

Pinecone AI Help Bot プロトタイプ キット specification

pineconeは文書やFAQを数値ベクトルとして保存・検索できるクラウド型のデータベース
AIチャットボットにおける「知識ベース」の役割を担います。

Pinecone はクラウド上でスケーラブルなベクトルデータベースを提供するサービスで、
外部の埋め込みモデル（例: OpenAI Embeddings）で生成したベクトルを格納し、効
検索・管理するために利用されます。

本システムではユーザーの質問に対して的確な回答をさせるためにpineconeのベクトル
ベクトルデータベースとは、文章・画像・音声などから抽出された特徴量（埋め込み）
類似度検索を効率的に行う事が出来ます。

従来のリレーショナルデータベースがキーワード一致検索に強いのに対し、ベクトル
計算できる点に特徴があります。

1. Pineconeの主な仕様

1-1. Index

ベクトル検索の基本単位です

作成時に以下を指定する必要があります。

Dimensions: ベクトルの次元数（例: 1537）

Metric: 類似度計算方式（cosine / dotproduct / euclidean）

Index 内に格納されるベクトルはすべて同じ次元数で統一される

1-2. Namespace

Index 内の論理的な区分け

顧客ごとや用途ごとにデータを分離できる

同じ ID でも Namespace が異なれば別データとして扱われる

1-3. Metric

類似度の計算方式を定義する

cosine: 方向性を重視

dotproduct: 内積を利用

euclidean: ユークリッド距離

1-4. Dimensions

ベクトルが持つ特徴量の数

Index 作成時に固定される

登録されるベクトルは全てこの次元数に一致する必要がある

1-5. 操作機能

Upsert: データ挿入・更新

Query: 類似検索（Top-K 検索）

Delete: ベクトル削除

Fetch: ID 指定で取得

Update: メタデータやベクトル更新

1-6. Metadata

各ベクトルに属性情報を付与可能（例: タイトル、タグ、日付）
検索時にフィルタ条件として利用できる

ここにZIP
フォルダ

2. 環境構築

2-1. Pythonインストール

① Microsoft Store からインストール（推奨・簡単）

`start mswindowsstore://pdp/?productid=9PJPW5LDXLZ5`

↑ このコマンドを実行すると、Microsoft Storeで「Python 3.10」が開きます。
「インストール」 ボタンを押すだけで完了です。

② Chocolatey 経由でインストール（開発者向け）

`SetExecutionPolicy Bypass Scope Process Force;`

``iex ((NewObject System.Net.WebClient).DownloadString('https://chocolatey.org/install.ps1'))`
`choco install python version=3.10.9 y`

③ 【Linux編】 Ubuntu / Debian系

Python 3.10 / 3.11 / 3.12 インストール（任意で選択）

`sudo apt update`

`sudo apt install y softwarepropertiescommon`

`sudo addaptrepository ppa:deadsnakes/ppa`

`sudo apt update`

`sudo apt install y python3.12 python3.12venv python3.12dev`

④ 【Linux編】 CentOS / RHEL 7, 8, 9系

Python 3.6～3.9（標準レポジトリ or EPELから）

`sudo yum update y`

`sudo yum install y epelrelease`

`sudo yum install y python3`

⑤ インストール確認コマンド

`python version`

例：Python 3.10.9 と表示されればOKです。

2-2. OpenAI APIキー取得手順

<https://platform.openai.com/signup> にアクセスし、アカウントを作成またはログイン後、右上のプロフィールアイコン → 「View API keys」を選択。
「Create new secret key」 ボタンを押してキーを生成、コピーして保管。

2-3. Pinecone APIキー取得手順

<https://www.pinecone.io/start/> にアクセスし、アカウントを作成またはログイン後、ダッシュボードに入り「API Keys」セクションへ移動。
「Create API Key」 ボタンを押して名前を入力し、キーを生成、コピーして保管。

2-4.config.txtに以下の様に設定しておきます。

`OPENAI_API_KEY=skxxxxxxxxxxxxxxxxxxxxxx`

PINECONE API KEY=xx

2-5. requirements.txtの作成

このシステムを構築する際に必要な Python ライブラリを一括でインストールする
以下の様に編集します。

```
openai>=1.2.0
```

pineconeclient>=3.0.0

tiktoken>=0.5.1

PyMuPDF>=1.23.0

pythondotenv>=1.0.0

以下のコマンドで自動インストール出来るので便利です。

```
> pip install r requirements.txt
```

2-6. 推奨ディレクトリ構成

```
/projectfolder
```

|——upload_embeddings.py

|——query_embeddings.py

|— index.html

└─ openai.web.js

└── config.txt (POC用)

└─Flask

```
|-----.env
```

|——app.py

|——config.py

2-7. セキュリティに関する注意事項

本システムでは初期構成として、`config.txt` から API キー等を読み込む形式を採用しています。この方法には**セキュリティ上の懸念**があります。

config.txt は HTML・JSと同一ディレクトリに置かれるため、公開領域に含まれる。ブラウザで **F12 キー** の押下でデベロッパーツールが起動し、**ネットワーク** タブの **APIキー** が見えてしまいます。

3. 各プログラムの用途

ファイル名	説明
upload_embeddings.py	ドキュメントをOpenAIでベクトル化し、Pineconeにベクトルデータベースの初期構築に使用。
query_embeddings.py	ユーザーの質問をベクトル化し、類似度検索を行う。Pineconeと連携して、最適な回答候補を抽出。

openai.web.js	フロントエンドとOpenAI APIとの接続処理をJavaScript WebページからAPIを呼び出す際に利用。
---------------	---

4. ユースケースから実際の動きを検証（POC編）

あなたの会社はモバイルバッテリーを製造・販売しているとします。

そこでよくある質問（FAQ）を自動で回答できるAIヘルプボットをWebページにベクトル化して考えています。そのためにはまず「specification.txt」を学習させて、実際に「バッテリーの容量は？」という質問に答えられるように構築してみましょう。

4-1. pineconeに「specification.txt」を読み込ませます

本システムではユーザーの質問文をベクトル化し、登録済みのテキストやPDFと照合して最も関連性の高い情報を抽出します。

手元にある「specification.txt」からベクトルを生成して、pineconeにアップロード
※ 尚、ファイル名は英語名にしてください。pinecone の仕様ではVector IDはASCII
ファイル名からVector IDを取得している本システムでは日本語のファイル名にす
> python upload_embeddings.py ./docs

処理開始: specification.txt

specification.txt → chunk 0 アップロード完了

specification.txt → chunk 1 アップロード完了

4-2. 実際に質問してみましょう

「バッテリーの容量は？」という質問を query_embeddings.py にパラメータで渡
42 で登録した specification.txt と最も関連性の高い情報を Pinecone から抽出し
OpenAI にその情報を渡して自然言語で回答してもらいます。

以下のように2行で実行できます：

> python -c "from query_embeddings import ask_direct_answer; print(ask_direct_answer('バッテリーの容量は？'))"

実行結果例：

バッテリー容量は10,001mAhです。

4-3. Webページから質問してみましょう

4-2までで「バッテリーの容量は？」という質問に答えられる準備が整いました。
今度は実際にユーザーが使うWebページを使って、質問を入力・送信し、AIから
ここでは、開発・検証用の最小構成として、index.html と openai.web.js の2つ
これらはプロダクトパッケージに同梱されますので動作検証用に使用できます。
尚、APIキーなどの機密情報は 直接 openai.web.js に書かず、別ファイル config.js

① 事前準備

index.html と openai.web.js を同一ディレクトリに配置してください。

どちらのファイルも、4章の構築工程に含まれます。

② 操作手順

1. ターミナルを開き、index.html があるディレクトリで以下のコマンドを実行し

> python m http.server 8000

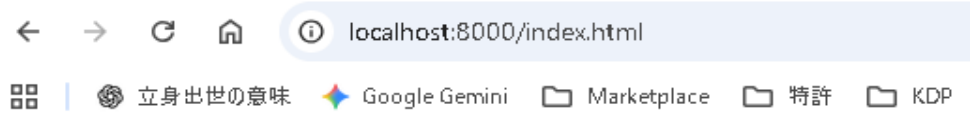
2. ブラウザで次のURLにアクセスします。

<http://localhost:8000/index.html>

- 3. 表示されたページの質問フォームに「バッテリーの容量は？」と入力し送信を
- 4. 検索された内容にOpenAIが自然言語で回答します。

③ 表示例

以下の様に回答を返せば成功です。



モバイルバッテリー AIヘルプボット

質問を入力してください：

送信

バッテリー容量は10,000mAhです。

5. パラメータ調整 よりよい回答のために

5-1. Pinecone 側の主な調整項目

パラメータ	意味	推奨値
minScore	類似検索のスコア閾値	0.6～0.8
topK	類似文書の取得件数	3～5
namespace	データセット識別子	製品名カテゴリ等

5-2. OpenAI 側の主な調整項目

パラメータ	意味	推奨値
temperature	回答の多様性・創造性	0～0.5
max_tokens	出力最大トークン数	200～500
system prompt	モデルへの役割・制約指示	用途に応じ設定
response_format	出力形式の制御	JSON推奨

を解答した状態の
構成を書く

Pythonが最も適合する理由を
書く

るための設定ファイルです。

採用していますが

まれやすく
やソースのタブから

アップロード。

で記述。

トル化
ーザから

いった文書と

・ドします。
l文字と指定されているため
るとエラーとなります。

ますと、
、

ct_answer('バッテリーの容量は?'))"

自然言語の回答を得る処理を行います。
ファイルのみで構成します。
ます。
fig.txt に記述して読み込ませます。

ます。

クリック。



📁 新

注意点
高すぎると「情報不足」、低すぎるとノイズ回答
多すぎるとノイズ、小さすぎると文脈不足
整理・精度向上に有効、一貫性のある命名を推奨

注意点
0=厳格、1=バランス、>1=創造的
短すぎると切れる、長すぎると冗長
指示が曖昧だと誤答や揺れが増える
{"answer":"...", "evidence":"..."}などで根拠も返せる