

1. 医療における生成AIの現状と今後の展望

藤田 広志 岐阜大学工学部

ディープラーニング(深層学習)により人工知能(AI)の話も少し落ち着いたと思っていたら、OpenAI社の「ChatGPT」が2022年11月に発表されるや、翌2023年は生成AIに関する新しいサービスが次々とリリースされた。この2023年は「生成AI元年」とも称せられ、いまや「第4次AIブーム」に突入したとも言われる(図1)。その生成AI技術の進化のスピードには目を見張るものがあり、日替わりで新しい情報が飛び交うため、最新情報についていくのも簡単ではない昨今である。本稿では、医療における生成AIの最近の情報の一端を、筆者の興味に準じて紹介する。

認識AIから生成AIへ

図1は、歴史的なAIの進化と期待値について示したもので、過去3度のAIブームがあり、現在は生成AIにより4度目のブームとも言われている。図中には、コンピュータ支援診断(CAD)の開発に関するエポックも記載してある。ディープラーニングによるAIは識別AI、または認識AIであったのに対して、生成AIでは新しいコンテンツが作成される(図2)。

OpenAI社が開発した話題の会話型AIサービスのChatGPTは、大規模言語モデル(large language model: LLM)の一種であるGPTが使われている。GPTは「Generative Pre-trained Transformer」の略であることが示すように、生成型の大量のデータで事前学習されたTransformer(トランスフォー

マー)と呼ばれる入力シーケンスを出力シーケンスに変換する、ニューラルネットワークアーキテクチャを用いた言語モデルである。Transformerは、言語処理に限らず、音声認識やタンパク質配列分析まで、あらゆる種類の配列変換に使用されている。なお、このトレーニングにおける特徴は、一般的に正解ラベルのないデータを使用した「自己教師あり学習」で行われることである。

GPTのような大規模モデルは、2021年8月にスタンフォード大学のHAI(人間中心のAI研究所)により、基盤モデル(foundation model)と呼ばれるようになり(図3, 4)、テキストや画像、音声、表形式データなど、多様なデータ(=モダリティ)から事前学習する場合には、マルチモーダル基盤モデルと表現される。

基盤モデルでは²⁾、幅広い下流タスクに適応できるように再トレーニング(主にファインチューニング)されるため、2段階にわたる訓練工程がある。よって、1つの基盤モデルで多様なタスクに適応できること(マルチタスク能力)が大きな特長となっており、まさしくあらゆるタスクの基盤となるモデルである。テキストだけではなく画像データも事前学習した「GPT-4V(GPT-4 with Vision)」は2023年9月にリリースされているので、現在のChatGPTは画像データ(そして音声データ)にも対応可能となり、マルチモーダル基盤モデルとなっている。このように、今の時流は、LLMから発して、大規模マルチモーダルモデル(large multimodal model: LMM)への方向に向かっていく。

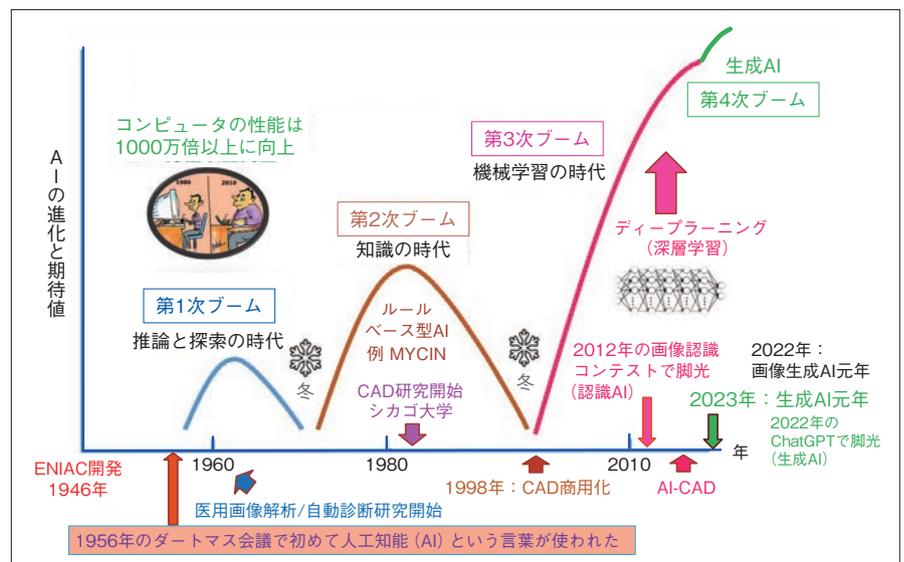


図1 歴史的に見るAI技術の進化の変遷