
東北芸術工科大学 紀要

BULLETIN OF TOHOKU UNIVERSITY OF ART AND DESIGN

第31号 2024年3月

基本図形の作成を生成AIで作る試み、その利用と応用について
Attempts to Create Basic Shapes with Generative AI: Their Use and Application

村山 秀明 | MURAYAMA Hideaki

【論文】

基本図形の作成を生成AIで作る試み、その利用と応用について Attempts to Create Basic Shapes with Generative AI: Their Use and Application

村山 秀明 | MURAYAMA Hideaki

Since the advent of "chatGPT," AI that automatically generates text from text input has become popular.

Image generation is also becoming mainstream, where images are generated from text.

In this paper, we focused on four image generation service applications: DALL-E, Stable Diffusion, Bing Image Creator, and Adobe Firefly, and compared the results of image generation using the same prompts.

We have summarized the differences in the functions of each application, consideration of the text used to generate images, and thoughts and considerations regarding prompts for creating basic shapes such as rectangles and circles using the four services.

Keywords:

人工知能, AI, 画像生成AI, プロンプト, Adobe Firefly, Stable Diffusion, DALL-E, Image Creator, 基本図形, Artificial Intelligence, AI, Image Generation AI, Prompt, Adobe Firefly, Stable Diffusion, DALL-E, Image Creator, Basic Shapes.

はじめに

openAI社から2022年11月に一般公開した「ChatGPT」により、生成AIの利用が注目されるようになった¹⁾。

「ChatGPT」の発表をきっかけに、さまざまな種類の生成AIが誰でも使えるサービスやアプリケーションとして登場している。

その中でも、言語から画像を生成するAIの成長は著しい。2018年ごろからAIが画像生成するサービスは色々と登場している。2023年9月現在までの間に有料・無料を問わず誰でも利用できる画像生成AIのサービスやアプリケーションは、次の通りだ。

- 2018年 Artbreeder (Morphogen)²⁾³⁾
- 2018年7月 StyleGAN (NVIDIA社)⁴⁾⁵⁾
- 2019年 generated.photos (Generated Media)⁶⁾⁷⁾
- 2021年1月 DALL-E 2 (openAI社)⁸⁾⁹⁾
- 2022年6月 Imagen (google社)¹⁰⁾¹¹⁾
- 2022年7月 Midjourney (Midjourney社)¹²⁾
- 2022年8月 Stable Diffusion (Stability AI社)¹³⁾
- 2022年8月 Craiyon (旧名:DALL-E mini) (Craiyon社)¹⁶⁾
- 2023年2月 Bing ImageCreator (Microsoft社)¹⁷⁾¹⁸⁾
- 2023年9月 Adobe Firefly (Adobe社)¹⁹⁾²⁰⁾

生成AI (Generative AI) とは、人工知能の一分野で、機械学習という大量のデータからパターンや特徴を学習させるプログラムとそこから新たなコンテンツを生成するものである。画像生成AIに関しては当初、「敵対的生成ネットワーク (GAN)」や「変分オートエンコーダ (VAE)」などの技

術を使い、画像を用意されたパラメータで変化させたり、2つ以上の画像を組み合わせる新たな画像を生成するArtbreeder、大量の画像を学習させ、学習した画像から類似したものを再び画像生成させるStyleGANやgenerated.photosのような画像から画像を生成するAIが多かった⁵⁾。しかし、「ChatGPT」のような自然言語処理のモデルが登場したことで、DALL-E 2のように、入力したテキストから画像を生成するモデル(Text To Image Diffusion Models)が増えている¹⁰⁾。

とくに最近では、Stability AI社のStable Diffusionがオープンソースでコードを提供しているため、利用者が増加している¹⁴⁾¹⁵⁾。Stable Diffusionが入力したテキストから画像を生成するモデルの代表的なアプリケーション・サービスになりつつあると言える。

今回は、Stability AI社のStable Diffusion¹³⁾、openAI社のDALL-E⁸⁾、Microsoft社のBing Image Creator¹⁸⁾、そしてAdobe社のAdobe Firefly²⁰⁾の4社が提供するサービス・アプリケーションでの画像生成結果を比較し、効率の良いテキストの命令(プロンプト)、結果の違いを調べる。

画像生成の結果の違いを比較するため、入力するテキストの命令(「プロンプト」と呼ぶ)は、可能な限りすべて英語で同じものを利用する。Stable Diffusion、DALL-E、Adobe Fireflyは生成した画像を元にして、さらに画像を生成する機能を持つため、画像から生成する機能はとくに利用しない。

なお、Stable DiffusionはStable Diffusion web UIを使い¹⁴⁾、WindowsとMac両方の環境で確認を行った。

画像生成AIでよくある使い方の比較

生成の元になる画像データの権利について

通常、画像生成AIは大量の画像データを解析し、解析された画像データから、テキストに近い画像データを生成する。画像生成AIの利用者は自分が思い描いているイメージを生成するためにこれを利用する。しかし、生成に用いられる解析に利用された画像データの準備方法や著作権などの権利状況は不明だ。

今回調査する4つの画像生成AIの中で、Adobe Fireflyのみが、商用利用を認めている¹⁹⁾。DALL-E、Bing Image Creatorは個人が楽しむ範囲での利用を認めている¹⁷⁾。

Stable Diffusionは、利用規約上では、学習データ、生成したものを商用利用も含め、利用者が責任を取るのであれば、自由に利用できることになっている。しかし、Stable Diffusionの画像生成の元になる画像データは、著作権者の許諾を得たものかわからないため、実際、商用利用は難しいだろう¹⁵⁾。

Stable Diffusionでの画像生成

Stable Diffusionを自身のPCに設定できるStable Diffusion Web UIで画像生成AIを制作する環境を構築し¹⁴⁾、東北芸術工科大学から見る風景に近い画像の作成を試みた。生成したいイメージを想定し、テキストを考えたい。山形市や東北芸術工科大学という固有の名称は、生成AIの画像データには含まれていないと仮定し、「小高い丘に立ち、眼下に広がる市街地の風景」「視界の奥に山並み」「空は青空」といった内容でテキストの命令(プロンプト)を設定した。

設定したプロンプト:

The view from a small hill, a residential area below, the residential area spreading out to the horizon, high mountains in the distance, greenery spreading near the mountains, blue sky.

(日本語訳:小高い丘からの眺め、眼下に広がる住宅街、地平線まで広がる住宅街、遠くに高い山々、山の近くに広がる緑、青い空。)

画像を生成した結果は図1から図4までで、いずれも指定したテキストの内容を含んだ画像が生成され、大きく異なる画像にはなっていない。

図1:DALL-Eによる生成結果の例

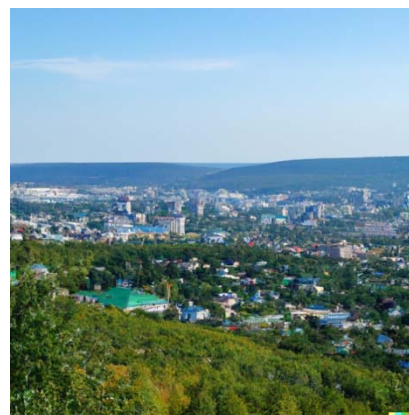


図2:Stable Diffusionによる生成結果の例



図3:Bing Image Creatorによる生成結果の例



図4:AdobeFireflyによる生成結果の例



画像生成AIの問題

画像生成AIで作成した画像は、学習した膨大な画像データから入力したテキストの命令(プロンプト)に近い画像を生成しているだけだ。画像生成AIの行う機械学習の多くは、画像データの色情報をピクセル単位やそれに近い単位で解析、解析した情報を数値化し、配列、JSON、ベクトルなどのデータに変換する。実際、アプリケーション・サービスが解析した画像を利用する際には、その数値化された画像データを、アプリケーション・サービスごとに持つ特定のルー

ルで画像生成を行うが、生成する画像には、意図しない画像情報、いわゆるノイズが含まれることもある。

意図しない画像情報(ノイズ)が含まれることを前提に、DALL-E、Stable Diffusion、Adobe Fireflyでは生成した画像の一部を選択し、テキストの命令(プロンプト)を新たに指定することで、一度生成した画像の一部のみを変更した画像を作成することもできる。例えば、DALL-Eで先ほどの生成例として示した図1の山の部分を塗りつぶす機能を使い、置き換えると、図5のように画像の一部だけを生成し直すことができる。

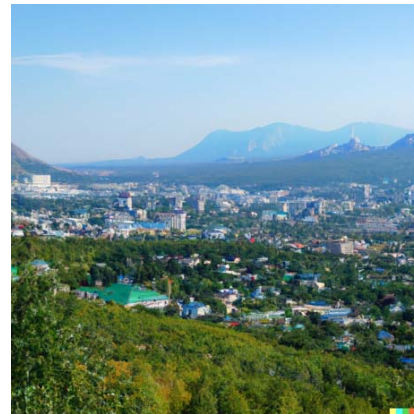


図5:DALL-Eで実際に生成した画像の一部を指定し、再生成を行った画像

Stable Diffusionでは画像を生成すると、生成した画像の情報の中に「シード値」という固有の数値がつく(図6)。この「シード値」を指定したうえで、テキストの命令(プロンプト)を新たに入力し、画像を生成すれば、画像の情報を一部引き継いだ別の画像を生成が可能だ。

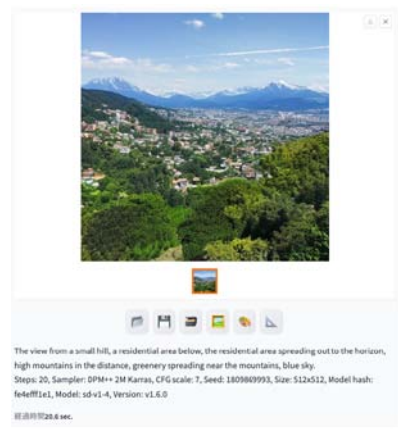


図6:Stable Diffusionでは生成した画像の様々な情報が付与される

また、Stable Diffusionでは、意図しない画像生成を減らすために、画像を作るためのテキストの命令(プロンプト)のほかに、画像中に含めたくないテキスト(ネガティブプロン

プト)を事前に設定してから画像生成を行うことができる。さらに、生成したい画像のイメージの下書きを事前に読み込んだ上で、テキストの命令(プロンプト)を実行する方法でも画像を生成できる。アプリケーション・サービスによっては、一度生成した画像から画像を生成できるかは異なるので、注意が必要だ。

基本図形を生成AIで作る試み

同じ色の画像を生成する

風景などのイメージでは意図しない画像情報(ノイズ)が気になることは少ないが、人物や小物が入った画像を生成すると、意図しない画像情報(ノイズ)が気になることがある。例えば、図7のように「見える部分の色がすべて青一色の画像」をAdobe Photoshopを操作して作成するのは簡単だ。しかし、図7と同じ画像を画像生成AIを使って、テキストの命令(プロンプト)で生成するのは難しい。

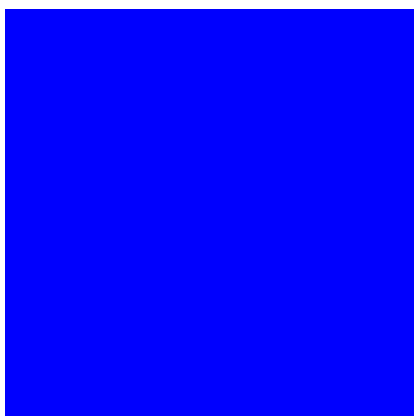


図7:Adobephotoshopで作成した青い画像

画像生成AIで「見える部分の色がすべて青一色の画像」を作るために、次のようなテキストの命令(プロンプト)を入力した。

設定したプロンプト:

blue,#0000FF ,on blue,R=0,G=0,B=100%,flat,blurry

それぞれの生成AIで、このテキストの命令(プロンプト)で複数回生成を試み、生成された画像はおおよそ以下の図8から図11までの4つの画像だ。

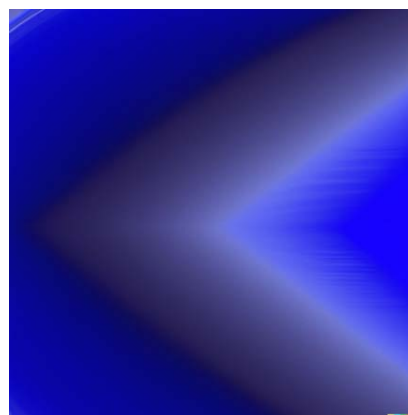


図8:DALL-Eによる生成結果の例

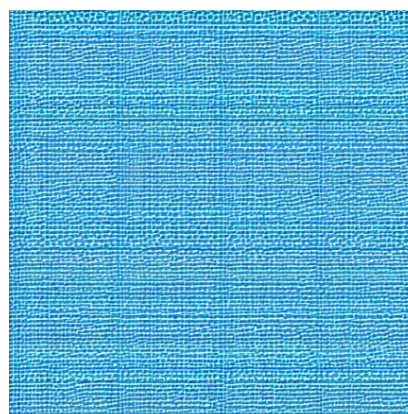


図9:Stable Diffusionによる生成結果の例

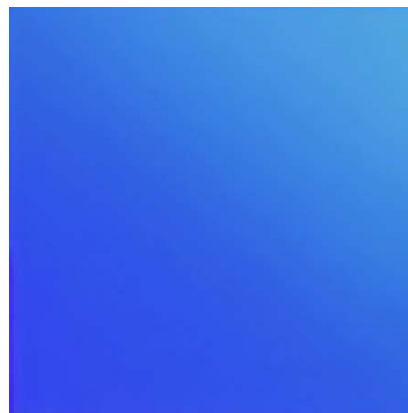


図10:Bing Image Creatorによる生成結果の例

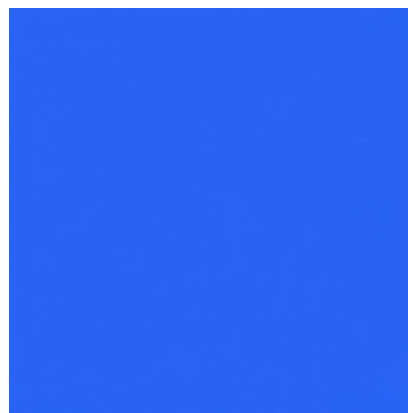


図11:AdobeFireflyによる生成結果の例

DALL-E(図8)、Stable Diffusion(図9)は、複数回繰り返し画像作成を行なっても、DALL-Eでは図8、のようなグラデーションを含んだ画像を、Stable Diffusionでは図9のように、意図しない画像情報(ノイズ)で画像全体がデコボコとした画像を生成し、単色の画像を生成できなかった。一方、Bing Image Creator(図10)、Adobe Firefly(図11)では、指摘されなければ気がつかない程度のグラデーションや色味が若干異なる画像を生成できた。Bing Image Creator、Adobe Fireflyでは、複数回画像を生成しても、Adobe Photoshopで作成した図7のように「見える部分の色がすべて青一色の画像」を生成できた。

基本図形を生成する

画像生成AIで四角形、丸、三角形などの基本図形の生成を試した。試しに、四角形を意味する「square, rectangle」というテキストの命令(プロンプト)で画像生成を行った。その結果が図12から図15である。

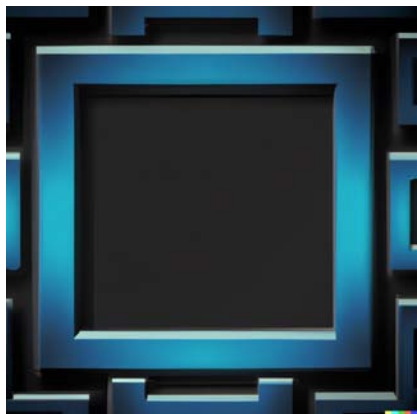


図12:DALL-Eによる生成結果の例



図13:Stable Diffusionによる生成結果の例



図14:Bing Image Creatorによる生成結果の例

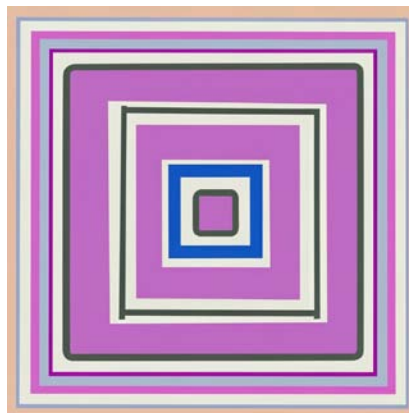


図15:AdobeFireflyによる生成結果の例

「square, rectangle」というテキストの命令(プロンプト)で図16のような画像が生成されることを想定していたが、全く意図していない画像が生成された。



図16:テキストの命令で生成されることを想定していた画像

基本図形の生成に限らず、画像生成AIで想定したイメージに近い画像を生成するためには、テキストの命令(プロンプト)で、①描く被写体の具体的な形状②被写体の輪郭、面・塗りなど細部の状態③被写体の背景の色④作風の4項目をテキストで明確に記述する必要がある。図16のような基本図

形、シンプルな画像を作るために、先ほどの4項目を含んだ次のようなテキストの命令(プロンプト)を作文した。

図16を想定したプロンプトの日本語:

背景は白地。

作成する画像の50%の大きさの正方形。

正方形の線は黒い細い直線。

正方形の内側は白い色。

設定したプロンプト:

background is white.

one square 50% the size of the image.

The square is a thin black straight line.

inside of the square is white.

数回同じテキストの命令(プロンプト)で画像生成を試みたが、想定する図16に近い画像は生成されなかった。どのサービス・アプリケーションも生成した画像(図17から図20)では、1cmや1ピクセルという単位、%や割合など画面に対し描くものの比率を設定しても無視され、そしてまた、線の数、図形の形状、背景の色も指定したテキストの命令(プロンプト)通りに描かれることはなく、意図しない画像情報(ノイズ)が入った画像を生成した。

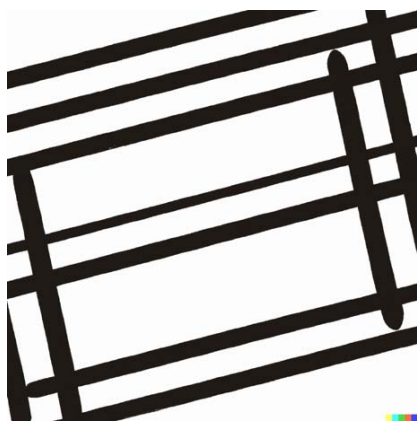


図17:DALL-Eによる生成結果の例

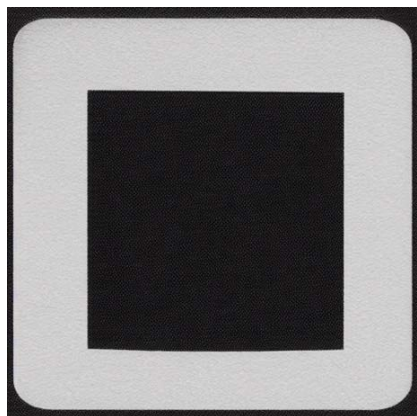


図18:Stable Diffusionによる生成結果の例

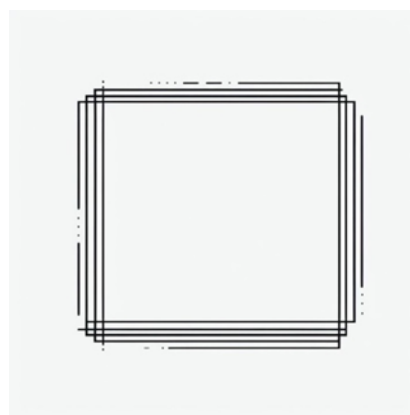


図19:Bing Image Creatorによる生成結果の例

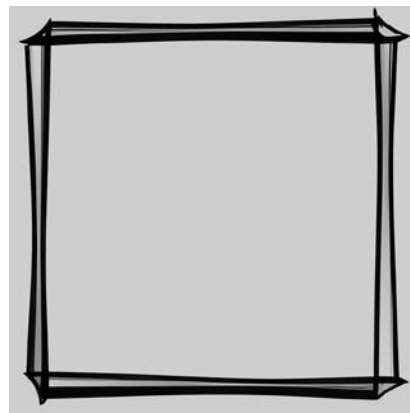


図20:Adobe Fireflyによる生成結果の例

そのため、次に先ほどの命令と同じ意味の文章で、以下のような短いテキストの命令(プロンプト)に変更して生成を試みた。

新たに設定したプロンプトの日本語:

白地に黒い細い線で1つの正方形を描く

設定したプロンプト:

draw one square with one thin black line on white

その生成した画像が、図21から図24である。

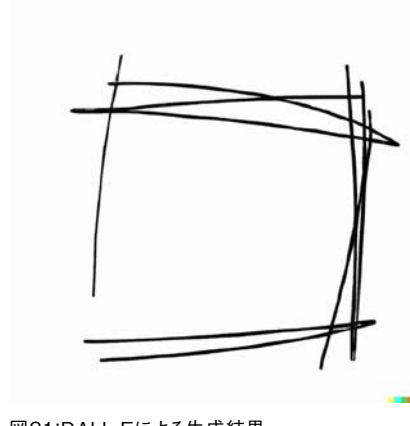


図21:DALL-Eによる生成結果

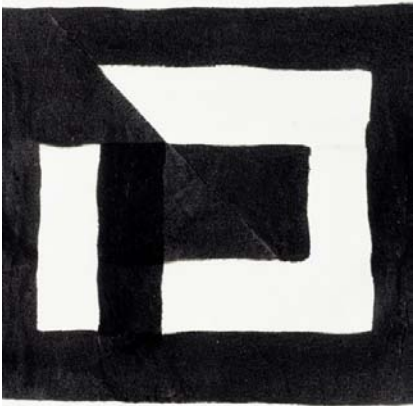


図22:Stable Diffusionによる生成結果

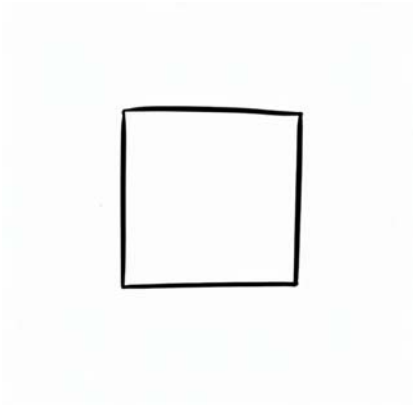


図23:Bing Image Creatorによる生成結果

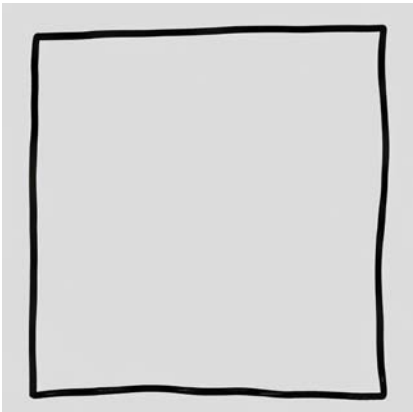


図24:AdobeFireflyによる生成結果

同じテキストの命令(プロンプト)で画像生成を繰り返し試したが、短いテキストの命令(プロンプト)の方が、意図しない画像情報(ノイズ)の少ない画像を生成することができた。とくにBing Image Creator(図23)、Adobe Fire Fly(図24)は想定していた図16の画像に近い画像を生成した。

画像生成AIで基本図形を作成するには、①描く被写体の具体的な形状②被写体の輪郭、面・塗りなど細部の状態③被写体の背景の色④作風の4項目を詳細なテキストの命令(プロンプト)より、短いテキストの命令(プロンプト)の

ほうが、意図しない画像情報(ノイズ)を少なくすることができ、想定する図形に近いものを生成できるようだ。

テキストの命令(プロンプト)の検証

画像生成AIで基本図形を作成するには、「短いテキストの命令(プロンプト)のほうが、意図しない画像情報(ノイズ)を少なくすることができる」ということを検証するため、「丸」の図形の画像生成で試みた。

「四角形」の図形を作るために使ったテキストの命令(プロンプト)の「square」の部分「perfect circle」に置き換えて、画像生成を試みた。

新たに設定したプロンプトの日本語:

白地に黒い細い線で1つの完全な丸を描く

新たに設定したプロンプト:

draw oneperfect circle with one thin black lineon white

それぞれの画像生成AIで図25から図28のような画像を生成できた。図26のStable Diffusionの生成結果を除けば、DALL-E(図25)は線の形状が大きく異なるものの、Bing Image Creator(図27)、Adobe Fire Fly(図28)では、短いテキストの命令(プロンプト)で意図した画像を生成できた。

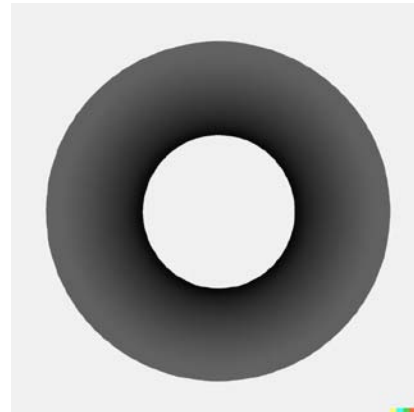


図25:DALL-Eによる生成結果



図26:Stable Diffusionによる生成結果

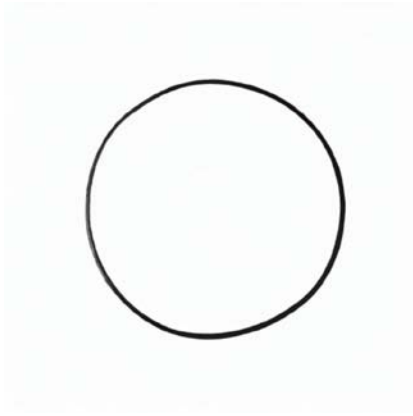


図27:Bing Image Creatorによる生成結果

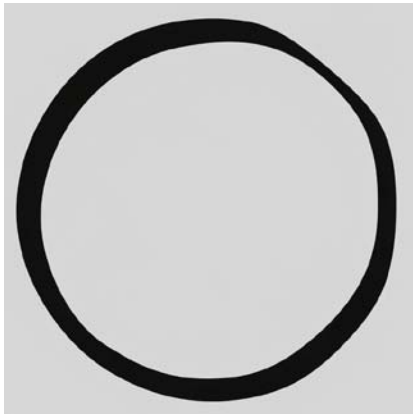


図28:AdobeFireflyによる生成結果

さらにもう一つ同じテキストの命令(プロンプト)を修正し、「直線」で画像生成を試みた。

新たに直線を設定したプロンプト:

draw one black horizontal line on white

画像を生成した結果は、図29から図32である。複数回同じテキストの命令(プロンプト)で実行したが、Bing Image Creator(図31)、AdobeFirefly(図32)は、ほぼ図のような画像生成した。ところが、DALL-E(図29)、Stable Diffusion(図30)は、意図しない画像情報(ノイズ)として、

画像の一部に歪みや、斜線、命令に指示した「直線」でない、まったく異なる画像を生成し、想定した図形を描く結果を得ることができなかった。

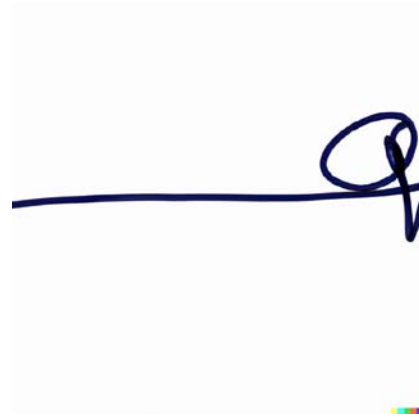


図29:DALL-Eによる生成結果



図30:Stable Diffusionによる生成結果

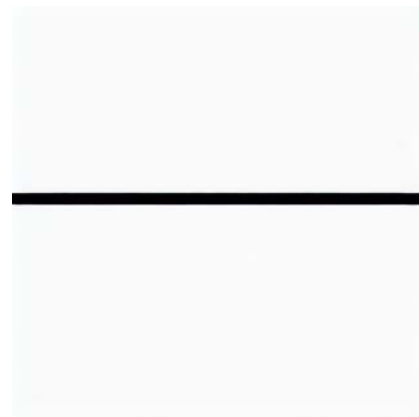


図31:Bing Image Creatorによる生成結果

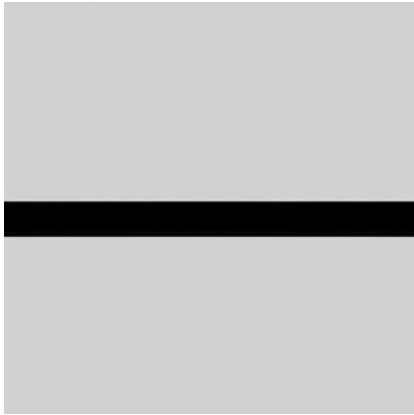


図32:AdobeFireflyによる生成結果

終わりに

その他に三角形、五角形、六角形などの多角形、五芒星の生成も「四角形」を生成した短いテキストの命令(プロンプト)を修正して画像生成を試みたが、DALL-E、Stable Diffusionは思ったような図形を生成することができなかった。今回比較した4つの画像生成AIの中で、DALL-E、Stable Diffusionが基本図形の画像生成が想定した通りにならない点について、さらに時間をかけて検証する必要がある。また、基本図形の画像生成を行う際、同じテキストの命令(プロンプト)で結果を検証したが、テキストの命令(プロンプト)は、長いテキストの命令より、短い言葉のテキストの命令のほうが意図した画像を生成するか、違う単語の組み合わせでの画像生成のほうが意図した画像を生成するかについても回数を重ねて検証の必要がある。

また、今回は同じテキストの命令(プロンプト)で結果を検証したが、それぞれのアプリケーション・サービスごとに、テキスト以外に設定できるパラメータもあるため、テキストの命令(プロンプト)だけでなく、それらの設定を利用して意図するイメージ画像を生成できるかの検証も必要だ。

結論として、現在の画像生成AIがAdobe PhotoshopやIllustratorなどの画像制作アプリケーションで作るものより精度の高い図形を作ることは難しい。

ただし画像生成AIの生成の元になる画像データの学習が進み、AI自身も進化を続けているため、同じテキストの命令(プロンプト)でも今後の進化次第では、比較した4つの画像生成AIのサービス・アプリケーションで言葉だけで意図した基本図形を作ることができる可能性がある。とくにAdobe FireflyやBing Image Creatorはより多くの画像データを学習しているので、可能性は高いと予想される。

今後、より簡単なテキストの命令で画像生成AIが画像を作成できれば、資料作成の図の作成や説明図を作ることも簡単になると思うので、有効なテキストの命令(プロンプト)の研究と様々な画像生成AIでの画像生成を試していく。

[参考資料]

1. ChatGPT: Optimizing Language Models for Dialogue (OpenAI,2022年11月30日),<https://openai.com/blog/chatgpt>
2. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio (2014年6月10日). “Generative Adversarial Networks” (pdf). arXiv,<https://arxiv.org/abs/1406.2661>
3. Artbreeder,Morphogen,<https://morphogen.io/>
4. StyleGAN — Official TensorFlow Implementation,<https://github.com/NVlabs/stylegan>
5. StyleGAN/generated.photos,<https://en.wikipedia.org/wiki/StyleGAN>
6. generated.photos,Generated Media, Inc,<https://generated.photos/>
7. A Style-Based Generator Architecture for Generative Adversarial Networks,<https://arxiv.org/abs/1812.04948>
- 8.DALL-E 2,<https://openai.com/dall-e-2>
9. DALL-E, <https://en.wikipedia.org/wiki/DALL-E>
10. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding,<https://arxiv.org/abs/2205.11487>
11. Imagen,<https://blog.google/technology/research/how-ai-creates-photorealistic-images-from-text/>
12. Midjourney,<https://en.wikipedia.org/wiki/Midjourney>
13. Stable_Diffusion, https://en.wikipedia.org/wiki/Stable_Diffusion
14. Stable Diffusion web UI, <https://github.com/AUTOMATIC1111/stable-diffusion-webui>
15. Stable Diffusionの利用規約,<https://huggingface.co/spaces/CompVis/stable-diffusion-license>
16. Craiyon, <https://www.craiyon.com/>
17. 言葉で画像が作成可能に—新しい Bing に Bing Image Creator が登場 (Microsoft,2023年3月24日),<https://news.microsoft.com/ja-jp/2023/03/24/230324-create-images-with-your-words-bing-image-creator-comes-to-the-new-bing/>
18. Bing Image Creator, <https://www.bing.com/images/create?FORM=IRPGEN>
19. アドビ、Adobe Fireflyの一般提供開始と「Adobe Firefly web版」を発表 (Adobe,2023年9月13日),https://www.adobe.com/jp/news-room/news/202309/20230913_commercial-release-of-gen-ai.html
20. Adobe Firefly, <https://firefly.adobe.com/>

(2023年9月29日 原稿受理、2024年1月18日 採用決定)