What Does AI Do for Cultural Interpretation? A Randomized Experiment on Close Reading with Exposure to AI

Jiayin Zhi
jzhi@uchicago.edu
University of Chicago
Chicago, Illinois, USA

Hoyt Long
hoytlong@uchicago.edu
University of Chicago
Chicago, Illinois, USA

Richard Jean So
richard.so@duke.edu
Duke University
Durham, North Carolina, USA

Mina Lee
mnlee@uchicago.edu
University of Chicago
Chicago, Illinois, USA

Abstract

AI demonstrates unprecedented reasoning capabilities, but its increasing integration into reading via automated summarization and interpretation has provoked debate about its use for cultural interpretation. Close reading—the practice of understanding, analyzing, and critiquing cultural texts for pleasure—is at the core of such interpretation. To test AI's impact on close reading, we conducted a preregistered randomized experiment ($n = 400$) investigating the impact of AI assistance by presenting single or multiple AI interpretations, on close reading poems, compared to no AI assistance. We found that a single AI interpretation boosted both interpretive performance and pleasure, while multiple AI interpretations only improved performance. Further exploration revealed a trade-off: participants who heavily relied on AI showed better performance but lower pleasure. These findings contribute to discussion on how to design AI tools that augment both the interpretive performance and experiential value of reading cultural texts: *"less is more."* This paper is an encore of a CHI '26 full paper.[1]

1 Introduction

The rise of Generative AI has provoked both optimism and fear around its impacts on human reasoning [29, 32, 35]. The debate revolves around the question of how much of human reasoning can or should be delegated to AI without making our own contributions irrelevant or unnecessary in the process.

While much of this debate has focused on the implications of AI for how we reason via writing [18–20, 26], less attention has been given to the kinds of reasoning that happen via reading—especially when the materials are poems, songs, stories or movies, which demand focused and complex forms of interpretive attention. Recent work showed the benefits of large language models (LLMs) *simplifying* large volumes of text to make it more accessible [2, 37], but do these same benefits accrue to our ability to *interpret* culture and arts? And if the point of reading a poem or watching a movie is because you find pleasure in how it makes you think and feel, is there really much to be gained by having an LLM do it for you?

This specific kind of interpretive skill is called *close reading*: the ability to understand, explain, interpret, evaluate, and critique culture works [1, 31]—in textual form like poems and novels, or in other media like songs and films. Though often associated with and widely taught within humanities education, people also practice close reading in everyday life. For instance, social media influencers rack up millions of followers by simply explaining popular TV shows and films [28]. In this broader sense, close reading becomes a marker of social engagement and cultural awareness [5, 11].

Recent work in AI has begun exploring LLM capabilities in domains requiring aesthetic judgment [6, 16, 30, 31, 33, 38], demonstrating the capabilities of state-of-the-art AI models for interpretive reasoning. The possibility that AI might automate close reading is controversial and has opened up a fierce debate among writers, creators, journalists, humanities professors, as well as common consumers of culture [21, 36]. Few doubt that AI is a useful tool to automate instrumental tasks. But many question whether AI should be used to automate the task of interpretation—finding meaning in cultural texts—because it is typically imagined to be the thing that makes us human, one that cannot benefit from automation since its primary value lies in providing personal pleasure. Many fear that AI will somehow corrupt or diminish what is posited to be an exclusive human skill [36]. As a matter of fact, there is already evidence of AI-generated interpretations being incorporated alongside poems on poetry platforms [24], making this investigation into AI's impact on close reading particularly timely.

To push the boundary of this discussion, this work seeks to better understand how close reading is affected by the integration of AI into our reading and reasoning. Given the importance of close reading as a social skill, our study focuses on lay readers, and on poems as cultural texts. We investigate the impact of AI assistance powered by an LLM on close reading by examining two essential elements: *interpretive performance* and the *pleasure* derived from the process. We focus on identifying stylistic features and explaining their effects within the text, which is the first necessary step and foundation for effective close reading [34], and operationalize interpretive performance through feature identification, interpretation quality, and writing quality. Building on close reading scholarship [1, 4, 13] and intrinsic rewards theory [7, 8], we conceptualize the pleasure of close reading as arising from discovering personally resonant meanings, enjoying the interpretive puzzle-solving process, and feeling empowered to make sense of complex texts. By this view, we operationalize these three sources of pleasure as three interrelated subjective experience constructs: appreciation, enjoyment, and self-efficacy [1, 20].

[1] https://doi.org/10.1145/3772318.3791727

2 Method

2.1 Study Design

In a preregistered[2] randomized controlled experiment ($n = 400$) with crowdworkers on Prolific, we examined how different amounts of AI assistance influence individuals' interpretive performance and the pleasure they derive from close reading by comparing three conditions:

(1) AI-Single, which presented a single AI interpretation (Figure 1, center);
(2) AI-Multiple, which offered multiple AI interpretations stacked on top of each other (i.e., shown one at a time, with the top interpretation fully visible; Figure 1, right);
(3) Control, which did not provide any AI interpretation (Figure 1, left).

The AI-Single condition reflects the design used on poetry platforms, which display one AI interpretation beneath the poem [24]. The AI-Multiple condition tests the effect of providing *more* AI assistance by having additional interpretations accessible, aligning with the open-ended nature of close reading.

2.1.1 Procedure. In the study, each participant completed interpretation tasks for three poems in random order in their randomly-assigned condition. The interpretation tasks, adapted from the Critical Reader's Interpretive Toolkit (CRIT) [34], focused on identifying stylistic features from the poems and analyzing their effects. After completing the task for each poem, they rated their appreciation and enjoyment of the poem, as well as their sense of self-efficacy in interpretation. A post-task survey asked participants about their approach and rationale for using or not using the AI assistance provided, followed by overall study feedback.

2.1.2 Reading Materials. We selected three poems through expert curation and iterative group discussion: *"Love Poem"* [23], *"Dusting"* [22], and *"Theme for English B"* [15]. Selection criteria prioritized readability for lay readers while ensuring diversity across multiple dimensions: themes and topics (love, nature, social identity, etc.), poetic styles, and author backgrounds.

2.1.3 AI Interpretations Curated. For the AI-Single condition, we used a widely-used general purpose model at the time (GPT-4o) to generate one response for each poem using the exact instructions given to participants in the interpretation tasks. In other words, the LLM automated the same close reading task that participants were asked to complete. For the AI-Multiple condition, we prompted GPT-4o to generate three distinct interpretations. Each AI interpretation identified different stylistic features and explained their effects accordingly. Our graders evaluated all AI interpretations using the same rubrics applied to participants.

2.2 Dependent Variables

2.2.1 Interpretative Performance. Participants completed three interpretation tasks per poem. Two trained graders evaluated participants' response to each interpretation task based on the scoring rubric, assigning three types of scores for Interpretive Performance: **Feature Identification**: Whether the identified feature was correct (0 = incorrect/missing, 1 = correct); **Interpretation Quality**: Depth and insight of the explanation (1 = poor, 3 = average, 5 = excellent); **Writing Quality**: Clarity and coherence of expression (1 = poor, 3 = average, 5 = excellent). For analysis, we averaged each score type across the three tasks within each poem, yielding three Interpretative Performance measures per poem per participant.

2.2.2 Subjective Experience. After completing the interpretation tasks for each poem, participants rated their Subjective Experience on three 7-point Likert scales: **Appreciation**: Rating on the question "To what degree did this poem resonate with you?" (1 = strongly negative reaction, 7 = strongly positive reaction); **Enjoyment**: Rating on the question "How much did you enjoy reading the poem?" (1 = very unenjoyable, 7 = very enjoyable); **Self-efficacy**: Rating on the question "How confident are you in your ability to interpret the poem?" (1 = strongly unable, 7 = strongly able). Thus, we had three Subjective Experience measures (Appreciation, Enjoyment, Self-efficacy) per poem per participant.

2.3 Participants

We recruited 405 participants from Prolific. We have 400 participants in all after excluding those who completed any of the poems in excessively short time (3 SDs below the mean). Based on a power analysis using pilot and simulated data, this sample size is needed for 80% power, a medium effect size, using a significant level of 0.05. The overall study took around one hour to complete and each was paid £9.5. Participants had a 100% approval rate on Prolific, based in the US, using English as their primary language and fluent in English. The study was approved by the Institutional Review Board (IRB) of our institution. Our randomized experiment yielded 400 participants in all: 141 in the Control, 115 in the AI-Single, and 144 in the AI-Multiple ($\chi^2(2) = 3.815$, $p = 0.149$).

2.4 Analysis

2.4.1 Grading Process. We employed two expert graders, both with undergraduate degrees in English and one a PhD student in English. For inter-rater reliability, we stratified a validation set of 70 participants (17.5% of the total sample) drawn equally from all three experimental conditions. For Feature Identification accuracy, the agreement between graders was 85.4%. For Interpretation Quality and Writing Quality, the correlation coefficient ICC(2,1) is 0.76, exceeding the recommended threshold of 0.70 [12, 14].

2.4.2 Statistical Analysis. For Interpretive Performance measures, we fitted mixed-effects linear regression models with conditions, expertise, poem types, poem positions as fixed effects, and participant as a random effect. For Subjective Experience measures, we used mixed-effects ordinal regression models with the same fixed and random effects structure.

3 Results

Our findings reveal that exposure to AI assistance in the form of a single AI interpretation boosted both performance and pleasure, while multiple AI interpretations improved performance but did not increase pleasure.

[2] https://aspredicted.org/555c-y7kz.pdf

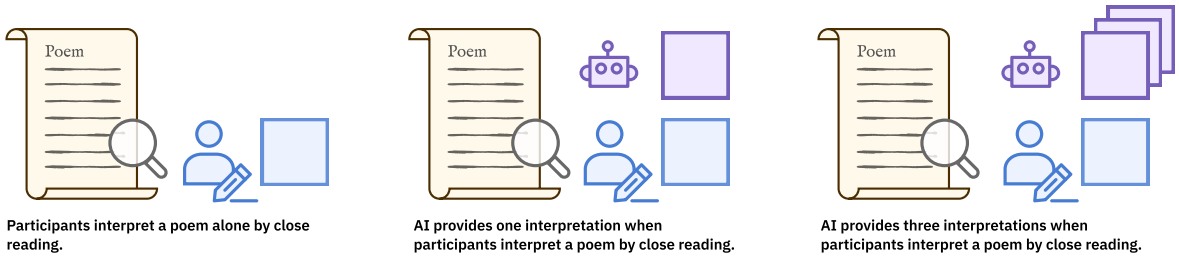


Figure 1: Illustration of the three experimental conditions. Participants in all conditions interpret poems through close reading, with randomly assigned AI assistance: (left) Control condition where participants interpret alone, (center) AI-Single condition where one AI interpretation is provided, and (right) AI-Multiple condition where three AI interpretations are provided.

	Feature Identification	Interpretation Quality	Writing Quality
Condition (ref: Control):			
AI-Single	0.048* [0.002, 0.094]	0.865*** [0.664, 1.065]	1.035*** [0.817, 1.253]
AI-Multiple	0.051* [0.008, 0.095]	0.593*** [0.404, 0.782]	0.784*** [0.578, 0.989]
Poem Position (ref: Position 1):			
Position 2	0.017 [-0.009, 0.043]	0.006 [-0.059, 0.071]	0.040 [-0.022, 0.102]
Position 3	0.010 [-0.015, 0.036]	0.084* [0.019, 0.149]	0.111*** [0.049, 0.173]
Poem Type (ref: Love Poem):			
Dusting	-0.003 [-0.029, 0.023]	-0.017 [-0.082, 0.048]	-0.024 [-0.086, 0.038]
Theme for English B	0.074*** [0.048, 0.100]	-0.110*** [-0.175, -0.045]	-0.089** [-0.151, -0.027]
Participant Expertise (ref: Inexperienced):			
Experienced	0.046* [0.010, 0.083]	0.061 [-0.099, 0.220]	0.055 [-0.118, 0.228]

Note: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1: Mixed-effects linear regression models predicting Interpretive Performance measures (Feature Identification, Interpretation Quality, and Writing Quality) from conditions, poems and participant expertise. We report coefficients with 95% confidence intervals in brackets and highlight significant effects that are consistent on all the three performance measures.

	Appreciation	Enjoyment	Self-Efficacy
Condition (ref: Control):			
AI-Single	0.899*** [0.522, 1.277]	0.969*** [0.531, 1.407]	0.900** [0.356, 1.443]
AI-Multiple	-0.070 [-0.416, 0.276]	-0.248 [-0.645, 0.149]	-0.352 [-0.856, 0.152]
Poem Position (ref: Position 1):			
Position 2	-0.127 [-0.384, 0.130]	-0.042 [-0.303, 0.219]	-0.163 [-0.436, 0.110]
Position 3	-0.096 [-0.356, 0.164]	0.010 [-0.256, 0.276]	-0.118 [-0.393, 0.157]
Poem Type (ref: Love Poem):			
Dusting	-0.125 [-0.379, 0.129]	-0.360** [-0.623, -0.097]	-0.196 [-0.471, 0.079]
Theme for English B	0.802*** [0.537, 1.067]	0.431** [0.164, 0.698]	0.199 [-0.073, 0.472]
Participant Expertise (ref: Inexperienced):			
Experienced	-0.114 [-0.409, 0.181]	0.148 [-0.191, 0.487]	0.191 [-0.237, 0.619]

Note: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2: Mixed-effects ordinal regression models predicting Subjective Experience measures (Appreciation, Enjoyment, and Self-Efficacy) from conditions, poems and participant expertise. We report coefficients as log-odds with 95% confidence intervals in brackets and highlight significant effects that are consistent on all the three subjective experience measures.

3.1 Interpretive Performance

Table 1 shows a consistent pattern that both AI-Single and AI-Multiple significantly enhanced all Interpretive Performance measures. Compared to the Control condition, the AI-Single condition showed a significant positive effect on Feature Identification (0.048, 95% CI: [0.002, 0.094], $p = 0.042$), Interpretation Quality (0.865, 95% CI: [0.664, 1.065], $p < 0.001$), and Writing Quality (1.035, 95% CI: [0.817, 1.253], $p < 0.001$). The AI-Multiple condition also showed significant positive effects compared to Control across all performance measures: Feature Identification (0.051, 95% CI: [0.008, 0.095], $p = 0.020$), Interpretation Quality (0.593, 95% CI: [0.404, 0.782], $p < 0.001$), and Writing Quality (0.784, 95% CI: [0.578, 0.989], $p < 0.001$). It is worth noting that the AI-Single condition showed larger effect sizes than AI-Multiple across all measures.

3.2 Pleasure Derived From Interpretation

Table 2 presents a consistent pattern across Enjoyment, Appreciation, and Self-Efficacy: AI assistance by presenting a single interpretation consistently enhanced all three aspects related to pleasure, while AI assistance by presenting multiple interpretations had no statistical significance. Compared to the Control condition, the AI-Single condition showed a significant positive effect on Appreciation ratings (0.899, 95% CI: [0.522, 1.277], $p < 0.001$), Enjoyment ratings (0.969, 95% CI: [0.531, 1.407], $p < 0.001$), and Self-Efficacy ratings (0.90, 95% CI: [0.356, 1.443], $p = 0.001$). There was no significant difference between the AI-Multiple and the Control.

4 Behavioral Engagement

To better understand these results, we further investigate how participants engaged with AI assistance using descriptive analyses of both behavioral logs and self-reported data.

4.1 Did Participants View the Multiple AI Interpretations?

In the AI-Multiple condition, participants were shown three AI interpretations stacked together with one fully visible by default and two more signaled by labeled buttons that participants could click to view. Table 3 shows that participants viewed only the default AI interpretation in 42.1% of instances, clicked to view a second AI interpretation in 12.3% of instances, and clicked through all three in 45.6% of instances, suggesting limited engagement with the multiple AI interpretations. While participants who viewed multiple AI interpretations showed better Interpretive Performance, they had lower Subjective Experience compared to those who viewed only one AI interpretation in the AI-Multiple or those in the AI-Single and Control. Notably, those who viewed only one AI interpretation in the AI-Multiple showed lower Interpretive Performance and Subjective Experience, compared to those in the AI-Single. Results of exploratory analyses aligned with this: after adjusting for the number of AI interpretations viewed, participants in the AI-Multiple showed lower Interpretive Performance and Subjective Experience than AI-Single. This pattern suggests that the mere visible presence of multiple AI interpretations may have negative effects.

The availability of multiple AI interpretations did not translate into a better subjective experience for participants, as one might expect given the wider selection. Instead, multiple AI interpretations might trigger feelings of inadequacy and competition, harming the subjective experience, as P27 in the AI-Multiple noted in their open-ended responses reflecting on their use of the AI assistance: *"All the answers by the AI again interpreted the poem much better than I could've. I only had a surface level interpretation. [...]"* Many chose not to view additional interpretations to preserve their self-confidence. P36 in the AI-Multiple explained this preservation strategy: *"I was tempted to read the other ones to make sure we hadn't come up with the same examples. I felt like the AI was a threat that I had to do better than. However, I rationalized that if I didn't read what it wrote at all, then I would know for certain that it had not influenced my interpretation whatsoever."* The limited engagement with and reluctance to explore multiple AI interpretations may explain the absence of pleasurable benefits for AI-Multiple.

4.2 Did Participants Report Using AI?

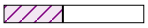
After interpreting each poem, participants responded to a question about whether they used the AI for interpretation. Table 4 shows that a proportion of participants reported not using AI assistance despite its availability (AI-Single: 45.2%; AI-Multiple: 53.0%). Within the AI-Single or AI-Multiple condition, participants who reported using AI demonstrated better Interpretive Performance but consistently had lower Subjective Experience compared to those who did not report AI use. This pattern reflects a performance-pleasure tradeoff: those who acknowledged using AI showed better analytical output but diminished subjective rewards. Notably, participants in both AI-Single and AI-Multiple who did not report using AI showed better Interpretive Performance than those in Control. This aligns with our main analyses and suggests that participants may have learned from the AI implicitly, even when they did not explicitly acknowledge "using" it.

4.3 How Did Participants Incorporate AI Interpretation?

To understand how participants might incorporate AI interpretation(s), we examined textual overlap between participant answer and the AI interpretations they were exposed to, as well as copy-paste behavior. We identified three different categories: (1) complete overlap (instances where participants' interpretations were identical to the AI interpretation), (2) high overlap (instances where textual overlap exceeded the medium rate of 57.1% among all participants in the AI-Single and AI-Multiple), and (3) low overlap (instances below the medium textual overlap rate). Table 5 reveals a consistent trade-off between performance and experience. Participants with high textual overlap achieved substantially better Interpretive Performance scores, yet had lower Subjective Experience ratings. This pattern was most pronounced in Complete overlap instances, where these participants achieved the highest Interpretive Performance scores, but the lowest Subjective Experience ratings, particularly in the AI-Multiple. Conversely, participants with Low overlap had the highest Subjective Experience ratings despite lower Interpretive Performance scores. This inverse relationship between performance and experience suggests that closely following AI interpretations improves objective performance on the surface, but at the expense of personal satisfaction and confidence.

5 Discussion

Reading serves many purposes, and augmenting it requires attending to what readers actually value. For close reading, and cultural interpretation in general—where the process of making meaning is itself the reward—more AI assistance can improve interpretive performance while undermining the experiential value that motivates people to read. The performance-pleasure trade-off we observed suggests that the challenge is about how to present AI's interpretive capability. Presenting maximal AI capabilities risks outsourcing not just the cognitive work but the pleasure of discovery and meaning-making that make cultural interpretation worthwhile. This points to a core design insight: *"less is more."* Designers should carefully calibrate AI support to maintain human engagement, leaving sufficient interpretive space for personal discovery—even if this means accepting lower objective performance.

Condition	**Viewing Behavior** (%)	*Interpretive Performance* Feature	Interpretation	Writing	*Subjective Experience* Appreciation	Enjoyment	Self-Efficacy
AI-Single	 100.0%	0.85 (0.23)	2.52 (0.95)	2.71 (1.03)	5.99 (0.98)	5.91 (1.01)	5.92 (1.11)
AI-Multiple	 42.1%	0.82 (0.24)	1.91 (0.85)	2.11 (0.88)	5.37 (1.63)	5.48 (1.50)	5.53 (1.48)
	 12.3%	0.91 (0.17)	2.62 (0.92)	2.86 (1.05)	5.32 (1.59)	5.22 (1.52)	5.09 (1.92)
	 45.6%	0.87 (0.22)	2.46 (1.01)	2.68 (1.11)	5.30 (1.37)	5.28 (1.47)	5.20 (1.70)
Control	N/A	0.80 (0.27)	1.65 (0.75)	1.68 (0.76)	5.43 (0.98)	5.49 (1.15)	5.59 (1.11)

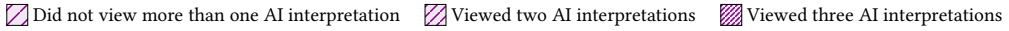


Table 3: Percentage of instances viewing different numbers of AI interpretations in the AI-Multiple condition, and summaries of Interpretive Performance and Subjective Experience measures (means with standard deviations in brackets). The AI-Single and Control are included for reference.

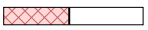
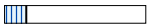
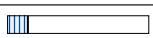
Condition	**Self-Reported AI Use** (%)	*Interpretive Performance* Feature	Interpretation	Writing	*Subjective Experience* Appreciation	Enjoyment	Self-Efficacy
AI-Single	 45.2%	0.92 (0.40)	2.92 (0.89)	3.22 (0.95)	5.81 (1.49)	5.67 (1.70)	5.65 (1.69)
	 54.8%	0.79 (0.36)	2.19 (0.88)	2.29 (0.91)	5.96 (1.08)	6.11 (1.29)	6.14 (1.04)
AI-Multiple	 53.0%	0.89 (0.21)	2.58 (0.96)	2.83 (1.05)	5.25 (1.47)	5.19 (1.55)	5.04 (1.82)
	 47.0%	0.91 (0.34)	1.87 (0.86)	2.04 (0.88)	5.43 (1.55)	5.57 (1.39)	5.66 (1.36)
Control	N/A	0.80 (0.27)	1.65 (0.75)	1.68 (0.76)	5.43 (0.98)	5.49 (1.15)	5.59 (1.11)

Table 4: Percentage of instances of self-reported AI use in each condition, and summaries of Interpretive Performance and Subjective Experience measures (means with standard deviations in brackets). The Control is included for reference.

Condition	**Overlap with AI (%)**	*Interpretive Performance* Feature	Interpretation	Writing	*Subjective Experience* Appreciation	Enjoyment	Self-Efficacy
AI-Single	 15.4%	0.93 (0.14)	3.56 (0.47)	4.03 (0.33)	5.36 (1.81)	4.96 (2.01)	5.04 (2.09)
	 25.2%	0.83 (0.22)	2.38 (0.84)	2.56 (0.95)	5.87 (1.28)	6.00 (1.43)	6.01 (1.32)
	 59.4%	0.83 (0.24)	2.30 (0.92)	2.44 (0.93)	6.04 (1.07)	6.12 (1.28)	6.11 (1.08)
AI-Multiple	 14.4%	0.95 (0.12)	3.36 (0.65)	3.81 (0.70)	5.27 (1.55)	5.16 (1.77)	4.76 (2.09)
	 31.0%	0.91 (0.17)	2.40 (0.85)	2.53 (0.86)	5.25 (1.47)	5.30 (1.57)	5.21 (1.85)
	 54.6%	0.79 (0.25)	1.87 (0.86)	2.07 (0.92)	5.40 (1.52)	5.47 (1.36)	5.55 (1.33)
Control	N/A	0.80 (0.27)	1.65 (0.75)	1.68 (0.76)	5.43 (0.98)	5.49 (1.15)	5.59 (1.11)

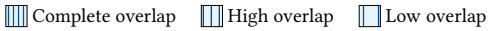


Table 5: Percentage of instances of different extent of textual overlap (complete overlap, high overlap, and low overlap), and summaries of Interpretive Performance and Subjective Experience measures (means with standard deviations in brackets). The Control is included for reference.

Two theoretical lenses support these insights. First, the gradual release of responsibility model [3, 9, 10, 25, 39] describes effective learning as a progression from observing a demonstration to practicing independently. A single AI interpretation may function like a teacher's demonstration—providing scaffolding that readers can internalize before transitioning to independent discovery, where personal satisfaction emerges. Multiple interpretations, by contrast, may keep readers in the observation phase, exhausting the interpretive space and leaving little room for personal meaning-making. Second, participants who reported not using the AI still outperformed the control group, consistent with abstract modeling [3, 17, 27, 39]—implicitly extracting principles from observed examples without conscious adoption. This suggests that modest exposure to AI may be enough to elevate reading without displacing the reader's own interpretive work.

Our findings also open broader questions for future research. First, our study examined one interaction paradigm: AI-generated interpretations presented alongside poems, reflecting how poetry platforms currently provide AI assistance online, and tested whether providing more AI assistance in this form helps. But readers may

also actively seek AI support for interpretation, such as through freeform AI chatbots. How should we design AI support for close reading when the interaction shifts from passive exposure to active dialogue, and can such designs enhance interpretive performance while still preserving the pleasure? Second, reading spans a spectrum of purposes—from utility-driven tasks like scanning legal documents to deeply interpretive engagement with cultural texts. Our findings suggest that what counts as good augmentation can depend on what readers value in the first place. For close reading, that value extends well beyond comprehension and interpretation to include enjoyment, discovery, and interpretive agency. Other forms of reading may carry different but equally important experiential stakes—for instance, the cognitive exercise of learning through reading, or the skill development through engaging with challenging texts. How might research on augmenting reading better account for the diverse values and purposes of what readers seek, and design AI support that enhances rather than bypasses these valued processes?

References

[1] Meyer Howard Abrams, Geoffrey Galt Harpham, and Geoffrey Galt Harpham. 1999. *A glossary of literary terms*. Vol. 369. Harcourt Brace College Publishers Fort Worth.

[2] Elliott Ash, Aniket Kesari, Suresh Naidu, Lena Song, and Dominik Stammbach. 2024. Translating legalese: enhancing public understanding of court opinions with legal summarizers. In *Proceedings of the 2024 Symposium on Computer Science and Law*. 136–157. https://doi.org/10.1145/3614407.3643700

[3] Albert Bandura. 1986. *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Englewood Cliffs, NJ. doi:10.4135/9781446221129.n6

[4] Don Bialostosky. 2006. Should college English be close reading? *College English* 69, 2 (2006), 111–116. https://doi.org/10.58680/ce20065837

[5] Donal Carbaugh. 1991. Communication and cultural interpretation. (1991), 336–342. https://doi.org/10.1080/00335639109383965

[6] Kate Crawford. 2021. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

[7] Mihaly Csikszentmihalyi. 2014. Play and intrinsic rewards. In *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*. Springer, 135–153.

[8] Mihaly Csikszentmihalyi. 2015. Intrinsic rewards and emergent motivation. In *The hidden costs of reward*. Psychology Press, 205–216.

[9] Nell K Duke and P David Pearson. 2009. Effective practices for developing reading comprehension. *Journal of education* 189, 1-2 (2009), 107–122. https://doi.org/10.1177/0022057409189001-208

[10] Douglas Fisher and Nancy Frey. 2021. *Better learning through structured teaching: A framework for the gradual release of responsibility*. ASCD.

[11] Clifford Geertz. 2017. *The Interpretation of Cultures*. Basic Books.

[12] Natasa Gisev, J Simon Bell, and Timothy F Chen. 2013. Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy* 9, 3 (2013), 330–338. https://doi.org/10.1016/j.sapharm.2012.04.004

[13] John Guillory. 2025. On close reading. In *On Close Reading*. University of Chicago Press.

[14] Kevin A Hallgren. 2012. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology* 8, 1 (2012), 23. doi:10.20982/tqmp.08.1.p023

[15] Langston Hughes. 2002. Theme for English B. In *The Collected Works of Langston Hughes*. https://www.poetryfoundation.org/poems/47880/theme-for-english-b Originally published in 1951. Reprinted by permission of Harold Ober Associates, Inc..

[16] Jessica Hullman, Ari Holtzman, and Andrew Gelman. 2023. Artificial intelligence and aesthetic judgment. *arXiv preprint arXiv:2309.12338* (2023). https://doi.org/10.48550/arXiv.2309.12338

[17] John F Kihlstrom. 1987. The cognitive unconscious. *Science* 237, 4821 (1987), 1445–1452. doi:10.1126/science.3629249

[18] Bart Lamiroy and Emmanuelle Potier. 2022. Lamuse: Leveraging artificial intelligence for sparking inspiration. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*. Springer, 148–161. https://doi.org/10.1007/978-3-031-03789-4_10

[19] Byung Cheol Lee and Jaeyeon Chung. 2024. An empirical investigation of the impact of ChatGPT on creativity. *Nature Human Behaviour* 8, 10 (2024), 1906–1914. https://doi.org/10.1038/s41562-024-01953-1

[20] Jack McGuire, David De Cremer, and Tim Van de Cruys. 2024. Establishing the importance of co-creation and self-efficacy in creative collaboration with artificial intelligence. *Scientific Reports* 14, 1 (2024), 18525. https://doi.org/10.1038/s41598-024-69423-2

[21] Rod J. Naquin. 2024. Close reading with NotebookLM. Substack. https://rodjnaquin.substack.com/p/close-reading-with-notebooklm Accessed: 2025-08-25.

[22] Marilyn Nelson. 1994. Dusting. Published in *Magnificat*. https://poets.org/poem/dusting Reprinted by the Academy of American Poets. Used with permission.

[23] Linda Pastan. 2005. Love Poem. The Writer's Almanac with Garrison Keillor. https://writersalmanac.publicradio.org/index.php%3Fdate=2005%252F02%252F12.html Originally from *The Imperfect Paradise*, © W.W. Norton. Reprinted with permission.

[24] Linda Pastan. n.d.. Love Poem. https://allpoetry.com/poem/14373834-Love-Poem-by-Linda-Pastan. AllPoetry.com.

[25] P David Pearson and Margaret C Gallagher. 1983. The instruction of reading comprehension. *Contemporary educational psychology* 8, 3 (1983), 317–344. https://doi.org/10.1016/0361-476X(83)90019-X

[26] Nitin Rane and Saurabh Choudhary. 2024. Role and challenges of ChatGPT, Google Bard, and similar generative Artificial Intelligence in Arts and Humanities. *Studies in Humanities and Education* 5, 1 (2024), 1–11. https://doi.org/10.48185/she.v5i1.999

[27] Arthur S Reber. 1996. *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford University Press, New York. https://doi.org/10.1093/acprof:oso/9780195106589.001.0001

[28] Screen Daily. 2023. How Social Media Influencers Are Transforming Film Marketing. https://www.screendaily.com/features/how-social-media-influencers-are-transforming-film-marketing/5187462.article Accessed: 2025-08-25.

[29] Anjali Singh, Karan Taneja, Zhitong Guan, and Avijit Ghosh. 2025. Protecting human cognition in the age of AI. *arXiv preprint arXiv:2502.12447* (2025).

[30] Kevin Stowe, Prasetya Utama, and Iryna Gurevych. 2022. IMPLI: Investigating NLI models' performance on figurative language. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, 5375–5388. doi:10.18653/v1/2022.acl-long.369

[31] Peiqi Sui, Juan Diego Rodriguez, Philippe Laban, J Dean Murphy, Joseph P Dexter, Richard Jean So, Samuel Baker, and Pramit Chaudhuri. 2025. KRISTEVA: Close reading as a novel task for benchmarking interpretive reasoning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 32829–32849.

[32] Lev Tankelevitch, Elena L Glassman, Jessica He, Aniket Kittur, Mina Lee, Srishti Palani, Advait Sarkar, Gonzalo Ramos, Yvonne Rogers, and Hari Subramonyam. 2025. Understanding, Protecting, and Augmenting Human Cognition with Generative AI: A Synthesis of the CHI 2025 Tools for Thought Workshop. *arXiv preprint arXiv:2508.21036* (2025).

[33] Xiaoyu Tong, Rochelle Choenni, Martha Lewis, and Ekaterina Shutova. 2024. Metaphor Understanding Challenge Dataset for LLMs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 3517–3536. doi:10.18653/v1/2024.acl-long.193

[34] University of Texas at Austin, Department of English. 2024. The Critical Reader's Toolkit. Web page. https://liberalarts.utexas.edu/english/the-critical-reader-s-toolkit.html Accessed: 2025-08-25.

[35] Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. 2024. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour* 8, 12 (2024), 2293–2303. doi:10.1038/s41562-024-02024-1

[36] Marc Watkins. 2023. AI Reading Assistance: A Revolutionary Tool or a Threat to Close Reading Skills? Substack. https://marcwatkins.substack.com/p/ai-reading-assistance-a-revolutionary Accessed: 2025-08-25.

[37] Ziang Xiao, Tiffany Wenting Li, Karrie Karahalios, and Hari Sundaram. 2023. Inform the uninformed: improving online informed consent reading with an AI-powered Chatbot. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17. https://doi.org/10.1145/3544548.3581252

[38] Ruibin Yuan, Hanfeng Lin, Yi Wang, Zeyue Tian, Shangda Wu, Tianhao Shen, Ge Zhang, Yuhang Wu, Cong Liu, Ziya Zhou, et al. 2024. ChatMusician: Understanding and Generating Music Intrinsically with LLM. In *Findings of the Association for Computational Linguistics: ACL 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 6252–6271. doi:10.18653/v1/2024.findings-acl.373

[39] Barry J Zimmerman and Anastasia Kitsantas. 1997. Developmental phases in self-regulation: Shifting from process to outcome goals. *Journal of Educational Psychology* 89, 1 (1997), 29–36.