

GenerativeAIによるデジタルコンテンツ作成演習の 授業実践

渋沢良太

鹿児島県霧島市国分中央 1-10-2 第一工科大学 工学部 情報電子システム工学科

E-Mail: r-shibusawa@daiichi-koudai.ac.jp

Digital Content Creation Exercise Class with GenerativeAI

Ryota SHIBUSAWA

Department of Information and Electronic Systems Engineering, 1-10-2, KokubuChuo, Kirishima, Kagoshima, 899-4395, Japan

Abstract: The department of information and electronic systems engineering at Daiichi institute of technology has been offering Digital Content Creation Exercises I and II, in which digital content creation using video and image editing software has been the subject matter. In response to recent major developments in GenerativeAI technology, we have reviewed the content of the classes and decided to focus on the creation of digital contents using GenerativeAI in the 2022 academic year. This paper describes the practice of the class.

Key words: GenerativeAI, Diffusion model, Stable diffusion, Digital content

1. はじめに

近年, Deep Learning [1]を応用した技術が急速に発展しており, ほぼ毎日のように新しい技術が発表されている. その中でも 2022 年頃から急速に進展している技術分野として, 静止画, 動画, 音楽, テキスト等を自動生成する Generative AI [2]がある. これらの技術では, ユーザが生成したいコンテンツを連想させるテキストや画像を入力すると, それに適したコンテンツをソフトウェアが自動生成する. これらの技術の進展には, 潜在拡散モデル(Latent diffusion model [3], [4])が大きく寄与している.

本学の情報電子システム工学科のビジネス分野では, デジタルコンテンツ作成演習 I と II の授業を開講し, これまでに動画や画像編集ソフトを使った, デジタルコンテンツの制作を題材としてきた. 上記の Generative AI 技術の大きな進展を受けて授業内容を見直し, 2022 年度は Generative AI を使ったデジタルコンテンツ作成 II の題材とすることにした. 本論文では, その授業の実践について述べる.

2. 授業の概要

2022 年度の後期に, 本学の工学部情報電子システム工学科の情報工学ビジネス分野で開講した, デジタルコンテンツ作成演習 II の授業の概要を表 1 に示す.

表 1 授業の内容

回	内容
1	イントロダクション
2	デジタルコンテンツの市場環境
3	Deep Learning の基礎
4	Deep Learning の基礎
5	Stable Diffusion [5]を使った text-to-image による 2D 画像生成
6	Midjourney [6]を使った text-to-image による 2D 画像生成
7	Mubert Text to Music [7]を使った text-to-music による音楽生成
8	RunwayML [8]を使った画像生成と編集
9	Stable Diffusion による image-to-image による 2D 画像生成
10	Stable DreamFusion [9]を使った text-to-3D image による 3D 画像の生成
11	Phenaki [10]を使った text-to-video による動画の生成
12	コンテスト応募作品の制作
13	コンテスト応募作品の制作
14	コンテスト応募作品の制作
15	まとめ

本授業は, 簿記などの経営学に関する科目を必修で学ぶ学生を対象に開講した授業である. そのため単に技術的な内容のみを教えるのでは

なく、デジタルコンテンツに関する現在の市場環境と今後の展開についての学習を第1回と2回で行った。また技術の詳細について知らなくても Generative AI のツールは使用できるが、原理と技術の限界を理解するために、Generative AI に関わる Deep Learning の技術を第3回と4回で学んだ。その後デジタルコンテンツ作成の演習に取り組み、学習のアウトカムとして制作した最終課題の作品を学外のコンテストに応募した。

3. 演習で作成した作品例

3.1. 2D 画像

Stable Diffusion[5]を使った text-to-image の生成では、Google の Colaboratory [11]上で Python のプログラムを実行して画像を生成した。その際、学習済みの機械学習モデルやライブラリを公開している Hugging Face [12]の rinna/japanese-stable-diffusion モデル[13]を使って実装した。同モデルでは、日本語でキーワードの列を入力して画像を生成する。生成にかかる時間は、無料版の Colaboratory で1~5分程度である。また、1枚ずつ生成するのではなく、複数枚を同時に生成することも可能である。同じキーワード列であっても実行の度に、モデル中のアルゴリズムで使用されているシード値がランダムに変わるため、異なる画像が生成される。そのため、生成された画像が気に入らなければ何度も生成し直すことができる。画像生成時のキーワードとシード値を控えておけば、同じ画像を再び生成することも可能である。図1に「子供 笑顔 虹 油絵」というキーワードで生成した画像を示す。また、図2に「子供 笑顔 竹やぶ 水彩画」というキーワードで生成した画像を示す。

写真の様な画像を生成することもできるが、この例のように「油絵」や「水彩画」のように、どのように描かれた画像にするかを指定して出力させることもできる。同様に、図3、図4にも別のキーワードで生成した画像を示す。

このように、テキストによるキーワードの列によって画像を生成する text-to-image においては、ユーザの望みに近い画像を生成するためには、どのようにキーワードを入れるかが重要になる。一般に、画像以外を生成する場合も含めて、Generative AI に対するテキスト入力をどのように上手く行わせるかについてのノウハウは、

プロンプトエンジニアリング[14]と呼ばれている。



図1「子供 笑顔 虹 油絵」



図2「子供 笑顔 竹やぶ 水彩画」

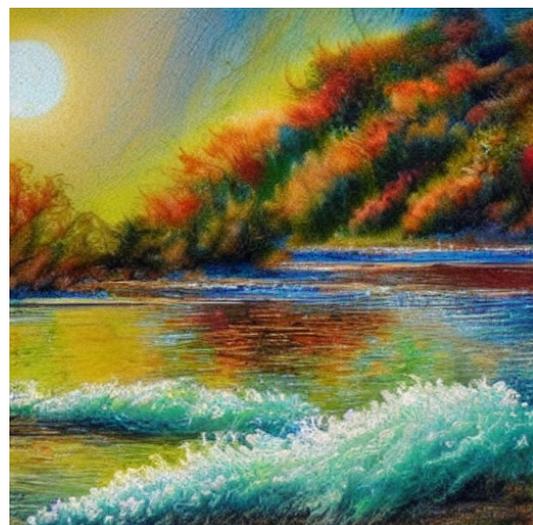


図3「森 川 海 光 いろいろな生物 油絵」



図 4 「魚川鳥海蝶森花光油絵」

次に、Midjourney[6]によって生成した画像を図 5, 図 6, 図 7, 図 8に示す。Midjourneyは、チャットサービスである Discord[15]を通して画像の生成を行う。Discordのチャットルームに画像生成のコマンドとオプション、キーワード列を入力することで画像生成できる。画像生成時のオプション指定により、アスペクト比や解像度を細かく指定できる他、HDRI(High Dynamic Range Image)や、パノラマ画像の生成も可能である。そのため、Midjourneyで生成したパノラマ画像をUnity[16]などで球体に貼り付けて3D空間を生成し、Meta Quest[17]等のVRデバイスで体験すること等も可能である。



図 5 Midjourneyで生成した馬の画像



図 6 「medical equipment cool cartoon」



図 7 「medical equipment cool cartoon」

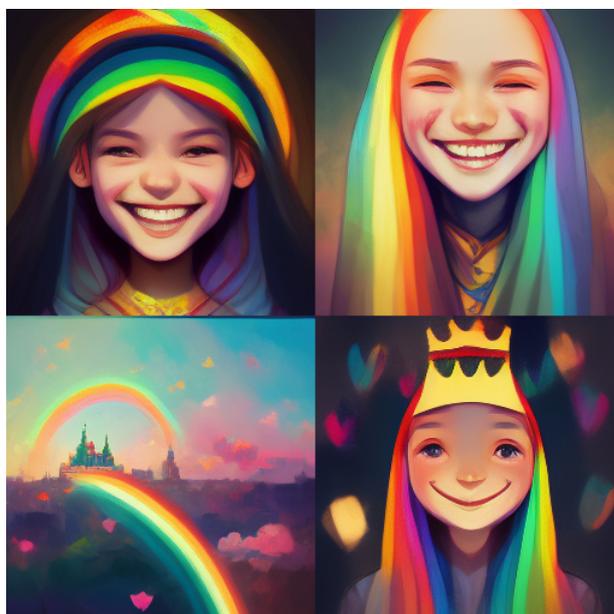


図 8 「smiling princess rainbow」

これらの演習成果を応用して、第12回から14回の授業では、「第43回全国豊かな海づくり大会～おんせん県おおいた大会～」[18]の大会ロゴマークデザインに応募する作品の制作を行った。同大会は「多様な水産物の味力（みりょく）発信」、ひいては「大分県産魚の消費拡大」に資することを目的としている。ロゴマークデザインの募集では、その目的に沿って、大分県の豊かな自然環境で生まれた多様な水産物をPRできるようなわかりやすいロゴマークデザインのイラスト画が求められた。制作したロゴマークデザインの例を図9、図10、図11に示す。図11の作品は、ベースとなるいくつかの魚の画像をGenerative AIで生成した後、画像編集ソフトで加工して背景画像を制作して、合成されたものである。

また、同大会の大会PRキャラクターとして「めじろん」が既に制作されているが、同様に大会PRのマスコットキャラクターの制作を希望する受講生は、図12の作品を制作した。



図9 ロゴマークデザイン作品例1



図10 ロゴマークデザイン作品例2

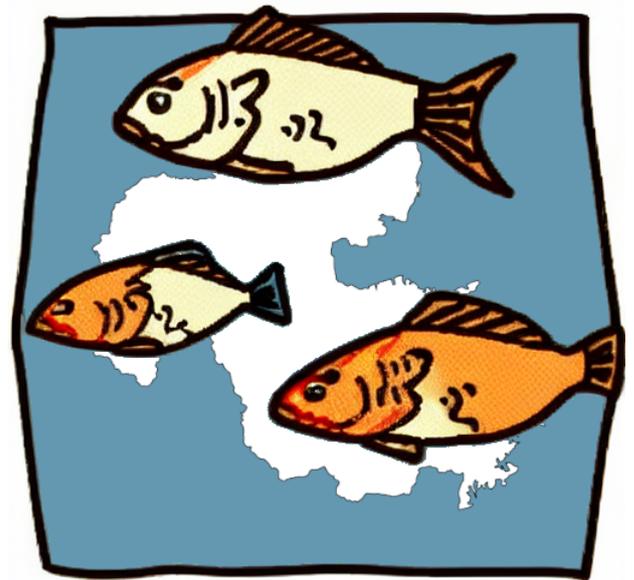


図11 ロゴマークデザイン作品例3



図12 マスコットキャラクターの作品例

3.2. その他

本論文では、論文という形式に掲載できるコンテンツである2D画像を主に示したが、授業ではStable DreamFusion [9]による3D画像の生成、Mubert Text to Music [7]による音楽の生成、Phenaki [10]による動画の生成も取り扱った。Mubert Text To Musicによる音楽生成ではColaboratoryでコードを作成、実行して音楽生成した。音楽の時間、キーワードを列挙することで音楽の生成が可能である。

4. 今後の課題

本論文では、Generative AIを使ったデジタルコンテンツ演習の授業の実践について報告した。Generative AIのプロンプトエンジニアリングにより、ユーザが求めているものに近いコンテンツを生成できる一方で、生成されたコンテンツに人間が加工してコンテンツを生成したい場面も多く見られた。このような用途のために、Photoshopのデータ形式のように、レイヤー等に別れたコンテンツの生成が上手くできるようになると、人間とGenerative AIが協調してデジタルコンテンツを作成しやすくなると考えられる。

参考文献

- [1] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015). <https://doi.org/10.1038/nature14539>
- [2] Cao, Yihan, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun. "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt." *arXiv preprint arXiv:2303.04226* (2023).
- [3] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15)*. JMLR.org, 2256–2265.
- [4] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer: High-Resolution Image Synthesis with Latent Diffusion Models. *CVPR 2022*: 10674-10685.
- [5] Stable Diffusion, <https://github.com/Stability-AI/stablediffusion>, (2023年5月6日参照).
- [6] Midjourney, <https://www.midjourney.com/>, (2023年5月6日参照).
- [7] Mubert-Text-to-Music, <https://github.com/MubertAI/Mubert-Text-to-Music>, (2023年5月6日参照).
- [8] RunwayML, <https://app.runwayml.com/>, (2023年5月6日参照).
- [9] Stable-Dreamfusion, <https://github.com/ashawkey/stable-dreamfusion>, (2023年5月6日参照).
- [10] Ruben Villegas, Mohammad Babaeizadeh, Pieter-Jan Kindermans, Hernan Moraldo, Han Zhang, Mohammad Taghi Saffar, Santiago Castro, Julius Kunze, Dumitru Erhan: Phenaki: Variable Length Video Generation From Open Domain Textual Description. *CoRR abs/2210.02399* (2022).
- [11] Google Colaboratory, <https://colab.research.google.com/>, (2023年5月6日参照).
- [12] Hugging Face, <https://huggingface.co/>, (2023年5月6日参照).
- [13] rinna/japanese-stable-diffusion, <https://huggingface.co/rinna/japanese-stable-diffusion>, (2023年5月6日参照).
- [14] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 384, 1–23. <https://doi.org/10.1145/3491102.3501825>
- [15] Discord, <https://discord.com/>, (2023年5月6日参照).
- [16] Unity, <https://unity.com/>, (2023年5月6日参照).
- [17] Meta Quest, <https://www.meta.com/jp/quest/>, 2023年5月6日参照).
- [18] 第43回 全国豊かな海づくり大会～おんせん県おおいた大会～, <https://www.pref.oita.jp/soshiki/16360/umidukuritop.html>, (2023年5月8日参照)