

生成 AI による物語教材の視覚的教材作成プロセスの設計と プロンプト評価の課題

Designing a Visual Instructional Material Generation Process for Narrative Texts Using Generative AI: Issues in Prompt Evaluation

應供 亮生^{*1}, 緒方 思源^{*1} 小川 修史^{*1}

Ryosho OGU^{*1}, Shigen Fang OGATA^{*1}, Hisashi OGAWA^{*1}

^{*1} 兵庫教育大学 学校教育研究科

^{*1} Graduate School of Education, Hyogo University of Teacher Education

あらまし：本研究では生成 AI を用いて物語教材の特定場面を視覚化する三段階の画像生成プロセスを設計し、プロンプト評価に関する課題を検討した。プロンプト評価を生成 AI に委ねた際に生じる文字数バイアスを検証した結果、情報量の多寡が評価得点に強い相関を示すことが確認された。さらに、生成画像には情報過多に起因する矛盾も観察され、人間の認知処理における「状況モデル構築の困難性」との類似性が示唆された。

キーワード：生成 AI, 画像生成, プロンプト設計, 認知心理学, 教材開発

1. はじめに

2024 年 12 月に文部科学省から公表された「初等中等教育段階における生成 AI の利活用に関するガイドライン (Ver.2.0)」⁽¹⁾ では、教育目的での生成 AI 活用に一定の方向性が示され、実践への検討が広がっている。生成 AI、とりわけ大規模言語モデル (LLM) のマルチモーダル化が進む中、教育においても AI を活用した教材生成や学習支援の設計が検討されつつある。

国語教育において、これまでの物語教材の読解は論理的な言語的理解を中心に構成されてきた。しかし、認知心理学の観点からは、読解とは文章内の情報を基に登場人物の心情や場面の構造、時間、空間的な状況などを高次で統合した、「状況モデル (Situation model)」を構築する過程であるとされる⁽²⁾⁽³⁾。このモデルの構築が困難な場合、視覚的イメージによる補助が理解を促進することが知られている⁽³⁾。メイヤーによるマルチメディア学習の認知理論⁽⁴⁾においてもテキストと画像の併用が理解を助け

るという知見が示されている。それらの知見を踏まえ、著者らは文章と画像を統合した認知モデル⁽⁵⁾を提案した。

本研究では、こうした理論的背景を踏まえ、物語教材に対して文章から描画方針 (登場人物、場面、心情、時間、空間など) を LLM で抽出し、それを画像生成のプロンプトとして用いて視覚的教材を作成する一連のプロセスを、教育支援システムとして設計した (図 1)。このプロセスは、従来の挿絵では実現が難しかった柔軟な視覚支援を可能にし、学習者の認知特性や発達段階に応じた教材のカスタマイズを実現する。

さらに本研究では、画像生成の際に中心的な役割を果たすプロンプトに着目し、その評価を LLM に委ねた場合に生じる構造的バイアスについても検討した。特に、出力されたプロンプトの情報量が評価結果に影響を与える傾向⁽⁶⁾が確認されており、LLM を活用した教材開発においては、プロンプト設計とその評価手続きの妥当性が重要な検討課題となる。

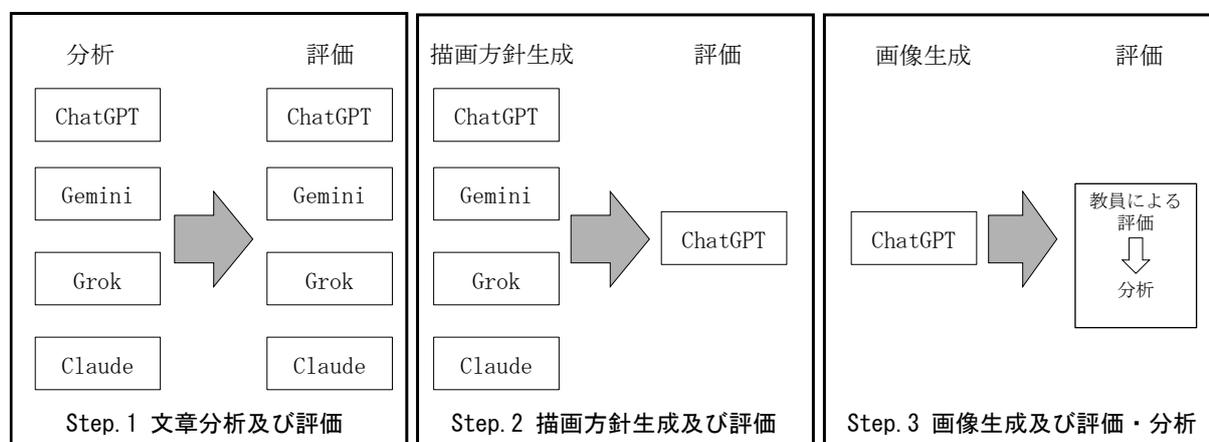


図 1 生成 AI を用いた物語教材に対する視覚支援教材作成プロセスと検証手順

2. 画像生成プロセス

本研究では、物語教材の読解支援を目的として、文章からの情報抽出と画像生成を連携させた支援プロセスを構築した。複雑なタスクの構造を明確化し、問題を把握しやすくするために ReAct アプローチ⁽⁷⁾を導入し、Step.1 物語文全体の分析、Step.2 特定場面の描画方針（プロンプト）の生成、Step.3 画像生成、という3段階で構成した。この時、Step.1 の分析や Step.2 の描画方針に対する評価を生成 AI が行う設計とした。また、AI が自己生成物を高く評価しやすいという自己バイアス⁽⁸⁾の回避のため複数の LLM で比較検討を行った。

3. バイアスの検討

生成 AI による描画方針（プロンプト）の評価において、出力されたテキストの文字数が評価に与える影響、すなわち情報量に基づくバイアスの存在を検証した。

2種類の物語教材（「ごんぎつね」と「走れメロス」）に対する4種類の LLM（ChatGPT, Gemini, Grok, Claude）による文章分析と描画方針作成の生成文書に対し、同じ4種類の LLM による評価を行った。そこで出力された文字数と評価との関係を検討するため、相関分析を実施した。評価プロンプトによって評価得点にばらつきがあったため、前処理としてセットごとに z スコアによる標準化を行った。分析の結果、表1に示す通り、文字数と評価得点の間には強い正の相関が認められた（文章分析 $n = 64, r = .777, p < .001$ ；描画方針 $n = 32, r = .794, p < .001$ ）。これは、出力された文章が長文になるほど、AI による評価得点も高くなる傾向があることを示している。

この結果は、文字数＝情報量そのものに AI が引きずられる可能性を示唆しており、プロンプトの設計段階において、必要以上の情報追加や過剰な冗長性が評価に影響しうることを示している。生成 AI を評価者として活用する場合には、このバイアスに対する設計上の対策が求められる。

表1 文字数と評価の相関

	n	相関係数 (r)	p
文章分析 文字数-評価	64	.777	<.001
描画方針 文字数-評価	32	.794	<.001

4. 教員による評価

生成した画像を用いて教員を対象にオンラインフォームによる調査を実施し、回答を得た。評価の結果、生成 AI が出力した画像には、「視覚化により多様な特性を持つ児童にも理解が促進される」といった肯定的な評価が多く得られた。一方で、「絵に引きずられ、自らの想像力で読解する努力を妨げる懸念」も指摘され、教材提示の工夫が今後の課題となる。なお、調査の詳細については紙幅の都合により本稿では割愛し、発表にて報告する。

5. まとめと今後の課題

本研究では、物語教材における生成 AI を用いた視覚支援プロセスを構築し、AI が評価する時の文字数バイアスの存在を実証的に検討した。プロンプトの情報量が多いほど高評価を得る傾向は設計上の対策の必要性を示している。

また、情報量が過剰な描画方針は、表現空間の限界による矛盾やノイズの増加を引き起こす可能性が考えられる。実際、AI によって高く評価された描画方針を用いても、矛盾がある画像が生成されることが確認された。これは、人間が状況モデルを構築する際に直面する注意資源やワーキングメモリの限界、情報統合の困難さと類似しており、情報量の最適化と情報の取捨選択の難しさは AI と学習者に共通する課題といえる。つまり、「本当に必要な情報」を見出し、情報を統合する能力の育成が学習支援の要点だと考えることができる。

本研究では、プロセスの段階化によって ReAct アプローチを導入したことで、より焦点化されたプロンプトの生成と評価が可能になったが、プロンプトの評価を AI に委ねる際のバイアスや、生成画像のばらつきといった技術的な限界も残る。加えて、視覚情報が児童の想像を制限する可能性があることも教員から指摘された。

今後は描画方針の情報量の最適化、生成 AI の出力精度を高めるプロンプトの設計、さらには教育現場への実践的な導入可能性について、総合的に検討を進める必要がある。

参考文献

- (1) 文部科学省：“初等中等教育段階における生成 AI の利活用に関するガイドライン (Ver.2.0)”，https://www.mext.go.jp/content/20241226-mxt_shuukyo02-000030823_001.pdf (参照 2025.5.20)
- (2) Kintsch, W.: “Text Comprehension, Memory, and Learning”. *American Psychologist*, 49(4), 294-303 (1994)
- (3) Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., and Curiel, J. M.: “Constructing multidimensional situation models during reading”. *Scientific Studies of Reading*, 2(3), 199-220 (1998)
- (4) Mayer, R.E. “Multimedia learning”, Cambridge University Press, Cambridge (2002)
- (5) 應供亮生・緒方思源・小川修史：“認知モデルに基づく画像が文章理解に及ぼす感性的影響の理論的検討および画像生成 AI による教育支援の可能性”，第20回日本感性工学会春季大会 (2025)
- (6) 應供亮生，緒方思源，小川修史：“生成 AI による物語教材への学習支援の可能性”，第148回全国大学国語教育学会（三重大会）(2025)
- (7) Yao, Shunyu, Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. “React: Synergizing reasoning and acting in language models.”, *International Conference on Learning Representations* (2023)
- (8) Xu, W., Zhu, G., Zhao, X., Pan, L., Li, L., and Wang, W. Y. “Pride and prejudice: LLM amplifies self-bias in self-refinement”, *arXiv preprint, arXiv:2402.11436*(2024)