E-Structural (English Studies on Translation, Culture, Literature, and Linguistics)
Vol 8 No 02
p-ISSN: 2621-8844, e-ISSN: 2621-9395
http://publikasi.dinus.ac.id/index.php/estructural

A Comparative Study on the Reception of AI-Generated and Human-Written Flash Fiction by the Students of the English Literature Study Program

[1]Maya Kurnia Dewi, [2]Ratih Laily Nurjanah, [3]Deswandito Dwi Saptanto, [4]Adam Zidan Fahrezi, and [5]Irene Sharon Margaretha

[1,2,3,4,5]Universitas Ngudi Waluyo, Ungaran, Indonesia
[1]mayakurnia@unw.ac.id
[2]ratih.laily@gmail.com
[3]deswanditodwi@unw.ac.id
[4]zydanfahrezi72@gmail.com
[5]irenesharon272@icloud.com

Article History: Submitted November 23rd, 2025; Accepted January 8th, 2026; Published January 13th, 2026

Abstract. The emergence of generative artificial intelligence (AI) fundamentally challenges traditional literary paradigms that burden the modern reader with authenticating the text's source. This study empirically investigates the reception of AI-generated flash fiction (Text A) and human-written flash fiction (Text B) by 16 4th Semester English Literature students at Universitas Ngudi Waluyo. The research utilizes a mixed-methods approach grounded in Hans Robert Jauss's reception theory, particularly the concept of the horizon of expectations. The core research methodology employs a "mirrored prompt" approach to ensure high internal validity, giving the human and AI authors the same core narrative and emotional task. The questionnaire analyzed the students' literary experience, Technological Horizons, and Interpretative Horizons. The results show a consistent and significant preference for Text B (Human-written) across all measured dimensions of the Interpretative Horizon, particularly in terms of emotion and stylistics, compared to Text A. Eleven out of 16 students (68.75%) accurately identified Text A as AI-generated and Text B as human-written. Qualitative data reveal that students critique text A for its lack of affective resonance, while text B has a "natural and flowing style". This finding empirically validates that the reader's interpretative horizon, particularly the expectation for deep emotion and unique style, is the primary factor in determining the perceived authenticity of a text, thus updating Jauss's theory to include the challenge of algorithmic works. The accuracy rate (68.75%) is significantly higher than previously reported research, suggesting that academic literary competence may increase the ability to discern AI-generated fiction.

Keywords: AI-generated; flash fiction; horizon of expectation; human-written; reception theory

Abstrak. *Munculnya kecerdasan buatan (AI) generatif secara fundamental menantang paradigma sastra tradisional. Studi ini secara empiris menganalisis resepsi* flash fiction *(fiksi kilat) yang dihasilkan AI (teks A) dan tulisan manusia (teks B) oleh 16 mahasiswa semester empat Program Studi Sastra Inggris di Universitas Ngudi Waluyo. Penelitian ini menggunakan pendekatan metode campuran yang didasarkan pada teori resepsi Hans Robert Jauss, khususnya konsep horizon harapan. Metodologi penelitian intinya menggunakan* "mirrored prompt" *untuk memastikan validitas internal yang tinggi, memberikan tugas naratif dan emosional inti yang sama kepada penulis manusia dan AI. Kuesioner menganalisis horizon pengalaman sastra, teknologi, dan interpretatif. Hasilnya menunjukkan preferensi yang konsisten dan signifikan untuk teks B (tulisan manusia) di semua dimensi horizon interpretatif yang diukur. Teks B mencapai skor rata-rata yang jauh lebih tinggi dalam emosi dan stilistika dibandingkan dengan teks A. 11 dari 16 mahasiswa (68,75%) secara akurat mengidentifikasi teks A sebagai AI dan teks B sebagai tulisan manusia. Data kualitatif mengungkapkan bahwa mahasiswa mengkritik teks A karena kurangnya resonansi afektif, sebaliknya, teks B dipuji karena gaya yang alami dan mengalir. Temuan ini secara empiris memvalidasi bahwa horizon interpretatif pembaca, terutama ekspektasi terhadap emosi yang mendalam dan gaya yang unik, adalah faktor utama*

dalam menentukan keaslian teks yang dirasakan, sehingga memperbarui teori Jauss untuk mencakup tantangan karya algoritmik. Tingkat akurasi (68,75%) secara signifikan lebih tinggi daripada penelitian yang dilaporkan sebelumnya, menunjukkan bahwa kompetensi sastra akademis dapat meningkatkan kemampuan untuk membedakan fiksi yang dihasilkan AI.

Kata kunci: *flash fiction; horizon harapan; kecerdasan buatan generatif; teori resepsi; tulisan manusia*

INTRODUCTION

The digital era has given birth to an unexpected new literary agent: generative artificial intelligence (AI). The emergence of Large Language Models (LLMs) capable of producing creative text with increasing sophistication has fundamentally challenged traditional literary paradigms. It challenges human creativity and cognition (Rahmeh, 2023), human ingenuity (Pavlik, 2023) and human processing power (Perkins, 2023) giving rise to significant problems over academic integrity. This phenomenon demands in-depth scholarly investigation into how these algorithmic works are received, interpreted, and evaluated by readers, especially by a critically trained audience such as literature students. The algorithmic era now adds a new layer of complexity: an ontological distance. The question facing the modern reader is no longer just "What does this text mean?" but also, "Who or what created this text?". The uncertainty regarding the nature of consciousness behind the work, whether human or machine, inherently changes the entire communicative act and the reading process itself (Willis, 2021).

The act of reading, traditionally a hermeneutic exercise in making meaning of a text, is now burdened with the additional task of authenticating its source. This burden of authentication transforms the reader into both a critic and a detective, implicitly having to evaluate the authenticity, intent, and "soul" of the text they face. This cognitive process, which was previously absent, has the potential to affect the reader's trust, emotional investment, and ultimately, their aesthetic judgment. In the terms of literary reading, some forms of prose can be used to investigate the interpretation of the readers, such as by reading flash fiction. Flash fiction is a term for all short fiction under 1000 words, which does not require narrative with much movement or dynamism (Roche-Jacques, 2024). Masih adds that flash fiction is a story in miniature, a work of art carved on a grain of rice. The idea of flash fiction as a story in miniature is aesthetically pleasing and simple, but it does not quite get to the heart of the form. (Masih, 2009). McCormack adds the characteristics of flash fiction as employing relatively closed endings, concluding that '[t]he combination of brevity, flat characterization, brief time frame, and a single generic setting combine to limit the writer's options for introducing the ambiguities, nuances, and symbolic resonances that function so effectively in most open endings (McCormack, 2021). While Fish (2022) in Roche-Jacques focuses on the idea of 'compression' in flash fiction by saying:

> Compression and distillation are so important in flash [fiction]. You can have a very short story that is not very distilled […]. The story must "feel" larger than the space it takes up on the page. That means cutting anything the reader doesn't need. Trusting the reader to understand. Taking out unnecessary words.

Readers' interpretation of the flash fiction can be analyzed by using reception theory similar to the research done by Liliani et al., who used reception theory to analyze micro fictions (Liliani et al., 2022). Before the advent of generative AI, the digital era had already significantly altered the

landscape of literary reception. The internet and social media have created new circuits of reception operating outside traditional literary institutions such as publishing and academic criticism. Digital platforms allow a fragmented audience to engage with texts in new ways, through online comments, discussion forums, and social media, enabling them to dialogue not only with the text but also with fellow readers and sometimes even directly with the author (Kor, 2018). Furthermore, the adaptation of literary works into other media formats such as film, audiobooks, graphic novels, and even video games has broadened the boundaries of literary perception and interpretation (Hrytsak et al., 2025). Each new medium brings its own conventions and expressive possibilities, which in turn shape how a narrative is received and understood. Perception of an artwork is no longer shaped only by individual interpretation, but also by the complex influence of contemporary media integrating textual, visual, and auditory elements (Hrytsak et al., 2025). This transformation indicates that the reader's horizon of expectations is not static but is constantly being reshaped by the dominant communication technology.

Emerging in the 1960s, primarily through the work of thinkers at the University of Konstanz such as Hans Robert Jauss and Wolfgang Iser, reception theory radically shifted the focus of literary criticism from the author (biographical criticism) and the text itself (New Criticism) toward the reader and the act of reading (Willis, 2021). Reception theory asserts that a literary work does not have a fixed and inherent meaning; rather, meaning is "actualized" or realized in the dynamic process of interaction between the text and the reader. The text, in this view, is a linguistic or semiotic artifact designed to be read, and its cognitive and non-cognitive effects only emerge through the encounter with the reader (Willis, 2021). Hans Robert Jauss's aesthetics of reception emphasizes that the meaning of a text is not inherent in the text itself but is formed through the active interaction between the reader and the work. Jauss (1982) introduced the concept of the horizon of expectations, which is the set of experiences, values, and literary knowledge the reader brings when engaging with a text. This horizon determines how the reader interprets the structure, style, events, and emotions within a literary work. In other words, reception is historical and subjective: two readers with different backgrounds can produce different interpretations of the same text. This horizon is shaped by various factors, including the reader's understanding of genre conventions, their familiarity with previous literary works, as well as the social and cultural norms of their time period. According to Jauss (1982), the aesthetic value of a work is often determined by the "aesthetic distance," that is, the extent to which the work challenges, modifies, or even "destroys" the reader's existing horizon of expectations. A work that merely fulfills expectations tends to be considered light entertainment, while a work that radically changes the reader's perspective is considered to have high artistic value (Jauss, 1982). In the context of this research, Jauss's theory is used to explain how English Literature students, as readers with academic literary competence, evaluate texts written by AI and humans. Their literary experience horizon (e.g., reading frequency and familiarity with narrative structures), their technological horizon (familiarity with AI), and their interpretative horizon (expectations regarding emotion, language complexity, coherence, or stylistics) are crucial factors in the formation of reception. All the research questionnaire instruments are designed based on these components of the horizon of expectations, making Jauss's theory not only a conceptual but also a methodological foundation.

Previous studies have examined how readers respond to AI-generated texts compared to human texts. The research of Fiedler and Döpke analyzes whether human experts can identify AI-generated academic texts more accurately than current machine-based detectors. Conducted as a survey experiment at a German university of applied sciences, 63 lecturers in engineering, economics, and social sciences were asked to evaluate short excerpts (200–300 words) from both human-generated and AI-generated texts. These texts varied by discipline and writing level (student vs. professional)

with the AI-generated content. The results show that both human evaluators and AI detectors correctly identified AI-generated texts only slightly better than chance, with humans achieving a recognition rate of 57 % for AI texts and 64 % for human-generated texts (Fiedler & Döpke, 2025). While Schmitz & Sonnlaitner administered texts with similarly high quality, showing reviewers being unable to consistently identify authorship origins (Schmitz & Sonnleitner, 2025). Related to the emotional resonance and authenticity, Paulson & Reeds highlight human-only composition resulting in deeper emotional resonance and authenticity, particularly in expressing nuanced feelings and personal experiences (Paulson & Reeds, 2025). Another research by Makki says that AI-generated content has important limitations like the lack of lived experience, cultural context, and emotional authenticity (Makki, 2025). For the complexity of the writing, Durak's findings show that human-generated articles exhibit higher average singular word usage and longer sentence lengths compared to AI-generated articles, suggesting a more complex and nuanced language structure in human writing (Yildiz Durak et al., 2025).

Some researchers have explored the impacts of AI on English literature students. 22 students from Universitas Negeri Medan were examined regarding their frequency of AI usage, purposes, and perceptions of its impact on academic skills. The findings indicate that AI tools are predominantly used for essay writing, grammar correction, idea generation, and translation. (Fahira et al., 2024). A study in the English literature classroom at the undergraduate level in India argues that the facilitator's efficient pedagogical design can transmute Artificial Intelligence into a collaborating agent in the teaching-learning process and mitigate the epistemological and ethical questions that it potentially entails (Roy & Putatunda, 2023). The subjects of this study are the students of the English literature study program in Universitas Ngudi Waluyo who receive several subjects of English literature during their study. Some lecturers of the English literature study program at Universitas Ngudi Waluyo conducted research using fictional media such as movies or games with the students as the subjects. The studies show that the movie can improve students' listening competence (Susilowati et al., 2023) and the game can increase their speaking skill (Saptanto & Wibowo, 2018). However, it still lacks of the study about the reception of literary works

The previous studies have investigated some comparisons between text made by humans and AI. However, they have not applied the reception theory of Jauss to explain how the reader's background experience and expectations influence the reception of AI and human texts. Therefore, this research offers a new insight into the reception of the academic readers with literary analysis competence, such as English Literature students. Therefore, this research aims to empirically investigate the impact of this crisis of interpretation of literary works through a comparative study. Using a mixed-methods approach, this research will design and administer a questionnaire intended to capture rich quantitative and qualitative data. The primary objective is to answer a series of fundamental research questions:

1. What are the descriptions of the experience and technological horizons of the students?
2. What is the description of the interpretative horizon of the students?
3. How do the experience, technological, and interpretative horizons influence their interpretation and evaluation of a literary text?

METHOD

This research employed a mixed-methods approach, combining quantitative and qualitative analysis, and used a descriptive research design. The theoretical framework was based on Jauss's concept of

the horizon of expectations, which was directly applied in the construction of the questionnaire as the research instrument. The participants consisted of 16 4th-semester of English Literature students. This group was selected because they have possessed literary competence and reading experience that can influence their aesthetic horizon. It used two flash fictions made by Guy Fletcher (2025) entitled "Anna" and AI, as the research instruments, followed by a questionnaire based on the flash fictions. The complete text of "Anna" is as follows:

It was the second anniversary of her departure and Josh had never recovered. Suddenly he saw her across the busy street, unmistakably the woman he had adored: slight frame and stunning long blonde locks.

He attempted to cross the road but was thwarted by an angry motorist. Josh had lost her.

He surveyed the shoppers and to his relief spotted her. He strode briskly tapping her on the shoulder. "Great to see you again Anna." However, he realised it was not Anna, fate had constructed her. Josh apologised then viewed another blonde woman…believing it was Anna (Fletcher, 2025).

This text is highly suitable for this research due to several characteristics. It has a clear, self-contained narrative with a simple plot arc: unresolved grief, mistaken sighting, failed attempt, rediscovery, realization, and repetition of the delusion. The emotional core is strong and clearly identifiable, centering on the grief, longing, and self-deception experienced by the protagonist[1]. Most importantly, the style is straightforward and accessible. The language used does not employ complicated literary devices or unusual stylistic choices, making it a fair basis for comparison against AI-generated text, which may not yet master more idiosyncratic styles.

Creating a methodologically comparable AI text is the most critical step in this research design. It is not sufficient to simply ask AI to "write a sad story". To ensure the validity of the comparison, the AI text must be generated in response to the same creative task implicitly completed by the human text. Therefore, a "mirrored prompt**"** approach will be used. This process began by analyzing the core components of "Anna": theme (grief, mistaken identity), plot (seeing someone similar to the deceased person), and constraint (around 100 words). Based on this analysis, a highly specific and controlled prompt was designed to be given to the AI model. This prompt was not intended to imitate Fletcher's style but to emulate the underlying narrative and emotional task of his story.

Proposed mirrored prompt:

"Write a 100-word flash fiction story. The plot should be about a character who is still grieving the loss of their partner exactly two years after their departure. In the present moment, the character is in a public place and suddenly sees a stranger who they mistake for their deceased partner due to a similar appearance. The character approaches them, only to realize their mistake, which reinforces their sense of ongoing loss and delusion."

This prompt served as a proxy for authorial intent. In traditional literary analysis, authorial intent is a debated concept. However, in the context of AI generation, where the model lacks "intent" in the human sense but responds to instructions based on pattern recognition, the human-crafted prompt becomes the locus of intent for the AI's creative act. By carefully mirroring the core plot and emotional beats of "Anna" in the prompt, the research gave the same "task" to both the human and AI authors. Consequently, any resulting differences in the two texts could be more confidently attributed to fundamental differences in their creative processes (human cognition vs. algorithmic

generation) rather than to different starting points or objectives. This would significantly strengthen the internal validity of the study.

To generate the text, a sophisticated and up-to-date large language model (LLM), such as OpenAI's GPT-4 or equivalent, was used. This ensures that the AI output represents the current capabilities of the technology and not an older or less capable version. Several outputs were generated from the same prompt, and the one most faithful to the prompt constraints without requiring additional human editing was selected. This step is important to maintain the integrity of pure AI authorship. The following is the text generated from the prompt above:

```
The park air was still, marking two years. Then, through the crowd, I saw the back
of his head—that familiar chestnut hair. He was here.

"Ben!" I cried, stumbling forward. My heart hammered with impossible joy.

The man turned. His face was a canvas of polite confusion, not the kind eyes I'd
memorized. "I... I'm sorry," I whispered, shrinking back.

He was just a stranger. The hope evaporated, leaving only the cold, sharp reality.
The delusion was a cruel trick, and the loss felt new, and absolute, again.
```

The flash fiction made by AI was labelled as text A, while the one made by Fletcher was labelled as text B. The questionnaire consisted of four parts: demographic information of the respondents, evaluation of text A, evaluation of text B, and respondents' attribution. Demographic information asked about the frequency of the students in reading literary works and the frequency in using AI, while the evaluation of both texts was organized to know the respondents' emotional response, text complexity, coherence, and stylistics. The last part of the questionnaire was used to know the respondents' answers to choose which text was made by humans and AI and their reasons in deciding their choice. Demographic information and questions in the evaluation were the quantitative data to know the students' literary experience horizon and technological horizon, analyzed by using a Likert scale. Demographic information as the quantitative data, was measured by stating very rarely to very often for the frequency in reading literary works, and using a Likert scale not familiar at all to very familiar, in measuring the frequency of using AI. The closed questions in the evaluation part were the quantitative data measured by using 1-5 Likert scale, ranging from 1 meaning strongly disagree, and 5 meaning strongly agree, to know their interpretative horizon. Based on the background and interpretation of the respondents, the respondents' attribution questions decided whether they could decide that text A was made by AI and text B was made by a human by stating the reasons of their choice as the qualitative data.

RESULTS AND DISCUSSION

Jauss's theory is used to explain how English Literature students in Universitas Ngudi Waluyo, as readers with academic literary competence, evaluate texts written by AI and humans. Their literary experience horizon (e.g., reading frequency and familiarity with narrative structures), their technological horizon (familiarity with AI), and their interpretative horizon (expectations regarding emotion, language complexity, coherence, or stylistics) are crucial factors in the formation of reception. The results of the study can be seen in Table 1.

E-Structural (English Studies on Translation, Culture, Literature, and Linguistics)
Vol 8 No 02
p-ISSN: 2621-8844, e-ISSN: 2621-9395
http://publikasi.dinus.ac.id/index.php/estructural

Table 1 Students' Horizon of Expectations

Student	Text	Literary experience and Technological horizons		Interpretative Horizon				Accuracy	Reasons
		Reading Freq.	Familiarity with AI	Emotion	Language Complexity	Coherence	Stylistics		
Student 1	text A	very often	quite familiar	2	3	3	1	accurate	Text A feels less natural, B really conveys longing after a long wait
	text B			4	4	4	5		
Student 2	text A	often	quite familiar	3	2	2	2	accurate	A: The plot is predictable. B: There is a plot twist at the end of the story.
	text B			5	4	4	5		
Student 3	text A	often	quite familiar	1	3	3	3	accurate	A: emotions feel shallow. B: disappointment is felt so deeply that it causes hallucinations.
	text B			5	4	4	4		
Student 4	text A	often	quite familiar	2	2	3	2	accurate	A: too neat, B: natural
	text B			3	4	4	5		
Student 5	text A	often	very familiar	2	1	3	2	accurate	A: there is a typical AI sentence style such as "head-that"
	text B			4	5	4	4		
Student 6	text A	often	quite familiar	2	1	3	1	accurate	A: The diction is too coherent, the emotion is less pronounced. B: The diction is easier to understand and capture the meaning, the emotion is more pronounced because the words are easier to comprehend
	text B			5	4	4	5		
Student 7	text A	often	quite familiar	2	2	4	2	accurate	A: neat and too stiff sentence structure. B: natural and flowing style
	text B			4	5	3	5		
Student 8	text A	sometimes	a bit familiar	4	4	3	5	not accurate	A: easy to understand, B: difficult to understand
	text B			2	2	4	2		
Student 9	text A	sometimes	a bit familiar	2	3	2	3	accurate	A: The writing style is stiff, doesn't convey any emotional feelings, and I feel empty when reading the text. B: The writing style is more varied, so I'm carried away by the atmosphere and emotional feelings the author describes
	text B			5	4	4	4		
Student 10	text A	sometimes	quite familiar	1	2	2	2	accurate	A: Formulated and non-dramatic sentences. B: Human writing because it seems to tell a real event and text B is sadder.
	text B			4	4	5	5		
Student 11	text A	sometimes	quite familiar	2	2	3	2	accurate	A: The writing style and language are very neat, the emotional feeling is not felt, there is nothing unique. B: The language is unique, very poetic, and the ideas are unexpected
	text B			4	5	4	4		
Student 12	text A	sometimes	quite familiar	3	3	4	3	accurate	A: less deep emotions
	text B			4	4	3	4		

Student 13	text A text B	sometimes	a bit familiar	4 3	4 2	4 4	4 2	not accurate	A: warmer in emotion. B: hollow and colorless
Student 14	text A text B	sometimes	a bit familiar	4 2	4 3	3 2	4 3	not accurate	A: coherent language. B: the end of the story seems to be separated from the main story
Student 15	text A text B	very rarely	a bit familiar	5 3	4 2	3 3	4 4	not accurate	A: uses subtle, more poetic emotional descriptions, with a variety of sentence lengths. B: is AI-generated because the flow is too fast, some transitions feel stiff.
Student 16	text A text B	rarely	a bit familiar	4 3	5 3	3 3	3 2	not accurate	A: simple and easy to understand language. B: complicated language

Experience and Technological Horizons of the Students

The experience horizon, represented by reading frequency, shows that most students possess moderate literary exposure. Out of 16 students, 7 students (43.75%) report reading "often", 7 students (43.75%) report reading "sometimes", 1 student (6.25%) reads "very often", and 2 students (12.5%) read "rarely" or "very rarely." This distribution indicates that nearly 87.5% of participants possess at least a moderate level of reading exposure, while only a small minority have minimal literary experience.

Students who read *often* or *very often* demonstrate a more refined awareness of narrative structure, emotional layering, and stylistic variation indicated by 8 of the 11 of them gave accurate identifications (72.7%). Their evaluations frequently reference abstract literary qualities such as predictability, emotional subtlety, stylistic stiffness, or narrative flow, indicating an internalized set of literary norms formed through repeated reading experiences. In contrast, among the 5 students who produced inaccurate identifications, 4 students (80%) read "sometimes," "rarely," or "very rarely.". For these readers, a text that is simple, clear, and immediately understandable is more likely to be valued positively, even if it lacks stylistic complexity or emotional depth. This difference highlights how the experience horizon shapes expectations about what literature should accomplish: while experienced readers expect aesthetic complexity and emotional resonance, less experienced readers prioritize clarity and directness.

Alongside literary experience, the technological horizon, measured through familiarity with AI, plays a crucial role in shaping interpretation. Of the 16 students, 8 students (50%) report being "quite familiar" with AI, 7 students (43.75%) are "a bit familiar," and 1 student (6.25%) is "very familiar." Students with higher AI familiarity tend to articulate more precise critiques related to linguistic patterns, such as "typical AI sentence style," excessive coherence, or formulaic diction. These readers appear to have developed an awareness of algorithmic writing tendencies, enabling them to recognize textual features they associate with artificial generation. Conversely, students who are only *a bit familiar* with AI often rely on assumptions rather than experiential knowledge. Their comments reveal generalized beliefs that AI writing is either overly complex, overly fast-paced, or emotionally hollow. Because their technological horizon is limited, these students interpret texts primarily through emotional immediacy and readability rather than stylistic or structural markers of artificiality. Thus, the technological horizon functions as a cognitive filter that shapes how textual cues are decoded and evaluated.

Taken together, the experience and technological horizons are unevenly distributed among the students, producing diverse starting points for interpretation. These horizons do not operate independently; instead, they intersect to form a composite framework through which each reader approaches and evaluates the texts.

Interpretative Horizon of the Students

The interpretative horizon refers to the criteria and expectations students use when assigning meaning and value to the texts. In this study, it is reflected in students' ratings of emotion, language complexity, coherence, and stylistics, as well as in their qualitative justifications.

1. **Emotional Evaluation**

Emotion is the most decisive interpretative dimension. Across all students, Text B consistently receives higher emotional scores, typically ranging from 4 to 5, while Text A's emotional scores cluster between 1 and 3 for most accurate readers. Among the 11 accurate students, 10 students (90.9%) give Text B an emotion score of 4 or higher, whereas Text A receives scores of 2 or lower from 8 of these students (72.7%). Among the 5 inaccurate readers, however, 4 students (80%) assign Text A higher emotional scores than Text B, with emotion ratings for Text A ranging from 4 to 5. This reversal demonstrates an alternative interpretative horizon in which explicit emotional cues or warmth are valued more highly than emotional subtlety or depth. Text B consistently receives higher emotional scores and is frequently described as "sad," "deep," "longing," and "atmospheric." Many students report being emotionally carried by Text B, suggesting that emotional resonance is central to their conception of literary quality. Text A, by contrast, is often described as emotionally shallow, flat, or empty. Even when it receives moderate emotional scores, students frequently qualify their evaluations by noting a lack of depth or authenticity. This pattern indicates that, within the dominant interpretative horizon, emotion is not merely about presence but about intensity, subtlety, and the ability to evoke affective engagement. For many students, emotional depth functions as a key marker of human authorship and literary legitimacy.

2. **Language Complexity and Coherence**

Language complexity further differentiates interpretative horizons. For accurate readers, Text B receives complexity scores of 4–5 from 9 students (81.8%), while Text A receives scores of 1–2 from 7 students (63.6%). In contrast, inaccurate readers frequently reverse this pattern. Among the 5 inaccurate students, 4 students (80%) rate Text A's language complexity at 4 or higher. Students who accurately identify the texts tend to value balanced, complexity language that is expressive and varied without being obscure. They criticize Text A for being "too neat," "formulated," or "predictable," suggesting that excessive regularity disrupts their aesthetic expectations. In contrast, Text B is often praised for its poetic language, varied sentence structures, and unexpected expressions. These evaluations reflect an interpretative horizon that associates human writing with imperfection, irregularity, and stylistic flexibility. However, this interpretative horizon is not shared by all students. Some participants, particularly those with lower reading frequency, interpret complexity as difficulty rather than aesthetic value. For them, Text B is described as "difficult to understand" or narratively disjointed, while Text A is praised for being clear and coherent. This indicates an alternative interpretative horizon in which accessibility and straightforwardness are prioritized over stylistic nuance.

Coherence shows a similar numerical ambivalence. While Text B generally scores 4–5 in coherence among accurate readers, several students explicitly criticize Text A for being "too coherent." This

paradox indicates that within the dominant interpretative horizon, excessive coherence (perceived as mechanical regularity) is evaluated negatively, whereas coherence combined with narrative tension is valued positively. Comments describing Text A as "too coherent" or "too neat" suggest that minor inconsistencies, ambiguities, or narrative disruptions are expected features of human storytelling. Text B's occasional unpredictability and narrative tension align more closely with these expectations. Nevertheless, students who misidentify the texts often criticize Text B for perceived incoherence, indicating that their interpretative horizon favors linearity and structural clarity.

3. Stylistics as a Decisive Interpretative Marker

Stylistics emerges as the strongest differentiating variable. Among accurate readers, Text B receives stylistic scores of 4–5 from all 11 students (100%), while Text A receives scores of 1–2 from 8 students (72.7%). Among inaccurate readers, stylistic evaluations are more evenly distributed. 3 of the 5 inaccurate students (60%) give Text A stylistic scores of 4 or higher, indicating that stylistic regularity aligns with their expectations of good writing. Numerically, this confirms that stylistics functions as a key interpretative battleground where different horizons collide. Text A is frequently associated with stiffness, rigid sentence structures, and uniform diction, all of which trigger associations with AI-generated writing. Text B, on the other hand, is described as natural, flowing, and poetic. These stylistic perceptions strongly influence emotional engagement and authorship judgments, demonstrating that stylistics is a central component of the students' interpretative horizon.

Influence of Experience, Technological, and Interpretative Horizons on Interpretation

When the numerical patterns are considered holistically, a clear relationship emerges between horizons and interpretation. Students with higher reading frequency and higher AI familiarity are significantly more likely to accurately identify authorship (11 of 16 students, or 68.75%), assign higher emotional, stylistic, and complexity scores to Text B, and critique Text A for excessive coherence and stylistic rigidity.

Students with higher reading frequency and greater familiarity with AI tend to demonstrate accurate identification and more nuanced evaluations. Their experience horizon equips them with sensitivity to literary conventions, while their technological horizon enables them to recognize algorithmic patterns. As a result, their interpretative horizon aligns closely with dominant literary norms that associate human authorship with emotional depth, stylistic variation, and narrative unpredictability. For these students, Text A's excessive neatness and coherence create an aesthetic distance, leading to lower emotional engagement and negative stylistic evaluations. Text B, by contrast, fulfills their expectations of literary authenticity, resulting in higher scores across interpretative dimensions. Their accuracy, therefore, is not accidental but emerges from the alignment between their horizons and the textual features of the human-authored text.

In contrast, students with lower reading frequency and limited familiarity with AI are overrepresented among inaccurate readers (5 of 16 students, or 31.25%). Their interpretations prioritize clarity, explicit emotion, and linear coherence, leading them to evaluate Text A more positively and misidentify its authorship. Students with limited reading experience and lower AI familiarity are more likely to produce inaccurate identifications. Different priorities, such as ease of comprehension, explicit emotion, and structural clarity guide their evaluations. Because their interpretative horizon does not emphasize stylistic irregularity or emotional subtlety, they may perceive Text A as more human-like and Text B as overly complex or confusing. These misidentifications should not be interpreted as interpretative failure, but rather as evidence of a different horizon of expectations.

Overall, the findings demonstrate that students' interpretations and evaluations of literary texts are shaped by the convergence of their experience, technological, and interpretative horizons. The experience horizon influences sensitivity to literary aesthetics, the technological horizon shapes awareness of artificial writing patterns, and the interpretative horizon mediates how emotional, linguistic, and stylistic cues are valued. Together, these horizons determine whether a text is perceived as authentic, emotional, artificial, or mechanical. From a reception-theory perspective, the study confirms that literary meaning and authorship judgments are not inherent in the text but are constructed through reader–text interaction. The same textual features can be interpreted as poetic or artificial, coherent or rigid, depending on the reader's horizons. This underscores the importance of considering reader diversity when examining responses to AI-generated literature and highlights the fluid, negotiated nature of literary interpretation in the age of artificial intelligence.

CONCLUSION

This study, framed by Hans Robert Jauss's reception theory, provides compelling empirical evidence regarding how academic readers evaluate literary flash fiction created by human and algorithmic authors. The analysis of the interpretative horizon clearly demonstrates that the human-written text (text B) is consistently perceived as qualitatively superior across all dimensions: emotion, stylistics, language complexity, and coherence compared to the AI-generated text (text A). The most significant aesthetic distance resides in emotion and stylistics. Readers primarily judge authenticity based on the text's affective resonance, rejecting text A for its "shallow emotions" and "too neat" or "formulated" structure. This indicates that the current capabilities of AI language models still fail to meet the reader's deeply entrenched expectation for affective depth and stylistic idiosyncrasy that characterizes human creativity. By applying Jauss's theory, the study concludes that the reader's interpretative horizon, shaped by their literary experience horizon, is the crucial locus of authentication. The finding that 11 out of 16 students accurately identify the AI text (68.75% accuracy), which is significantly higher than previously reported rates in less specialized fields. This suggests that academic literary competence functions as a highly effective filter, training the reader to be both a "critic and a detective" capable of discerning the aesthetic cues of algorithmic generation.

Despite offering valuable insights into students' literary reception of human- and AI-generated flash fiction, this study has several limitations that should be acknowledged when interpreting the findings. First, the sample size is limited, involving only 16 students**.** While this number allows for in-depth qualitative interpretation and close reading of individual responses, it restricts the generalizability of the results. The numerical patterns observed, such as the relationship between reading frequency, AI familiarity, and interpretative accuracy, should therefore be understood as indicative rather than representative of broader student populations. Second, the study relies on self-reported measures of reading frequency and familiarity with AI. Such measures are inherently subjective and may not accurately reflect students' actual literary engagement or technological competence. For instance, students who report being "quite familiar" with AI may differ significantly in their practical exposure or understanding of AI-generated texts, potentially influencing the consistency of the technological horizon variable. In light of the limitations identified in this study, several directions for future research are recommended to strengthen and extend the findings. First, future studies should involve a larger and more diverse sample size**.** Expanding the number of participants across different academic disciplines, educational levels, or institutions would enhance the generalizability of the results. A broader sample would also allow researchers to examine whether patterns observed in this study, such as the relationship between reading frequency, AI familiarity, and interpretative accuracy, remain consistent across different reader populations. Second, to address the limitations of self-

reported data**,** future research could incorporate objective measures of literary and technological experience. For instance, reading experience can be assessed through reading logs, comprehension tests, or literary exposure inventories, while familiarity with AI technology can be measured through task-based assessments or brief knowledge tests. Such approaches would provide a more reliable representation of students' experience and technological horizons.

REFERENCES

Fahira, A. N., Siagian, C. R., Simarmata, D. Y., Manik, S. F., & Natsir, M. (2024). The impact of AI usage in supporting English Literature students' learning (2022). *Indonesian Journal of Education and Development Research*, *3*(1), 756–762. https://doi.org/10.57235/ijedr.v3i1.4921

Fiedler, A., & Döpke, J. (2025). Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses. *International Review of Economics Education*, *49*, 100321. https://doi.org/10.1016/j.iree.2025.100321

Hrytsak, N., Bartish, S., Kuchma, N., Liubinetska, M., & Nehodiaieva, S. (2025). Intermediality and literary reception in the digital age: The impact of modern media on interpretation. *International Journal on Culture, History, and Religion*, *7*(SI1), 225–241. https://doi.org/10.63931/ijchr.v7iSI1.163

Jauss, H. R. (1982). *Toward an aesthetic of reception*. Univ. of Minnesota Pr.

Kor, R. (2018). The commenting persona: Reception theory and the digital rhetorical audience. *Journal of Media Research*, *11*(1 (30)), 55–70. https://doi.org/10.24193/jmr.30.4

Liliani, E., Kusmarwanti, K., & Budiyanto, D. (2022). Perkembangan fiksimini Indonesia 2010 – kini. *LITERA*, *21*(2), 227–239. https://doi.org/10.21831/ltr.v21i2.45039

Makki, M. (2025). Modern poetry and artificial intelligence investigating how Al-generated poetry is shaping perceptions of creativity and authorship. *International Journal for Scientific Research*, *4*(5), 46–64. https://doi.org/10.59992/IJSR.2025.v4n5p3

Masih, T. L. (Ed.). (2009). *Field guide to writing flash fiction: Tips from editors, teachers, and writers in the field*. Rose Metal Press.

McCormack, L. (2021). Flash fiction: A study in temporality. *New Writing*, *18*(1), 2–20. https://doi.org/10.1080/14790726.2019.1710217

Paulson, D., & Reeds, M. (2025). *A comparative study of AI-assisted and human-only poetic composition techniques*. Computer Science and Mathematics. https://doi.org/10.20944/preprints202506.2320.v1

Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, *78*(1), 84–93. https://doi.org/10.1177/10776958221149577

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, *20*(2). https://doi.org/10.53761/1.20.02.07

Rahmeh, H. (2023). Digital verses versus inked poetry: Exploring readers' response to AI-generated and human-authored sonnets. *Scholars International Journal of Linguistics and Literature*, *6*(09), 372–382. https://doi.org/10.36348/sijll.2023.v06i09.002

Roche-Jacques, S. (2024). Flash fiction as a distinct literary form: Some thoughts on time, space, and context. *New Writing*, *21*(2), 171–189. https://doi.org/10.1080/14790726.2023.2293767

Roy, D., & Putatunda, T. (2023). From textbooks to chatbots: Integrating AI in English literature classrooms of India. *Journal of E-Learning and Knowledge Society*, 65-73 Pages. https://doi.org/10.20368/1971-8829/1135860

Saptanto, D. D., & Wibowo, T. A. (2018). Relasi game warspear dan kemampuan berbicara serta kosakata mahasiswa Sastra Inggris Universitas Ngudi Waluyo. *PHILOSOPHICA Jurnal Bahasa, Sastra, Dan Budaya*, *1*(1), 31. https://doi.org/10.35473/po.v1i1.120

Schmitz, L. M. R. Y., & Sonnleitner, P. (2025). Evaluating AI-generated vs. human-written reading comprehension passages: An expert SWOT analysis and comparative study for an educational large-scale assessment. *Large-Scale Assessments in Education*, *13*(1), 20. https://doi.org/10.1186/s40536-025-00255-w

Susilowati, E., Saptanto, D. D., Alkaromah, L., & Bila Fauziyah Tamami, S. (2023). Improving students' ability in analyzing a work using popular movie. *Philosophica: Jurnal Bahasa, Sastra, Dan Budaya*, *6*(1), 75–83. https://doi.org/10.35473/po.v6i1.2548

Willis, I. (2021). Reception theory, reception history, reception studies. In I. Willis, *Oxford Research Encyclopedia of Literature*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190201098.013.1004

Yildiz Durak, H., Eğin, F., & Onan, A. (2025). A comparison of human-written versus AI -generated text in discussions at educational settings: Investigating features for ChatGPT, Gemini and BingAI. *European Journal of Education*, *60*(1), e70014. https://doi.org/10.1111/ejed.70014