2025. Proc Ling Soc Amer 10(1). 5885. https://doi.org/10.3765/plsa.v10i1.5885.

Are we biased against AI-made haiku poems?

Koichi Tateishi & Shinobu Mizuguchi*

Abstract. Haiku is a traditional Japanese unrhymed short poem of 17 syllables with 'a seasonal word' and 'a cut letter' to divide 17 syllables into smaller units of [5][75], [57][5], or [5][7][5]. Due to its brevity, haiku often lack arguments and predicates, and a cut letter separates the text into smaller parts, making it difficult to establish local coherence. However, Gilbert (2024) claims that the play of disjunction and coherence is a taproot of haiku. Haiku, or a version of it, is now written by poets worldwide and even by AI. This paper examines how humans evaluate AI-generated haiku poems through experiments.

Recently, AI-generated text has made much progress in art and literary works. Kawamura and his team (2019, 2021, among others) have generated 100 million AI-generated haiku via three steps: (i) deep learning of 400,000 haiku by Long Short-Term Memory (LSTM) and making word sequences, (ii) filtering out the word sequences for the traditional 17-syllable rule and a seasonal word, and (iii) scoring the generated haiku (cf. Hirata et al. 2022). They report that the scoring process is the most challenging.

Chamberlain et al. (2017) claim that humans have a bias against computer-generated art, and we will investigate whether this bias also applies to AI-generated haiku in this paper. Three in-person experiments were conducted. Experiment 1 aimed to establish a standard for evaluating human-generated haiku. Experiments 2 and 3 investigated how humans evaluate AI-generated and human-generated haiku based on the evaluation standard gained by Experiment 1. We recruited 140 university students and asked them to evaluate 30 human-generated and AI-generated haiku on a 5-point Likert scale. In Experiment 3, but not in Experiment 2, the subjects were informed which haiku was AI-generated. After Chamberlain et al., we predicted human-generated haiku would be more highly evaluated, but the results show that human cognition is more complex.

Keywords: haiku; AI-generated haiku; human-generated haiku; bias; inference

1. Introduction. Haiku, a traditional Japanese unrhymed short poem of 17 syllables/morae, originated in the 17th century. (1a) is a famous haiku by a haiku master Matsuo Basho (1644–1694).

* We would like to express our gratitude to the participants of our experiments for their cooperation. We would also like to thank the Laboratory of Harmonious System Engineering Research Group of Synergetic Information Engineering (Harmo-lab) of Hokkaido University for making their AI-generated haiku poems open to researchers. Our thanks go to the comments and suggestions for our presentations at the 11th Conference on Language, Discourse, and Cognition in Taipei in May 2024 and at the 5-Minute Linguist session at the 2025 Annual Meeting of the Linguistic Society of America in Philadelphia in January 2025. We also like to thank Ayumi Matsuo and Susan Jones of Koge College for their comments and our mentors from LSA for preparing the talk. This work is partially supported by the research fund of Kobe College given to the first author. The remaining errors are, of course, all ours. Authors: Koichi Tateishi (tateishi@mail.kobe-c.ac.jp) & Shinobu Mizuguchi (mizuguti@kobe-u.ac.jp).

(1) a. Furuike-ya kawazu tobikomu mizu-no oto[1] (MATSUO Basho)
old pond-CL frog (OJ) jump in water-GEN sound
(where CL=cut letter, OJ=Old Japanese, GEN=genitive)
b. Old pond – frogs jumped in – sound of water (translated by Lafcadio Hearn)

Haiku's reception outside of Japan has always been mediated by translators (cf. Shea & Caldwell 2024:6). Lafcadio Hearn (1850–1904)[2] translated Basho's famous haiku (1a) in a monostich as in (1b). Nowadays, haiku is generated globally, and the Haiku Society of America defines haiku as in (2).

(2) The definition of haiku by The Haiku Society of America (2004)
A haiku is a short poem that uses imagistic language to convey the essence of an experience of nature or the season intuitively linked to the human condition.

Haiku, or a version of it, is now written by poets worldwide, and even by AI. Kawamura and his team (2019, 2021, among others) have generated 100 million AI-generated haiku via three steps: (i) deep learning of 400,000 haiku by Long Short-Term Memory (LSTM) and making word sequences, (ii) filtering out the word sequences for the traditional 17-syllable rule and a seasonal word[3], and (iii) scoring the generated haiku (cf. Hirata et al. 2022). They report that the scoring process was the most challenging.

Chamberlain et al. (2017) claim that humans have a bias against computer-generated art, and we will investigate whether this bias also applies to AI-generated haiku through evaluation experiments. The structure of this paper is as follows: Section 2 briefly introduces haiku, especially its structure and the traditional evaluation method. In Section 3, we conduct Experiment 1 on human-generated haiku and establish our standard to evaluate haiku. In Sections 4 and 5, we conduct in-person evaluation experiments on human-generated and AI-generated haiku, with (Experiment 3) and without (Experiment 2) author identification. Section 6 is general discussions, and the conclusion follows in Section 7.

2. Structure and Evaluation of Haiku Poem. We will briefly introduce haiku poems, especially their structure and the traditional evaluation method, in this section.

2.1. STRUCTURE OF HAIKU. Japanese haiku has a formulaic limitation of 5-7-5 syllables/morae. It usually contains a seasonal word and a cut letter (CL) to divide 17 syllables into smaller units of [5][75], [57][5], or [5][7][5]. The literature (cf. Toyama 2003, Arima 2018, among others) says that the number of cut letters found in haiku varies from 13 to 22, depending on when haiku poems were written. The most typical ones are *ya* (cf. (1)), *keri,* and *yo* today.

Japanese is an agglutinative language, and a case particle is attached to a head noun. The basic word order is SOV, but Japanese allows scrambling so that the word orders of SVO, OVS, OSV, VSO, and VOS are all possible since a case particle like *ga* or *o* realizes the grammatical function of subject or object.

Due to its brevity and poetic character, expressions in haiku often lack syntactic/semantic arguments and predicates, and case particles are often missing, too; in (1a), for example,

---

[1] The method of Romanization of Japanese follows the Hepburn method unless otherwise specified.

[2] Lafcadio Hearn was a Greek-Irish writer, translator, and teacher. His collections of legends and ghost stories and his book *Glimpses of Unfamiliar Japan* (1894) offered unprecedented insight into Japanese culture. He became a Japanese citizen and his Japanese name was Koizumi Yakumo.

[3] There are more than 8,000 seasonal words used in haiku. For the online database, visit http://www.haiku-data.jp/kigo.php.

Subjective case marker *ga* is missing after *kawazu* ‘frog’. Readers are expected to infer that the subject of *tobikomu* ‘jump in’ is *kawazu*.

A cut letter (CL) brings readers another difficulty in reproducing a poet’s experience since it splits the whole text of 17 syllables/morae into more than one part: [5][75], [57][5], [5][7][5]. (3a) is a famous haiku by Yosa Buson (1716–1783); the cut letter *ya* separates the haiku into [57][5], and the second [5] *huyu-kodachi* ‘winter grove’ is considered to be the topic of this haiku as a result.

(3) a. Ono-irete ka-ni odoroku-ya fuyu-kodachi (Yosa Buson)
axe-cut scent-with surprised-CL winter-grove
b. Cutting (a winter grove) with an axe, (I) was surprised with its scent
(translated by one of the authors)

(3a) contains two predicates *ireru* ‘cut’ and *odoroku* ‘be surprised’, but the subjects are not explicitly expressed. Readers are expected to infer the appropriate subjects by themselves, i.e., the author in this case. Also, readers are expected to infer the object of *ireru* ‘cut’, i.e., the topic of the haiku *fuyu-kodachi*.

Haiku is a short poem written in 17 syllables/morae that uses few verbs, few pronouns, no rhymes, and no title, which is not evident to Westerners. Haiku usually includes one word to suggest a season and one indicator of a break. Hokenson (2024: 152) claims that haiku traits are brevity, simplicity, and suggestiveness, which are most startling for the West.

2.2. EVALUATION OF HAIKU. There are tremendous numbers of both amateur and professional haiku-writing circles and posting sites in Japan and other parts of the world (e.g., https://haikutown.jp for human-generated haiku, https://aihaiku.org for AI-generated haiku). Usually, haiku is evaluated within haiku-writing circles and posting sites. Since the members of the writing circles are closed, how haiku is evaluated within a circle is unknown to the public (cf. Kawamura & Otsuka 2022). There are many case studies on professional human-generated haiku poems from the literary point of view, but only a few studies are found on the evaluation method of haiku. To the authors’ knowledge, Kuwabara (1946) was the first to research the evaluation of amateur and professional human-generated haiku poems. Hitsuwari et al. (2021, 2023) and Hirata et al. 2022 conducted evaluation experiments on AI-generated haiku poems. Below, we will introduce previous evaluation studies on human-generated and AI-generated haiku poems briefly.

2.2.1. EVALUATION OF HUMAN-GENERATED HAIKU. Kuwabara (1946) was the first study on the evaluation of human-generated haiku. Kuwabara was not a haiku poet but a scholar of French literature. He researched how human-generated haiku was evaluated, as given in (4).

(4) Kuwabara’s Experiment (Kuwabara 1946, reprinted in Kuwabara 2024:58)
materials:
One haiku each by ten contemporary poets with reputations as masters of the form and five haiku by unknown or relatively unknown poets were chosen. The names of the poets were removed from all the materials, and haiku poems were introduced anonymously.
method:
Kuwabara showed the materials to a small number of educated readers, and asked them to 1) rank the haiku in order of evaluative rating, 2) try to guess which haiku were written by famous poets, regardless of the rating, and 3) consider whether a valid distinction can be made between the ten verses composed by professionals and the five written by amateurs.

results:
A single verse provides little basis for judging poetic talent, making it impossible to distinguish first-rate poets from the amateurs.

Kuwabara (1946) argued that art is meaningless unless the work allows the artist's experience to be reproduced in the person who admires it. Kuwabara claimed haiku is 'incomplete' and a 'second-class art'. As discussed in the previous section, haiku does not express everything in itself and requires readers to 'infer' what is not expressed in words so that they can share a sophisticated aesthetic experience of the poet. Kuwabara was a scholar with deep knowledge of Western culture, and he labeled haiku as 'incomplete.' Kuwabara was, however, much criticized by haiku poets who could not accept his idea that haiku is 'a second class art.' His research was limited in the materials and the number of participants, and his study is far from acceptable as an experiment. However, to the authors' knowledge, no universal standard to evaluate human-generated haiku has been proposed before or after him.

2.2.2. EVALUATION OF AI-GENERATED HAIKU. Hitsuwari et al. (2021, 2023) researched AI-generated haiku from a psychological point of view. They conducted psychological experiments and examined aesthetic evaluations of AI-generated haiku and people's beliefs. They divided AI-generated haiku with and without human intervention. They recruited 385 participants and asked them to evaluate 40 haiku (20 human-generated, 10 AI-generated haiku without human intervention, and 10 AI-generated haiku with human intervention) on 21 items such as beauty, valence, empathy, nostalgia, and novelty. They also asked the participants to judge whether the haiku was human-generated or AI-generated. Their results were (i) the beauty rating of the AI-generated haiku with the human intervention was the highest, and those of the human-generated and AI-generated haiku without human intervention were equal, and (ii) participants could not distinguish human-generated haiku from AI-generated haiku.

Though Hituwari et al.'s studies are more convincing than Kuwabara's, how they chose their materials and intervened in the AI-generated haiku in their experiments is unclear.
Hirata et al. (2022) studied how a haiku generator was evaluated from the informatic point of view. Harmo-lab of Hokkaido University, Japan, generated over one trillion haiku by deep-learning 400,000 human-generated haiku via Long Short-Term Memory (LSTM) and GPT-2. They later used novels and articles available online (https://github.bom/aozorabunko/) for deep-learning and they made three haiku generators: LSTM that deep-learned human-generated haiku, GPT-2 that deep-learned human-generated haiku, and GPT-2 that deep-learned both human-generated haiku and novels and articles.

For evaluation, they asked amateur haiku poets to label their AI-generated haiku as either haiku or non-haiku. They calculated the Area Under the Curve (AUC) based on the likelihood of AI-generated haiku and their binary label. Their experiment in 2021 showed that GPT-2 that deep-learned not only haiku but also novels and articles hit the highest AUC (0.56) among the three generators.

Hirata et. al (2022) is not a study of the evaluation of human-generated haiku, and we do not know how to interpret their experiment's relatively low AUC scores. However, what interests us is that not only human-generated haiku but also other types of literary works function effectively to generate AI-generated haiku that is more human-generated-like.

2.2.3. HOW TO EVALUATE HAIKU. Haiku has been written for over 400 years and draws a common life in a short and simple form. Due to its brevity, haiku sometimes misses arguments or

verbs syntactically. (5) is an example of a professional haiku poet intentionally generating 'incompleteness.'

(5)[4] a. Ame-no hi-no ame-no hikari-no susuki-no-ho
rain-GEN day-GEN rain-GEN light-GEN pampas-GEN-ear
(Kakutani Masako)
'The rainy light on a pampas ear on a rainy day.'
(translated by one of the authors)

b. Ame-no hi-no ame-no hikari-o susuki-no-ho
light-ACC
(revised by Kuroda Kyoko)
'On a rainy day, rain (brings) light on a pampas ear.'

(5a) is the original haiku poem by Kakutani Masako, a professional haiku poet. (5b) is a revised version of (5a) by Kuroda Kyoko, a master haiku poet. She changed Genitive *no* in (5a) to Accusative *o*. This change makes readers infer that a predicate is missing in (5b), so they must fulfill the predicate themselves. We tentatively insert *bring* in our translation, but there may be other possibilities, depending on the interpretation of readers. Haiku generates 'incompleteness' in this way, which makes haiku sound more thought-provoking.

Also, due to its disjunctive function of a cut letter, the topic of haiku is sometimes hard to combine with the other parts of haiku semantically. Gilbert (2024) claims that the play of disjunction versus coherence is a taproot of haiku. The disjunctive method related to *kireji* 'cut letter' highlights a 'topic' of haiku (cf. (1), (3)), helping to catalyze the reader's aesthetic perception of haiku as an art form and evoking a sense of 'depth'. Some consider these characteristics of haiku as 'incomplete' (cf. Kuwabara 1946). Others claim that juxtaposition alone does not confer poetic power, and 'incompleteness, absence, and ambiguity' are necessary (cf. Kawamoto 2019). This is where the difficulty of evaluating haiku poems lies. Due to its development, AI is good at generating and summarizing verbal texts. The evaluation standard for texts is clear: correctness. We, however, do not know how to evaluate poems yet. As a first step in the search, we will investigate what affects the evaluation of human-generated haiku in Section 3 through an experiment. Then, we will conduct evaluation experiments on AI-generated and human-generated haiku and see whether humans are biased against AI-generated art in Sections 4 and 5.

3. Experiment 1 on human-generated haiku. To establish a standard for evaluating human-generated haiku, we will conduct a follow-up experiment of Kuwabara's one (1946).

3.1. METHOD. Our materials are the 15 haiku poems (10 by professional poets, 5 by amateur poets) used in Kuwabara's experiment. We added 2 human-generated and 3 AI-generated haiku, so the total number of professional haiku and amateur haiku to be evaluated is 10 each.

We recruited 15 volunteer amateur haiku poets from the same haiku-writing circle and asked them to evaluate the randomized 20 haiku on a 5-point Likert scale. After evaluation, they were informed that three materials were AI-generated and were asked to identify the haiku they thought were AI-generated. After Hitsuwari et al. (2021, 2023), we predicted that our participants could not distinguish AI-generated haiku from human-generated haiku. The

[4] The original haiku and the revised version are taken from *Haiku* (May 2024:133), a haiku journal, published by Kadokawa Publishers, Tokyo.

experiment, including the exercise session, was conducted in person and took about 30 minutes to finish.

3.2. RESULTS. We have three main findings. First, we found no significant difference in evaluation between haiku written by famous poets (mean 3.33) and haiku written by amateur poets (3.67) on one-way ANOVA ($F(1, 16)=4.54$, $p=.14$). Second, we found no significant difference ($F(1,18)= 4.41$, $p=.69$) between haiku with a cut letter (mean 3.52) and haiku without a cut letter (mean 3.43), either. Third, AI-generated haiku was more highly evaluated (mean 3.86) than human-generated haiku. Nobody except the leader of the haiku-writing circle identified all three AI-generated haiku as they are. The leader said he could identify all the AI-generated haiku poems because each word perfectly coheres with the following ones, as in (6).

(6) Miidera-no monzen aogu oborozuki (AI-generated)
Mii Temple-GEN gate look up hazy moon
'The hazy moon, looking up the gate of the Mii Temple' (translated by one of the authors)

3.3. DISCUSSION. Experiment 1 shows that the evaluation of famous haiku poets and amateur poets is not distinguishable, suggesting that Kuwabara's claim is correct. However, we find problems in his experiment design. First, the standard for the selection of haiku materials is not clear. We might have had different results if we had chosen another haiku. Second, the numbers of haiku in each group are not equal; Kuwabara (1946) chose 10 haiku written by famous poets and 5 haiku by amateur poets, the reason of which is not given in the paper.

We need to establish a standard for evaluating haiku. Gilbert (2024) claims that the play of disjunction and coherence is a taproot of haiku. Disjunction is realized by a cut letter, and coherence is established by semantic combination. In the search for an evaluation measure, we first examined whether a cut letter like *ya* or *keri* affects evaluation. A cut letter divides the whole text into a smaller unit as in (3a), sometimes making it difficult for readers to reproduce a poet's experience. However, as seen above, we found no evidence that a cut letter affects the evaluation of haiku in Experiment 1.

As a second step to search an effective evaluation variable, we focused on the number of predicates. Due to its brevity, haiku often omits arguments or predicates, which makes it difficult to establish local coherence. Haiku poets let readers infer that a predicate is missing by a case particle (cf. (5b)), which they believe makes haiku more sophisticated. There are even haiku without a predicate, like (7), which consists only of nouns.

(7) Nara nanae shichidô[5] garan yaezakura
Nara sevenfold tower seven temples cathedrals double cherry blossom
(MATSUO Basho)
'The sevenfold tower in Nara, seven temples and cathedrals, double cherry blossoms'
(translated by one of the authors)

It is not easy for readers to interpret haiku like (7) in terms of how the three nouns are semantically combined and are coherent.

[5] We follow the official rule of the Agency of Cultural Affairs of the Japanese Government regarding the Romanization of long vowels. According to the rule, long /a, o, u/ is with a circumflex accent (*â, ô, û*), long /e/ is *ei*, and long /i/ is *ii*. The rule of accentuation, however, does not apply to proper names, following the rule by the Japanese Ministry of Foreign Affairs.

If we are on the right track to suspect that the inference of the missing predicate and the global coherence between predicates may provoke deeper interpretation of haiku, the number of predicates may be a key to evaluate haiku; haiku is classed into the ones with zero predicate, like (7), those with one predicate, like (1) and (6), and those with two predicates, like (3).

On this assumption, we reconsidered our materials used in Experiment 1 and classed them into three groups: Group 1 without a predicate (N=2), Group 2 with one predicate (N=9), and Group 3 with two predicates (N=9). We found a significant difference in evaluation between Group 2 (mean 3.74) and Group 3 (mean 3.32) on One-Way ANOVA ($F(1,17)=4.49$, $p=.038$). Group 3 consists of haiku with two predicates, so readers are expected to process how the two predicates are combined semantically. This type of coherence is called 'global coherence' in the literature (cf. Halliday and Hasan 1976) and is harder to process than local coherence. We suspect the lower evaluation of Group 3 than Group 2 is due to this type of processing difficulty.

Based on the results of Experiment 1, we tentatively assume the number of predicates is one of the evaluation variables of human-generated haiku. The other evaluation variable we assume is 'cut letter'; though Experiment 1 did not find any significant difference in evaluation between human-generated haiku with a cut letter and those without a cut letter, the literature traditionally treasures the function of a cut letter. We decided to include 'cut letter' in evaluation variables and conduct evaluation experiments on a larger number of materials than in Experiment 1. In Experiment 2, we examine how humans evaluate AI-generated haiku and human-generated haiku with the materials varied in the number of predicates and the presence of cut letter.

4. Experiment 2 without author identification. Experiment 2 aims to investigate how humans evaluate human-generated haiku and AI-generated haiku. Chamberlain et al. (2017) claim that humans are biased against computer-generated art, and we predict that human-generated haiku would be more highly evaluated than AI-generated haiku.

4.1. METHOD. We used two sets of materials. Set 1 contained haiku with a cut letter, while Set 2 did not. We arranged each set to contain 15 AI-generated and 15 human-generated haiku, with zero predicate, one predicate, and two predicates, respectively (cf. Appendix for the materials).

We recruited 59 (for Set 1) and 46 (for Set 2) university students without haiku background and asked them to evaluate 30 haiku on a 5-point Likert scale in person. After evaluation, they were informed that half of the materials were AI-generated and were asked to identify the haiku they thought were AI-generated. The experiment, including the exercise session, took about 30 minutes, and the students were given a course credit for participation.

4.2. RESULTS. AI-generated haiku were evaluated (mean 3.22) higher than human-generated haiku (mean 3.18). Still, the difference was not significant on One-way ANOVA ($F=(1,59)=4.0$, $p=.38$). No significant difference was observed between Set 1 with a cut letter (mean 3.21) and Set 2 without a cut letter (mean 3.16), either ($F=(1,59)=4.01$, $p=.64$).

Secondly, 70% of AI-generated haiku were incorrectly judged as human-generated, and 30% of human-generated haiku were misjudged as AI-generated.

Third, the number of predicates affected the evaluation of haiku; One-way ANOVA among 0, 1, and 2 predicates was significant ($F(2, 29)=3.35$, $p=.033$). Further examination reported a significant difference between AI-generated and human-generated haiku with one predicate ($F(1, 17)=4.49$, $p=.03$). Observe Figure 1.

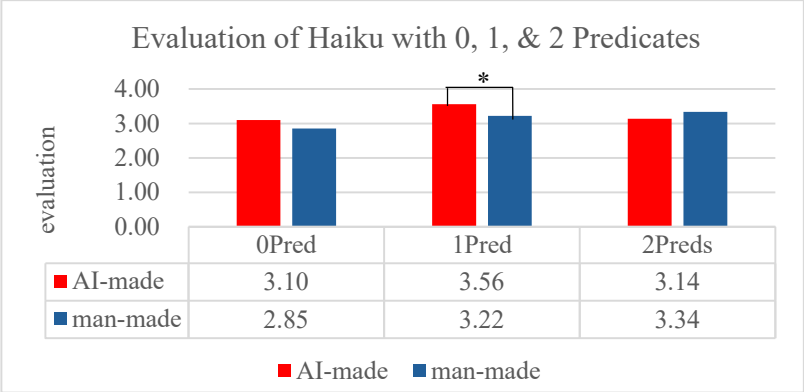

Figure 1. Evaluation of haiku with 0, 1, and 2 predicates

4.3. DISCUSSION. Experiment 2 supports the finding of Experiment 1 that a cut letter does not affect the evaluation of haiku; there was no significant difference in evaluation between Set 1 and Set 2. We can claim that a cut letter does not affect the evaluation of both human-generated and AI-generated haiku.

Figure 1 shows that the number of predicates affects the evaluation of haiku. Haiku without a predicate consist of nouns only (cf.(7)), and it is difficult for readers to interpret how they are semantically combined. This probably explains the lower evaluation of haiku without a predicate (mean 3.10 for AI-generated haiku, mean 2.85 for human-generated haiku) than haiku with predicates. Haiku with one predicate makes establishing local coherence easier, and AI-generated haiku with one predicate records the highest score (mean 3.56) in this experiment; *AI-Issakun* generates haiku by LSTM and calculates the accessibility of each argument and predicate, which probably makes AI-generated haiku easier to process than human-generated haiku. On the other hand, human-generated haiku with two predicates are evaluated higher (mean 3.34) than AI equivalents (mean 3.14). Haiku with two predicates (cf. (3)) require combining two predicates semantically and demand global coherence in addition to local coherence. We suspect that AI has not learned how to establish global coherence yet (p.c., Soichiro Yokoyama of Harmo Lab, Hokkaido University), and this may lead to a lower evaluation of AI-generated haiku with two predicates. In the next section, we will come back to the discussion of the number of predicates used in a haiku.

Experiment 2 shows that there was no significant difference in evaluation between AI-generated haiku (mean 3.22) and human-generated haiku (mean 3.18), contrary to our prediction. We will see whether humans evaluate AI-generated haiku differently when they are informed which haiku is AI-generated before they evaluate and compare the results of Experiments 2 and 3 in the next section.

5. Experiment 3 with author identification.

5.1. METHOD. The materials were the same as those used in Experiment 2. We recruited 35 university students without haiku background and asked them to evaluate 30 haiku on a 5-point Likert scale. In Experiment 3, the participants were informed which haiku was AI-generated before evaluation. The experiment, including the exercise session, took about 20 minutes, and the participants were given a course credit for participation. We predicted that human-generated haiku, not AI-generated haiku, would be more highly evaluated in Experiment 3.

5.2. RESULTS. The results were different from those of Experiment 2. First, human-generated haiku was more highly evaluated (mean 3.39) than AI-generated haiku (mean 3.18), and the difference was significant on One-way ANOVA (F(1,59)=4.0. p=.015).

Second, the difference between Set 1 and Set 2 is significant on One-way ANOVA (F(1,59)=4.01, p=0.034).

Third, the number of predicates haiku contains brings a difference in evaluation; haiku with two predicates shows a significant difference between AI-generated and human-generated haiku (F=(1,22)=4.32, p=.01), but haiku with zero and one predicate do not. Observe Figure 2.

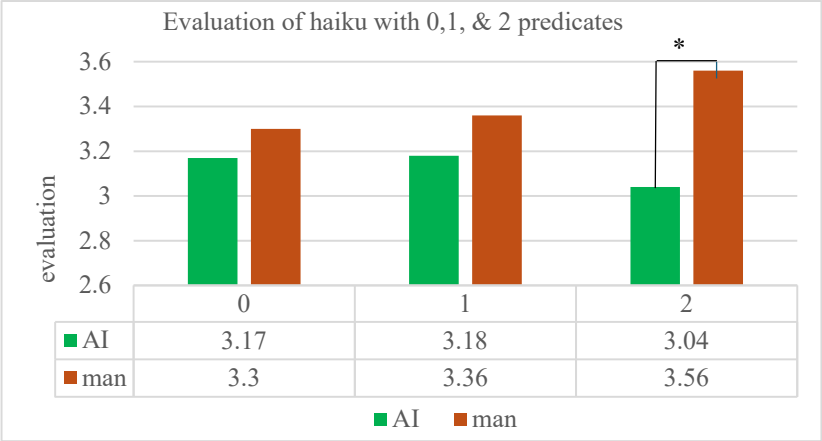

Figure 2. Evaluation of haiku with 0, 1, and 2 predicates

5.3. DISCUSSION. When informed, humans evaluate human-generated haiku (mean 3.39) highly than AI-generated haiku (mean 3.18). The differences are significant in haiku with two predicates. Our prediction that human-generated haiku would be more evaluated is borne out.

Contra Experiment 2, Experiment 3 shows a significant difference in evaluation between Set 1 and Set 2. This means that a cut letter now affects the evaluation of haiku with author identification. Observe Figure 3. We will further compare the results of Experiments 2 and 3 and assess how humans process human-generated and AI-generated haiku in the next session.

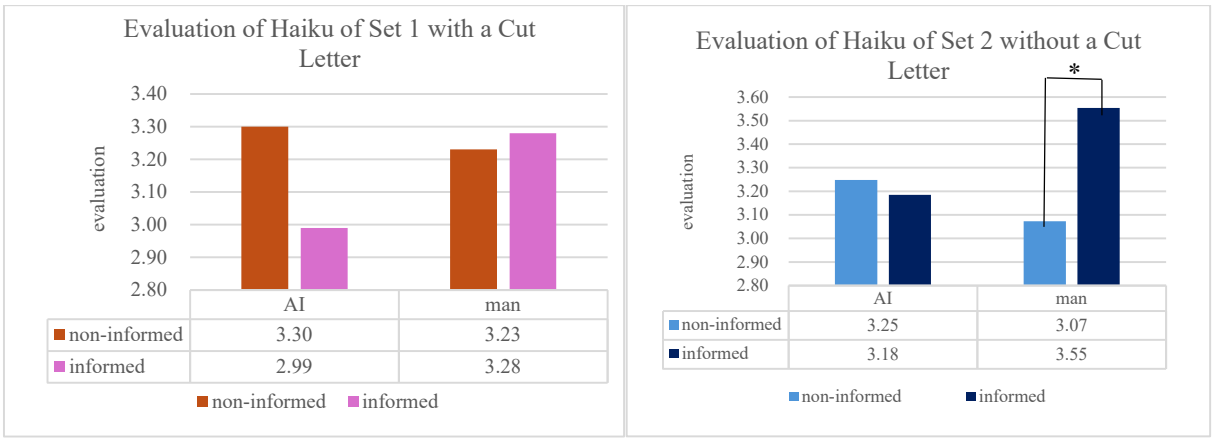

Figure 3. Evaluations of AI-generated and Human-generated Haiku of Set 1 (Left) and Set 2 (Right)

6. General Discussion. Our research question was to see whether humans are biased against AI-generated art. We conducted three evaluation experiments in person. In Experiment 2, where

participants were not informed which haiku was generated by AI, AI-generated haiku (mean 3.32) was evaluated higher than human-generated haiku (mean 3.18). Still, the difference was not significant: $F(1, 59)=4.01$, $p=.64$) . After evaluation, we asked the participants to identify the haiku they thought were AI-generated, and one-third of AI-generated haiku was correctly judged as AI-generated. One-third of human-generated haiku was incorrectly judged as AI-generated. These results suggest that humans do not distinguish between the two types of haiku.

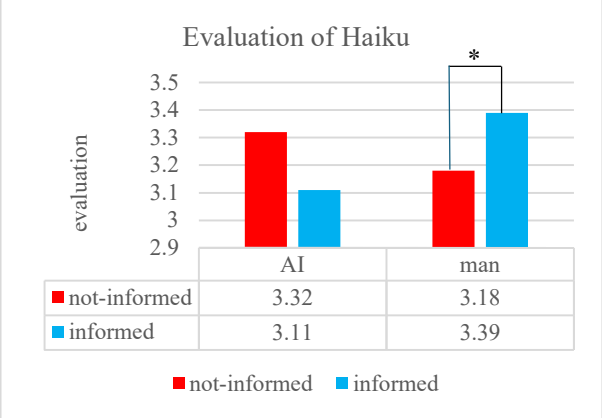

Figure 4. Evaluation of Haiku with/without author identification

In Experiment 3, where the author identification was informed, human-generated haiku (mean 3.39) was evaluated as significantly higher than AI-generated haiku (mean 3.11): ($F(1,59)=4.0$. $p=.015$). Figure 4 shows that the evaluation of AI-generated haiku was lower when informed as is, while the opposite was true with human-generated haiku. In a word, the evaluations of AI-generated haiku are not much different whether or not the author is identified, while human-generated haiku is more highly evaluated when the author is identified. These facts suggest that humans appreciate human-generated haiku more than AI-generated haiku. Our answer to our research question is that we are not biased against AI-generated haiku, but we are biased to believe that humans are superior in writing haiku to AI.

We would like to account for how humans process haiku poems. Since human-generated haiku without a cut letter is more highly evaluated when the author is identified (cf. Figure 3), we assume author identification and the number of predicates as explanatory variables and conducted a multiple regression analysis between the evaluation values on Set 2 without a cut letter. Table 1 shows the result.

	human-generated (Set 2)		AI-generated (Set 2)	
	non-informed t	informed t	non-informed t	informed t
one-predicate	2.2*	0.91	0.87	0.23
two-predicates	4.57***	2.5*	-0.13	-0.03
Adj.R^2	0.578	0.25	-0.047	-0.03

Table 1. Results of regression analyses between the evaluation of Set 2 and one/two predicates (where *: $p<0.05$, ***: $p<0.001$)

Table 1 shows that when the author is not identified, predicates, especially two predicates, are good predictors of the evaluation of human-generated haiku of Set 2 without a cut letter. When the author is identified, two-predicates are weaker predictors.

The story is different with AI-generated haiku. Table 1 shows that predicates do not work as explanatory variables to predict the evaluation of AI-generated haiku. This is not a surprise when

we see how AI generates haiku; AI deep-learns a large amount of haiku and prose and then generates word sequences by LSTM or GPT-2. The AI-haiku-generating system calculates the accessibility of each argument and predicate and establishes local coherence. It has not, however, considered how to combine predicates yet (p.c., Soichiro Yokoyama of Harmo Lab). For evaluating haiku, the AI-generating system calculates Bert exchange value (cf. Harmo-lab). As far as the authors know, BERT models look at the surrounding words to understand the context and deliver state-of-the-art results in natural language processing. To see whether Bert exchange values predict the evaluation of haiku, we conducted Pearson's Correlation between human evaluation scores gained by Experiments 2 and 3 and the Bert exchange values, which are open by Harmo-lab (http://harmo-lab.jp/?page_id=4052). The correlation is very low (r=0.047), and we claim that Bert exchange values are not correlated with human evaluation. AI-generating system has to search for other measures to reflect human cognition.

We have seen that human-generated haiku and AI-generated haiku are generated completely differently and that the evaluation of AI-generated haiku does not differ much between with and without author identification. In other words, as far as each argument and predicate are calculated to establish coherence, evaluation does not change. On the other hand, human-generated haiku is intentionally vague without subjects or predicates; it is evaluated higher with two predicates (cf. Figure 2) and when the author is identified (cf. Figure 3). Table 1, however, suggests that predicates alone cannot explain the tendency of human cognition. We need further research.

7. Conclusion. This paper aimed to see whether we are biased against AI-generated haiku. We studied how humans process the human-generated and AI-generated haiku with and without author identification. We found that our processing is subject to change, depending on the information given when we evaluate. When identified as human-generated, human-generated haiku are evaluated more than when they are not. AI-generated haiku show no such difference in evaluation. We conclude that we are biased not against AI-generated haiku but against human-generated haiku.

Human-generated haiku is sometimes 'incomplete'; it is intentionally vague, often lacking subjects or predicates. Readers are encouraged to infer missing arguments and predicates, which makes haiku thought-provoking and sophisticated. Conversely, AI generates haiku by calculating the accessibility of each argument and predicate and establishing local coherence. Hence, each word perfectly coheres with the following word, and we do not need to infer when we read. Experiment 3 shows that human-generated haiku is significantly highly evaluated with author identification. This is probably because humans try to find 'incompleteness' in haiku and to infer 'hidden' entailment in human-generated haiku. However, people do not encourage themselves to deepen their interpretation of AI-generated haiku. We suspect that the evaluation difference between the two experiments is due not to haiku but to the human processing attitude. In Experiment 2, where no author identification is given, AI-generated haiku is evaluated higher than human-generated haiku. This is probably because they are 'easier' to process.

Our study is a small step in investigating the human cognitive system through haiku poems. We need further study on what is activated to process human-generated art. Also, AI needs to generate art to activate the human cognitive system more. When these two successfully develop and are combined, our evaluation of AI-generated art will change. We will see when.

References

Arima, Michiko. 2018. *Kigôron-kara Mita Haiku* [Semiotical Approach to Haiku]. Tokyo:

Kaitakusha.
Barzilay, Regina & Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics* 34(1). 1–34.
Chamberlain, Rebecca., Caitlin R. Mullin & Johan Wagemans. 2017. Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics Creativity and the Arts*. 12(2). 177–192. https://doi.org/10.1037/aca0000136.
Gilbert, Richard. 2024. The disjunctive dragonfly: A study of disjunctive method and definitions in contemporary English-language haiku. In James Shea & Grant Caldwell (eds.), 197–224. https://doi.org/10.4324/9781003493309-16.
Halliday, M.A.K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
Harmo-lab [Laboratory of Harmonious Systems Engineering Research Group of Synergetic Information Engineering, Hokkaido University]. http://harmo-lab.jp/?page_id=4052.
Hirata, Kodai, Soichiro Yokoyama, Tomohisa Yamashita & Hidenori Kawamura. 2022. Evaluation of a haiku generator with autoregressive models based on deep learning [in Japanese]. *Proceedings of the 84th General Meeting of Information Processing Society of Japan*. 2-813–814.
Hitsuwari, Junpei & Michio Nomura. 2021. Ambiguity and beauty: Japan-Germany cross-cultural comparison on aesthetic evaluation of haiku poetry. *Proceedings of the 38th Annual Meeting of Japanese Cognitive Science Society*. 919–923.
Hitsuwari, Junpei, Yoshiyuki Ueda, Yun Woojin & Michio Nomura. 2023. Does Human-AI collaboration lead to more creative art? Aesthetic evaluation of human-generated and AI-generated haiku poetry. *Computers in Human Behavior,* 139. https://doi.org/10.1016/j.chb.2022.107502.
Hokenson, Jan W. 2024. Haiku as a Western genre: Fellow traveler of Modernism. In James Shea & Grant Caldwell (eds.), 148–173. https://doi.org/10.4324/9781003493309-14.
Kawamoto, Koji. 2019. *Haikai-no Shigaku* [Poetics of Haiku]. Tokyo: Iwanami.
Kawamura, Hidenori & Gai Otsuka. 2022. *Hito-wa Naze Haiku-o Yomu-no-ka?* [Why do humans make haiku?]. Chiba: dZero.
Kawamura, Hidenori, Tomohisa Yamashita, & Soichiro Yokoyama. 2021. *Jinkô-Chinô-ga Haiku-o Yomu: AI Issa-kun-no Chôsen* [AI generates Haiku: A Challenge of AI-Issa]. Tokyo: Ohmusha.
Kuwabara, Takeo. 1946. A second-class art: On contemporary haiku. Reprinted in James Shea and Grant Caldwell (eds.), 57–73. https://doi.org/10.4324/9781003293309-8.
Shea, James, & Grant Caldwell (eds.). 2024. *The Routledge Global Haiku Reader.* Abingdon: Routledge.
Toyama, Shigehiko. 2003. *Tanshi-kei-no Bungaku* [Literature in Short Verse]. Tokyo: Misuzu Shobo.
Yokoyama, Soichiro, Tomohisa Yamashita & Hidenori Kawamura. 2019. Generation and selection of haiku poems using deep learning [in Japanese]. *Artificial Intelligence* 34-4, 467–474.

Appendix: Due to space limitations, we will only present the experiment materials of Set 2. The numbers in parentheses indicate the set number and the number of predicates haiku contains. (2-0), for example, means haiku with zero predicate in Set 2. The author's name is given in a parenthesis with the last name first and the second name last. Haiku is translated by one of the

authors. Abbreviations: ACC=accusative; CL=cut letter; CONJ=conjunction; GEN=generic; HON=honorific; IO=indirect object, LOC=locative, SUB=subjective, PL=plural marker.

SET 2 (without a cut letter)

HUMAN-GENERATED HAIKU

#1(2-0) Haru-no yami O-Shikoku-no yami ware-no yami
spring-GEN gloom HON-Shikoku-GEN gloom I-GEN gloom
'The gloom of spring, of Shikoku, and of my own.'
(Kuroda Kyoko)

#2(2-0) Amenbo-to ame-to amenbo-to ame-to
water strider-and rain-and water strider-and rain-and
'Water strider, rain and water strider, and rain.'
(Fujita Shoshi)

#3(2-0) Kagerô-ni guigui neko-no ibiki-kana
heat haze loud cat-GEN snore-CL
'Cat is snoring loudly in the heat haze.'
(Kobayashi Issa)

#4(2-0) Saigoku-no aze manjushage manjushage
Western Japan-GEN shore red spider lily red spider lily
'On the shore of Western Japan is planted red spider lily, red spider lily'
(Mori Sumio)

#5(2-1) Odoriko-no tsuma-ga nagarete ikini-keri
dancer-GEN wife-SUBJ float go-CL
'My wife, a dancer, is dancing away.'
(Nishino Kirin)

#6(2-1) Konna mi-mo hirofu kami ari-te haru-no hana
like this body-too save god be-CL spring-GEN flower
'There is a god, saving one like me, spring flower.'
(Kobayashi Issa)

#7(2-1) Shi-go-nin-ni tsuki ochi-kakaru odori-kana
four-five person-LOC moon pour down dance-CL
'The moonlight is pouring down over four, five dancing people'
(Yosa Buson)

#8(2-1) Tokeiya-no tokei haru-no yoru dore-ga honto
watch store-GEN clock spring-GEN evening which-SUBJ correct
'Clocks at a watch store in the spring evening, which one is correct?'
(Kubota Mantaro)

#9(2-2) Iwashi-gumo hito-ni tsugu-beki koto-nara-zu
Cirrocumulus cloud peope-IO tell-should thing-be-NEG
'Cirrocumulus clouds, not a thing to tell to others.'
(Kato Shuson)

#10(2-2) Basu-o machi ôji-no haru-o utagawa-zu
bus-ACC wait big street-GEN spring-ACC doubt-NEG
'Waiting for a bus, I do not doubt the arrival of spring on the street.'
(Ishida Hago)

#11(2-2) Botan chiri-te uchikasanari-nu ni, san-ben
peony fall-CONJ pile up-PERFECT two, three-Classifier
'Peonies fell, and two, three are piled up on the ground.'
(Yosa Buson)

#12(2-2) Yoku mire-ba nazunabana saku kakine-kana
well see-when sheperd's purse bloom hedge-CL
'A closer look finds a shepherd's purse blooming on the hedge.'
(Matsuo Basho)

#13(2-2) Rokugatsu-o kireina kaze-no fuku-koto-yo
June-ACC beautiful wind-SUBJ blow-COMP-CL
'A pleasant breeze blows in June.'
(Masaoka Shiki)

#14(2-2) Nagasu-beki ryûto ware-no mune terasu
float away-should lantern I-GEN heart shine
'Lanterns to be floated are shining on my heart.'
(Terayama Shuji)

#15(2-2) Naki hito-ni atara-nu-yoni mame-o maku
loved one-DAT hit-NEG-such as bean-ACC sprinkle
'I sprinkle beans, caring not to hit the loved one.'
(Oki Amari)

AI-GENERATED HAIKU

#16(2-0) Taiyô-to suiheisen-to manjushage
the sun-and the horizon-and red spider lily
'The sun, the horizon, and red spider lily.'

#17(2-0) manjushage jigoku gokuraku jôdo-kana
red spider lily the hell the heaven the pure land
'Red spider lily, the hell, the heaven, and the pure land.'

#18(2-0) Kuri-no hana shônen-no hi-no mizutamari
chestnut-GEN flower boy-GEN day-GEN puddle
'A chestnut flower, and a puddle of the boy's day.'

#19(2-0) Tenpyô-no iraka-no shita-no fukinotô
The Tenpyo Era-GEN roof-GEN below-GEN butterbur
'Butterburs below the roof of the Tenpyo Era.'

#20(2-1) Shagâru-no koi-no hajimaru natsu-bôshi
Chagall-GEN love-GEN begin summer-hat
'Chagall's summer hat begins love.'

#21(2-1) Shirasagi-no mai-hajime-taru mizu-no ue
white heron-SUBJ fly-start-CL water-GEN surface
'White herons begin flying on the water.'

#22(2-1) Tsuyo-jimo-no hikari-no naka-o aruki-keri
Strong-frost-GEN light-GEN in-ACC walk-CL
'I walked in the light of strong frost.'

#23(2-1) Tsuma-mo ko-mo barentain-no hi-nari-keri
wife-too child-too St.Valentine's Day-GEN day-be-CL
'For my wife and my daughter is the St. Valentine's Day.'

#24(2-1)	Aibiki-no	koe-no	kuragari	sakuranbo	
	Rendezvous-GEN	voice-GEN	darkness	cherry	
	‘Lovers’ rendezvous in the low voice, and cherries.’				
#25(2-1)	Kanashimi-no	futto	kie-taru	fuyu kodachi	
	sadness-GEN	suddenly	disappear-CL	winter grove	
	‘Sadness suddenly disappears, winter grove.’				
#26(2-2)	Kibukure-te	kanpô-yaku-o	kahini-keri		
	get gloated	herbal medicine-ACC	buy-CL		
	‘Got bloated, I bought herbal medicine.’				
#27(2-2)	Masuku-shi-te	hito-no	ushiro-o	tôri-keri	
	mask-wear-CONJ	people-GEN	behind-ACC	pass-CL	
	‘Wearing a mask, I passed behind people.’				
#28(2-2)	Uguisu-no	mi-o	hososhi-te	naruni-keri	
	Japanese nightingale-SUBJ	body-ACC	slim-CONJ	sing-CL	
	‘Japanese nightingale, slimming herself, is singing.’				
#29(2-2)	Tenohira-o	taisetsunishi-te	hina	osame	
	palm-ACC	care for-CONJ	Hina doll	store	
	‘Caring for the palms, I stored Hina dolls.’				
#30(2-2)	Ankô-no	okuchi	aku-te	warahi-keri	
	angler-SUBJ	big mouth	open-CONJ	laugh-CL	
	‘An angler laughs with its mouth wide open.’				