左上のアプリメニューは「チャット」にしておいてください。

1. 画像ファイルを画面下のチャットボックスにドラッグ & ドロップします。

(Windowsの場合はエクスプローラー、Macの場合はファインダーからもチャットボックスに直接ドラッグ & ドロップすることができます)



2. チャットボックスの中にドロップした画像が表示されるので、質問を入力し、送信します。



カスタムLLMの設定

3. AIからの回答が表示されます。

あなたの質問
これはなんですか？



□ コピー

OpenAI GPT-4 Turboが回答を作成しました

これはエッフェル塔です。フランス・パリにある有名な観光名所で、1889年の万国博覧会のために建設されました。エッフェル塔は鉄製で、高さは約300メートルあります。

保存 □ コピー

おすすめプロンプト コンサルタント 営業 プロダクトマネージャー システムエンジニア

チャット ここに質問を入力してAIに聞いてみましょう

[次のページへ >](#)

[一覧に戻る](#)

文書ファイルを要約する。

文書ファイルをインデックスに登録すると、登録した文書ファイルに記載された情報についてAIに要約させることができます。

ファイルをインデックス登録する方法については[こちら](#)をご確認ください。

1. 画面左上のアプリメニューから「テキスト要約」を選択します。



2. 要約したいファイルを右クリックし、「要約」をクリックします。



3. AIが要約した結果が表示されます。

[次のページへ >](#)

[一覧に戻る](#)

文書ファイルを翻訳する

文書ファイルをインデックスに登録すると、登録した文書ファイルをAI機能によって翻訳することができます。画面中央下のチャットボックスに文書を直接入力することで、その文書を翻訳することもできます。

ファイルをインデックス登録する方法については[こちら](#)をご確認ください。

1. 画面左上のアプリメニューから「テキスト翻訳」を選択します。



2. 翻訳したいファイルを右クリックし、「翻訳」をクリックします。対応しているファイルフォーマットはpdf, txt, md, docx, xlsx,です。



3. 翻訳元のテキストに並んでAIが翻訳した結果が表示されます。翻訳結果の右下にある「保存」ボタンを押すと、翻訳結果をファイルとして保存することができます。

⑧ あなたの質問

次のファイルを翻訳してください：
D:\Document\Folder01\Asakusa.txt

□ コピー

● OpenAI GPT-4o miniが回答を作成しました

浅草寺（せんそうじ）は、東京都台東区に位置する日本で最も古く、有名な仏教寺院の一つです。この寺院は628年に創建され、浅草の地域はその周囲を中心に発展しました。浅草寺は日本の伝統や文化に興味がある外国人観光客にとって非常に人気のある観光地です。

主な特徴として、仁王門（雷門とも呼ばれる）が有名であり、巨大な提灯が下がっている様子は多くの写真やポストカードで見ることができます。この門をくぐると、長い商店街「仲見世通り」が続いており、色々な日本土産や食べ物を販売する店が並んでいます。仲見世を抜けると本堂にたどり着きます。

Senso-ji Temple is one of the oldest and most famous Buddhist temples in Japan, located in Taito Ward, Tokyo. This temple was founded in 628, and the Asakusa area developed around it. Senso-ji is a very popular tourist destination for foreign visitors interested in Japanese traditions and culture.

One of its main features is the Niōmon Gate (also known as the Thunder Gate), which is famous for the large lantern hanging from it and can be seen in many photographs and postcards. Passing through this gate, you will find the long shopping street called Nakamise Street, lined with shops selling various Japanese souvenirs and food. After passing through Nakamise, you will arrive at the main hall.

□ 保存

□ コピー

翻訳 ▾

プロンプトで指定 (LLM) ▾

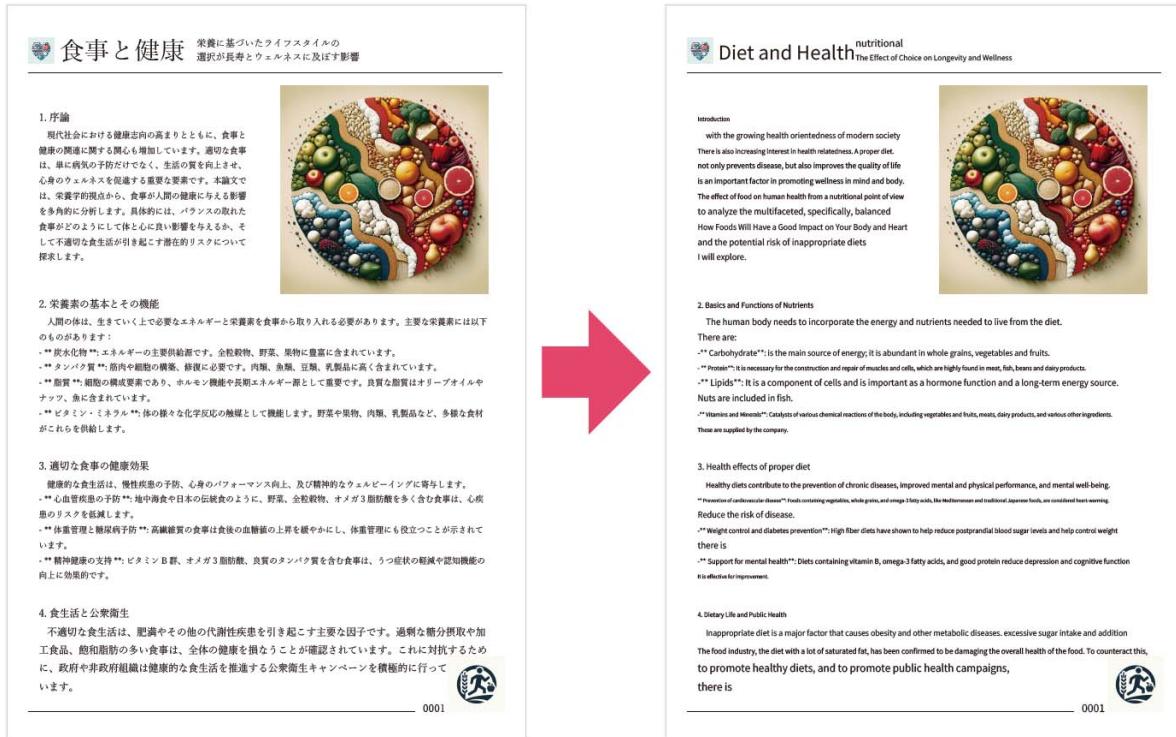


ここに質問を入力してAIに聞いてみましょう



カスタムLLMの設定

4. PDFファイルは翻訳をすると、PDFのまま出力することができます。(*1)



(*1)なお、PDFは非常に複雑な構造を持っており、ailia DX Insightの内部でレイアウトを推定しています。PDFの種類によっては、翻訳後のレイアウトが崩れる場合があります。

文章を翻訳する

日本語で書かれた文章は英語へ、日本語以外の言語で書かれた文は日本語へ翻訳されます。

1. 画面左上のアプリメニューから「テキスト翻訳」を選択します。



2. 翻訳したい文書を画面中央下のチャットボックスに入力し、送信します。

3. AIが翻訳した結果が表示されます。

② あなたの質問

Please translate the following text. If the text is in Japanese, please translate it into English. If the sentence is not in Japanese, please translate it into Japanese. Answer only the result of the translation without including any additional formatting or characters:
...

浅草寺（せんそうじ）は、東京都台東区に位置する日本で最も古く、有名な仏教寺院の一つです。この寺院は628年に創建され、浅草の地域はその周囲を中心に発展しました。浅草寺は日本の伝統や文化に興味がある外国人観光客にとって非常に人気のある観光地です。

主な特徴として、仁王門（雷門とも呼ばれる）が有名であり、巨大な提灯が下がっている様子は多くの写真やポストカードで見ることができます。この門をくぐると、長い商店街「仲見世通り」が続いており、色々な日本土産や食べ物を販売する店が並んでいます。仲見世を抜けると本堂にたどり着きます。

浅草寺は観音菩薩を本尊としており、多くの人々が願い事をし、お参りをするために訪れます。また、寺の境内や周辺では季節ごとにさまざまな祭りやイベントが開催され、特に春の浅草三社祭や新年の初詣は大変多くの人で賑わいます。

また、浅草寺周辺は古き良き日本の雰囲気を感じられるエリアとしても知られており、芸者文化や伝統的な日本建築を楽しむことができます。外国人観光客はこの地域の魅力を十分に体験することができ、日本文化にふれる貴重な機会となるでしょう。
...

 コピー

③ ローカル (ailia LLM) gemma-2-2b-it-Q4_K_Mが回答を作成しました

The Sensoji temple (浅草寺) is one of the oldest and most famous Buddhist temples in Japan, located in Tokyo's Taitō ward. This temple was founded in 628 and the surrounding area of Asakusa has developed around it. The Sensoji temple is very popular among foreign tourists interested in Japanese tradition and culture.

Its main feature is the To-do-men (雷門) which is famous for its giant lanterns. The scene of the lanterns has been documented in photos and postcards. Passing this gate, a long shopping street called "Nakamise-dori" continues, with shops selling various Japanese souvenirs and snacks. After passing the Nakamise, you reach the main hall.

The Sensoji temple is dedicated to Kannon Bodhisattva and many people visit to make wishes and pray. Various festivals and events are held throughout the year in the temple's grounds and surrounding areas, particularly the Spring Asakusa Three Temples Festival and New Year's First Visit are extremely popular.

The area around the Sensoji temple is also known for its nostalgic old Japan atmosphere, with opportunities to experience traditional Japanese architecture and geisha culture. Foreign tourists can experience this area's charm, making this a valuable opportunity to understand Japanese culture.

 保存

 コピー

[次のページへ >](#)

[一覧に戻る](#)

文書ファイルを校正する

文書ファイルをインデックス登録すると、登録した文書ファイルに記載されている内容をAIに校正させることができます。

ファイルをインデックス登録する方法については[こちら](#)をご確認ください。

1. 画面左上のアプリメニューから「テキスト校正」を選択します。



2. 校正したいファイルを右クリックし、「校正」を選択します。

The screenshot shows a file list in a grid view. The columns are 'File Name', 'Size', 'AI Status', and 'Last Updated'. A specific row for 'Health.pdf' is selected and highlighted with a red box. The 'Correct' option in the context menu for this file is also highlighted with a red box and has a cursor icon over it. Other files listed include 'FileView.txt', 'hsl.pdf', 'hsl_appendix.pdf', 'overview.md', 'PhysicsII-5.pdf', 'Rerank.ai', 'Rerank.pdf', 'Rerank_EN.pdf', 'Rerank_vvv.pdf', 'water_healing.pdf', 'water_healing_0.pdf', and 'water healing 02.pdf'. Each file has a 'PDF' status indicator and a date/time column.

3. AIが文書を校正し、校正前の文章が左に、校正結果の文章が右に表示されます。

The screenshot shows a comparison between the original text on the left and its corrected version on the right. The original text discusses the relationship between diet and health, mentioning the importance of various nutrients like carbohydrates, proteins, and fats. The corrected text on the right is a simplified summary of this information.

● OpenAI GPT-4o miniが回答を作成しました

食事と健康 栄養に基づいたライフスタイルの 選択が長寿とウェルネスに及ぼす影響

1. 序論 現代社会における健康志向の高まりとともに、食事と健康の関連に関する関心も増加しています。適切な食事は、単に病気の予防だけでなく、生活の質を向上させ、心身のウェルネスを促進する重要な要素です。本論文では、栄養学的視点から、食事が人間の健康に与える影響を多角的に分析します。具体的には、バランスの取れた食事がどのようにして体と心に良い影響を与えるか、そして不適切な食生活が引き起こす潜在的リスクについて探求します。

2. 栄養素の基本とその機能 人間の体は、生きていく上で必要なエネルギーと栄養素を食事から取り入れる必要があります。主要な栄養素には以下 のものがあります：

- 炭水化物: エネルギーの主要供給源です。全粒穀物、野菜、果物に豊富に含まれています。
- タンパク質: 筋肉や細胞の構築、修復に必要です。肉類、魚類、豆類、乳製品に高く含まれています。
- 脂質: 細胞の構成要素であり、ホルモン機能や長期エネルギー源として重要です。良質

食事と健康
栄養に基づいたライフスタイルの
選択が長寿とウェルネスに及ぼす影響
1. 序論
現代社会における健康志向の高まりとともに、食事と
健康の関連に関する関心も増加しています。適切な食事
は、単に病気の予防だけでなく、生活の質を向上させ、
心身のウェルネスを促進する重要な要素です。本論文で
は、栄養学的視点から、食事が人間の健康に与える影響
を多角的に分析します。具体的には、バランスの取れた
食事がどのようにして体と心に良い影響を与えるか、そ
して不適切な食生活が引き起こす潜在的
リスクについて
探求します。
2. 栄養素の基本とその機能
人間の体は、生きていく上で必要なエネルギーと栄養素を食事から取り入れる必要が
あります。主な栄養素には以下 のものがあります：

[次のページへ >](#)

[一覧に戻る](#)

画像を検索する

画像ファイルをインデックス登録することで、類似画像検索を行うことができます。インデックス登録された画像ファイルを対象に検索が実行されます。テキストによる検索と、画像による検索の2通りの検索方法があります。

ファイルをインデックス登録する方法については[こちら](#)をご確認ください。

テキストで検索

検索ワードに類似した画像を検索することができます。

1. 画面左上のアプリメニューから「画像検索」を選択します。



2. チャットボックスに検索ワードを入力し、送信します。
3. 類似度順に検索結果が表示されます。表示された検索結果をクリックすると、全ての画像が類似度順に並べられグリッド表示されます。

画像で検索

選択した画像に類似した画像を検索することができます。画像はインデックス登録済である必要があります。

1. 「画像検索」モードの状態で検索したい画像ファイルを右クリックし、「この画像で類似検索」を選択します。



2. 類似度順に検索結果が表示されます。表示された検索結果をクリックすると、全ての画像が類似度順に並べられグリッド表示されます。

[次のページへ >](#)

カスタムLLMの設定

[一覧に戻る](#)

画像を生成する

チャットボックスに入力した文書を元にAIによる画像生成を行うことができます。

また、ailia DX Insightでは生成された画像を編集し、より望むものに近づけることができる他、著作権保護や改ざん防止の為にウォーターマークを埋め込んで画像を生成させることができます。

※画像生成にはDALL-Eを使用するため、GPT-4以降で動作します。

※GPTのバージョンは設定  の中の「チャットAI」の項目の中で「OpenAI」をクリックすると選択できます。

プロンプトから画像を生成する

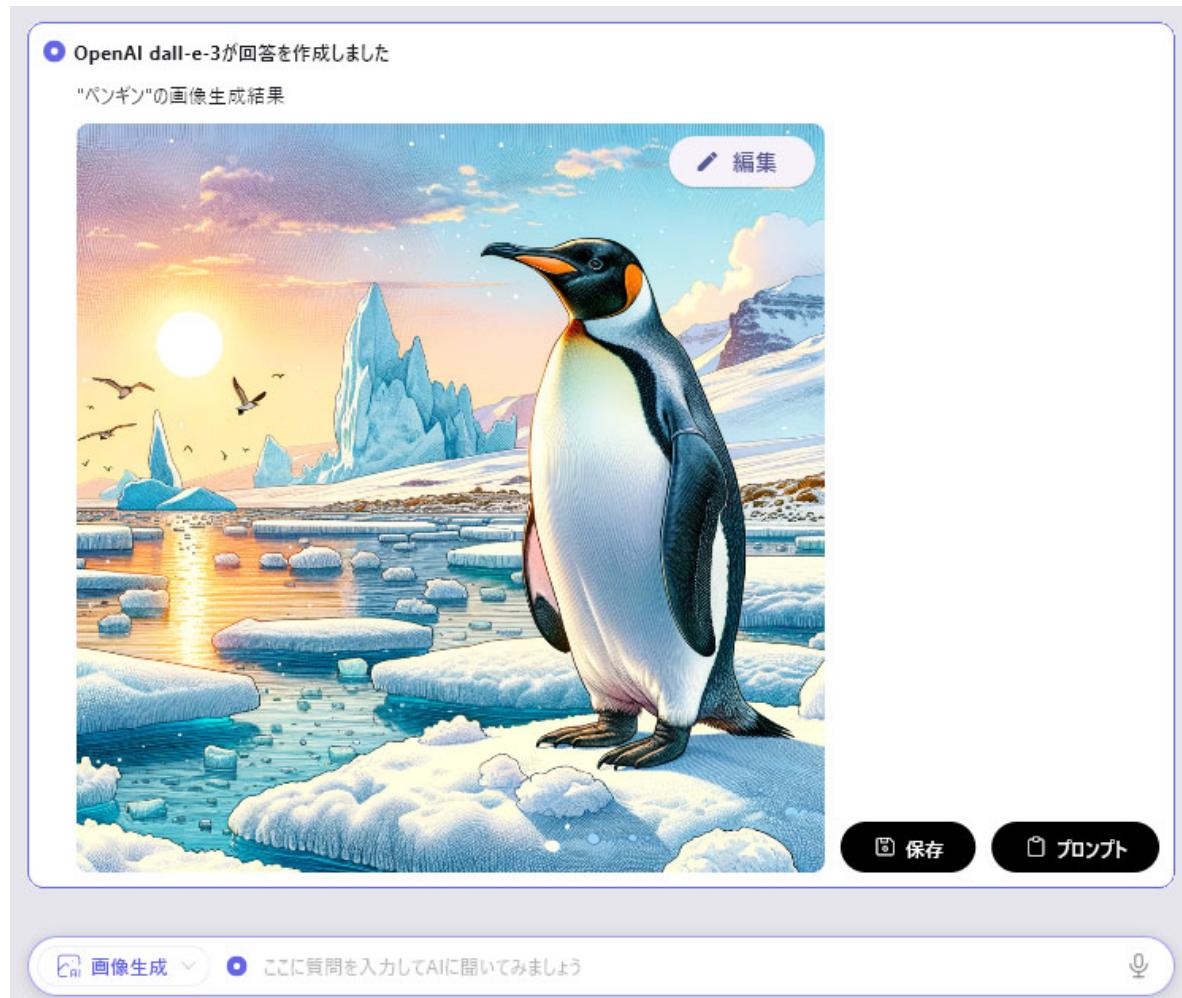
キーワード等の文字から画像を生成することができます。

1. 画面左上のアプリメニューから「画像生成」を選択します。



2. 生成したい画像の概要を入力し、送信します。

3. AIが画像を生成した結果が表示されます。

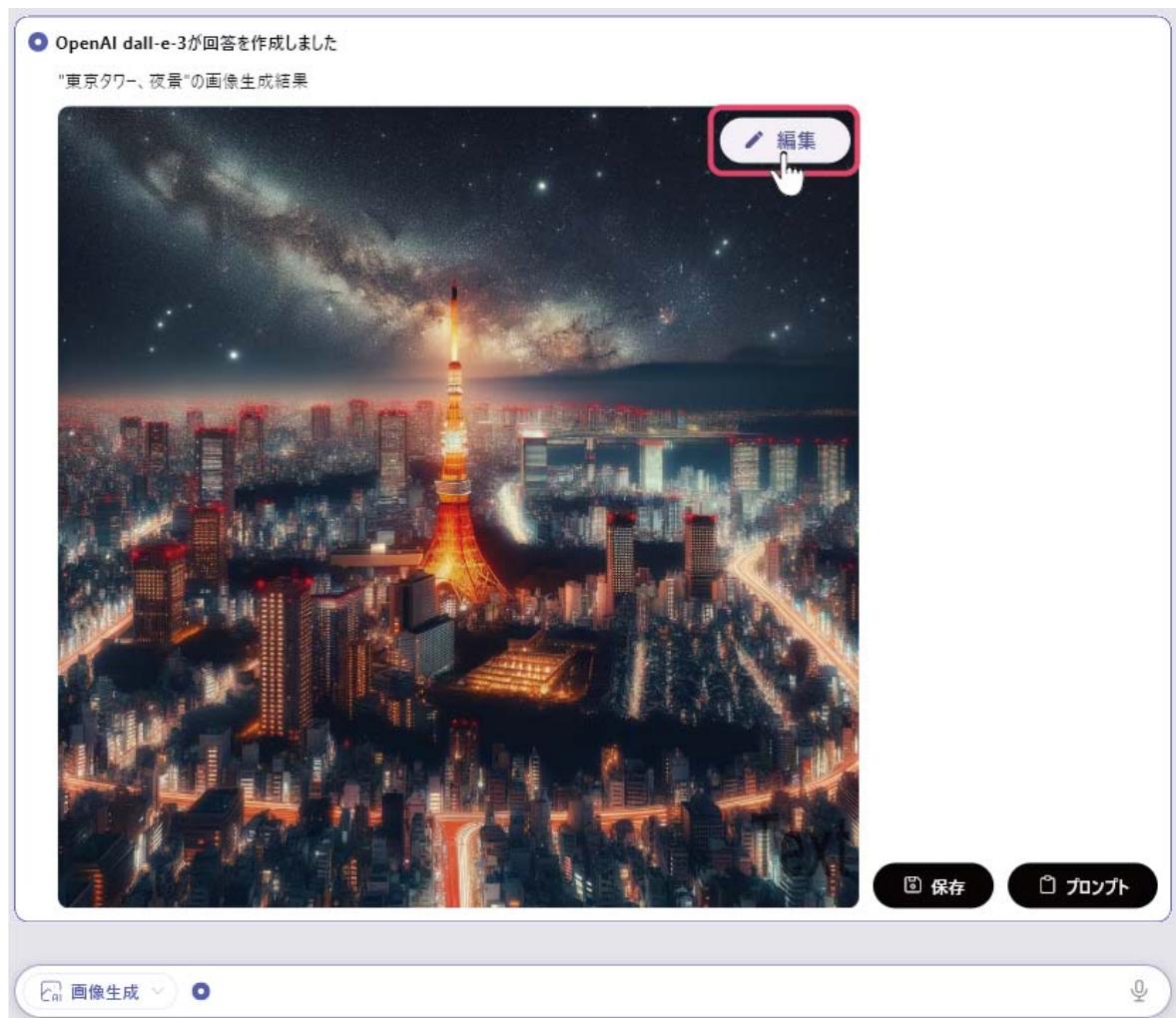


生成された画像を編集する

生成された画像を編集し、より望むものに近づけることができます。

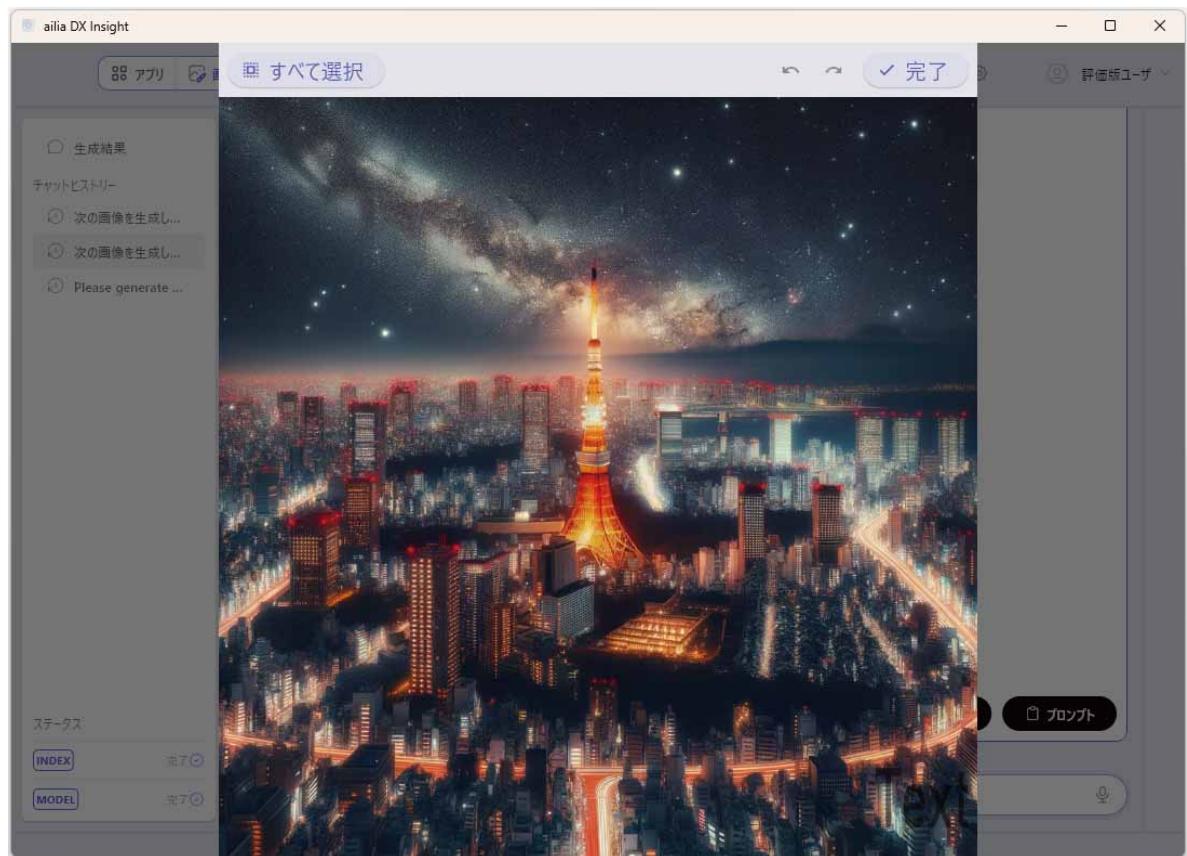
カスタムLLMの設定

- 生成された画像の右上に表示されている「編集」を選択します。



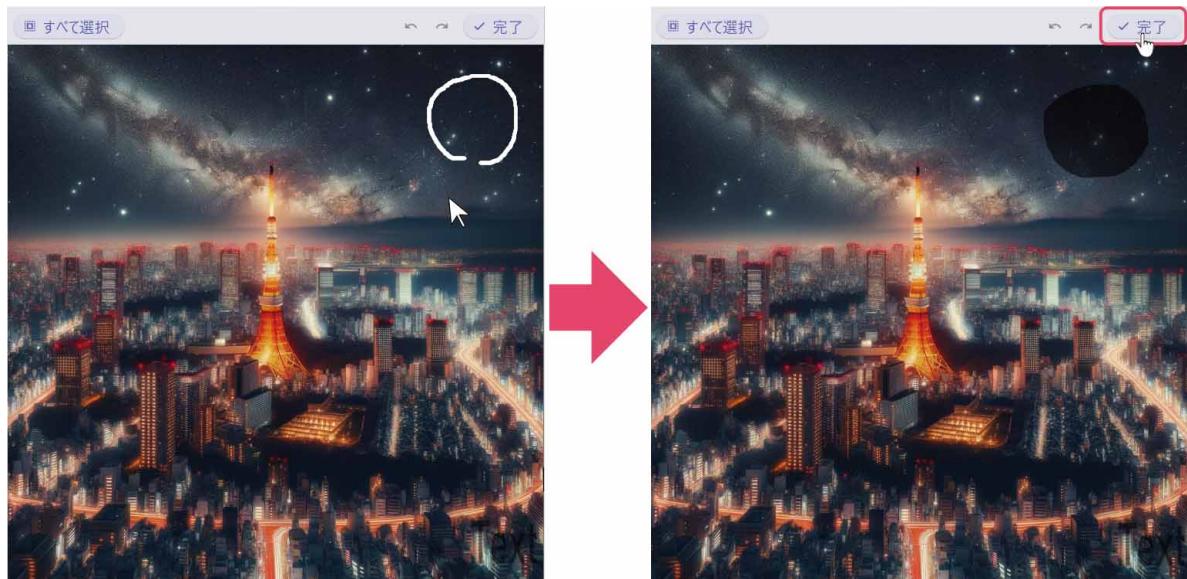
カスタムLLMの設定

2. 編集画面が表示されます



3. 編集したい箇所をマウスでドラッグして囲み、マスクを作成します。

マスクを作成した状態で「Done」を選択し、マスクの作成を完了します。



4. 画面下のチャットボックスに編集したい内容を入力し、送信します。

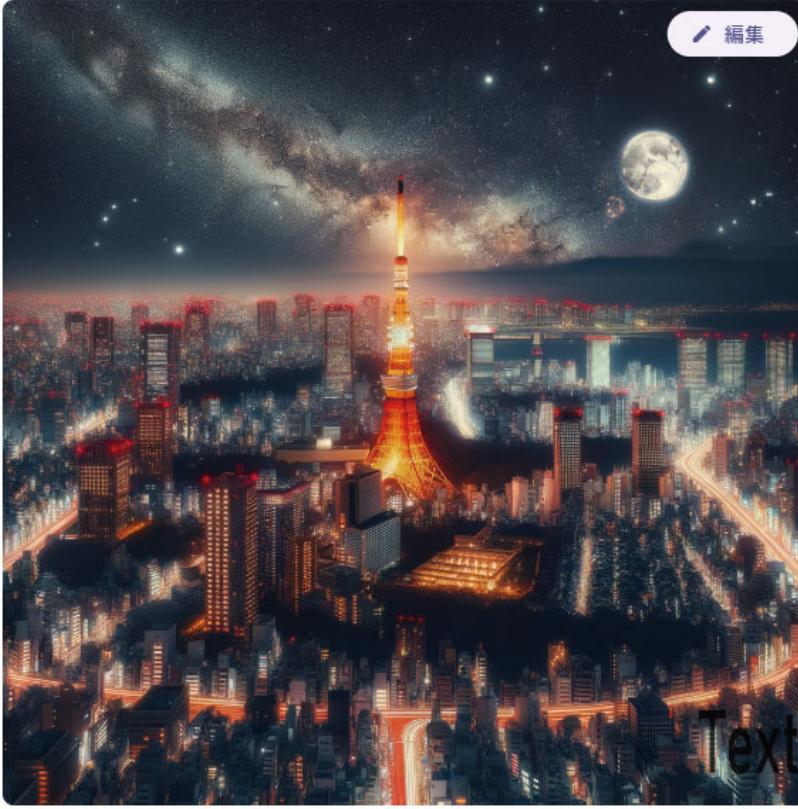


カスタムLLMの設定

5. AIが画像を再生成した結果が表示されます。

● OpenAI dall-e-2が回答を作成しました

"雲がかった満月を追加して"の画像生成結果



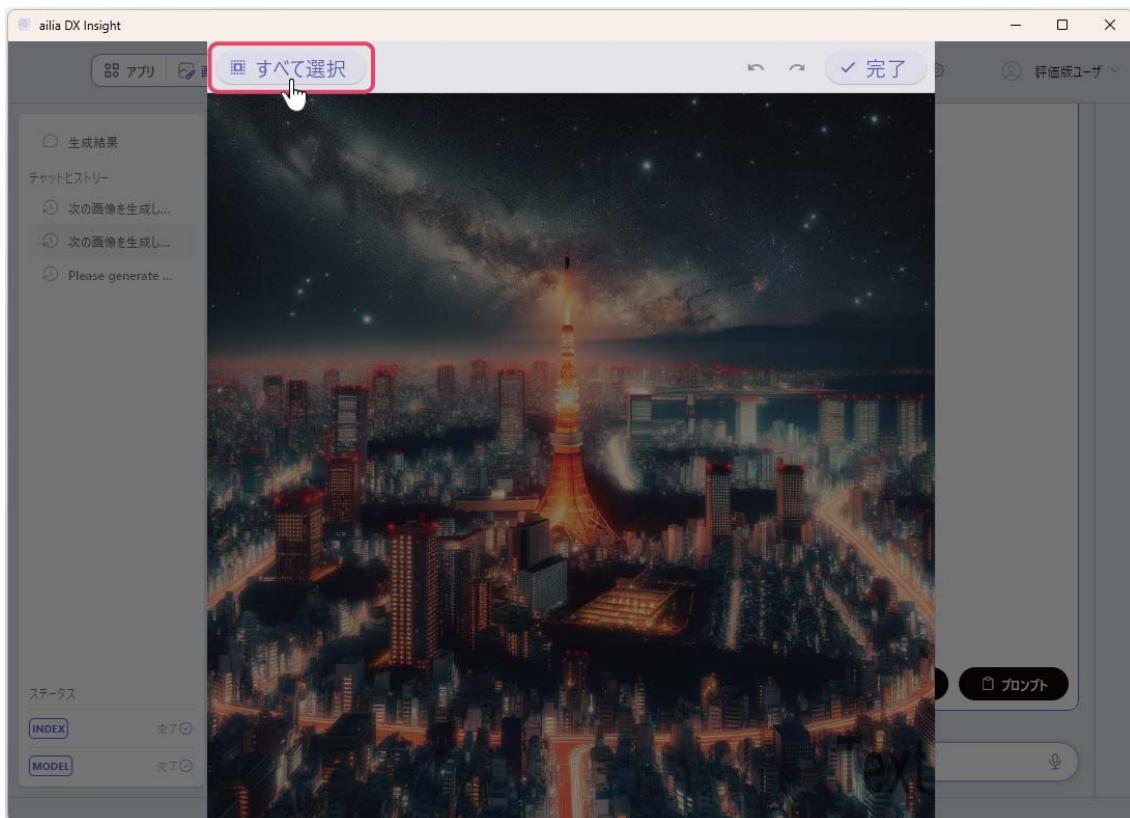
編集

保存 プロンプト

画像生成 ここに質問を入力してAIに聞いてみましょう

カスタムLLMの設定

- 画像の編集画面で左上の「すべてを選択」を押すことで、画面全体を選択することができます。



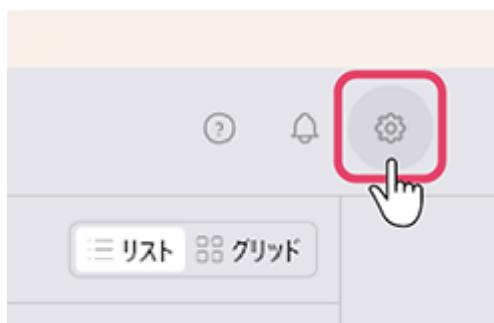
生成される画像にウォーターマークを追加する

ailia DX Insightでは生成する画像に、ウォーターマークを手軽に追加することができます。

ここでは見本としてこちらのロゴマークを50%透過させた画像を使用します。

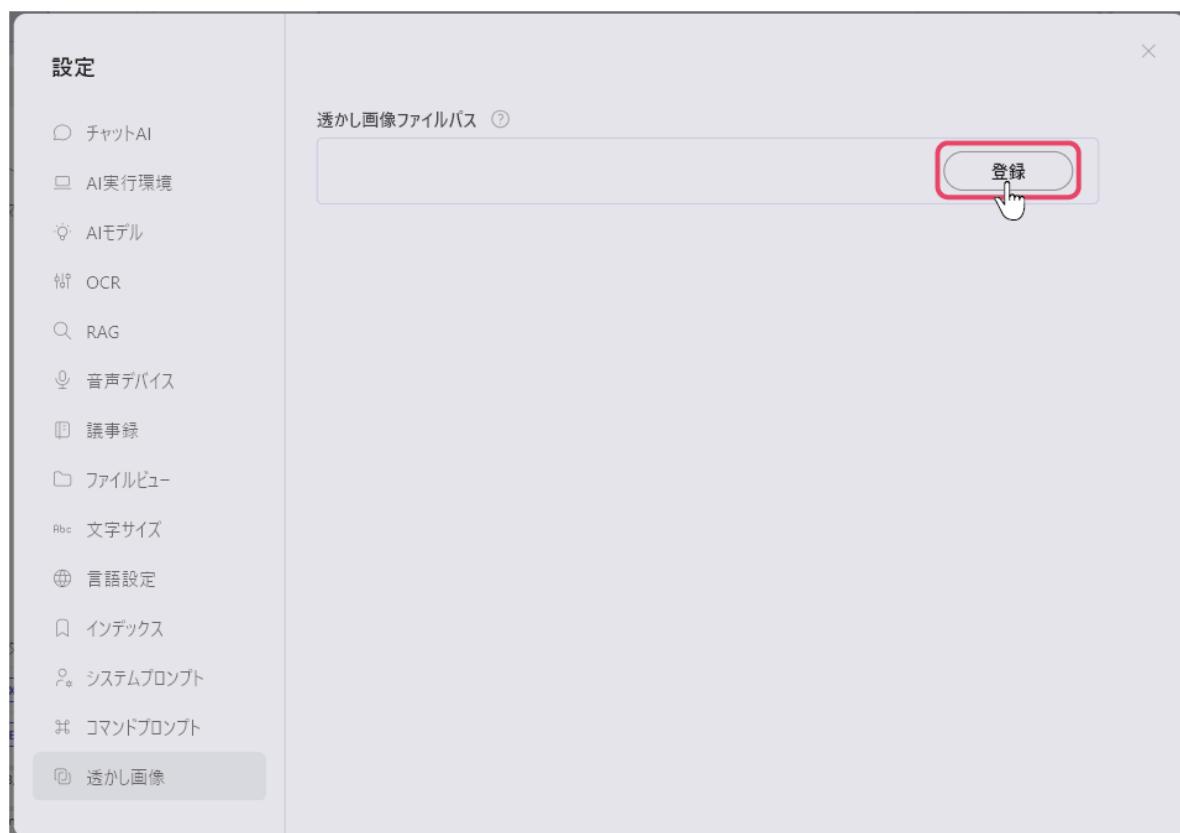


1. 画面右上にある歯車アイコンを押します。

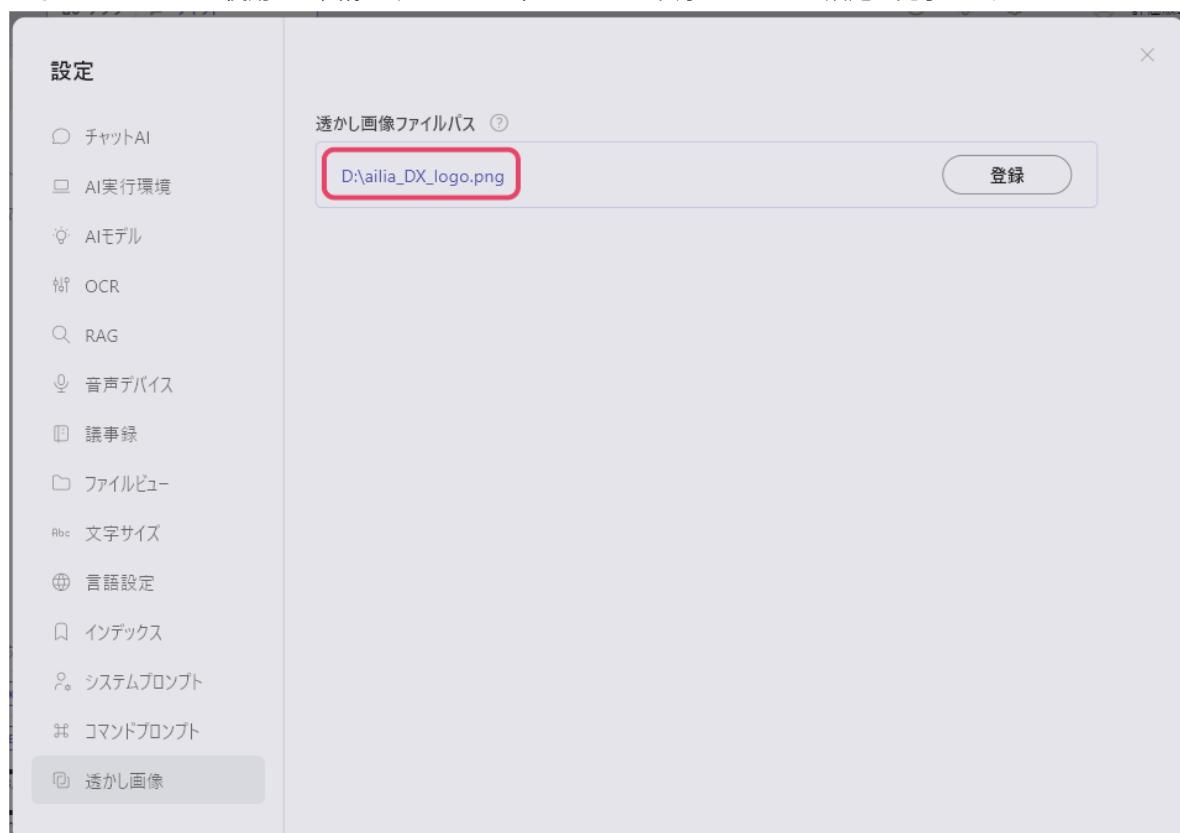


2. 「透かし画像」の項目内にある透かし画像ファイルパスの「登録」を選択し、表示されるエクスプローラー内にて使用する画像を選択します。

カスタムLLMの設定

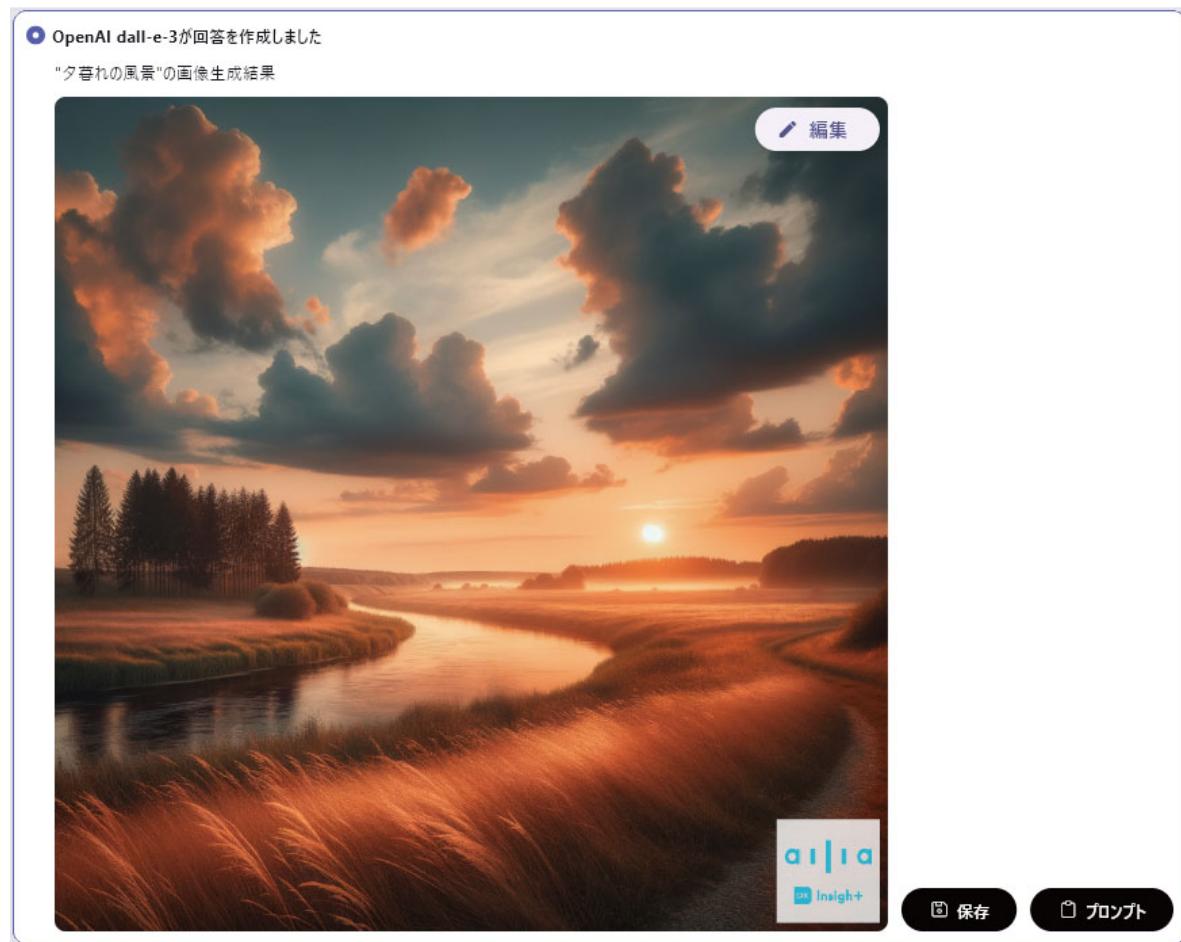


3. ウォーターマークに使用する画像ファイルのディレクトリが表示されれば、設定は完了です。



カスタムLLMの設定

4. 画像を生成すると、画像の右下にウォーターマークが追加されます。



[次のページへ >](#)

[一覧に戻る](#)

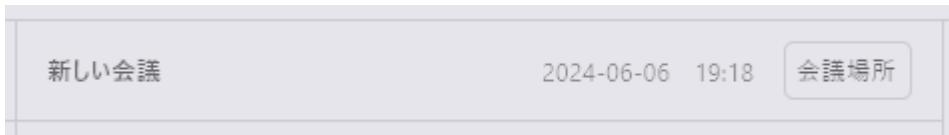
議事録を取る

音声認識AIによって、手を動かすことなく議事録を取ることができます。

1. 画面左上のアプリメニューから「議事録作成」をクリックし、議事録画面を開きます。



2. 画面右上の「会議場所」に会議場所を入力します。会議名を変更することもできます。



3. 画面下部の「記録を開始」ボタンをクリックし、議事録の記録を開始します。



4. 音声認識に必要なAIモデルのダウンロードがお済みでない場合は、ダウンロードが開始されます。ダウンロード完了後より利用可能となります。

5. 記録を終えるタイミングで、画面下部の「記録を停止」ボタンをクリックし、議事録を保存します。



記録した議事録を確認する

カスタムLLMの設定

議事録画面の左側に並ぶ議事録をクリックすると、記録した議事録を見ることができます。



記録された文の右側にあるボタンをクリックすると、文の訂正ができます。



また、画面下部の「記録を開始」ボタンをクリックすることで、再度議事録を記録することができます。

音声ファイルから議事録を作成する

クエリバーの「マイク」をクリックし、「ファイル」に変更した後に「記録を開始」を選択することで、音声ファ

イルから議事録を作成することができます。



対応形式

議事録機能で標準で読み込めるファイルはwavファイルのみとなります。

別途、ffmpegをインストールすることで、mp3とmp4の読み込みが可能となります。

Windowsの場合、環境変数のPATHの通った場所、もしくは下記のフォルダにffmpeg.exeを配置してください。

```
/Users/[UserName]/Documents(*)/ailia DX Insight/ffmpeg/
```

macOSの場合、環境変数のPATHの通った場所、もしくは下記のフォルダのいずれかにffmpegを配置してください。

```
~/Documents(*)/ailia DX Insight/ffmpeg/
```

```
/usr/local/bin/
```

```
/opt/homebrew/bin/
```

```
/opt/local/bin/
```

議事録を書き出す

議事録画面の左側に並ぶ議事録を右クリックし、「内容をエクスポート」を選択すると、議事録をtxt形式のファイルとして書き出すことができます。



[次のページへ >](#)

[一覧に戻る](#)

音声入力を利用する場合

- 文章入力前のチャットボックス右のマイクアイコンをクリックします。



※当アプリをインストール後、初めてマイクアイコンをクリックした場合、音声認識に必要なAIモデルのダウンロードが開始されます。ダウンロード完了後より音声入力が利用可能となります。



- 質問内容をマイクに向かって話します
- 再度マイクボタンをクリックし音声入力を完了します。

[次のページへ >](#)

[一覧に戻る](#)

ローカルLLMの設定

GGUF形式で圧縮されたLLMモデルを使用することで、ailia DX Insightの中だけで簡単にローカルLLMとの接続をすることができます。

ローカルLLMを実行するために必要なPC条件は下記となります。

Windowsの場合

- 2bモデルの動作条件

GPU実行：Vulkan 1.1以降に対応したVRAMを4GB以上搭載したGPU

CPU実行：AVX2に対応したCPU、8GB以上のメモリ

- 9bモデルの動作条件

GPU実行：Vulkan 1.1以降に対応したVRAMを6GB以上搭載したGPU

CPU実行：AVX2に対応したCPU、8GB以上のメモリ

MacOSの場合

- 2bモデルの動作条件

8GB以上のメモリを搭載したApple Silicon

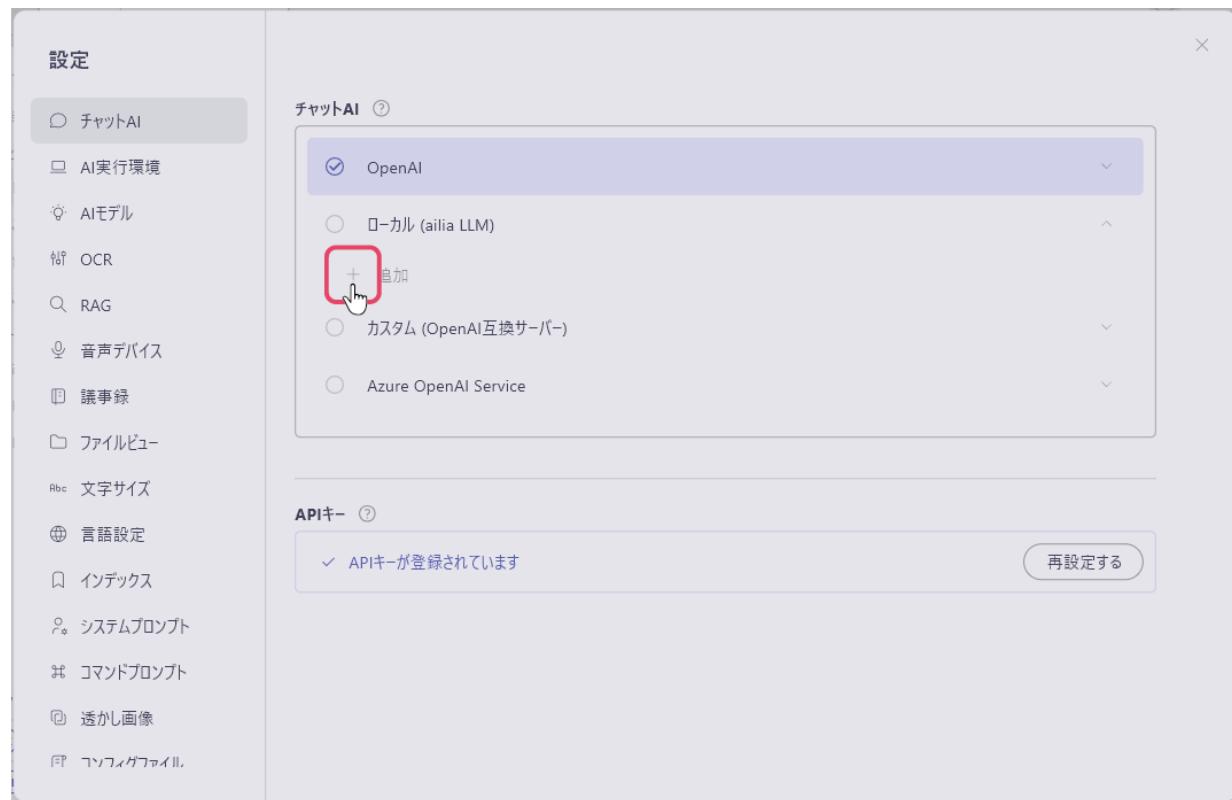
- 9bモデルの動作条件

16GB以上のメモリを搭載したApple Silicon

ailia DX Insight内のローカルLLMの構築

- 設定  中のチャットAIの項目を選択し、「ローカル（ailiaLLM）」の「+追加」を選択します。

カスタムLLMの設定



1. ローカルLLMモデル登録用ウィンドウが開きます。まずははじめに「Local LLM モデルタイプ」の項目を選択します。



規定のLLMから選択する

規定のLLMから指定をするだけで、ローカルモデルをダウンロードすることができます。

alia DX Insightでは「gemma2 2b / gemma2 9b」の二つから選択可能です。

LLMを選択すると、名前と最大トークン長が自動で入力されます。（説明の項目は必要に応じて任意で入力してください）



カスタムLLMの設定

右下の完了をクリックすると、「モデルのダウンロードを開始します。」というメッセージウィンドウが表示され、「はい」を選択すると、ダウンロードが開始されます。



GGUFファイルを使用する

カスタムLLMの設定

「Local LLM モデルタイプ」の項目を「GGUF model」に変更することで、「GGUF モデルパス」の項目が表示さ



れます。

カスタムLLMの設定

各項目を入力します。

Local LLM モデル

Local LLM モデルタイプ

GGUF model

GGUF モデルパス

D:\LLM\gemma-2-2b-it-Q4_K_M.gguf

名前

GGUF model

説明

最大トークン長

8192

キャンセル 完了

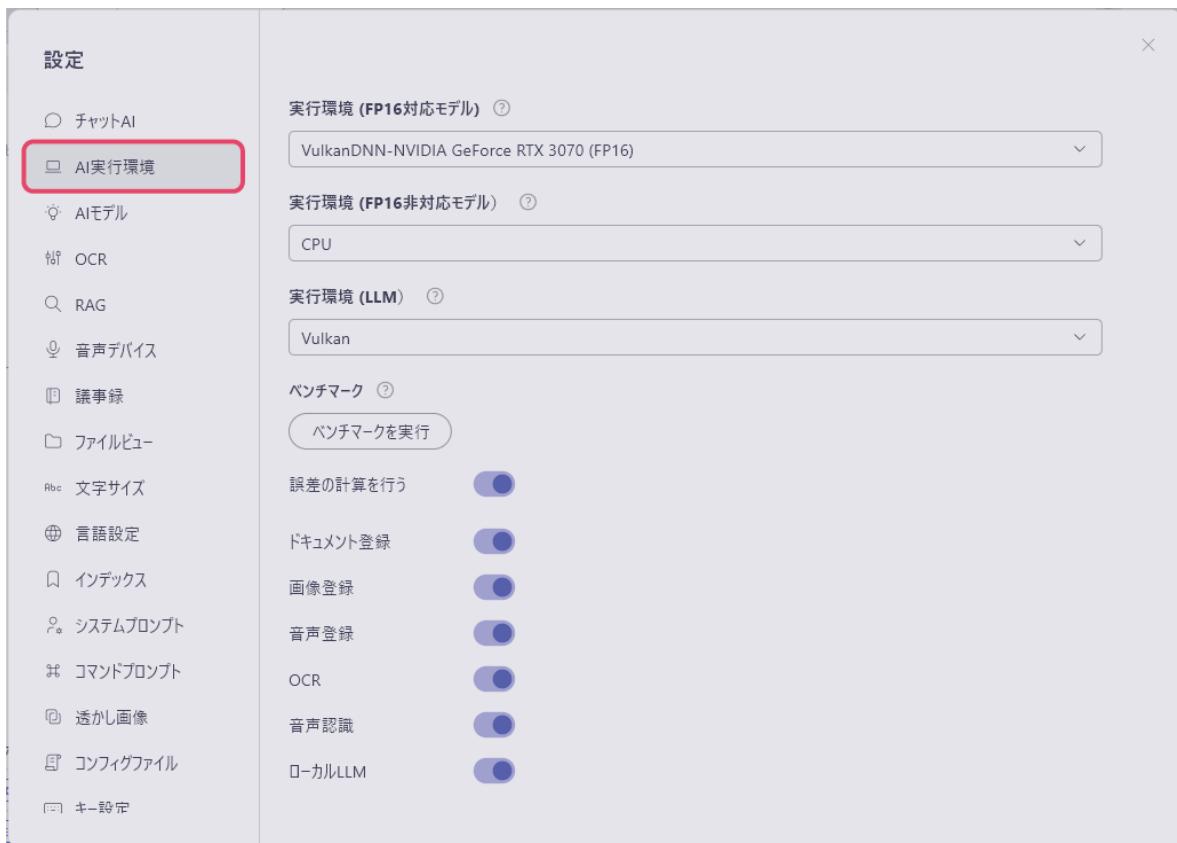
- local LLM モデルタイプ : GGUF modelを選択
- GGUFモデルパス : GGUFファイルのディレクトリを入力
- 名前 : 使用するLLMの名称 (UI表示にのみ使用するため何でも構わない)
- 説明 : 必要に応じてメモとして使用
- 最大トークン長 : ダウンロードしたモデルに応じたものを入力

[次のページへ >](#)

ベンチマーク機能

ailia DX Insight内の機能を使い、お使いのハードウェアでのAI実行環境のベンチマークを測定することができます。数多くあるLLMの中からご利用されるLLMに迷った場合や、ご利用されるPCを決める際の指標にご活用ください。

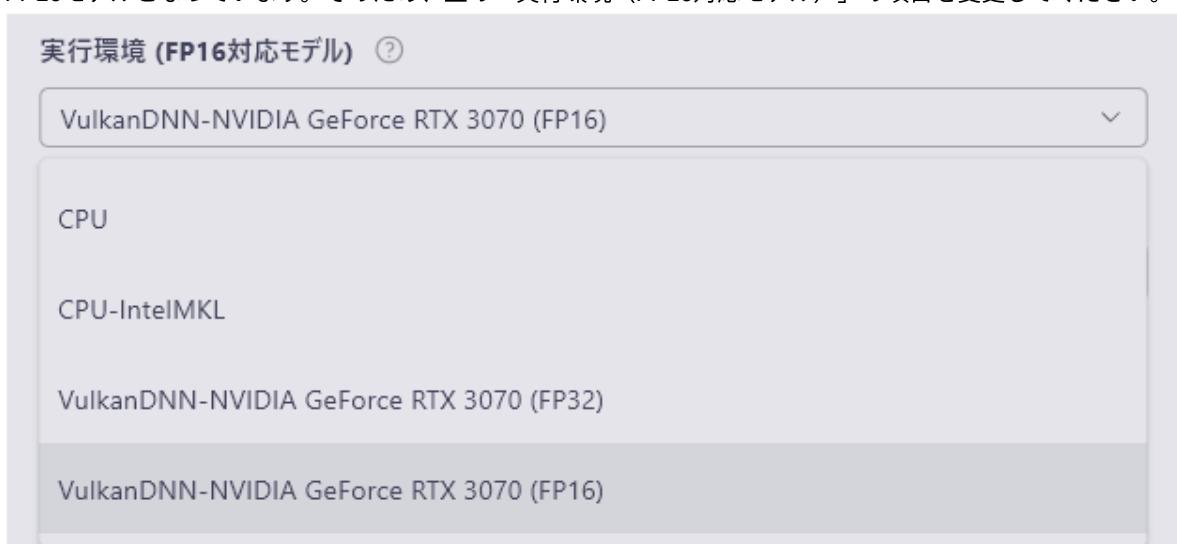
1. 設定 の「AI実行環境」の項目を選択します。



2. 実行環境を選択します。

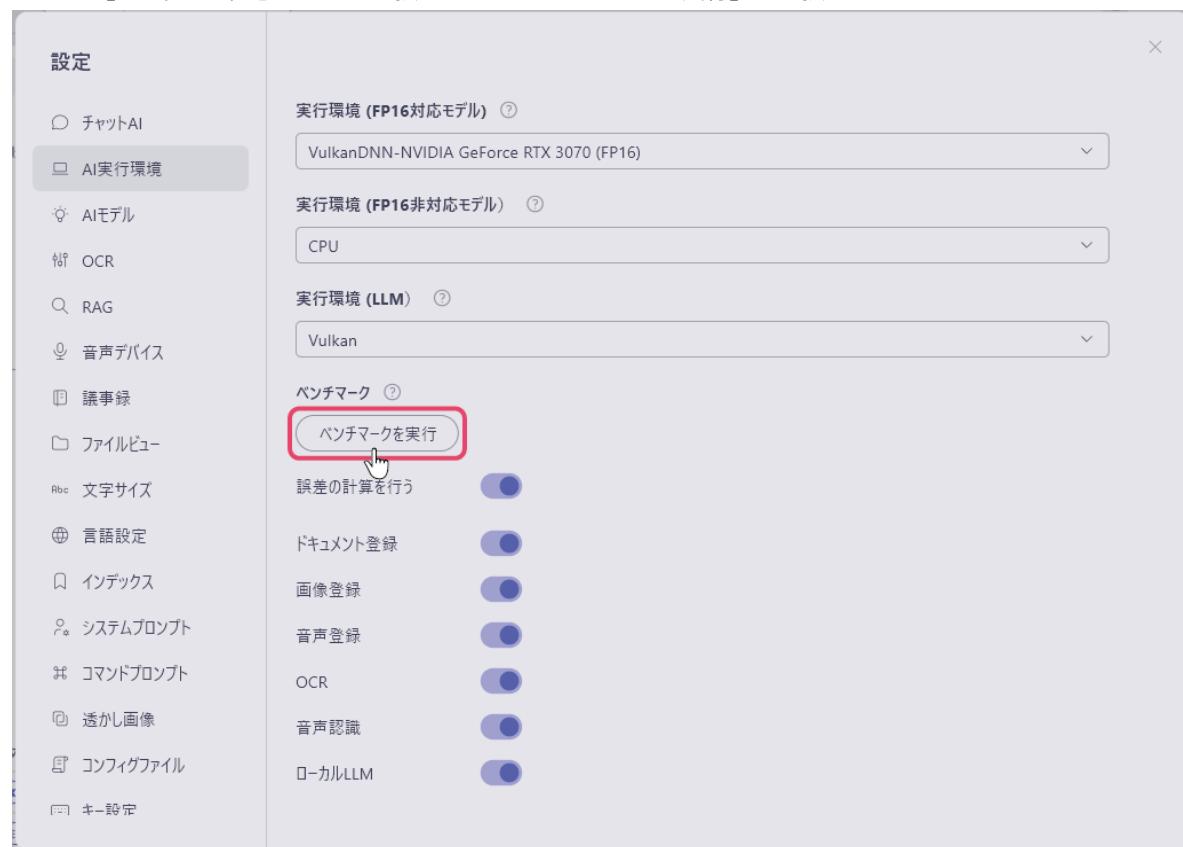
GPUとCPUの選択肢が表示されるので、生成時に使用するものを選択します。

画像検索および音声検索のクエリ処理、リランクはFP16に非対応のモデルです。それ以外の、文章登録、画像登録、音声登録、OCR、音声認識のモデルはFP16に対応しています。ベンチマーク対象のモデルはすべてFP16モデルとなっています。そのため、上の「実行環境（FP16対応モデル）」の項目を変更してください。

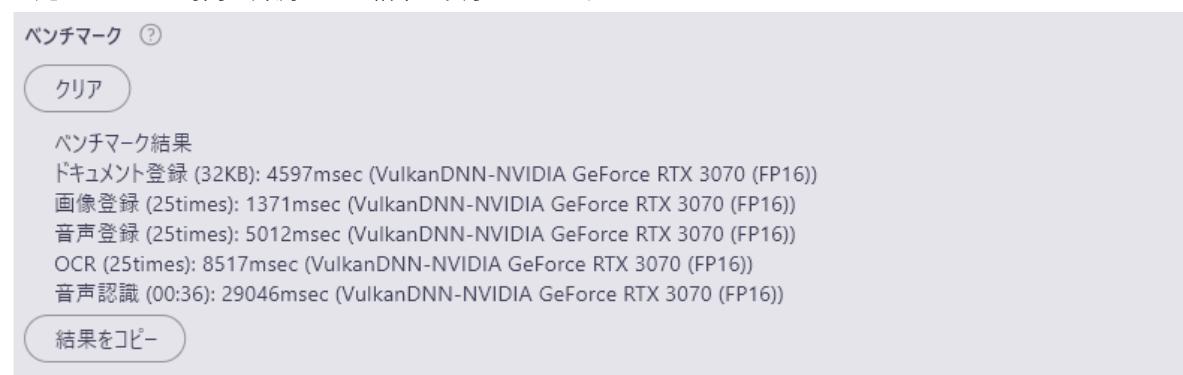


カスタムLLMの設定

ベンチマークを計測する項目を「ドキュメント登録」「画像登録」「音声登録」「OCR」「音声認識」「ローカルLLM」の中から任意のものを選択して、「ベンチマークを実行」を選択します。



3. AI処理にかかる時間が計測され、結果が表示されます。



[次のページへ >](#)

セキュアな環境構築

ailia DX Insightはオフラインとオンラインのハイブリッドです。

質問に対する情報検索はオフラインで動作するため安全かつ低成本です。



カスタムLLM

ailia DX Insightは、文章の生成をChatGPTで行います。 ChatGPTを使用する場合、クラウドへのデータ転送を伴うので、使用を制限せざるを得ないケースがあります。 ailia DX insightでは、ローカルに構築したカスタムLLMを接続することができます。これにより、完全にセキュアな環境でailia DX Insightを動作させることができます。



カスタムLLMの設定方法は[こちら](#)から

次のページへ >

カスタムLLMの設定

ailia DX Insightではローカルに構築したカスタムLLMを使用することで完全にオフライン環境でより安全な環境で動作させることができます。カスタムLLMはailia DX Insightを動作させるPC上で動かすことも出来ますし、別のPCやサーバーなどで動作させているものに接続することも可能です。

ChatGPTの互換サーバについて

ailia DX InsightはChatGPTの互換サーバに対応しています。

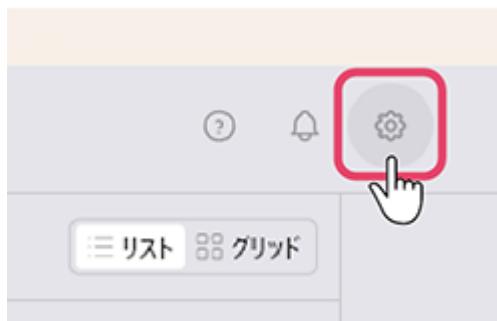
ローカルLLMを使用する方法には以下のものがあります。

- ・ [「LM Studio」を使用する方法](#)
- ・ [「Ollama」を使用する方法](#)
- ・ [Pythonの「fastchat」を使用する方法](#)
- ・ ailia DX Enterpriseを使用する方法（7月末リリース予定）
- ・ ailia DX Insight内蔵ローカルLLMを使用する方法（8月末リリース予定）

カスタムLLMのクライアントの設定

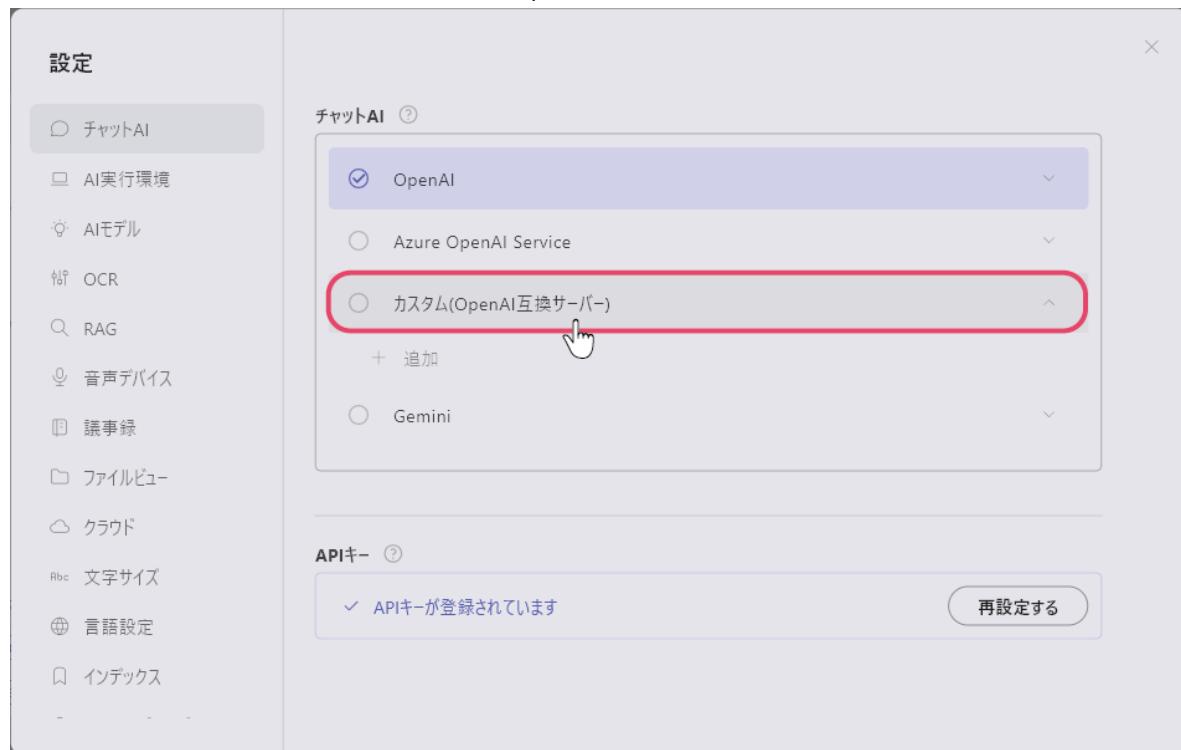
設定画面の呼び出し

1. ailia DX Insightの初期画面にて、右上の歯車アイコンをクリックして設定ウィンドウを表示させます。



カスタムLLMの設定

2. 「チャットAI」の項目の中にある「カスタム（OpenAIサーバー）」をクリックし、「+追加」を選択します。



カスタムモデル登録

カスタムLLMの設定

カスタムモデル登録ウィンドウが開きます。各項目を記入します。

カスタム(OpenAI互換サーバー)

名前

説明

モデル

URL

最大トークン長

キャンセル 追加する

- 名前：使用するLLMの名称 (UI表示にのみ使用するため何でも構わない)
- 説明：必要に応じてメモとして使用
- モデル：OpenAI互換APIサーバーで指定したモデル名 (gpt-3.5-turbo等)
- URL(*)：OpenAI互換APIサーバーが発信しているIPアドレスとport番号 (サーバー公開が–host 192.168.1.10 – port 8000の場合、<http://192.168.1.10:8000>)
- 最大トークン長：モデルの対応している最大トークン数を設定 (4096以上の値を設定する、この値に応じてRAGのtopKが決定される)

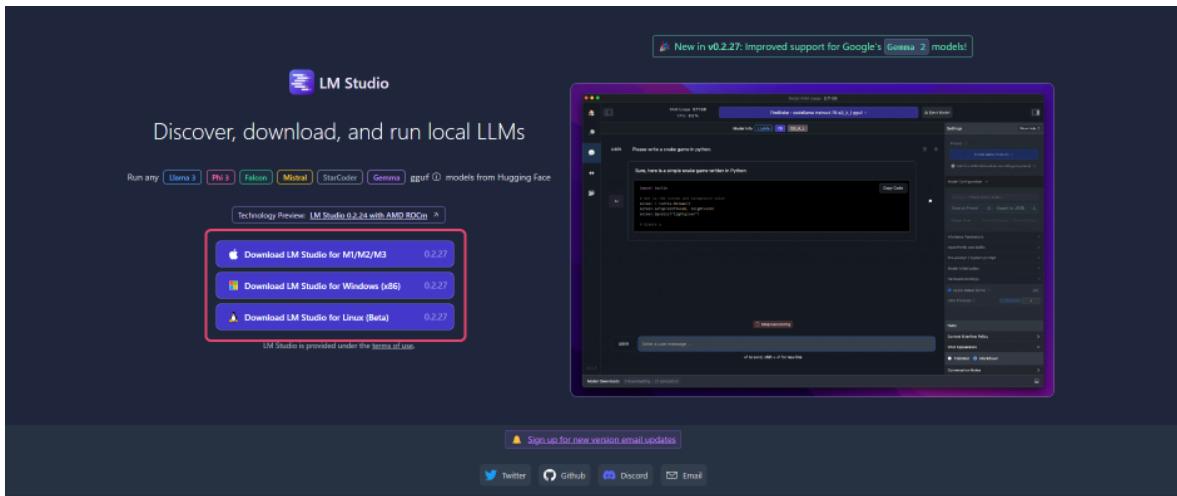
(*)未記入の場合、HTTPの接続エラーになります。

[次のページへ >](#)

LMスタジオを使用したカスタムLLM構築

LMスタジオのインストール

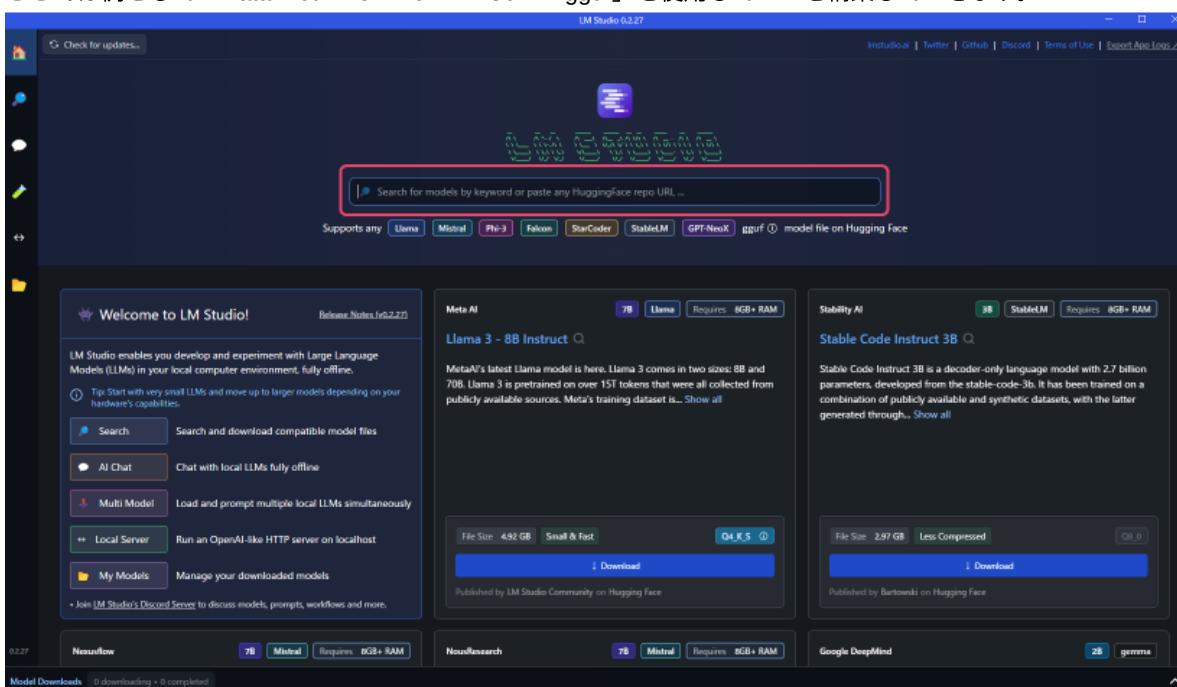
1. <https://lmstudio.ai/>よりLMスタジオをインストールします。



カスタムLLMのダウンロード

1. LMスタジオを起動させ、画面中央にある検索バーからLLMを検索します。

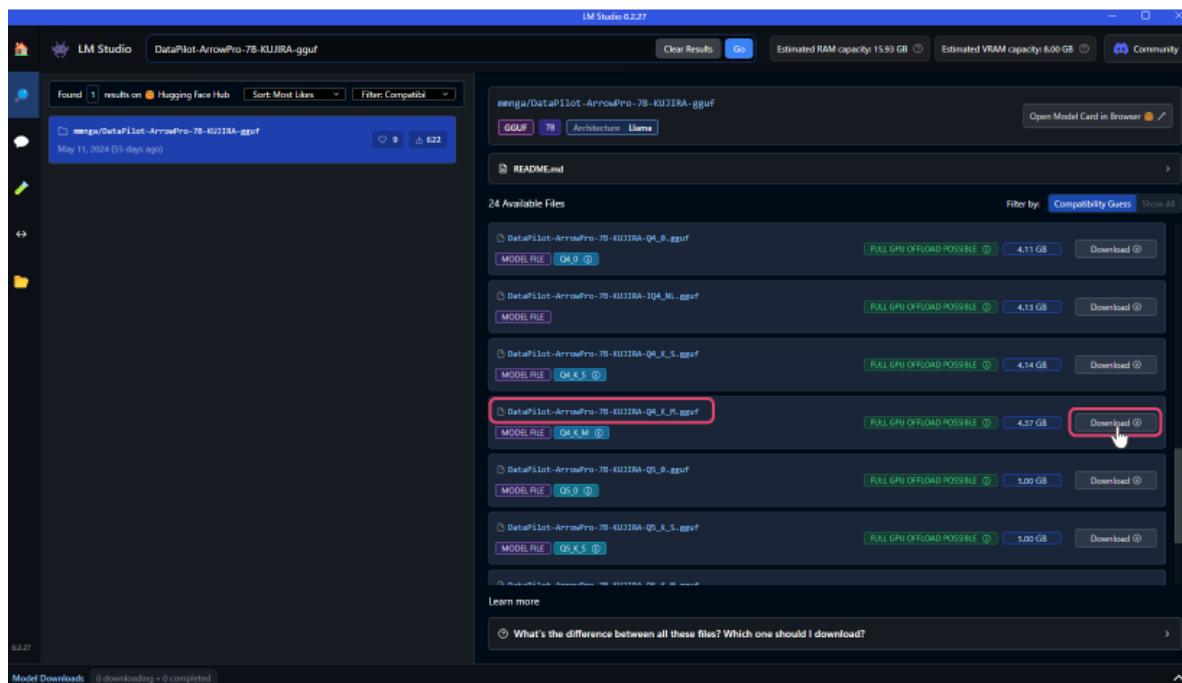
ここでは例として「DataPilot-ArrowPro-7B-KUJIRA-gguf」を使用してLLMを構築していきます。



2. 検索をすると画面右側に該当する検索候補が複数表示されます。

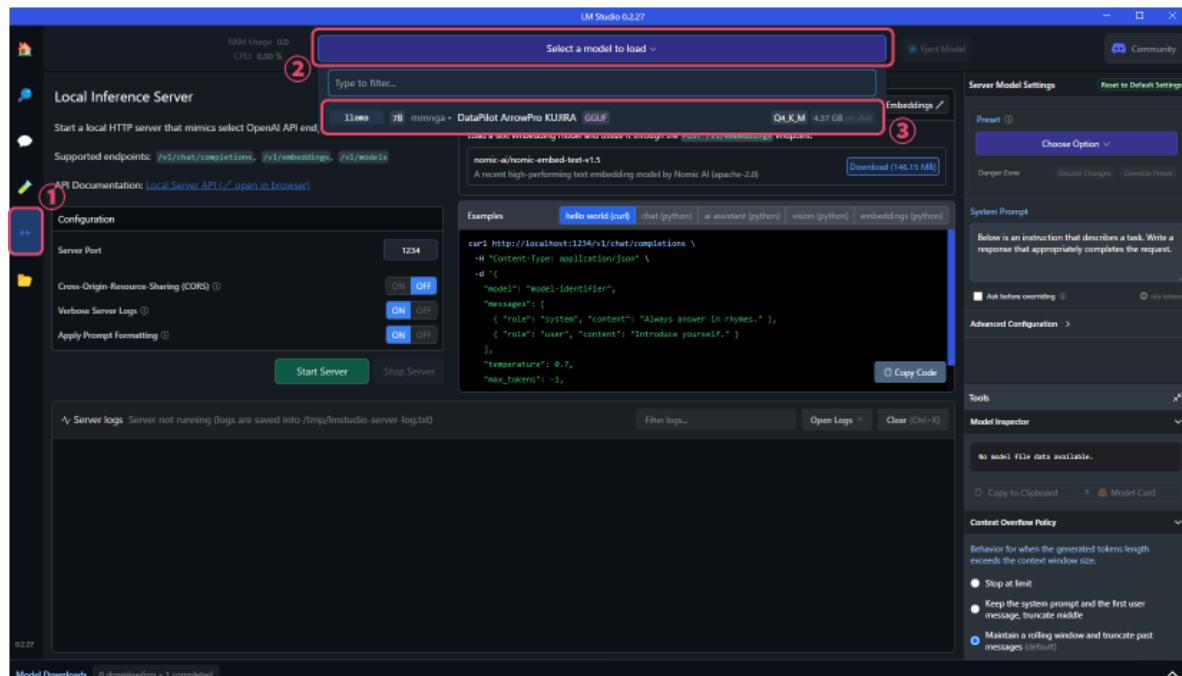
目的のLLMを見つけたら「Download」をクリックしてモデルのダウンロードを開始します。

カスタムLLMの設定



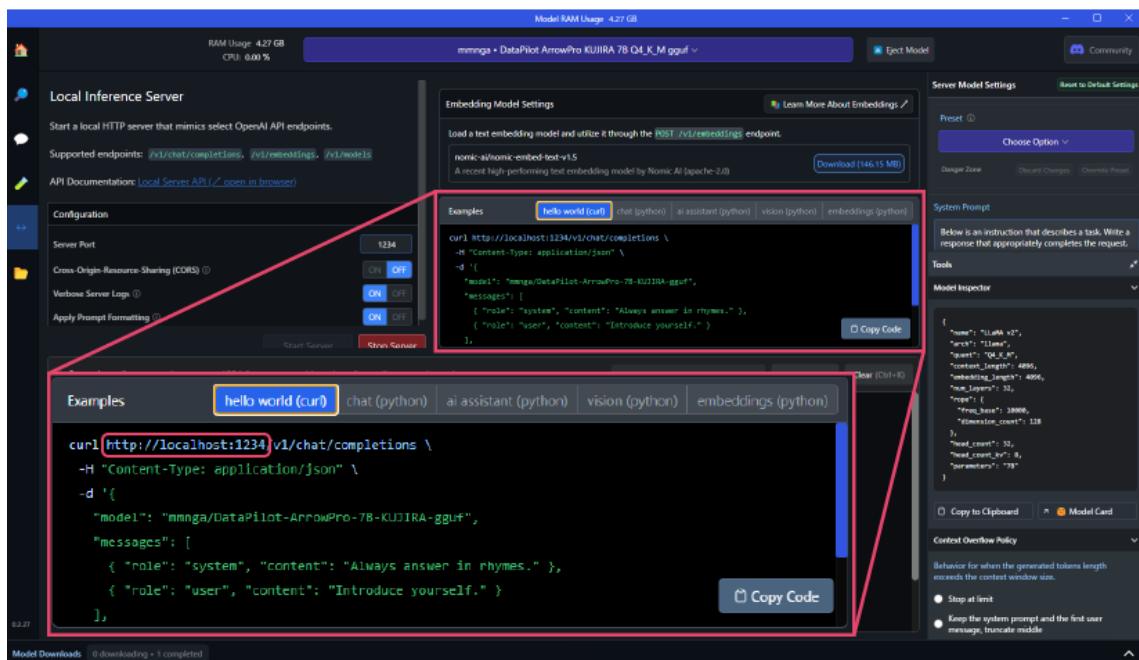
カスタムLLMの立ち上げ

- ダウンロード完了後、左サイドバー内にある「↔」ボタンをクリックして画面のモードを変更します。画面上部にある「Select a model to load」をクリックすると、プルダウンでダウンロード済みのLLMが一覧で表示されるので、ダウンロードしたモデルを選択します。

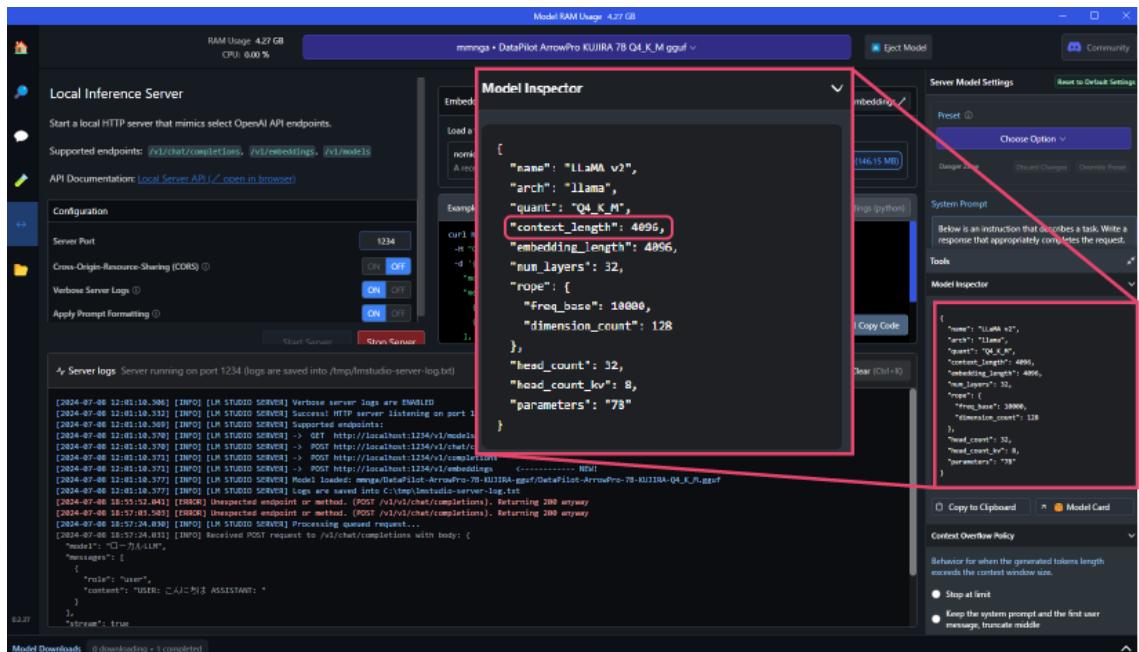


- 画面中央付近にailia DX Insight内の設定で使用するカスタムLLMのURLが表示されます。（通常はlocalhost:1234に設定されます）

カスタムLLMの設定



- 画面右側「Model Inspector」内に記載の context_length の数値がailia DX Insight内の設定で使用する「最大トークン数」となります。



aillia DX InsightでカスタムLLMの登録

カスタムLLMの設定

詳しい設定画面の呼び出し方は[こちら](#)をご参照ください。

カスタム(OpenAI互換サーバー)

名前

説明

モデル

URL

最大トークン長

キャンセル完了

- 名前：使用するLLMの名称 (UI表示にのみ使用するため何でも構わない)
- 説明：必要に応じてメモとして使用
- モデル：ダウンロードしたモデルの名称
- URL(*)：カスタムLLMの立ち上げで表示されたURL（通常はlocalhost:1234に設定されます）
- 最大トークン長：カスタムLLMの立ち上げで表示された context_length の数値

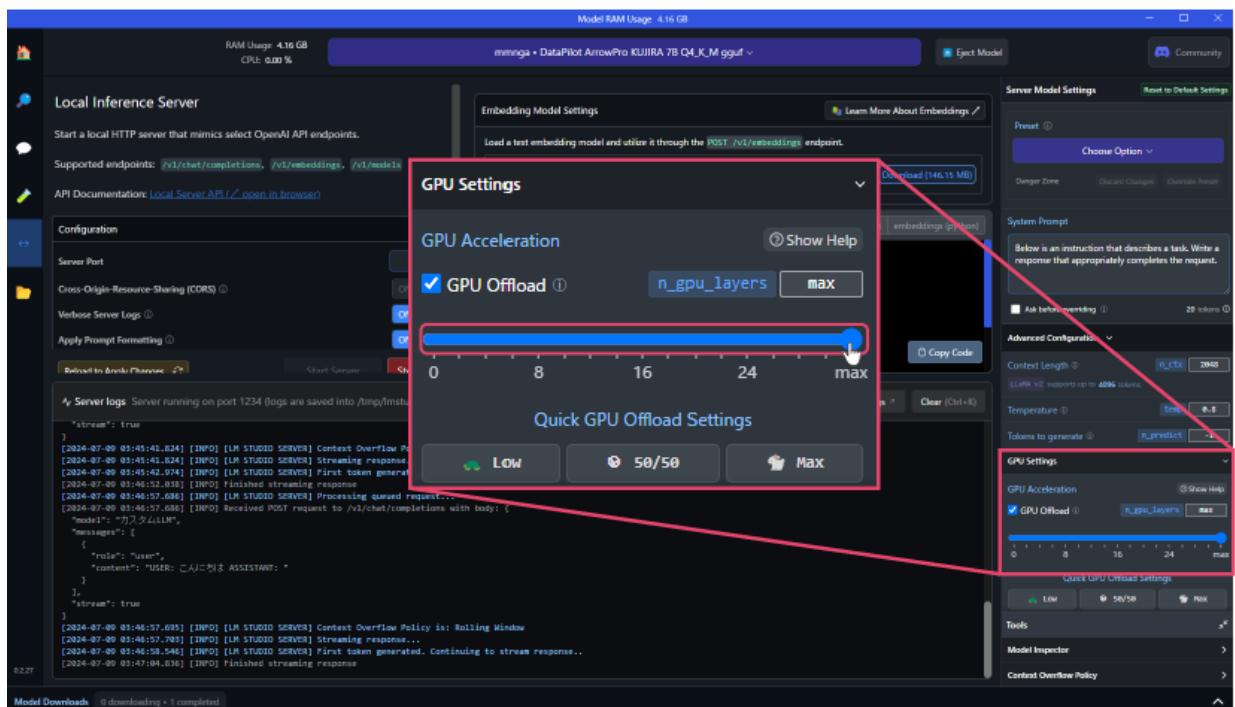
(*)未記入の場合、HTTPの接続エラーになります。

ローカルLLMの処理速度を向上させる方法

画面右側の「Advanced Configuration」内にある「GPU Settings」のパラメータ値を変更することでGPUメモリを使用する割合を変更することができます。

GPU Offload/パラメータを最大に設定すると、GPUメモリを使用した高速処理をおこなうことができます。

カスタムLLMの設定



次のページへ >

Ollamaを使用したカスタムLLM構築

Ollamaのインストール

<https://ollama.com>よりOllamaをインストールしてください。

The screenshot shows the Ollama website's download page for Windows. At the top, there is a navigation bar with links for Blog, Discord, GitHub, a search bar labeled "Search models", and buttons for "Models", "Sign in", and "Download". Below the navigation is a large cartoon llama logo. The main text reads "Get up and running with large language models." followed by a subtext: "Run [Llama 3](#), [Phi 3](#), [Mistral](#), [Gemma 2](#), and other models. Customize and create your own." A prominent "Download ↓" button is centered below the text. Below the button, a note says "Available for macOS, Linux, and Windows (preview)".

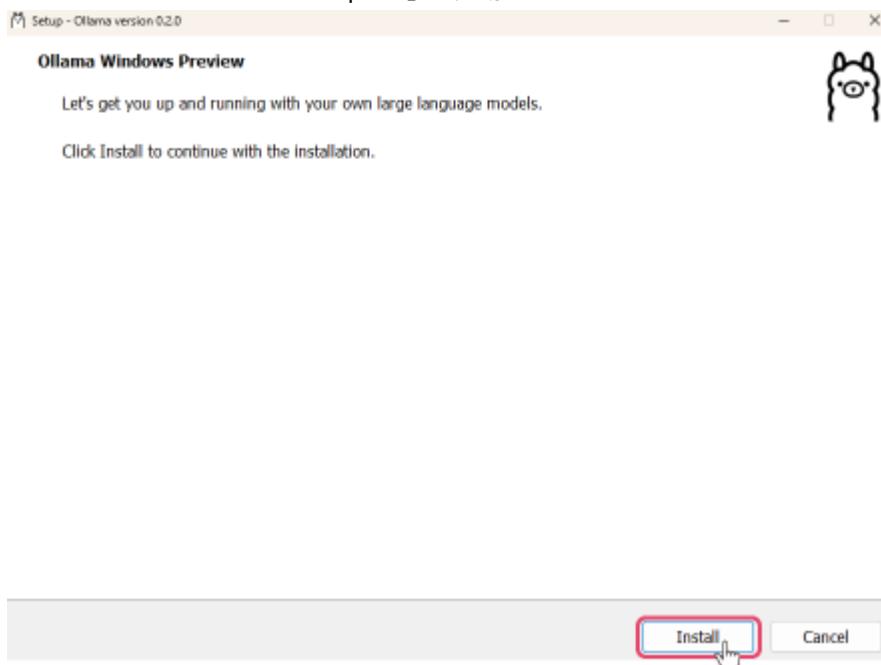
© 2024 Ollama

[Blog](#) [Docs](#) [GitHub](#) [Discord](#) [X \(Twitter\)](#) [Meetups](#)

Ollamaの起動（Windows）

カスタムLLMの設定

1. インストールした「OllamaSetup.exe」を起動し、インストールをします。



2. インストールが完了したら、コマンドプロンプトを起動してパスが通っていることを確認します。
コマンドプロンプトで `ollama --help` を実行し、次のようなヘルプ画面が表示されることを確認してください。

```
Large language model runner

Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve      Start ollama
  create     Create a model from a Modelfile
  show       Show information for a model
  run        Run a model
  pull       Pull a model from a registry
  push       Push a model to a registry
  list       List models
  ps         List running models
  cp         Copy a model
  rm         Remove a model
  help       Help about any command

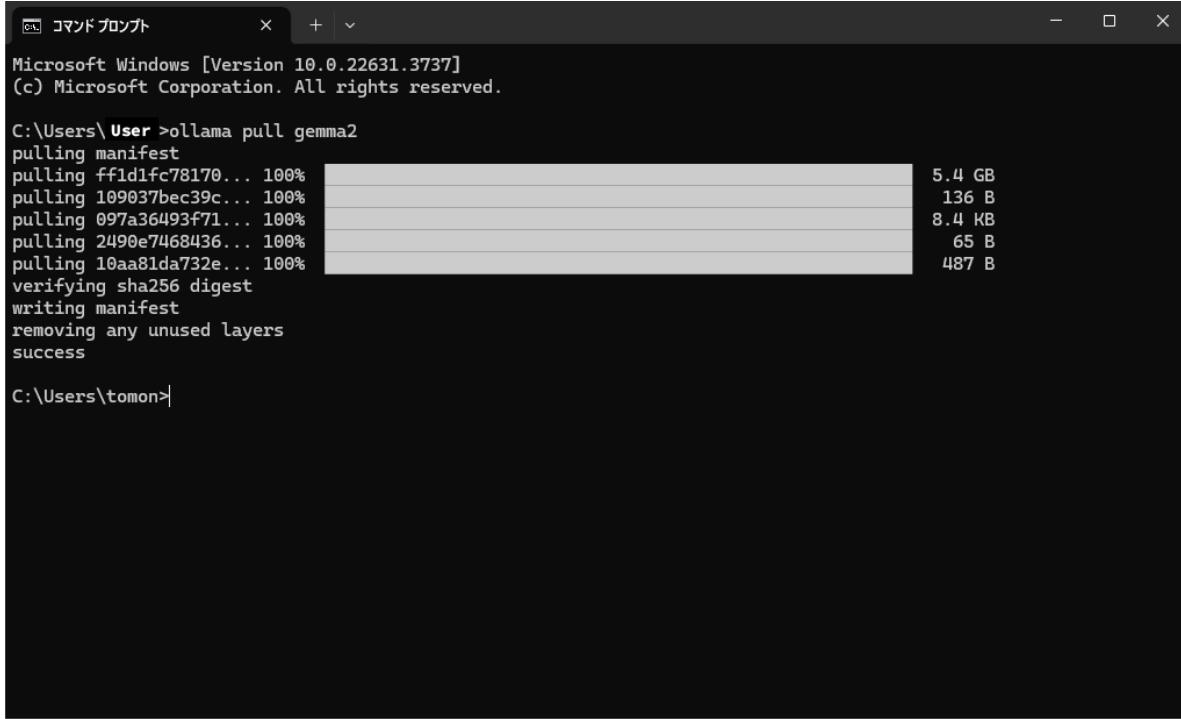
Flags:
  -h, --help    help for ollama
  -v, --version Show version information

Use "ollama [command] --help" for more information about a command.
```

LLMのダウンロード

カスタムLLMの設定

- ローカルで動かしたいLLMを ollama pull のコマンドでダウンロードします。例：Googleのgemma-2-9Bをダウンロードする場合、コマンドプロンプトに ollama pull gemma2 と入力します。



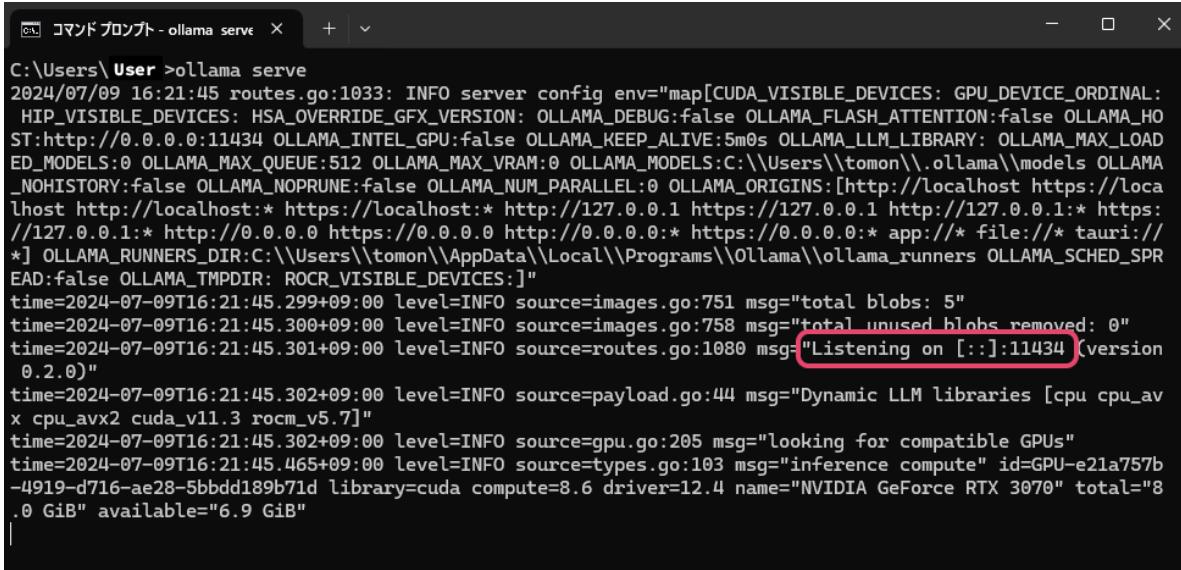
```
Microsoft Windows [Version 10.0.22631.3737]
(c) Microsoft Corporation. All rights reserved.

C:\Users\User>ollama pull gemma2
pulling manifest
pulling ff1d1fc78170... 100%
pulling 109037bec39c... 100%
pulling 097a36493f71... 100%
pulling 2490e7468436... 100%
pulling 10aa81da732e... 100%
verifying sha256 digest
writing manifest
removing any unused layers
success

C:\Users\tomon>
```

- 他のPCからもこのローカルLLMにアクセスしたい場合はコマンドプロンプトで set OLLAMA_HOST=0.0.0.0 を実行し、OLLAMA_HOSTの環境を「0.0.0.0」に設定します。

- コマンドプロンプトで ollama serve を実行することでLLMを起動させます。



```
C:\Users\User>ollama serve
2024/07/09 16:21:45 routes.go:1033: INFO server config env="map[CUDA_VISIBLE_DEVICES: GPU_DEVICE_ORDINAL: HIP_VISIBLE_DEVICES: HSA_OVERRIDE_GFX_VERSION: OLLAMA_DEBUG:false OLLAMA_FLASH_ATTENTION:false OLLAMA_HOST:http://0.0.0.0:11434 OLLAMA_INTEL_GPU:false OLLAMA_KEEP_ALIVE:5m0s OLLAMA_LLM_LIBRARY: OLLAMA_MAX_LOAD_ED_MODELS:0 OLLAMA_MAX_QUEUE:512 OLLAMA_MAX_VRAM:0 OLLAMA_MODELS:C:\\\\users\\\\tomon\\\\.ollama\\\\models OLLAMA_NOHISTORY:false OLLAMA_NOPRUNE:false OLLAMA_NUM_PARALLEL:0 OLLAMA_ORIGINS:[http://localhost https://localhost http://localhost:* https://localhost: http://127.0.0.1 https://127.0.0.1 http://127.0.0.1:* https://127.0.0.1:* http://0.0.0.0 https://0.0.0.0 http://0.0.0.0:* https://0.0.0.0:* app:///* file:///* tauri:///*] OLLAMA_RUNNERS_DIR:C:\\\\users\\\\tomon\\\\AppData\\\\Local\\\\Programs\\\\Ollama\\\\ollama_runners OLLAMA_SCHED_SPAED:false OLLAMA_TMPDIR: ROCR_VISIBLE_DEVICES:]"
time=2024-07-09T16:21:45.299+09:00 level=INFO source=images.go:751 msg="total blobs: 5"
time=2024-07-09T16:21:45.300+09:00 level=INFO source=images.go:758 msg="total unused blobs removed: 0"
time=2024-07-09T16:21:45.301+09:00 level=INFO source=routes.go:1080 msg="Listening on [:]:11434 (version 0.2.0)"
time=2024-07-09T16:21:45.302+09:00 level=INFO source=payload.go:44 msg="Dynamic LLM libraries [cpu cpu_avx cpu_avx2 cuda_v11.3 rocm_v5.7]"
time=2024-07-09T16:21:45.302+09:00 level=INFO source=cpu.go:205 msg="looking for compatible GPUs"
time=2024-07-09T16:21:45.465+09:00 level=INFO source=types.go:103 msg="inference compute" id=GPU-e21a757b-4919-d716-ae28-5bddd189b71d library=cuda compute=8.6 driver=12.4 name="NVIDIA GeForce RTX 3070" total="8.0 GiB" available="6.9 GiB"
```

- 「Listening on [:]:11434」と表示されている場合は他のPCからもアクセス可能です。
- 「Listening on 127.0.0.1:11434」の場合はそのPC内からのみアクセスできます。ここの「11434」の数値がポート番号となり、aillia DX Insight内の設定に使用します。

aillia DX InsightでカスタムLLMの登録

カスタムLLMの設定

詳しい設定画面の呼び出し方は[こちら](#)をご参考ください。

カスタム(OpenAI互換サーバー)

名前

説明

モデル

URL

最大トーカン長

キャンセル追加する

- 名前：使用するLLMの名称 (UI表示にのみ使用するため何でも構わない)
- 説明：必要に応じてメモとして使用
- モデル：ダウンロードしたモデルの名称 (今回の例ではgemma2)
- URL(*)： `ollama serve` で表示されたポートのURL (今回の例では<http://localhost:11434>)
- 最大トーカン長：gemma-2-9Bの場合、「8192」となります。

(*)未記入の場合、HTTPの接続エラーになります。

ダウンロードしたLLMの削除

LLMが容量を圧迫する場合や、間違えたモデルをダウンロードしてしまった場合、以下の手順でLLMを削除することができます。

1. コマンドプロンプトで `ollama list` を実行すると、ダウンロード済みのモデルが一覧で表示されます。
2. コマンドプロンプトで `ollama rm モデル名` を実行することでダウンロード済みのモデルを削除することができます。

カスタムLLMの設定

```
C:\Users\ User >ollama list
NAME          ID          SIZE      MODIFIED
gemma2:latest ff02c3702f32  5.4 GB  9 minutes ago

C:\Users\ User >ollama rm gemma2
```

[次のページへ >](#)

FastChatのインストール

サーバーを立てる端末にpythonモジュールとしてfschatをインストールします。 `pip3 install fschat`

サーバの起動

vicuna-7b-v1.5をChatGPT API互換サーバーとして起動します。

```
python3 -m fastchat.serve.controller  
python3 -m fastchat.serve.model_worker -model-names "vicuna" -model-path lmsys/vicuna-7b-v1.5 -load-8bit  
python3 -m fastchat.serve.openai_api_server --host localhost --port 8000
```

aillia DX InsightでカスタムLLMの登録

詳しい設定画面の呼び出し方は[こちら](#)をご参考ください。

カスタム(OpenAI互換サーバー)

名前

説明

モデル

URL

最大トークン長

キャンセル 追加する

- 名前：使用するLLMの名称 (UI表示にのみ使用するため何でも構わない)
- 説明：必要に応じてメモとして使用
- モデル：OpenAI互換APIサーバーで指定したモデル名 (gpt-3.5-turbo等)

カスタムLLMの設定

- URL(*) : OpenAI互換APIサーバーが発信しているIPアドレスとport番号 (サーバー公開が–host 192.168.1.10 – port 8000の場合、<http://192.168.1.10:8000>)
- 最大トークン長 : モデルの対応している最大トークン数を設定 (4096以上の値を設定する、この値に応じてRAGのtopKが決定される)

(*)未記入の場合、HTTPの接続エラーになります。

[次のページへ >](#)

弊社推奨LLM

GPUメモリ8GB未満～8GBのPCをご利用になる場合に推奨するLLMの一覧をLLMのパラメータサイズ毎に超小型/小型/中型(それ以上)に分けて紹介します。

モデルの中にはパラメータ数の違う複数のモデルを提供しているものがあり、各モデルの名前の中にあるB(Billion)で表記されている数字から確認することができます。GPUメモリに余裕のある場合はLM Studio等からさらに上位モデルを探して使用することができます。

パラメータサイズ、RAM使用量、特徴から利用目的とPC環境にあったモデルをお選びください。（表記されているRAM使用量は量子化方式によって増減します）

超小型（0.5B～4B）

Qwen/Qwen2.5-0.5B-Instruct-GGUF (RAM使用量 1.18GB～)

Qwen/Qwen2.5-1.5B-Instruct-GGUF (RAM使用量 1.27GB～)

Qwen/Qwen2.5-3B-Instruct-GGUF (RAM使用量 1.44GB～)

特徴：アリババクラウドが提供するLLMとなっております。英語と中国語に加えて27の言語での学習データを活用し、自然言語理解、知識獲得、コーディング、数学、多言語対応などのベンチマークで高いパフォーマンスを示しています。0.5Bと非常に小さなモデルですので、知識量は少ないですが低スペックのPCで使用する際に最適です。

Qwen2.5では様々な専門分野での性能を大きく向上させ、多様なタスクに対応できる言語モデルとして進化しています。

DeepSeek-R1-Distill-Qwen-1.5B-GGUF (RAM使用量 1.37GB～)

特徴：中国のDeepSeekが開発した推論能力に特化した大規模言語モデルです。推論タスク、数学問題、コーディング、長文推論、クリエイティブライティング等の主要なベンチマークにおいて主要モデルに匹敵する高いパフォーマンスを示しています。

サイバーエージェントによって日本語データによる追加学習を行った「DeepSeek-R1-Distill-Qwen-14B/32B-Japanese」が公開されています。

TinySwallow-1.5B-Instruct-GGUF (RAM使用量 1.13GB～)

特徴：日本に拠点を置くAI企業の「Sakana AI」が発表した小規模日本言語モデルです。TAID（時間適応型保管蒸留）と呼ばれる新しい知識蒸留手法を用いて開発され、同規模の言語モデルの中で日本語機能を評価するベンチマークテストで最高性能を示しています。小規模言語モデルとしてスマートフォンでも動作できるほどに軽量で、リソースの限られた環境での動作が可能となっています。

lmstudio-community/gemma-2-9b-it-GGUF (RAM使用量 2.32GB～9.15GB)

bartowski/gemma-2-2b-it-GGUF (RAM使用量 1.79～3.05GB)

カスタムLLMの設定

特徴： Gemma2は、Googleの最先端のAIモデルであるGeminiと同じアーキテクチャを用いて構築されており、軽量でありながら高い性能を発揮します。特に270億パラメータモデルは、そのサイズクラスで最高の性能を発揮し、2倍以上の規模のモデルに匹敵する性能を実現しています。また、90億パラメータモデルも、同サイズの他のオープンモデルを上回る優れた性能を示しています。また、20億パラメータモデルはサイズに対してクラス最高のパフォーマンスとなりノートPC等での活用に適します。

microsoft/Phi-3-mini-4k-instruct-gguf (RAM使用量 2.99GB程)

QuantFactory/Phi-3-mini-128k-instruct-GGUF (RAM使用量 2.08～4.54GB)

特徴： Microsoft社が提供するLLMとなっております。3.8B（38億個のパラメータ）という非常に小型なモデルながら、2倍以上大きいモデルと同等の性能を有しています。現在、4Kトークンと128Kトークンに対応した2つのモデルが公開されています。

小型（7B～9B）

Qwen/Qwen2.5-7B-Instruct-1M-GGUF (RAM使用量 1.97GB～)

特徴： アリババクラウドが提供するLLMとなっております。英語と中国語に加えて27の言語での学習データを活用し、自然言語理解、知識獲得、コーディング、数学、多言語対応などのベンチマークで高いパフォーマンスを示しています。Qwen2.5では様々な専門分野での性能を大きく向上させ、多様なタスクに対応できる言語モデルとして進化しています。

Qwen2.5-1MはQwen2.5-Turboを元に開発されており、最大100万トークンのコンテキストの処理が可能となりました。

Qwen/Qwen2.5-7B-Instruct-GGUF (RAM使用量 1.14GB～)

特徴： アリババクラウドが提供するLLMとなっております。英語と中国語に加えて27の言語での学習データを活用し、自然言語理解、知識獲得、コーディング、数学、多言語対応などのベンチマークで高いパフォーマンスを示しています。Qwen2.5では様々な専門分野での性能を大きく向上させ、多様なタスクに対応できる言語モデルとして進化しています。

DeepSeek-R1-Distill-Qwen-7B-GGUF (RAM使用量 2.23GB～)

DeepSeek-R1-Distill-Llama-8B-GGUF (RAM使用量 2.04GB～)

特徴： 中国のDeepSeekが開発した推論能力に特化した大規模言語モデルです。推論タスク、数学問題、コーディング、長文推論、クリエイティブライティング等の主要なベンチマークにおいて主要モデルに匹敵する高いパフォーマンスを示しています。サイバーエージェントによって日本語データによる追加学習を行った「DeepSeek-R1-Distill-Qwen-14B/32B-Japanese」が公開されています。

mmnga/umiyuki-Umievo-itr012-Gleipnir-7B-gguf (RAM使用量 2.33～6.34GB)

特徴： Japanese-Starling-ChatV-7B、Ninja-v1-RP-expressive-v2、Vecteus-v1、Japanese-Chat-Umievo-itro04-7b4の4モデルを統合することで完成した日本語性能の高いLLMです。

mmnga/DataPilot-ArrowPro-7B-KUJIRA-gguf (RAM使用量 2.26～3.49GB程)

特徴： オープンソースのLLM「NTQAI/chatntq-ja-7b-v1.0」を基にして、AIを活用したバーチャルYouTuber（AITuber）やAIアシスタントでの使用を目的に開発されました。日本語での性能が高く、会話の品質においても高評価を得ています。「ArrowPro-7B-KillerWhale」は「DataPilot/ArrowPro-7B-KUJIRA」の強化版という位置づけです。

mmnga/Llama-3.1-8B-Instruct-gguf (RAM使用量 2.26～5.79GB程)

特徴： Meta社が提供するこれまで最も先進的かつ高性能なLLMとなっております。Llama 3.1は128Kのコンテキスト長、最先端のツール使用、推論機能の強化により、長文テキストの要約、多言語会話エージェント、コーディング支援などの高度なユースケースをサポートしています。

mmnga/Llama-3-ELYZA-JP-8B-gguf (RAM使用量 2.73～8.75GB程)

特徴： 東大松尾研究室発のAIカンパニー株式会社ELYZA(イライザ)が提供する日本語に特化したLLMとなっております。Meta社のLlama 3 8B-Instructをベースに日本語の大規模データセットを用いて訓練され、日本語の文法、語彙、文化的背景に精通しており、日本語特有の表現やニュアンスを正確に理解し、洗練された日本語文章を生成することができます。

中型（それ以上）

Qwen/Qwen2.5-14B-Instruct-1M-GGUF

特徴： アリババクラウドが提供するLLMとなっております。英語と中国語に加えて27の言語での学習データを活用し、自然言語理解、知識獲得、コーディング、数学、多言語対応などのベンチマークで高いパフォーマンスを示しています。Qwen2.5では様々な専門分野での性能を大きく向上させ、多様なタスクに対応できる言語モデルとして進化しています。Qwen2.5-1MはQwen2.5-Turboを元に開発されており、最大100万トークンのコンテキストの処理が可能となりました。

Qwen/Qwen2.5-14B-Instruct-GGUF

Qwen/Qwen2.5-32B-Instruct-GGUF

Qwen/Qwen2.5-72B-Instruct-GGUF

特徴： アリババクラウドが提供するLLMとなっております。 英語と中国語に加えて27の言語での学習データを活用し、自然言語理解、知識獲得、コーディング、数学、多言語対応などのベンチマークで高いパフォーマンスを示しています。 Qwen2.5では様々な専門分野での性能を大きく向上させ、多様なタスクに対応できる言語モデルとして進化しています。

DeepSeek-R1-Distill-Qwen-14B-GGUF

DeepSeek-R1-Distill-Qwen-32B-GGUF

DeepSeek-R1-Distill-Llama-70B-GGUF

特徴： 中国のDeepSeekが開発した推論能力に特化した大規模言語モデルです。 推論タスク、数学問題、コーディング、長文推論、クリエイティブライティング等の主要なベンチマークにおいて主要モデルに匹敵する高いパフォーマンスを示しています。 サイバーエージェントによって日本語データによる追加学習を行った「DeepSeek-R1-Distill-Qwen-14B/32B-Japanese」が公開されています。

mmnga/ELYZA-japanese-Llama-2-13b-fast-instruct-gguf (RAM使用量 6.13 ~14.31GB)

特徴： 東大松尾研究室発のAIカンパニー株式会社ELYZA(イライザ)が提供する日本語に特化したLLMとなっております。 Meta社のLlama 2 13Bをベースに日本語の大規模データセットを用いて訓練され、 日本語の文法、語彙、文化的背景に精通しており、日本語特有の表現やニュアンスを正確に理解し、 洗練された日本語文章を生成することができます。

※まもなくLlama3ベースの新しいELYZAモデルが公開される予定です（2024年6月27日現在）

lmstudio-community/gemma-2-27b-it-GGUF (RAM使用量 8~29.65GB)

特徴： Gemma2は、Googleの最先端のAIモデルであるGeminiと同じアーキテクチャを用いて構築されており、 軽量でありながら高い性能を発揮します。 特に270億パラメータモデルは、そのサイズクラスで最高の性能を発揮し、 2倍以上の規模のモデルに匹敵する性能を実現しています。 また、90億パラメータモデルも、 同サイズの他のオープンモデルを上回る優れた性能を示しています。

andrewcanis/c4ai-command-r-v01-GGUF (RAM使用量 8~10.27GB)

特徴： Command-Rは、CohereForAIが作ったLLMで、35B（350億）のパラメータが含まれています。 このモデルは、大量の情報を元に、文章を生成したり、要約したり、質問に答えたりするのが得意です。 また、最大128kトークンの長文コンテキストに対応します。

pmysl/c4ai-command-r-plus-GGUF (RAM使用量 8~212.71GB)

カスタムLLMの設定

特徴： Command R+は、command Rの強化版となる104B（1,040億）のパラメータを含むLLMです。このモデルはエンタープライズ向けに優れた性能を発揮するようにデザインされ、誰でも使えるオープンソースでありながら、GPT-4 Turboに迫る性能を持っています。また、最大128Kの長文コンテキストに対応します。

(補足) command-r-v01／Command R Plusの利用について： command-r-v01／Command R Plusは非営利目的での使用を前提に、モデルの変更や配布をおこなうLLMとなります。商用利用（ビジネス利用）については、提供元の有償のAPIサービスを利用する必要がありますのでご注意ください。有償APIの取得については右記サイトよりおこなっていただくことができます（<https://cohere.com/command>）

Qwen/Qwen2-72B-Instruct-GGUF (RAM使用量 8~148.98GB)

アリババクラウドが提供するLLMとなっております。英語と中国語に加えて27の言語での学習データを活用し、自然言語理解、知識獲得、コーディング、数学、多言語対応などのベンチマークで高いパフォーマンスを示しています。7B・72Bモデルについては、最大128Kトーケンまでの長文コンテキストに対応します。

mmnaga/Llama-3.1-70B-Instruct-gguf (RAM使用量 7.66~9.07GB)

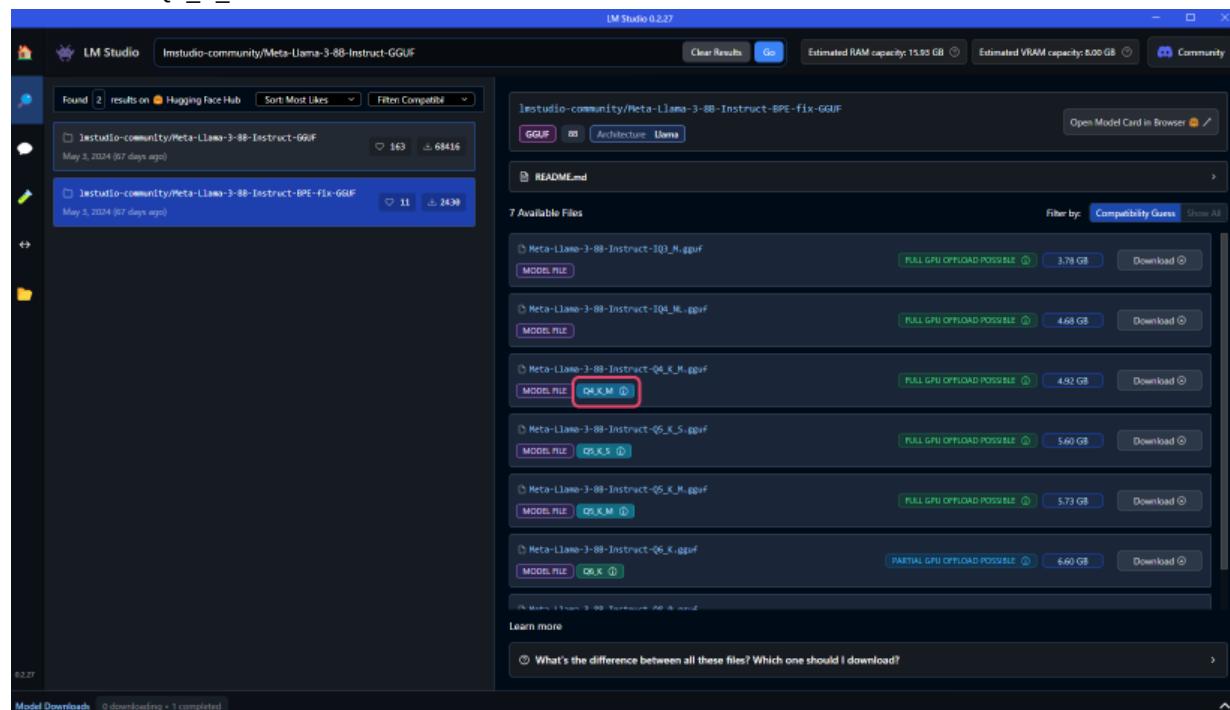
特徴： Meta社が提供するこれまで最も先進的かつ高性能なLLMとなっております。Llama 3.1は128Kのコンテキスト長、最先端のツール使用、推論機能の強化により、長文テキストの要約、多言語会話エージェント、コーディング支援などの高度なユースケースをサポートしています。

弊社推奨LLMについての補足

推奨LLMをLMスタジオで検索いただくと、以下のように、多くのモデルが表示されるかと思います。

弊社では、サイズと性能のバランスに優れた4ビット量子化モデルの利用を推奨させていただいております。

※具体的にはQ4_K_Mの記載のあるもの、またはそれに近い量子化がされているものとなります。



LinuxサーバーでLMスタジオを動作させる場合

カスタムLLMの設定

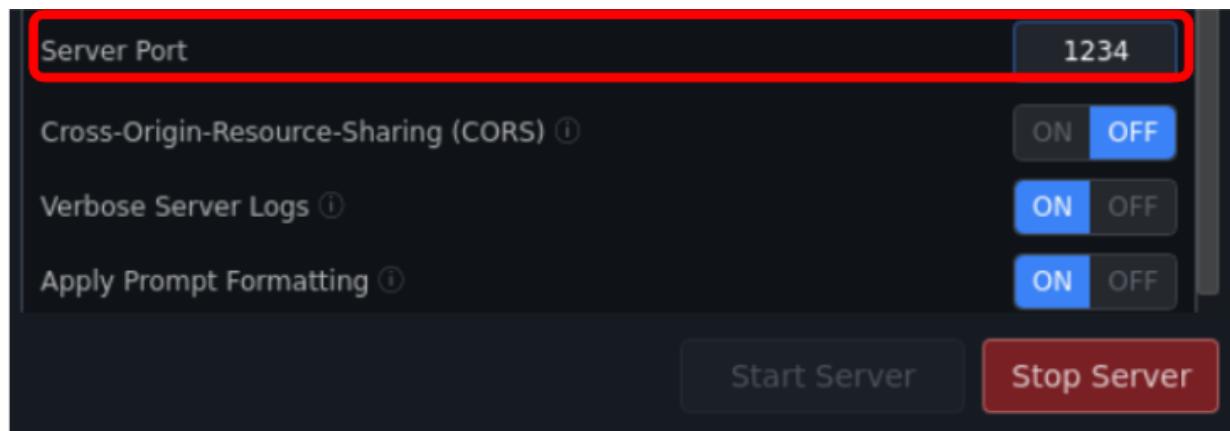
LinuxサーバーでLMスタジオを動作させてailia DX insightから接続しようとした際に以下のようなエラーが表示された際はLinuxサーバーのファイヤーウォール設定でポートへのアクセスが禁止されている可能性があります。



以下のコマンドでポートをアクセスを許可してください。

```
sudo ufw allow 1234/tcp
```

※上記のポート番号「1234」はLMスタジオで設定した番号に置き換えてください。

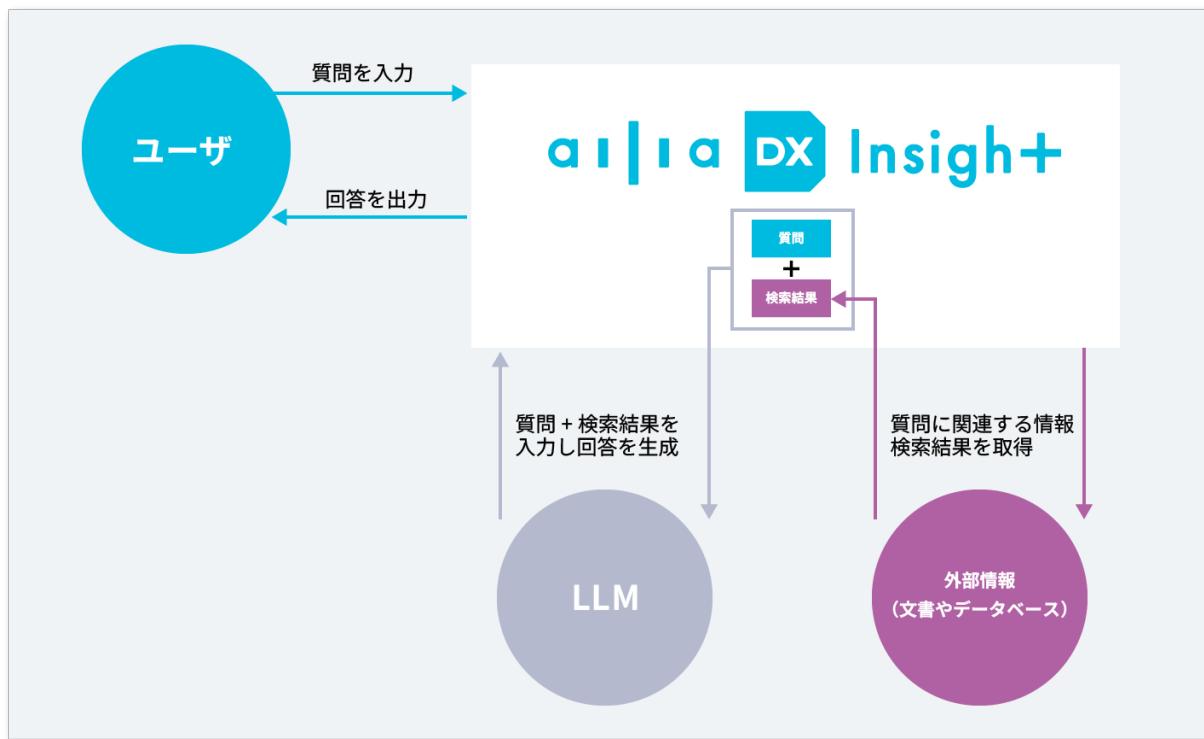


[次のページへ >](#)

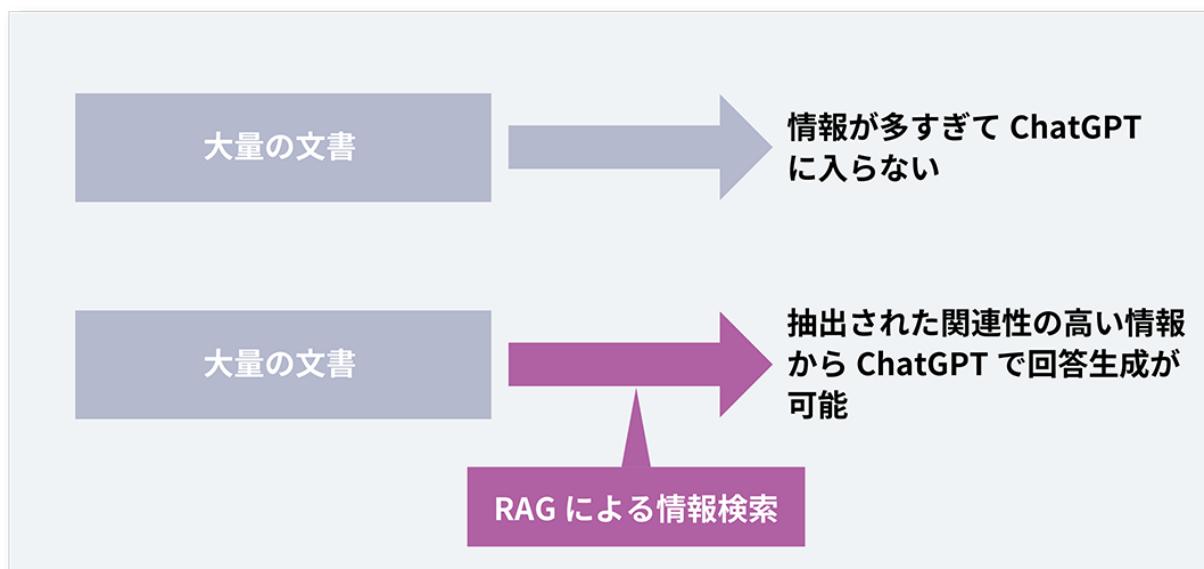
RAGによる回答精度の向上

RAGとは

Retrieval-Augmented Generation (RAG) とは大規模言語モデル（LLM）によるテキスト生成と外部情報の検索を組み合わせることで、事実に基づかない情報の生成を抑制し、AIによる回答精度を向上させる技術のことです。



ChatGPTにはトークン数に上限があるため、大量の文書をそのまま供給することはできません。そのため、ailia DX InsightではRAGによる情報検索を事前に行い、その一部を元に回答を生成されています。



最大トークン数について

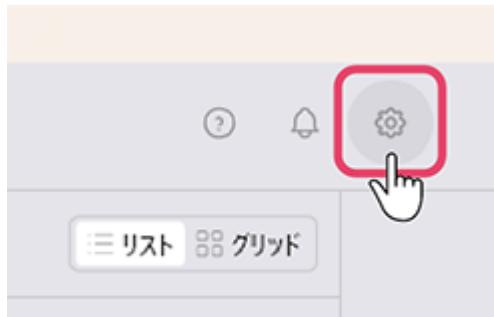
カスタムLLMの設定

```
chatgpt-3.5 : 2k  
chatgpt-3.5-turbo (16kと統合された) : 16k  
chatgpt-4 : 8k  
chatgpt-4-turbo : 128k
```

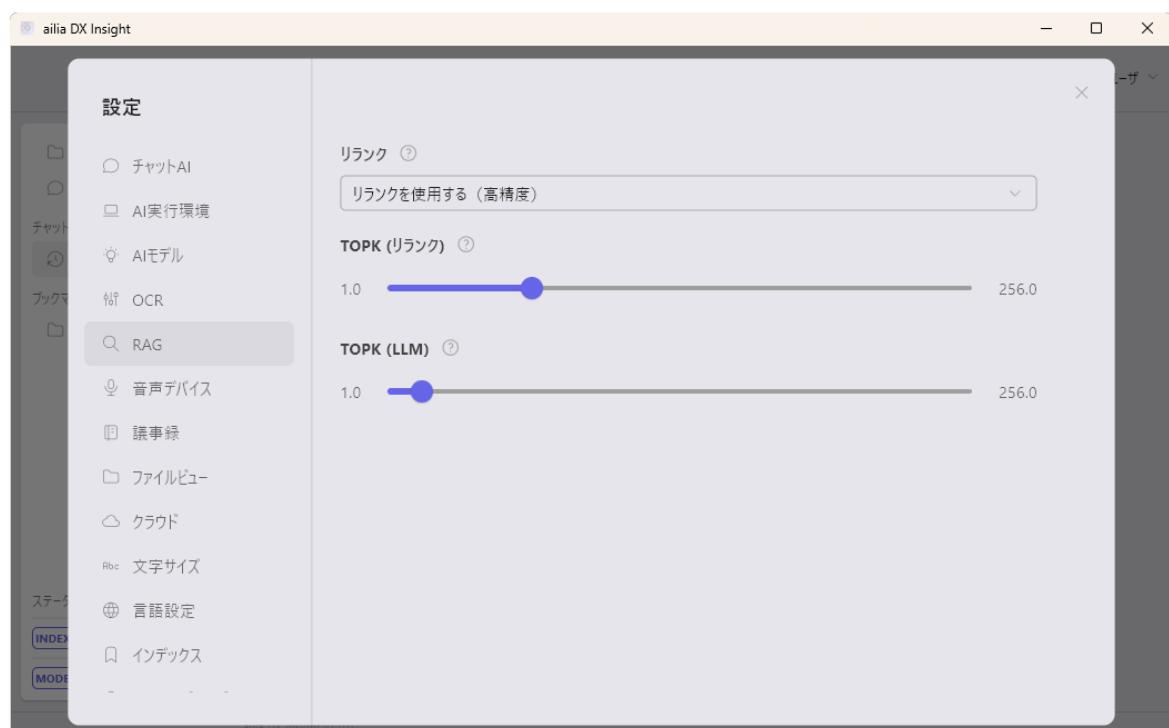
トークン数が多いほど、複数チャンクの情報をまとめて投入可能で、精度が向上します。ただし、その分回答が遅くなったり、コストが高くなります。

RAGの設定

1. ailia DX insightの初期画面にて、右上の歯車アイコンをクリックして設定ウィンドウを表示させます。



2. 左側の項目から「RAG」を選択します。

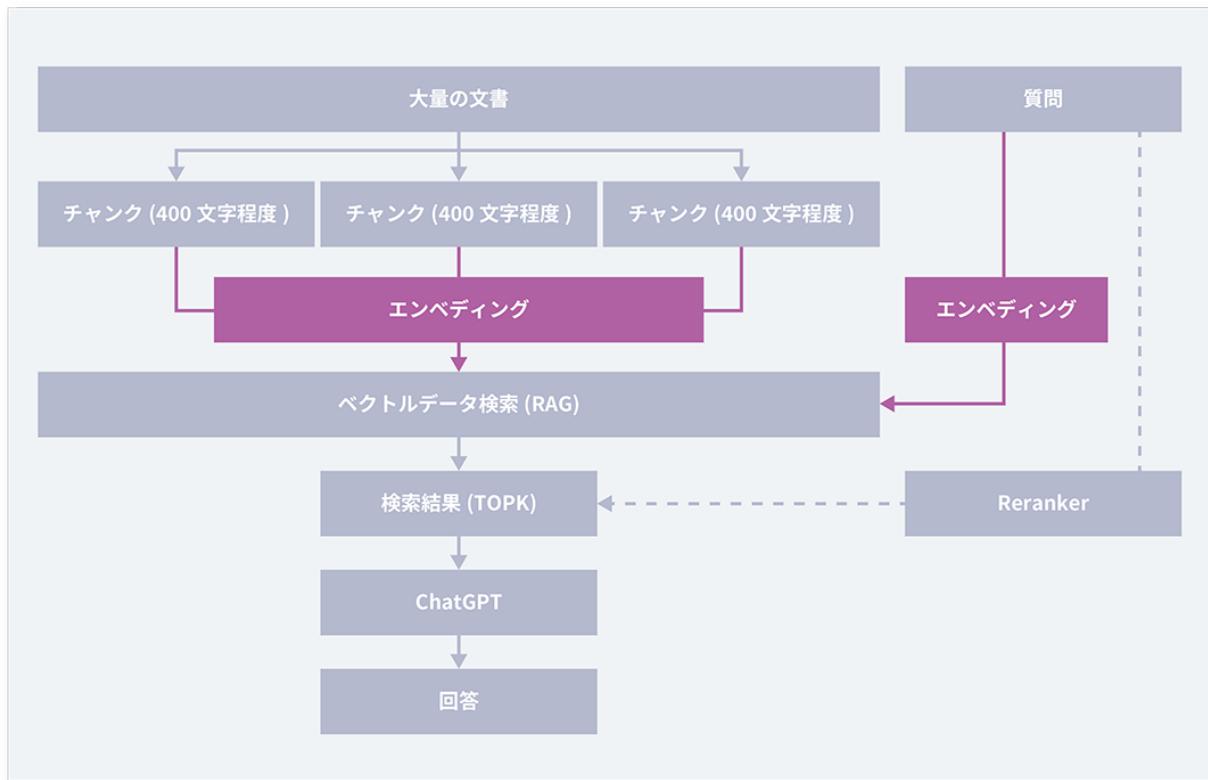


3. リランク、TOPKについて設定することができます。

[次のページへ >](#)

エンベディングとは

エンベディング（Embedding）とはチャンクのテキストをベクトルデータに変換する技術です。ベクトル間の距離を計算することで、テキストとテキストの意味が近いかを判定可能です。RAGでは、質問文のテキストのベクトルと、チャンクのベクトルの距離を計測することで、質問文に近いチャンクを検索します。



[次のページへ >](#)

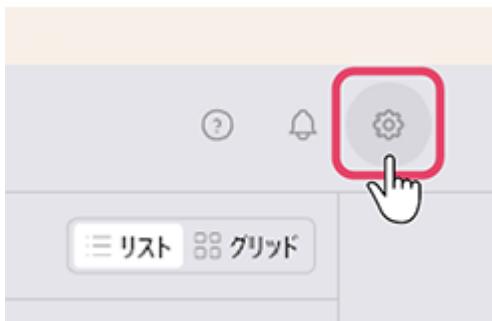
Azure OpenAI Service

ailia DX insightは、文章の生成をChatGPTで行います。OpenAI APIだけでなく、Azure OpenAI Serviceを介してChatGPTを利用することも可能です。Azure OpenAI Serviceについての詳細は[こちら](#)をご覧ください。

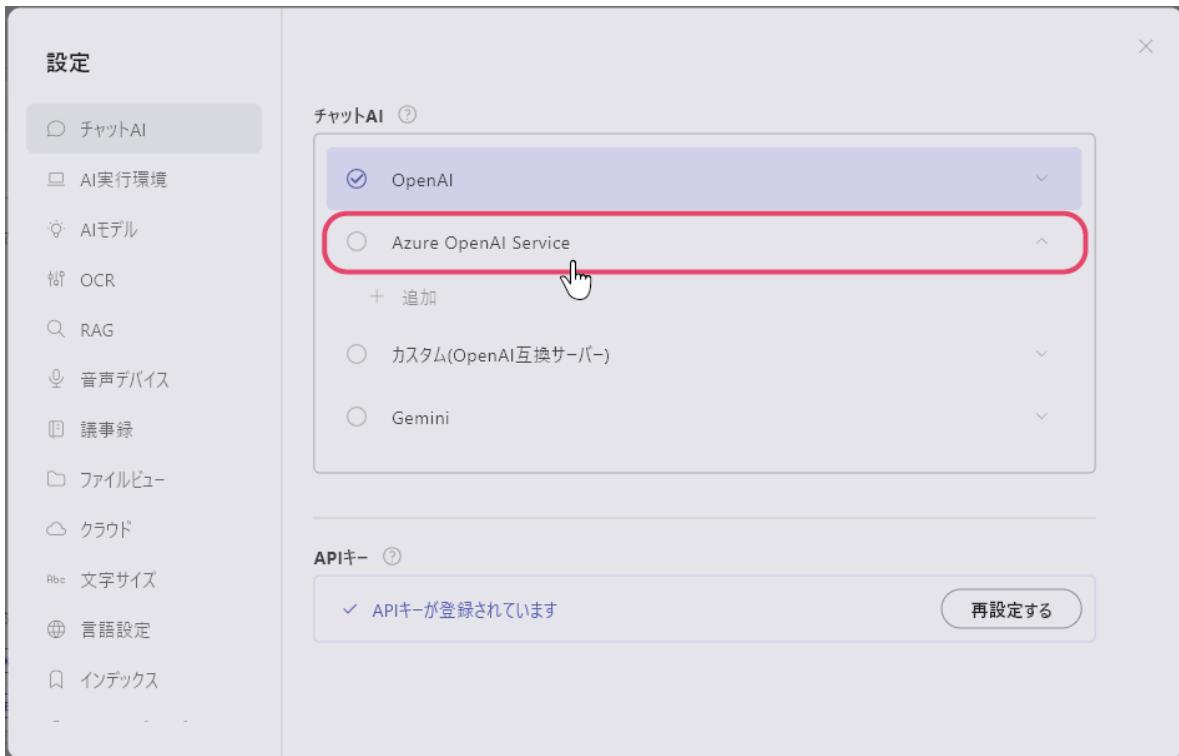
Azure OpenAI Serviceへの接続方法

設定画面の呼び出し

- ailia DX insightの初期画面にて、右上の歯車アイコンをクリックして設定ウィンドウを表示させます。



- 「チャットAI」の項目の中にある「Azure OpenAI Service」をクリックし、「+追加」を選択します



Azure OpenAI Serviceの設定を登録

カスタムLLMの設定

AzureにデプロイしたChatGPTを登録するウィンドウが開きます。各項目を記入します。

Azure OpenAI 設定

名前

APIキー

デプロイ名

APIバージョン

リソース名

キャンセル 完了

- 名前：UI表示に使用する名称
- APIキー: ChatGPTをデプロイしたリソースのAPIキー
- デプロイ名: ChatGPTをデプロイした際に設定したデプロイ名
- APIバージョン: 利用するAPIバージョン
 - [こちら](#)のページの、「サポートされているバージョン」が利用可能です。
- リソース名: ChatGPTをデプロイしたリソースの名前
 - ChatGPTをデプロイしたエンドポイント https://RESOURCE_NAME.openai.azure.com/ のRESOURCE_NAMEにあたる部分です。
- モデル: デプロイしたモデル名(gpt-3.5-turbo等)
- 説明: 必要に応じてメモとして使用

[次のページへ >](#)

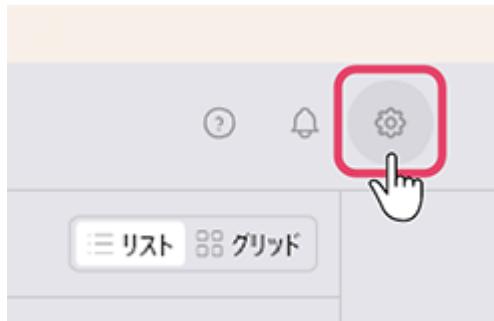
Gemini

ailia DX insightは、Geminiを使用して文章の生成を行うこともできます。 Geminiについての詳細は[こちら](#)をご覧ください。

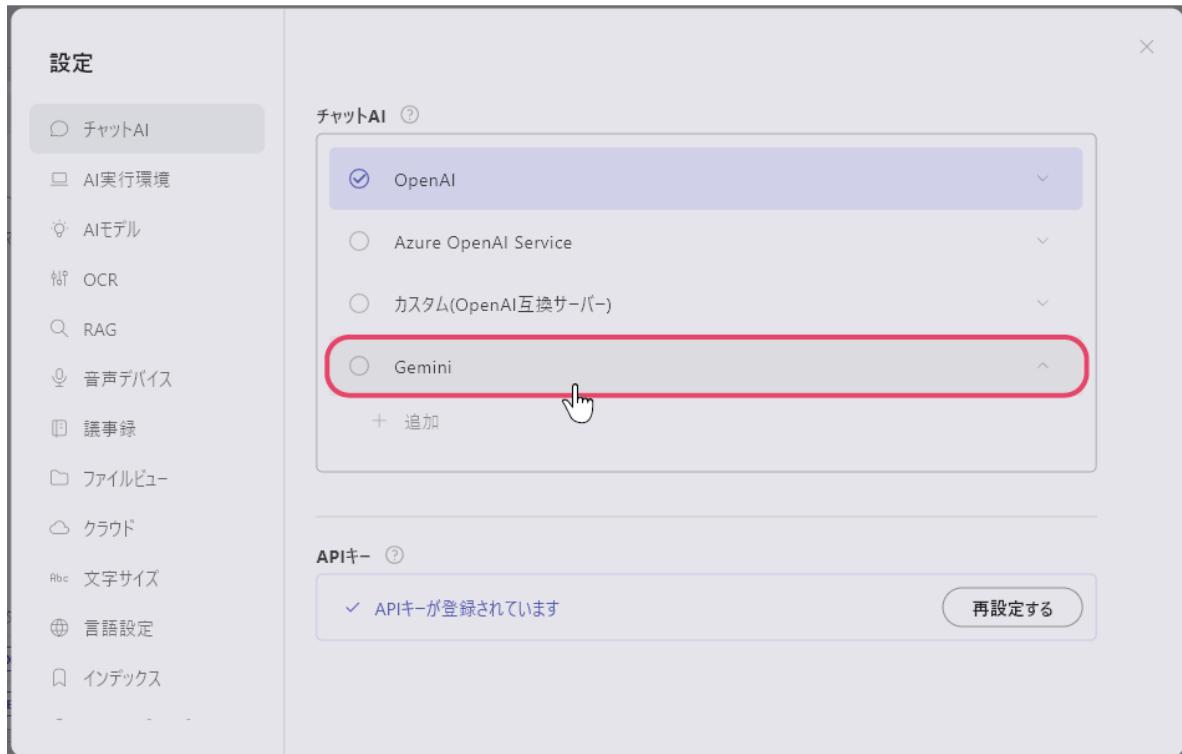
Geminiへの接続方法

設定画面の呼び出し

1. ailia DX insightの初期画面にて、右上の歯車アイコンをクリックして設定ウィンドウを表示させます。



2. 「チャットAI」の項目の中にある「Gemini」をクリックし、「+追加」を選択します



Geminiの設定を登録

カスタムLLMの設定

1. GeminiにデプロイしたChatGPTを登録するウィンドウが開きます。各項目を記入します。

Gemini モデル

名前

説明

モデル

最大トーケン長

キャンセル 追加する

- 名前: UI表示に使用する名称
- 説明: 必要に応じてメモとして使用
- モデル: デプロイしたモデル名
- 最大トーケン長: モデルの最大トーケン数

2. APIキーを登録してくださいとAPIキーのセクションに表示されているので、設定します。

APIキー、モデル名については[こちら](#)をご覧ください。

Geminiの最新モデルを追加する

Geminiの最新モデルは以下のように項目を記入することで追加することができます。現在利用できるモデルについては[GoogleAI for Developers](#)をご確認ください。

Gemini 2.0 Flash

Gemini モデル

名前

gemini-2.0-flash

説明

gemini-2.0-flash

モデル

gemini-2.0-flash

最大トークン長

1048576

キャンセル

完了

- 名前：UI表示に使用する名称
- 説明：必要に応じてメモとして使用
- モデル: gemini-2.0-Flash
- 最大トークン長: 1048576

Gemini 2.0 Flash-Lite プレビュー

Gemini モデル

名前

gemini-2.0-flash-lite-preview-02-05

説明

gemini-2.0-flash-lite-preview-02-05

モデル

gemini-2.0-flash-lite-preview-02-05

最大トークン長

1048576

キャンセル

完了

- 名前：UI表示に使用する名称
- 説明：必要に応じてメモとして使用
- モデル: gemini-2.0-Flash-lite-preview-02-05
- 最大トークン長: 1048576

Gemini 1.5 Flash

Gemini モデル

名前

gemini-1.5-flash

説明

gemini-1.5-flash

モデル

gemini-1.5-flash

最大トークン長

1048576

キャンセル

完了

- 名前：UI表示に使用する名称
- 説明：必要に応じてメモとして使用
- モデル: gemini-1.5-Flash
- 最大トークン長: 1048576

Gemini 1.5 Flash-8B

Gemini モデル

名前

説明

モデル

最大トークン長

キャンセル

完了

- 名前：UI表示に使用する名称
- 説明：必要に応じてメモとして使用
- モデル: gemini-1.5-Flash-8b
- 最大トークン長: 1048576

Gemini 1.5 Pro

Gemini モデル

名前

説明

モデル

最大トークン長

[キャンセル](#) [完了](#)

- 名前：UI表示に使用する名称
- 説明：必要に応じてメモとして使用
- モデル: gemini-1.5-pro
- 最大トークン長: 2097152

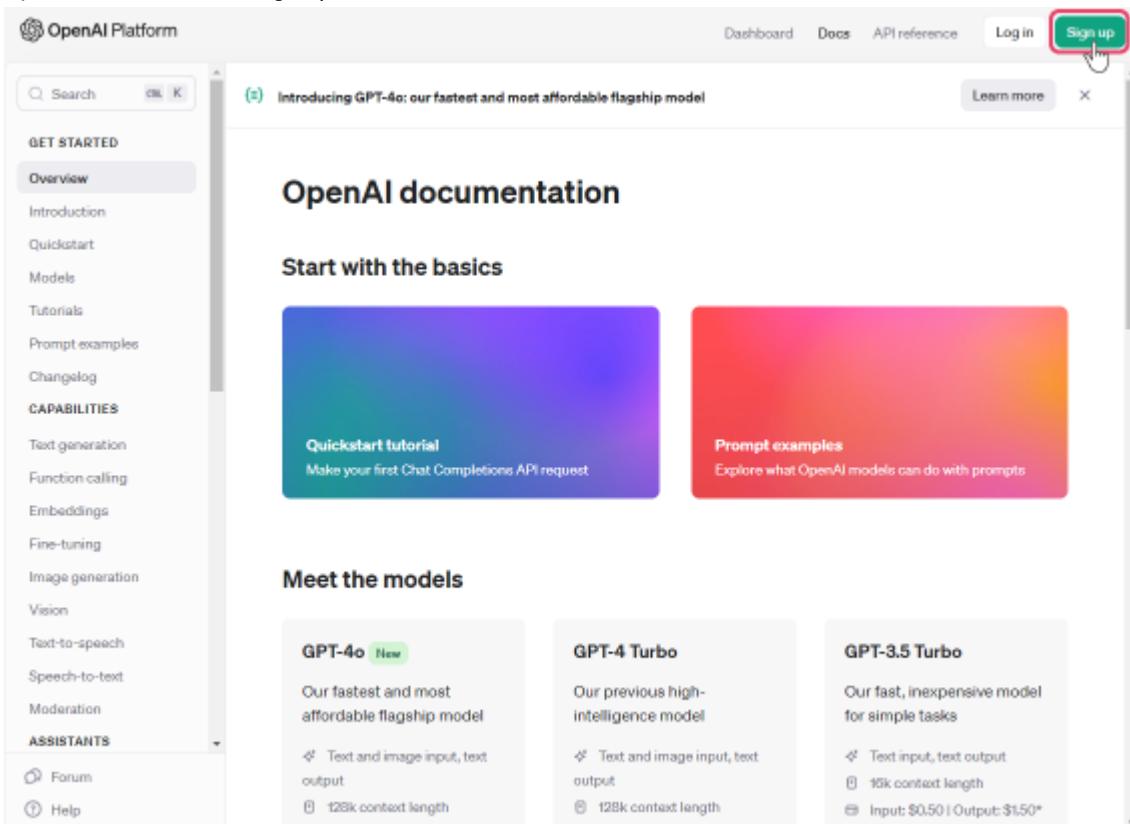
[次のページへ >](#)

OpenAI APIキーの取得方法

「ailia DX insight」内でOpenAIの機能を使用する場合にはOpenAIのAPIキーの取得が必要となります。また、OpenAIではレート制限を設けている為、特に多人数で利用される場合はレート制限の上限を上げることをお勧めいたします。それには規定日数（最低7日～）の経過が必要となる為、お早目の対応が必要となります。

OpenAIアカウントを作成する

1. OpenAI Platformから「Sign up」を選択します。



The screenshot shows the OpenAI Platform homepage. On the right side, there is a navigation bar with 'Dashboard', 'Docs', 'API reference', 'Log in', and a green 'Sign up' button, which is highlighted with a red box and a cursor icon. Below the navigation bar, there is a banner with the text '(?) Introducing GPT-4o: our fastest and most affordable flagship model'. To the right of the banner is a 'Learn more' button and a close button ('X'). The main content area is titled 'OpenAI documentation' and features a section titled 'Start with the basics' with two cards: 'Quickstart tutorial' (blue card) and 'Prompt examples' (red card). Below this is a section titled 'Meet the models' with three cards: 'GPT-4o New' (gray card), 'GPT-4 Turbo' (light gray card), and 'GPT-3.5 Turbo' (light gray card). The left sidebar contains a search bar, a 'GET STARTED' section with 'Overview' selected, and several other sections like 'CAPABILITIES', 'ASSISTANTS', 'FORUM', and 'HELP'.

カスタムLLMの設定

2. 案内に沿ってメールアドレス、パスワードを入力し、続けるを選択します。

アカウントの作成

OpenAI Platformへ移動するためにOpenAI用のパスワードを設定してください。

メールアドレス 編集

パスワード @

続ける

アカウントをお持ちですか？ ログイン

または

Google で続ける

Microsoft Account で続ける

Apple で続ける

3. 登録したメールアドレス宛に「OpenAI - Verify your email」という件名のメールが届くので、メールアドレスの確認を選択すると、ブラウザが起動します。



メール アドレスの確認

OpenAI アカウントの設定を続けるには、これがあなたのメールアドレスであることを確認してください。

メール アドレスの確認

このリンクは3時間後に有効期限が切れます。このリクエストを行っていない場合は、このメールを無視してください。サポートが必要な場合は、[ヘルプセンター](#)を通じてお問い合わせください。

4. 起動されたブラウザで「氏名」「組織名（オプション）」「生年月日」を入力し、規約、プライバシーポリシーを確認後、同意するを選択します。

カスタムLLMの設定



電話番号を登録する

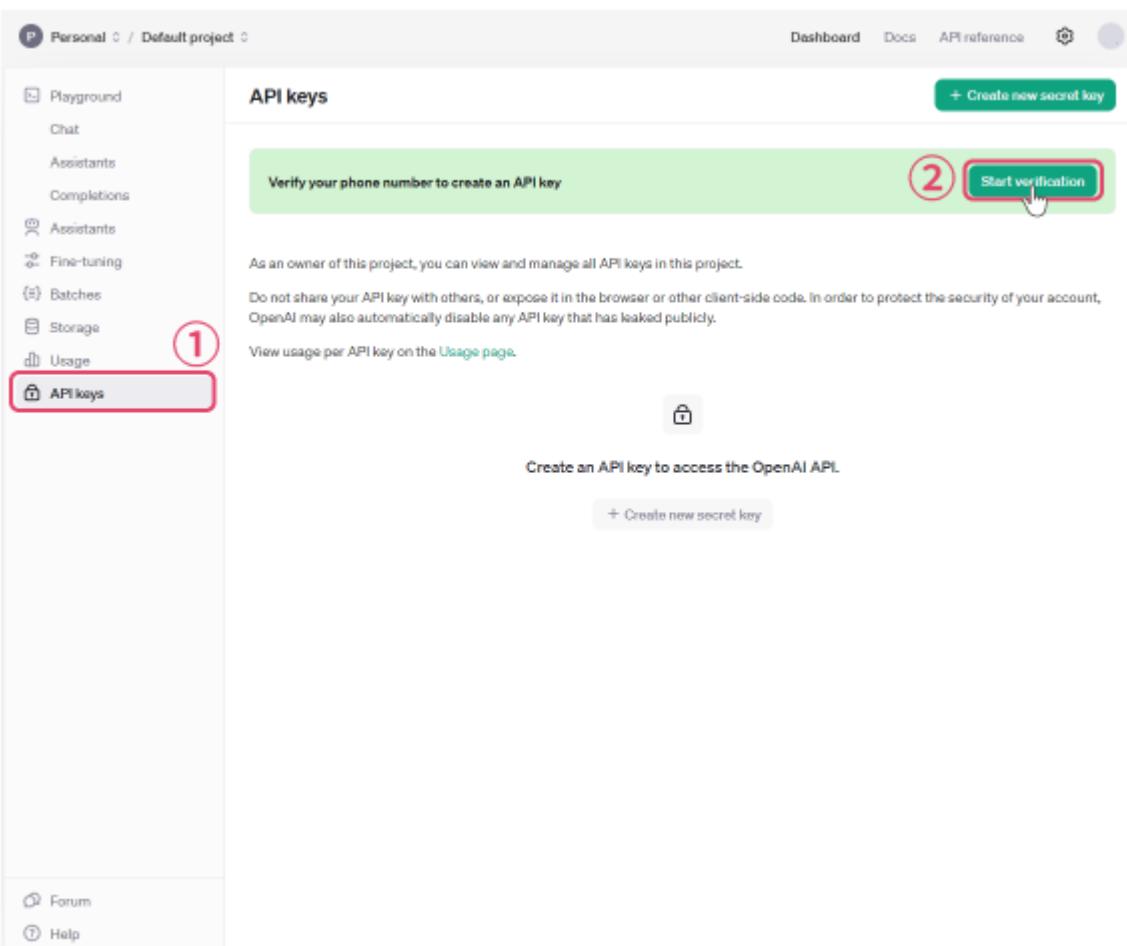
APIキーを取得するにはショートメッセージを受け取ることのできる電話番号の登録が必要となります。

1. OpenAIのDocsページ右上の「Dashboard」を選択します。

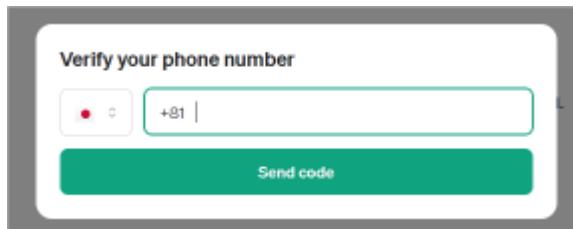
A screenshot of the OpenAI Documentation dashboard. The top navigation bar shows "Personal / Default project". The "Dashboard" tab is highlighted with a red box and a cursor arrow. The main content area features the heading "OpenAI documentation" and "Start with the basics". It includes two large cards: "Quickstart tutorial" (blue gradient background) and "Prompt examples" (red gradient background). Below this, there's a section titled "Meet the models" with three cards: "GPT-4o" (green "New" badge), "GPT-4 Turbo", and "GPT-3.5 Turbo". Each card provides a brief description and some technical details. At the bottom right, there's a note: "* prices per 1 million tokens". On the left side, there's a sidebar with a "GET STARTED" section containing links like "Overview", "Introduction", "Quickstart", etc., and a "CAPABILITIES" section listing various AI services. At the very bottom of the sidebar, there are links for "Forum" and "Help".

カスタムLLMの設定

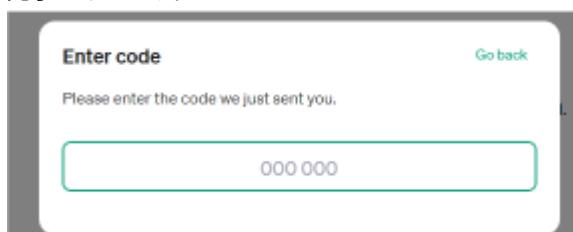
2. 左側のメニューから「API keys」を選択して、「Verify your phone number to create an API key」という案内の横にある「Start verification」を選択します。



3. 電話番号を入力し、「Send code」を選択します。



4. 電話番号宛てにOpenAI認証コードがショートメッセージで届くので、次の画面で入力し、電話番号の登録が完了となります。



APIキーを生成する

カスタムLLMの設定

- 「+ Create new secret key」を選択します。

The screenshot shows the 'API keys' section of the OpenAI dashboard. On the left sidebar, 'API keys' is selected. A prominent green button at the top right says '+ Create new secret key'. Below it, a message states 'Project API keys have replaced user API keys.' and provides a link to learn more. The main area contains instructions for creating an API key and a note about security. A small lock icon is present.

- 「Create new secret key」のウィンドウが表示されるので、任意の名称を入力してAPIキーを生成します。

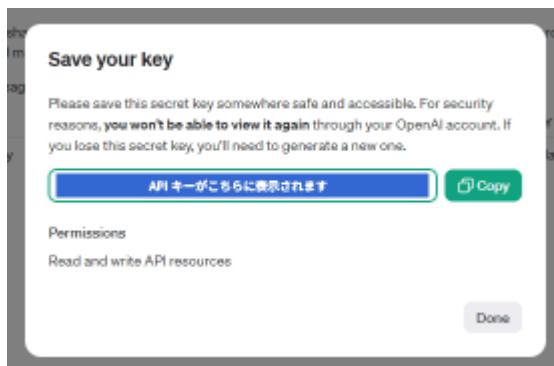
The screenshot shows the 'Create new secret key' dialog box. It has fields for 'Name' (set to 'My Test Key'), 'Project' (set to 'Default project'), and 'Permissions' (set to 'All'). At the bottom are 'Cancel' and 'Create secret key' buttons.

- APIキーが表示されます。

※生成されたAPIキーが表示されるのは一度きりとなるので、この時点でコピーして保存しておくことをお勧め

カスタムLLMの設定

めします。



支払い情報を登録する

1. OpenAIのホーム画面右上のご自身のアカウントのアイコンをクリックし、「Your profile」を選択します。

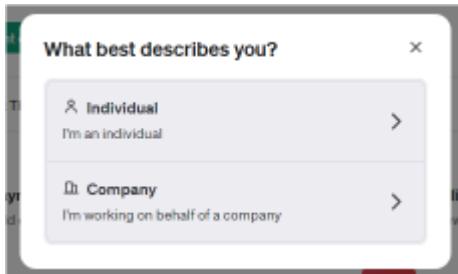
The screenshot shows the OpenAI documentation homepage. On the left, there's a sidebar with various navigation links under sections like 'GET STARTED', 'CAPABILITIES', 'ASSISTANTS', etc. The main content area features the title 'OpenAI documentation' and a section 'Start with the basics' with two cards: 'Quickstart tutorial' (blue card) and 'Prompt examples' (red card). Below this is a section 'Meet the models' with three cards: 'GPT-4o' (green 'New' badge), 'GPT-4 Turbo', and 'GPT-3.5 Turbo'. In the top right corner, there's a user profile dropdown menu. The 'Your profile' option is highlighted with a red box and a cursor arrow pointing to it. Other options in the menu include 'Terms & policies' and 'Log out'. The URL at the bottom of the page is <https://platform.openai.com/settings/profile>.

2. Settingsという画面になります。左のメニューから「Billing」を選択し、「Add payment details」を選択します。

カスタムLLMの設定

The screenshot shows the 'Billing' section of the ChatGPT settings. The sidebar on the left has a 'SETTINGS' category with options like 'Your profile', 'Organization', 'General', 'Members', 'Billing' (which is highlighted with a red box and circled with a red number 1), 'Limits', 'Project', 'General', 'Members', 'Limits', and '+ Create project'. The main content area is titled 'Billing' with tabs for 'Overview', 'Payment methods', 'Billing history', and 'Preferences'. It shows a 'Free trial' section with 'Credit remaining: \$0.00' and a green button 'Add payment details' which is also highlighted with a red box and circled with a red number 2. Below this, there's a note: '(?) Note: This does not reflect the status of your ChatGPT account.' To the right, there are four cards: 'Payment methods' (green icon), 'Billing history' (purple icon), 'Preferences' (pink icon), and 'Pricing' (orange icon).

3. 「What best describes you?」（あなたに当てはまるものを選択してください）というウィンドウが表示されます。「individual」（個人）か「Company」（会社）を選択します。



4. 表示されるウィンドウにてカード情報/請求先住所を入力し、「Continue」を選択することで、支払い情報の登録が完了します。

カスタムLLMの設定

Add payment details

Add your credit card details below. This card will be saved to your account and can be removed at any time.

Card information

カード番号 月 / 年 セキュリティコード

Name on card

Billing address

Country
Address line 1
Address line 2
City Postal code
State, county, province, or region

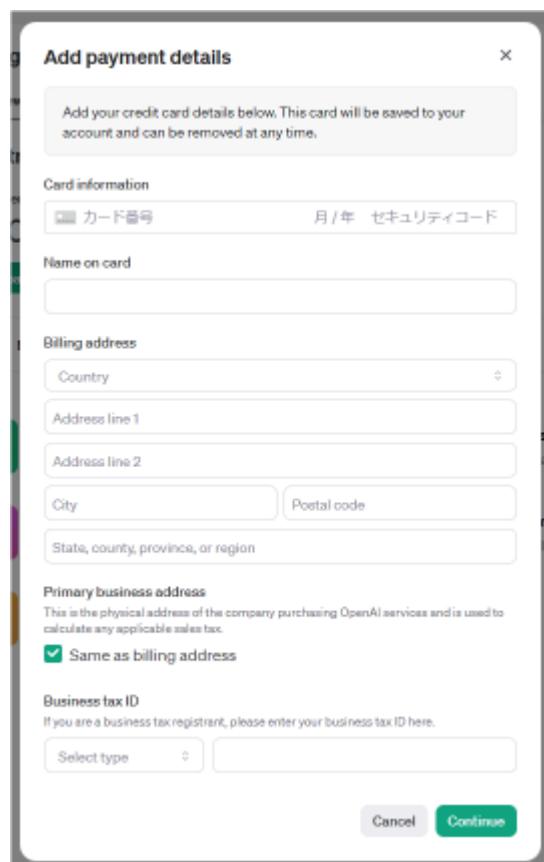
Primary business address
This is the physical address of the company purchasing OpenAI services and is used to calculate any applicable sales tax.

Same as billing address

Business tax ID
If you are a business tax registrant, please enter your business tax ID here.

Select type

Cancel Continue



5. 月の支払い限度額を設定することで、APIにかかる費用を管理することができます。

[次のページへ >](#)

OpenAIの料金体系

ailia DX InsightにてOpenAIの機能（ChatGPT/dall-e等）を使用する場合、料金体系とレート制限について留意いただく必要がございます。レート制限とはOpenAIではサービスの安定性等の観点から設けているもので、ユーザーのTier（利用レベル）によって、リクエスト数の制限がかかります。特に多人数で1アカウントのご利用を考えている場合はレート制限を解除する為に準備期間として1週間以上（場合によっては3週間ほど）が掛かる場合がある為、ご注意ください。

ChatGPTのトークンについて

OpenAIの各モデルはトークンという単位で従量課金のお支払い金額が決まります。入力したトークン数に応じた単価と出力したトークン数に応じた単価を足した金額が利用料金となります。

トークンとは、英語の場合は1単語が1トークン、日本語の場合は1文字が基本的には1～2トークンに該当します。また、日本語の場合、「です」「ます」等が2文字で1トークンとなることがあります。（ChatGPTのバージョンによってトークン数の判定が異なります）

また、句読点やクエスチョンマーク等もそれぞれ1トークンとしてカウントされます。

例) アメリカの首都はどこですか? → 13トークン

例) What is the capital of the United States? → 9トークン

OpenAIが提供している「Tokenizer」というツールを使用すると、入力したトークンがどのようにカウントされているか確認することができます。

GPT-4o (coming soon)

GPT-3.5 & GPT-4

GPT-3 (Legacy)

アメリカの首都はどこですか?

Clear

Show example

Tokens

13

Characters

14

アメリカの首都はどこですか?

Text Token IDs

カスタムLLMの設定

モデル毎の利用料金に関しては変動がある為、[こちら](#)からご確認ください。

レート制限

OpenAIではサービスの安定性等の観点からレート制限を設けています。

レート制限はRPM(リクエスト数/1分)、RPD(リクエスト数/1日)、TPM(トークン数/1分)、TPD(トークン数/1日)、IPM(イメージ数/1分)の5つの方法で測定されます。

レート制限は、最初に何が起こるかに応じて、どのオプションでもヒットする可能性があります。たとえば、100トークンのみを含むリクエストを20回ChatCompletionsエンドポイントに送信すると、その20回のリクエスト内で150,000トークンを送信していなくても(TPM制限が150,000の場合)、制限に達してしまいます(RPMが20の場合)。

また、レート制限はユーザーのTierと使用するモデルに毎に異なります。

レート制限について詳しくは[OpenAIのHP](#)よりご確認ください。

Tier

Tierは支払金額と最初の支払いからの経過日数によってTier1～Tier5の5段階に分かれています。（ユーザー登録のみの無料枠もあります）

Tier	条件
無料	
Tier1	\$5の支払い実績
Tier2	\$50の支払い実績と最初の支払いから7日間以上の経過
Tier3	\$100の支払い実績と最初の支払いから7日間以上の経過
Tier4	\$250の支払い実績と最初の支払いから14日間以上の経過
Tier5	\$1,000の支払い実績と最初の支払いから30日間以上の経過

※Tierの条件については変動の可能性がある為、[OpenAIの公式HP](#)も合わせてご参照ください。

レート制限の引き上げをリクエストする

Tier2以上になると、レート制限の引き上げをリクエストすることができます。

カスタムLLMの設定

1. Your profile内の「Limits」の項目内、一番下にある「Request an exception」を選択します。

The screenshot shows the 'Personal' settings page for a 'Default project'. The 'Limits' section is highlighted with a red box. At the bottom of this section, there is a button labeled 'Request an exception' which is also highlighted with a red box and has a cursor icon pointing to it.

SETTINGS

Limits (highlighted with a red box)

Increasing your limits

Your organization is currently in **Usage tier 2**. Your limits will automatically be increased once you move to the next usage tier based on the criteria outlined below. Visit our [usage tiers documentation](#) to learn more about the limits associated with each tier.

Current tier
Usage tier 2

Once the following criteria are met, you'll automatically move to the next tier:

- At least **\$100** spent on the API since account creation.
- At least **7 days** passed since first successful payment.

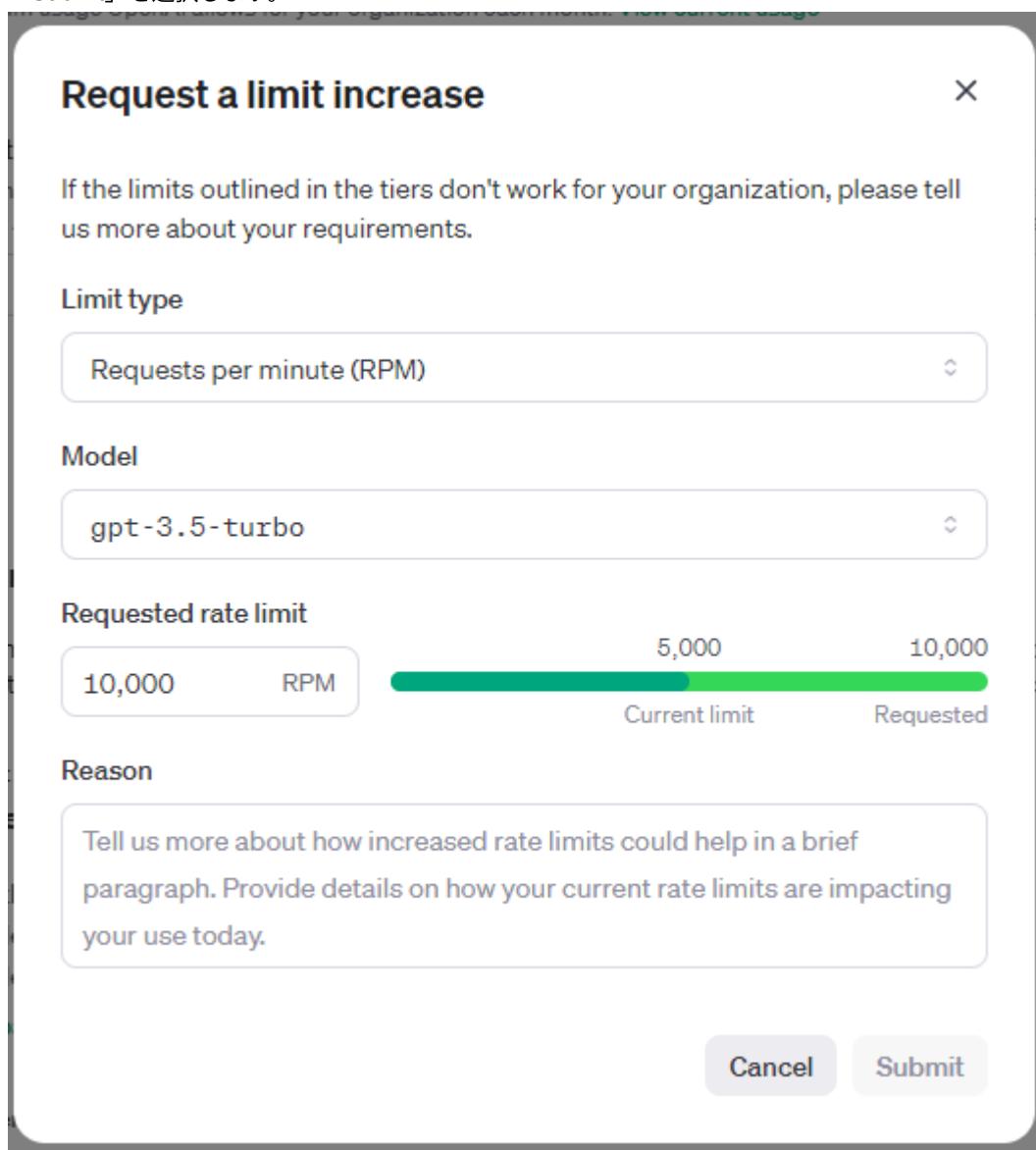
[View payment history](#) [Buy credits](#)

Next tier
Usage tier 3

Request an exception (highlighted with a red box and a cursor icon)

カスタムLLMの設定

2. 「Request a limit increase」のウィンドウが表示されますので、引き上げたい制限の要項を入力し、「Submit」を選択します。



- **Limit type**
制限タイプ：1分あたりのリクエスト数（RPM）/1分あたりのトークン数（TPM）/1分あたりの画像数（IPM）/月間使用制限（Monthly usage limit）の中から一つを選択します。
 - **Model (*)**
制限を引き上げたいモデルを選択します。
(チャットとしての使用がメインであれば gpt-4 / 画像生成の使用がメインであれば dalle-e となると思います)
 - **Request rate limit**
希望するレート制限の数値を入力します。
 - **Reason** レート制限の引き上げがどのように役立つか、現在のレート制限によってどのような不便があるか等について、簡単な文章で説明を入力してください。
- (*) **Limit type** でMonthly usage limitを選択している場合はモデルの選択はありません。

支払い限度額を設定する

OpenAIのモデルは従量課金となっていて、使用制限をつけることで利用料金を管理することができます。

カスタムLLMの設定

1. OpenAIのホーム画面右上にアカウントのアイコンがあるので、クリックし、「Your profile」を選択します。

The screenshot shows the OpenAI documentation homepage. On the right side, there is a user profile sidebar with options like 'Dashboard', 'Docs', 'API reference', 'Theme', 'Your profile' (which is highlighted with a red box), 'Terms & policies', and 'Log out'. The main content area features sections for 'OpenAI documentation', 'Start with the basics' (with 'Quickstart tutorial' and 'Prompt examples'), and 'Meet the models' (listing GPT-4o, GPT-4 Turbo, and GPT-3.5 Turbo). A sidebar on the left contains links for 'GET STARTED', 'CAPABILITIES', 'ASSISTANTS', and other documentation categories.

2. Settingsという画面になるので、左のメニューから「Billing」を選択し、「Usage limits」を選択します。

The screenshot shows the 'Settings' page in the OpenAI interface. The left sidebar has a 'SETTINGS' section with options like 'Your profile', 'Organization', 'Billing' (which is highlighted with a red box and circled with a red number 1), 'Project', 'Members', and 'Create project'. The main content area is titled 'Billing' and includes tabs for 'Overview', 'Payment methods', 'Billing history', and 'Preferences'. It shows a 'Free trial' section with '\$0.00' and buttons for 'Add payment details' and 'View usage'. Below this is a note: 'Note: This does not reflect the status of your ChatGPT account.' At the bottom, there are four buttons: 'Payment methods', 'Billing history', 'Preferences', and 'Usage limits' (which is highlighted with a red box and circled with a red number 2). The 'Usage limits' button has a tooltip 'Set monthly spend limits'.

カスタムLLMの設定

[HOMEへ戻る](#)