

生成AIの作る画像はなぜ違和感があるのか

～フロイトの視点とプロンプトの工夫からのアプローチ～

頌栄女子学院高等学校 川原碧月

背景・目的



昔から絵を描くことが好きだったため、AI画像生成に興味があった。AIの生成する絵を見たときに違和感を感じ、生成AIによる画像生成について調べてみたくなった。



人間の顔や体、背景などをうまく生成できないことがある。また、プロンプト(AIへの指示)の趣旨をAIが十分に汲み取れないことがある。(図1)

現時点での生成AIの課題



図1 Dream Studioを利用し「School festival (文化祭)」と入力して作成

調査

【課題点の克服】

テキスト生成AIではプロンプトに「ステップバイステップで考えてください」と付け加えることで、付け加えなかった場合に比べて文章生成の精度が向上するという[1]。このように、少しの工夫で生成AIをより効率的に使用できる場合がある。今回は、画像生成AIにおいてプロンプトの工夫による上記の課題点の克服を目指す。

【違和感の分析】

違和感を「不自然さ」と「不気味さ」に分解して考える。

【不自然さ】

AI生成画像の不自然さの原因として、以下が考えられる。

1. 状況とマッチしていない

プロンプトで表現しきることができなかった情報について、AIが自己解釈を加えて画像を生成することで、プロンプトの示す状況とミスマッチの画像が生成されてしまう。

2. 人の予想の範疇を外れる

プロンプトの不備やその他の理由によって、ユーザーが生成したい画像のイメージと実際に生成された画像との間にギャップが生じ、違和感を感じる原因となっている。

【不気味さ】

・ 不気味なもの

オーストリアの心理学者のジークムント・フロイトの「不気味なもの」という論文によると、不気味さとは馴染みのあるものに由来するのだという[2]。生成AI画像の中にも馴染みのある部分とそうでない部分(=不自然さ)があり、それが不気味さに繋がっているのではと推測した。



ジークムント・フロイト

・ 「不気味の谷」現象

森政弘氏によって提唱された理論。人工物の見た目や動作が人間に近づくにつれて好感が上がるが、図2のように、ある程度まで人間に近づく急激に好感が下がり、そこから更に人間に近づくことで好感が最大になるといふもの。

現在、AI生成画像の印象はこの谷の中にいると考えられるため、谷に落ちる前か谷の向こう側に好感を持ってゆく必要がある。今回は、谷の手前に好感を持ってゆくことを目標とする。

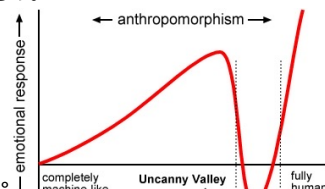


図2 不気味の谷 Wikipediaより引用 (© 2004 Matthew Trump)

【プロンプトの工夫】

学校の文化祭のポスターに使われる画像を生成する想定で、DreamStudio, Dalle-3を使用して実験を行った。

【工夫1】可能な限り詳細な指示をする

細かく状況を指定することで、どんな画像を生成したのかという理想図をより正確に伝えることができる。

【工夫2】イラスト風の画像の生成を指示

イラスト風にすることで、写実的な画像に比べて必要とされるディテールの量が減る。その結果、不気味の谷に落ちる前の好感の最大値に近づくことができる。

【結果】

- ・ 図3は「文化祭の様子」とだけ入力したもの、
 - ・ 図4は「文化祭で日本の高校生が笑顔でこちら側に手を振っている。制服を着ている。屋外で、校舎が風船やカラフルなテープで飾り付けられている。人々はランダムに立っている。みんな違う表情をしている。」と入力したもの、
 - ・ 図5では図4のプロンプトに「イラスト調」と加えて入力したものである。
- 上記のようなプロンプトの工夫は効果的であった。段々と画像の違和感、不気味さが軽減されていることが分かる。
- また、プロンプトによって加えられるディテールについて、色に関する指示は拾われやすいが、奥行き表現が必要だったりするのは反映されにくいことが分かった。



図3「文化祭の様子」(Dalle-3)



図4 工夫1を適用して生成した画像 (Dalle-3)



図5 工夫2を適用して生成した画像 (Dalle-3)

【その他の得られた結果】

小説の一節をプロンプトとして使用したところ、まとまりのある印象の画像が高確率で生成された。図6では、川端康成の「雪国」の一節をプロンプトとして使用した。「文章を通じて情報を他者へ伝える働きをする」という点でプロンプトと小説は共通していると言えるが、小説の一節をそのままプロンプトとして使用した場合と、同じ意味になるように要素を羅列したプロンプトで画像を生成した場合では、前者の方がより雰囲気のある画像が生成された。



図6 川端康成『雪国』より「明かりをさけてゆくり雪を踏んで来た男は、襟巻で鼻の上まで包み、耳に帽子の毛皮を垂れていた。」(Dalle-3)

結論



AI生成画像に違和感を感じるのは、生成画像が持つ不自然さによって人が不気味な印象を持つからである。また、その違和感の多くは細かい状況の指定やイラスト調の画像を指示するなどプロンプトの工夫で解決可能である。

今後の課題



AIは奥行きや構造理解を必要とする三次元的な画面構成が苦手である。3Dモデルを使用することで改善できないかと思ったが、個人で取り組むには技術面で現実的ではなく断念した。今後何かしらの形でまた生成AIの三次元構造理解について研究したい。

【参考文献】

- [1] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa, Large language models are zero-shot reasoners. In Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22). Article 1613, 2022
- [2] ジークムント・フロイト、不気味なもの、フロイト全集17、1919-22年、岩波書店、2006
- [3] 森政弘、不気味の谷、エナジー誌、Vol.7、No.4、pp.33-35、1970