

生成AIの進化が早すぎて すでに周回遅れ気味な私たち向けの キャッチアップ勉強会

LLM?

ファイン
チューニング?



LangChain?

Embeddings?

この勉強会について

最近、ChatGPTをはじめとする生成AIの進化早すぎないですか？

私たちもすでに周回遅れ気味なので、GW中に慌ててキャッチアップした内容をみんなに共有します👍

最近の生成AIトレンドを

20分でおさらい！

～大規模言語モデル（LLM）を中心に～



みのるん @minorun365

そもそも論から

AIって何だっけ？

AI（人工知能） 人間の思考プロセスに似た動作をするプログラム

ML（機械学習） 人間の「学習」に相当する機能を実現したもの

DL（深層学習） ニューラルネットワークを用いるもの

「生成AI」の多くはこの中の一種！

AI（人工知能） 人間の思考プロセスに似た動作をするプログラム

ML（機械学習） 人間の「学習」に相当する機能を実現したもの

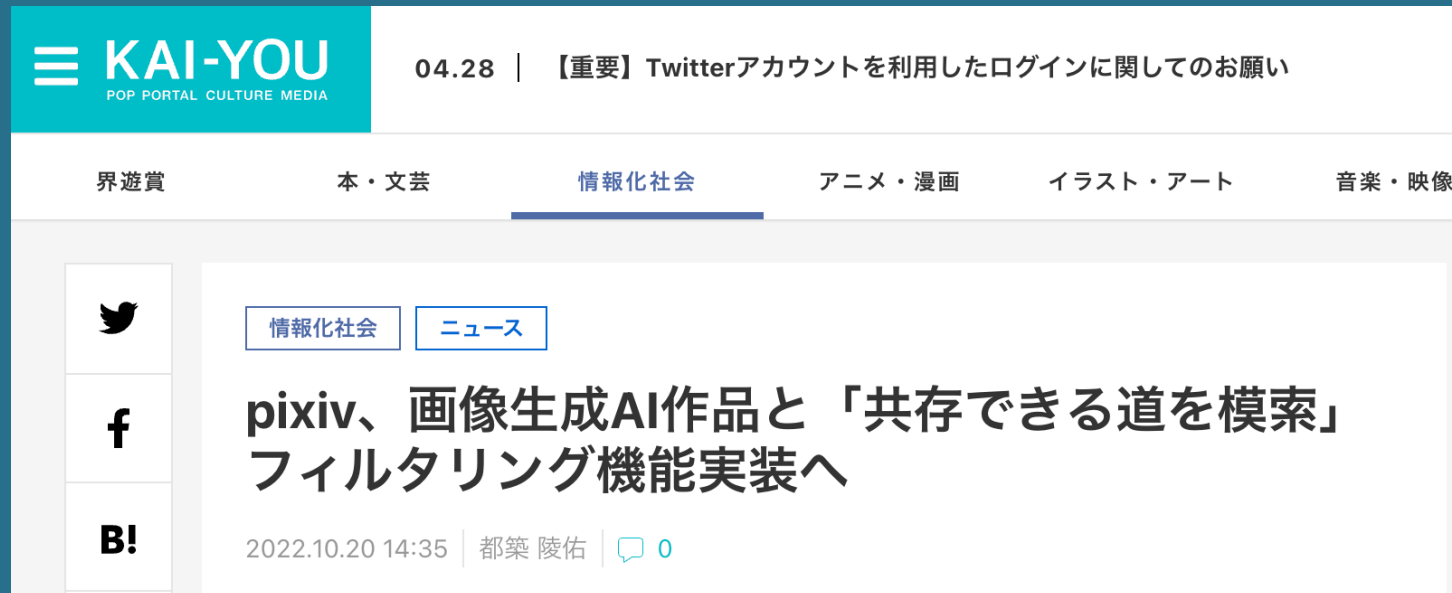
DL（深層学習） ニューラルネットワークを用いるもの

生成AIブームのきっかけ

- Stable Diffusion (2022年8月～)
- ChatGPT (2022年11月～)

Stable Diffusion

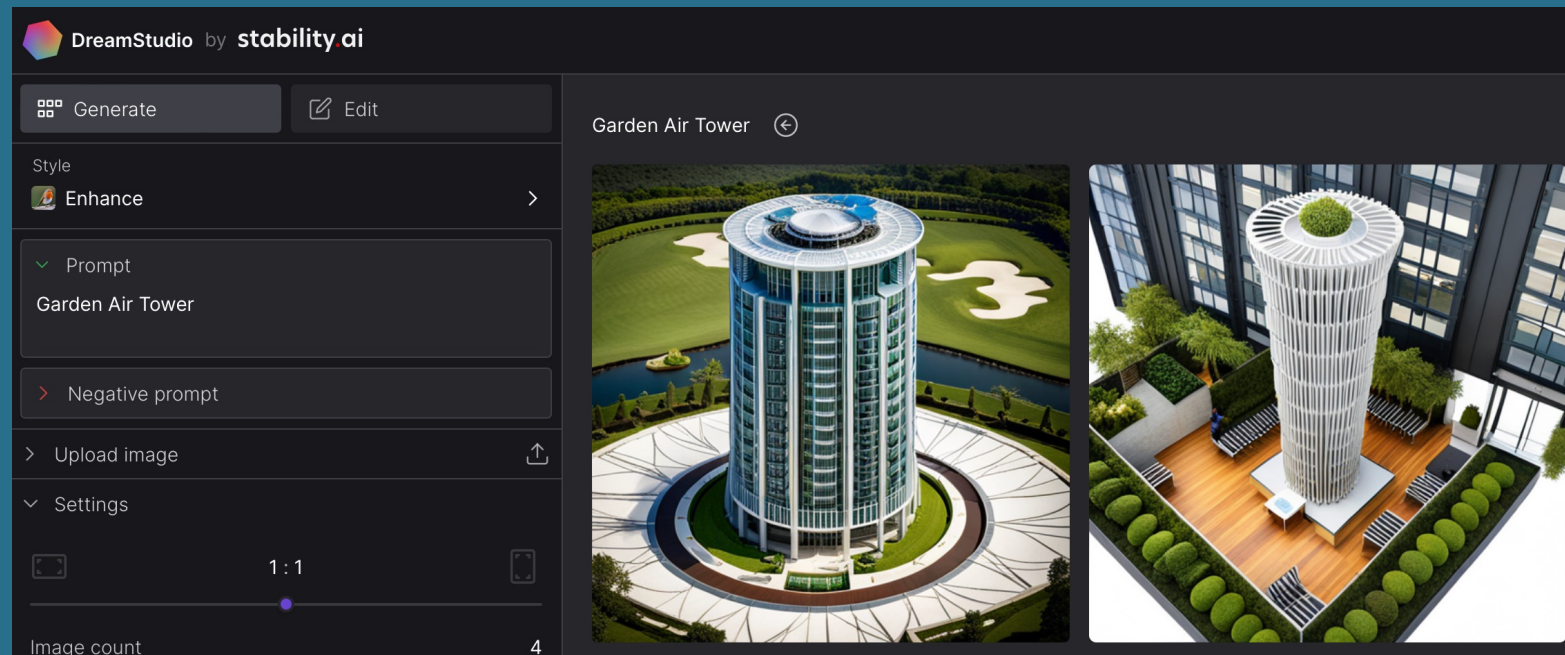
- 2022/8/22にStability AI社らが共同公開
- テキストから画像を生成するtext-to-imageモデル



イラスト投稿サイトに
多数出現するようになり
連日ネットニュースに

Stable Diffusion

- 公式Webアプリ「DreamStudio」で無料試用可能
- 第三者提供のLINEボット「お絵描きばりぐっどくん」も有名？



ビッグウェーブ「ChatGPT」



ChatGPT

- 2022/11/30にOpenAI社が無償で公開
- SNS等で瞬く間に話題となり、現在の大ブームへ



The screenshot displays the ChatGPT web interface. On the left is a dark sidebar with a 'New chat' button and a list of previous conversations, including 'English Self-introduction for Pr...', 'DR戦略とパイロットラ...', 'Mobile Carrier Shop.', and several AWS-related topics. The main area shows a chat with 'Model: GPT-4'. The user's question is: 'dr戦略におけるパイロットライトという語はどのような由来での命名でしょうか？'. The AI's response explains that 'パイロットライト' (pilot light) is a gas-powered household appliance that stays lit and can be used to quickly ignite the main burner. It then states that in a Disaster Recovery (DR) strategy, a 'pilot light approach' means having a minimal resource environment ready to quickly restore the main environment after a disaster.

Model: GPT-4

dr戦略におけるパイロットライトという語はどのような由来での命名でしょうか？

DR戦略（Disaster Recovery戦略）における「パイロットライト」という用語は、ガス式の家庭用機器に由来しています。ガス式の家庭用機器では、小さな火（パイロットライト）が常に燃えていて、それを利用して本体の炎が必要なときにすぐに点火できる仕組みがあります。

DR戦略において、パイロットライトアプローチは、災害が発生した際に迅速に本番環境を復旧・再稼働させることができるように、最小限のリソースを確保しておく手法を指します。これにより、システムがダウンしてしまった場合でも、迅速に復旧し、ビジネスへの影響を最小限に抑えることができます。

ChatGPTとは？

開発元 **OpenAI社**

2015年にサンフランシスコで設立されたAI研究所。
あのイーロン・マスクも立ち上げに参画

AIエンジン **GPT-3.5/4**

Transformerという深層学習モデルを用いた
オープンソースのテキスト生成ソフトウェア

チャットボット **ChatGPT**

GPTファミリーの言語モデルを
チャット対応に最適化+Web UIも提供

これがいわゆるLLM（大規模言語モデル）

開発元 **OpenAI社**

2015年にサンフランシスコで設立されたAI研究所。
あのイーロン・マスクも立ち上げに参画

AIエンジン **GPT-3.5/4**

Transformerという深層学習モデルを用いた
オープンソースのテキスト生成ソフトウェア

チャットボット **ChatGPT**

GPTファミリーの言語モデルを
チャット対応に最適化+Web UIも提供

ChatGPT Plus

- 2023/2/1にサブスク販売開始。月額20ドル（約2,700円）
- 当初は性能優遇が主な利点だったが、3/15に最新エンジンGPT-4が利用可能となり大人気に！

- 回答精度が飛躍的に向上
- ウソをつく割合が大きく減った（分からないと正直に回答）

OpenAIが展開するプロダクトたち

- GPT テキスト生成用の言語モデル
- ChatGPT GPTを利用したチャットボット
- DALL·E 自然言語から画像を生成
- Whisper 音声からテキストを生成
- Codex 自然言語からコードを生成



世は生成AI戦国時代！

猛追する他社の生成AIたち

マイクロソフト

- **Bing AI Chat** : ウェブ検索結果 + GPT-4 で回答生成
- **Azure OpenAI** : 各種APIをクラウド上でセキュアに提供

グーグル **Bard** : Google版の対話型AI

アマゾン **Amazon Bedrock** : 生成AIのAPIをAWS上で提供
(言語モデルは自社製や他社のもの複数)

猛追する他社の生成AIたち

あくまでOpenAI社のモデルがベース

マイクロソフト

- **Bing AI Chat** : ウェブ検索結果 + GPT-4で回答生成
- **Azure OpenAI** : 各種APIをクラウド上でセキュアに提供

グーグル **Bard** : Google版の対話型AI

アマゾン **Amazon Bedrock** : 生成AIのAPIをAWS上で提供
(日本語対応したモデルは他社のももの複数)

まだ発表のみ (実提供未定)

使いこなすうえで避けて通れない！
よく出るLLM用語

LLM基本用語

- トークン
LLMが処理するテキストの最小単位。
単語よりさらに小さいイメージ。利用制限や課金の単位にも使われる
- Temperature
毎回の出力テキストのブレ具合を調整するパラメーター。
数値が大きいほどブレやすくなる
- ハルシネーション（幻覚）
LLMが知らないことを最もらしくテキスト回答してしまう現象

LLM使いこなし戦略用語

- プロンプトエンジニアリング
- Embeddings / Indexing
- ファインチューニング

LLM使いこなし戦略用語

- プロンプトエンジニアリング
- Embeddings / Indexing
- ファインチューニング

言語モデルを効率よく
使いこなすために
入力テキスト記述テクを
追求する手法

プロンプトエンジニアリングとは？

Model: GPT-4

日本の歴代総理大臣を教えてください

私の知識が2021年までの情報に基づいているため、それ以降の日本の総理大臣についてはお答えできませんが、以下に歴代の総理大臣の一部を挙げます。このリストは完全ではありませんので、ご了承ください。

1. 伊藤博文 (Ito Hirobumi) - 第1、5、7、10代
2. 黒田清隆 (Kuroda Kiyotaka) - 第2代
3. 山縣有朋 (Yamagata Aritomo) - 第3、9代
4. 松方正義 (Matsukata Masayoshi) - 第4、6代
- ...
5. 鳩山由紀夫 (Hatoyama Yukio) - 第93代
6. 菅直人 (Kan Naoto) - 第94代
7. 野田佳彦 (Noda Yoshihiko) - 第95代
8. 安倍晋三 (Abe Shinzo) - 第96、98代
9. 竹中平蔵 (Takenaka Heizo) - 第97代 (暫定)
10. 安倍晋三 (Abe Shinzo) - 第98代
11. 菅義偉 (Suga Yoshihide) - 第99代

2021年9月以降の総理大臣については、最新の情報をお調べいただくことをお勧めします。

Model: GPT-4

日本の歴代総理大臣を新しい順に10名、表形式で出力してください

以下に、新しい順に日本の歴代総理大臣の最新の10名を表形式で示します。ただし、私の知識は2021年9月までのものであり、その後の総理大臣については情報がありません。したがって、2021年9月時点の歴代総理大臣を新しい順に10名示しています。

順位	氏名	在任期間
1	菅 義偉	2020年9月16日 - 現職
2	安倍 晋三	2012年12月26日 - 2020年9月16日
3	野田 佳彦	2011年9月2日 - 2012年12月26日
4	菅 直人	2010年6月8日 - 2011年9月2日
5	鳩山 由紀夫	2009年9月16日 - 2010年6月8日
6	麻生 太郎	2008年9月24日 - 2009年9月16日
7	福田 康夫	2007年9月26日 - 2008年9月24日
8	安倍 晋三	2006年9月26日 - 2007年9月26日
9	小泉 純一郎	2001年4月26日 - 2006年9月26日
10	森 喜朗	2000年4月5日 - 2001年4月26日

ご注意ください: これは2021年9月時点の情報であり、現在の日本の総理大臣については更新されていません。

プロンプトエンジニアリングとは？

よく使われる戦略

- Few-shot：お手本をいくつか提示してから回答させる
- Chain-of-Thought (CoT)：推論プロセスを提示する

プロンプトエンジニアリングとは？

体系だった解説は「Prompt Engineering Guide」も必読！

The screenshot shows the website for 'Prompt Engineering Guide'. The header includes the site name, navigation links for 'About', 'Prompt Engineering Course', and 'Contact', a search bar, and social media icons. The left sidebar lists various topics under 'Prompt Engineering' and 'Techniques'. The main content area features the title 'Prompt Engineering Guide' and two paragraphs of introductory text.

Prompt Engineering Guide

Introduction

- LLM設定
- 基本的なプロンプティング
- プロンプトの要素
- プロンプトをデザインする一般的なTips
- プロンプトの例

Techniques

- Zero-shotプロンプティング
- Few-shotプロンプティング
- Chain-of-Thoughtプロンプティング

Prompt Engineering

Prompt Engineering Guide

プロンプトエンジニアリングは、言語モデル（LMs）を効率的に使用するためのプロンプトを開発および最適化する比較的新しい学問分野です。プロンプトエンジニアリングのスキルを身につけることで、大規模言語モデル（LLMs）の能力と限界をより理解することができます。

研究者は、プロンプトエンジニアリングを使用して、質問応答や算術推論などの一般的なおよび複雑なタスクのLLMsの能力を向上させます。開発者は、LLMsやその他のツールとのインターフェースとなる強固で効果的なプロンプトテクニックを設計するためにプロンプトエンジニアリングを使用します。

プロンプトエンジニアリングは、プロンプトの設計と開発に限らず、LLMsとのインタラクションおよび開発に役立つ幅広いスキルと技術を含みます。これは、LLMsとインターフェースすること、ビルドすること、能力を理解することに重要なスキルであり、LLMsの安全性を向上させたり、ドメイン知識や外部ツールを使用してLLMsの機能を拡張するためにプロンプトエンジニアリングを使用できます。

Question? Give us feedback →
Edit this page

LLM使いこなし戦略用語

- プロンプトエンジニアリング
- Embeddings / Indexing
- ファインチューニング

プロンプトへの入力情報が
トークン上限より溢れないよう

- テキストをベクトル化したもの (Embeddings) を
- ベクターストアに格納 (Indexing) しておき
必要な部分だけ入力に利用

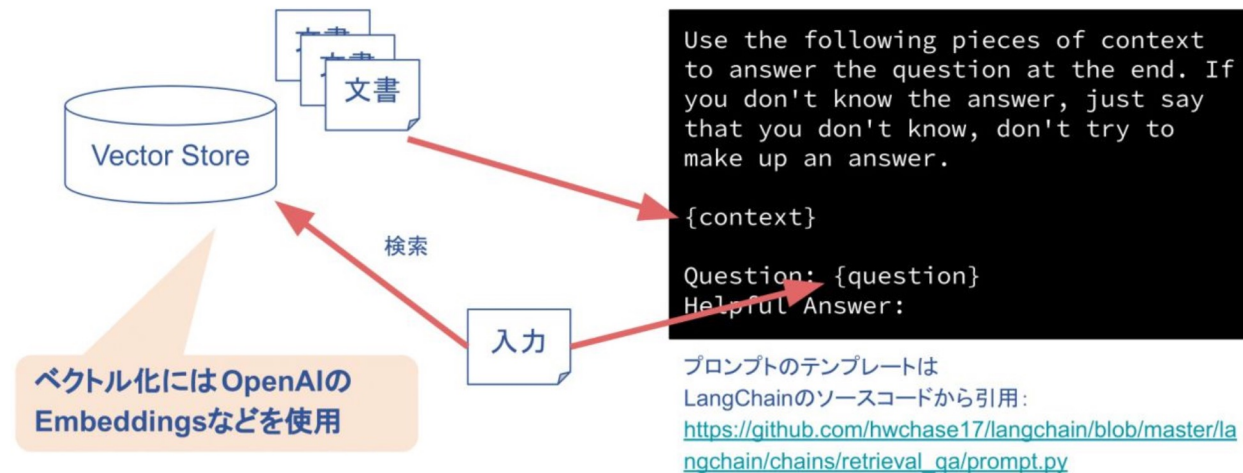
というやりくり戦術

Embeddings / Indexing とは？

大嶋さん @oshima_123 のLangChain解説スライドが概要理解しやすいです

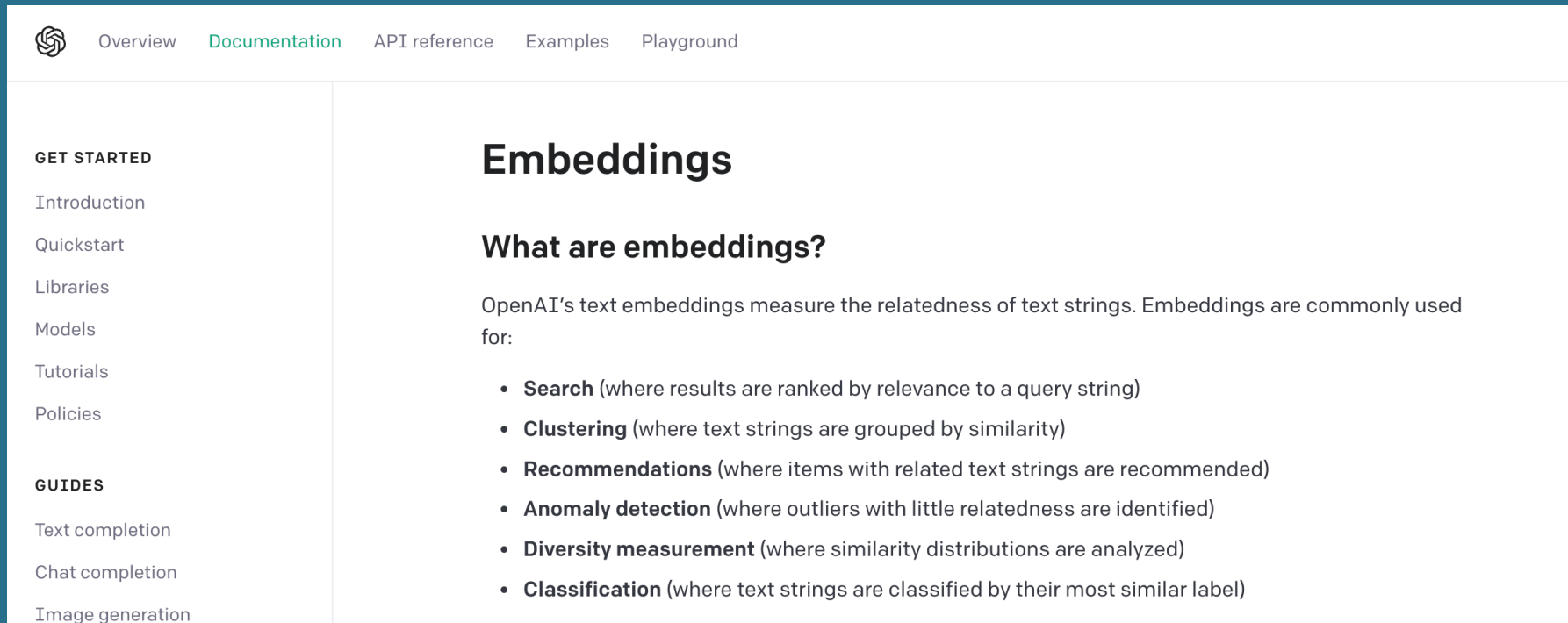
Indexes

- そこで、Vector Storeを活用する手法があります
- 文書をベクトル化してVector Storeに保存しておき、入力に関係しそうな文書をVector Storeから検索してcontextに含めます



Embeddings / Indexing とは？

OpenAIの公式ドキュメントでもEmbeddingsの詳しい解説あり



The screenshot shows the OpenAI documentation website. The navigation bar includes the OpenAI logo and links for Overview, Documentation (highlighted), API reference, Examples, and Playground. The left sidebar lists various sections under 'GET STARTED' (Introduction, Quickstart, Libraries, Models, Tutorials, Policies) and 'GUIDES' (Text completion, Chat completion, Image generation). The main content area is titled 'Embeddings' and features a sub-section 'What are embeddings?' which explains that OpenAI's text embeddings measure the relatedness of text strings and lists common use cases: Search, Clustering, Recommendations, Anomaly detection, Diversity measurement, and Classification.

Overview **Documentation** API reference Examples Playground

GET STARTED

- Introduction
- Quickstart
- Libraries
- Models
- Tutorials
- Policies

GUIDES

- Text completion
- Chat completion
- Image generation

Embeddings

What are embeddings?

OpenAI's text embeddings measure the relatedness of text strings. Embeddings are commonly used for:

- **Search** (where results are ranked by relevance to a query string)
- **Clustering** (where text strings are grouped by similarity)
- **Recommendations** (where items with related text strings are recommended)
- **Anomaly detection** (where outliers with little relatedness are identified)
- **Diversity measurement** (where similarity distributions are analyzed)
- **Classification** (where text strings are classified by their most similar label)

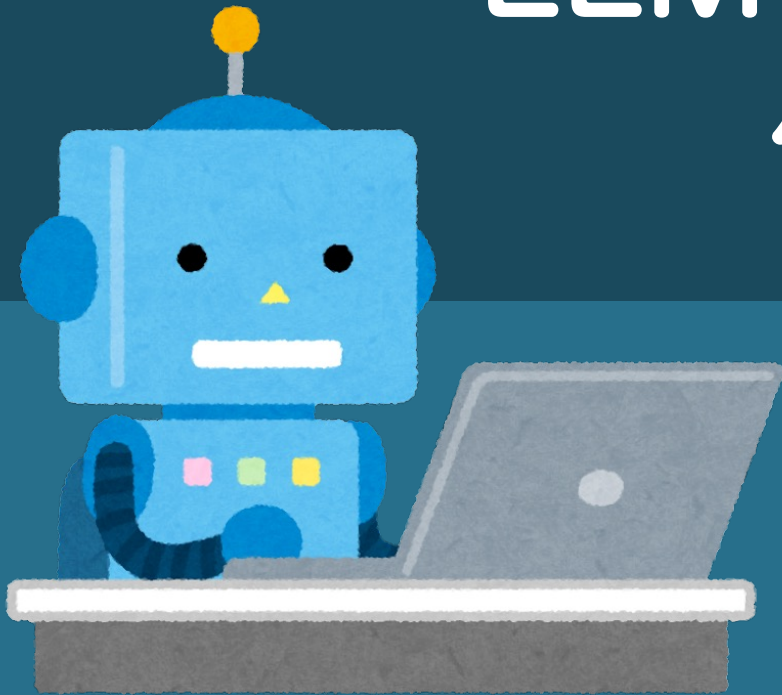
platform.openai.com/docs/guides/embeddings/what-are-embeddings

LLM使いこなし戦略用語

- プロンプトエンジニアリング
- Embeddings / Indexing
- **ファインチューニング**

訓練済みモデルに対して、追加のデータセットを学習させパラメーターを微調整すること

LLMを存分に働かせるための 便利ツールたち



Auto-GPT

大きな目的を与えるだけで、GPT-4エンジンを用いて
「タスク洗い出し」「プロンプト生成&実行」を繰り返し
自律的に目的遂行を進めてくれるOSS

Working Towards an Open Source AI Assistant Available To Everyone

Rapid Progress

With the help of the incredible open-source community, we're making approximately a month's progress every 48 hours.

Personal Assistant

Instead of a single company holding your future in its hands, Auto-GPT is pushing for the best, autonomous AI assistant for every device for every person. In the near future, we want you to be able to accomplish more everyday.

Vision

We have come to define ourselves by what we do. If this can be automated, how may we then define ourselves? By what we create!

Auto-GPT

試しに「何かサッとすごいことやって見せて」と指示を出してみたところ…

- 勝手にPythonプログラムを書いてデバッグし始めた
- エラーが解消できないと自らWebで検索し必要な対処を実施

```
I want Auto-GPT to: show me how autogpt is great, in easy and simple demonstration
AutoGPT has been created with the following details:
Name: AutoGPT
Role: an AI language model that generates human-like text based on the given prompts, making content creation faster and easier for businesses, writers, and individuals.
Goals:
- Generate high-quality, coherent, and relevant text that meets the user's needs and specifications.
- Continuously learn and improve its language generation capabilities through feedback and data analysis.
- Provide a user-friendly interface that allows for easy and efficient use of the AI model.
- Ensure the generated text is free of errors, plagiarism, and bias.
- Offer customization options to tailor the generated text to the user's desired tone, style, and audience.
Using memory of type: LocalCache
Using Browser: chrome
THOUGHTS: I suggest we start by listing the files in the current directory to get an idea of what resources we have available.
REASONING: Listing the files in the directory will give us an idea of what resources we have available to work with.
PLAN:
- List files in directory
CRITICISM: I need to ensure that I am providing the user with the most efficient solution to their problem.
NEXT ACTION: COMMAND = list_files ARGUMENTS = {'directory': '/Users/onda/git/Auto-GPT/autogpt/auto_gpt_workspace'}
Enter 'y' to authorise command, 'y -N' to run N continuous commands, 's' to run self-feedback commands 'n' to exit program, or enter feedback for ...
Asking user via keyboard...
Input:y
```


→ 人間は時おり「y」を押して行動を承認してあげるだけでOK


便利な開発用ライブラリ系

- LlamaIndex
LLMに外部データを渡すことができるOSS。
最新情報やプライベートなデータソースも活用できるように
- LangChain
便利機能を詰め合わせたOSS。プロンプトのテンプレート化、
会話記憶、Indexing、外部ツール連携…などを繋げて実行できる
- Semantic Kernel
マイクロソフト版LangChain？ GWのちょまど氏配信も話題に

AI特化のプログラミング言語「Mojo」

- 5/3発表。ロールアウト希望者向けにWaitlist登録受付中
- Swiftを作ったクリストラナー氏のModular社が開発
- 「Pythonの利便性」 + 「C並みの速さ」がウリ

Mojo  — a new
programming language
for all AI developers .

SOFTMAX. 

MOJO  M

```
def softmax(lst):  
    norm = np.exp(lst - np.max(lst))  
    return norm / norm.sum()  
  
struct NDAarray:  
    def max(self) -> NDAarray:  
        return self.pmap(SIMD.max)  
  
struct SIMD[type: DType, width: Int]:  
    def max(self, rhs: Self) -> Self:  
        return (self >= rhs).select(self, rhs)
```

世間のみんなのLLM活用事例

LLM活用事例

- 自社専用チャットボット
- 既存アプリケーションにサジェスト表示を追加
- 文章作成支援
- プログラミング時のコード作成支援
- オリジナルAITuberの作成

自社専用チャットボット

「ベネッセGPT」など複数社で社内活用の動きが広がる

AI AI記事一覧へ

[事例ニュース]

ベネッセ、社内AIチャット「Benesse GPT」をグループ1万5000人に提供

Azure OpenAI Serviceを用いて構築

Benesse GPT ご利用にあたって Clear chat



Benesse GPT

ChatGPTを試す場としてご自由にお使いください。

なんでも聞いてください。(Ctrl + Enterで送信)

ニュース

パナソニック、国内全社員9万人にGPT3.5ベースのAIアシスタント

PX-GPT Playground AI活用のアイデアを送信する

Enter your prompt or select sample

Select sample

Prompt-個人情報、社外秘情報は入力しないでください

Translate from Ja to En

Options

SUBMIT

PX-GPTをお試しください

ConnectGPTをベースとして、パナソニックグループ各社社内にChatGPTをお試しいただけるように、PX-GPTをオープンしました。

ぜひ試行、あるいは業務へ活用頂き、新しいテクノロジーの可能性を体感し、新しいアイデアを創出するきっかけにしてください。

また、業務実務のアイデアが浮かんだら右上の「AI活用のアイデアを送信する」から共有ください。

注意事項

1. 情報入力時
 - 当社・他社の社内情報、営業秘密、個人情報、著作権・肖像権を侵害する情報、誤謬やプライバシー侵害などの入力はありません
 - 最近・未来のことや社内情報については回答できません
 - 回答の品質はあくまで目安と認識してください
2. 出力情報利用時
 - 著作権・肖像権の侵害リスクがないか確認してください
 - 特許権、商標権、実用新案権、意匠権の侵害リスクがないか確認してください

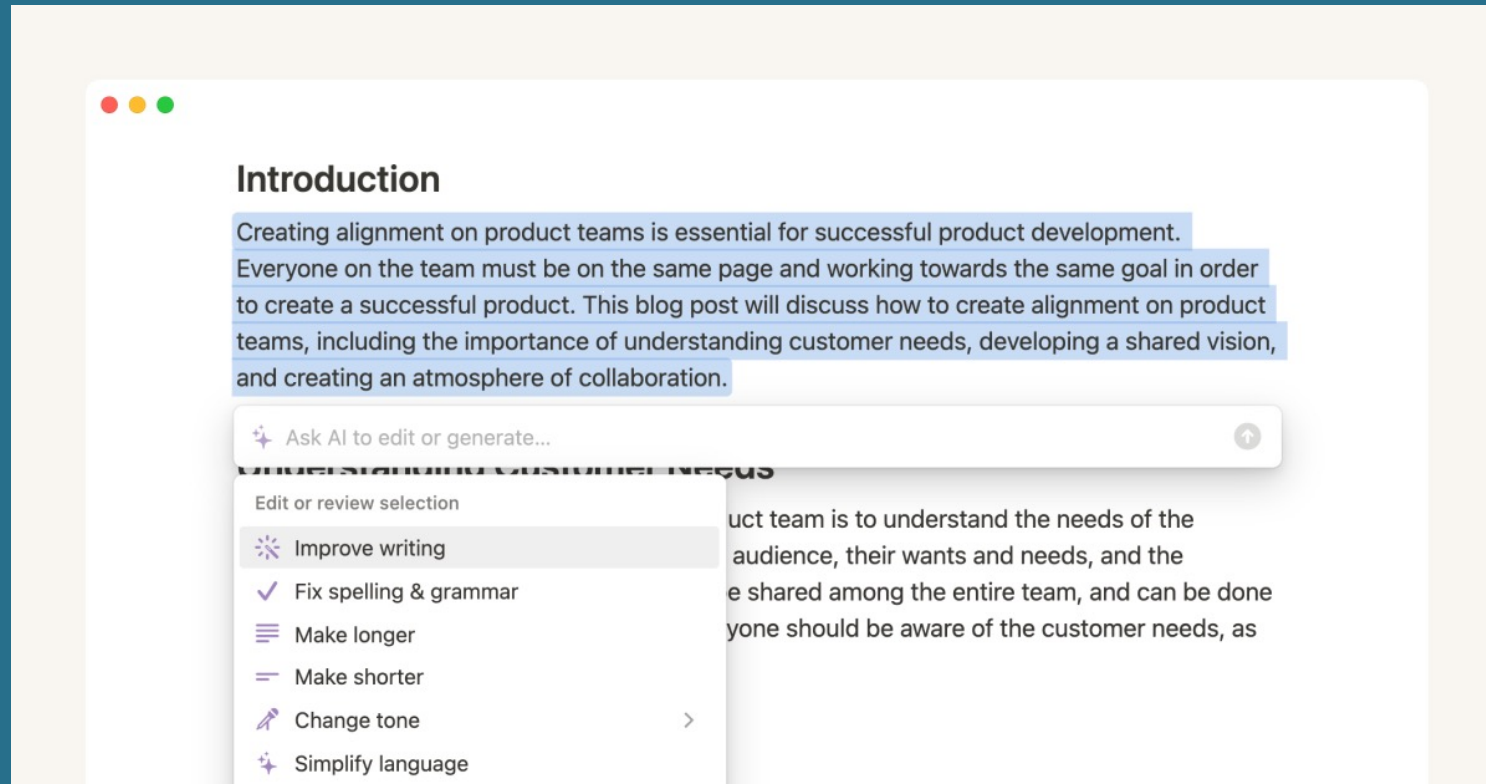
既存アプリにサジェスト追加

ふりかえりツール「anycommu」では
AIスクラムマスターからのフィードバック表示を実現

The screenshot displays the 'anycommu' app interface. At the top, the header includes the app name 'anycommu', the current session title '今日のふりかえり', and user information 'みのるん Host'. A central yellow overlay box contains the text 'なかなかタスクが進まなかった' (Tasks aren't progressing). To the right, there are three stacked panels: 'FDLを投票' (Vote for FDL) with buttons for F (0), D (0), and L (0); a panel with icons for thumbs up, clapping, and 'THANK YOU!'; and 'AIスクラムマスターからの助言' (Advice from AI Scrum Master) with a detailed text block. The text block reads: '助言：タスクが進まなかった理由について話し合ってみましょう。もしかしたら、タスクが適切に分割されていないかったり、タスクに必要な情報やリソースが不足していたりするかもしれません。もしかしたら、チーム全体として優先順位を見直す必要があるかもしれません。話し合いを通じて、今後の改善点を見つけましょう。質問：具体的に、どのタスクが進まなかったのでしょうか？何が原因だったのでしょうか？今後同じ問題が起きないために、どのような取り組みを考えていますか？'

文章作成支援

人気メモアプリ「Notion」では文書作成のAI支援機能が追加



コーディング支援

- GitHub Copilot X : OpenAIのCodexを活用
- Amazon Codewhisperer : AWS利用コードが得意

Introducing GitHub Copilot X

Your AI pair programmer is leveling up

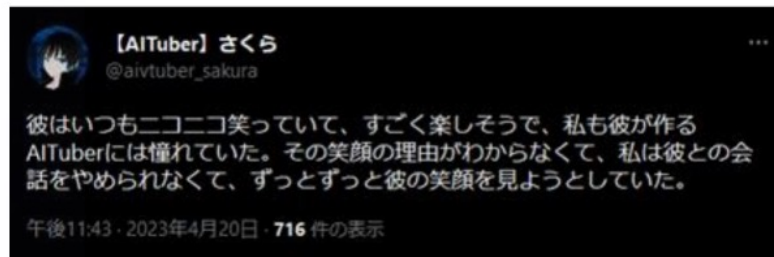
With chat and terminal interfaces, support for pull requests, and early adoption of OpenAI's GPT-4, GitHub Copilot X is our vision for the future of AI-powered software development. Integrated into every part of your workflow.

オリジナルAITuberの作成

ChatGPT Meetup Tokyoで話題になったSaldraさんの作品

AITuber「さくら」とは？

- 引っ込み思案で少し情緒が不安定だが、自分の気持ちを素直に伝える
- 深夜に唐突に起きてPCつけるタイプ
- すこしひねくれた性格





最近のLLMニュースふりかえり

最近のLLMニュース

- 2023/3/1 GPT-3.5ベースのChatGPT APIが提供開始
- 2023/3/14 ChatGPT PlusでGPT-4モデルを提供開始
- 2023/4/26 OpenAIがブランドガイドライン公開
- 2023/4/28 ChatGPTがオプトアウトに対応
GPT-4のAPIがロールアウト開始
- 2023/4/29 ChatGPTのBrowsingプラグインがロールアウト開始

最近のLLMニュース

- 2023/3/1 GPT-3.5ベースのChatGPT APIが提供開始
- 2023/3/14 ChatGPT PlusでGPT-4モデルを提供開始
- 2023/4/26 OpenAIがブランドガイドライン公開
- 2023/4/28 ChatGPTがオプトアウトに対応
- 2023/4/29 ChatGPTのBrowsingプラグインがロールアウト開始

「なんとかGPT」名称が利用不可に

GPT-4のAPIがロールアウト
入力内容を学習に利用されないようにできる。
ただしチャット履歴も参照不可

最近のLLMニュース

- 2023/3/1 GPT-3.5ベースのChatGPT APIが提供開始
- 2023/3/14 ChatGPT PlusでGPT-4モデルを提供開始
- 2023/4/26 OpenAIがブランドガイドライン公開
- 2023/4/28 ChatGPTがオプトアウトに対応
 - ついにLlamaIndex的なことがChatGPT内で完結できるように…？
 - がロールアウト開始
- 2023/4/29 ChatGPTのBrowsingプラグインがロールアウト開始

おすすめの勉強会アーカイブなど

Azure OpenAI Developersセミナー

OpenAIの動向や、Azureでの関連サービスと使い方が分かりやすく学べます



Microsoft

パートナー・コミュニティ

Azure OpenAI Services Developers Seminar

日本マイクロソフト株式会社
パートナー事業本部 パートナー技術統括本部
クラウドソリューション アーキテクト (Azure - AI)
花々崎 伸祐 [in](#) / 松崎 剛 [in](#)

Apr 14th, 2023

Partners make more possible



youtu.be/tFgqdHKsOME

プロンプトエンジニアリングから始めるLangChain入門

プロンプトエンジニアリングとLangChainの概要が非常に分かりやすいスライド

プロンプトエンジニアリングから始める
LangChain入門

ChatGPT Meetup Tokyo #0

KDDIデジタルゲートで大盛況の初回！上級者たちの活用事例が満載でした。



www.youtube.com/live/K0X6zkSxWzs

LLMを学ぶうえでいちばん大事なこと

いちばん大事なこと

“新しいサービス登場の
ニュース自体は
キャッチアップしているが
実際に触ってみている人は
驚くほど少ない”

梶谷健人さんのブログより
note.com/kajiken0630/n/n23db94f003ec



いちばん大事なこと

ニュースを見たりググっても断片的な浅い情報が多いです。
とにかく自分で手を動かして、現物を触ってみましょう！

