

世界を席卷するジェネレーティブAI

青山 祐輔 ●ITジャーナリスト

「Midjourney」や「ChatGPT」など誰でも簡単にコンテンツを生成できるサービスが流行し、AI技術が身近に。作風や創造性の学習モデルをめぐる議論も起こり、人間による知的作業の本質を突き付ける。

ジェネレーティブ (Generative) AIは、2022年に話題としてもサービスとしても最も急拡大した技術と言えるだろう。その中心にあったのが「Midjourney」と「Stable Diffusion」、そして「ChatGPT」だ。

ジェネレーティブAIとは、人工知能 (AI) によって文章や画像、音声といったコンテンツを生成する技術である。主にニューラルネットワークを用いて膨大なデータを学習させることでモデルを作り、そのモデルに基づいてさまざまなコンテンツを生成することができる。

コンピューターによるコンテンツ生成は、AIの利用方法の一つとして古くから研究されてきた。例えば、音楽においては音楽理論に基づいたアルゴリズム化が容易なことから、古くから楽曲の自動生成ツールやサービスが存在した。ディープラーニング技術による画像認識技術の応用としては、画像を劣化させずに拡大したり、画像の中の余計なオブジェクトを消去したりといったレタッチ機能が、米アドビのPhotoshopという画像編集ソフトウェアなどに搭載されている。また、画像や動画をAIによって加工し人の顔を入れ替えることで偽の映像を作り出すディープフェイクは、フェイクニュースに利用されたり、外交や政治に悪影響を与えたりするなど、大きな問題となっ

ている。ところが2022年に、それまでのコンテンツの生成・編集とは一線を画すレベルのジェネレーティブAIが登場した。

■ Midjourney

2022年7月に公開された「Midjourney」は、同名の研究所が開発した画像生成サービスだ。Midjourneyは有償のサービス (月額10ドルで200枚まで、30ドルで枚数無制限) だが、25枚までは無償で試せるとあり、多くの人が実際に触れその性能の高さを体験したことで、瞬く間に大きな流行となった。

Midjourneyが大きな反響を呼んだのは、主に3つの理由がある。「高品質な画像」を「誰でも簡単」に生成でき、それを「インターネット上で共有」できたからだ。Midjourneyで生成した画像は、プロのイラストレーターが描いたようなグラフィックやデジタルカメラで撮影したような画像であったり、最新の3DCGエンジンによるレンダリング画像に匹敵するクオリティを持ったりしている。もちろん細かく見れば欠点はあるが、多くの人にはAIが生成したものか人間が描いたものかの見分けが付かないレベルにあった。

さらに、そうした画像を生成するのに必要な作業は、プロンプトと呼ばれるキーワードの入力の

みだ。画像の主題、描く対象物、構図、絵柄、背景、空模様、画角、レンズ効果など、およそ画像に関わるさまざまな要素を考え組み合わせることで、ありとあらゆる画像を生成することができる。このプロンプトの組み合わせにはコツがあり、思い通りの画像を生成するには工夫やノウハウも必要となることから日本では「呪文」、英語圏では「magic words」と呼ばれている。こうした呪文と画像を共有するウェブサイトも登場しており、AIにうまく絵を描かせるためのノウハウが、急速にインターネット上に蓄積しつつある。

■ Stable Diffusion

続く2022年8月には、AIスタートアップの英Stability AIが、AIによる画像生成ソフトウェア「Stable Diffusion」をオープンソースとして公開した。Stable Diffusionは、インターネット上でソフトウェアと学習済みモデルが無償で公開されているため、自分のパソコンやクラウド上の仮想マシンにインストールすることで自由に使うことができた。さらに、画像生成に用いる学習済みモデルを切り替えることができるため、自分が描きたい画像のタイプに合ったモデルを用いることができるのが大きな利点となっている。なお、Stable Diffusionの標準モデルは、写真や3DCG的なリアル志向のイラストは得意だが、日本のアニメ的な絵柄を苦手としている。このため、インターネットのAI開発者コミュニティにおいて、標準モデルが苦手としている絵柄に特化した学習済みモデルの作成と提供が行われている。

Stable Diffusionは商業利用も可能であることから、これを用いた画像生成サービスやアプリも登場している。アニメ風イラストに特化したものや、写真をアップロードするとイラスト風に加工してくれるものなど、さまざまなタイプのサービスが登場している。ユニークなものとして、無償

で利用できるイラストを提供している「いらすとや」とコラボし、いらすとや風の画像を生成できる「AIいらすとや」がある。

● 画像生成AIの学習における問題

Stable Diffusionを利用したサービスの一つ、ラディウス・ファイブが2022年8月に公開した「mimic」は、自分で描いたイラスト画像をアップロードすることでオリジナルの学習モデルを用意し、自分の画風に基づいた画像を生成できる。同社は当初、クリエイターが自ら利用し、自身の創作活動を効率化するためのツールとしての利用を想定していた。だが、この「自分の画風をAIに学習させる」というのが物議を醸すこととなった。

そもそもAIに画像を学習させること自体は、日本の著作権法において正当な利用として認められている。これは、検索エンジンがウェブサイトのインデックスを作成するためにウェブサイトのデータを利用することが許されているのと同様だ。それに対して前述のmimicは、クリエイターの画像を用いて画風をまねた画像を生成できるため、例えば著名なイラストレーターの画像を勝手に用いて、そのイラストレーターの画風をまねた画像を勝手に生成することも可能だ。実際、画像生成AIのモデルの学習には過去の著名な画家の作品も含まれているため、プロンプトにその画家の名前を入れることで、その画家の画風に似た画像が生成できる。例えば、ミュシャ風やモネ風といった、よく知られた画風でいろいろなイラストを作り出すことができる。

一方で、著作権法で保護されるのはクリエイターが自ら描いた個別の絵であり、画風やスタイルと呼ばれるものは保護されない。しかし、AIによって勝手に自分の画風を模倣した画像が生成されることに対して心理的な抵抗を感じるクリエイターは少なくない。特にmimicの場合は、生成さ

れた画像の著作権はラディウス・ファイブに帰属するという利用規約もあったため、mimicに対し厳しい批判がなされた。その結果、同社はサービスを開始したものの直後に公開を一時停止した。その後、同社はサービスへの利用登録時に審査を行うとともに、利用者が著作権を保持しない画像の許諾なしの学習やmimicで生成された画像の利用範囲を逸脱する行為などを禁止すること、生成した画像の権利が利用者に帰属することなど、利用規約やルールを整理し、約2か月後にサービスを再開した。

こうした画像生成AIの学習における権利処理については、海外でも議論が起きている。2023年1月には、サラ・アンダーセン氏、ケリー・マッカーナン氏、カーラ・オルティス氏という3人のアーティストが、インターネット上で公開されている画像を学習したAIによって画像を作成することが無許可の二次創作に当たるとして、MidjourneyとStability AI、Stable Diffusionを利用した画像生成ツールを提供しているDeviantArtなどを訴えている。前述の通り、他人の著作物をAIに学習させること自体は日本だけでなく多くの国で認められており、そのAIが生成した画像は利用者の著作物として扱われる。しかし、そうやって学習させたAIが学習に用いた作品によく似た画像を生成した場合は、権利侵害と認められる可能性はある。例えば、利用者が元の作品を知っており、それと似た画像を生成することを意図してプロンプトを入力したり、元の画像自体を直接AIによって加工したりする場合は、利用者に元の画像を模倣する意図があるためだ。

現状では、前述の訴訟の結果はまだ出しておらず、AIによる画像生成に関する判例も少ない。従って、AIによって生成した画像の取り扱いについては、今後しばらくは注意深く見守る必要がある。

■ ChatGPT

ジェネレーティブAIで話題になったのは画像だけではない。2022年11月に米OpenAIがチャット形式での対話型AI「ChatGPT」を公開した。OpenAIはイーロン・マスク氏らが出資して設立されたAI研究組織で、現在のジェネレーティブAIブームのきっかけをつくったと言える。これまでに画像生成の「DALL・E」とその後継の「DALL・E2」、言語生成モデル「GTP-3」、音楽を生成する「Jukebox」などのAIを開発している。

ChatGPTはチャット形式のインターフェースで、質問をしたり指示を与えたりすると、それに答えたり指示に従った文章やデータを返したりする。これだけ聞くと、SiriやAlexaなどエージェント型のAIを想像するかもしれない。しかし、SiriやAlexaに質問すると単純にWikipediaなどのすでに存在する文章を読み上げるだけなのに対して、ChatGPTは学習した文章を独自に要約・解釈した形で回答する。例えば「ジェネレーティブAIとは？」とChatGPTに尋ねたところ、次のような答えが返ってきた。

「ジェネレーティブAIは、人工知能によって生成されたデータを意味します。一般的に生成モデルは、人間が生成したデータに似たものを生成するためにトレーニングされます。生成モデルには、言語生成、画像生成、音声生成などのさまざまなタイプがあります。これらの生成モデルは、自然言語処理、画像認識、音声認識などのタスクにも使用されます」

もちろん、見ての通り内容に正確ではない部分や、日本語としてこなれていないところもある。しかし、大意は理解できる文章で、また、ここからさらに深掘りして調べるのに必要十分な用語も提示されている。

なお、ChatGPTは日本語や英語以外に複数の言語に対応しているが、もともと英語ベースで開発されているため英語が最も得意となっている。同じ質問でも、日本語でするよりも英語で行った方が、よりの確な答えが返ってくることが多い。

他にも、日本語の文章を英訳したり、逆に和訳したり、長い文章を要約させたり、プログラミングのコードを生成させたりと、高度な知的会話・知的作業を実行でき、ある意味で人間以上に「使える」AIと言える。このため、米マイクロソフトがOpenAIに数十億ドルの追加出資を発表したほか、米グーグルが検索サービスに対する破壊的イノベーションと見なしAI戦略を見直すと報じられた。

■ AIと共存する時代に向けて

2010年代にニューラルネットワークによるディープラーニングで、コンピューターによる画像認識が人間以上の能力を持つようになった。これにより、第3次AIブームが起き、さまざまな領

域でAI活用のブレークスルーが起きている。そして、このジェネレーティブAIの登場は、AIの能力をより人間に近づける新たな段階に入ったといわれている。

一方で、こうしたAIの普及によって、職を失う人間が増えたり、人間自体の能力が劣化したりといった危惧を抱く人も少なくない。実際に、画像生成AIの能力を見て絵を描くことのモチベーションを失ったと発言するクリエイターもいる。だが、画像生成にしろ文章生成にしろ、一定のクオリティの成果を求めるためには、利用者側にも創意工夫や試行錯誤が求められることには変わらない。求める結果の品質に対して人間がインプットするためのコストが下がり、作業や時間の効率化が進むことは間違いないが、あくまで人間が使う道具であることに変わりはないと筆者は考える。従って、AIの脅威におびえるよりも、積極的に使い、それによって自らのアウトプットを向上させていくべきではないだろうか。



1996, 1997, 1998, 1999, 2000...

[インターネット白書ARCHIVES] ご利用上の注意

このファイルは、株式会社インプレスR&Dおよび株式会社インプレスが1996年～2023年までに発行したインターネットの年鑑『インターネット白書』の誌面をPDF化し、「インターネット白書 ARCHIVES」として以下のウェブサイトで公開しているものです。

<https://IWParchives.jp/>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、データ、URL、名称など)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真・図の作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は掲載されていない場合があります。
- このファイルの内容を改変したり、商用目的として再利用したりすることはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用される際は、出典として媒体名および年号、該当ページ番号、発行元などの情報をご明記ください。
- オリジナルの発行時点では、株式会社インプレスR&Dおよび株式会社インプレスと著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

お問い合わせ先

インプレス・サステナブルラボ

✉ iwp-info@impress.co.jp