

スタートアップマニュアル（概要編）

本書は、Generative AI FW（以降、本サービス）の概要を理解するための情報および利用に関する注意事項について説明します。

提供機能

本サービスはAPI機能とUI機能を提供します。

API機能

本サービスが提供するAPIは推論用のAPIとベクトルDB管理APIがあります。

推論API

以下は本サービスが提供するLLMを利用するための推論APIです。APIの詳細はAPIリファレンスを参照してください

- OpenAI API準拠のAPI
- 履歴付き一般対話API（履歴対話機能）
- 登録済のテンプレートプロンプトでの対話API（テンプレート対話機能）
- 文書検索を利用した対話API（検索対話機能）
- 添付したファイルやWebコンテンツを利用した対話API（拡張対話機能）

ベクトルDB管理API

以下は本サービスが提供するインデックスを操作するためのAPIです。APIの詳細は「ベクトルDB管理APIチュートリアル」を参照してください。

UI機能

- 管理ポータル
- チャットUI
- チャットボット

用語定義

本サービスのガイドに記載する用語の一覧です

用語	当文書内での意味
cotomi	NECが提供するLLM（大規模言語モデル）

APIキー	APIを利用するための認証キー
管理ポータル	本サービスを利用するために必要な管理操作を行うWebアプリケーションツール
チャットUI	本サービスのLLMを用いて対話を行うためのWebアプリケーション
ユーザ	本サービスの利用者。管理ポータルにアクセスできるシステム管理者（組織管理者）と、チャットUIを利用する一般ユーザが存在する
アカウント	本サービスの機能を利用するためのユーザ認証の情報
インデックス	文書検索機能を利用するために作成する、検索対象文書データの集合
ファインチューニング	あらかじめ存在するモデルに利用者が追加学習させて、新しいモデルを生成すること
履歴付き対話APIでの「会話」	同一のhistoryIdを用いて行う一連のやりとり
履歴付き対話APIでの「対話」	会話内の1回のやりとり（LLMへの入力と回答の1回分）
テンプレート	モデルとの会話でよく利用するシステムプロンプト／ユーザプロンプトの定型文を予め作成したもの
システムプロンプト	モデルに会話の背景・目的・ルールなどを指示するために入力するテキスト
ユーザプロンプト	モデルとの対話において利用者が入力するテキスト
システムテンプレート	サービスが提供するテンプレート
ユーザテンプレート	システム管理者が作成したテンプレート
グループ	ユーザ、インデックス、テンプレートが所属できる集合
システムグループ	あらかじめサービスに登録されているグループ
システムユーザ	あらかじめサービスに登録されているユーザ

Keycloak管理者ユーザ	認証(Keycloak)に関する管理画面を操作するための管理者ユーザ
拡張対話	以下の機能を備えた対話の総称 <ul style="list-style-type: none"> 拡張対話（ファイル添付）機能 拡張対話（Webコンテンツ）機能 拡張対話（Web検索）機能
回答根拠の確認	LLMの生成文章に対するハルシネーション対策機能。回答根拠の確認機能には「推論を用いない回答根拠の確認」と「推論を用いた回答根拠の確認」があり、それぞれ以下のような機能の違いがあります。 <ul style="list-style-type: none"> 推論を用いない回答根拠の確認 <p>回答根拠になった文書の内容と関連度のみを確認できる</p> <ul style="list-style-type: none"> 推論を用いた回答根拠の確認 <p>回答根拠になった文書の内容に加え、根拠となった判断理由を提供します。また判断根拠がないなど矛盾がある個所も指摘します。</p>
カスタム認証	デフォルトの認証方式(本サービスに登録したユーザ情報に基づき、メールアドレスとパスワードで認証する方式)以外の認証方式の総称。 <p>詳細は「カスタム認証利用ガイド」を確認してください。</p>
IdP連携	カスタム認証のひとつとして選択できる、外部のIDプロバイダーと連携し認証管理を行うための連携機能。詳細は「カスタム認証利用ガイド」を確認してください。