

生成 AI の発展と新たなコンテンツ経済圏

栄藤 稔¹ (大阪大学)

要 旨

生成 AI の技術進化が 2022 年から急激に加速し、ユーザーを取り込んだ新たなコンテンツ経済圏が形成されようとしている。2014 年に発明された GAN (生成的敵対ネットワーク) という技術により、人が区別できないほど精巧な画像が自動生成されるようになった。さらに、2016 年、データを高度に抽象化する深層学習技術、トランスフォーマーが登場した。これは「データを与えさえすれば機械が自動学習する」という教師なし学習の大きなブレークスルーとなった。そして 2022 年、ChatGPT に代表されるコンテンツを自動生成する技術が登場し、多様なメディアを対象として急速に進化しようとしている。深層学習の進化、音声や画像認識の実用化、そしてこれらの技術組み合わせることで、従来人間が行っていた文章の執筆、絵の描画、楽曲の制作、動画の撮影や編集といったクリエイティブな作業が AI によって置き換えられる時代が到来した。コンテンツ制作の主体がプロのクリエイターから一般の人々へと移行する可能性が出てきた。従来のクリエイター中心の視点から、ユーザー中心の視点へのコンテンツ経済圏のシフトが予測される。

AI が作成したコンテンツを AIGC (AI Generated Content) と呼ぶ。それがどのような経済圏を作るかを議論したい。脚本の生成や俳優の演技のデジタル複製・変更が簡単に行えるようになり、これが脚本家や俳優の役割や権利への影響をもたらすことが予想される。このような変化は、クリエイターとして知られる脚本家、アニメーター、俳優などの様々な分野の専門家たちの生態系に大きな変動を引き起こす可能性がある。日本には、ポケモンに代表されるキャラクターコンテンツを中心とした世界的に成功を収めているメディアフランチャイズ事業や、ユーザー主導でのコンテンツの流通を特徴とするコミュニケーションマーケットなどの独自の文化が存在する。その代表例として初音ミクを取り上げる。デジタル技術の進化、ユーザーの積極的な参加、ファンの熱狂、そして柔軟な著作権管理を組み合わせたビジネスモデルが、日本において生成 AI を効果的にビジネスに取り入れるための良い土壌を形成している。今後、ソーシャルメディアと生成 AI の組み合わせによって、ユーザー生成コンテンツ (UGC) が AIGC と一体化し、世界的に広がっていくことが期待される。一方で、生成 AI の技術の利用には、著作権法の問題や倫理的な課題など、様々な問題が伴う。特に、人間の感性や独自性を持つコンテンツの生成に関しては、AI とのバランスをどのように取るかが重要となる。生成 AI 技術と人間のクリエイターが対立するのではなく、互いに共存し、新しい形のコンテンツを共に生み出すことが、今後のコンテンツ産業の発展の鍵となる。

キーワード：生成 AI、ユーザー生成コンテンツ、メディアフランチャイズ、著作権法、初音ミク

1. はじめに

2023 年初夏に始まった全米脚本家組合および映画およびテレビ俳優の労働組合のストライキ[1]は人工知能 (AI) の進化によって変わる生態系の変化を象徴している。もちろん、これらのストライキは、生成 AI の脅威に対抗することだけが主旨ではない。映画やテレビ

¹ 大阪大学先導的学際研究機構教授

業界は、近年のストリーミングサービスの利用を加速させている。映画やドラマの制作・配信方法が大きく変わり、脚本家や俳優の報酬体系や雇用の安定性に問題が生じている。特に、配信限定の作品は再放送やコンテンツ販売の利益を再配分しないため、脚本家、俳優たちの収入機会を損なっている可能性がある。そして、もう一つの視点として、AIによるコンテンツ制作の効率化が挙げられる。AIの進化により、脚本の生成や俳優の演技のデジタル複製・変更が可能となりつつある。これにより、脚本家や俳優の役割が脅かされ、その待遇や権利に対する懸念が高まっている。様々な娯楽やアートの分野で創作活動を行う人々を「クリエイター」と呼ぶことにする。クリエイターの仕事には、文章の執筆、絵の描画、楽曲の制作、動画の撮影や編集があるが、今後、数年内に、彼らの仕事内容が大きく変わらざるを得なくなるであろう。娯楽やアートの配信プラットフォームの変化[2]と、ソーシャルメディアとAIによる高品質なユーザー生成コンテンツ(UGC)[3]の登場によりクリエイターの生態系が大きく変化することになる。ここに生じる軋轢を社会が受容しながら、時間をかけて新しいコンテンツ経済圏が登場することになる。背景となるのは、2022年から始まった生成AI[4]の急速な進展だ。ChatGPT(チャットジーピーティー)[5]に代表される生成AIと呼ばれる、文章や画像を自動生成するAIにより作られるコンテンツをAIGC(AI Generated Content)と略す。生成AIにより、コンテンツ提供者が、高度に訓練されたクリエイターから、市井の人々(ユーザー)に変わりつつある。そのユーザー作成コンテンツはUGC(User Generated Content)と呼ばれ、AI時代はAIGCとUGCが同一化して見られている。そこに問われる創造性は、画才や演奏力などの従来の提供者側の視点ではなく、芸術を鑑賞する目利き力や、文章の表現力などのユーザー側の視点になる。ユーザー参加なしに、次のコンテンツはない。これまでの2年間で起きている技術環境変化とこれから2年間で起こりえる事業環境変化を、生成AIの技術発展(2章)、メディアフランチャイズと影響を受けるコンテンツ経済圏(3章)、AIGCの新経済圏と権利処理の解釈(4章)で見えていく。

2. 生成AIの進化

生成AIは、2022年以降のAI技術の進化の中で特に注目されるようになった分野である。それまでのAIは、定型化された作業の自動化やデータの整理・分類を主な目的としていたが、生成AIはその枠組みを超え、データのパターンや関係を学習して新しいコンテンツを創出することが新しい。

深層学習(ディープニューラルネットワーク、DNN)は多層のニューラルネットワークを使用して複雑なタスクを学習する能力を持つようになった。図1に2010年以降の歴史を示す。最初の実用化は2011年ごろの音声認識である。それまでサイエンスフィクションに出ていた「コンピュータとの対話」が現実になったのはこの頃である。多層畳み込みネットワーク(多層CNN)の登場により、画像認識の精度が大幅に向上した[6]。今では空港で顔認証に使われている画像認識であるが、実用化になったのは数年前のことだ。CNNは、画像の局所的な特徴を捉える能力を持ち、多くの画像認識タスクで高い性能を示した。同じ時期2014年に、GAN(生成的敵対ネットワーク)が提案された[7]。GANは新しいデータを生成するための技術として登場している。これにより、高品質な画像や音声の生成が可能で、生成AIの幕開けとなった。一方、時系列データの処理には、LSTM[8]が使

用されるようになった。これにより、音声やテキストのような連続的なデータの処理が向上した。加えて、Attention（注視機構）と呼ばれるデータ相関の重要度を動的に重み付けする技術との組み合わせが機械翻訳の本命技術とみなされるようになったのが 2014 年ごろである。Attention は特に系列データの処理において、入力データの特定の部分に焦点を当てることで、モデルの性能を向上させることができる。Attention は、元々機械翻訳タスクにおいて、ソース文とターゲット文の間の関連性を捉えるために導入された²。2016 年、自然言語処理(NLP)は、Attention の多層・多重的な組み合わせで大発展した。それが Transformer[9]である。これにより、機械翻訳の精度が大幅に向上することになる。Transformer の中心的な要素である Attention は、入力データの異なる部分に「注意」を向けるメカニズムである。具体的には、シーケンス内の各単語やトークン³が他の全てのトークンとどれだけ関連しているかを評価し、その関連性に基づいて新しい表現を生成する。例を挙げると、文章の中で「彼」という単語が出てきた場合、Attention メカニズムは「彼」が指す実体や内容に関連する他の単語や情報に「注意」を向ける。これにより、文脈を考慮した情報の抽出や表現が可能となる。この Attention の仕組みは、長文において、各トークン間の関連性や依存関係を捉えるのに非常に効果的である。Transformer は、この Attention メカニズムを複数重ねることで、より複雑な関連性やパターンを捉えることができる。そして、それが事前学習（Pre-Training）という大発展につながる[10]。Pre-Training の手法は、大量のデータでモデルを事前学習し、特定のタスクに個別チューニングを行う方法の基礎として注目された。これにより、少ないデータで異なるタスクにも高い性能を達成することができるようになった。NLP における発展系が ChatGPT[11]である。この大規模言語モデル（Large Language Model, LLM）が登場し、より自然な文章の生成や、高度な質問応答タスクが可能となり、さらに画像生成技術と結合することにより AIGC 技術が急速に立ち上がった。技術の萌芽は 2014 年ごろであり、コンテンツ経済圏に影響をもたらす状況になったのは、2022 年と言える。

² 例えばソース文は英語であり、ターゲット文は翻訳された日本語となる。

³ ニューラルネットの設計によって異なる、単語・文字を表現するデータ単位。

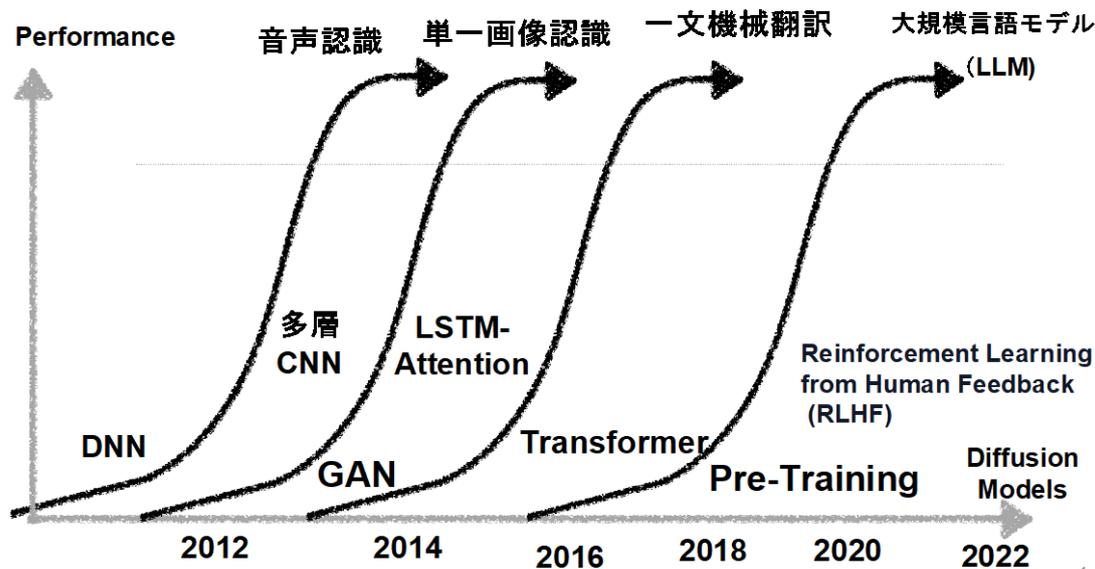


図1 深層学習の発展と生成AIの登場

2. 1. 自然言語処理：大規模言語モデル

AIGCの説明に入る前にまず、自然言語処理(NLP)を説明する。連続するN個のアイテム(通常は文字や単語)の列を表す統計モデルとしてN-Gram[12]がある。大規模言語モデルにより、文脈を反映して可能になったとはいえ、そもそも、言語モデルとは文字の連続から次の文字、あるいは欠落した文字を予測する確率過程を表したものである。テキストの生成や文章の補完などのタスクにも使用される。例えば、ユーザーが文章の途中まで入力すると、モデルはそれを元に文章を完成させる使い方は正しい。また翻訳に用いることも言語モデルの使い方として正しい。規模が大きくなったとはいえ、与えられた文章に対してもっともらしい文章を紡いでいく機械にすぎない。したがって、大規模言語モデルに問い合わせを入力して、その答えが合っていた、間違っていたというのは、見当違いな使い方だ。

ただ、Transformerという文字、単語、文章の相関関係をモデル化できる技術が登場したことにより飛躍的に予測効率が向上した。そのために膨大な文章を入力とすることから、知識ベースとしての機能が期待されるようになった。知識ベースとしての有用性が認められるようになったのがGPT-3だろう。GPT-3により、大規模言語モデルは、ユーザーの質問に答えるためのツールとして使用される。これにより、ユーザーは特定のトピックや情報に関する質問を簡単に行うことができる。企業や組織では、顧客サポートや質問・応答の自動生成などのタスクにも利用される。これにより、顧客の問い合わせに迅速に対応することができる。またプログラミングに利用することも知られるようになった。

大規模言語モデルの解説論文として文献[13]を取り上げる。ここに引用されている大規模現モデルの中で代表的なモデルを

表 1 に示す。これらのモデルは、数十億から数兆のパラメータを持つことで知られている。オープンソースのライセンス条件に注目したい。Google が開発した T5 は、多言語タスクに対応したトランスフォーマーベースのモデルであり、Apache-2.0 ライセンスで公開された。Galactica は著作物再利用の促進を目的とした国際的非営利団体であるクリエイティブ・コモンズが定める CC-BY-NC-4.0⁴で公開されている。これは著作者情報を表示する必要があり、利用は非営利に限られる。

Llama[14]は 2023 年 2 月に Meta の研究機関がリリースした。パラメータ数は 70 億、130 億、330 億、650 億のモデルが存在する。ソースコードが公開されているが、Llama のライセンスは非商用利用に限定されている。理由は、InstructGPT のように人間のフィードバックによる訓練をされておらず、誤った回答や攻撃的なコンテンツを生成するリスクがあると言われている。Meta は、2023 年 7 月に Llama2[15]を公開した。このモデルを基盤として、コードに特化したデータセットで追加訓練を行った結果、新たなモデル「Code Llama」も開発されている。Code Llama は、Python、C++、Java、PHP、JavaScript、Typescript、C#、Bash などの主要なプログラミング言語に対応している。提供されるモデルのサイズは 3 種類あり、1 つの GPU で実行できる 7B、13B、コーディング支援が可能な 34B のモデルがある。加えて、Python に特化した「Code Llama - Python」と、人間の自然言語入力に対する理解を深める「Code Llama - Instruct」の 2 つのファインチューニングモデルも提供されている。

表 1 代表的な大規模言語モデル

モデル名	開発者	サイズ (Billion)	特徴	ソースコードライセンス条件
T5	Google	11	言語タスクに対応した事前学習モデルの定番。	Apache-2.0
GPT-NeoX	EleutherAI	20	Megatronアーキテクチャに基づくモデル。	Apache 2.0
AlphaCode	DeepMind	41	GitHubデータとCodeforceの問題・解答を学習に使用。	クローズソース
LLaMA	Meta	65	20言語の大規模コーパスでトレーニング。パラメータの少なさでの高性能を目指す。	非商用オープンソース
Chinchilla	DeepMind	70	より多くのデータでトレーニングされた低パラメータモデル。Sparrow botで使用。	クローズソース
Galactica	Meta	120	科学的テキストとモダリティでトレーニング。	CC-BY-NC-4.0
LaMDA	Google	137	会話応答生成に特化したモデル。	クローズソース
GPT-3	OpenAI	175	一般的な目的のモデル。	クローズソース
Gopher	DeepMind	280	MassiveTextという10.5TBコーパスでトレーニング。	クローズソース
PaLM	Google	540	モデルスケールの実用的な限界を目指す。	クローズソース
GLaM	Google	1200	Sparse mixture-of-expertsモデル。GPT-3と比較してトレーニングは高価だが、推論は安価。	クローズソース

⁴ <https://creativecommons.org/licenses/by-nc/4.0/>

GPT-3.5 世代の言語モデルは、2022年11月以降に登場し、GPT-3を主なベンチマークにしているが、その開放性に関して、

1. 学習用、推論用ソースコードが公開され改変が自由であるか。
2. 学習済みモデルが公開されて追加学習ができるか。
3. 商用利用に制限があるか。

の3点が重要となる。Llama2は「月間アクティブユーザーが7億人以上のサービスでの利用には特別なライセンス取得が必要」と制限があるが、上記、オープンな言語モデルの代表作だ。実際に日本語の派生モデル[16]も登場している。この流れは加速するだろう。

大規模言語モデルの利用には以下の3形態が考えられる⁵。

1. 利用者が該当サービスを Web から利用する。
2. 利用者が該当サービスの API を利用して他サービスと組み合わせて利用する。
3. 第三者が当該サービスのソースコード、学習済みモデルを取得して、改良して自サービスとして他者に提供する。

上記の真のオープンな言語モデルは3の利用形態だろう。1と2の利用者にとっては、ソースコード非開放でも、1と2で得られたAI成果物が商用に利用できるかどうか課題となる⁶。そこでは、商用利用の可否と利用条件及び制限事項に注意する必要がある。OpenAI社の利用規約によれば、サービスから得られる出力に関する知的財産権はユーザーに帰属し、商用利用を含むあらゆる目的での利用が許可されている。したがって、OpenAI社のサービス、例えばChatGPTの出力は、商用利用が認められている。利用規約には、サービスの使用に関する制限事項が明記されている。具体的には、他人の権利を侵害する方法でのサービス利用、モデルのソースコードの取得、出力を人間が作成したものとして表現することなどが禁止されている。また、個人データをサービスで処理する場合、ユーザーは法的に適切なプライバシー通知を提供し、データ処理に関する同意を取得し、関連する法規に従ってデータを処理していることをOpenAI社に通知する必要がある。このようなガイドラインは、他社の大規模言語モデルによるWebサービス提供者と共通するものである。

大規模言語モデル(LLM)の構築において、データの収集は極めて重要である。以下に、各データソースの特性とその応用、課題についてまとめる。

● Web ページ:

- ▶ 応用: インターネットの普及により、多様なデータが生み出されており、LLMはこれらのデータを通じて多様な言語知識を獲得し、一般化能力を強化する。
- ▶ データ源: CommonCrawl⁷などの Web クロールデータ。

⁵ 後述する画像生成も同じ分類ができる。

⁶ <https://blog.brainpad.co.jp/entry/2023/05/16/153000> に ChatGPT 利用の注意点が明記されている。

⁷ <https://commoncrawl.org/>

- 課題:クロールされた Web データには、ウィキペディアのような高品質なテキストからスパムメールのような低品質なテキストまで含まれているため、データ品質を向上させるためのフィルタリング処理が必須となる。
- 会話テキスト:
 - 応用:会話データは LLM の会話能力を強化し、質問応答タスクのパフォーマンスを向上させる。
 - データ源:PushShift.io Reddit コーパスなどの公開会話コーパスやオンラインソーシャルメディアからの会話データ。
 - 課題:オンラインの会話データは多くの参加者間の議論を含むため、会話をツリー構造に変換する処理が必要。また、対話データの過度な統合は、LLM が指示や質問を誤って会話の開始と認識する原因となる可能性がある。
- 書籍:
 - 応用:書籍は正式な長文の重要なソースとして、LLM が言語知識を学び、長期の依存関係をモデル化し、物語性のある一貫したテキストを生成する。
 - データ源:Books3 や Bookcorpus2 データセットなどのオープンソースの書籍データ。
 - 課題:書籍のデータを使用する際には、著者や出版社からの心理的抵抗や懸念も考慮する必要がある。著作権者の許諾が望まれる。
- 専門的なテキストデータ:
 - 多言語テキスト:ターゲット言語のテキストの他に、多言語コーパスを統合することで、言語の理解と生成の多言語能力を強化する。
 - 科学的テキスト:人間による科学の探求により、科学的出版物の増加する。LLM の科学的知識の理解を強化するために、モデルの事前学習のために科学的コーパスを組み込む。
 - コード: プログラム合成は、コード上で訓練されたプログラミング言語モデルが使用され、特に研究コミュニティで広く利用されている。しかし、現時点で、これらのモデルで高品質で正確なプログラムを生成することが困難と言われている。

以上のデータソースは、LLM の一般的な言語モデリング能力と特定のタスク解決能力の両方を向上させるための基盤となっている。これにより、LLM は知識ベースとして、翻訳機として、プログラミング支援ツールとして動作する。GPT-3 も当初は、Web ページデータからなる対話システムであったが、GPT-4 になりプログラミングにも利用されるようになっていく[17]。

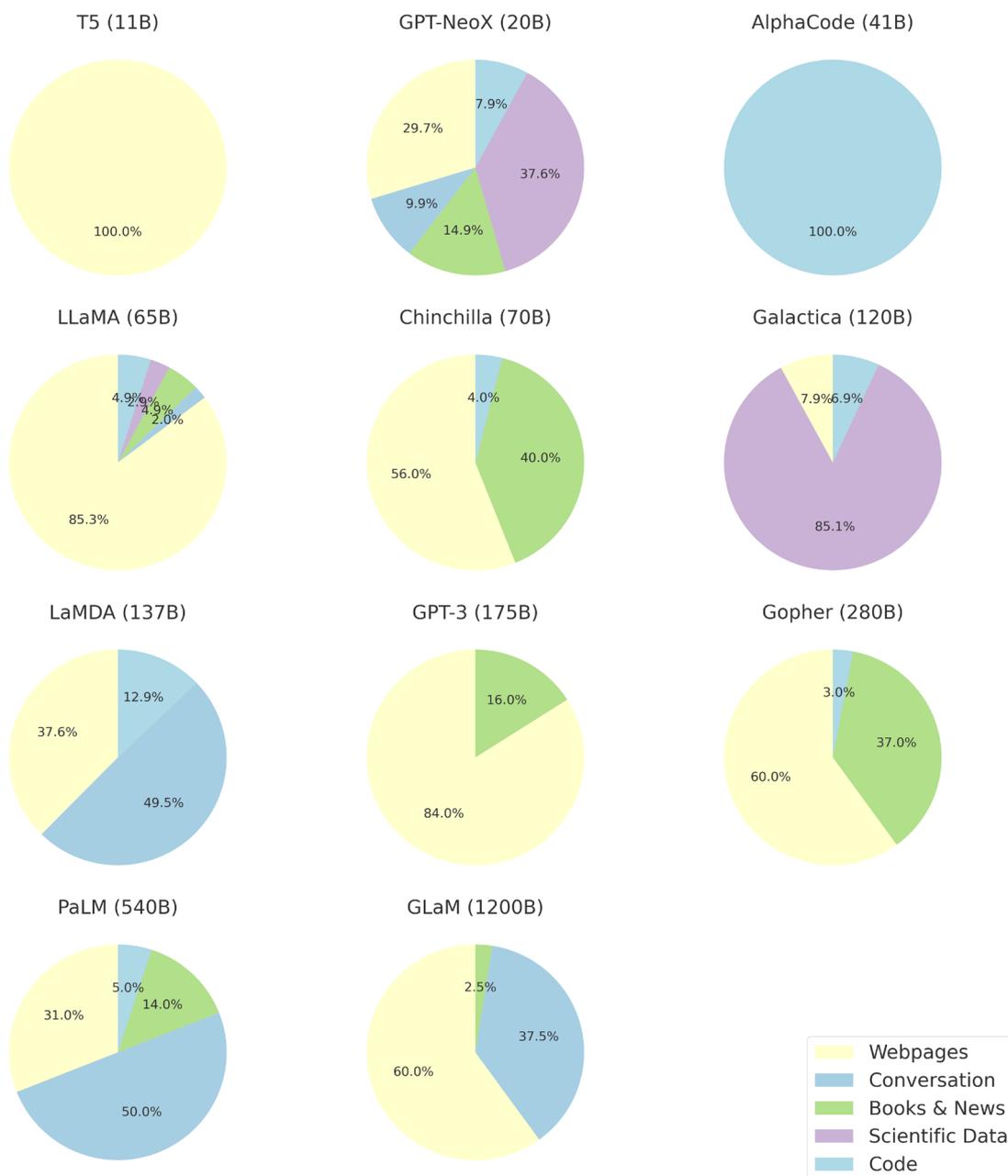


図2 既存のLLMの事前学習データにおける様々なデータソースの比率、[5]より作図

これだけのデータを集める際に、訓練データとして使用されるテキストには、二次利用を許さないとするコンテンツ、個人情報や機密情報、公序良俗に反するコンテンツの混入が避けられない[18]。したがって、データの収集と前処理の段階で、プライバシーに関する厳格なガイドラインとフィルタリングプロセスを適用することが不可欠となる。

図3にChatGPTのチューニング構造⁸を示す。OpenAIの言語モデル、GPT-3からGPT-3.5、そしてその後のChatGPTへの急速な進化は、特に人のフィードバックからの強化学

⁸ <https://lifearchitect.ai/chatgpt/>

習 (RLHF) [19]によるモデルのトレーニング手法の進歩によって支えられている。2022年1月には、OpenAIはGPT-3.5としても知られる InstructGPT を導入した。この上で、2022年11月には、短期間で驚異的なユーザーベースを築いた対話中心のモデル、ChatGPT が外部提供された。文章の生成モデルを対話モデルに大転換させることに成功したのは、人間がモデル出力を教導する RLHF の適用である。さらに RLHF は対話性能を向上させるばかりか、厳格なガイドラインに従い、有害または不適切なコンテンツを生成するリスクを減少させる。RLHF の導入は NLP 分野で 2022 年の最大貢献の一つと言えよう。

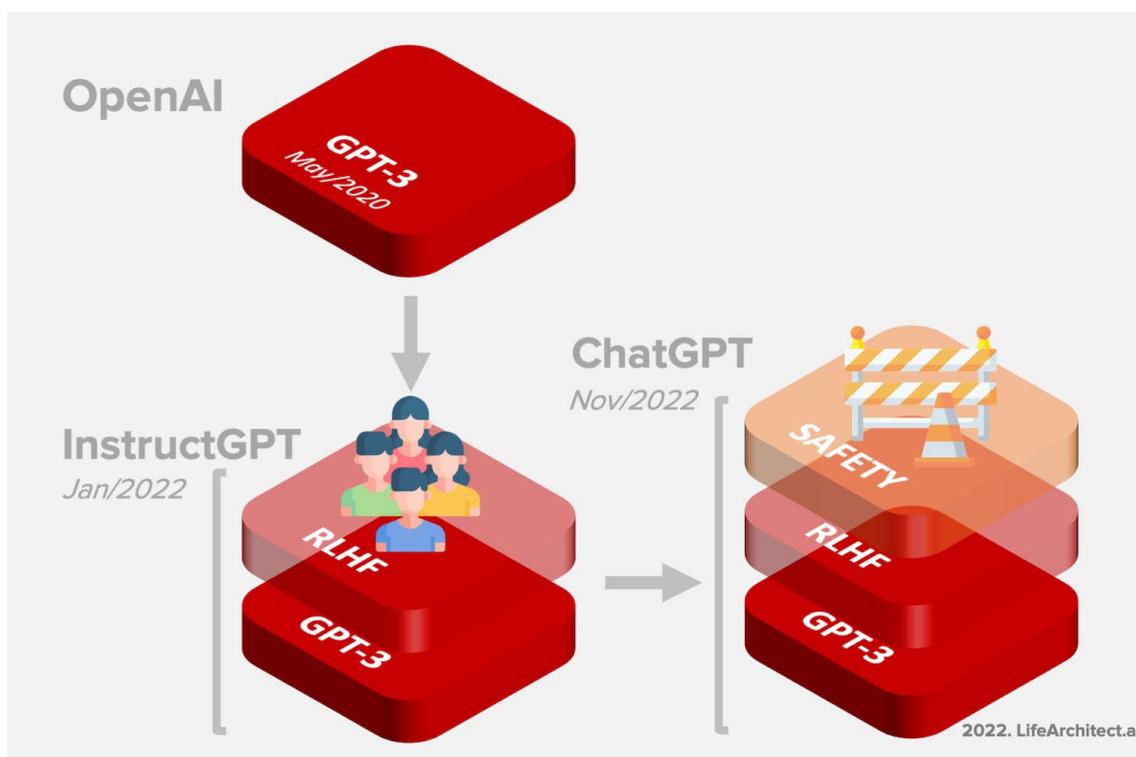


図3 ChatGPTのチューニング構造⁹

2. 2. 画像生成モデル

画像認識は自然言語処理と独立した研究分野として、ここ 50 年進化してきたが、2015 年ごろから、急速に統合処理の研究対象となるようになった。図 4 に画像認識 (CV)、自然言語処理 (NLP)、画像言語処理 (VL) から見た生成 AI の歴史[5]を示す。

画像認識[20]は、画像処理の中で、画像の意味を理解することを目標とした研究分野である。猫の画像をコンピュータに入力して、「猫」という言葉を出力させる技術を開発することであるが、GAN[7]の登場から、「猫」という言葉から猫の画像を生成する画像生成という研究分野と統合されて研究されるようになった。GAN は、生成器と判定器の 2 つの部分から成り立っている。生成器は、実例から虚例を生成することを試み、判定器は、

⁹ <https://lifearchitect.ai/chatgpt/>

入力の実例かどうかを判断する。異なる深層学習の2つのモデルで、片方が騙す、片方が真贋を見極めるといった競争により、画像を生成するというもので、その印象的な結果のために、生成AIのマイルストーンとなった¹⁰。Progressive GAN[21]やStyleGAN[22]などのGANの後継技術は、さらに性能を向上させている。

GANに加えて画像生成AIの主要構成画像処理技術として変分オートエンコーダ (VAE) [23]と拡散モデル[24]がある。

VAEは、データを低次元の確率的な分布に反映させ、元の入力に近い再構成を学習する試みとして提案された生成モデルである。VAEの発展として、DiffuseVAE[25]のような統合モデルが、VAEの能力をさらに拡張するための技術として提案されている。拡散モデルにはDenoising Diffusion Generative Model (DDPM) [26]がある。DDPMはノイズを徐々に追加し、その後徐々に除去するというプロセスを使用しているが、考え方は、データ分布の勾配を直接推定して、あるべき画像を自動生成するというスコアベースの生成モデル[24]と考え方は同じである。

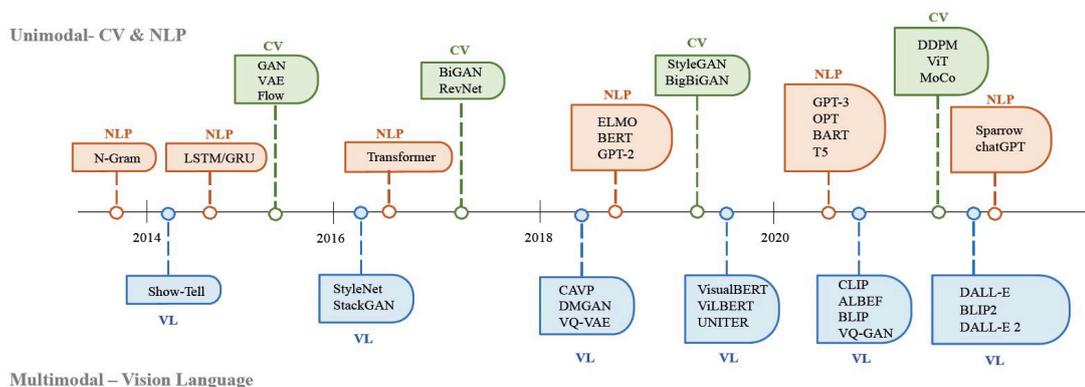


図4 画像認識 (CV)、自然言語処理 (NLP)、画像言語処理 (VL)から見た生成AIの歴史[5]

さて、ここで、画像生成モデルに大きな飛躍をもたらせたのは、画像言語処理 (VL : Vision and Language) という画像と言語の情報を同時に扱う研究分野であり、2019年を境に急速な進展が見られる。この背景には、上記のVAEと拡散モデルの登場、そして、NLPに革新をもたらしたTransformerの画像分野への適応である。

以下にVLの代表的な事例を列举する。

- VisualBERT (2019): ビジョンと言語のタスクを同時に学習するためのトランスフォーマーベースのモデルである。画像とテキストの結合表現を学習するために、事前学習の段階で画像と文章の両方のデータを用いている[27]。
- CLIP(2021): 異なる視覚的タスクに転送可能なビジョンモデルを学習するための手法である。文章と画像を同時に扱い、両方のモダリティ間での関連性を学習する[28]。

¹⁰ 後の章で説明する著作権法第30条4項が成立する2019年1月の前である。

- **Stable Diffusion(2022)**: 画像生成のための新しい手法である。Diffusion モデルとトランスフォーマーアーキテクチャを組み合わせ、高品質な画像を生成する[20]。
- **DALL-E 2(2022)**: テキストの記述に基づいて画像を生成するモデルである。トランスフォーマーアーキテクチャをベースにしており、多様な画像の生成が可能である[29]。
- **Midjourney(2022)**: Midjourney, Inc.は、Leap Motion の共同設立者である David Holz によってサンフランシスコで設立された。Midjourney の画像生成プラットフォームは、2022 年 7 月 12 日にオープンベータとしてリリースされている。アーキテクチャは Stable Diffusion に近いと想像されるが仕様は公開されていない。

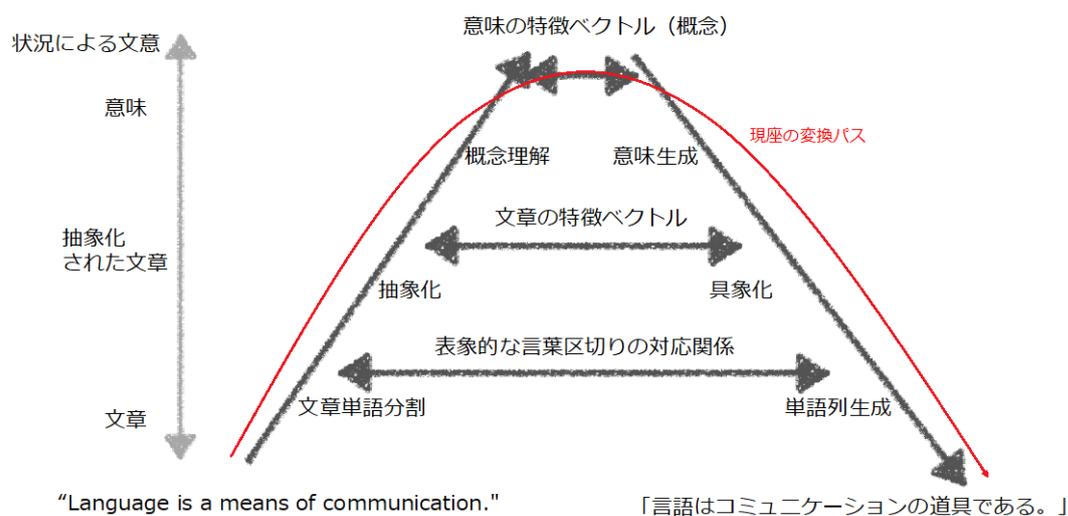


図 5 Vauquois の三角形 (著者により加筆・修正)

生成 AI の内部で何が起きているのかを直観的に理解するために、機械翻訳のアナロジーを使う。図 5 は、機械翻訳の理想を示すために 1960 年代に書かれた三角形のモデル[30]である。これを用いて、2020 年前後に起きたことを説明したい。左下の入力“Language is a means of communication.”を日本語に翻訳する際、Transformer 出現以前は、単語を分割し、対応する日本語に変換し並び替えるという表象的な言葉区切りの対応関係を統計的に行っていた。Transformer 出現以降は、深層学習により意味に相当する特徴ベクトルに変換され、意味から文章が生成され、翻訳されるようになった。三角形上位にある文章や意味は深層学習の特徴ベクトルとして表現される。BERT や 2016 年の機械翻訳は一文の特徴ベクトルの対応レベルであるが、GPT-3 以降は文脈を反映していることから最上位の意味の特徴レベルに近づいた抽象化が行われている (図 5 中の赤い経路)。ChatGPT では日本語の問い合わせに英語や中国語で応答することができるが、これは、知識処理が各言語で行われて、翻訳されているのではなく、抽象化された知識で翻案されたものが、必要な言語で具象化されているにすぎない¹¹。

同様に、画像生成では VAE が画像の抽象化を行っており、VisualBERT や CLIP が行っているのは、文章 + 画像の同時抽象化である。画像生成では、抽象化された特徴ベクトル

¹¹ <https://blog.modernmt.com/making-generative-ai-multilingual-at-scale/>

ルにノイズ付加するなどの学習処理の汎化に工夫が必要となるが、本質的に抽象化と具象化を扱うフレームワークは同じである。AI のモデルは図 5 の特徴ベクトルから学習される重みづけで構成されるが、そこには原文、原画像の直接的な表現はない。AI が古典的な k 近傍法[31]を使わない限りは、元データがモデルに含まれることはない。2016 年頃では機械翻訳で英語から日本語への変換レベルだった AI 技術が、2023 年頃に文章から画像への変換、文章から楽音への変換を行うようになったと言える。

2. 3. AIGC 技術

生成 AI の産業化は McKinsey のレポート[32]によれば小売と消費財産業、創薬、医療製品産業、銀行におけるマーケティング、R&D、基幹システムの置換に数%の効率化インパクトがあると述べられている。一方で、大企業から市井の人々まで、生成 AI を手に入れたことより、2022 年以降に急速にプロフェッショナル生成コンテンツ (PGC) とユーザー生成コンテンツ (UGC) が AIGC で占められるようになってきている[33]。これには、AI による絵画、AI による執筆、AI による音楽作曲、AI によるビデオ生成、AI による音声合成、AI によるプログラミングなどが含まれている。前記した生成 AI の技術により、デジタルコンテンツの新しい形態の生成とインタラクションを生み出している。市場調査会社 Precedence Research のデータによれば、2022 年の全世界の AIGC 市場の規模は 16 兆円であり、2030 年までに 106 兆円に達すると予想されている¹²。

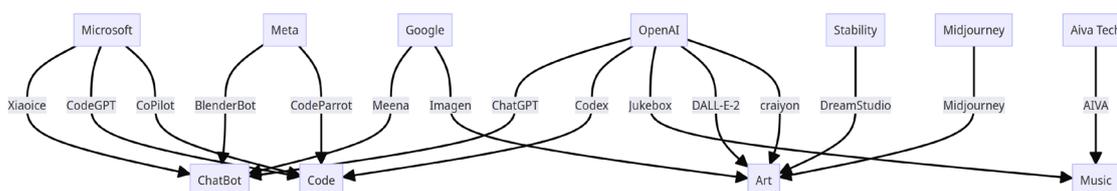


図 6 AI 生成コンテンツの開発者と応用分野 (文献[5]を修正)

文献[5]に登場する AIGC を開発者、サービス名（モデル名）、応用分野を図 6 に整理した。主要開発者が米国に集中していることがわかる¹³。

これらの技術の中で、Midjourney と Stable Diffusion を代表サービスとして比較した（表 2）。Midjourney は Web サービスとして提供されていて、ソースコード、生成モデルのデータは公開されていない。最新版の利用規約¹⁴によれば、生成した画像の商用利用は可能である。ただし、有料会員でない場合の画像は CC BY-NC 4.0¹⁵となり商用利用できない。また、年間 100 万ドル以上の収入のある企業による商用利用は、高額プランに加入する必要がある。非常に微妙な権利処理となっている。想定ユーザーは画像生成を趣味とする一般ユーザーと予想される。

¹² <https://www.chinadaily.com.cn/a/202306/14/WS64898027a31033ad3f7bc3b0.html>

¹³ Stability.ai は英国、Aiva Tech はルクセンブルク登記である。

¹⁴ <https://docs.midjourney.com/docs/terms-of-service>

¹⁵ <https://creativecommons.org/licenses/by-nc/4.0/deed.ja>

一方で **Stable Diffusion** を見てみよう。開発者である **Stable.ai** は自ら、**dream studio** という Web サービスを提供しているが、そのサービスの元になった画像生成 AI 技術をオープンソースで公開している。そのモデルも公開しており、急速にその派生技術が進化している。**Stable Diffusion** の商用利用¹⁶についての制約が、後述の AIGC がもたらす新経済圏を象徴していて興味深い。まず、**Stable Diffusion** は **CreativeML OpenRAIL** ライセンス¹⁷で提供されている。生成した画像に対する開発者の権利は主張されず、ユーザーが自由に使用できる。ただし、法律に違反する内容や人に危害を与える内容など、ライセンスで定められた規定に反する使用は禁止されている。商用利用が認められているが、以下の2つのケースでは商用利用が制約されている。

- 他者の画像から新たな画像を生成する場合：その他者からの画像に著作権がある場合は2次利用が制約される。**Stable Diffusion** では画像から画像を2次生成できるために、生成モデルだけでなく直接画像を改変することは商用利用に注意が必要となる。
- 他者のモデルを用いて追加学習させる場合：LoRA[34]という追加学習ができる機能が**Stable Diffusion** では多用されている。他人が作成した LoRA モデルを利用することにより、特定のアニメ風のイラストや特定の容貌をもった顔画像生成が容易になるが、その LoRA モデルが違法に作成された場合は、商用利用はできない。

LoRA は Low-Rank Adaptation の略で **Stable Diffusion** モデルの最も重要な部分であるクロスアテンション層に微小な変更を適用することで動作する。この部分は、画像とプロンプトが交差する部分であり、このセクションの微調整が優れたトレーニング結果をもたらす[35]。LoRA モデルは、Civitai¹⁸や HuggingFace¹⁹などの様々な場所で見つけることができる。特に Civitai は LoRA モデルの大規模コレクションとして、キャラクターLoRA、スタイル LoRA、コンセプト LoRA、ポーズ LoRA、衣服 LoRA、オブジェクト LoRA などがあり、学習済みモデルへの差分ファイルとしてダウンロードできる。

¹⁶ <https://yuryoweb.com/stable-diffusion-commercial-use/>

¹⁷ <https://github.com/CompVis/stable-diffusion/blob/main/LICENSE>

¹⁸ <https://civitai.com/>

¹⁹ <https://huggingface.co/>

表2 Midjourney と Stable Diffusion の比較

項目	Midjourney	StableDiffusion
想定ユーザー	スマートフォン利用者。ソーシャルメディア投稿や画像生成を趣味とするユーザー	高度な機能を求めるユーザー、技術的背景を持ち、詳細な調整を必要とするプロフェッショナル。
導入障壁	低い。直感的で短期間での習得が可能。月額20ドルの手頃な価格。	高い。技術的知識が必要で、1ヶ月以上の学習時間。GPUの手当てが必要。
対象画像	風景、物体。	キャラクター、グラフィックなどの詳細な描写が得意。
表現力	ソーシャルメディア投稿やブログ用途に適した標準的なクオリティ。ネガティブな指示を含むプロンプトには対応が弱い	高解像度で精緻な画像。領域単位の修正、動画への拡張。ネガティブ指示に対応。
開放性	ウェブサービスとして提供されているのみ。	オープンソース、オープンデータ。ユーザーが派生モデルを作ることが可能。
拡張性	機能追加やカスタマイズがユーザー側にはない。	機能追加やカスタマイズが容易。動画や他のメディア連携への拡張性が高い。
商用利用	有料であれば可能。	原則、可能。

Stable Diffusion は 50 億以上の画像とテキストのペアを含む「LAION-5B」というデータセットを使用している。LAION-5B は、ドイツの非営利団体 Large-scale Artificial Intelligence Open Network(LAION)²⁰によって 2022 年 3 月に研究用としてリリースされた。このデータセットは、58 億 5000 万の画像とテキストのペアで構成されており、多言語のデータが含まれている。LAION は、インターネット上のデータを解析し、前述した CLIP を用いて類似性の高い画像とテキストのペアを抽出している。このデータには原著者から著作権が主張されるものと個人情報が含まれている可能性がある。Stable Diffusion はオープンソースである、元の配布された画像生成モデルに個人、個社で追加学習を行うことができる。その上で、LoRA モデルという軽量の追加学習モデルが加わるので、強力な画像生成経済圏が構築されている。Civitai のサイトを訪れればそれが確認できる。以上のデータの流れを図 7 に整理した。

Stable Diffusion は、生データを LAION-5B、学習用データを Stable.ai が用いたデータ、学習用プログラムを Stable Diffusion の学習器、学習済みモデルを流通している Checkpoint²¹、楢円内学習済みモデルを LoRA、AI 直接生成物を LoRA で作成された画像として理解すると良い。これは開発者側から見れば、楽園のような世界だ。一般的に AI の開発は、商用可能なライセンス条件の下で、ソースコードの公開だけでなく相互運用可能なモデルが流通することが理想となる。Stable Diffusion の世界ではそれが実現されている。一方で、大規模言語モデルでは、Stable Diffusion のような開発者から見た理想郷は実現されていない。倫理的ガバナンスが必要であること、技術的には、学習、推論時の計

²⁰ <https://laion.ai/>

²¹ Checkpoint とは、Stable Diffusion の学習済みモデルを保存したファイルのことを示す。

算資源が膨大で図 6 に登場する巨大プラットフォームしか追加学習ができないことが原因である²²。

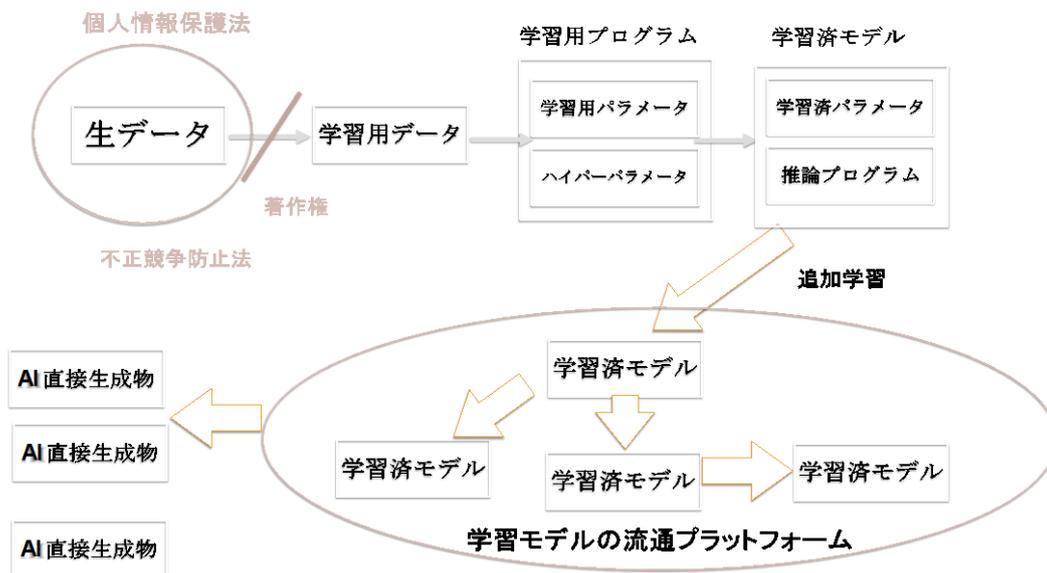


図7 Stable Diffusionの流通モデルの一般化

3. メディアフランチャイズのコンテンツ経済圏

AIGCの技術進展が2022年以降、急速に加速したことを前章で説明した。本章では、それによって影響を受けるコンテンツ経済圏に話を移す。既存のメディアフランチャイズ²³の経済圏にどうAIGCがインパクトを与えるかを見ていこう。メディアフランチャイズは小説、映画、マンガ、アニメ、コンピュータゲームなどのエンターテインメント分野において、ある商業作品が市場を持つ際に、その作品から派生した商品を多数の娯楽メディアを通じて製作する手法として知られている。この手法は、キャラクターや著作権といった知的財産(IP)を中心に、様々なメディアや企業を超えて活用する特性を持つため、「IPコンテンツ」や「IPもの」と業界用語で呼ばれることもある。この分野で日本は健闘している。次に、そのゆりかごとしてのコミックマーケットのゆるい権利処理について述べる。そして、技術とゆるい権利処理の組合せによる成功例として初音ミクを議論する。

3. 1. 日本のメディア・フランチャイズ

日本は数多くのメディアフランチャイズを生み出してきた国である。これらのフランチャイズは、アニメ、マンガ、ビデオゲーム、映画、そしてその他のメディアフォーマットを含む幅広いカテゴリーに跨るものである。特に、ポケットモンスターやモンスターストライクのようなフランチャイズは、国内外での成功を収めており、それらは商品販売やゲ

²² 機械翻訳では、SYSTRAN社の機械翻訳言語モデルのマーケットプレイスがある。
<https://www.systran.net/marketplace-catalog/>

²³ 日本では和製英語でメディアミックスと呼ばれる。

ームの売り上げなど、多岐にわたる収益源からの利益を上げている。日本のメディアフランチャイズは、その独自の文化とクリエイティブなアイデアにより、世界中の多くのファンを魅了してきた。

ウィキペディア²⁴にあるメディアフランチャイズの2023年8月における累積売り上げを図8のグラフにした。ディズニーはミッキーマウス、プーさん、スター・ウォーズ、ディズニープリンセス、ライオンキングなどのフランチャイズによる強力な存在感はあるが、総売上のトップフランチャイズはポケモンであり為替レートで上下するが1.5兆円を超える。それにハローキティ、マリオ、アンパンマン、ドラゴンボール、トランスフォーマー、ガンダムが上位に入っている。フランチャイズの収益は長寿であることと、地域でヒットし、グローバルに展開することが重要である。主なソースは、商品販売、興行収入、ビデオゲームの売上、ホームエンターテインメントの組み合わせとなっている。

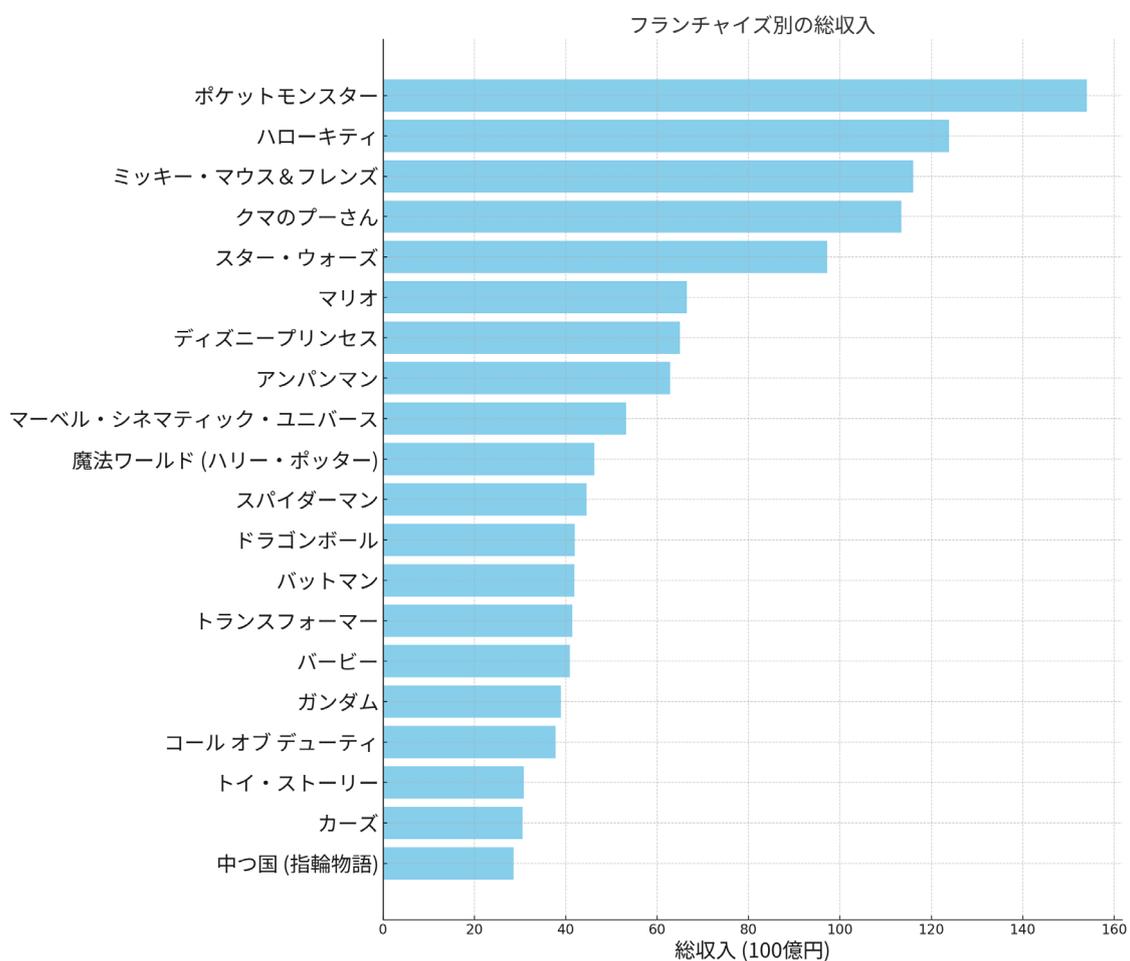


図8 メディアフランチャイズの累計売上高

²⁴ https://en.wikipedia.org/wiki/List_of_highest-grossing_media_franchises

3. 2. コミックマーケット（同人誌即売会）

コミックマーケット（通称コミケ）は、1975年に始まった世界最大の同人誌即売会である。年2回、夏と冬に開催され、出展者や参加者の数は年々増加している。このイベントは、多くの出展者が経済的利益を追求するのではなく、自分の作品を表現・共有する場として存在している。

コミケは、二次創作の最大の市場として知られる。その会場において、オリジナルの作品と二次創作が頒布される中、約7～8割の出店者が二次創作に従事しているとの報告がある[36]。コミケの人気の増大するにつれ、著作権の問題が焦点となってきた。特に、既存作品をベースにした同人誌の販売は、翻案権や同一性保持権のような著作権を侵害するリスクが指摘されている。翻案権は、オリジナル作品のアレンジを行う権利で、原作者のみがその権利を持っている。一方、同一性保持権は、作品のキャラクターやストーリーを変更することを制限する著作権の一部として存在する。これらの権利が侵された場合、原作者は該当の同人誌の販売停止を要求できる。

著作権侵害に関して、二次創作は親告罪の対象となるため、被害者が告訴しない限り、法的手段が取られることは少ない。事実、多くの二次創作が市場に出回っており、特に悪質と判断されるケース以外では、多くの著作権者がこれを容認しているとされる。コミケは、表現の自由の場としての役割を果たし続けており、原作者が少数の売り上げを得るクリエイターを訴えることは稀である。メディアフランチャイズの視点から見れば、コミケにおける二次創作の黙認は、ファン層の拡大や商品の売上げ向上の機会として捉えられている。「コミケは我が国が世界に誇るコンテンツのゆりかごの役割を果たしている」、「作品の受け手が送り手に容易に変わり得る、それが入れ替わり続けることで多様性と創作の再生産を可能にしている」という指摘[37]がある。

コミケの位置付けは、日本の特有な二次創作文化として重要であり、これは新しい技術、例えば画像生成 AI による UGC という文化との親和性が高い。

3. 3. 初音ミクに見る経済圏の構成要素

初音ミク（Hatsune Miku）は、クリプトン・フューチャー・メディアが（CFM）発売しているバーチャル・シンガーソフトウェアのキャラクターであり、「電子の歌姫」として知られている。彼女は「ボカロ」という文化・ジャンルを築き上げ、ネット上の文化や音楽シーンに大きな影響を与えた。日本史の教材では、初音ミクは「現代の IT 技術が生み出した新たな文化の象徴的存在」として評価されている。初音ミクは音声合成ソフトを超えて多くのクリエイターによって様々な創作活動が行われ、「初音ミク現象」として知られるムーブメントを起こしている。3DCG ライブ、ゲーム、フィギュアなどでメディアフランチャイズ事業が成立している。初音ミクは UGC によって成立していると言え、CFM が発行したライセンスに基づいて無償の二次創作活動が許可されている。

これまでの議論を元に、ユーザー生成コンテンツとメディアフランチャイズの関係を図9に整理した。これの構成要素を以下に列挙することで初音ミクを理解したい。

- IP コンテンツ：初音ミクは、CFM によって開発されたバーチャルシンガーである。彼女はボーカロイドという音声合成技術を使用して、ユーザーが入力した歌詞とメロディに従って歌を歌うことができる。
- オーディエンス構築：CFM は公式サイト「Piapro」を立ち上げた。このサイトは、ユーザーが非営利でコンテンツを共有し、新しいコンテンツを生み出すための場として機能する[38]。初音ミクの人気は、CFM の積極的なマーケティング戦略によって急速に高まった。特に、ユーザーが自由に楽曲を作成し共有することを奨励するアプローチは、大きなコミュニティの形成を促進した。
- 作成ツール：初音ミクの声は、ボーカロイドという音声合成技術を使用して生成される[39]。その後、MikuMikuDance(MMD) [40]などのツールが登場し、ユーザーは初音ミクを用いたアニメーションやダンスの動画を作成できるようになった。
- ルールと基準：CFMは、初音ミクの使用に関するガイドラインを設定している。特に、CC BY-NC 3.0 (表示・非営利 3.0) のライセンスの下での使用が奨励されている。これにより、誕生から現在にかけて、ユーザーによって無数のイラストや楽曲でアレンジされ、自由に使われてきた。
- 配信プラットフォーム：初音ミクの成功には、動画配信サービス「ニコニコ動画」の存在が大きい。ニコニコ動画は、ユーザーが動画をアップロードし、コメントを共有することができるサービスであり、初音ミクの動画が多数アップロードされることで、彼女の人気は急速に広がった[41]。
- 製作者：多くのユーザーが初音ミクを自由に使用して楽曲や動画を作成しており、これらの作品は二次創作、三次創作、N次創作として広く認識されている。
- 利用者：初音ミクのコンテンツは、音楽ファン、アニメファン、クリエイターなど、様々な背景を持つ人々に楽しまれている。

初音ミクは、情報処理技術の進歩により、これまでの肉声でなければ歌手ではないという常識を破り、伝統的なキャラクタービジネスのセオリーを覆し、ユーザー主導のクリエイティブな活動を促進することで、独自のビジネスモデルを築き上げた。コミケに代表される日本の UGC 文化と、図 9 の技術側面である作成ツールと配信プラットフォームの進化により、第2、第3の初音ミクが誕生する可能性を示唆している。

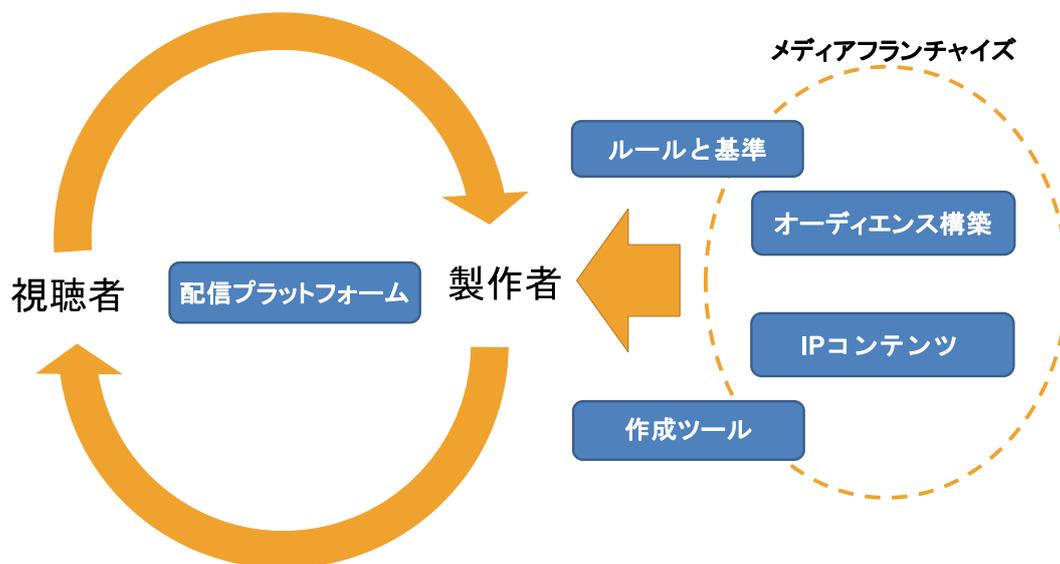


図9 ユーザー生成コンテンツとメディアフランチャイズの構成

4. AI 生成コンテンツ (AIGC) による新経済圏と課題

生成 AI という新たな破壊的イノベーションが舞台に登場し、AIGC によって、コンテンツ産業全体が進化する機会が訪れている。しかしながら、この新技術の進化速度にガイドライン、社会の受容性、法制度が追いつけず、多くの問題点が浮上している。AIGC によって生まれる新たなビジネスモデルを次節で紹介し、その後、権利処理を考慮して、クリエイターの生態系に現れている課題を概観する。

4. 1. AIGC による事業機会

多くの Web サイト²⁵に画像生成 AI で対価を得る手法が紹介されている。その中で、今後の技術進化を見て拡大しそうな事業候補を以下に示す。

- ゲーム開発：ゲームの視覚資産の生成を迅速化でき、すでに利用が進んでいる。
- 映画とアニメーション：映画やアニメーションのコンセプトアートを迅速に生成する。2023 年時点での不十分な動画生成能力も、今後の進展によっては破壊的な結果をもたらす可能性がある。
- キャラクターコンテンツの育成：初音ミクはボーカロイドであったが、これがバーチャル芸人として進化し、オンラインやオフラインの両方で展開されると考えられる。すでにバーチャルユーチューバー[40, 42]ビジネスがあり、これらの拡大が今後期待できる。
- 建築デザイン支援：建築、土木、不動産の設計素材の生成、顧客説明に利用する。

²⁵ 例えば <https://ambcrypto.com/blog/how-to-make-money-with-midjourney-15-easy-ways-to-check-out/>

- 衣装ファッションデザイン：カスタマイズファッションコンテストがすでに開催されている²⁶。
- 広告コンテンツの作成：Web サイト、チラシ作成で実際の利用が進んでいる。
- アートプリントの販売：ロゴデザイン、プリントアートのオンデマンド産業²⁷が出現する。
- グラフィックノベルや子供の本の出版：AI 生成アートを特長とした自己出版の方法を探求する。
- グラビア雑誌作成：すでにこれは Stable Diffusion の応用として進んでおり、日本発のコンテンツが多い²⁸。
- オンライン教育コースの作成と販売：AI アート生成スキルを活用し、オンラインコースを設計・販売する。
- プロンプトの販売：AIGC の制御文であるプロンプトを市場で販売する。

以上を俯瞰すると、AIGC はまずゲーム、映画、アニメの世界はコミケに参加している少人数グループレベルから大規模制作会社まで広い範囲でコンテンツ制作の効率化に貢献する。建築、ファッションや e コマースの分野では、企業の内部利用と対顧客のマスクスタマイゼーション[43]と呼ばれる個人向け価値提案を行う利用が増えていくだろう。そして最後は個人事業としてできるコンテンツ、ノウハウ販売だ。AI でイラストを作成する者の総称として「AI 絵師」という言葉があり、それが職業として成立しようとしている。生成 AI のプロの利用は洋の東西を問わず、大企業では内部で静かに進むだろう。UGC ではどうか。日本は先頭を走ることが可能だ。是非は別として 3. 2. で述べた二次創作について権利意識が厳格でないコミケ文化がある。さらに 3. 3. で述べた初音ミクのように新しい技術とキャラクターを組み合わせるユーザーがいる。

4. 2. AIGC の課題

AIGC を作るために何が必要か。言語から画像を生成するためには画像と同時に学習した言語モデルが必要だ。Stable Diffusion に議論を絞ろう。Stable Diffusion が使用するデータセットは、不適切な内容を含む可能性が指摘されており、商用利用には不向きであるとの警告がなされている。それにもかかわらず、一部の企業はこのデータセットを商用ライセンス下で公開し、収益化している。Stable Diffusion の組成で前述した"LAION-5B"は、インターネットから収集した画像とキャプションを大量に含むものである。このデータセットのリリースにより多言語大規模トレーニングと視覚言語モデル研究の可能性が広がった。

米国では、画像生成 AI を巡り、Stability AI、Midjourney、DeviantArt の 3 社に対して集団訴訟が提起されている²⁹。これらの AI はアーティストやクリエイターの著作権を侵

²⁶ <https://fashionchallenge.ai/ja>

²⁷ <https://sozomuseum.com/entry/>

²⁸ 日本の大手出版社がグラビア作成を断念したという事例がある。

<https://www.itmedia.co.jp/news/articles/2306/07/news150.html>

²⁹ <https://gigazine.net/news/20230116-stable-diffusion-midjourney-litigation/>

害する形で製品化されたとされる。特に、**Stable Diffusion** は多くの著作権で保護された画像を訓練データとして使用し、アーティストの許可や報酬なしに画像がダウンロードされて使用されているとの指摘がある。米国のフェアユースの概念には曖昧性があり、**LAION-5B** データセットを用いた AI 画像生成利用企業は、その曖昧性によるリスクを抱えている。原告の主張によれば、研究目的で収集されたデータを商用に利用する行為は、フェアユースの原則に反する可能性がある。この問題は、AI 技術の商用展開において、新たな法的規範やガイドラインの確立が必要であることを示している。フェアユースは米国の著作権法における例外的な概念で、特定の条件の下で、著作権によって保護される作品を許可なく使用することを許容する。この原則の存在理由は、著作権の保護と公共の利益との間のバランスを取るため使用の目的や性質、使用される作品の性質、使用の量や範囲、そして使用による潜在的な市場や価値への影響という 4 つの要因が考慮される。これらの要因は、具体的なケースごとに検討され、一つの要因だけでフェアユースが決定されるわけではない。裁判結果が出るのはこれからである。

一方で日本はどうか。2019 年 1 月に発布された著作権法第 30 条の 4 [44] という AI 学習に寛容な画期的な法律がある³⁰。

この著作権法第 30 条の 4 に基づくと、AI の開発に関しては「AI 開発・学習段階」と「生成・利用段階」の二つの段階に分けて考慮される。AI の「開発・学習段階」では、著作物に表現された思想や感情を目的としない利用行為として、著作権者の許諾なしに著作物を利用することが可能とされている。しかし、これは「著作権者の利益を不当に害することとなる場合」を除くとされている。学習用データとしての著作物の収集・複製や、AI 学習目的でのデータベースの著作物の複製は、著作権者の利益を不当に害することがない限り、許される。これは米国のフェアユースが判例なしに認められていることに等しい。一方、AI の「生成・利用段階」では、通常の著作権侵害の規定が適用される。AI を用いても、特別な扱いはされず、AI によって生成されたコンテンツが著作物に「依拠性」と「類似性」が共に認められない場合は著作権侵害とならない。

結論として、AI の「開発・学習段階」では、一定の条件下で著作物を無許可で利用できるが、「生成・利用段階」では、通常の著作権の規定が適用されることとなる。深層学習により巨大パラメータ数の中に拡散してしまったデータを使うことに、依拠性があるかどうかは判断が分かれるが、AIGC の一作品が、明らかに他の著作物に類似していなければ、その AIGC 作品の使用は著作権侵害とはされない。

このような現状に対して、クリエイター団体から以下の提言が出ている[45]。

1. **Stable Diffusion** に代表される画像生成 AI の機械学習における著作物の使用は、著作権法 30 条の 4 に規定する著作物の権利制限の対象外とすること。
 - ・画像生成 AI の学習における著作物の使用は著作権の原則に従いオプトイン方式とし、著作者から著作物の使用許可を事前に得ること。
2. 画像生成 AI の使用において、AI の機械学習に使用した著作物の著作者に対し、学習への使用及びその AI が消費者に使用された回数等に応じた使用料を支払うこと。

³⁰ 動画解説が理解しやすい。 <https://www.youtube.com/watch?v=eYkwTKfxyGY>

3. 著作権は、これまで通り、思想または感情の創作的表現に与えること。
 - ・画像生成 AI 生成物においては、すべて又は大部分が AI 生成物である制作物を著作権の保護の対象とせず、創作的寄与が明確に認められるもののみ保護の対象とすること。
4. 画像生成 AI 生成物は、AI 生成物であること及びその起源の明示を義務付けること。
5. 画像生成 AI を、著作権法にとどまらず、人権侵害や安全保障を含む幅広い観点でリスク評価し、国際的な枠組みに沿って適切に規制すること。

1と2の実現には現行著作権法の法改正が必要になる。裏返せば、現行法の開発者への寛容度は、世界的に見て稀有な位置付けになっている。3については、現時点で、AIGCに著作権を与えるべきでないとの意見が大勢に見える。一方で、生成 AI を用いたとしても、人が思想・感情を表現しようとする創作意図があり、創作的寄与と認めるに足る行為を行い、生成物が思想感情の創作的表現として評価される場合は、著作権が与えられる可能性がある[46]。生成 AI の利用が一般的になった社会では、判断が変わるかもしれない。4と5は社会倫理設計の重要な課題になる。

4. 3. AIGCの課題に対する考察

Stable Diffusion の流通モデル (図 7) を振り返ると、現行法では公開されている生データから学習済みモデルは著作者の同意なく作成することができる。おそらく米国もフェアユースの概念で認められることになると予測する。Stable Diffusion では追加学習がユーザー側で自由に行えることから、「低賃金労働で作られ、個人情報を含み、違法に取得されたデータ」が含まれる倫理的・法的問題があったとしてもそれを検査することが非常に難しい。つまり、データ利用の倫理が保証されないことが課題になる。一方で、図 3 の ChatGPT のチューニング構造は、RLHF³¹という訓練された人の指導が生成 AI の制御機構として組み込まれており、さらにその上に、倫理的に許されるコンテンツのみを出力するというフィルタリング機構が構築されている。Stable Diffusion のオープンソース・オープンデータのモデル流通システムは、開発者にとって理想だが、一方で ChatGPT の管理されたクローズドシステムの良さも評価する必要がある。

インターネット上の偽造コンテンツを自動発見するツール開発[47-49]は必須だろう。しかし、矛と盾の逆説と同じように、偽造コンテンツ生成と摘発の能力は拮抗していく。いずれは合法・違法、倫理的是非に関係なく AIGC は広まっていく。オープンソース化した Stable Diffusion の拡張技術の発展は目覚ましい。例えば ControlNet[50]は画像中の人物、物体の姿勢を反映した動画像を半自動で生成することができる。まだまだ、生成 AI による動画像生成は発展途上だが、数年内にはストーリーと一致した違和感のない動画像自動生成が可能になるだろう。

³¹ 人間からのフィードバックを用いた強化学習という直訳どおり、人が介在するというフアクターが重要である。

AI による文章、動画、楽曲のゴーストライティングが広まる可能性がある。それはまるで 2000 年に起きたナップスター事件を超えるコンテンツ経済圏の変化をもたらすことになるかも知れない。ナップスター[51]は 1999 年に創業し、音楽コンテンツを自由にダウンロードできるようにした。友達が 1 枚のコンパクトディスク (CD) を買えば、友人が全て共有できる仕組みだ。創業から 2001 年にかけて、Napster の存在により CD の売上が毎年 10% 以上減少した。1999 年を境に、全世界の CD 売上は 5 年で半減している。しかし、違法性を経営陣が認識していたことからナップスターは 2002 年に操業停止となっている。

それからどうなったか。コンテンツ収入は CD から Apple の iTunes、Spotify と言った音楽配信サービスに移行し、音楽家たちの収入はライブ市場に移って行った。同時に配信サービス音楽コンテンツのグローバル化を促すこととなった³²。MP3 の違法ダウンロードをアナロジーとして考えると、AI ゴーストライターによる UGC の拡散が、これまでの著作権ビジネスを破壊しそうな状況を作り、そこから生成 AI を使いこなすプロフェッショナルクリエイター、熱烈なユーザークリエイター、配信システム、ソーシャルメディアを巻き込んで、新たなコンテンツ経済圏が生まれるだろう。

5. まとめ

コンテンツ生成における新たな可能性が拓かれつつある。AI による文章生成、映像制作、音楽作成など、多岐にわたる領域での活用が進められている。本稿では、2 章で生成 AI の技術進化が急加速していることを報告した。そして 3 章でメディアフランチャイズのコンテンツ経済圏を俯瞰し、4 章でユーザー主導の AIGC による新経済圏の想定を述べた。本稿執筆時点 (2023 年夏) の生成 AI 関連の引用文献の多くは 2022 年から 2023 年の新技術である。生成 AI の技術進化は目覚ましく、来年に引用すべき文献は一変しているだろう。生成 AI の技術開発について日本は米国、中国に大きく劣後している。一方で、日本企業はキャラクターと呼ばれる IP コンテンツを軸としてメディアフランチャイズの事業化に成功している。映像媒体、ゲーム、イベント、商品化などの多様な展開を通じて、海外での IP 価値を高めている。この理由として、戦略的なオーディエンスチャネルの構築、アニメの広範な配信、現地企業との柔軟な IP ライセンス契約があるが、強調したいのはファンコミュニティ活動の促進である。コミックマーケットという二次創作に寛容な文化的背景があるからこそ、画像生成 AI の日本での活用が見えてくる。

日本のコンテンツ産業は米国や中国と異なり、コンテンツ産業とメディア・テクノロジー産業との統合が進んでいないように見える。統合することが伸び代になる。初音ミクに見られるように、作成ツール、配信プラットフォームの利用、クリエイター育成にユーザーを巻き込んでいくことにより、新たな IP コンテンツを中心とした多元的経済圏の形成が可能になる。日本には生成 AI の利用によりコンテンツ産業が発展する素地ができている。

現時点で、AI によるコンテンツ生成は、効率的で一貫性のある出力を得ることができる一方、人間の感性や独自性を持つコンテンツを生み出すことは難しい。逆に、人間のクリ

³² <https://news.yahoo.co.jp/expert/articles/6688469800a4d8acf7f40370070297e86c8e432f>

エーターはその独自の視点や感性を活かしたコンテンツを生み出すことができるが、生産性やコスト面での課題が存在する。このような背景から、新たな経済圏と古い経済圏が対立するだけでは共倒れが危惧される。創造的コンテンツの供給として生成AIを使いこなすクリエイターが必要である。具体的には、AIを活用した効率的なコンテンツ生成と、人間の感性や独自性を活かしたコンテンツ制作を組み合わせることで、より質の高いコンテンツを生み出すことが可能となる。また、AI技術の進化をクリエイターたちが積極的に取り入れることで、新しい表現方法やコンテンツの形態が生まれる可能性もある。AI技術と人間のクリエイターが単なる対立軸として捉えられるのではなく、共存共栄の関係を築くことが、コンテンツ産業の未来を拓く鍵となる。

参考文献

- [1] J. Hsu, "Who will control AI-created'digital twins'?", ed: Elsevier, 2023.
- [2] O. Budzinski, S. Gaenssle, and N. Lindstädt-Dreusicke, "The battle of YouTube, TV and Netflix: an empirical analysis of competition in audiovisual media markets," *SN Business & Economics*, vol. 1, no. 9, p. 116, 2021.
- [3] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business horizons*, vol. 53, no. 1, pp. 59-68, 2010.
- [4] M. Jovanovic and M. Campbell, "Generative artificial intelligence: Trends and prospects," *Computer*, vol. 55, no. 10, pp. 107-112, 2022.
- [5] Y. Cao *et al.*, "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt," *arXiv preprint arXiv:2303.04226*, 2023.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [7] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [9] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [11] OpenAI, "Introducing chatgpt," *OpenAI Blog*, 2022.

- [12] C. E. Shannon, "Prediction and entropy of printed English," *Bell system technical journal*, vol. 30, no. 1, pp. 50-64, 1951.
- [13] W. X. Zhao *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.
- [14] H. Touvron *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.
- [15] H. Touvron *et al.*, "Llama 2: Open foundation and fine-tuned chat models," *arXiv preprint arXiv:2307.09288*, 2023.
- [16] "ELYZA-japanese-Llama-2-7b-instruct-demo."
<https://github.com/camenduru/ELYZA-japanese-Llama-2-7b-instruct-demo-hf>
(accessed).
- [17] B. Yetiştirilen, I. Özsoy, M. Ayerdem, and E. Tüzün, "Evaluating the Code Quality of AI-Assisted Code Generation Tools: An Empirical Study on GitHub Copilot, Amazon CodeWhisperer, and ChatGPT," *arXiv preprint arXiv:2304.10778*, 2023.
- [18] N. Kandpal, E. Wallace, and C. Raffel, "Deduplicating training data mitigates privacy risks in language models," in *International Conference on Machine Learning*, 2022: PMLR, pp. 10697-10707.
- [19] D. M. Ziegler *et al.*, "Fine-tuning language models from human preferences," *arXiv preprint arXiv:1909.08593*, 2019.
- [20] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," *arXiv:2112.10752 [cs.CV]*, 2022.
- [21] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.
- [22] O. Patashnik, Z. Wu, E. Shechtman, D. Cohen-Or, and D. Lischinski, "Styleclip: Text-driven manipulation of stylegan imagery," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2085-2094.
- [23] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [24] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems*, vol. 32, 2019.
- [25] K. Pandey, A. Mukherjee, P. Rai, and A. Kumar, "Diffusevae: Efficient, controllable and high-fidelity generation from low-dimensional latents," *arXiv preprint arXiv:2201.00308*, 2022.
- [26] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840-6851, 2020.
- [27] L. H. Li, M. Yatskar, D. Yin, C.-J. Hsieh, and K.-W. Chang, "Visualbert: A simple and performant baseline for vision and language," *arXiv preprint arXiv:1908.03557*, 2019.

- [28] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*, 2021: PMLR, pp. 8748-8763.
- [29] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, vol. 1, no. 2, p. 3, 2022.
- [30] B. Vauquois, "A survey of formal grammars and algorithms for recognition and transformation in mechanical translation," in *Ifip congress (2)*, 1968, vol. 68, pp. 1114-1122.
- [31] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, 2022.
- [32] M. Chui *et al.*, "The economic potential of generative AI: The next productivity frontier," McKinsey & Company, 2023. [Online]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction>
- [33] *Reshaping E-Commerce: The Influence Of AI-Generated Content*. 2023.
- [34] E. J. Hu *et al.*, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.
- [35] S. Ryu, "Low-rank adaptation for fast text-to-image diffusion fine-tuning," ed, 2023.
- [36] 河原優子, "< 論文 > 二次創作文化の集団論的検討," *京都社会学年報: KJS*, vol. 28, pp. 127-148, 2020.
- [37] 文化審議会著作権分科会法制・基本問題小委員会, "第6回議事録," 文部科学省東館3階講堂, 2015. [Online]. Available: https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoki/h27_06/
- [38] 伊藤博之, "CGMの現在と未来: 初音ミク, ニコニコ動画, ピアプロの切り拓いた世界: 3. 初音ミク as an interface," *情報処理*, vol. 53, no. 5, pp. 477-482, 2012.
- [39] 後藤真孝, "「初音ミク」はなぜ注目されているのか," *電気学会誌*, vol. 132, no. 9, pp. 630-633, 2012.
- [40] L. K. Le, "Examining the rise of Hatsune Miku: The first international virtual idol," *The UCI Undergraduate Research Journal*, vol. 13, no. 1, pp. 1-12, 2014.
- [41] 濱崎雅弘, 武田英明, and 西村拓一, "動画共有サイトにおける大規模な協調的創造活動の創発のネットワーク分析 ニコニコ動画における初音ミク動画コミュニティを対象として," *人工知能学会論文誌*, vol. 25, no. 1, pp. 157-167, 2010.
- [42] X. Zhou, "Virtual youtuber kizuna ai: co-creating human-non-human interaction and celebrity-audience relationship," 2020.
- [43] G. Lanza, S. Peukert, and G. L. Steier, "Latest advances in cloud manufacturing and global production networks enabling the shift to the mass personalization

- paradigm," in *Design and operation of production networks for mass personalization in the era of cloud technology*. Elsevier, 2022, pp. 39-77.
- [44] 文化庁, "著作権法の一部を改正する法律 (平成 30 年法律第 30 号) について," *知財ぶりずむ: 知的財産情報*, vol. 18, no. 214, pp. 154-169, 2020.
- [45] クリエイターと AI の未来を考える会, "画像生成 AI の適正使用及びそれに伴う著作権制度の整備等に関する提言 (第 2 版) ," 2023.08.11 2023. [Online]. Available: <https://support-creators.com/archives/87>
- [46] 愛知靖之, "AI 生成物・機械学習と著作権法," *パテント*, vol. 73, no. 8, pp. 131-146, 2020.
- [47] L. Hu, S. Wei, Z. Zhao, and B. Wu, "Deep learning for fake news detection: A comprehensive survey," *AI Open*, 2022.
- [48] R. Chauhan, R. Popli, and I. Kansal, "A Comprehensive Review on Fake Images/Videos Detection Techniques," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, 2022: IEEE, pp. 1-6.
- [49] Y. Hamid, S. Elyassami, Y. Gulzar, V. R. Balasaraswathi, T. Habuza, and S. Wani, "An improvised CNN model for fake image detection," *International Journal of Information Technology*, vol. 15, no. 1, pp. 5-15, 2023.
- [50] L. Zhang and M. Agrawala, "Adding conditional control to text-to-image diffusion models," *arXiv preprint arXiv:2302.05543*, 2023.
- [51] A. Bridy, "Why pirates (still) won't behave: Regulating P2P in the decade after Napster," *Rutgers LJ*, vol. 40, p. 565, 2008.