I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17646

AI Bias for Creative Writing: Subjective Assessment Versus Willingness to Pay

Martin Abel
Reed Johnson

JANUARY 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17646

AI Bias for Creative Writing: Subjective Assessment Versus Willingness to Pay

Martin Abel
Bowdoin College and IZA

Reed Johnson
Bowdoin College

JANUARY 2025

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

AI Bias for Creative Writing: Subjective Assessment Versus Willingness to Pay*

How do perceptions of AI versus human authorship affect engagement with creative work? In an incentivized experiment, participants (N=654) assessed the content of a short story labeled as either human or AI-generated and reported their willingness to pay and work to finish reading it. Consistent with prior research, the AI-labeled story received significantly lower content assessments. However, the time people invest in reading the story and their willingness to pay and work did not differ between the labels, even for the 36% of participants who profess to value human over AI writing. These findings raise questions about whether subjective assessments and aspirations to favor human authorship translate into actions.

JEL Classification: D12, O33, Z11

Keywords: Artificial Intelligence, creative writing, willingness to pay, AI bias

Corresponding author:
Martin Abel
Bowdoin College
255 Maine Street
Brunswick
Maine 04011
USA

E-mail: m.abel@bowdoin.edu

* This paper greatly benefited from comments by Simeon Abel, Jason Brown, Jeff Carpenter, Eric Chown, Christopher Robert, and Daniel Stone. Carly Davey and Aidan Singer provided outstanding research assistance. The experiment and analysis plan were registered under registry number AEARCTR-0014949. IRB approval was obtained from Bowdoin College. All errors and omissions are our own. Declarations of interest: none.

1 Introduction

We are experiencing an explosion of AI-generated content. From professional and social media platforms [1, 2] to journalism [3, 4, 5] and poetry [6, 7, 8], Large Language Models (LLMs) are producing text that is increasingly indistinguishable from human writing. Recent evidence suggests that AI may even be perceived as "more human than human" [9, 1, 8].

The AI revolution has the potential to reshape the market for creative works and creative labor. Similar to the introduction of automation in industries that previously relied on skilled human labor, the adoption of AI promises to dramatically reduce production costs and thus expand access to creative products to broader audiences. While economists have long argued that such transformations increase overall welfare, it would come at the cost of substantial employment losses among creative workers and wider concerns about the erosion of meaning and human creativity.

Consumer behavior will ultimately determine the impact of emerging AI systems capable of generating vast quantities of text at near-zero cost [10]. One key question is whether consumers' preferences and willingness to pay (WTP) depend on the *process* of how a product was created.

For visual art, recent studies have found that humans have difficulty distinguishing between AI and human-created work and, moreover, that when they are unlabeled, humans actually tend to prefer AI-generated works [11, 12, 13]. By contrast, findings indicate holding the artwork constant while changing the labeling as AI- or human-generated causes assessments biased against AI along various qualitative measures such as authenticity, novelty, and meaningfulness [14, 15, 16, 17, 18]. Similar patterns emerge from studies on creative writing, which have focused more on poetry. One recent study finds that "ratings are significantly lower when told the poem is AI-generated but are significantly higher when the poem is actually AI-generated" [8].

These studies consistently document an AI bias in people's assessments and stated valuation of creative content, even when engaging with the same work. However, to our knowledge, previous studies have not yet measured people's actual willingness to invest time or money in consuming AI-generated creative works. Addressing this knowledge gap is important given the well-documented divergence between hypothetical and revealed valuations [19, 20].

2 Methods

We recruited a sample of 654 participants based in the US through Prolific in December 2024.[1] We informed people at the beginning that the purpose of the study was to "better understand how people engage with and value creative writing." The expected time was twelve minutes for which they received a base pay of 2.50 USD.

Table 1: Summary Statistics and Balance

Variable	N	Full Sample Mean/SE	AI Mean/SE	Human Mean/SE	P-value AI-Hum
Age	630	45.61	45.01	46.218	0.340
		(0.632)	(0.892)	(0.896)	
Female	654	0.488	0.517	0.459	0.138
		(0.020)	(0.028)	(0.028)	
College Degree	654	0.509	0.502	0.517	0.696
		(0.020)	(0.028)	(0.028)	
Liberal	654	0.401	0.401	0.401	1.000
		(0.019)	(0.027)	(0.027)	
Conservative	654	0.332	0.312	0.352	0.281
		(0.018)	(0.026)	(0.026)	
Follow AI (0-6)	653	3.597	3.581	3.613	0.796
		(0.063)	(0.088)	(0.089)	
Nr Books	649	9.236	8.926	9.545	0.540
		(0.504)	(0.687)	(0.739)	
F-test of joint significance (p-value)					0.730

Notes: The table presents summary statics for characteristics specified in our analysis plan. The last column presents p-values from a test of whether characteristics are balanced in the AI and human writing group.

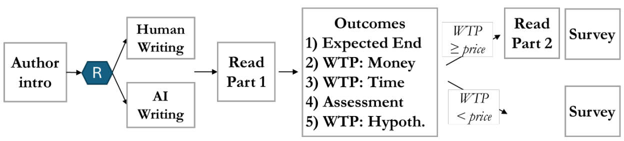
The sample is nationally representative in terms of key demographic characteristics such as age, race, and gender. The average age is 45.6 years, 48.8% identify as female, and 50.9% completed a college degree (Table 1). 40.1% identify as liberal and 33.2% as conservative. People

[1] About 6% of the targeted sample size of 700 did not pass the attention or bot detection check.

report 3.6 (on a 0-6 scale) in terms of how much they follow news about AI, and the average (median) number of books read per year is 9.2 (4).

Figure 1 summarizes our experimental research design. All participants first read a short biographic summary of the acclaimed author Jason Brown (see Figure C1 for details) and are informed that they will "read a short story that is representative of Jason Brown's writing." We additionally note that the story is unpublished.

Figure 1: Experimental Design



The short story was, in fact, written by OpenAI's GPT-4. We first prompted GPT-4 to summarize Brown's writing style and published work. After documenting its familiarity with his writing, we asked GPT-4 to create a short story of less than a thousand words that is representative of Jason Brown's work. We specified that the story should be about a professor struggling with the rise of AI.

While all participants read the *identical* story, participants were randomly assigned in equal shares into one of two groups. One group does not receive any additional information and is thus made to believe that the story was composed by the author. The other group is informed that it "was written by Artificial Intelligence (ChatGPT)" (see Figure C2).[2] Table 1 suggests randomization was successful and that the two groups are balanced.

The story begins with a professor reading an email from a student who has used AI to generate ideas for her fiction-writing assignment, which makes him reflect on his own struggles with completing his novel. At the point where the professor starts to reply to the student, we interrupt the story and administer a set of questions. First, we ask participants how they thought the story would continue: what the professor will advise the student to do, whether he himself will use AI for his writing, and if so, how satisfied he will be with his decision.

Next, we elicit participants' maximum WTP to read the end of the story along two dimensions: i) what part of their fifty-cent bonus they are willing to have deducted, and ii) how much time (0-6 minutes) they are willing to work transcribing text. Measuring both WTP dimensions provides a more comprehensive understanding as people may value money and time differently due to socioeconomic circumstances.[3] To make the WTP elicitation incentive-compatible, we inform participants that we will randomly choose either the money or time category and then randomly select the price level within this category (Figure C5). If their WTP is at least as large as this price, they will read the end of the story and will be "charged" the price at the end of the study.[4]

Participants next assess the story content by choosing how much they agree (on a 0 to 10 scale) with the following seven statements presented in random order: "I was interested in the struggles of these characters," "The writing was predictable," "The writing was atmospheric; I felt like I was on the scene," "The characters are nuanced and complex," "The idea of the story was creative and original," "The narrator sounds authentic," and "This story deserves to be published in a top literary journal." These present common evaluative dimensions of fiction writing—from characterization to narratorial voice and atmosphere/transportation—found in both creative writing pedagogy [23] and research on narrative engagement [24, 25]. And indeed, these measures are highly predictive

[2]To further emphasize the treatment variation, we included an image of a person writing on a typewriter in the control group, and an image of a robot writing on a laptop under each message.

[3]Time-based WTP reveals the opportunity cost people associate with reading the story. Monetary WTP may disproportionately reflect the valuation of participants who can afford to spend money. The correlation coefficient of 0.5 suggests that these two metrics measure distinct dimensions of WTP. Research also suggests that time-based WTP is more reliable predictor of behavior than monetary-based WTP [21]. We randomized the order of price categories and asked participants to move a slider that is initially set at the mean value.

[4]This Becker, DeGroot and Marshak (BDM) method is a commonly used incentive compatible way to elicit willingness to pay [22]. To facilitate participants' understanding of the process, we provide them with two examples.

of participants' WTP (p-v.<0.001).[5]

Participants with a WTP equal to or larger than the randomly selected price of one minute transcription work next read the end of the story before answering survey questions related to their familiarity with and attitudes toward AI.They then transcribed a provided text for one minute. Participants whose WTP was below one minute proceeded to the survey questions without reading the story. Last, we provide all participants with information about the experimental design and research questions.

3 Results

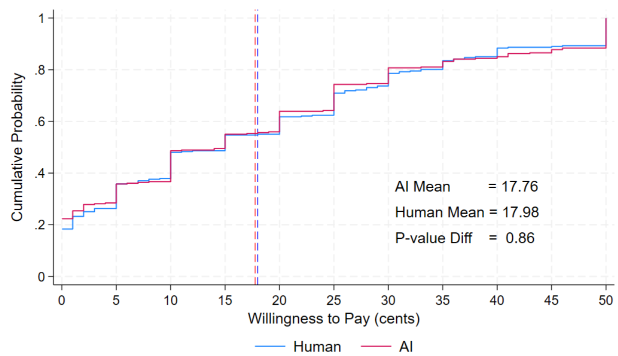
The experimental design allows us to test the effect of changing perceptions of the author of the story, while holding the content constant. As specified in our pre-registration, we divide results into incentivized and non-incentivized outcomes.[6]

3.1 Willingness to Pay and Time Investment

We test the effect on three primary incentivized outcomes: monetary WTP, work WTP, and the amount of time participants invest in reading the story. Figure 2 shows the cumulative distribution and mean values of monetary WTP for the human and AI group. The difference in means of 0.2 cents (1.1%) is not significant (p-v.=0.86) and we can reject that the two distributions are different (p-v.=0.96) The same pattern holds for the time people are willing to work: the difference in means of 0.04 minutes (1.7%) is small and not significant (p-v.=0.74). Results are robust to controlling for demographic characteristics (Table A1).

We collect data on the exact time participants spend reading the story, which serves as our primary measure of attention.[7] We find a small and insignificant drop in reading time of around 2% in the AI group (Table A4).[8]

Figure 2: Willingness to Pay - Money: CDF



Notes: The figure shows the cumulative density function of monetary WTP for human and AI writing. The dashed lines present mean values.

To compare effect sizes, we standardize outcomes and present differences between the AI and human writing group in standard deviations (sd). The top of Figure 3 shows that differences for all incentivized outcomes are less than 0.05 sd. The WTP estimates are relatively precise: we can rule out that effects of AI are larger than 0.18 sd with 95% confidence.

3.2 Story Assessment

Figure 3 shows that there are large differences in the quality assessments of writing: AI is perceived as more predictable and scores lower on measures of emotional engagement, creativity/originality, authenticity, atmosphere and literary merit. Effects are sizable and largest for authenticity, literary merit and atmosphere. The effect on a standardized quality index, our pre-specified primary outcome, is 0.28 sd (p-v.<0.001).

We hypothesized that one of the reasons why people may value the story differently is that they think that AI

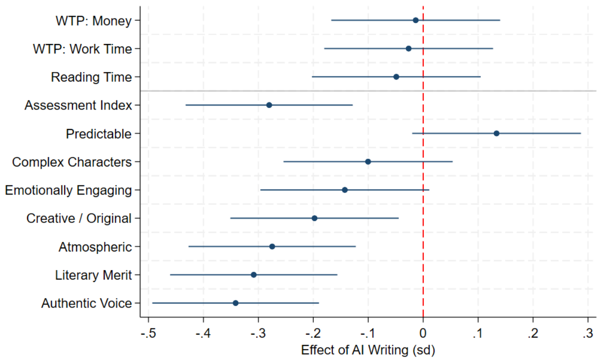
[5]A one sd increase in the assessment index is associated with a 4.8 cent (27%) increase in WTP (Figure B2)

[6]The analysis plan, registered before data collection started, can be found at www.socialscienceregistry.org/trials/14949.

[7]Participants are not informed we collect this data. We impose a minimum reading time of 30 seconds. This is unlikely to affect many readers as the median reading time is close to 2.5 minutes and less than 6% spent less than 1 minute. For the analyze we winsorize the time measure at the 5% level to account for outliers.

[8]Measuring effects on time in addition to monetary investment is important, given that fiction competes as a form of entertainment in a crowded attentional space. In contrast to our findings about time spent reading, we find that the share of participants who remember a detail of the story (related to who said a certain quote) is 6.6 pp (7.7%) lower in the AI group, suggesting that they read the AI-labeled story differently — a question that is outside the scope of the present study.

Figure 3: Effects of AI Writing (standardized)



Notes: The figure presents treatment coefficients of AI writing with 95% confidence intervals for standardized outcome measures. Effects are measured in standard deviations.

changes the *content*. We indeed find that participants in the AI group are 8.5 pp (25%) more likely to believe that the professor will use AI for his writing and that he will be happier with the result (Table A3). We also find a small increase in the share believing that the professor will advise the student to use AI, although this difference is not significant. We interpret this as suggestive evidence that while participants think AI will change the content of writing (in a self-serving way), it does not lead to a difference in WTP.

4 Discussion

4.1 Awareness of Preferences

To assess whether people are aware of their preferences, following questions about WTP and story assessment, we asked participants in the AI (human) writing group whether they thought their WTP would have been different had the story been written by a human (AI). If they answered in the affirmative, we asked for the reason. Participants could select from a pre-specified list of reasons or enter their own response (Figure C6).

61.9% of respondents say their WTP would not have changed. Another 36.6% state they would be willing to pay more for human writing, while only 1.5% say they would have paid more for AI writing. Among those who value human writing more, the most common reasons were that it is more meaningful (81.1%), that they care about the livelihood of human writers (63.9%) and that they have ethical or environmental concerns about AI (32.5%).

Figure 4 presents how results vary with people's professed preferences. Consistent with their stated beliefs, participants who claim to value human writing more (top panel) give a 0.5 sd lower content assessment in the AI group.[9] While they spend slightly less time (0.2 sd) reading the story, their difference in WTP is close to zero and not significant for both money and time. By contrast, results for those professing to value AI and human writing equally (bottom panel) are more consistent: differences for all incentivized outcomes and the assessment index are close to zero and not significant. Comparing assessment dimensions between these two groups provides insights into why people state they value human writing more. We find the largest differences for predictability and authenticity, while effects for creativity/originality and emotional engagement are small and not significant.

4.2 Study Limitations

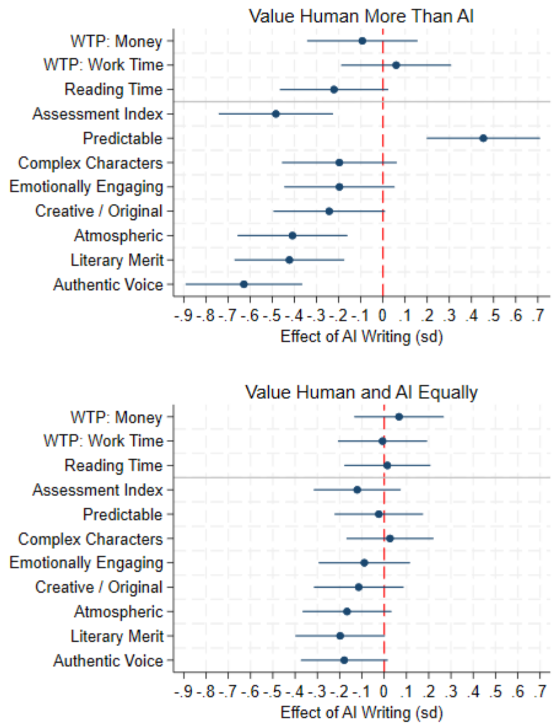
While our sample is nationally representative in terms of gender, age, and race, online samples tend to be more educated, technologically proficient, and may be distinct in their reading habits, which could affect the generalizability of our findings. However, our analysis reveals no significant differences in the effect of AI writing across participants' gender, age, education levels, readership, and familiarity with AI (Table A6). The latter result also mitigates concerns that our results are driven by short-lived "novelty effects" among people unfamiliar with AI.[10]

A second study limitation is that we elicit outcomes at the midpoint of the story, after participants have engaged with the writing. Our results may thus not extend to decisions whether to *begin* reading a story labeled as

[9] The correlation between "anthropocentric views of creativity" and anti-AI bias is similarly documented in visual art [15].

[10] The fact that less than 1% of participants reported they would be willing to pay more for an AI-generated story because they "are curious about AI" provides further evidence against this explanation.

Figure 4: Results by Stated Preferences



Notes: The figure presents treatment coefficients of AI writing with 95% confidence intervals for standardized outcome measures. The top panel includes people who say they value human more than AI writing. The sample in the bottom panel are people who say they value human and AI writing equally.

AI-generated, especially for those who hold an AI bias. We attempt to address this limitation by measuring reading time, as participants make this decision starting from the outset. Another feature of our study design that may limit the generalizability is that people's WTP decisions were made privately. In real-world contexts, people may decide to pay a premium for human-generated stories as a means to publicly signal certain values or virtues.

A final potential concern is that our incentivized outcomes may not fully capture the satisfaction people derive from reading the story and that those in the AI group were, in fact, less satisfied. To address this, we analyze participants' effort levels during the transcription task. Since we specified only the work time (Figure C7), the number of words transcribed serves as a measure of effort not contractually specified, which is often linked to intrinsic motivation [26]. We find that the number of words in the AI group is 19.4% (p-v.=0.019) *higher*, suggesting that participants did not express dissatisfaction with the reading experience by exerting less effort.

5 Implications

Consistent with previous studies, we find that many people hold a bias against AI when assessing creative writing. However, our findings reveal that this bias may not translate into differences in WTP or the time they invest in reading.

Before considering the far-reaching implications this could have for creative work markets, a note of caution. By keeping the story constant, we isolate the effects of labeling. However, this paradigm precludes any head-to-head comparisons between actual human- and AI-generated stories. While non-experts struggle to distinguish AI- and human-generated creative work in a variety of media—such as visual art [27] and poetry [6]—similar comprehensive studies have yet to be done for fiction writing. AI's current capacity to generate longer-format texts is still limited, including its ability to craft complex plots, develop nuanced characters and maintain thematic consistency and authorial voice [28].[11] At the same time, LLMs are evolving rapidly and AI-generated works are already flooding the market [30].[12]

One possible scenario is that AI-generated content will come to dominate markets. As AI models continue to improve, they can offer products perceived as equivalent (or even superior) at a fraction of the price, displacing human writers. Alternatively, we may witness a segmentation of markets, where some consumers focus solely on the

[11] AI stories in the aggregate also do not show the same creative range as human works, which may limit the ability of LLMs to create work in genres that place high value on originality [29].

[12] The influence of gatekeepers such as retailers and editors who can control market access may further limit the adoption of AI-generated creative writing. For example, online retailers have started to require labeling for AI-generated books, though the effects of labeling on sales has not yet been studied.

price while others also take into account the *process* of how a product is created. A similar dynamic has occurred in the demand for hand-made goods, which command a premium over machine-made products even when their functionality is the same. This focus on the process of creation may be understood as a cultural response to the widespread adoption of machine manufacturing, similar to the Arts and Crafts movement around the turn of the 19th century [31].[13]

In short, the shape of our future literary marketplace will likely be shaped by the collective decisions of consumers. While attitudes toward AI are still evolving, our study provides timely evidence that widely documented professed AI biases may not be reliable indicators of willingness to pay for human creative labor.

References

[1] Maurice Jakesch, Jeffrey T Hancock, and Mor Naaman. Human heuristics for ai-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11):e2208839120, 2023.

[2] Kristina Radivojevic, Nicholas Clark, and Paul Brenner. Llms among us: Generative ai participating in digital discourse. In *Proceedings of the AAAI Symposium Series*, volume 3, pages 209–218, 2024.

[3] Vaclav Moravec, Nik Hynek, Marinko Skare, Beata Gavurova, and Matus Kubak. Human or machine? the perception of artificial intelligence in journalism, its socio-economic conditions, and technological developments toward the digital future. *Technological Forecasting and Social Change*, 200:123162, 2024.

[4] Sarah Kreps, R Miles McCain, and Miles Brundage. All the news that's fit to fabricate: Ai-generated text as a tool of media misinformation. *Journal of experimental political science*, 9(1):104–117, 2022.

[5] Anja Wölker and Thomas E Powell. Algorithms in the newsroom? news readers' perceived credibility and selection of automated journalism. *Journalism*, 22(1):86–103, 2021.

[6] Nils Köbis and Luca D Mossink. Artificial intelligence versus maya angelou: Experimental evidence that people cannot differentiate ai-generated from human-written poetry. *Computers in human behavior*, 114:106553, 2021.

[7] Jimpei Hitsuwari, Yoshiyuki Ueda, Woojin Yun, and Michio Nomura. Does human–ai collaboration lead to more creative art? aesthetic evaluation of human-made and ai-generated haiku poetry. *Computers in Human Behavior*, 139:107502, 2023.

[8] Brian Porter and Edouard Machery. Ai-generated poetry is indistinguishable from human-written poetry and is rated more favorably. *Scientific Reports*, 14(1):26133, 2024.

[9] Daphne Ippolito, Daniel Duckworth, Chris Callison-Burch, and Douglas Eck. Automatic detection of generated text is easiest when humans are fooled. *arXiv preprint*, 2019.

[10] Xiang Hui, Oren Reshef, and Luofeng Zhou. The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Organization Science*, 35(6):1977–1989, 2024.

[11] Harsha Gangadharbatla. The role of ai attribution knowledge in the evaluation of artwork. *Empirical Studies of the Arts*, 40(2):125–142, 2022.

[12] Simone Grassini and Mika Koivisto. Understanding how personality traits, experiences, and attitudes shape negative bias toward ai-generated artworks. *Scientific Reports*, 14(1):4113, 2024.

[13] Jeongeun Park, Hyunmin Kang, and Ha Young Kim. Human, do you think this painting is the work of a real artist? *International Journal of Human–Computer Interaction*, 40(18):5174–5191, 2024.

[14] Lucas Bellaiche, Rohin Shahi, Martin Harry Turpin, Anya Ragnhildstveit, Shawn Sprockett, Nathaniel Barr, Alexander Christensen, and Paul Seli. Humans versus ai: whether and why we prefer human-created compared to ai-created artwork. *Cognitive Research: Principles and Implications*, 8(1):42, 2023.

[15] Kobe Millet, Florian Buehler, Guanzhong Du, and Michail D Kokkoris. Defending humankind: Anthropocentric bias in the appreciation of ai art. *Computers in Human Behavior*, 143:107707, 2023.

[16] C Blaine Horton Jr, Michael W White, and Sheena S Iyengar. Bias against ai art can enhance perceptions of human creativity. *Scientific reports*, 13(1):19001, 2023.

[17] Alwin de Rooija. Bias against artificial intelligence in visual art: A meta-analysis. 2024.

[18] Martin Ragot, Nicolas Martin, and Salomé Cojean. Ai-generated vs. human artworks. a perception bias towards artificial intelligence? In *CHI conference on human factors in computing systems*, pages 1–10, 2020.

[19] James J Murphy, P Geoffrey Allen, Thomas H Stevens, and Darryl Weatherhead. A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30:313–325, 2005.

[20] Jerrod M Penn and Wuyang Hu. Understanding hypothetical bias: An enhanced meta-analysis. *American Journal of Agricultural Economics*, 100(4):1186–1206, 2018.

[21] Ned Augenblick, Muriel Niederle, and Charles Sprenger. Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115, 2015.

[22] Christoph Breidert, Michael Hahsler, and Thomas Reutterer. A review of methods for measuring willingness-to-pay. *Innovative marketing*, 2(4), 2006.

[23] Janet Burroway, Elizabeth Stuckey-French, and Ned Stuckey-French. *Writing fiction: A guide to narrative craft*. University of Chicago Press, 2019.

[24] Rick Busselle and Helena Bilandzic. Measuring narrative engagement. *Media psychology*, 12(4):321–347, 2009.

[25] Maryam Vaezi and Saeed Rezaei. Development of a rubric for evaluating creative writing: a multi-phase research. *New Writing*, 16(3):303–317, 2019.

[26] Roland Bénabou and Jean Tirole. Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520, 2003.

[27] Yikang Sun, Cheng-Hsiang Yang, Yanru Lyu, and Rungtai Lin. From pigments to pixels: a comparison of human and ai painting. *Applied Sciences*, 12(8):3724, 2022.

[28] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. Creative writing with an ai-powered writing assistant: Perspectives from professional writers. *arXiv preprint arXiv:2211.05030*, 2022.

[13] Under this market segmentation scenario, different fiction genres and book categories ("commercial", "upmarket", "literary") may adopt AI at different rates in accordance with their varying genre conventions and readers' imperatives, likely with commercial fiction most vulnerable to near-term automation.

[29] Anil R Doshi and Oliver P Hauser. Generative ai enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):eadn5290, 2024.

[30] Nina Beguš. Experimental narratives: A comparison of human crowdsourced storytelling and ai storytelling. *Humanities and Social Sciences Communications*, 11(1):1–22, 2024.

[31] Christoph Fuchs, Martin Schreier, and Stijn MJ Van Osselaer. The handmade effect: What's love got to do with it? *Journal of marketing*, 79(2):98–110, 2015.

A Online Appendix

A.1 Tables

Table A1: Willingness to Pay: Money and Time

	Willingness to Pay		Willingness to Work	
	(1)	(2)	(3)	(4)
1=AI Writing	-0.226	-0.226	-0.043	-0.043
	(1.300)	(1.300)	(0.127)	(0.127)
Observations	654	654	654	654
Control Mean	17.98	17.98	2.34	2.34
SD	16.5	16.5	1.64	1.64
R-Square	0.00	0.00	0.00	0.00
Control Var	N	Y	N	Y

Notes: Outcomes are willingness to pay in cents (Col. 1-2) and minutes work (Col. 3-4). Control variables include age, education, political leaning and gender. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Writing Assessment

	Index	Interest	Predict	Complex	Creative	Authen	Publish	Atmosp
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1=AI Writing	-0.280***	-0.366*	0.298*	-0.240	-0.458**	-0.791***	-0.778***	-0.645***
	(0.077)	(0.201)	(0.174)	(0.188)	(0.181)	(0.179)	(0.195)	(0.182)
Observations	654	650	652	650	651	652	651	654
Control Mean	0.14	6.18	4.51	4.85	6.21	6.88	4.58	6.76
SD	.964	2.5	2.33	2.4	2.28	2.19	2.51	2.19
R-Square	0.02	0.01	0.00	0.00	0.01	0.03	0.02	0.02
Control Var	N	N	N	N	N	N	N	N

Notes: Outcomes are a standardized index (Col. 1) of the individual outcomes and separate measures for content assessment (Col. 2-8) measured on a 0 to 10 scale. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: AI Effect on Story Content

	Advice: Use AI		Prof: Use AI		Happy Use AI	
	(1)	(2)	(3)	(4)	(5)	(6)
1=AI Writing	0.061	0.068	0.076**	0.075**	0.141**	0.145**
	(0.066)	(0.067)	(0.038)	(0.038)	(0.064)	(0.065)
Observations	654	654	654	654	654	654
Control Mean	-0.29	-0.29	0.35	0.35	-0.70	-0.70
SD	.845	.845	.476	.476	.789	.789
R-Square	0.00	0.02	0.01	0.02	0.01	0.02
Control Var	N	Y	N	Y	N	Y

Notes: The dependent variable in Col. 1-2 measures whether the professor advises the student to use AI (-1=No, 0=Unsure, 1=Yes). Col 3-4 measures whether the professor will use AI himself (1=Yes, 0=No) and whether he will be happy with the outcome (-1=No, 0=No Use, 1=Unsure, 2=Yes). Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Attention

	Time Read (Part 1)		Time Read (Part 2)		Remember Details	
	(1)	(2)	(3)	(4)	(5)	(6)
1=AI Writing	-4.125	-4.125	-1.368	-1.368	-0.061**	-0.061**
	(6.594)	(6.594)	(2.548)	(2.548)	(0.030)	(0.030)
Observations	654	654	542	542	654	654
Control Mean	162.19	162.19	54.94	54.94	0.86	0.86
SD	84.9	84.9	29.9	29.9	.351	.351
R-Square	0.00	0.00	0.00	0.00	0.01	0.01
Control Var	N	Y	N	Y	N	Y

Notes: Outcomes are the seconds people spend reading part 1 (Col. 1-2) and part 2 (Col 3-4) of the story, winsorized at the 5 percent level and a binary variable measuring if the reader remembers a detail of the story (Col. 5-6). Control variables include age, education, political leaning and gender. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Transcription Performance

	Nr. Words		Nr. Words (winsor)		Nr. Words = 0		Share Mistakes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1=AI Writing	3.326**	2.911*	1.675	1.248	-0.035	-0.022	-0.002	-0.002
	(1.545)	(1.521)	(1.254)	(1.215)	(0.033)	(0.032)	(0.006)	(0.006)
Observations	542	542	542	542	542	542	440	440
Control Mean	19.17	19.17	18.94	18.94	0.19	0.19	0.06	0.06
SD	14.8	14.8	14.2	14.2	.397	.397	.0567	.0567
R-Square	0.01	0.07	0.00	0.09	0.00	0.06	0.00	0.01
Control Var	N	Y	N	Y	N	Y	N	Y

Notes: The sample is limited to participants with a WTP at least as high as the time transcription price. Column 1 to 4 uses the number of words transcribed as the dependent variable. Column 3 and 4 winsorized the outcome at the 5 percent level. The outcome in Column 5 and 6 us a binary variable measuring if participants transcribed zero words. The outcome in Column 7 and 8 is the share of incorrectly transcribed words. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Subgroup Analaysis: WTP Money

	(1)	(2)	(3)	(4)	(5)	(6)
1=AI Writing	-1.266	0.506	0.220	0.985	-0.696	-0.573
	(1.870)	(1.902)	(1.830)	(1.940)	(2.219)	(1.827)
College Degree	1.011					
	(1.826)					
AI x College	2.103					
	(2.599)					
Female		1.346				
		(1.818)				
AI x Female		-1.569				
		(2.604)				
Age $<$ 45			-2.890			
			(1.830)			
AI x Age $<$ 45			-0.153			
			(2.602)			
Conservative				3.885*		
				(2.111)		
AI x Conservative				-4.412		
				(3.029)		
Moderate				1.295		
				(2.291)		
AI x Moderate				0.933		
				(3.250)		
Follow AI					0.096	
					(1.921)	
AI x Follow AI					0.740	
					(2.740)	
Read $>$ 5 Books						1.445
						(1.829)
AI x Read $>$ 5 Books						1.126
						(2.609)
Observations	654	654	630	654	654	649
Control Mean	17.98	17.98	17.98	17.98	17.98	17.98
SD	16.5	16.5	16.5	16.5	16.5	16.5
R-Square	0.00	0.00	0.01	0.01	0.00	0.00
P-value	0.64	0.55	0.97	0.14	0.98	0.77

Notes: Subgroup indicators interacted with the AI dummy include dummies for female, under 45, politically conservative and moderate, following AI (above median), and reading more than 4 books (median) annually. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2 Figures

Figure B1: Histogram: WTP - Money

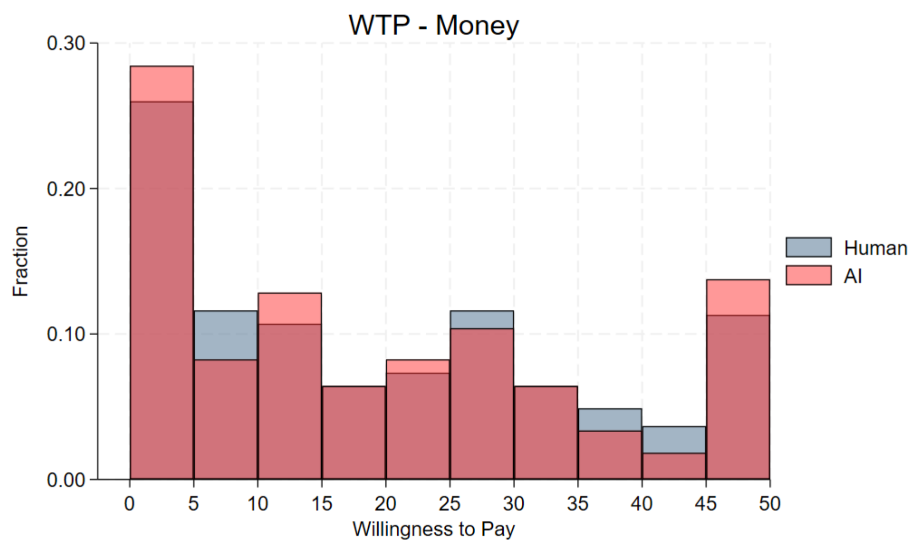
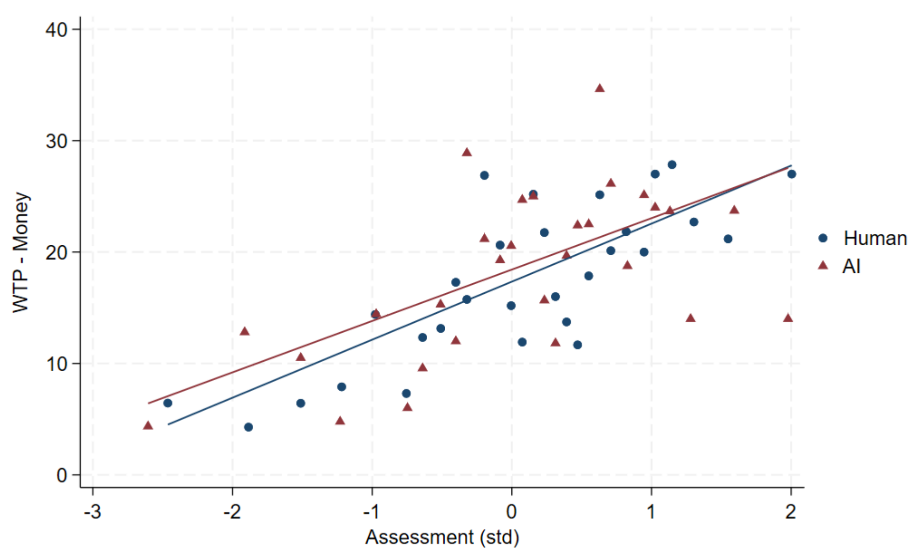


Figure B2: Correlation: Assessment and WTP (Money)



A.3 Research Design

Figure C1: Author Introduction

Jason Brown is an acclaimed short story writer. He was a Stegner Fellow and Truman Capote Fellow at Stanford University. His stories and essays have won several awards and appeared in The New Yorker, The Atlantic, Best American Short Stories, The L.A. Times, and The Guardian.

Figure C2: Experimental Variation

You will next read a short story that is representative of Jason Brown's writing.
(It is not yet published so you can't find it online.)

(a) Human Writing Information

You will next read a short story that is representative of Jason Brown's writing.
It was **written by Artificial Intelligence (ChatGPT).**
(It is not yet published so you can't find it online.)

(b) Artificial Intelligence Information

Figure C3: Short Story (Part 1)

The Quieting (Part 1)

It had been snowing for days, softening the edges of the world into a gray hush. Peter stood at his office window, watching the flakes build over the quad, the once well-worn paths now smooth and untrodden. The buildings of the university crouched against the weather, red brick turned brown in the dim light.

His students would trickle in soon, their footsteps barely making a sound on the packed snow. These days, fewer and fewer arrived. He blamed it on the weather, but in his gut, he knew better. It wasn't the snow; it was everything else.

He walked back to his desk, rubbing his hands to warm them, even though the office was plenty warm. His computer screen sat before him, blinking softly, a single email opened and waiting for a response.

Dear Professor Walsh,
Thank you for the feedback on my story. I tried something new this time—I used AI to help me generate some ideas. The program gave me three different endings for the story, and I combined them into the final one you read. I hope that's okay.
Best,
Megan

Peter stared at the screen, unsure how long he'd been sitting there, frozen by the note.

Megan's story hadn't been bad, better than her last one by miles. But it had felt hollow. Something about the way the characters spoke to each other had an unnatural clarity, too clean and resolved. He had sensed it while reading, a tickle in the back of his mind telling him something wasn't quite right.

The thought that the AI had written those words tugged at him, pulling him into an old memory. A professor from his graduate days—what was his name, Dr. Dwyer?—had once said to him, "A writer's voice is their soul, Peter. You can't fake that."

He had nodded, dutifully agreeing, but now those words rattled in his mind like loose change. Maybe, Peter thought, that was what the AI had done—faked it. But then again, wasn't that what most writing was? Trying to fake experience, to make believable the things you hadn't lived? Was Megan's use of the AI any different than his own reliance on the same tired themes, the same rural New England settings that he pulled out, year after year, like an old sweater?

And then there was his own work. The half-finished novel sat in a folder on his computer, untouched for months. Each time he tried to write, the words stayed locked inside him. Was it any wonder his students were turning to AI when even he couldn't push through his own creative blocks?

After a long pause, he began typing a response, slow at first, then with more certainty.

Figure C4: Short Story (Part 2)

The Quieting (Part 2)

After a long pause, he began typing a response, slow at first, then with more certainty.

Dear Megan,
Thank you for letting me know. It's important to experiment, and AI is an interesting tool. But writing is more than putting words together in the right order. It's about what those words mean to you. Ask yourself if the ending you chose feels true to your voice, your perspective. If it does, then it's a good ending. If not, keep searching.
Best,
Professor Walsh

He hit send, then closed his laptop, suddenly exhausted. Outside, the snow continued to fall, blanketing the campus in stillness. For a moment, he imagined what it would be like to stop wrestling with the words, to let the machine take over, to give in to the ease of it. But then he shook his head and reached for his coat.

There was something sacred in the struggle, he thought, something human. And for now, at least, that was still worth holding onto.

As he walked out into the snow, his footprints left shallow impressions behind him, soon to be covered by the falling flakes. The world around him was quiet, and though the temptation of the machine lingered in the back of his mind, the cold air cleared his thoughts, if only for a little while.

Figure C5: Instructions: WTP

The short story is almost over. There is a price you need to **pay in return for reading the end of the story**.

There are two dimensions in which we measure your willingness to pay: **money and time** willing to work.

Here is the process:
1) Willingnes to pay:
a) **Money**: You receive a bonus of $0.50 and choose your maximum willingness to pay from this bonus.
b) **Time**: You also choose the maximum time you would be willing to work transcribing text.

2) Determining the price:
a) We randomly picked one the two price dimensions: **money or time**
b) We randomly picked a **level** within that price category.

3) Purchase transaction:
a) If your willingness to pay / work is **above** that price --> you will read the end of the story and need to pay the randomly choisen price (not your stated willingness to pay/work!)
4) If your willingness to pay / work is **below** that price --> you will not get to read the end. You will keep the $0.50 and don't need to transcribe.

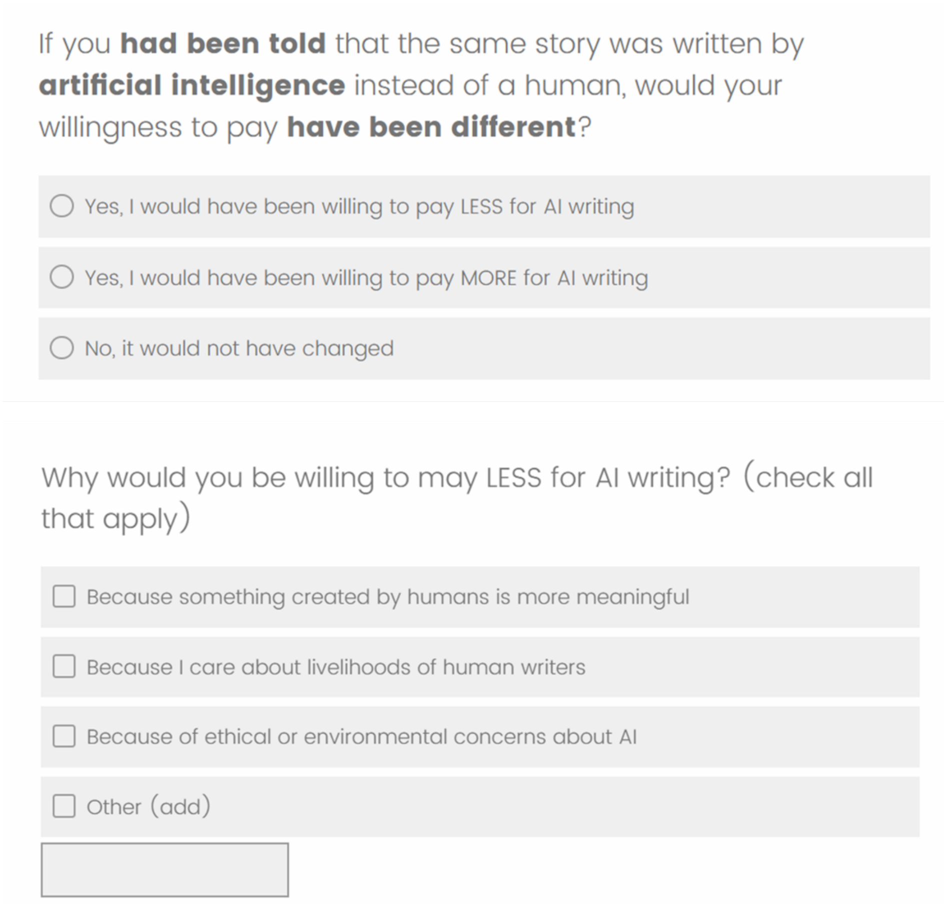
Here is an example:

1) You decide your willingness to pay is:
a) money: 25 cents
b) time: 3 minutes

2) We randomly pick time as the price category and the price level at 2 minutes

3) Your willingness to pay is above the price --> you will need to work for 2 minutes.

Figure C6: Instructions: Hypothetical WTP



If you **had been told** that the same story was written by **artificial intelligence** instead of a human, would your willingness to pay **have been different**?

- ◯ Yes, I would have been willing to pay LESS for AI writing
- ◯ Yes, I would have been willing to pay MORE for AI writing
- ◯ No, it would not have changed

Why would you be willing to may LESS for AI writing? (check all that apply)

- ☐ Because something created by humans is more meaningful
- ☐ Because I care about livelihoods of human writers
- ☐ Because of ethical or environmental concerns about AI
- ☐ Other (add)

Figure C7: Instructions: Transcription

The randomly chosen price category was "**time**".

The randomly drawn "**price**" of reading the story was **1 minute**.

Below is the text that we ask you to transcribe. The page will advance after 1 minute.

I hire 2,700 workers for a transcription job, randomly assigning the gender of their (fictitious) manager and provision of performance feedback. While praise from a manager has no effect, criticism negatively impacts workers' job satisfaction and perception of the task's importance. When female managers, rather than male, deliver this feedback, the negative effects *double* in magnitude. Having a critical female manager does not affect effort provision but it does lower workers' interest in working for the firm in the future. These findings hold for both female and male workers. I show that results are consistent with gendered expectations of feedback among workers. By contrast, I find no evidence for the role of either attention discrimination or implicit gender bias.

Transcription: