

研究報告

画像生成AIの学習モデルの違いに関する研究

A Study on Differences in Learning Models for Image Generation AI

吉 岡 章 夫 Yoshioka Akio
宝塚大学東京メディア芸術学部

抄録

本紀要では、Midjourneyとにじジャーニーという異なる学習モデルを持つシステムの違いを明らかにし、同じpromptを入力した場合でも得られる結果が異なることを示す。これにより、学習モデルの違いがどのように影響するかを理解することができる。また、両システムを効果的に使うためのpromptについて調査し、適切なシチュエーションやタスクに応じてpromptの最適化方法やシステムの選択を提案する。さらに、異なる学習モデルの相互補完的な活用や今後の研究や開発に向けた展望について考察する。この紀要を通じて、Midjourneyとにじジャーニーの特性を最大限に活かす方法や活用の可能性を探る。

In this bulletin, we clarify the difference between systems with different learning models, Midjourney and Niji Journey, and show that the results obtained are different even if the same prompt is entered. This allows us to understand how different learning models affect us. In addition, we investigate prompts to effectively use both systems, and propose optimization methods and system selection for prompts according to appropriate situations and tasks. In addition, we will consider the mutually complementary use of different learning models and the prospects for future research and development. Through this bulletin, we will explore ways to maximize the characteristics of Midjourney and Niji Journeys, as well as potential uses.

1. はじめに

画像生成AIは、近年急速に発展し、多様な分野での用途が拡大している。画像生成AIの成り立ちは、ディープラーニングと呼ばれる機械学習技術をベースにしており、大量の画像データを学習し、新たな画像を生成することが可能である。特に、敵対的生成ネットワーク（GAN）は、生成モデルと識別モデルが競合することで、高品質な画像を生成する技術として注目されている。

画像生成AIの用途は多岐にわたる。アートやデザイン分野では、新しいスタイルやイメージを生成することで、アーティストの創作活動を助けることができる。また、広告業界では、ターゲット層に合わせたビジュアルを生成し、効果的な広告戦略を立てることができる。さらに、建築や都市計画分野では、未来の都市や建物のビジュアルを生成することで、より効率的な計画やデザインが可能となる。エンターテインメント産業では、映画やゲームの世界観をAIが生成し、制作コストや時間を削減することが期待されている。医療分野では、医療画像デー

タを利用して、病気の診断や治療計画の支援ができる。現在の画像生成AIは、技術の進歩に伴い、より高解像度でリアルな画像を生成することができるようになってきている。しかしながら、著作権やプライバシーの問題、悪用の懸念など、新たな課題も生じている。今後は、技術の進化と共に、適切な規制や倫理基準を設定し、画像生成AIの持つポテンシャルを最大限に活用できる環境を整えることが求められるであろう。

また、画像生成AIの発展は、他のAI技術との連携によって、更なる進化が期待されている。例えば、自然言語処理(NLP)と組み合わせることで、文章から直接画像を生成することが可能となり、ストーリーテリングや教育分野での応用が広がるであろう。また、音声や動画データとの連携により、マルチメディアコンテンツの生成も容易になる。

このように、画像生成AIは多くの分野で活用され、人々の生活や産業を革新する可能性を秘めている。今後の技術の進化や、適切な利用方法が見つかれば、私たちの生活はさらに豊かで創造的なものとなるであろう。画像生成AIの持つ無限の可能性を追求することは、今後の研究や開発において重要なテーマとなる。さらなる技術革新や応用範囲の拡大により、画像生成AIは、エンターテインメント、ビジネス、科学、教育など、あらゆる分野で革新的な変化をもたらすことが期待されている。

本研究では、代表的な画像生成AIであるMidjourneyとにじジャーニーに着目し、学習モデルの違いが結果に与える影響を分析する。

2. Midjourneyとにじジャーニーの概要

Midjourneyは、David Holz氏が代表を務めるAI研究チームが開発した画像生成AIであ

る。ユーザーが打ち込んだテキストから画像を生成するAI画像生成サービスの1つである。チャットサービス「Discord(ディスコード)」を通して利用できる。promptを使ったユーザーインタフェースが特徴で、効果的にpromptを扱う事がシステムを最大限に使うポイントとなる。また同じpromptを入力しても、2度と同じものは生成できない。

現在はフリーで使えるv4とサブスクリプションで使えるv5が存在する。v5はv4に比べ、画像の鮮明さ、写実的な表現が飛躍的に向上している。

アニメ風、水彩風の出力画像デザインの方向性がv4とv5で異なる。AIが不得意とされる手の表現については、v5は向上している。

にじジャーニーは、SpellbrushとMidjourneyが開発したアニメ系の絵柄に特化した画像生成AIである。年代別、水彩、浮世絵、kawaii、水墨画、アール・ヌーヴォーの絵柄学習モデルが特徴である。こちらもniji4(v4)とniji5(v5)があり、niji5(v5)では画像の鮮明さ、コミックイラスト(pixiv画)的な表現が飛躍的に向上している。

3. 学習モデルの違いによる異なる結果の事例

同じシチュエーションを描くpromptを両システムに入力した場合の結果の比較を行う。それに加え、絵柄を制御するpromptを入力した場合の結果の比較を行う。

- (1) 描くシチュエーションを規定するpromptを作成する。「透明感のある夏、2人のイケメンサーファーが浜辺で遊んでいる姿」というテーマにする。ただし、このままだとpromptの情報が少ないため、ChatGPTを使って想定イメージを拡張する。

ChatGPTへの依頼：

system：あなたはプロのイラストレーターで、イラストや画像を言語化することに長けている。

prompt：あなたに依頼する。「透明感のある夏、2人のイケメンサーファーが浜辺で遊んでいる姿」というテーマで絵を描こうと思っている。詳細化のためにこのテーマを膨らませてください。

ChatGPTの返答：

このイラストでは、夏の日差しが強く、空には青く澄んだ空気感を感じることができる。透明感のある夏というテーマに沿って、海はエメラルドグリーンから透明な青色にかけてグラデーションが施され、水面には陽光がキラキラと反射している。波は穏やかで、白い泡が砂浜に柔らかく押し寄せている。

2人のイケメンサーファーは若く、活力に満ちている。彼らは茶色の髪を持ち、髪のはらは太陽によって明るくなっている。彼らは筋肉質でありながら、洗練された身体つきをしている。どちらも鮮やかな色のサーフボードを持っており、それぞれの個性が表れている。片方は赤とオレンジのストライプが施されたボードを持ち、もう片方は青と白の幾何学模様が描かれたボードを持っている。

サーファーたちは浜辺で楽しそうに遊んでおり、サーフボードを持って波を追いかける様子が描かれている。彼らは時折、お互いに親しげに声を掛け合っている。周囲には鮮やかな色のパラソルやビーチチェアが配置されており、夏のビーチシーンを盛り上げている。また、遠くの海には、他のサーファーや遊泳客が楽しんでいる。

空には白い雲がぽっかりと浮かんでおり、太陽は雲に隠れることなく輝いている。背景に

は、緑豊かな植物や木々が描かれ、夏の自然が息づいていることを感じさせる。全体的に明るく、さわやかな雰囲気が漂っており、観る人に夏のエネルギーと楽しさを伝えることができるでしょう。

以上より作成したprompt：

Summer day beach, 2 boys in swim suits, handsome, blonde hair, healthy tanned skin, beautiful beach, clear blue sea, yacht, surfer, shining sun, riding waves, surfing, beach parasol, beach chairs, surfboard, footprints, palm trees, seabirds, tourists, looking at the camera, half body shot, 8K --ar 16:9

(2) 同じpromptをMidjourney v4, v5, にジジェーニー v4, v5に入力した結果を次に示す。



Image1：Midjourney v4の出力結果



Image2：Midjourney v5の出力結果



Image3：にじジャーニー v4の出力結果



Image4：にじジャーニー v5の出力結果

4. 学習モデルの違いが結果に与える影響の考察

本研究では、Midjourneyとにじジャーニーに同じpromptを入力した際の結果を比較・分析することで、両システムの学習モデルの違いがどのように影響するかを明らかにすることを旨とした。例として、prompt「Summer day beach, 2 boys in swim suits, handsome, blonde hair, healthy tanned skin, beautiful beach, clear blue sea, yacht, surfer, shining sun, riding waves, surfing, beach parasol, beach chairs, surfboard, footprints, palm trees, seabirds, tourists, looking at the camera, half body shot, 8K --ar 16:9」を両システムに入力した。

Midjourneyの結果について、v4の出力画像(Image1)では、写実的なイラストレーションが提示された。v5の出力画像(Image2)では、カメラで撮影された実写のようなイメージが提示された。一方で、にじジャーニーの結果について、v4の出力画像(Image3)では、マン

ガのイラストのようなイメージが提示された。v5の出力画像(Image4)では、画像投稿サイトpixivでよく見られるようなコミックイラストのようなイメージが提示された。

この例から、Midjourneyはイラストや実写のような出力の幅があるものの写実的な絵柄に重点を置く傾向があり、一方でにじジャーニーは日本的なコミックイラストのような絵柄の傾向があることがわかった。これらの違いは、両システムの学習モデルが異なるデータセットやアルゴリズムを用いて構築されていることから来していると考えられる。

この分析を通じて、Midjourneyとにじジャーニーはそれぞれ異なる視点やアプローチを提供するため、両システムを組み合わせることで、より幅広いテイストや多様な表現に対応できることが示唆された。

5. Midjourneyとにじジャーニーの効果的な使い方

- (1) 絵柄をコントロールするpromptを追加し、それぞれのシステムが適したシチュエーションやタスクを特定する。
- (2) 絵柄をコントロールするために追加したprompt

photo：写真らしくなる。

Lomography：少しレトロな風合いになりやすく、まわりがぼんやりとにじむこともある。

watercolor：水彩。

ukiyo-e：浮世絵風。

cel anime：日本のアニメらしくなる。

Ink painting：水墨画。

※効果を分かりやすく判断するために水墨画のみ使用したprompt: Ink painting, 2 boys in swim suits

以下にそれぞれの結果と分析を示す。

photoを追加した場合：



Image5：photoを追加したMidjourney v4の出力結果



Image6：photoを追加したMidjourney v5の出力結果



Image7：photoを追加したにじジャーニー v4の出力結果



Image8：photoを追加したにじジャーニー v5の出力結果

Midjourneyは、“photo”promptを追加すると、より実写に近い表現と構図に変化した(Image5, 6)。一方にじジャーニーは、ほとんど変化が見られない事から、写実的な表現は不得意であると結論付けられる(Image7, 8)。

Lomographyを追加した場合：



Image9：Lomographyを追加したMidjourney v4の出力結果



Image10：Lomographyを追加したMidjourney v5の出力結果



Image11：Lomographyを追加したにじジャーニー v4の出力結果



Image12: Lomographyを追加したにじジャーニー v5の出力結果

Midjourney、にじジャーニーともに、“Lomography”promptを追加すると、特有のレトロな表現とわずかなにじみを持つ絵柄に変化した(Image9, 10, 11, 12)。特ににじジャーニーは、イラストテイストが無くなり写実的な表現への変化が見られた。

Watercolorを追加した場合：



Image13 : watercolorを追加したMidjourney v4の出力結果



Image14 : watercolorを追加したMidjourney v5の出力結果



Image15 : watercolorを追加したにじジャーニー v4の出力結果



Image16 : watercolorを追加したにじジャーニー v5の出力結果

Midjourney、にじジャーニーともに、“watercolor”promptを追加すると、筆のタッチはそれぞれ異なりだが、総じて水彩の絵柄に変化した(Image13, 14, 15, 16)。その中にじジャーニー v5は、コミックイラストのテイストが重く、水彩のニュアンスが薄く反映されているようだ(Image16)。

ukiyo-eを追加した場合：



Image17 : ukiyo-eを追加したMidjourney v4の出力結果



Image18 : ukiyo-eを追加したMidjourney v5の出力結果



Image19 : ukiyo-eを追加したにじジャーニー v4の出力結果



Image20 : ukiyo-eを追加したにじジャーニー v5の出力結果

Midjourney v4、v5、にじジャーニー v4は“ukiyo-e”promptを追加すると、浮世絵のテイストを含む絵柄に変化した(Image17, 18, 19)。特に波や雲の表現にその特徴が表れている。一方にじジャーニー v5は、ほとんど変化が現れず、“ukiyo-e”promptの影響は受けないと結論付けられる(Image16)。

cel animeを追加した場合：



Image21 : cel animeを追加したMidjourney v4の出力結果



Image22 : cel animeを追加したMidjourney v5の出力結果



Image23 : cel animeを追加したにじジャーニー v4の出力結果



Image24 : cel animeを追加したにじジャーニー v5の出力結果

Midjourney v4、v5、にじジャーニー v4は“cel anime”promptを追加すると、年代の方向性はそれぞれであるが日本のアニメ風の絵柄に変化した(Image21, 22, 23)。一方にじジャーニー v5は、ほとんど変化が現れず、“cel anime”promptの影響は受けないと結論付けられる(Image24)。

Ink paintingを追加した場合：

※効果を分かりやすく判断するために水墨画のみ使用したprompt: Ink painting, 2 boys in swim suits



Image25 : Ink paintingを追加したMidjourney v4の出力結果



Image26 : Ink paintingを追加したMidjourney v5の出力結果

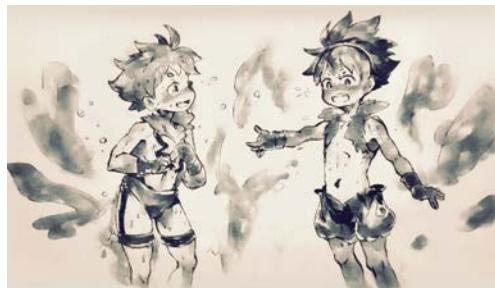


Image27: Ink paintingを追加したにじジャーニー v4の出力結果



Image28: Ink paintingを追加したにじジャーニー v5の出力結果

Midjourney v4、v5、にじジャーニー v4は“Ink painting”promptを追加すると、筆のタッチはそれぞれ異なるが、総じて水墨画の絵柄に変化した(Image25, 26, 27)。一方にじジャーニー v5は、ほとんど変化が現れず、“Ink painting”promptの影響は受けないと結論付けられる(Image28)。

(3) 以上の結果よりMidjourney v4、v5、にじジャーニー v4、v5、それぞれのシステムの適したシチュエーションやタスクを特定する。

Midjourney v4は写実的な表現(Image5)を得意とし、水彩(Image13)や水墨画(Image25)、アニメ調(Image21)など特定の絵柄を表現できる事に特徴がある。

Midjourney v5は、v4をより進化させた出力になっており、より緻密で繊細な表現(Image14, Image22, Image26)が可能になっている。

にじジャーニー v4は、絵柄の指定をしないと特徴のない魅力の少ない表現(Image3)となるが、水彩(Image15)や水墨画(Image27)、アニメ調(Image23)など特定の絵柄を表現できる事に特徴がある。

にじジャーニー v5は、絵柄の指定に影響されることがほとんどなく、今日的なコミックイラスト(pixiv)のような絵柄を表現できる事に特徴がある。

6. Promptの工夫と最適化の実践例

Midjourneyは写実的な表現が得意であることが分かったので、よりそれを強化するようなpromptを工夫することが、その能力をより引き出しやすい事は容易に想像できる。

最適化の実践として、より写実的な絵柄になるよう“cinematic realism”をpromptとして使用した。また実写のカメラにあるパラメータをpromptに加えることで対象物にフォーカスするような効果が得られるため“35mm lens”をpromptに使用した。

これらを実践するためのpromptは次の通りである。

cinematic realism, beautiful 2 boys in swim suits, blonde hair, healthy tanned skin, beautiful beach, clear blue sea, yacht, surfer, shining sun, Summer day beach, riding waves, surfing, beach parasol, high contrast, looking at the camera, full body shot, 35mm lens, light bloom, 8K --ar 16:9

得られた結果は次の通りである。



Image29 : Midjourney v4の出力結果



Image30 : Midjourney v5の出力結果

対象へのフォーカスの当たり方や画像の質感など実写のそれと見間違えるほどの精度で出力された。

にじジャーニー v4は、システムが学習している絵柄を指定した表現が得意であることが分かったので、それを強化するようなpromptを工夫することが、システムの能力をより引き出しやすいと考える。

最適化の実践として、よりエモーショナルな絵柄になるよう“pixiv illust”, “pixiv masterpiece”をpromptとして使用した。また絵柄の年代をpromptに加えることで画風をコントロールする事が出来るため“2020s”をpromptに使用した。

これらを実践するためのpromptは次の通りである。

2020s,high quality,pixiv illust, pixiv masterpiece::1.2, 2 boys in swim suits, Summer

day beach, looking at the camera, half body shot, 8K --ar 16:9

得られた結果は次の通りである。



Image31：にじジャーニー v4の出力結果

“pixiv illust”や“pixiv masterpiece”の影響が大きいのか、予想を超える現実離れした幻想的な絵柄が出力された。“2 boys in swim suits”や“looking at the camera”が無視される傾向もあるようだ。これが画像生成AIの醍醐味とも言える。

最後に、にじジャーニー v5は、pixivのような今日的コミックイラストに特化している事が分かったので、それを強化するようなpromptを工夫することが、システムの能力をより引き出しやすいと考える。

最適化の実践として、よりエモーショナルな絵柄になるよう“pixiv illust”をpromptとして使用した。また絵柄の年代をpromptに加えることで画風をコントロールする事が出来るため“2020s”をpromptに使用した。

これらを実践するためのpromptは次の通りである。

2020s,high quality,pixiv illust, 2 boys in swim suits, Summer day beach, looking at the camera, half body shot, 8K --ar 16:9

得られた結果は次の通りである。



Image32：にじジャーニー v5の出力結果

アウトラインや髪の毛の描き方、着色の仕方に今日的なコミックイラストの特徴がある画像が出力された。

7. まとめ

(1) 本研究では、以下の主要な点を取り上げ、検討した。

Midjourneyとにじジャーニーという異なる学習モデルを持つシステムの基本的な概要と違いについて説明した。これにより、両システムの性質とそれぞれの特徴を理解することができた。

同じpromptを両システムに入力した場合の事例を比較・分析し、学習モデルの違いが結果に与える影響を考察した。この結果から、それぞれのシステムが独自の強みを持ち、その強みに応じた指示に対して適切な出力を行えることが確認された。

Midjourneyとにじジャーニーを効果的に使うためのpromptについて調査し、それぞれのシステムが適したシチュエーションやタスクの特定promptの最適化方法を提案した。これにより、両システムの利点を最大限に活用するための具体的なアプローチが明らかになった。

これらの主要な点を再確認することで、今後のMidjourneyとにじジャーニーの活用において、最適なシステムを選択し、それぞれの特性を最大限に活かす方法を実践することが可能となる。

(2)

今後のMidjourneyとにじジャーニーの活用に関しては、以下のような可能性が考えられる。

より高度なタスクへの応用：両システムの特徴を活かして、複雑な問題解決やクリエイティブなアイデア生成など、より高度なタスクに対応できるようになることが期待される。

マルチモーダルな活用：Midjourneyとにじジャーニーをテキストだけでなく、画像や音声などのマルチモーダルなデータに対応させることで、より幅広い分野での活用が可能となる。

相互補完的な連携：両システムの強みを組み合わせることで、それぞれのシステムが持つ限界を補完し、より高品質な結果を生み出すことが期待される。例えば、Midjourneyによる出力をにじジャーニーに入力することで（その逆もあり得る）、新しい絵柄や表現を創出できるであろう。

人間との協働：両システムが人間の専門家やクリエイターと協働することで、人間の知識や経験を補完し、より効果的な意思決定や革新的なアイデアを生み出すことが可能となる。

システムの進化：両システムは継続的にデータを学習し、アルゴリズムが改善されることで、より正確で独創的な結果を提供する能力が向上し、新たな活用分野が拓かれることが期待される。

これらの展望を踏まえ、今後のMidjourneyとにじジャーニーの活用においては、システムの進化や相互補完的な連携を通じて、多様な分野での応用や人間との協働がさらに進展することが予想される。

参考文献

- 1) Midjourney (2023/7/31アクセス)
<https://www.midjourney.com/app/>
- 2) にじジャーニー (2023/7/31アクセス)
<https://nijijourney.com/ja/>
- 3) 【画像生成AI】Midjourneyでプロンプトによく使う言葉まとめ (2023/7/31アクセス)
<https://blogcake.net/aiart-prompt/>
- 4) niji・journeyの絵柄変更系呪文3種類の紹介と使い方 (2023/7/31アクセス)
<https://www.youtube.com/watch?v=NZjWRhqCF7g>