

# 基礎研 レポート

## 生成 AI は人間を代替するか ～生成 AI と人間の知能の違いとは～

金融研究部 准主任研究員・ESG 推進室兼任 原田 哲志  
(03)3512-1860 harada@nli-research.co.jp

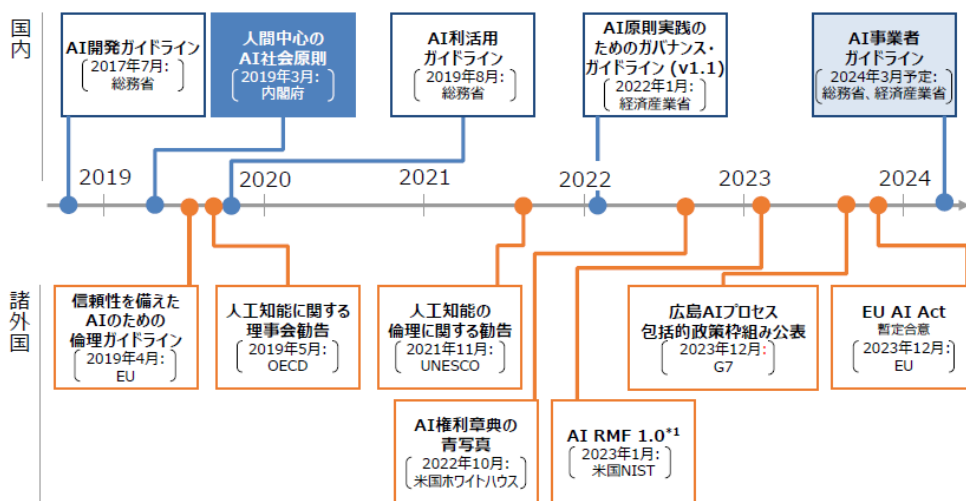
### 1—生成 AI の活用

近年では生成 AI（人工知能）の性能向上などにより社会の様々な分野での生成 AI の活用に注目が集まっている。生成 AI の社会的な活用が推進されており、2023 年 11 月には政府は生成 AI の開発者向けに政府保有データの提供を始めることを決定した<sup>1</sup>。生成 AI の開発において米国など海外が先行する中、国産の生成 AI 開発を促進する狙いだ。

一方で、生成 AI の普及には様々なリスクも指摘されており、適切な利用が肝要と言われている。政府は AI の開発や利活用に関するガイドラインの策定を進めている<sup>2</sup>。「広島 AI プロセス包括的政策枠組み」や「人間中心の AI 社会原則」といった国内外の諸原則や規制を踏まえたガイドラインの策定が行われている（図表 1）。

生成 AI の活用を巡っては、活発な議論が行われており、様々な意見がある。生成 AI の活用や今後を考える上では、その仕組みや性質について理解することが必要だろう。本稿では、生成 AI の仕組みや性質について説明したい。

図表 1 AI に関する主な諸原則



(資料)総務省 経済産業省、「『AI 事業者ガイドライン』案概要」

<sup>1</sup> 日本経済新聞、「生成 AI 学習向け、政府データを提供へ 国内事業者に」、2023 年 11 月 7 日

<sup>2</sup> 経済産業省、「AI 事業者ガイドライン案」

## 2—生成 AI の仕組みと特徴

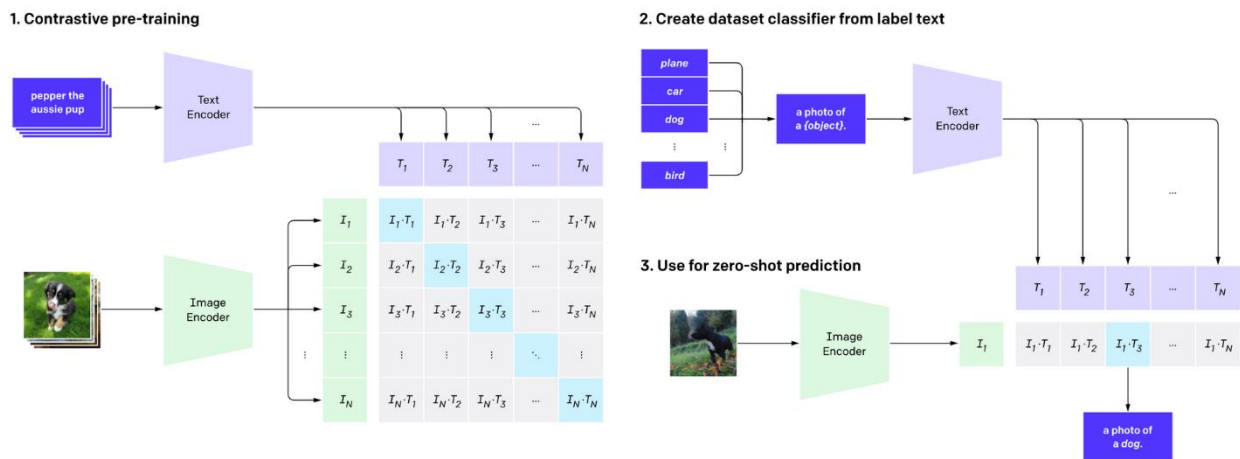
代表的な生成 AI サービスである ChatGPT を提供する OpenAI は入力された文章から画像を生成する処理の基盤となる CLIP (Contrastive Language-Image Pre-training) モデルの論文を公表している<sup>3</sup>。CLIP は、テキストと画像の対応関係をとらえるモデルであり、①与えられた学習データの文章や画像の特徴(特徴量)を抽出、②それらの特徴量によって文章と画像を関連付ける、③画像が入力された際には、その特徴量を抽出、あらかじめ学習したテキストと画像の対応関係に基づいて関連性が高い文章を提示することができる(図表 2)。つまり、簡単に言うと入力された画像の内容について文章で回答することができる。

画像や文章から特徴量を抽出することを「エンコード」と呼ぶ。逆に特徴量をもとに具体的な画像や文章を生成することを「デコード」と呼ぶ。

文章をもとに画像を生成するように指示された場合は、与えられた文章の特徴量をもと CLIP モデルにより、関連性が高い画像の特徴量を取得し、その特徴量をもとに画像を生成するというプロセスを行っている(図表 3)。

このように生成 AI は大きく①画像や文章から特徴量を抽出、②これらの特徴量を関連付け、③関連性から得られた特徴量をもとに画像や文章を生成する処理から成り立っている。これにより、入力された画像や文章と関連が強い画像や文章を出力している。また、文章による質問に文章で回答する場合も同様に関連性が高い情報を探し提示するプロセスを行っている。

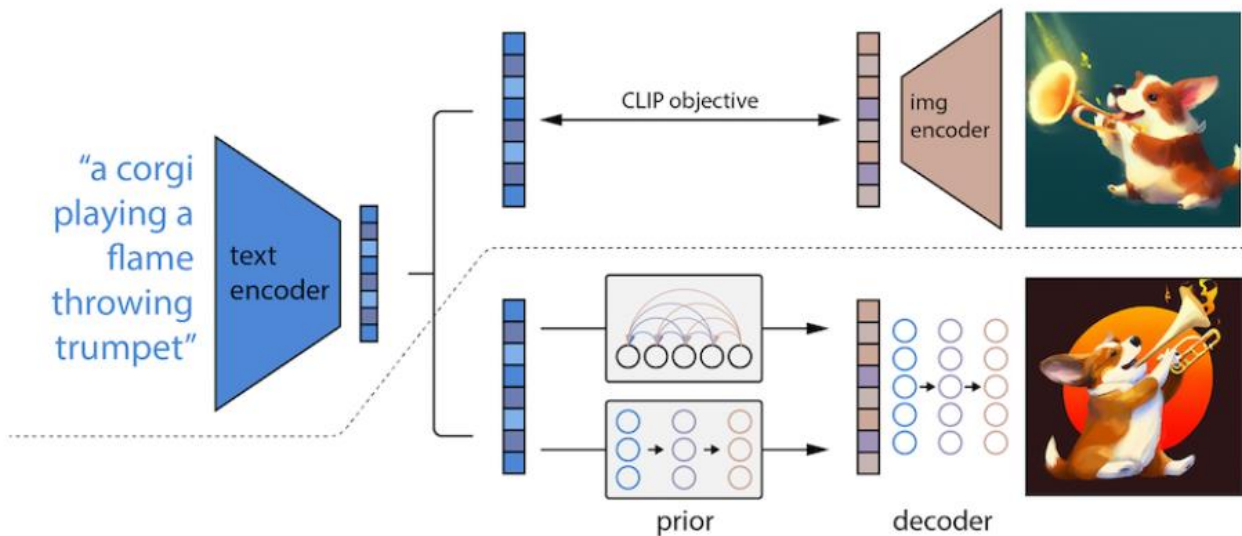
図表 2 CLIP モデルによる画像と文章の関係の学習



Alec Radford et al., "Learning Transferable Visual Models From Natural Language Supervision", 2021/2/26

<sup>3</sup> OpenAI, "CLIP: Connecting text and images" <https://openai.com/research/clip>

図表3 CLIPモデルによる画像生成の流れ



(資料)Ramesh et al., ” Hierarchical Text-Conditional Image Generation with CLIP Latents” 2022/4/13

生成AIはこのようにして与えられた文章から画像を生成しているが、どのような処理が得意・不得意だろうか。

生成AIは「手」の画像の生成が不得意だと言われている(図表4)。人間は片手の指が基本的には5本だということをあらかじめ知っており、そうした前提知識をもとに絵を描くことができる。一方で、生成AIはそうした前提知識はなく、与えられた手の画像の学習データのみからそれを判断し、画像を生成する。しかし、多くの手の画像は一部の指が隠れていることなどから、本来の手の形状・性質の学習が困難となっている。生成AIの学習データは一般的に2次元の平面的な画像であり、3次元の「立体的な情報」は得られないことも影響している。

また、生成AIの特徴として「ハルシネーション(幻覚)」が指摘される。ハルシネーションとは生成AIが事実と異なる情報や、実際には存在しない情報を生成する現象を指す。

ハルシネーションは生成AIが関連の強い情報を提示する仕組みに基づいて回答文を生成しており、事実かそうでないかという観点は基本的に持ち合わせていないことが原因となっている。

このようにAIが単に関連性によって回答しているのに対して、人間は例えば「事実かそうでないか」、「対象に関する知識」、「立体的な情報」といった多面的な観点や知識から判断を行っていることが分かる。

図表4 AIが生成した手の画像



(資料)ChatGPT をもとに筆者作成

### 3—AIが抱え続けるフレーム問題とは

上記の点と関連するAIの課題点として「フレーム問題」が指摘されている。フレーム問題とは、AIの処理能力は有限であるため、現実で起こりうる無数の出来事に対処することはできないことを意味する。チェス、将棋といったルールが規定されたボードゲーム、製品の組立作業といった対象となる範囲が限定されている課題では、フレーム問題は生じずAIは対処ができる。

対象や問題の範囲が限定されている状況では、それに沿って特定のモデルにより回答することができるが、人間が普段何気なく行っているように問題の範囲が限定されない中で様々な観点をもとに回答するようなことを、特定のモデルにより実現することは困難である。

AIの分類として、「強いAIと弱いAI」という分類がされる場合がある。弱いAI(特化型AI)とは、対象となる課題の範囲内で回答を行うAIを指す。強いAI(汎用型AI)とは、対象となる課題が限定されず、人間のように様々な課題に対して対応・回答を行うAIを指す。弱いAIが特定の課題にしか対処できず、道具のような存在であるのに対して、強いAIは人間に近いものと言える。

しかし、強いAIである汎用型AIは現時点では原型となる有力なモデルすら存在しない、想像上の存在である。様々な要素を考慮に入れるようにAIのモデルや性能が改善される可能性はあるが、フレーム問題が解決しない限り、強いAIの実現は難しい。

米国の大手IT企業IBMは「学術機関や民間のAI研究者たちが、汎用人工知能(AGI)の開発に投資を行っている一方で、それは現時点では、理論的な概念としてしか存在しておらず、具体的な現実にはいたっていません」と述べている<sup>4</sup>。また、米国のコンピュータ科学者Marvin Minsky氏は、今後数十年のうちに汎用型AIが実現するという見方は楽観的すぎると指摘している。

<sup>4</sup> IBM, 「強いAIとは」 <https://www.ibm.com/jp-ja/topics/strong-ai>

今後、AI の性能が向上し続けたとしてもフレーム問題は現状では解決の目途は立っておらず、AI が苦手な分野は長期間残り続けるかもしれない。

生成 AI は現状では期待が先行しているが、こうした特徴や得意・不得意はよく知られていない面もある。しかし、けっして生成 AI が万能でないことは理解すべきである。生成 AI の社会実装を進めていく上では、その得意・不得意を踏まえた適切な利用が不可欠であり、政府や各企業等の具体的で分かりやすいガイドラインの早期策定が求められる。生成 AI 活用の今後の動向に注目したい。

---

(お願い) 本誌記載のデータは各種の情報源から入手・加工したものであり、その正確性と安全性を保証するものではありません。また、本誌は情報提供が目的であり、記載の意見や予測は、いかなる契約の締結や解約を勧誘するものではありません。