



# Stable Diffusion

出典: フリー百科事典『ウィキペディア (Wikipedia)』

**Stable Diffusion**（ステイブル・ディフュージョン）は、2022年に公開されたディープラーニング（深層学習）のtext-to-imageモデルである。主にテキスト入力に基づく画像生成（text-to-image）に使用されるが、他にもインペインティング（英: inpainting）、アウトペインティング（英: outpainting）、テキストプロンプトによって誘導される画像に基づく画像生成（image-to-image）にも使用される<sup>[3]</sup>。

Stable Diffusionは、ミュンヘン大学のCompVisグループが開発した潜在拡散モデル（英: latent diffusion model）であり、深層生成ニューラルネットワーク（英: deep generative neural network）の一種である<sup>[4]</sup>。このモデルは、EleutherAIとLAIONの支援を受け、Stability AI、CompVis LMU、Runwayの三者が共同で公開した<sup>[5][1][6]</sup>。2022年10月、Stability AIは、Lightspeed Venture PartnersとCoatue Managementが主導するラウンドで1億100万米ドルを調達した<sup>[7]</sup>。

Stable Diffusionのコードとウェイトは一般に公開されており<sup>[1]</sup>、少なくとも8GBのVRAMを持つGPUを搭載したほとんどの消費者向けハードウェアで実行可能である。そのため、DALL-EやMidjourneyなど、クラウドサービス経由でのみアクセス可能だった従来のプロプライエタリなtext-to-imageモデルとは一線を画すものであると評されている<sup>[8][9]</sup>。

Stable Diffusionは、教師データを確保するためにインターネット上から何十億もの画像をスクレイピングしているが、その画像の大半は著作権で保護されており同意を得ていないわけではないため、著作権法に違反するのかどうか、フェアユースに

## Stable Diffusion

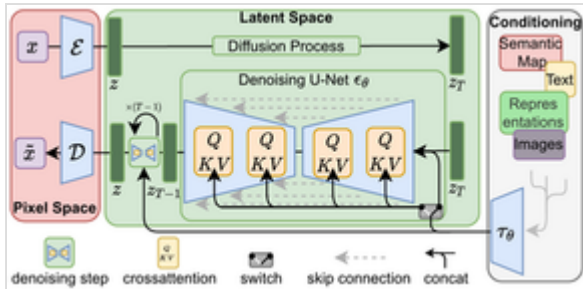


「乗馬する宇宙飛行士の写真<sup>[注 1]</sup>」と指定して出力された画像

開発元	CompVisグループ (ミュンヘン大学) Runway Stability AI <sup>[1]</sup>
初版	2022年8月22日
最新版	SDXL 1.0 (モデル) <sup>[2]</sup> / 2023年7月26日
リポジトリ	<a href="https://github.com/Stability-AI/stablediffusion">github.com/Stability-AI/stablediffusion</a> ( <a href="https://github.com/Stability-AI/stablediffusion">https://github.com/Stability-AI/stablediffusion</a> )
プログラミング言語	Python
対応OS	CUDAカーネルをサポートしている任意のOS
種別	text-to-imageモデル
ライセンス	Creative ML OpenRAIL-M
公式サイト	<a href="https://stability.ai">stability.ai</a> ( <a href="https://stability.ai">https://stability.ai</a> )

該当するかどうか、倫理的に適切であるのかといった法的・倫理的な論争が生じており、提供元のStability AIに対して複数の訴訟が提起されている<sup>[10]</sup>。

## 技術



Stable Diffusionのアーキテクチャ。

### アーキテクチャ

Stable Diffusion は、拡散モデル（英: diffusion model、DM）の一種である潜在拡散モデル（英: latent diffusion model、LDM）を使用している<sup>[1]</sup>。2015年に登場した拡散モデルは、学習元の画像に連続的に付与されたガウシアンノイズを除去する目的で学習されており、デノイズング・オートエンコーダ（英: denoising autoencoder）のシーケンス

として考えることが可能である。

Stable Diffusionは、変分オートエンコーダ（英: variational autoencoder、VAE）、U-Net、およびオプションのテキストエンコーダの3つによって構成されている<sup>[11]</sup>。VAEは、画像をピクセル空間からより低次元の潜在空間（英: latent space）へと圧縮し、画像のより基本的な意味を捕らえる<sup>[12]</sup>。圧縮された潜在表現（英: latent representation）には、順方向拡散過程（英: forward diffusion process）の際にガウシアンノイズが連続的に付与される<sup>[11]</sup>。ResNetで構成されるU-Netブロックは、潜在表現を得るために順方向拡散過程の出力をデノイズ（ノイズを除去）する。この過程を逆方向拡散過程（英: reverse diffusion process）という。最後に、VAEデコーダが潜在表現をピクセル空間へ逆変換し、最終的な画像を生成する<sup>[11]</sup>。

デノイズの過程は、文字列・画像・その他のモダリティによって柔軟に条件づけることが可能である。エンコードされた条件づけのためのデータは、cross-attention機構によってDenoising U-Netに与えられる<sup>[11]</sup>。テキストによる条件づけの場合、学習済みの固定的なCLIP ViT-L/14テキストエンコーダを使用して、テキストプロンプトが埋め込み空間へ変換される<sup>[1]</sup>。研究者らは、LDMの利点として、学習と生成のための計算効率の向上を挙げている<sup>[13][14]</sup>。

### 学習データ

Stable Diffusionは、LAION-5Bから取得した画像とキャプションのペアから学習している。これは、ウェブ上から収集されたコモン・クロールデータから派生した一般に利用可能なデータセットである。50億枚の画像とテキストのペアが言語に基づいて分類されており、解像度、ウォーターマークを含んでいる可能性の尤度、および「美学スコア」（英: aesthetic score）の予測に基づいて別々のデータセットにフィルタリングされている<sup>[15]</sup>。このデータセットは、Stability AIから資金提供を受けているドイツの非営利団体LAIONによって作成されたものである<sup>[15][16]</sup>。Stable Diffusionモデルは、LAION-5Bの3つのサブセット（laion2B-en、laion-high-resolution、laion-aesthetics v2 5+）から学習している<sup>[15]</sup>。このモデルの学習データを第三者が分析したところ、使用したオリジナルの広いデータセットから抽出した1200万枚の画像のうち、

サンプルサイズの約47%が100の異なるドメインから取得された画像であり、Pinterestがサブセットの8.5%を占め、WordPress、Blogspot、Flickr、DeviantArt、ウィキメディア・コモンズなどのウェブサイトがそれに続いていることが確認されたという<sup>[17][15]</sup>。

## 学習プロセス

このモデルは、最初に「laion2B-en」と「laion-high-resolution」から学習し、最後の数回に「LAION-Aesthetics v2 5+」から学習している。「LAION-Aesthetics v2 5+」は、LAION-Aesthetics Predictor V2が、人間に対してどの程度（その画像が）好きかを尋ねたときに、平均して10点中5点以上をつけると予測した6億枚のキャプション付き画像のサブセットである<sup>[18][15][19]</sup>。「LAION-Aesthetics v2 5+」サブセットは、低解像度の画像、およびLAION-5B-WatermarkDetectionが80%以上の確率でウォーターマークを検出した画像が除外されている<sup>[15]</sup>。最後に行われた学習では、Classifier-Free Diffusion Guidanceを改善するために、さらにテキストの条件づけ（キャプション）が10%削除された<sup>[20]</sup>。

このモデルの学習には、AWS上の256個のNVIDIA A100が使用され、60万米ドルの費用と合計15万のGPU時間が費やされた<sup>[21][22][23]</sup>。

## 制限

Stable Diffusionには、特定の条件下における劣化や不正確さなどの問題がある。このモデルの初期バージョンは、512×512の解像度の画像からなるデータセットで訓練されたため、ユーザーの要求がその期待される512×512の解像度から逸脱すると、生成される画像の品質が著しく低下する<sup>[24]</sup>。バージョン2.0のアップデートでは、768×768の解像度で画像をネイティブに生成する機能が導入された<sup>[25]</sup>。もう一つの課題は、LAIONデータセットにおける四肢のデータ品質が低いため、人間の手足の生成に支障が見られることである<sup>[26]</sup>。データセットには代表的な特徴がないため、人間の手足や顔を理解するための学習が不十分となっており、そのような画像の生成を指示すると、モデルが混乱してしまうのである<sup>[27]</sup>。人間の手足だけでなく、動物の四肢の生成も困難であることが確認されており、馬の画像を生成しようとする25%の確率で失敗するという<sup>[28]</sup>。

また、個人開発者のアクセシビリティも問題となっている。アニメキャラクターの生成（waifu diffusion）<sup>[29]</sup>など、データセットに含まれていない新しいユースケースに対してモデルをカスタマイズするには、新たな学習データと訓練が必要である。しかし、この微調整プロセスは、新しいデータの品質に影響を受けやすい。低解像度の画像や、元の学習データと異なる解像度の画像は、新しいタスクを学習できないだけでなく、モデル全体のパフォーマンスを低下させる可能性がある。また、高品質な画像でモデルを追加学習させたとしても、一般的な消費者向けのコンピュータでモデルを動かすことは困難である。たとえば、waifu-diffusionの学習プロセスには最低でも30GBのVRAMが必要であるが<sup>[30]</sup>、一般的な消費者向けのGPU、たとえば、NVIDIAのGeForce 30シリーズのVRAM容量（約12GB）<sup>[31]</sup>を優に超えており、一般人が利用可能な通常のリソースを超えている。

Stable Diffusionの開発者は、（このモデルは）主に英語の説明文が付与されている画像で学習したため、アルゴリズムバイアスが生じる可能性があることを認めている<sup>[22]</sup>。開発者は、他のコミュニティや文化圏からのデータが不足していると指摘しており、その結果、生成される画像

は、西洋の視点に立っており、社会的偏見を助長させるものとなっているという。また、他の言語で書かれたプロンプトよりも、英語で書かれたプロンプトの方がより正確な画像が生成され、西洋（または白人）の文化がデフォルトの表現となることが多いという<sup>[22]</sup>。

## 機能

Stable Diffusionモデルは、出力に含まれる要素、あるいは含まれない要素を記述するテキストプロンプトを入力することによって、ゼロから新しい画像を生成する機能を備えている<sup>[1]</sup>。また、既存の画像を与えることも可能であり、diffusion-denoising機構によって、テキストプロンプトに記述された新しい要素を取り込んで再描画することができる（たとえば、guided image synthesis<sup>[32]</sup>など）<sup>[1]</sup>。さらに、インペインティング（英: inpainting）とアウトペインティング（英: outpainting）によって既存の画像を部分的に変更することも可能であり、多くの異なるオープンソースの実装が存在する<sup>[33]</sup>。

Stable Diffusionは、10GB以上のVRAMでの動作が推奨されているが、VRAMが少ない場合は、デフォルトのfloat32ではなくfloat16の精度でウェイトを読み込むことができる。ただし、モデルの性能とVRAM使用量はトレードオフの関係にある<sup>[24]</sup>。

## テキストに基づく画像生成（text-to-image）

テキストから画像を生成するサンプリングスクリプトは、「txt2img」（text-to-imageやt2iとも言う）として知られている。これには、テキストプロンプトに加え、サンプリング方式、出力画像の解像度、シード値などのさまざまなオプションのパラメータが使用される。このスクリプトは、プロンプトに対するモデルの解釈に基づいて画像ファイルを出力する<sup>[1]</sup>。生成された画像には、Stable Diffusionによって生成された画像であることをユーザーが識別できるように、目に見えない電子透かしが付与されるが<sup>[1]</sup>、画像のサイズを変更したり回転したりすると効力を失う<sup>[34]</sup>。

各txt2imgの生成には、出力画像に影響を与える特定のシード値が含まれる。ユーザーは、さまざまな出力を探るためにシード値をランダムにするか、以前に生成された画像と同じ出力を得るために同じシード値を使用するかを選択できる<sup>[24]</sup>。また、サンプラーの推論ステップ数を調整することも可能である。値を大きくすると生成に時間がかかり、値を小さくすると視覚的な瑕疵が生じる可能性が生じる<sup>[24]</sup>。CFGスケール（英: classifier-free guidance scale）の値を調整することで、出力画像がプロンプトに対してどれだけ忠実であるかを調整することも可能である<sup>[20]</sup>。実験的なユースケースには低いスケール値を使用したり、具体的な出力を得ることが目的のユースケースには高い値を使用したりできる<sup>[24]</sup>。



画像生成におけるネガティブプロンプトの影響。

- 上段: ネガティブプロンプトなし
- 中段: "green trees"
- 下段: "round stones, round rocks"



その他のtxt2imgの機能は、Stable Diffusionのフロントエンド実装によって提供される。これには、テキストプロンプトの特定の部分に与えられる重みを調整する機能が含まれる。強調マークは、キーワードを括弧で囲むことによって、そのキーワードを強調したり、抑制したりできる<sup>[35]</sup>。プロンプトの部分的な重みを調整する別の方法として、「ネガティブプロンプト」（英: negative prompt）と呼ばれる機能がある。ネガティブプロンプトは、いくつかのフロントエンド実装に含まれる機能であり、画像生成時にモデルが避けるべきプロンプトをユーザーが指定できる。この機能を使用することで、ユーザーによって指定されたポジティブプロンプト（英: positive prompt）によって（あるいはモデルが元々どのように訓練されたかによって）出力画像内に存在することになる望ましくない画像の特徴を指定・排除できる<sup>[33]</sup>。

## 画像に基づく画像生成（image-to-image）

Stable Diffusionには、「img2img」（image-to-imageやi2iとも言う）として知られているもう一つのサンプリングスクリプトがある。これには、テキストプロンプト、既存の画像のパス、0.0から1.0までの強度値が使用される。このスクリプトは、テキストプロンプトで指定された要素を含む、オリジナルの画像に基づいた新しい画像を出力する。強度の値は、出力画像に加えられるノイズの量を表している。強度の値が高いほど画像内に新しい要素が増えるが、指定されたプロンプトと意味的に一致しない画像が生成される可能性がある<sup>[1]</sup>。

img2imgは、元画像にノイズを加えることができるため、画像データの視覚的特徴を変化させて匿名化する、データ匿名化やデータ拡張に有効である可能性がある<sup>[36]</sup>。また、画像の解像度を上げ、より詳細な情報を画像に加えるアップスケーリングにも利用できる可能性がある<sup>[36]</sup>。さらに、Stable Diffusionは、画像圧縮のツールとしても実験されている。

JPEGやWebPと比較すると、Stable Diffusionの画像圧縮で使用されている近年の方法は、小さな文字や顔を保存するのに限界がある<sup>[37]</sup>。

img2imgを使用した画像修正のユースケースは、Stable Diffusionモデルの多くのフロントエンド実装によって提供されている。インペインティング（英: inpainting）は、ユーザーが提供するマスキレイヤーで区切られた既存画像の一部を選択的に変更するものであり、指定されたプロンプトに基づいて新たに生成されたコンテンツでマスク空間を埋めるものである<sup>[33]</sup>。Stable Diffusion 2.0のリリースに伴い、インペインティング用途に特化して微調整された専用モデルがStability AIによって作成された<sup>[25]</sup>。また、インペインティングとは逆のを行うアウトペイ



画像生成におけるimg2imgの影響。

- 上段: Stable Diffusion 1.5で作成したオリジナル画像
- 下段: Stable Diffusion XLで作成した修正画像

ンティング（英: outpainting）という機能もあり、画像を元の解像度以上に拡張し、指定されたプロンプトに基づいて生成されたコンテンツで拡張された空白の部分を埋めることもできる<sup>[33]</sup>。

2022年11月24日にリリースされたStable Diffusion 2.0では、「depth2img」と名付けられたdepth-guidedモデルが導入された。このモデルは、提供された入力画像の深度を推測し、テキストプロンプトと深度情報の両方に基づいて新しい画像を生成するものであり、入力画像の一貫性と深度を出力画像に維持させることができる<sup>[25]</sup>。

## 用途と論争、訴訟

---

### 著作権

Stable Diffusionは、生成された画像に関する権利を主張しておらず、画像の内容が違法であったり、個人に有害でない限り、モデルから生成されたあらゆる画像の使用権をユーザーに自由に与えている。Stable Diffusionと他の生成モデルは、著作権のある画像から権利者の同意なしに学習を行っているため、画像の使用に関してユーザーに提供される自由について、所有権の倫理をめぐる論争が引き起こされている<sup>[38]</sup>。

ジャーナリストのナオミ・クラインは生成AI技術と関連企業を厳しく批判しており、「人類史上最大の盗み」「囲い込みと横領を行う機械」としており、「個人的な生および集合的な知的・芸術的遺産の双方を食いにし、私有化する」としている。つまり、AI企業はインターネット上に公開された人類の知識全体を、製品の中に囲い込んで専売しており、人々の生涯をかけた労働が同意や許可を得ずに、訓練データに使われているというのである。クラインは以下のように述べた<sup>[39]</sup>。

こんなことが合法であるはずがない。AIモデルを訓練するのに使われたことが判明している著作権保護の素材（この『ガーディアン』紙も含まれる）に関しては何件もの訴訟が申請されており、明白に違法な行為が行われたという主張がそこでなされるだろう。例えば、営利企業が生身の芸術家たちの絵画や写真をStable DiffusionやDALL-E 2のようなプログラムに注ぎ込み、それを使ってまさにその芸術家たちの作品のドッペルゲンガー版を作成できるようにするなどということが、いったいどうして許されるのだろうか？その利益が芸術家たち自身にだけは還元されないというのに？<sup>[39]</sup>

### 日本における著作権問題

Stable Diffusionの登場に関連して、日本においては2018年に成立した改正著作権法が問題視されている。この30条4項では人工知能の訓練データに文章や画像を利用する場合に、営利・非営利を問わず著作物を利用できることを定めており、先進国で最も緩い法規制の枠組みになっている。「著作権者の利益を不当に害する場合」は利用できないとしているが、その具体的にどのようなケースに該当するかは不明瞭である。法改正に当たっては一部の弁護士や大学教授らで構成されたワーキングチームが主導したが、そのリスクは十分に十分に説明されなかった。

『読売新聞』の取材によれば、権利者団体から権利侵害を前提としたビジネスモデルへの懸念が示されたが、ワーキングチームはリスクは「ゼロか軽微」と説明した<sup>[40]</sup>。

この著作権法の規定に多くの漫画家、イラストレーター、音楽家、俳優、芸能人、出版社が懸念を示しており、俳優や音楽家らで結成された日本芸能従事者協会はアンケートを実施し、94.1%がAIによる権利侵害に懸念を示し、「創作者の尊厳が軽んじられないような法制度」を求める意見などが表明された。日本音楽著作権協会も、「営利目的で行われる情報分析サービスまで権利制限の対象とすることは不当」と主張した。主要な出版社で作る日本書籍出版協会、学術論文の著作権を管理する学術著作権協会も改正当時は生成AIを前提とした説明が行われなかったと回答している<sup>[40]</sup>。

## データ倫理

### ウェブ上の画像のスクレイピング

『ハーバード・ビジネス・レビュー』は、ネット上に公開された無数の著作物を訓練データとして利用するStable Diffusionをはじめとする現行の画像生成AIについて、知的財産権を侵害している可能性があるとして、潜在的な法的リスクを指摘している。そして、訓練データが権利者のライセンスを受けていることが望ましく、生成されたコンテンツの出所を示す手段が必要であるとしている<sup>[41]</sup>。

開発企業はアーティストの権利を尊重すると称して、著作物を訓練データから除外するためのオプトアウトの申請を受け付けている。しかし『ハーバード・ビジネス・レビュー』は、このオプトアウトは知的財産権を保護する責任を権利者に押し付けているとして、オプトアウトよりもオプトインが適切であると批判している<sup>[41]</sup>。

芸術家らは、Stable Diffusionのような画像生成ソフトが普及すると、写真家・モデル・撮影監督・俳優といった人間の芸術家が、AIを使用した競合他社に対して徐々に商業的価値を失っていくのではないかと懸念を表明している<sup>[42]</sup>。『名探偵ピカチュウ』のクリーチャー・デザインに携わったR・J・パーマーは、Stable Diffusionが多くの現役クリエイターの作品を学習データに取り込んでおり、潜在的に彼らの仕事を奪いかねず権利侵害にあたると非難している<sup>[43]</sup>。DALL・E 2など他の画像生成サービスは、現代の風景写真家の作品を学習データから排除するなど、クリエイターとの協調関係に配慮を見せている<sup>[44]</sup>。

2023年1月13日、オレゴン州在住の漫画家サラ・アンダーソンら3人の原告はAIアートの違法性を訴え、Stability AIとStable Diffusionを利用した画像生成サービスを展開するMidjourney、DeviantARTの3社に対して訴訟を提起した。画像生成AIは、教師データを確保するためにインターネット上から何十億もの画像をスクレイピングしているが、その画像の大半は著作権で保護されており、同意も得ていないため、画像生成AIには倫理的・法的問題が生じている。アンダーソンは『ワシントン・ポスト』紙の取材に対し、「AIは私の作品を盗んだ」と答えた一方、Stability AI、Midjourney、DeviantARTの3社はいずれも取材を拒否した<sup>[10][45]</sup>。

計算機科学者でプリンストン大学教授のアーヴィンド・ナラヤナンは、「画像生成AIを開発する企業は同意や補償なしに訓練用画像を収集するなど、アーティストに敵対するような方法で開発・デプロイを行っている」とした上で、「特定のアーティストの画風に寄せた画像生成ツールを許容することは、アーティストの労働や視覚的な独自性を明確に流用しているケースのように思える」と述べ、「開発者は、アーティストを訓練用の素材ではなく、パートナーや利害関係者として扱うこともできたはずだ」と画像生成AIの現状に異を唱えた上で、「この現状が必然だったと主張する人物は、企業が責任ある技術開発をできなかったことの言い訳をしているに過ぎない」と結論付けた<sup>[10]</sup>。

2023年4月3日、東京大学は理事・副学長の太田邦史の署名付き文章で、全学生・教員向けにMidjourney、Stable Diffusion等の生成AIの利用に関する注意喚起を行った。画像生成AIが、インターネット上のコンテンツを取り込んで学習し、画像を生成しており、これらの元データの作成者が知らないうちに著作権を侵害されたとして、問題提起を行っている現状を指摘。将来的に画像生成AIが生み出したコンテンツが訴訟の対象になる可能性に言及した<sup>[46]</sup>。

Stable Diffusionの訓練データの収集方法については「データ・ロンダリング」がなされているとの批判を受けている。つまり、提供会社が直接データを収集・利用するのではなく、非営利の研究機関を間に挟むことで、「研究目的」として著作物を収集することができ、著作物の使用に対する対価を払うことなく、商用利用を行っているとされる<sup>[47]</sup>。

Stable Diffusionなどの画像生成AIが抱える倫理的・法的問題に対して、Adobeは独自の画像生成AIのFireflyをリリースした。これはパブリックドメイン、オープンライセンスおよびAdobe Stockの画像を訓練データとして利用しており、知的財産権を巡る問題を大幅に軽減している。Nvidiaも独自の生成AIであるNvidia Picassoをリリースしたが、著作権者に正当な使用料を支払っているとしている<sup>[48][49][50]</sup>。

電子フロンティア財団の弁護士キット・ウォルシュは、ウェブから画像をスクレイピングする行為について、ある行為が著作権侵害になる可能性があっても、その行為が非侵害的使用を行う上で不可欠なステップである場合、その行為自体はフェアユースに該当する傾向にあることを指摘している<sup>[51]</sup>。そのため、中間的な使用と分析のための使用の双方において、スクレイピングが著作権法に違反するとは考えにくいという<sup>[51]</sup>。また、Stable Diffusionは訓練画像を圧縮（保存）していると批判されているが、Stable Diffusionのモデルは50億枚以上の画像に4GB分の観察情報（ピクセル配置とテキスト注釈の単語との相関に関する情報）を保存しているにすぎないため、分析した画像1枚につき1バイト以下の情報しか含まれておらず、数学的にはStable Diffusionは訓練画像の複製を保存しているわけではないと述べている<sup>[51]</sup>。そして、「15000行の大作を10行にまとめたものは二次的著作物ではなく、書籍を説明するための要約も大半が二次的著作物ない」ことや訓練セットに含まれる各画像から抽出された著作権で保護されうる表現の量は、「侵害と認定するにはあまりにも小さい」（デ・ミニミス）とみなされる可能性があることなどから、Stable Diffusionは侵害的な二次的著作物を生成・保存していないと指摘している<sup>[51]</sup>。

## 出力画像の生成

英国のストックフォト企業であるGetty Imagesも画像生成AIが膨大な量の画像を盗用し、知的財産権を侵害しているとしてStability AIへの損害賠償請求訴訟を提起した。画像生成AIはアーティストの知的財産である作品をベースに画像を生成するが、その作品の作者には補償されないし、著作権者の許可も得ていないことが問題視されている<sup>[52]</sup>。

電子フロンティア財団の弁護士キット・ウォルシュは、プライバシー関連の研究<sup>[53]</sup>から「拡散モデルが訓練データ中の画像を再現できるような情報を保存する可能性は、当該画像が訓練中に何度も複製された場合にはわずかながらある」としながらも、「訓練データ中の画像が出力される確率は、たとえその出力を引き出すために特別にしつらえたプロンプトをもってしても、文字通り100万分の1以下の確率である」と述べている<sup>[51]</sup>。そのため、著作権を主張できる



権利者は訴訟の原告らが含まれるとは考えにくく、また、仮に侵害の責任を負うとすれば、それは既存のクリエイティブ・ツールと同じようにツールの製造者や提供者ではなく、ユーザーになるだろうという見解を示している<sup>[51]</sup>。

## ディープフェイク等の問題

Stable Diffusionは生成AIに基づく他の商業製品と比べて、ユーザーが生成できるコンテンツの種類（暴力や性描写など）に寛容であることが特徴的である<sup>[54]</sup>。Stability AIのCEOであるEmad Mostaqueは、このモデルが悪用されるのではないかと懸念に対して、「この技術の用途が倫理的・道徳的・合法的であるかどうかはユーザーに責任がある」と説明し<sup>[9]</sup>、潜在的にはマイナスの影響もあるが、Stable Diffusionの機能を一般人の手に委ねることは、この技術が純粋な利益を提供することにつながると述べている<sup>[9]</sup>。さらに、Stable Diffusionをオープンにした背景には、クローズドな画像合成AIシステムしか開発してこなかった企業による、このような技術に対する支配と統制を終わらせる意図があると述べている<sup>[9][54]</sup>。この理念は、Stable Diffusionが学習済みのウェイトとともにオープンソースとして公開されているために、ユーザーが生成するコンテンツにStability AIが課す成約を容易に回避できるという事実にも反映されている<sup>[38][55][1]</sup>。

また、Discordサーバーを通じて使えるStable Diffusionのホストバージョンは、他の画像生成サービスが禁止している政治的・宗教的にセンシティブな多くのテーマに関する画像<sup>[注2]</sup>の生成を野放しにしているとTechCrunchは非難したが、モスタークは「違法なコンテンツは別として、フィルタリングは最小限に抑えられており、どのようにStable Diffusionを使用するかはユーザー次第」とコメントするに留まった<sup>[43]</sup>。

OpenAIが提供する画像生成AIはディープフェイクに悪用されないように、特定のリクエストの入力をブロックするような対策がされている一方で、Stable Diffusionは名目上ポルノ生成を抑止するような方策が取られているものの、実際にはそれを回避した性的なコンテンツの生成が可能である。カリフォルニア大学バークレー校教授のハニー・ファリドによれば、インターネット上に公開された何億もの画像データを学習する過程で、女性を性的に扱うバイアスが反映されるため、モデルの出力を100%制御することは困難であるとされる<sup>[56][57]</sup>。

Kickstarterでは、Stable Diffusionによるポルノの作成を目指す出資プロジェクトが、集めた資金で既存の画像を学習データとして利用することが、多くのアーティストから批判されていることを受け、AIアートに関する条項を変更し、プロジェクトをブロックした<sup>[58]</sup>。

生成AIの負の側面として、詐欺や思考誘導、世論操作を行う目的で、捏造した顔画像や映像、フェイク音声、フェイク文章といったフェイクメディアを流通させる可能性が懸念されている。すでに国内においてもStable Diffusionを使って捏造された災害画像が拡散する事例もあり、AIを悪用した詐欺や詐称、名誉毀損は現実の脅威となっている。このため国内外で対策のための研究が進められている<sup>[59]</sup>。

## その他の批判

ナラヤナンは、メディアによる画像生成AIの宣伝に関しても厳しく批判しており、それを過度に擬人化するなど誤解を招くような印象を蔓延させ、誇大広告に加担していると指摘している。また、ナラヤナンは「AI報道で気をつけるべき18の落とし穴」として、AIが人間と同じよ

うに学習すると暗示して人間の知能とAIを比較したりすることや、AIを電気の発明や産業革命のような歴史的な大転換に安易になぞらえることを批判している<sup>[60]</sup>。

## 脚注

---

### 注釈

1. ^ 原文: "a photograph of an astronaut riding a horse"
2. ^ ウクライナでの戦争、中国の台湾侵攻、預言者ムハンマドの姿など<sup>[43]</sup>

### 出典

1. ^ *abcdefghijklmnopqrstuvwxyz* "Stable Diffusion Repository on GitHub (<https://github.com/CompVis/stable-diffusion>)". CompVis - Machine Vision and Learning Research Group, LMU Munich (2022年9月17日). 2022年9月17日閲覧。
2. ^ "Announcing SDXL 1.0 (<https://web.archive.org/web/20230726215239/https://stability.ai/blog/stable-diffusion-sdxl-1-announcement>)". *stability.ai*. 2023年7月26日時点のオリジナル (<https://stability.ai/blog/stable-diffusion-sdxl-1-announcement>)よりアーカイブ。2023年7月26日閲覧。
3. ^ "Diffuse The Rest - a Hugging Face Space by huggingface (<https://web.archive.org/web/20220905141431/https://huggingface.co/spaces/huggingface/diffuse-the-rest>)". *huggingface.co*. 2022年9月5日時点のオリジナル (<https://huggingface.co/spaces/huggingface/diffuse-the-rest>)よりアーカイブ。2022年9月5日閲覧。
4. ^ Rombach; Blattmann; Lorenz; Esser; Ommer (June 2022). *High-Resolution Image Synthesis with Latent Diffusion Models* ([https://openaccess.thecvf.com/content/CVPR2022/papers/Rombach\\_High-Resolution\\_Image\\_Synthesis\\_With\\_Latent\\_Diffusion\\_Models\\_CVPR\\_2022\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2022/papers/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.pdf)) (PDF). International Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA. pp. 10684–10695. arXiv:2112.10752 (<https://arxiv.org/abs/2112.10752>).
5. ^ "Stable Diffusion Launch Announcement (<https://web.archive.org/web/20220905105009/https://stability.ai/blog/stable-diffusion-announcement>)". *Stability.Ai*. 2022年9月5日時点のオリジナル (<https://stability.ai/blog/stable-diffusion-announcement>)よりアーカイブ。2022年9月6日閲覧。
6. ^ "Revolutionizing image generation by AI: Turning text into images (<https://www.lmu.de/en/newsroom/news-overview/news/revolutionizing-image-generation-by-ai-turning-text-into-images.html>)". *LMU Munich*. 2022年9月17日閲覧。
7. ^ Wiggers, Kyle (2022年10月17日). "Stability AI, the startup behind Stable Diffusion, raises \$101M (<https://techcrunch.com/2022/10/17/stability-ai-the-startup-behind-stable-diffusion-raises-101m/>)" (英語). *Techcrunch*. 2022年10月17日閲覧。
8. ^ "The new killer app: Creating AI art will absolutely crush your PC (<https://web.archive.org/web/20220831065139/https://www.pcworld.com/article/916785/creating-ai-art-local-pc-stable-diffusion.html>)". *PCWorld*. 2022年8月31日時点のオリジナル (<https://www.pcworld.com/article/916785/creating-ai-art-local-pc-stable-diffusion.html>)よりアーカイブ。2022年8月31日閲覧。
9. ^ *abcd* Vincent, James (2022年9月15日). "Anyone can use this AI art generator — that's the risk (<https://www.theverge.com/2022/9/15/23340673/ai-image-generation-stable-diffusion-explained-ethics-copyright-data>)". *The Verge*. 2022年11月30日閲覧。

10. ^ a b c “Artists are alarmed by AI — and they’re fighting back (<https://www.washingtonpost.com/comics/2023/02/14/ai-in-illustration/>)” (英語). *The Washington Post*. 2023年2月18日閲覧。
11. ^ a b c d Alammar, Jay. “The Illustrated Stable Diffusion (<https://jalammar.github.io/illustrated-stable-diffusion/>)”. *jalammar.github.io*. 2022年10月31日閲覧。
12. ^ “High-Resolution Image Synthesis with Latent Diffusion Models (<https://ommer-lab.com/research/latent-diffusion-models/>)” (英語). *Machine Vision & Learning Group*. 2022年11月4日閲覧。
13. ^ “Stable Diffusion launch announcement (<https://stability.ai/blog/stable-diffusion-announcement/>)” (英語). *Stability.AI*. 2022年11月2日閲覧。
14. ^ Rombach; Blattmann; Lorenz; Esser; Ommer (June 2022). *High-Resolution Image Synthesis with Latent Diffusion Models* (PDF). International Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA. pp. 10684–10695. arXiv:2112.10752.
15. ^ a b c d e f Baio, Andy (2022年8月30日). “Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator (<https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/>)” (英語). *Waxy.org*. 2022年11月2日閲覧。
16. ^ “This artist is dominating AI-generated art. And he's not happy about it. (<https://www.technologyreview.com/2022/09/16/1059598/his-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>)” (英語). *MIT Technology Review*. 2022年11月2日閲覧。
17. ^ Ivanovs, Alex (2022年9月8日). “Stable Diffusion: Tutorials, Resources, and Tools (<https://stackdiary.com/stable-diffusion-resources/>)” (英語). *Stack Diary*. 2022年11月2日閲覧。
18. ^ Schuhmann, Christoph (2022-11-02), *CLIP+MLP Aesthetic Score Predictor* (<https://github.com/christophschuhmann/improved-aesthetic-predictor>) 2022年11月2日閲覧。
19. ^ “LAION-Aesthetics | LAION (<https://web.archive.org/web/20220826121216/https://laion.ai/blog/laion-aesthetics/>)” (英語). *laion.ai*. 2022年8月26日時点のオリジナル (<https://laion.ai/blog/laion-aesthetics>)よりアーカイブ。2022年9月2日閲覧。
20. ^ a b Ho, Jonathan; Salimans, Tim (25 July 2022). “Classifier-Free Diffusion Guidance”. arXiv:2207.12598 (<https://arxiv.org/abs/2207.12598>) [cs.LG (<https://arxiv.org/archive/cs.LG>)].
21. ^ Mostaque, Emad (2022年8月28日). “Cost of construction (<https://web.archive.org/web/20220906155426/https://twitter.com/EMostaque/status/1563870674111832066>)” (英語). *Twitter*. 2022年9月6日時点のオリジナル (<https://twitter.com/emostaque/status/1563870674111832066>)よりアーカイブ。2022年9月6日閲覧。
22. ^ a b c “CompVis/stable-diffusion-v1-4 · Hugging Face (<https://huggingface.co/CompVis/stable-diffusion-v1-4>)”. *huggingface.co*. 2022年11月2日閲覧。
23. ^ Wiggers, Kyle (2022年8月12日). “A startup wants to democratize the tech behind DALL-E 2, consequences be damned (<https://techcrunch.com/2022/08/12/a-startup-wants-to-democratize-the-tech-behind-dall-e-2-consequences-be-damned/>)” (英語). *TechCrunch*. 2022年11月2日閲覧。
24. ^ a b c d e “Stable Diffusion with 🖌️ Diffusers ([https://huggingface.co/blog/stable\\_diffusion](https://huggingface.co/blog/stable_diffusion))”. *huggingface.co*. 2022年10月31日閲覧。
25. ^ a b c “Stable Diffusion 2.0 Release (<https://stability.ai/blog/stable-diffusion-v2-release>)”. *stability.ai*. 2022年11月24日閲覧。
26. ^ “LAION (<https://laion.ai/>)” (英語). *laion.ai*. 2022年10月31日閲覧。

27. ^ “Generating images with Stable Diffusion (<https://blog.paperspace.com/generating-images-with-stable-diffusion/>)” (英語). *Paperspace Blog* (2022年8月24日). 2022年10月31日閲覧。
28. ^ François Chollet. “(If you were wondering how often Stable Diffusion will give you a horse with more than 4 legs (or sometimes less) when you ask it for a photo of a horse: in my experience it's about 20-25% of the time.) (<https://twitter.com/fchollet/status/1573879858203340800>)” (英語). *Twitter*. 2022年10月31日閲覧。
29. ^ “hakurei/waifu-diffusion · Hugging Face (<https://huggingface.co/hakurei/waifu-diffusion>)”. *huggingface.co*. 2022年10月31日閲覧。
30. ^ Mercurio, Anthony (2022-10-31), *Waifu Diffusion* (<https://github.com/harubaru/waifu-diffusion/blob/6bf942eb6368ebf6bcbbbb24b6ba8197bda6582a0/docs/en/training/README.md>) 2022年10月31日閲覧。
31. ^ Smith, Ryan. “NVIDIA Quietly Launches GeForce RTX 3080 12GB: More VRAM, More Power, More Money (<https://www.anandtech.com/show/17204/nvidia-quietly-launches-geforce-rtx-3080-12gb-more-vram-more-power-more-money>)”. *www.anandtech.com*. 2022年10月31日閲覧。
32. ^ Meng, Chenlin; He, Yutong; Song, Yang; Song, Jiaming; Wu, Jiajun; Zhu, Jun-Yan; Ermon, Stefano (2 August 2021). “SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations”. *arXiv:2108.01073* (<https://arxiv.org/abs/2108.01073>) [cs.CV (<https://arxiv.org/archive/cs>. CV)].
33. ^ **a b c d** “Stable Diffusion web UI (<https://github.com/AUTOMATIC1111/stable-diffusion-webui-feature-showcase>)”. *GitHub* (2022年11月10日). 2022年11月30日閲覧。
34. ^ *invisible-watermark* (<https://github.com/ShieldMnt/invisible-watermark/blob/9802ce3e0c3a5ec43b41d503f156717f0c739584/README.md>), Shield Mountain, (2022-11-02) 2022年11月2日閲覧。
35. ^ “stable-diffusion-tools/emphasis at master · JohannesGaessler/stable-diffusion-tools (<https://github.com/JohannesGaessler/stable-diffusion-tools>)” (英語). *GitHub*. 2022年11月2日閲覧。
36. ^ **a b** Luzi, Lorenzo; Siahkoohi, Ali; Mayer, Paul M.; Casco-Rodriguez, Josue; Baraniuk, Richard (21 October 2022). “Boomerang: Local sampling on image manifolds using diffusion models”. *arXiv:2210.12100* (<https://arxiv.org/abs/2210.12100>) [cs.CV (<https://arxiv.org/archive/cs>. CV)].
37. ^ Bühlmann, Matthias (2022年9月28日). “Stable Diffusion Based Image Compression (<https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202>)” (英語). *Medium*. 2022年11月2日閲覧。
38. ^ **a b** Cai, Kenrick. “Startup Behind AI Image Generator Stable Diffusion Is In Talks To Raise At A Valuation Up To \$1 Billion (<https://www.forbes.com/sites/kenrickcai/2022/09/07/stability-ai-funding-round-1-billion-valuation-stable-diffusion-text-to-image/>)” (英語). *Forbes*. 2022年10月31日閲覧。
39. ^ **a b** ナオミ・クライン、中村峻太郎（訳）「「幻覚を見ている」のはAIの機械ではなく、その製作者たちだ」『世界』岩波書店 2023年7月号 p85~95
40. ^ **a b** 『読売新聞』 2023年5月16日号 朝刊 1-2面
41. ^ **a b** “Generative AI Has an Intellectual Property Problem (<https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>)”. *Harvard Business Review*. 2023年4月12日閲覧。
42. ^ Heikkilä, Melissa (2022年9月16日). “This artist is dominating AI-generated art. And he's not happy about it. (<https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>)”. *MIT Technology Review*. 2022年11月30日閲覧。

43. ^ [abc](https://gigazine.net/news/20220815-stable-diffusion-controversy-ai-infringes-artists/) “まるで人間のアーティストが描いたような画像を生成するAIが「アーティストの権利を侵害している」と批判される (<https://gigazine.net/news/20220815-stable-diffusion-controversy-ai-infringes-artists/>)”. GIGAZINE. OSA (2022年8月15日). 2022年8月28日閲覧。
44. ^ “風景写真家がAI画像ジェネレーター「Stable Diffusion」で非実在美景を生成 (<https://gigazine.net/news/20220819-beautiful-landscapes-created-by-stable-diffusion/>)”. GIGAZINE. OSA (2022年8月19日). 2022年8月28日閲覧。
45. ^ 株式会社インプレス (2023年1月17日). “米画家ら、画像生成AI「Stable Diffusion」と「Midjourney」を提訴 (<https://pc.watch.impress.co.jp/docs/news/1470744.html>)”. PC Watch. 2023年1月18日閲覧。
46. ^ “生成系AI(ChatGPT, BingAI, Bard, Midjourney, Stable Diffusion等)について (<https://utelecon.adm.u-tokyo.ac.jp/docs/20230403-generative-ai/>)”. 東京大学. 2023年4月5日閲覧。
47. ^ “生成AI (Generative AI) の倫理的・法的・社会的課題 (ELSI) 論点の概観: 2023年3月版 ([https://elsi.osaka-u.ac.jp/system/wp-content/uploads/2023/04/ELSI\\_NOT\\_E\\_26\\_2023\\_230410.pdf](https://elsi.osaka-u.ac.jp/system/wp-content/uploads/2023/04/ELSI_NOT_E_26_2023_230410.pdf))”. 大阪大学. 2023年4月17日閲覧。
48. ^ “Nvidia launches new cloud services and partnerships to train generative AI (<https://techcrunch.com/2023/03/21/nvidia-launches-new-cloud-services-and-partnerships-to-train-generative-ai/>)”. TechCrunch. 2023年6月26日閲覧。
49. ^ “Adobe aims to sell businesses on its Firefly generative AI tools (<https://www.axios.com/2023/06/08/adobe-firefly-generative-ai-tools-licenses>)”. Axios. 2023年6月19日閲覧。
50. ^ “アドビ、生成AI「Firefly」でクリエイティブツールを強化--「Illustrator」などに搭載 (<https://japan.zdnet.com/article/35205161/>)”. ZDNet. 2023年6月26日閲覧。
51. ^ [abcdef](https://www.eff.org/deeplinks/2023/04/how-we-think-about-copyright-and-ai-art-0) Walsh, Kit (2023年4月3日). “How We Think About Copyright and AI Art (<https://www.eff.org/deeplinks/2023/04/how-we-think-about-copyright-and-ai-art-0>)” (英語). Electronic Frontier Foundation. 2023年5月14日閲覧。
52. ^ <https://www.gizmodo.jp/2023/02/ai-image-gen-getty-lawsuit.html>
53. ^ Carlini, Nicholas; Hayes, Jamie; Nasr, Milad; Jagielski, Matthew; Sehwag, Vikash; Tramèr, Florian; Balle, Borja; Ippolito, Daphne et al. (2023-01-30). “Extracting Training Data from Diffusion Models” (<http://arxiv.org/abs/2301.13188>). *arXiv:2301.13188 [cs]*.
54. ^ [ab](https://www.businessinsider.jp/post-258369) Ryo Shimizu (2022年8月26日). “Midjourneyを超えた？ 無料の作画AI「#StableDiffusion」が「AIを民主化した」と断言できる理由 (<https://www.businessinsider.jp/post-258369>)”. Business Insider Japan. 2022年11月30日閲覧。
55. ^ “Stable Diffusion Public Release (<https://web.archive.org/web/20220830210535/http://stability.ai/blog/stable-diffusion-public-release>)”. Stability.Ai. 2022年8月30日時点のオリジナル (<https://stability.ai/blog/stable-diffusion-public-release>)よりアーカイブ。 2022年8月31日閲覧。
56. ^ “AI porn is easy to make now. For women, that’s a nightmare. (<https://www.washingtonpost.com/technology/2023/02/13/ai-porn-deepfakes-women-consent/>)”. The Washington Post. 2023年2月20日閲覧。
57. ^ “Ready or not, mass video deepfakes are coming” (<https://web.archive.org/web/20220831115010/https://www.washingtonpost.com/technology/2022/08/30/deep-fake-video-on-agg/>). The Washington Post. (2022年8月30日). オリジナル (<https://www.washingtonpost.com/technology/2022/08/30/deep-fake-video-on-agg/>)の2022年8月31日時点におけるアーカイブ。 2022年8月31日閲覧。



58. <sup>^</sup> “AIでポルノ画像を作成するプロジェクト「Unstable Diffusion」がKickstarterでの出資募集をブロックされる (<https://gigazine.net/news/20221226-kickstarter-ai-porn-unstable-diffusion/>)”. *GIGAZINE*. 2023年2月8日閲覧。
59. <sup>^</sup> 越前巧、馬場口登、笹原和俊「インフォデミック時代におけるフェイクメディア克服の最前線」『人工知能』人工知能学界 2023年3月号 pp.189-196
60. <sup>^</sup> “インチキAIに騙されないために (<https://wirelesswire.jp/2023/01/83851/>)”. 2023年3月24日閲覧。

## 関連項目

---

- [コンピュータアート](#)
- [ジェネレーティブアート](#)
- [人工知能アート](#)
- [Midjourney](#)
- [DALL-E](#)
- [Imagen \(Google Brain\)](#)
- [Trinart](#)
- [NovelAI](#)

## 外部リンク

---



ウィキメディア・コモンズには、**[Stable Diffusion \(\[https://commons.wikimedia.org/wiki/Category:Stable\\\_Diffusion?uselang=ja\]\(https://commons.wikimedia.org/wiki/Category:Stable\_Diffusion?uselang=ja\)\)](https://commons.wikimedia.org/wiki/Category:Stable_Diffusion?uselang=ja)**に関するカテゴリがあります。

- [Stable Diffusion - a Hugging Face Space by stabilityai \(<https://huggingface.co/spaces/stabilityai/stable-diffusion>\)](https://huggingface.co/spaces/stabilityai/stable-diffusion)
- [DreamStudio \(<https://beta.dreamstudio.ai/>\)](https://beta.dreamstudio.ai/)
- 清水遼平：「DiffusionによるText2Imageの系譜と生成画像が動き出すまで」 (<https://speakeck.com/nhamanasu/diffusionniyorutext2imagenoxi-pu-tosheng-cheng-hua-xiang-gadong-kichu-sumade>)

---

「[https://ja.wikipedia.org/w/index.php?title=Stable\\_Diffusion&oldid=98710366](https://ja.wikipedia.org/w/index.php?title=Stable_Diffusion&oldid=98710366)」から取得

■