

基盤モデルで進化した生成AIをどう使うか

◆ロボットとのコミュニケーションは生成AIにより大きく変わるか

Hewlett Packard Enterprise (HPE) は2023年4月17～21日に開催されたハノーバーメッセにおいて、AIスタートアップAleph Alphaの生成AIを用いたデモを行った。工場作業員が産業用ロボットを操作する際、写真とテキストを産業用ロボットに送り、音声の指示をもらうコミュニケーションで作業を進める。

Aleph Alphaが独自に開発した大規模言語モデル「Luminous」は、テキストだけでなく画像のプロンプトも処理可能なマルチモーダル機能や、生成したコンテンツを検証し、そのソースまで追跡する説明可能性などを特長としている。さらに、クラウドだけではなくオンプレミス（自社で保有して運用しているシステム内）でも実装可能で、情報漏洩のリスクも避けられる。生成AIは、産業用ロボットの数百ページのマニュアルを用いてトレーニングされている。作業員が画像やテキスト、音声を使って質問すると自然言語で対応できる。ロボット操作の効率と安全性は格段に高まるものと期待される。

◆生成AIが登場するまでのAIの実力はどの程度のものだったのか

22年11月にOpenAIがChatGPTを発表してから、ChatGPTを始めとする生成AI関連のニュースが世間をにぎわしている。12年にディープラーニングで画像の特徴量を自動抽出できるようになって、画像認識精度が格段に向上し、人間の認識精度を超えて以来、画像認識AIは製造現場の製品検査や顔認識など幅広く社会実装されるようになった。しかしながら、個別システムごとに大量のデータセットを準備するには、多大な時間と人手がかかる（STEP1）。さらにAIモデルが実用レベルになるまで、データの特徴やパターンを発見し、別データの分析や予測を行えるよう学習させる必要がある

（STEP2）、新規システムの構築（STEP3）は非常に限られた領域のみであった（図.1）。

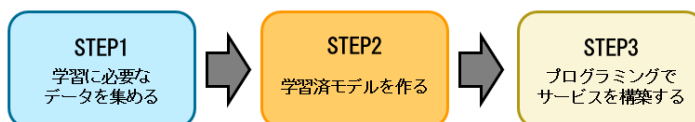


図.1 AIシステム構築手順

一方、自然言語処理AIにおいては、代表的なアルゴリズムとして時系列データ

処理を行う RNN（Recurrent Neural Network：回帰型ニューラルネットワーク）が採用された。これまで独立した単語として処理していた文章を単語の時系列データをして解析し、直前の言葉を記憶させることで、次の言葉の出現確率を上げた（図.2）。しかし、自然言語AIの精度は人間には遠く及ばず、社会実装されるのはチャットボットやSNS分析など、そこまで高い精度が求められない領域に限られていた。すなわち、基盤モデルが出現する前までのAIは、囲碁や将棋などある分野に限れば人間の能力を超えたが、学習していないことは全くできない、物体の認識機能に特化した特化型AIと呼ばれるものであった。

前の言葉から次の言葉を予測

僕 の 名前 は ?

僕 + 名前 = ? 男性の名前

図.2 RNNでの言葉予測の概念図

◆自然言語処理能力が格段に向上した大規模言語モデルとは

大規模言語モデルは、17年に登場したTransformerモデルをベースにしている（図.3）。基本的なモデル構造をシンプルにして、いろいろな組み合わせで構成できるようにした。Attention層と呼ばれるブロック（図.3のオレンジのブロック①～③）は、従来のRNNで課題となっていた、あまりにも長文になると精度が下がってしまう点や、時系列処理の学習が並列学習に向かず、大規模データを学習するには多大な時間が掛かるという問題を解決するために考案された。並列処理により文章全体を把握し、文の中で重要な単語には重み付けをすることで、単語間の関係性を理解する。Attention層では、複数カ所の重要な部分に注目することができるため、長文においても文脈を理解することができる。Transformerでは、どこに注目すべきかを分かった上で、出力の単語を推測するため、入力者が期待している単語を選択して返す確率が高まる。

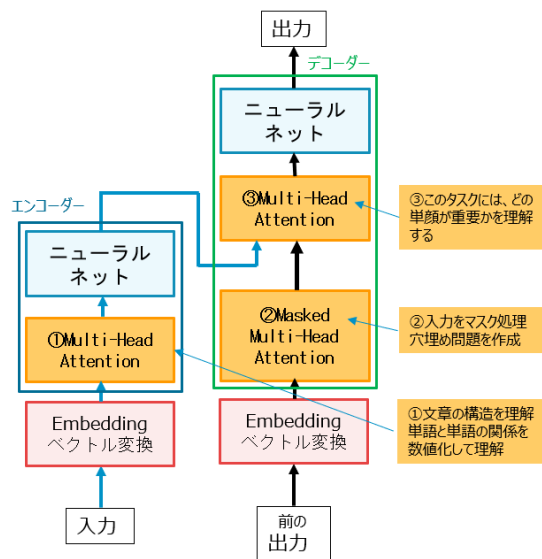


図.3 Transformerモデルの機能ブロック

このTransformerのモデルで構築した大規模言語モデルGPT (Generative Pre-trained Transformer) は、事前にラベル付けを行っていない広範囲で大量の文書データセットを「自己教師あり学習」でトレーニングしたものである。GPT-3.5では、Wikipediaなどインターネット上の3,000億の単語について、任意の文章を抽出し、自ら穴埋め問題を作成し、それを解いて機械的にデータにラベルをすることで言語を理解するように学習されたモデルである。AI自身が自動でラベル付け作業をしてくれるので、人間の工数が膨大に発生することはなくなった。また、データ数と、モデルのパラメータ数（モデルの規模）に比例してAIの性能が上がるスケール則という特徴がある。事前学習を十分にしたモデルを用意すると、少しのデータで学習する（転移学習）ことで、さまざまなアプリケーションを新規に構築することができる（図.4）。さらに、これまでのAI技術では、テキ

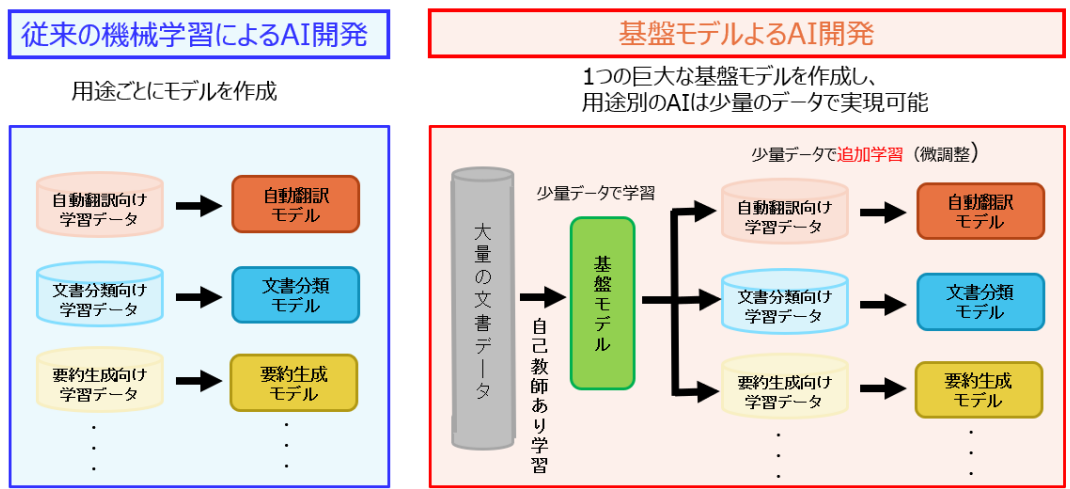


図.4 基盤モデルの利点 出典：日本IBMの資料を元にARC作成

ストや音声のみといった1種類の情報しか処理できなかったが、Transformer モデルは複数の種類の情報が混在した入力も処理可能なので、言語+画像を入力するようなマルチモーダルな処理が可能になる。

◆生成AIはチャットAI以外にもさまざまな分野でプロダクトが続々登場している

生成AIというと、ChatGPTやBing、Bardに代表される対話型（チャット）AIが有名であるが、さまざまな分野でプロダクトが登場している。DALL-E2、Stable Diffusion、Midjourneyに代表される画像生成AI、GEN-1、Flikiの動画生成AI、音楽生成AI、音声生成AI、文字起こしAI、文章生成AIなど数多くのアプリケーションがリリースされている（表.1）。

表.1 さまざまな領域で開発されている生成AI 各種資料よりARC作成

領域	プロダクト名	開発企業	特徴
対話型 (チャット) AI	ChatGPT	OpenAI	回答の精度は「3.5」(無料)より「4」(有料)の方が高い。テキストデータに特化したモデル。文章の作成やプログラミングが得意。画像や音声を取り扱うモデルは別途開発。
	Bing	Microsoft	会話のスタイルを「創造的」「バランス」「厳密」の3つから選べる。回答と同時に使ったウェブ上の出典を表示する(ファクトチェックが容易)。
	Bard	Google	1つの質問に対し3パターンの回答を出す。音声認識や画像認識にも対応し、グーグルドキュメントやGメール、Colabに連携し、回答を転用できる。
	Lumi	Aleph Alpha	独自開発したマルチモーダルな基盤モデル「Luminous」上に構成された会話モジュール。会話履歴を記憶し、雑談的な会話もできる。
画像生成AI	DALL-E2	OpenAI	テキストを元にオリジナルの画像を生成する。画像15枚/月まで無料。手持ちの画像から画像を生成、画像の一部をAIで編集など機能が豊富。
	Stable Diffusion	Stability AI	文字列から画像を生成する。入力するテキストや単語が多いほど、イメージに近い画像を出力する。コードが一般公開されており無料で使える。生成した画像の著作権は、画像を生成したユーザーに帰属する。
	Midjourney	Midjourney	テキストから画像を生成する(Text-to-Image) AIプログラム。Discordからメッセージを送信すると、テキストの内容から画像を生成してくれる。トライアル版では25枚まで無料。
	Bing Image Creator	Microsoft	Image Creatorは、Microsoft Edgeのサイドバーから直接DALL-Eを使用してテキスト入力に対応したAI画像を生成。
	Adobe Firefly	Adobe	「文章入力による画像生成」と「テキストエフェクト(文字装飾)の生成」、PhotoshopやPremiere Proなどと連携する。AIの学習データはAdobe Stockで著作権者が利用を許諾済、オープンライセンス、著作権期限切れのコンテンツのみ。
	Crypko	Preferred Networks	キャラクター生成プラットフォーム、高品質なキャラクターの創作をWebブラウザ上で簡単に行うことが可能。フリーミアムプランおよびスタンダードプランは商用利用ができない。
動画生成AI	GEN-1	Runway	画像やプロンプトの構図やスタイルを動画に適用し、新しい動画を生成する。お手本の動画を元に新しい動画を生成する、特定の動作を動画中に取り込むなど5つのモードがある。
	RunwayML	Runway	iOS向けアプリ。アプリ内で撮影した動画や端末に保存されている動画を、プロンプトを入力したり、画像やプリセットを選択したりしてまったく新しい動画に変換できる。
	Fliki	Nine Thirty Five	AI音声を使って、テキストを動画化できるサービス、Text-to-Speech機能だけでなく、動画も取り揃えている。75言語750人の音声に対応しており、日本語にも対応している。
	Vidon.AI	Vidon.AI	テキストコンテンツから自動で動画コンテンツを生成してくれるAIツール。ブログ記事からショートムービーを生成したり、カスタマーレビューをまとめた動画や、研修・教育用の動画を生成する。
音楽生成AI 音声生成AI	SOUNDRAW	SOUNDRAW	曲の長さ「15秒」やテンポ「Normal」、ムードやジャンル、テーマを選択するだけで、音楽が完成する。
	CoeFont	CoeFont	「AI音声プラットフォーム」を活用して人間の声のような自然な合成音声に変換したものを「AI音声」として提供。「AI音声」の種類は、アナウンサー、ナレーター、声優や著名人、2Dキャラクターなど5,000種類以上。
	Play.ht	Play.HT	入力したテキストを自然な発音で音声化するボイスジェネレーター。142の言語と907の豊富な音声パターン。生成された音声は、アクセント、速度、ピッチ、音量、発音や一時停止の箇所を手動で編集できるため、より自然に人間らしく聞こえるようカスタマイズして出力できる。動画やEラーニングのナレーションとして。
文字起こしAI	Otter	OTTER.AI	英語の音声をテキストにするAIサービス。マイクや「Zoom」などのビデオ会議から英語音声を入力するとリアルタイムでテキスト化する。発言者を区別して表示するので、インタビューや複数人の会議でも話者が分かりやすい。
	CLOVA Note β	LINE	長時間の録音データ解析に対応する音声認識モデル「CLOVA Speech」を用いた、AI音声認識アプリ。録音から文字起こしを行い、音声記録として残しつつ、議事録やノート作成をサポートする。参加者の音声を区別可。
	Rimo Voice	Rimo	会議後すぐに、日本語に特化したAIによる、自然な日本語の文字起こしが完成。AIが発言者を識別して自動で発言を分離し、読みやすく段落分けされた文章にまとめる。Zoom、Microsoft Teams、Google Meetのアプリに連携。
文章生成AI	BASE AI アシスタント	BASE	ネットショップ作成のBASEが、オーナーの業務効率化を目的として導入したChatGPTベースの新機能。商品登録画面で商品名を入力後、「商品の特徴」欄に素材やサイズ、カラーやセールスポイントなどを入力すると、その内容に沿った400文字前後の商品紹介文を作成。
	note AIアシスタント	note	コンテンツ投稿プラットフォーム「note」に実装されている創作支援機能で、オープンAI社の「GPT-3」を採用。文章のアイデアや構成、タイトルや書き出しの提案から、要約や翻訳まで、様々な文章作成機能を備える。
	Tome	Magical Tome	テキスト入力のみでプレゼンテーションで使えるスライドを1分程度で自動生成。任意の画像、動画、テキストボックス、表を追加。画像はDALL-Eを使用。
	AIのべりすと	Bit192	書き出し部分を入力すると、続きをAIが書いてくれるという小説作成サービス。総2テラバイト以上のコーパスからフルスクラッチで訓練した小説AI。
	AI editor	ミライセンス	文章の校正や校閲を自動化できるサービス。1万文字を3秒でチェックできる。公用語や医療法などプリセットされたルールも豊富にあり。独自のルールも作成でき、複数のルールをまとめたきめ細かい設定が可能。

今まで自然言語処理や画像処理にAIを適用したサービスを提供していた企業は生成AIの性能に注目し、自社サービスに取り入れる動きがみられる。

◆生成AIのリスクに対し各国が規制の動き

23年5月16日、ChatGPTを開発したOpenAIのサム・アルトマンCEOは、米上院司法委員会で「私たち全員がAI技術の大きな可能性を享受できるように、政府と産業界が一緒になってリスクを管理する必要がある」と証言し、政府による規制が必要だと訴えた。また、5月31日、欧州連合（EU）と米国はスウェーデンで閣僚級会合を開き、ChatGPTなどの生成AIについて、AIでビジネスを行っている企業に対して自主的な行動規範の作成を求めることで合意した。EUは現在、AI利用の包括的な規制法を準備中だが、想像を絶する速度で技術開発が進んでいることに注目し、法整備を待たずに対応強化に乗り出している。

日本政府は5月26日「AI戦略会議」で論点整理をし、AIの利活用に関する国際的なルール形成の議論をリードすべく、機密情報漏洩などリスクの列挙に力点を置き、7つのリスクを示した（表2）。同時に利用や開発の促進に注力することも訴えた。

懸念されるリスクの具体例と対応
1. 機密情報の漏洩や個人情報の不適正な利用のリスク
2. 犯罪の巧妙化・容易化につながるリスク
3. 偽情報などが社会を不安定化・混乱させるリスク
4. サイバー攻撃が巧妙化するリスク
5. 教育現場における生成AIの扱い
6. 著作権侵害のリスク
7. AIによって失業者が増えるリスク

表.2 生成AIの懸念されるリスク
資料よりARC作成

各国は生成AIについてのリスクを考慮し、規制の動きを見せている。

◆セキュリティや著作権侵害問題が指摘されているが生成AIの活用領域は広い

生成AIについては、多くのリスクが指摘されているなか、さまざまな領域での活用が見込まれている。長い文章を指定の文字数以内で要約する「文章の要約」や、プログラムコードの自動生成やコードチェック機能、アイデアを考える場合の「壁打ち」や論点の洗い出しなど、すでに活用している人は多い。

実際、BardでPythonのプログラムを作成してみたが、一瞬でプログラムを作成してくれた。プログラムの他に、ソースコードの解説が付いている。また、画面には、「コードは慎重に使用してください。」と詳細な注意事項を含めたコメントが付いている（図.5）。



図.5 Bardでのプログラム生成例

さらに、機械学習の教育及び研究用に提供しているGoogle Colaboratoryとの

連携が取れていて、コードの修正や再実行が簡単に行える環境になっている。

◆ 契約や法律チェックは生成AIで精度向上

23年5月30日、LegalOn Technologiesは、AI契約審査プラットフォーム「LegalForce」に「ChatGPT」APIを活用し、契約書の修正をサポートする「条文修正アシスト」の機能

オープンβ版の提供を開始した。「ChatGPT」を活用し、契約書の内容を一定程度反映した文案を表示することで、企業の法務担当者など法律専門家の修正文案検討業務効率



図.6 条文修正アシスト機能

出典：LegalOn Technologies

化を支援する（図.6）。これまで、契約類型別のチェックリストと契約書の照合を自然言語処理などの技術により自動的に行い、条文の抜け漏れや条項内の過不足を瞬時に提示し、抜け漏れや見落としを予防する機能をコアとしていた。自社の法務分野における知見を組み合わせた生成AIを連携させることで、これまでの技術では実現が難しかった、一定程度の実用に耐えうる文案を生成できる。

23年5月30日、Archaicは広告・ヘルスケア事業会社向け広告チェックAIにChatGPT連携サービスを開始した。Archaicは、ヘルスケアなどの専門的なクライアントデータやインターネットデータを長年にわたって解析することでAI技術を磨いてきた。これにChatGPTを連携することで、広告チェックの精度が向上し、判断が難しい表現も検知することが可能になる。21年8月に改正薬機法の施行により規制が強化され、規制対象となる広告主、代理店などは、広告表現のチェックにこれまで以上に時間をかける必要がある。薬機法に精通した人材が少なく、自社採用する余裕もない場合は、営業担当やマネージャーがチェックするためチェック漏れの発生が課題となっている。広告チェックAIは、わずか2分で広告表現をチェックし、不適切な表現や問題を検出する。また新たに追加した短文チェック機能では、広告キャプションやSNS投稿などの短文にも対応する。

実際に生成AIを使ってみて、AIの進化を実感してほしい。 【成田誠】