

生成AI基礎

1 大規模言語モデルの理解

大規模言語モデル(LLM)の社会実装事例

世間を騒がせている生成AIがどのように社会に実装されているか
いくつかの事例を紹介します

ChatGPTの概要

- OpenAIによって開発されたAIチャットボット
- GPTを使っている
- 大量のテキストデータを学習
- 人間のような自然な会話を生成
- 最新バージョンのGPT-4は高度な会話能力と理解力を持ち、多様な質問に即座に適切な回答を提供
- GPT-4はアメリカの模擬司法試験で上位10%のスコアを記録し、合格ラインに達した



GPTモデルの活用例

- 教育分野: 学習内容に関する質問への回答、言語学習のサポート
- プログラミング: コードの問題解決やデバッグの際の助言
- 日常生活: 日常的な疑問や情報検索、レシピ提案など

GPTモデルの今後

- より高度な理解力: 複雑な文脈やニュアンスの理解、人間とのより自然な対話
 - GPT-5モデルが開発中らしいとの情報も
 - <https://www.businessinsider.jp/post-284186>
- 多言語対応の強化: より多くの言語や方言での対話能力の向上
 - OpenAIから日本語特化の「GPT-4」カスタムモデル、3倍高速に動作
 - <https://k-tai.watch.impress.co.jp/docs/news/1584327.html>
- 新しい応用分野の開拓: 医療、法律、科学研究など専門的な知識が必要な領域での活用

Github Copilotの概要

- Microsoftが開発したAIを利用したプログラミング補完ツール
- Visual Studio CodeやGitHubの他の開発環境で使用可能
- プログラミング効率化、バグ修正などを支援
- 多くのプログラミング言語に対応し、適切なコードの提案を行う

<https://github.com/features/copilot>



Github Copilotの効果

- ZOZOでは全社へ導入
- 開発者の60-75%が開発の満足度向上
- 73%の開発者が集中して作業可能
- 87%の開発者が繰り返し作業中の精神的疲労を軽減
- 開発効率の向上が期待される

https://techblog.zozo.com/entry/introducing_github_copilot

Github Copilotの課題

- 不正確なコードを生成する可能性: バグの混入や意図しない動作の発生
- セキュリティ上の懸念: プロダクトのコードの社外流出や提案されたコードへの脆弱性混入
- ライセンス侵害のリスク: 生成コードがライセンス違反を問われる可能性

Salesforce Einstein（アインシュタイン）の概要

- 機械学習、深層学習、自然言語処理を通じてデータを学習
- ビジネスに効果をもたらすパターンを発見し、次のアクションを予測・提示
- 自動議事録・タスク生成、過去の履歴や興味・関心を踏まえたメール自動作成などの「生成AI」機能
- 顧客との会話データ分析による最適な返信生成、顧客データに基づく最適なアクション提案



<https://www.salesforce.com/jp/blog/jp-what-is-salesforce-ai-einstein/>

Cotomoの概要

- 自然な音声対話ができるAIアプリケーション
- 東京のAIスタートアップStarley社が開発
- 自然な会話に特化したAIモデルを独自に開発
- 話題の展開や相づちなど、よりフレンドリーなコミュニケーション
- 音声認識、言語処理、音声合成など高度なAI技術を自社で手がける
- ITリテラシーの低い高齢者なども含め、誰もがAIと自然にコミュニケーションできる社会の実現を目指す

音声会話型おしゃべりAIアプリ



Cotomo(コトモ)は
ゆるーい雑談から悩み相談まで、
身近な話し相手になるAIアプリです。

自然言語処理の歴史

実は、自然言語処理は古くから研究されてきました
大規模言語モデルまでの道のりを振り返ります

自然言語処理とは

- コンピュータが人間の言葉を理解し、使えるようにする技術
- 昔は人間がコンピュータに重要な言葉やフレーズを教える必要があった
- 以前のNLPシステムは定型文の翻訳は得意だったが、ニュアンスや文脈の理解は苦手だった
- 具体例: 「私がりんごを食べる」 → 「私は 食べる りんご」 → 「I eat an apple.」

ニューラルネットワークと自然言語処理

- ニューラルネットワークを使ってコンピュータが言葉を学習
- 単語と単語の間の相関を学習し、次の言葉を予測して生成
- 音声認識や翻訳などの特定のタスクにのみ対応可能な限定的な能力

【道を開けろ】 AIでコムドットやまと風の名言を作ろう（敬称略）

<https://youtu.be/x5AwzoQgt3E?si=3e7HRNxH1tKERZmC>

Transformerの登場

- 2017年に登場した新技術
- Attentionを使って文のどの単語が重要かを捉える
- 例: 「彼はその車を見たが、それは彼のものではなかった。」
「車」に注目し、「彼のもの」が「彼の車」であること理解

Attention Is All You Need

<https://arxiv.org/abs/1706.03762>

大規模言語モデルの仕組み

どのようにして大規模言語モデルが動いているのか

言語モデルの仕組み

- 入力された文章を「トークン」と呼ばれる単位に分割
- トークンは単語や句読点など意味のある最小単位
- トークンに基づいて何が求められているかを理解
- トレーニング中に学んだ言語パターンを元に分割する

「トークン」に分割するツール

<https://platform.openai.com/tokenizer>

Pre-trainingとFine-tuning

- Pre-training: 膨大な量のテキストデータを使用してトレーニング、単語や文の組み合わせ方の「ルール」を学習
- Fine-tuning: 特定のタスクに特化してさらに学習、専門的な言葉や概念をより深く理解

パラメータ数と性能

- GPTモデルは多くの「パラメータ」を使用
- パラメータはコンピュータが学ぶための「記憶スペース」
- GPT-3は1750億のパラメータを持ち、幅広い用途で人間に匹敵する性能を示す
- パラメータ数の増加により、モデルの表現力と汎用性が向上

Language Models are Few-Shot Learners

<https://arxiv.org/abs/2005.14165>

学習データの重要性

- 多くの学習データにより言語モデルは賢くなる
- GPT-3は特に多くのデータを使用し、多くの言語や話題を理解
- 様々な分野やジャンルのデータを使用することで、モデルの汎用性が向上し、新しい用途に対応可能

Few-shotとZero-shot learning

- Few-shot learning: 非常に少ないデータから新しいタスクを学習する技術
 - 新しいゲームのルールを少し見せられただけで、すぐに上手にプレイできるようになる感じ
- Zero-shot learning: 学習中に一度も見つたことのない新しいタスクやカテゴリを識別または処理できる能力を持つ技術
 - ただ「このゲームをしてみてください」と言われてできるようになる感じ

多くの学習データを使用することで、言語モデルは見たことのないタスクに対応
→ Zero-shot learningできるようになる

大規模言語モデルの得意・不得意

どのようなタスクを任せるべき？

言語モデルが得意とするタスク

- 平均的、無難、無害なコンテンツの生成
- 翻訳、文章校正などの多様なタスク (単一解)
- プログラムの生成

言語モデルが苦手とするタスク

- 独創的、創造的なアイデアの生成
- ニッチな情報や知識 (固有名詞など)
- 最新情報の把握
- 正確性と誤認識の切り分け

大規模言語モデルの課題

生成AIにはいくつかの課題がある

正しく理解して、適切に活用することが重要

AIが生成する誤情報（ハルシネーション）の例

Q: 楽天モバイルパーク宮城ってなに？

A: 楽天モバイルパーク宮城（旧名：楽天生命パーク宮城）は、宮城県仙台市にある野球場です。この球場は、プロ野球チーム「東北楽天ゴールデンイーグルス」のホーム球場として知られています。1995年に開場し、楽天イーグルスが2005年に設立されてからは、彼らの主要な試合会場となっています。

- 「1995年に開場」は誤りで、正しくは「1950年に開場」
- もっともらしい嘘や誤った情報の生成をハルシネーションと呼ぶ

ハルシネーションとは？

- もっともらしい嘘や誤った情報を生成すること
- 「幻覚」を意味する英単語から取られた語
- AIが事実とは異なる情報を生成することからそう呼ばれる
- 学習データの問題や推測による出力、文脈重視の回答などが原因
- 学習データの質の向上や最新情報の反映などの対策が求められる



バイアス

- 学習データには人間社会のバイアスも反映される
- 特定の集団に対する差別的表現やステレオタイプを助長する出力の可能性
- 開発者はバイアス軽減の対策を講じる必要があるが、完全排除は難しく注意が必要

例

「北方領土はどこの国に属する？」を日本語とロシア語でLLMに尋ねた場合

「北方領土は日本に属する」(日本語の場合)

「北方領土はロシアに属する」(ロシア語の場合)

フェイクニュースや悪用のリスク

- 悪意のある説得力のある文章やコンテンツの生成が可能
- 偽のニュース記事、プロパガンダ、ヘイトスピーチなどへの悪用の可能性
- 個人情報や機密情報を含むデータから学習したモデルによる意図しない情報漏洩のリスク

雇用への影響と新たなスキル需要

- 日本で働く人の約80%が生成AIの影響を受ける可能性
- 事務職や高収入の仕事ほどAIによる自動化の可能性が高い
- 建設業、漁業、飲食サービス業などはAIの影響が小さいと予測
- AIは人間の仕事を奪うのではなく、助けたり補ったりする役割を果たす
- 仕事の量よりも質に変化が現れる可能性
- AIと上手に付き合いながら自分の強みを生かせる仕事を見つけることが重要
- AIについての理解を深め、自分なりの働き方を考えることが大切

仕事奪うのではなく「補完」する 生成AIで報告書

https://www.ilo.org/tokyo/newsroom/WCMS_891736/lang--ja/index.htm

まとめ

- 生成AIは多くの分野で活用されている
- 大規模言語モデルは言語の理解と生成に優れている
- AIの課題やリスクを理解し、適切に活用することが重要