

～ 日本の次のトレンドを掴む ～

「米国企業の生成AIトレンド」から見るキーワード 1/2

エンタープライズ企業における、下記2つが進んでいる



マルチモデル化

マルチモデル化が進む背景

2023年は、大半の企業が1つのモデル (多くても2つ)を活用していたが…

多様なニーズに対応するため

パフォーマンス・サイズ・コストに基づいて、
活用シーンを最適化している

サービスのロックインを回避して、迅速に技術を取り込むため

複数モデルを試して、
日々進化する技術分野をいち早く活用しようとしている



オープンソース の利用

オープンソース普及の背景

2023年の市場シェアは 80%-90% がクローズドソースだったが
2024年では、46%がオープンソースを強く好む/好むという結果に

コストよりも、コントロール・カスタマイズ性を重視

- 自社データのセキュリティを確保し、モデルが特定の出力を生成する理由を理解できる
- ユースケースに応じた効果的なFine-tuningができる

～ 日本の次のトレンドを掴む ～

「米国企業の生成AIトレンド」から見るキーワード 2/2

ピックアップサービス

ディファイ

Dify

概要

直感的に生成AIサービスを開発できるプラットフォーム

マルチモデル

オープンソース

ノーコード

アプリはオープンソース、SaaSも提供

開発会社

2023年設立
米国のAIテクノロジー企業
LangGenius社

特徴



数百種類のAI言語モデルが使える

用途に合わせてモデルを選べる、複数モデルの回答を同時に比較できる

例) 難しい質問 → GPT4, Claude3 Opus、簡単な質問 → groq Llama3



プログラミング知識不要で生成AIアプリが作れる

簡易的なチャットボットから、複雑な条件を伴う処理を行う自動ツールまで幅広いニーズをカバー



RAGエンジンによるナレッジ機能がある

組織内に蓄積されたデータをもとにした回答するチャットボットが作成できる



柔軟性・拡張性が高い

様々なツール・APIとの連携・自由度の高いワークフロー設定により、独自のツール作成できる

ココがすごい

数分で、レゴブロックのように処理を組み合わせて、チャットボットが作れる

今ままで



生成AIを活用したアイデアは浮かぶのにアプリを作ることが出来ない…

これから



浮かんだアイデアをもとにプロトタイプをばぱっと作ってみた

5分で作れる!! 社内規定に関するチャットボット

1. アプリ名入力 社内規定お助け君
2. プロンプト入力 あなたはバックオフィスの…
3. モデルを選択 GPT4
4. ナレッジの選択 社内規定Wordデータ
5. “アプリ公開” ボタン押下!!

ユースケース

営業支援ツール

次の商談先の会社URLを渡すと、下記の2点をnotionページに保存

- 会社情報の要約
- 商談先の事業内容から、自社サービスの活用アイデア

社内QAの自動更新

CSVに入力された質問文をもとに、社内規定から情報収集し、Googleスプレッドシートを更新

論文検索の自動化

キーワードをもとに、arXivを検索して、関連性の高い5つの論文を探し、要約作成

生成AIの先駆者・チャレンジャーを追う

2024.09-10

OpenAI

より自然に会話できる高度な音声機能
「Advanced Voice」が利用可能に

- ✓ AIは、話し手の声のトーンから感情を感じ取り、それに応じて応答を調整する
- ✓ AIが話している途中で話し手が遮ったときは、AIは話すのを止めて聞いてくれる
- ✓ 日本語を含む50ヶ国語以上で会話ができる
- ✓ 音声は9種類から選択可能

ユースケース

モックインタビュー

面接官として、データサイエンティストとして応募した私を厳しめの面接をしてください。

言語学習

来週、ハワイに旅行に行きます。旅行で使える英語のフレーズをいくつか練習したいです。

GPT-4シリーズに比較して
より考えるモデル「o1」をリリース

出典: techcrunch, First impressions of OpenAI o1: An AI designed to overthink it, <https://techcrunch.com/2024/09/13/first-impressions-of-openai-o1-an-ai-designed-to-overthink-it/>

- ✓ “考えてから回答する”点が特徴。大きな問題を小さなステップに分解し、そのステップの1つが正しいか間違っているかを識別しようとするモデル (マルチステップ推論)
- ✓ 「o1」は、推論や複雑な質問への回答に優れているが、GPT-4oよりも約4倍コストが高い
- ✓ ユニークな価格設定で、入力トークン+ 出力トークン+ 推論トークンから計算される (多くのモデルは、入力トークン+ 出力トークンの価格設定)
- ✓ OpenAIはヘルプページで「ほとんどのプロンプトではGPT-4oがまだ最良の選択肢である」としている

Claude

Anthropic社による
大規模言語モデルを搭載した 生成型対話AI の最新モデルにて
Computer Use 機能がリリース

Claude 3.5 Sonnet

ひとつ前のモデル Claude 3は一般公開されたAIで、初めてIQ100を超えたAI

Computer Use 機能とは？

コンピュータを人間のように操作できる



例：
画面を見て、カーソルを動かし、
ボタンをクリックし、テキストを入力

主な活用例



- 反復的なプロセスの自動化
- ソフトウェアのテストとQA
- リサーチなどのオープンエンドなタスクの実行が可能

参考：
<https://www.maximtruth.org/p/ais-ranked-by-iq-ai-passes-100-iq>
<https://www.anthropic.com/news/3-5-models-and-computer-use>
<https://www.anthropic.com/claude/sonnet>
<https://docs.anthropic.com/en/docs/build-with-claude/computer-use>

生成AIの先駆者・チャレンジャーを追う - 1/3

先駆者 - OpenAI

1/2

開発者向けの新機能4つ - カスタマイズ可能性・作業効率UP, コストDOWN

モデル蒸留

大規模モデルの出力で小規模モデルの
ファインチューニングができる機能

従来は開発者が複数ツールを使ってやっており、エラーも出やすい
デリケートな作業であったが、その複雑な作業を1つのAPIで実現

ビジョンファインチューニング

テキストだけでなく画像を使用して
モデルをファインチューニングできる機能

モデルが画像を理解し認識する能力が向上するため、自動運転車の
オブジェクト検出や医療画像の解釈など、さまざまな用途での利用が可能

プロンプトキャッシング

開発者がAIに指示を出す際に使用する
命令文の一部(プレフィックス)を一時的に保存する機能

プレフィックスを保存することで、同じプレフィックスを使う新しい命令が
あった場合、その入力のコストを50%削減することができ、
AIアプリケーションの運用コストを削減できる

リアルタイムAPI

従来の音声対応アプリケーション開発をシームレスにできる機能

従来の音声対応アプリケーション開発は、音声からテキストへの変換、
テキストの言語モデルによる処理、テキストから音声への変換
といった複数のステップを必要とした



生成AIの先駆者・チャレンジャーを追う - 2/3

先駆者 - OpenAI

2/2

Canvas - ライティング・コーディングをするための新しいインターフェース

- ✓ ChatGPTのメインのチャットウィンドウとは別のウィンドウで、コードや長いテキスト等の大きいコンテンツを表示可能
- ✓ コードやテキストをキャンバスウィンドウに表示したまま、チャットウィンドウでChatGPTと会話を続けることができる

ライティング

The Value of Design

make it more creative

Introduction

In an increasingly competitive and fast-paced world, design has emerged as a critical differentiator that can make or break a product, service, or brand. For instance, Acme Co.'s focus on sleek, user-friendly design has been a major factor in the success of its products, helping it stand out in the crowded tech market. Gone are the days when design was considered merely an aesthetic addition; today, it's a

<https://openai.com/index/introducing-canvas/>

Canvas内で
直接テキストを
手動で編集できる

5つの便利ツール* (ショートカット機能)

- 編集提案: テキストに直接提案やフィードバックを表示
- 長さ調整: 文章の長さを短くしたり、長くしたりするオプションを提供
- 読書レベル調整: 幼稚園レベルから大学院レベルまで、読みやすさを調整
- 最終的な推敲: 文法、明確さ、一貫性をチェック
- 絵文字追加: 適切な場所に自動的に絵文字を挿入

Reading level

*キャンバスウィンドウの右下にあるコーディングアイコンから、アクセスできる機能

コーディング

```

30     .route("/greet", post(greet_user));
31
32     // Define the address to run the server on
33     let addr = SocketAddr::from(([127, 0, 0, 1], 3000));
34     println!("Listening on http://{}", addr);
35
36     // Run the server
37     axum::Server::bind(&addr)
38         .serve(app.into_make_service())
39         .await
40         .unwrap();
41 }
42
43 add routes for signup and login

```

<https://openai.com/index/introducing-canvas/>

- Canvasウィンドウで直接、生成されたコードを編集できる
- コードの特定のセクションを選択すると、その部分だけにフォーカスした小さなChatGPTプロンプトウィンドウが表示される

5つの便利ツール (ショートカット機能)

- コードレビュー: コード改善のための提案を受け取れる
- ログを追加: デバッグを容易にするprint文をコードに挿入
- コメントを追加: コードの適切なセクションに自動的にコメントを追加生成
- バグの修正: エラーを含むコードを自動的に書き換え
- 言語の変換: コード言語を別の言語に変換 (Python, Java, C++など)

Fix bugs

<https://openai.com/index/introducing-canvas/>

参考: pocket-lint, 「ChatGPT finally got the upgrade writers and coders have been waiting for」
<https://www.pocket-lint.com/how-to-use-the-canvas-feature-in-chatgpt/>

生成AIの先駆者・チャレンジャーを追う - 3/3

チャレンジャー - Nvidia

Nvidiaの革新的AIモデル「NVLM 1.0」の全貌

- ✓ Nvidiaは、オープンソースの強力な人工知能モデル「NVLM 1.0」を発表
- ✓ NVLM1.0は「フロンティアクラスのマルチモーダル大規模言語モデルファミリーであり、視覚・言語タスクにおいて最先端の結果を達成し、主要なプロプライエタリモデル(GPT-4oなど)やオープンアクセスモデルに匹敵する」と研究者らは論文で発表している

モデルの特徴

✓ マルチモーダル性能の強化・テキスト処理能力の進化

720億パラメータを持つ主要モデル「NVLM-D-72B」は、視覚と言語のタスクにおいて特に卓越したパフォーマンスを発揮

視覚情報の統合能力

- NVLM-D-72Bは画像やミームを解析し、それらの視覚情報を言語タスクに統合する能力が高い
- 複雑な視覚とテキストの入力を効率的に同時に処理することが可能

テキスト処理性能の進化

- マルチモーダルトレーニングを経た後のNVLM-D-72Bは、テキストのみのタスクでのパフォーマンスが向上
- 他のモデルがテキスト性能を落とす中、このモデルはキーテキストベンチマークで平均4.3ポイントの精度向上を達成

✓ オープンソース

- モデルのウェイトを公開するだけでなく、訓練コードの公開も予定している
- 研究者や開発者に前例のないアクセスを提供し、高度なAIシステムを広く共有することで、技術の民主化を図る

生成AI ピックアップNEWS

日立と静岡銀行、勘定系システム開発に生成AI 期間短縮

- ✓ 勘定系システムの追加機能の開発効率化へ生成AIの活用を使うと発表
- ✓ 金融機関の基幹業務を担うシステム開発への生成AI活用は国内初
- ✓ 2025年4月には実用化。将来は導入範囲を広げる予定

追加機能開発の工程

- ①新たな決済手段など追加したい機能の要件定義
- ②システムに落とし込むための設計
- ③設計に基づきプログラミングを書く「製造」
- ④個々のプログラムの動作を確認する単体テスト
- ⑤システム上に組み込んで正しく動くかを見るテスト

全工程の
約3割を占める
製造工程



情報のインプット・アウトプットの相関が
分かりやすい**工程から生成AI活用で省人化**

初期実験の成果

約2カ月間の初期実験の結果…



下記をもとに、弊社作成

出典：日本経済新聞, 日立と静岡銀行、勘定系システム開発に生成AI 期間短縮,

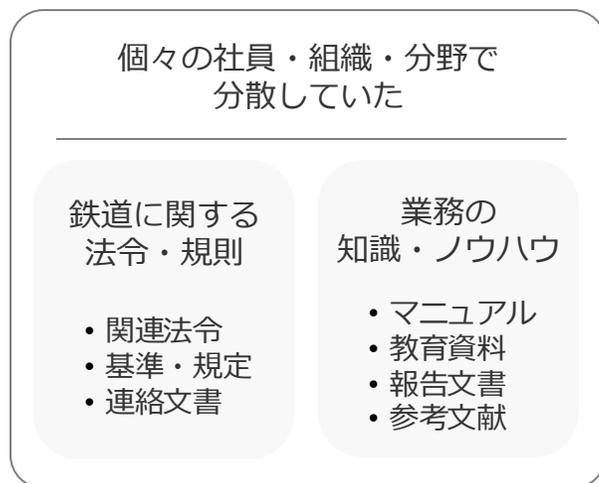
https://www.nikkei.com/article/DGXZQOCC115Y30R11C24A000000/?n_cid=SNSTW001&n_tw=1728969451

生成AI ピックアップNEWS

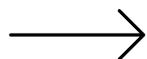
JR東日本、鉄道固有の知識を学習した「鉄道版生成 AI」開発へ

- ✓ 「鉄道版生成 AI」の本格的な開発が着手され、2027年度末の完成を目指しています
- ✓ これまで個々の社員や組織、分野ごとに分散していた業務知識やノウハウをAIで一元管理します
- ✓ AIの活用により日常業務を効率化し、社員が創造的な業務に注力できる環境を整備します
- ✓ 本プロジェクトの成果を他鉄道事業者にも展開する可能性が検討されています

「鉄道版生成 AI」で目指す姿



横断的に
情報集約



創造的業務に集中

- ・新事業開発
- ・地域活性化
- ・お客様サービスの充実

業務遂行の支援の例

問い合わせ対応サポート

お問い合わせやご案内時に、大量の社内の基準や規定を調べ、迅速かつ正確にご返答

社内調整業務の負担軽減

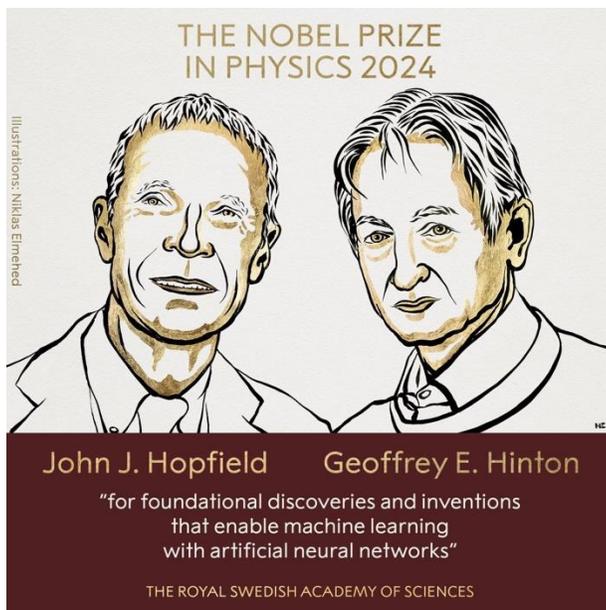
分野横断的な調整が必要な際、AIから他分野の知識を得られるため、業務負担を軽減(従来は各分野を担当する社員への問い合わせが必要だった)

2024年のノーベル賞

ノーベル物理学賞

ジョン・J・ホップフィールド氏とジェフリー・E氏
の下記に対して授与される

人工ニューラルネットワークを用いた
機械学習を可能にする
基礎的な発見・発明



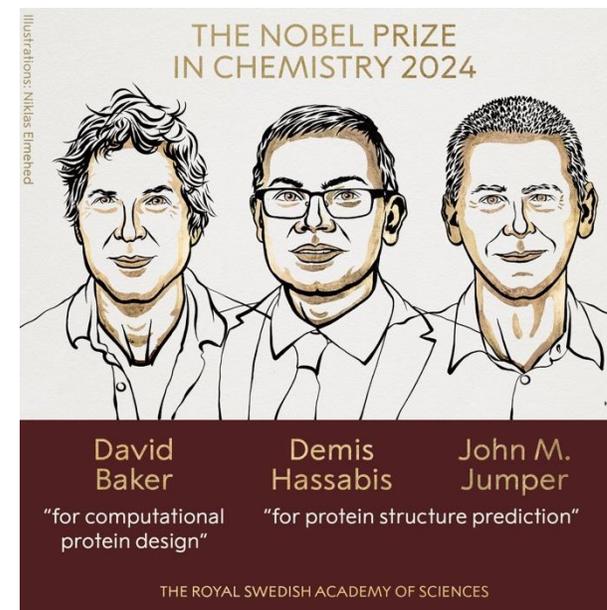
出典 : <https://x.com/NobelPrize/status/1843589140455272810>

ノーベル化学賞

それぞれに授与される

デイビッド・ベイカー氏
「計算によるタンパク質設計」

デミス・ハサビス氏とジョン・M・ジャンパー氏
「タンパク質構造予測」

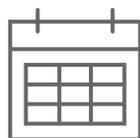


出典 : <https://x.com/NobelPrize/status/184395119796077760>

AIブームを牽引する会社のCEOが見据える世界

OpenAIのCEO サム・アルトマン

知能の時代 (The Intelligence Age)
というブログ記事にて、下記を述べている



数千日後に
人間の知能を超えた人工知能
“超知能AI” が実現するかもしれない

AnthropicのCEO ダリオ・アモデイ

愛の恵みである機械 (Machines of Loving Grace)
というブログ記事にて下記を述べている

2026年に強力なAIが実現すれば、
人間の寿命が150歳まで延びる可能性がある

AIを活用した生物学と医学の進展により、
人間の生物学者が50~100年かけて達成する
進歩を5~10年で実現できるため



下記の事実からも、十分に可能性はある

- カメなど一部の動物は200年生きる
- 20世紀に平均寿命が約2倍に
- ラットの最大寿命を25~50%延ばす薬も既に存在