

生成 AI の仕組みと教育・研究への活用、 そしてフィジカル AI の未来

1. 生成 AI(大規模言語モデル)の基礎と仕組み

生成 AI、特に ChatGPT などに代表される**大規模言語モデル(LLM)**の根幹にあるのは、「次に来る単語を確率的に予測する」というシンプルな原理です。

言語モデルの本質

例えば「このリンゴは〇〇」という文章に対し、前文の文脈(農家から貰った、等)を踏まえて「赤い」「黄色い」「美味しい」といった単語の出現確率を計算します。膨大なテキストデータから学習した知識に基づき、最も確率の高い単語を選択して繋げていくことで、自然な文章を生成します。

GPT の正体

GPT とは **Generative Pre-trained Transformer** の略称です。

- **Generative(生成)**: 文章を新しく作り出す。
- **Pre-trained(事前学習)**: インターネット上の膨大なデータを使い、穴埋め問題を解くような形で文法や世界知識をあらかじめ学習している。
- **Transformer(トランスフォーマー)**: 2017 年に発表されたニューラルネットワークの構造。単語間の関係性(アテンション)を効率よく捉えることができ、現在の AI の飛躍的発展の鍵となりました。

人間との対話を実現する調整

初期の GPT はインターネット上の不適切な情報も学習してしまいます。これを私たちが安心して使える「ChatGPT」というサービスにするために、**RLHF(人間からのフィードバックに基づく強化学習)**が行われています。人間が「良い回答」を評価し、AI がその価値観に沿うように微調整(ファインチューニング)されることで、有害な情報の抑制や対話能力の向上が図られています。

2. 生成 AI の課題と拡張技術「RAG」

生成 AI には、もっともらしい嘘をつく**ハルシネーション(幻覚)や、学習データの締め切り以降の最新情報を知らないという弱点があります。これを解決する手法として注目されているのが RAG(検索拡張生成)**です。

RAG の仕組み

ユーザーの質問に対し、いきなり AI が答えるのではなく、まず外部の信頼できるデータベースや最新のウェブ情報を検索します。その検索結果を「参考資料」として AI に与え、それに基づいて回答させることで、正確性と信頼性を劇的に高めることができます。

マルチモーダル化

最近では、テキストだけでなく画像も理解できる**VLM(Vision Language Model)**が登場しています。画像の中の矛盾点(例:車の上でアイロンをかけている男性など)を言語で説明できるようになっており、視覚と知能の統合が進んでいます。

3. 大学における活用事例:「藤吉 AI 先生」

藤吉教授は、これら AI 技術を大学の教育・研究現場に導入する試みを行っています。

研究の壁打ち相手としての活用

Google の「NotebookLM」などのツールを使い、教授自身がこれまでに執筆した論文や講演資料、ブログ記事などを RAG の外部データとして読み込ませた「研究版・藤吉 AI 先生」を構築しました。

- **メリット:** 学生は教授が不在の時でも、教授の思考スタイルや専門知識に基づいたアドバイス(アイデアの壁打ち)を受けることができ、研究の停滞を防げます。
- **汎用性:** これは企業の熟練技術者のノウハウ継承などにも応用可能な仕組みです。

教育現場での利活用: 講義動画との連携

多くの大学が抱える「アーカイブ動画が見られない」「学生が人前で質問するのを恥ずかしがる」という課題を、AI チャットボットが解決します。

1. 講義動画を解析し、スライドの切り替わりと発話をデータベース化。
2. 学生の質問に対し、AI が講義内容に沿って回答。
3. **動画連携**: 回答の根拠となった講義動画の「該当箇所」をワンクリックで再生可能に。

アンケートでは 9 割以上の学生がポジティブに評価しており、「24 時間いつでも、どんな初歩的な質問でも気兼ねなく聞ける」環境が、学びの質を向上させています。

4. 次なる波: フィジカル AI (VLA モデル)

現在、AI の最前線は画面の中(言語)から、現実世界での行動(物理)へと移っています。これがフィジカル AI です。

従来のロボットとの違い

これまでのロボット制御は「物体認識」「経路計画」「動作制御」といった工程が分業化(モジュール化)されていました。しかし、これでは最初の認識でミスをすると後の全工程が失敗します。最新の VLA (Vision-Language-Action) モデルは、これらを「一気通貫 (End-to-End)」で処理します。

VLA モデルの構造

- Vision (視覚): カメラ画像から状況を理解。
- Language (言語): 「ゴミを捨てて」といった人間の指示を理解。
- Action (行動): モーターをどう動かすかという数値(トークン)を直接出力。

例えば、「潰れたコーラの缶を捨てて」と指示すると、AI は「潰れているからゴミである」という世界知識(セマンティックな理解)に基づき、複雑なプログラミングなしにその物体を掴んでゴミ箱へ運ぶ動作を生成します。

ヒューマノイドへの応用

Google の「RT-2」や「Gemini Robotics」のようなモデルでは、学習したことのない未知の物体（例：ラグビーボール）に対しても、AI が持つ一般知識から「これはボールの一種だ」と判断して適切に扱うことが可能になりつつあります。将来的には、人間のように見て、聞いて、柔軟に行動するヒューマノイドロボットが家庭や現場に普及する土壌が整いつつあります。

5. 結び: AI と共生する未来への構え

生成 AI からフィジカル AI へと急速に進化する中で、藤吉教授は以下の 2 点が重要であると説きます。

1. **AI の限界を知る:** ハルシネーションのリスクを理解し、AI の回答を盲信しないこと。
2. **裏付けを確認する力:** AI が示した答えに対し、自ら情報の根拠（ソース）を確認し、判断する「人間側の能力」がこれまで以上に問われています。

技術の進歩を正しく理解し、活用することで、私たちの生活や研究、教育はより豊かなものになるでしょう。