

AI との共創で生まれた俳句の美しさに迫る

Discover beauty of haiku created through co-creation with AI

檀割 仁平[†], 上田 祥行[‡], 尹 優進[†], 野村 理朗[†]
 Jimpei Hitsuwari, Yoshiyuki Ueda, Yun Woojin, Michio Nomura

[†]京都大学大学院教育学研究科, [‡]京都大学人と社会の未来研究院
 Graduate School of Education, Kyoto University Institute for the Future of Human Society
hitsuwari.jimpei@gmail.com

概要

本研究では、385名が人間もしくはAIによって創作された俳句の評価と判別を行った。人間とAIの共創俳句の美しさが最も高く評価され、人間作とAI作の美しさは同等であった。また、参加者は人間とAIの作品を判別できなかった。これは人間とAIの共創が優れた創造性を持つことを示している。また、AI俳句の美的評価が高いほど、人間が作ったと誤認される傾向が見られた。これはAI芸術がアルゴリズム嫌悪により過小評価されている可能性を示唆している。

キーワード: AI art, haiku poetry, aesthetic evaluation, creativity, human-in-the-loop

1. はじめに

近年AIアートが隆盛しており、midjourneyに代表される生成系AIの一般化によって、その勢いはさらに増し続けている(e.g., Elgammal et al., 2017, レビューは Cetinic & She, 2022を参照)。しかしながら、絵画や写真などの視覚芸術に比べ、詩や文学といった言語芸術ではAIによる創作はまだ発展途上である。言語芸術を扱った最近の研究では、AI詩を、AIが生成したものから人間が選択したHuman-In-The-Loop (HITL)と生成時に人間が介入しなかったHuman-Out-of-The-Loop (HOTL)に分類した(Köbis & Mossink, 2021)。その結果、HITLの詩の好感度は、HOTLの詩よりも高く、人間が創作した詩と拮抗していた。さらに、HITL詩と人間創作詩は、HOTL詩と人間創作詩よりも区別が困難であることがわかった。こうした動向を背景に、本研究は、俳句を用いてKöbis & Mossink (2021)の再現と拡張を試みた。その特徴は、音数が17音(5-7-5)に制限された世界最短の詩型である俳句に着眼したことである。俳句は情報量が少ないため曖昧性を伴うことが多く、AIが曖昧な俳句を生成しても、読者は自分が想像できるイメージに従って自由に解釈する余地がある(Hitsuwari & Nomura, 2022a, 2022b)。そのため、俳句のこのような特徴が、先行研究とは異なる結果を導く可能性がある。また、先行研究では、好感度など単項目による評価しか行われてこなかったため、本研究では

哲学や心理学で美と関連すると考えられてきた多数の因子(Brielmann et al., 2021)を参照し、人間俳句とAI俳句、そしてその共創俳句を多面的に検討した。

2. 方法

本研究には385名(男性191名, 女性194名, $M_{age} = 40.9, SD = 10.1$)が参加し、2つの課題が行われた。1つは評価課題で、参加者は、人間が創作した俳句20句、HOTLによって創作された俳句10句、HITLによって創作された俳句10句の計40句を、21項目の観点から7件法で評価した。21項目は、美しさ、デジャブ、イメージの鮮明度、感情価、覚醒度、畏敬の念、ノスタルジア、新奇性、共感、意図、喜び、鑑賞し続けたい度合い、生き生きしている度合い、普遍的な美、作品とのつながりの度合い、憧れ、欲望からの解放、マインドワンダリング、驚き、理解したい度合い、ストーリーであった(詳しくは、Brielmann et al., 2021)。人間俳句は、歳時記から選択され、プロの俳人が作った俳句であった。これらの人間俳句は、俳句に対する特別な事前知識を持たない8名が参加した予備調査により、見慣れないものであることが確認された。AI俳句は、Long Short-Term Memoryアルゴリズム(川村他, 2021; 横山他, 2019)に基づく俳句生成システムによって作成され、指定した季語に対して大量生成されたものから、ランダムに選択されたか(HOTL)、著者らを含む3名によって高く評価されたものが選択された(HITL)。もう1つは判別課題で、参加者は40句の作者が人間かAIかを判別した。これらの課題に取り組む順序は、参加者間でカウンターバランスされた。最後に、参加者は、AIやロボットとの相互作用に関連することが示されているアニミズムや共感性など(Darling et al., 2015)の個人特性に関する質問紙に回答した。

3. 結果

美しさの評価はHITL俳句で最も高く($M = 4.56, SD$

= .74), 人間俳句 ($M=4.15, SD=.75$) と HOTL 俳句 ($M=4.14, SD=.78$) では同等であった ($F(384, 2)=212.41, p=.00, \eta^2=.06$; Figure 1)。同様に、イメージの鮮明度、畏敬の念などの多くの評価項目において、3条件間で有意差があった (Table 1)。

Figure 1. 人間俳句, HOTL 俳句, HITL 俳句の美の得点

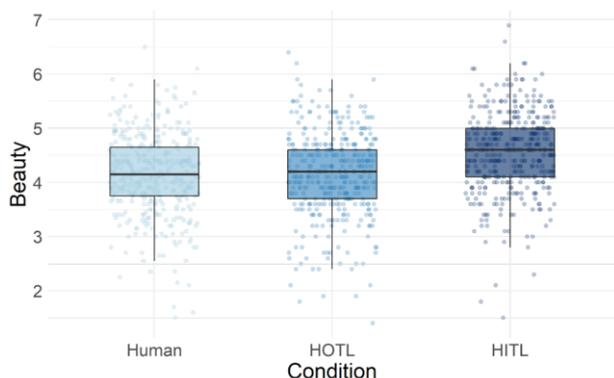


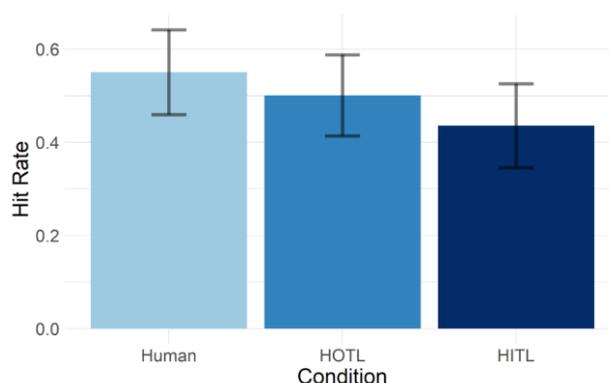
Table 1. 各評価項目における多重比較の結果

評価項目	多重比較の結果
美しさ	人間 =HOTL < HITL
デジャブ	HOTL < 人間 < HITL
イメージの鮮明度	HOTL < 人間 = HITL
感情価	HOTL < 人間 < HITL
覚醒度	HOTL < 人間 < HITL
畏敬の念	人間 =HOTL < HITL
ノスタルジア	HOTL < 人間 < HITL
新奇性	HOTL < 人間 = HITL
共感	HOTL < 人間 < HITL
意図	HOTL < 人間 < HITL
喜び	HOTL < 人間 < HITL
鑑賞し続けたい度合い	HOTL < 人間 < HITL
生き生きしている度合い	HOTL < 人間 < HITL
普遍的な美	人間 =HOTL < HITL
憧れ	HOTL < 人間 < HITL
欲望からの解放	人間 <HOTL < HITL
マインドワンダリング	人間 =HOTL < HITL
作品とのつながりの度合い	人間 =HOTL < HITL
驚き	HOTL < 人間 = HITL
理解したい度合い	HOTL < 人間 < HITL
ストーリー	人間 =HOTL < HITL

判別課題では、人間俳句, HOTL 俳句, HITL 俳句の正答率 (人間が作った俳句を人間が作ったと, AI が作

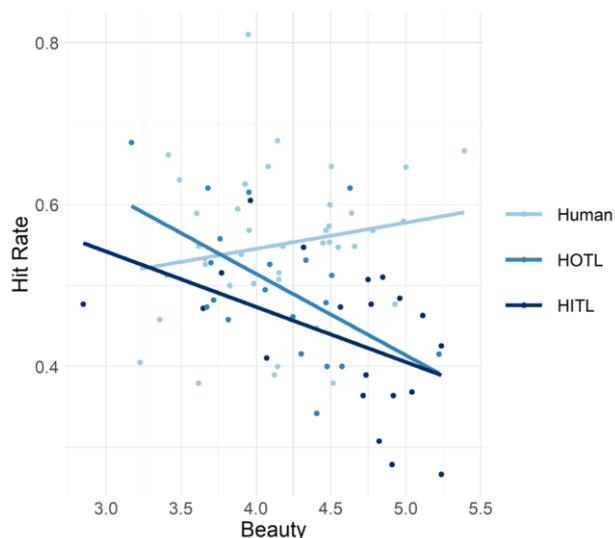
った俳句を AI が作ったと正しく判断する確率) は、それぞれ .55, .50, .43 であった (Figure 2)。チャンスレベル (0.5) と比較したところ、人間俳句の正答率はチャンスレベルより有意に高かったが ($t(39) = 3.51, p = .001$), HITL 俳句では有意に低かった ($t(19) = -3.19, p = .005$)。HOTL 俳句での正答率は、チャンスレベルと差がなかった ($t(19) = .03, p = .98$)。これらの結果は、参加者が HOTL 俳句について、人間が作ったものか AI が作ったものかを区別できず、HITL 俳句の作者は AI ではなく人間であると考えていたことを示している。

Figure 2. 条件ごとの判別課題正答率



また、美しいものは AI ではなく人間の手によって作られると考える人が多ければ、ヒット率と美しさ評価には有意な関係性があることが考えられる。各俳句の美しさ評価と正答率の関係を探索的に検討したところ、人間俳句では正の相関が見られたものの有意ではなかった ($r(39) = .18, p = .26$)。一方、HOTL 俳句と HITL 俳句では有意な負の相関が見られた (それぞれ $r(19) = -.54, p = .01$ と $r(19) = -.47, p = .04$, Figure 3)。AI 俳句の美しさの評価が高いほどヒット率が低いのは、参加者が「美しい俳句ほど人間が作った可能性が高い」と考えていることを反映している。

Figure 3. 条件ごとの美的評価と正答率の関係



4. 考察

本研究の結果、AI と人間の共創による俳句の美しさが最も高いと評価された。この結果は、人間が創作した詩が最も好まれたという先行研究 (Köbis & Mossink, 2021) と異なっていた。その原因の 1 つは、詩のスタイルの違いにあると考えられる。俳句は世界で最も短い詩の形式であり、5-7-5 音節の形式や季語を含めるといった明確な規則が特徴であるため、AI 作品の質が高まりやすかったかもしれない。また、Köbis & Mossink (2021) では、AI が作成した詩の最初の 2 行は人間が作成した詩と同じものであり、両方の詩を一緒に並べて、参加者にどちらが好きかを尋ねた (すなわち、二者択一強制選択)。二者択一は、本研究で採用した絶対評価に比べ、2 つの作品の微妙な違いに比較的敏感であるため、この方法論の違いによっても本研究の結果との相違を説明できるかもしれない。

また、参加者は、人間が作った詩と AI が生成した詩を識別することができなかった。この結果の背景には、俳句の生成技術が向上していることが大きな要因として考えられる (川村他, 2021; 横山他, 2019)。さらに、俳句は非常に短く、イメージに依存するなどの特徴があるため、人間が介在しない AI が生成した俳句でも、人間俳句との判別に支障をきたした可能性がある。

AI が生成した俳句における美的評価と正答率の間には負の相関があった。このことは、人々がアルゴリズム嫌悪を持っている可能性を示唆している (Burton et al., 2020)。アルゴリズム嫌悪は、人間と AI の詩を比較するという研究文脈において重要な概念であり、

Köbis & Mossink (2021) は、「車の運転」など異なるタスクに対して、そのタスクを実行するために人間と AI のどちらをどの程度信頼するかを示すアルゴリズム嫌悪尺度 (Castelo et al., 2019) を用いた。その結果、アルゴリズム嫌悪と人間の詩の選択率は弱いながら一貫した相関を見せ、詩を作るアルゴリズムに対する参加者の見方は嫌悪的であることが示唆された。今回の結果では、参加者が、AI の俳句生成の技術を見下していたと言えるかもしれない。本研究は、アルゴリズム嫌悪と作者判別、美しさ評価の関係を示した最初の研究であり、AI 芸術に対する現代人の態度が反映されていると考えられる。

5. まとめと展望

本研究では、AI が生成した俳句と人間が創作した俳句の美的評価と作者判別が行われた。美しさの評価は、人間が作った俳句も AI が作った俳句から無作為に選ばれた俳句も同等であった。一方、AI が作った俳句であっても、人間が選んだものは美しさの評価は最も高かった。さらに、人間が作った俳句と AI が作った俳句の判別は困難であった。また、AI 俳句では、美しさの評価と判別成績の間に負の相関が見られた。全体としてこれらの結果は、情報が最小限である俳句において、AI の芸術的質が人間に匹敵するレベルに達しており、人間と AI の共創によって、より創造的な作品を生み出すことができることを示唆している (Booten & Gero, 2021)。

本研究の今後の展望として、2 つの方向性を検討している。1 つ目は、俳句の熟達者を対象とした実験を行うことである。本研究では、HITL 俳句を選択したのも、俳句の評価をしたのも非熟達者であった。非熟達者は、イメージが鮮明で、曖昧性の低い俳句を好むことが知られている (Hitsuwari & Nomura, 2022a, 2022b)。その一方で、歳時記に載っているような俳人の俳句を鑑賞するには、俳句の知識や季語の理解などのリテラシーが求められる可能性がある。そのために熟達者が HITL 俳句を選択したり、評価したりすることで、AI アートが熟達者の創造行為にどれほど迫っているかを検討できるだろう。2 つ目は、AI が使用した句中の言葉に着眼して、俳句から人間が感じる美しさについてシステムティックな検討を行うことである。例えば、単語と単語の間の意味の距離を計算し、意味距離と作品の美しさ評価を検討することが可能である (c.f.,

Wang et al., 2023)。AI によって大量に、特徴量を統制した俳句を制作し、これらと評価を結びつけることで、芸術作品から喚起される美的感覚や、芸術鑑賞中のどのような要因が人々の印象に影響を与えるのかが明らかになることが期待される。

6. 謝辞

本研究で使用した AI 俳句は北海道大学調和系工学研究室の山下倫央准教授にご提供頂きました。この場を借りて感謝申し上げます。

文献

- [1] Briemann, A. A., Nuzzo, A., & Pelli, D. G. (2021). Beauty, the feeling. *Acta Psychologica*, 219, 103365. <https://doi.org/10.1016/j.actpsy.2021.103365>
- [2] Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239. <https://doi.org/10.1002/bdm.2155>
- [3] Booten, K., & Gero, K. I. (2021). Poetry machines: Eliciting designs for interactive writing tools from poets. *Creativity and Cognition*, 51, 1–5. <https://doi.org/10.1145/3450741.3466813>
- [4] Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, 56(5), 809–825. <https://doi.org/10.1177/0022243719851788>
- [5] Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(2), 1–22. <https://doi.org/10.1145/3475799>
- [6] Darling, K., Nandy, P., & Breazeal, C. (2015). Empathic concern and the effect of stories in human-robot interaction. *24th IEEE International Symposium on Robot and Human Interactive Communication*, 770–775. <https://doi.org/10.1109/ROMAN.2015.7333675>
- [7] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative adversarial networks generating “art” by learning about styles and deviating from style norms. *ArXiv*. <https://arxiv.org/abs/1706.07068>
- [8] Hitsuwari, J., & Nomura, M. (2022a). How individual states and traits predict aesthetic appreciation of haiku poetry. *Empirical Studies of the Arts*, 40(1), 81–99. <https://doi.org/10.1177/0276237420986420>
- [9] Hitsuwari, J., & Nomura, M. (2022b). Beauty and ambiguity: Japan–Germany cross cultural comparison on aesthetic evaluation of haiku poetry. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. <https://doi.org/10.1037/aca0000497>
- [10] Hopkins, J., & Kiela, D. (2017). Automatically generating rhythmic verse with neural networks. *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 1, 168–178.
- [11] 川村 秀憲・山下 倫央・横山 想一郎 (2021). 人工知能が俳句を詠む: AI 一茶くんの挑戦 オーム社
- [12] Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114. <https://doi.org/10.1016/j.chb.2020.106553>
- [13] Lau, J. H., Cohn, T., Baldwin, T., Brooke, J., & Hammond, A. (2018). Deep-speare: A joint neural model of poetic language, meter and rhyme. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.48550/arXiv.1807.03491>
- [14] Wang, X., Bylinskii, Z., Hertzmann, A., & Pepperell, R. (2023). A computational approach to studying aesthetic judgments of ambiguous artworks. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. <https://doi.org/10.1037/aca0000579>
- [15] 横山 想一郎・山下 倫央・川村 秀憲 (2019). 深層学習を用いた俳句の生成と選句. *人工知能*, 34(4), 467–474. https://doi.org/10.11517/jjsai.34.4_467