

ChatGPT を用いた 対話型小説生成アプリケーションの提案

及川郁也[†] 武田敦志[†]

近年、小説を AI によって作成する研究やアプリケーションが注目されている。この研究では、文章生成 AI とのインタラクションによって小説を作成するアプリケーションを作成し、インタラクションの有無によるシナリオの変化と、イメージ画像を生成する際の翻訳の有無による違いを評価した。

Application for Interactive Novel Generation using ChatGPT

Fumiya Oikawa[†] and Atsushi Takeda[†]

In recent years, research and applications that use AI to create novels have been attracting attention. In this study, we created an application that creates novels through interaction with a text-generating AI, and evaluated changes in scenarios with and without interaction, as well as differences with and without translation when generating images.

[†] 東北学院大学教養学部情報科学科
Department of Information Science, Tohoku Gakuin University

1. はじめに

近年、小説を AI によって作成する研究やアプリケーションが注目されている。2015 年の研究では AI との共作が星新一賞の一次審査を通過し^[1]、2022 年には AI を執筆に取り入れた個人の作品が星新一賞の優秀賞を獲得した。星新一賞は 2012 年に新設された、ショートショートおよび短編小説を対象とした公募文学賞である。特徴として、理系的な発想に基づいた課題にて募集されること、また人間以外(AI 等)の応募作品も受け付けていることから、執筆 AI 研究の対象になることが多い。現在の文章生成 AI を用いることで、人間が作成した文章と同じような自然な文章を自動的に作成することが可能となった。

そこで、この研究では、対話形式の文章生成 AI を活用することにより、ユーザーが意図したシナリオの小説を自動的に生成するアプリケーションを開発する。このアプリケーションでは、対話型生成 AI である ChatGPT に対して小説生成を目的とした入力を行うことで新しい小説を生成する。ただし、ChatGPT に対する入力はユーザーが作成するのではなく、ユーザーの操作にもとづきアプリケーションが自動的に適切な入力を作成する。この仕組みにより、ChatGPT の操作方法を知らないユーザーであっても小説を生成することが可能となっている。また、シナリオ分岐に関する選択肢を表示し、ユーザーがその選択肢からシナリオ進行を選ぶことにより、小説の内容が破綻しないようにシナリオを制御することが可能となる。さらに、ユーザーが小説の話の流れを制御できるため、出だしが同じ小説であっても、それぞれのユーザーの選択によって異なる結末の小説が生成される。

本稿では、対話型小説生成アプリケーションの構造について説明し、その有用性を実行結果から考察する。特に、ユーザーと小説生成 AI とのインタラクションがあることにより、小説のシナリオを制御することが可能となり、独自の小説が生成できることを確認する。さらに、このアプリケーションでは、小説の内容を画像生成 AI である DALL-E3 を用いることにより、小説に適した画像を自動的に生成する。このとき、小説の内容を英語に翻訳することにより、より小説に適した画像が生成できることを確認する。

2. 既存研究

以前より生成 AI を用いた小説生成の研究が行われており、研究[1]では「きまぐれ人工知能プロジェクト 作家ですよ」にて作成された AI と人間の共作が星新一賞の一次審査を通過したことが述べられている^[1]。一方、研究 [2]では、大規模文章データセットを学習させた自然言語モデルである GPT モデルが特定の自然言語処理のタスクに特化したモデルよりも優れた結果をもたらす場合があることが述べられており、この GPT モデルを使うことにより従来よりも自然な小説を生成できると考えられる。

ただし、この論文の GPT モデルは 1 億 1700 万個のパラメータのモデルであるが、現在広く使用されている自然言語モデル GPT-3.5 のパラメータ数は 3550 億個であることを考えると、GPT-3.5 を用いるとより自然な小説を生成できる可能性がある。

また、文章から画像を生成する AI の研究開発も行われており、OpenAI から画像生成 AI である DALL-E3 が発表されている。DALL-E3 は CLIP モデルと拡散モデルを使用してテキストから画像を生成する^{[3][4][5]}。そこで、本研究で開発するアプリケーションでは、画像生成 AI である DALL-E3 を活用することにより、小説に挿入するイメージ画像を生成する。

一方、研究[6][7]のような研究では、生成モデルに対して効果的なプロンプトを渡すために、元のプロンプトから曖昧な部分を減らし、具体的なプロンプトに修正する手法が提案されている。このアプリケーションは日本語での利用を想定しており、比較的プロンプトが曖昧になり生成の精度が安定しないことが予想されるため、このアプリケーションでは画像生成 AI に効果的にプロンプトを与えるために、一度日英翻訳を行いプロンプトを与える。

3. 対話型小説生成アプリケーション

本稿で提案する対話型小説生成アプリケーションは、ユーザーに対してシナリオ生成のための選択肢を提示するインターフェースであるため、ChatGPT などの文章生成 AI の操作方法を知らないユーザーであっても小説を生成できる特徴がある。まず、このアプリケーションを実行すると、ユーザーは小説のジャンルを選択することができる。アプリケーションはユーザーの選択にもとづいて ChatGPT に対する入力（プロンプト）を作成し、ChatGPT に小説の生成リクエストを送信する。このときの ChatGPT からの応答を画面に表示することにより、ユーザーは生成された小説を確認することができる。また、小説のシナリオの要所において、ユーザーに対して主人公の行動などの選択肢を表示する。ここでは、ユーザーが選んだ選択肢によって ChatGPT への入力を自動的に変更し、ユーザーが選んだ選択肢にもとづいてシナリオが進行するように文章が生成される。

また、本アプリケーションは小説を生成する前にあらかじめシナリオを指定し、そのシナリオにしたがって文章を生成するように指示することができる。ただし、シナリオを指示した場合であっても、シナリオの要所における選択肢は提示され、このときのユーザーの選択によって小説の文章は変化する。そのため、同じシナリオを指定して小説を生成したとしても、ユーザーの選択によって異なる小説が生成される。

さらに、本アプリケーションでは、DALL-E3 を用いて小説に挿入する画像を生成する。ただし、DALL-E3 は英語の文章を最も多く学習した画像生成モデルであり、日本語の文章を入力した場合は意図しない画像が生成されることが多い。そこで、生成さ

れた小説を機械翻訳サービスである DeepL を用いて英語に翻訳し、この英語文章を DALL-E3 に入力することにより、日本語の文章を入力した場合よりも正確に意図した画像を生成する。

4. 実行結果

4.1 アプリケーションの実行結果

アプリケーションの実行結果を図 1 に示す。アプリケーションを起動すると、シナリオの開始前にシナリオのジャンルを選択する画面が表示される。文章の生成を開始すると、選択したジャンルに応じてシナリオが進行し、シナリオの要所毎にユーザーに対してどんな行動を起こすかの回答が促される。シナリオ進行中は、ユーザーが回答し、その回答に応じてさらにシナリオが展開される。これをオプションで指定した回数分繰り返す。また、インタラクションの文章を入力せずに回答すると、AI によってシナリオが順当に進行される。シナリオ終了画面では、そのシナリオ全体のまとめが出力される。



図 1 アプリケーションの実行画面

4.2 インタラクションの有無による違い

インタラクションの有無による違いを調査するため、星新一の「悪魔」のあらすじを入力したときのシナリオ生成について検証する。具体的なあらすじとして、「エス氏が氷の張った湖の上で釣りをしていると、古いツボが釣れた。そのツボの中からエス氏の願いをかなえるという悪魔が出てくる。エス氏は金貨をくれと言うと、悪魔は一枚の金貨を渡した。エス氏がもう少しと言うと悪魔は一握りの金貨を渡した。そうやってエス氏が頼み続け金貨の山ができると、悪魔は『このぐらいでやめたらどうだ』という。しかしエス氏はあと一回と言い、その通りに悪魔は金貨を山の上に置いた。

すると、金貨の重みで氷にひびが入り始め、エス氏は大急ぎで岸へ逃げた。ほっとしながら振り返ると、金貨のツボも、かん高い声をあげている悪魔も、すべて湖の底へと消えていった。」とタイトル画面のあらすじに入力し、回数を5、ジャンルをショートショートに設定する。

図2にインタラクション無しの場合でシナリオを作成した結果を示す。インタラクション無しの場合では、入力したあらすじ通り、悪魔に対する願い事が質問され、回答もあらすじ通りに行われた。物語の終わりもあらすじの通り「すべて湖の底に消えていった。」となっており、あらすじ通りにシナリオが進行していることが確認できる。

図3にインタラクション有りの場合でシナリオを作成した結果を示す。インタラクション有りの場合では、質問に対してあらすじとは異なる回答、例えば悪魔に対する願い事に「金貨」ではなく「釣りの竿」などと回答を行った。一回答ごとにシナリオが少しずつ変化し、それに伴い質問の内容も変化している。物語の終わりのシナリオは「悪魔がエス氏に魚の知識を教えた」となり、回答によって物語が変化していることが確認できる。

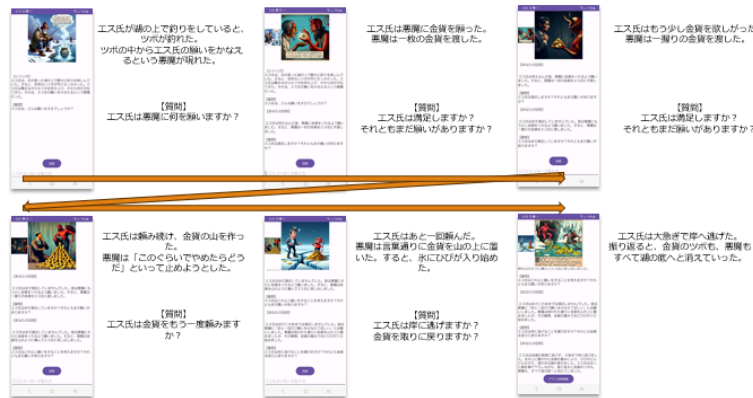


図2 インタラクション無し

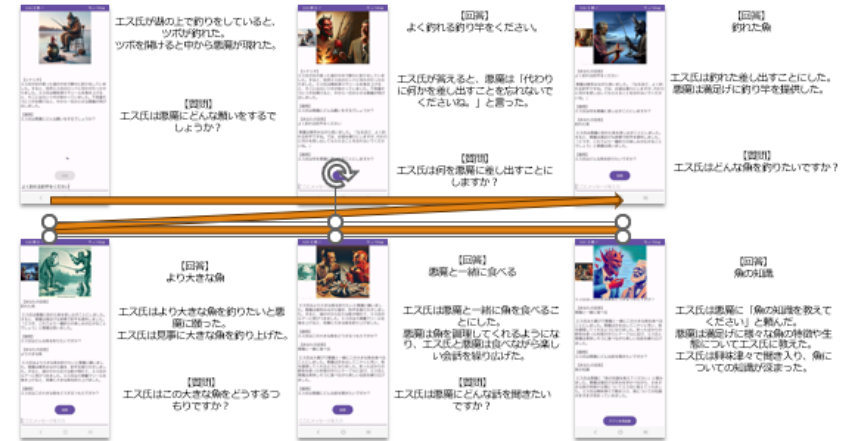


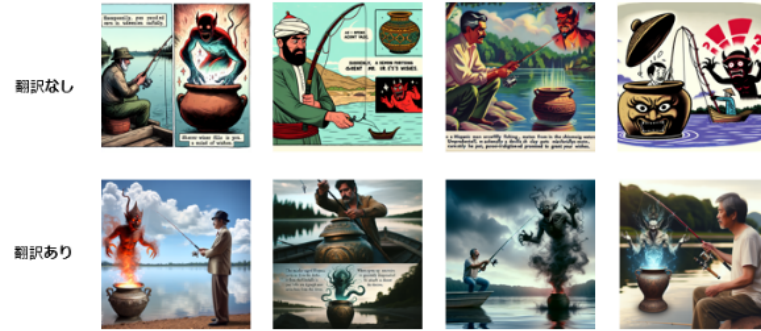
図3 インタラクション有り

4.3 翻訳による画像生成の精度比較

次に、翻訳による画像生成の精度比較を行う。本アプリケーションでは生成された文章をプロンプトとして画像生成 AI に入力しているが、この際に DeepL API を使用し日英翻訳を行ったものを画像生成 AI に入力している。同じ文章を入力する際に翻訳を行った場合と行わなかった場合の比較を図4に示す。プロンプトには、先述した星新一の「悪魔」から文章を抜き出して使用する。

相違点としては、主に画像内の文章の出現しやすさや構図の安定しやすさがみられる。翻訳を行わなかった場合は漫画のように、画像内に台詞のような部分が出現することやコマ割りされていたりすることが比較的多かった。翻訳を行った場合は漫画のようになることこそあれど、比較的登場人物の配置や行動が安定するような、プロンプト通りに画像が出力されることが多かった。

エス氏は湖の上で釣りをしていました。すると、古い壺が釣れました。ツボを開けると、中からエス氏の願いをかなえるという悪魔が現れました。



エス氏が頼み続け金貨の山ができると、悪魔は「このくらいでやめたらどうだ」という。



図 4 翻訳の有無による画像生成の精度の比較

5. まとめ

本稿では、ChatGPT を用いた対話型小説生成アプリケーションを提案した。このアプリケーションでは、ユーザーに対して小説生成のための選択肢を提示し、ユーザーが選んだ選択肢にもとづいて小説を生成するための ChatGPT への入力を自動的に作成する。そのため、ChatGPT などの生成 AI の利用方法を知らないユーザーであって

も、このアプリケーションを利用することで小説を生成できる。また、ユーザーはシナリオの要所において主人公などの行動を選択することができるため、ユーザーが意図したシナリオの小説を生成することが可能となる。また、画像生成 AI である DALL-E3 を用いることで小説に挿入する画像を生成することもできる。

本稿で提案したアプリケーションは ChatGPT や DALL-E3 などの現在の生成 AI を活用することで小説を生成する。しかし、現在も生成 AI の研究は行われており、近い将来により高い性能の生成 AI が発表される可能性は高い。これらの将来の生成 AI に対応したアプリケーションのあり方を模索することが今後の課題として挙げられる。

参考文献

- 1) 松原仁."AI が小説を書く", 情報法制レポート. 2022, 2 巻, p.20-24.
- 2) Radford, Alec, et al. "Improving language understanding by generative pre-training." (2018).
- 3) Ramesh, Aditya, et al. "Zero-shot text-to-image generation." International Conference on Machine Learning. PMLR, 2021.
- 4) Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with clip latents." arXiv preprint arXiv:2204.06125 1.2 (2022): 3.
- 5) Ramesh, Aditya, et al. "Improving Image Generation with Better Captions", OpenAI, 2023, <https://cdn.openai.com/papers/dall-e-3.pdf>, (2023).
- 6) Baek, Seungho, et al. "PromptCrafter: crafting text-to-image prompt through mixed-initiative dialogue with LLM." arXiv preprint arXiv:2307.08985 (2023).
- 7) Lian, Long, et al. "LLM-grounded Diffusion: Enhancing Prompt Understanding of Text-to-Image Diffusion Models with Large Language Models." arXiv preprint arXiv:2305.13655 (2023).