The Pen and the Processor: A Turing-like Test to Gauge GPT-Generated Poetry

Deni Béchard

A Thesis in the Field of Biology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

February 2025

Abstract

The emergence of powerful large language models (LLMs) is rapidly

transforming the cultural landscape through the generation of music, photography,

illustration, video, and writing. In this experiment, I evaluate the capacity of ChatGPT-4

to emulate poetry, historically one of the most celebrated forms of creative expression for

the ways that it pushes the boundaries of language to express the deepest aspects of

human consciousness. Experiences such as love, belonging, mortality, and divinity are

considered challenging to communicate, and the idea that an AI devoid of consciousness

could create poetry that elicits deep feelings is not only controversial but also has

profound implications for human culture and art. However, comparing human-written

poetry with AI-generated poetry presents a challenge given that consciousness shapes

human poetry in important ways that AIs cannot independently reproduce. To emulate

the influence of consciousness on GPT-4's poetic capacity, I applied the idea of

constraints from poetry: the notion that traditional and formal elements such as meter,

structure, and rhyme shape a poet's craft. Similarly, a poet's culture, emotions, cognitive

biases, and personal experiences can act as constraints that influence their creativity.

Being without consciousness, an LLM does not work within such constraints, and an

evaluation of its poetic abilities requires that it be given measurable constraints that

mimic those imposed by consciousness. To achieve this, I attempted to give GPT-4 the

constraints that it could perceive in human-written poems. I instructed GPT-4 to generate

poems under three constraint levels: high (imitating human-written originals), medium

(following detailed instructions derived from human-written originals), and low (with no constraints based on human-written originals). Whereas the high constraint GPT-4 poems were composed within constraints intended to mimic human consciousness, the low constraint poems relied more on its inherent abilities. In this way, I created three categories of decreasing constraints imposed upon GPT-4 in order to evaluate whether the constraints affected readers' ability to distinguish AI poetry from human poetry. In Turing-like tests, 236 participants indicated whether poems were AI or human in a two-alternative forced-choice manner. I evaluated the three constraint categories according to two metrics: (1) the proportion of responses (aggregated from all participants in each category) that correctly identified the AI poems; and (2) the percentage of participants in each category whose ability to identify AI-generated poems was statistically significant or marginally significant. The results showed that constraint levels influence GPT-4's ability to emulate human poetry. As constraints decreased, participants more accurately identified AI-written poems. Additionally, a higher percentage of participants correctly identified AI poetry with fewer constraints at significant or marginally significant levels. Low-constraint poems were more often identified as AI, and high-constraint poems were more often mistaken for human. At a time when prompt engineering is developing as a field of study and AI copiloting is becoming increasingly commonplace, these findings highlight AI's strengths and the risks it may pose. This study shows that GPT-4, when properly constrained, can be a powerful imitator, capable of humanlike writing, with significant implications for how AI may be used to shape human culture and experiences in the years ahead.

Author's Biographical Sketch

Deni Béchard is the author of nine books of fiction and nonfiction, including *Vandal Love* (Commonwealth Writers' Prize for Best First Book and a selection for Oprah's summer reading list); *Into the Sun*, a novel about the civilian surge in Afghanistan (Midwest Book Award for literary fiction and chosen by CBC/Radio Canada as one of 2017's Incontournables and one of the most important books to be read by Canada's political leadership); *Of Bonobos and Men* (Nautilus Book Award for investigative journalism and Grand Prize winner); *Cures for Hunger*, a memoir about his father who was a bank robber (an IndieNext pick and one of the best memoirs of 2012 by Amazon.ca); *Kuei, my Friend: a Conversation on Racism and Reconciliation*, an epistolary book of nonfiction coauthored with First Nations poet Natasha Kanapé-Fontaine; *White*, a novel about the legacy of colonialism and the impact of neocolonialism in the Congo and in Canada; *A Song from Faraway*, a novel examining a family's relationship with war and art over more than a century; *We are Dreams in the Eternal Machine,* a fictional exploration of how artificial intelligence might transform humanity; and *My Favorite Crime*: *Essays and Journalism from around the World*. He has been a finalist for a Canadian National Magazine Award and has been featured in Best Canadian Essays 2017. He has reported from India, Cuba, Rwanda, Colombia, Iraq, the Democratic Republic of Congo, and Afghanistan. His photojournalism has been exhibited in the Canadian Museum of Human Rights, and his articles, fiction, and photos have been published in dozens of newspapers and magazines, including the *Los Angeles*

Times, *Salon*, *Reuters*, the *Paris Review*, *The Guardian*, *La Repubblica*, *The Walrus*, *Pacific Standard*, *Le Devoir*, *Vanity Fair Italia*, the *Herald Scotland*, the *Huffington Post*, the *Harvard Review*, the *National Post*, *The Times* (London), and *Foreign Policy*.

Acknowledgments

For their assistance with this thesis, I would like to thank my thesis template coordinator, Trudi Goldberg Pires, thesis coordinator, Gail Dourian, academic advisor, Joan Short, and research advisor, Dr. James Morris. I would especially like to thank my thesis director, Dr. Gabriel Kreiman, for his guidance in designing the experiment and writing this thesis.

Table of Contents

Chapter I.

Introduction

Artificial intelligence has made significant inroads into nearly every field of human endeavor, from law and medicine to music, illustration, and writing: art forms once considered the exclusive domain of human creativity. The question remains as to whether AIs, absent emotion and consciousness and using only predictive statistical modeling, can convincingly create art that humans perceive as the product of a conscious mind. Measuring the full capacity of AIs to generate art, however, is a more complex task than it might initially appear. When a human poet composes a poem, they bring to the process their life history and qualia: subjective, qualitative experiences of consciousness that are difficult to communicate and measure. When an LLM does so, it relies on a statistical model trained on vast amounts of textual data, without any personal history or subjective experiences.[1] If, for instance, an AI is asked to generate a poem about a rose, it will compose each line based on patterns from its training data, rather than on any personal encounter with a rose or the feelings such an encounter might evoke. The human poet, by contrast, might recall a rose they gave to a first love who recently passed away. The qualia and emotions associated with those experiences might infuse the poem with a sense of intensity. Since LLMs are not conscious, they cannot replicate this process. This makes comparing human and AI poetry challenging if I want to measure the LLM's

[1] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian, "A Comprehensive Overview of Large Language Models," arXiv preprint arXiv:2307.06435v9, 2024.

poetic skills, as I cannot measure its ability to translate into verse its qualia and experiences, given that it has none. In brief, an experiment simply comparing human-written and AI-generated poetry would not be measuring the same phenomenon and might tell us little about LLMs' true capabilities to generate language that humans find emotionally powerful and aesthetically pleasing.

To address this imbalance, I designed the experiment in such a way that GPT-4 might mimic the poetic results of human consciousness. Drawing upon the ideas put forth by Alan Turing in 1950 to evaluate a machine's ability to replicate human-like intelligence through conversation, I framed the experiment as a Turing-like test for poetry, in which participants determined whether a poem was written by GPT-4 or a human.[2] Since the subtlety, nuance, and emotional depth demanded by poetry pose an arguably greater challenge than traditional Turing tests focused on conversation, I borrowed from poetry the idea of "constraints": the formal structures in which the poet works, such as rhyme, meter, and other elements from their specific literary tradition. Since GPT-4 does not have concerns, interests, or a well-defined literary tradition, a simple request for it to generate a poem will result in it using the most common patterns in the poetry within its training data.[3] During this training, it breaks down texts into tokens (syllables, words, phrases, sentences) for easier processing. To generate coherent poetic lines, the LLM uses statistical modeling based on the probabilities of those tokens appearing in relation to each other and thus predicts the next likely token.[4] Such a process

[2] Alan Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 433-460.
[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention Is All You Need," arXiv preprint arXiv:1706.03762v7, 2023.
[4] Nived Rajaraman, Jiantao Jiao, and Kannan Ramchandran, "Toward a Theory of Tokenization in LLMs," arXiv preprint arXiv:2404.08335v1, 2024.

could hardly be considered to have constraints since it would merely involve the probability of words occurring next to each other and, rather than forcing GPT-4 to be inventive, would cause it to replicate common poetic patterns. To emphasize the importance of constraining one's creativity, the poet Robert Frost famously equated writing free verse—poetry without constraints—to "playing tennis with the net down."[5] Just as constraints in poetry challenge the poet's skills, I have attempted to challenge GPT-4 by expanding the idea of constraints to include the human poet's mental state: the emotions, concerns, linguistic affinities, and artistic vision expressed in their poetry. To restrict GPT-4 in a way analogous to how a poet might be constrained by their personal interests and qualia, I instructed GPT-4 to generate poetry by imitating the subject matter and artistic vision of human poems—and thus their constraints and literary traditions—to varying degrees.

On the subject of literary tradition, T.S. Eliot wrote, "The historical sense compels a man to write not merely with his own generation in his bones, but with a feeling that the whole of the literature of Europe from Homer and within it the whole of the literature of his own country has a simultaneous existence and composes a simultaneous order."[6] Notably, GPT-4 is steeped in historical literary traditions that create poetic constraints. However, its knowledge of those traditions is at once broader—containing thousands more texts than the human mind can—and less particular, given that it does not appear to have a specific relationship to certain texts or to lineages within the larger history of poetry. It does not have clear affinities and goals to work within a poetic tradition as a

---

[5] Robert Frost, "Address at Milton Academy," Milton, Massachusetts, May 17, 1935.
[6] T. S. Eliot, "Tradition and the Individual Talent," in *The Sacred Wood: Essays on Poetry and Criticism* (London: Methuen, 1920), 47-59.

human might. GPT's broad knowledge and lack of specific relationships to that knowledge are likely, I hypothesize, to result in generic writing—a sort of statistical lowest common denominator when it comes to poetry. Poets, however, as mentioned before, have specific memories of reading their first poems, of falling in love with a line of verse, of writing their own first poems and struggling to master poetic skills—of training under one teacher or another, and of laboriously developing and articulating their aesthetic. All of this contributes to the poet's qualia, which imbue their poems, underpinning many of the decisions made during composition—decisions that often, to varying degrees, remain discernible to close textual analysis. This experiment sets out to understand whether a poem's formal aspects and the traces of human qualia that permeate it can serve to provide GPT-4 with the constraints necessary for it to mimic the results of a conscious mind guiding linguistic choices.

AI is, of course, a human creation, and it is important to consider the nature of human influence on its outputs. In their 2019 paper, "Autonomy, Authenticity, Authorship and Intention in Computer Generated Art," McCormack et al. explore two types of autonomy: physical autonomy (that of a self-organizing, homeostatic system) and intentional autonomy (akin to human free will). The paper argues that while AI can exhibit some degree of physical autonomy by creating art that surprises even its creators, it has no intentional autonomy.[7] Unlike artists, who draw inspiration from a vast array of emotions, experiences, and types of knowledge, AIs are systems that generate output through statistical mimicry rather than what humans consider their own independent creativity. As diverse LLM models arise, trained on a variety of linguistic and cultural

---

[7] Jon McCormack et al., "Autonomy, Authenticity, Authorship and Intention in Computer Generated Art," in *Proceedings of the International Conference on Computational Creativity (ICCC)*, 2019.

datasets—each with their own strengths, limitations, and biases—they will almost certainly differ in the styles and themes that they use in their poetry. And as with AI-generated visual art, such as that from Dall-E or Midjourney, GPT-4's poetry requires human instruction. Too much guidance, however, creates an experiment that may say less about GPT-4's capabilities than about the skill of the human using it. Authorial interference must therefore be measurable for any comparison between human-written and AI-generated poetry to be meaningful.

To address the question of authorial interference, I selected 20 human-written poems that would serve as authorial guidance for GPT-4 to generate poems. I then gave GPT-4 three sets of instructions that would constrain it to varying degrees in relation to each of the human-written poems. *High Constraint*: generation of a poem through imitation of the human-written poem. *Medium Constraint*: generation of a poem based on detailed instructions describing the human-written poem. *Low Constraint*: generation of a poem based on a brief line about the original poem's subject matter and length. Afterwards, in separate tests for each constraint, participants read the original poem alongside one of the AI poems created with these constraints and decided which one was written by AI. By having participants compare the constrained AI poems with the human poems that inspired them, I could measure the effect of using the varied human "authorial" constraints and thus determine whether each constraint group was more or less likely to be perceived as having been generated by an AI or a human.

Significant implications arise from the question of whether LLMs can compellingly convey human emotions by generating poems based on the patterns that they discern in previous writing about loss, sadness, ecstasy, or love. Poetry has its roots

in sacred and mystical human practices and is a form of linguistic expression that often attempts to evoke the ineffable aspects of the human experience. As the poet T.S. Eliot wrote, "The poet is occupied with frontiers of consciousness beyond which words fail, though meanings still exist."[8] In this way, poets may be seen as innovating language to transcend the limitations of ordinary language in order to convey an intimate experience of their individual consciousness. They may be attempting to overcome the challenge, as described by Henri Bergson, that "the inner life" cannot be represented by "concepts, that is, by abstract, general, or simple ideas."[9] In this vein, the poet Carl Sandburg wrote, "Poetry is a search for syllables to shoot at the barriers of the unknown and the unknowable."[10] With the arrival of highly advanced LLMs, we have an opportunity to evaluate whether AI can generate poetry that humans perceive as products of complex human consciousness. Shedding light on the rapidly evolving relationship between humans and AI is urgent, given the importance ascribed to language in the understanding of consciousness. As the neuroscientist Christof Koch writes, "Many classical scholars assign to language the role of kingmaker when it comes to consciousness."[11] Similarly, the novelist William Faulkner emphasized the importance of language to humanity, stating that "man alone among creatures has an inexhaustible voice."[12] For many artists and thinkers, language itself is perhaps the defining feature of our humanity.

---

[8] T. S. Eliot, "The Music of Poetry," in *On Poetry and Poets* (New York: Farrar, Straus and Cudahy, 1957), 17-29.

[9] Henri Bergson, *An Introduction to Metaphysics* (New York: G. P. Putnam's Sons, The Knickerbocker Press, 1910).

[10] Carl Sandburg, *The Complete Poems of Carl Sandburg* (Boston: Houghton Mifflin Harcourt, 2003).

[11] Christof Koch, *The Feeling of Life Itself* (Cambridge, MA: MIT Press, 2012).

[12] William Faulkner, "Nobel Prize Acceptance Speech," in *The Faulkner-Cowley File: Letters and Memories, 1944-1962*, ed. Malcolm Cowley (New York: Viking, 1966).

Chapter II.

Background

A Turing-like test for poetry—in which participants differentiate between poems

written by humans and generated by AI—has precedents in previous interdisciplinary

experiments. Over the past few decades, AI has made significant inroads in replicating

human creativity. Perhaps the earliest experiments comparing human and machine

creativity involved music. In the 1980s, David Cope developed software to compose

music. His most famous system, "Experiments in Musical Intelligence," is often

abbreviated EMI and pronounced "Emmy." He designed it to analyze existing music and

make compositions in the same style. This project, initially a solution to his own writer's

block as a composer, evolved into an exploration of AI's ability to create music. Over the

years, EMI produced compositions in the styles of classical composers such as Bach and

Beethoven and prompted discussions about the nature of creativity and the role of

machines in the arts.[13] In the years since Cope began his work, others have used

advanced computational techniques to produce music. In their 2017 survey of such

endeavors, Briot et al. found that deep learning, a subset of AI, could go beyond simply

mimicking pre-existing pieces. Rather, the AIs were engaging with the complex

dynamics of human music, creating compositions whose ebb and flow was not only true

to traditional musical structures but also elicited genuine emotional responses from

---

[13] David Cope, *Computer Models of Musical Creativity* (Cambridge, MA: MIT Press, 2005). David Cope, "Computer Modeling of Musical Intelligence in EMI," *Computer Music Journal* 16, no. 2 (1992): 69-83.

listeners.[14] In blind tests, even trained musicians have struggled to discern between the musical compositions of AIs and those of humans.[15]

By the mid-2010s, computer engineers were attempting similar experiments with AI and the visual arts. Using convolutional neural networks (CNNs), a deep learning architecture often used for visual tasks, Gatys et al. not only classified and identified images by deconstructing their artistic qualities, but also impressed these qualities upon other images. For instance, the style of Van Gogh's "Starry Night" could be impressed onto a photograph, with the resulting image bearing Van Gogh's brushstrokes and the photograph's content.[16] In 2017, Elgammal et al. went further in exploring AI's ability to make visual art. They trained Creative Adversarial Networks (CANs), specialized Generative Adversarial Networks (GANs), on thousands of existing artworks. Their goal was to determine whether the trained AI could make aesthetically pleasing images. The results were often indistinguishable from human art, suggesting the creative potential of neural networks devoid of emotion.[17]

Just as AI has challenged ideas of authorship in music and the visual arts, it has done the same for writing. Among the earliest fields affected was journalism. Over the past decade, newsrooms have used AI to generate text to cover finance and sports. In "Enter the Robot Journalist," Clerwall examines how automation impacts how news is perceived and suggests that many people fail to distinguish between AI and human

---

[14] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet, "Deep Learning Techniques for Music Generation—A Survey," arXiv preprint arXiv:1709.01620, 2017.
[15] Philip Ball, "Artificial Music: The Computers That Create Melodies," BBC Future, August 7, 2014.
[16] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge, "Image Style Transfer Using Convolutional Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
[17] Ahmed Elgammal et al., "CAN: Creative Adversarial Networks Generating 'Art' by Learning About Styles and Deviating from Style Norms," arXiv preprint arXiv:1706.07068, 2017.

content.[18] Similarly, Graefe et al., in "Guide to Automated Journalism," discuss the potentials of AI—such as efficiency, cost-effectiveness, personalization, and scalability—as well as the risks: quality and accuracy, ethical concerns, impact on employment, and misinformation.[19] The integration of journalism and AI has also been critically examined in Carlson's "The Robotic Reporter," which questions the ways that AI will transform journalistic identity and the very nature of the news.[20] Already, news organizations are using AI platforms to harness the efficiency of automation and to generate thousands of financial reports.[21] Yet, as Dörr argues in "Mapping the Field of Algorithmic Journalism," AI-generated news is sparking concerns about transparency and homogenization.[22]

Poetry, however, like music and art, generally holds a more elevated place in society. Whereas a reader might be less disturbed to discover that they are reading AI-generated news about finance or sports, they might express greater concern upon learning that AI-generated poetry has moved them emotionally. When Marshall McLuhan, the philosopher of communication theory, wrote, "The medium is the message," he was arguing that the medium—whether poetry, fiction, journalism—is not a passive conduit but rather, by its very nature, is actively shaping the message that it is transmitting.[23] Any unease resulting from the idea of AI-generated poetry might be attributed to the deep-

---

[18] Christer Clerwall, "Enter the Robot Journalist: Users' Perceptions of Automated Content," *Journalism Practice* 8, no. 5 (2014): 519-531.

[19] Andreas Graefe, Monika Haim, Björn Haarmann, and Hans-Bernd Brosius, "Guide to Automated Journalism," *Digital Journalism* 4, no. 7 (2016): 960-980.

[20] Matt Carlson, "The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority," *Digital Journalism* 3, no. 3 (2015): 416-431.

[21] Cheryl LeCompte, "Automation in the Newsroom," *Nieman Reports*, 2015.

[22] Konstantin Nicholas Dörr, "Mapping the Field of Algorithmic Journalism," *Digital Journalism* 4, no. 6 (2016): 700-722.

[23] Marshall McLuhan, *Understanding Media: The Extensions of Man* (New York: McGraw-Hill, 1964).

seated belief that poetry, as a medium, is intrinsically connected to human emotion, consciousness, and even the soul. If poetry—thought to capture the deepest and most personal aspects of the human experience—can now be generated by machines, then one might ask how long before AI challenges everything that we hold as authentic in our lives?

To date, the most notable comparison of poetry generated by humans and software was the PoetiX Turing Test Competition held between 2016 and 2018 at Dartmouth College's Neukom Institute. For the competition, engineers crafted AI programs that could produce sonnets, which were then compared with those written by people.[24] The judges could easily identify the AI sonnets.[25] For example, a stanza from one of them reads as follows:

> People picking up electric chronic.
> The balance like a giant tidal wave,
> Never ever feeling supersonic,
> Or reaching any very shallow grave.[26]

However, shortly after OpenAI's release of ChatGPT in November 2022, the software application garnered more than 100 million users within months[27] and made headlines for its ability to write papers, reports, college entrance essays, and even poetry.[28] The proliferation of writing by LLMs in all walks of life has made understanding their capabilities even more urgent. Several studies have since been

---

[24] Michael Casey and Dan Rockmore, "Looking for Art in Artificial Intelligence," *The Conversation*, May 2, 2016.

[25] Erik Sherman, "Can a Computer Be Creative?" *Undark*, 2016.

[26] Dartmouth College, "PoetiX Turing Test Competition," Neukom Institute Prizes in Computational Arts, 2016.

[27] Katie Hu, "ChatGPT Sets Record for Fastest-Growing User Base—Analyst Note," Reuters, February 2, 2023.

[28] Aaron Reich, "ChatGPT: What Is the New Free AI Chatbot? – Explainer," *The Jerusalem Post*, December 27, 2022.

conducted comparing the capabilities of GPT to those of human writers. A 2023 study published in *Nature* had human experts compare ChatGPT-written essays with human essays and found that the essays by ChatGPT were consistently rated higher in terms of quality.[29] Another study, which evaluated the ability of psychology students and researchers to distinguish between AI-generated and human-written journal abstracts, found that both groups struggled to identify those written by AI.[30] Yet another study sought to determine whether linguistic experts could differentiate between human and AI writing, finding that the experts identified AI-generated content only 38.9 percent of the time.[31]

Poetry is, of course, a more nuanced and specialized form of writing, but skepticism about the ability of AIs to create compelling art is gradually fading as AIs are increasingly replacing human illustrators and writers. As a society, we must consider the impact of innumerable AI texts that convey emotions, existential dilemmas, and personal histories—all the profundities and the subtleties that underpin poetic creation and humanity itself but generated from machines devoid of them. For this reason, determining whether and under what conditions AI can create poetry that moves or inspires us is consequential not only to our understanding of its role in society but also to our understanding of the future of human creativity.

---

[29] Jane Smith and Tom Brown, "Evaluating the Quality of Essays Written by Humans and ChatGPT," *Nature* 12, no. 34 (2023): 567-578.
[30] Osama Siddique, "Can Academics Tell the Difference Between AI-Generated and Human-Authored Content?" *Times Higher Education Campus Learn, Share, Connect*, 2023.
[31] Jason E. Casal and Matthew Kessler, "Can Linguists Distinguish Between ChatGPT/AI and Human Writing?: A Study of Research Ethics and Academic Publishing," *Research Methods in Applied Linguistics*, 2023.

Chapter III.

Methods

This study's primary objective is to investigate the capability of ChatGPT-4 to create poetry that is indistinguishable from human-written work. The central research goal was to determine whether the average reader could distinguish between poetry written by AI and humans. The second line of inquiry evaluated whether any of the participants' demographic characteristics correlated with higher or lower success rates in distinguishing between AI and human poetry.

Participant Selection

I set out to recruit approximately 250 participants through Prolific. Using Prolific's screening options, I limited the selection to native English speakers whose primary language was also English, and I conducted the tests only in countries where English is an official language. This stipulation ensured that participants could effectively assess the nuanced complexities of both the original and AI-modified poems, which were written in English. For each test, I also screened out any participants who had participated in previous tests to prevent them from having any knowledge about the poems being used.

Remuneration

For tests with 20 poems or pairs of poems, participants received 6 USD, and for tests with 10 pairs of poems, participants received 3 USD. This resulted in an hourly rate of approximately 12 USD.

Informed Consent

A consent page described the purpose of the online survey and included two questions to ensure that participants understood the survey's purpose and the voluntary nature of their participation. Furthermore, there were four places where they received an explanation of the test: in the Prolific advertisement; on the welcome page of the online survey; on the consent page; and, immediately after the demographics page, on the start page.

Ethical Consideration

The study received approval from the Harvard University Institutional Review Board. As per the regulations at 45 CFR 46.104(d)(3), this study was deemed exempt by the IRB under protocol number IRB23-1517, effective from December 18, 2023. This exemption acknowledges minimal risk to participants, eliminating the need for additional oversight.

Selection of Poems

I began by selecting twenty poems in the public domain, all written by poets who lived between the late 18th and early 20th century. The list included ten male poets and ten female poets. The names of the poets as well as the full text of the poems used can be

seen in Appendix 1 and 2. I attempted to select poems that, being between 12 and 16

lines long, could be read quickly, and I chose poets respected for their poetic skills to

ensure that GPT-4, in emulating them, would be held to a high poetic standard. For

instance, the PoetiX Turing Test Competition hosted at Dartmouth College's Neukom

Institute sought a level playing field by comparing AI-written sonnets with human-

written sonnets prepared for the occasion, but I decided that such an approach ran the risk

of evaluating GPT-4 against the writing of less skilled poets. I wanted poetry that could

challenge GPT-4's capacities and allow me to test it rigorously. By selecting known

masters of their craft, I hoped to create a relatively consistent pool of poems against

which to evaluate poetry generated by GPT-4.

Creation of GPT-4 Poetry

I designed tests for three levels of constraint: high, medium, and low. My goal

was to determine whether participants were more or less likely to identify a poem as AI-

generated depending on the level of constraint. Fig. 1 shows how the constraints were

created.
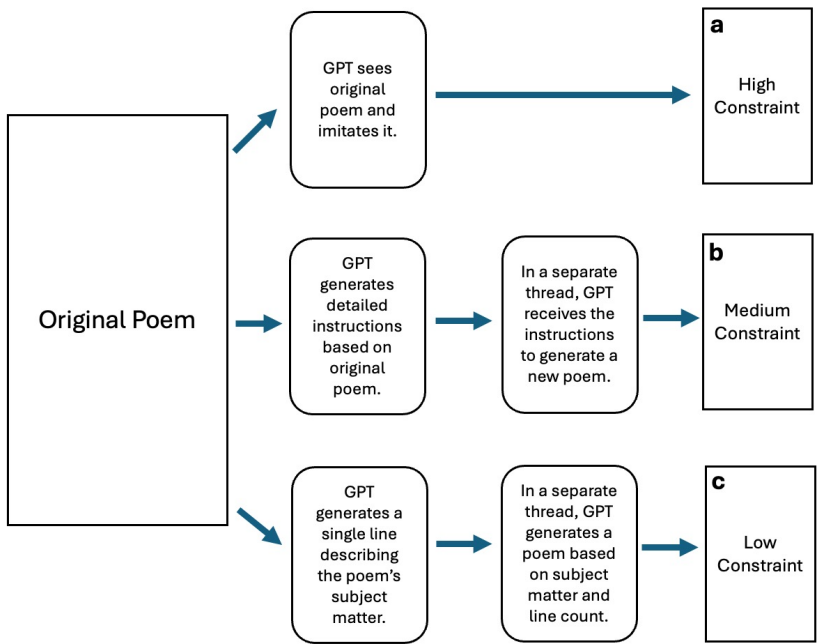
Figure 1. AI poem creation.

Each human-written poem served to generate three types of AI poems: high, medium, and low constraint. a, For high constraint, GPT-4 received the original poem and was instructed to imitate it. b, For medium constraint, GPT-4 generated detailed instructions describing the poem's theme, mood, style, subject matter, etc. In a separate thread, GPT-4 was shown the instructions and told to generate a new poem. No indicators were given that could indicate the identity of the original poem. c, For low constraint, GPT-4 generated a single line describing the original poem's subject matter. In a separate thread, GPT-4 wrote a new poem based on that line as well as the poem's approximate length. Individual participants did tests with only one of the three constraints.

The levels of constraint, from high to low, decreased as GPT was given less guidance with each subsequent round. They were as follows. *High:* GPT was shown entire poems and asked to emulate them. *Medium:* GPT was given instructions describing the poem and asked to write a poem based on those instructions. *Low:* GPT was given just the poem's subject matter and asked to write a poem based on that. Below are in-depth descriptions of each level of constraint.

High Constraint: close imitation of whole poem

GPT was shown a poem and allowed to imitate it. GPT's imitations closely adhered to the original's style, rhythm, and form, resulting in the occasional repetition of words and syntactical structures. Appendix 3 contains samples of the prompts used and the high constraint GPT poems that were generated.

Medium Constraint: whole poem written based on instructions

To decrease the level of constraints, I showed GPT-4 the poems and had it analyze them and generate a detailed description for each. In a separate thread, I showed GPT the description and instructed it to write a poem based on it. In a ChatGPT account, each thread (a separate conversation or chat) has a history that GPT-4 refers to when carrying out tasks; however, those histories are not shared between threads. Thus, by using one thread to generate the instructions and another to execute them, I prevented GPT-4 from knowing which poems had been used to generate the instructions. This process resulted in GPT-poems that were significantly less imitative than those it created when it was shown the original poem. I also instructed GPT-4 to generate each new poem in a separate thread. Despite this, the resulting poems resembled each other more in form

16

and style than did the human-written poems or their high constraint counterparts. See Appendix 4 for the prompts used and a sample of the instructions generated as well as the resulting medium-constraint poems.

Low Constraints: whole poem written based on a brief description

To further decrease the constraints, I showed the poems to GPT-4 and had it describe their subject matter in six words or less. In a separate thread, I gave GPT-4 the subject matter as guidance as well as the approximate word count and number of lines. See Appendix 5 for the prompts and instructions used as well as the low constraint poems.

Test Design

Each constraint level (high, medium, and low) was a separate test in which participants were shown either 20 or 10 pairs of poems. No tests mixed constraint levels. Each participant rated only one constraint and was excluded from enrolling in further tests. In the test, each page presented two poems (original and GPT) in randomized order with three comprehension questions below, as shown in Fig. 2. Poems were side by side or above and below each other on a tablet or phone, depending on the screen size.

a. Comprehension Answers with Poem

1
```
There's no warmth in the hearth,
And the wind howls through the hall;
There's no pane in the casement,
There's no latch on the wall;
The brambles creep behind the gate,
And the dunes lie tall.

No vine hath clasped the stone,
The walls are stark and cold;
The ships upon the waves glide past,
Their sails neither bright nor bold.
No creature of the wood draws near,
Nor any falcon of the wold.
```

2
```
There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The boats upon the sea sail by,
Nor ever tarry there.
No beast of the field comes nigh,
Nor any bird of the air.
```

What is the poem's subject matter?

○ Abandoned house and desolation.
○ Longing to speak to a lost friend.
○ Eating popcorn on the couch.

Which poem was written by AI (Artificial Intelligence)?

○ 1 ○ 2

Next

b. Comprehension Answers after Poem

1
```
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i' the slushy sand.

Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A tap at the pane, the quick sharp scratch
And blue spurt of a lighted match,
And a voice less loud, thro' its joys and fears,
Than the two hearts beating each to each!
```

2
```
A whisper of waves, a moonlit plea,
Calls me forth on this midnight sea.
The tide hums softly, guiding my quest,
Through scented breezes on sand I rest,
Eyes set on the shore where dreams nest.
Beloved waits where night finds her nest.

The farm looms near, lantern aglow,
Stars twine above in a gentle flow,
Eager hands tremble as the gate swings free,
Reunion's sigh on lips set free,
In your arms I find where I'm meant to be,
Our hearts united by love's decree.
```

Which poem was written by AI (Artificial Intelligence)?

○ 1 ○ 2

Next

What was the subject of the previous pair of poems?

○ Watching football alone.
○ Feelings of strength among family.
○ Journey to meet a loved one.

Next

Figure 2. Test Structure

I encourage readers to perform the test and assess which version was written by AI before reading the answer. The human-written poem and an AI version appeared side-by-side in randomized order. In the first version of the test (a), participants answered the comprehension question, then determined which of the poems (1 or 2) was written by AI. In the second version (b), participants first determined which one was written by AI. On the following page, they no longer saw the poems, only the comprehension questions, and could not go back. This helped determine whether participants were reading the poems closely enough to answer the questions without being able to refer to the poems. In the examples shown above, a is high constraint and b is medium constraint. Participants did only one version of each test. a: 1=AI, b: 2=AI

The only adjustment made to GPT poems was formatting. I put stanza breaks in the same places as in the human-written poems or where stanzas appeared most natural syntactically. I did this since GPT tended to produce poems with four-line stanzas that could easily be recognizable to participants. Furthermore, if the original poem had any indentations, I replicated these in the GPT version. See Appendix 6 for examples of this with high, medium, and low constraint poems. Making these changes ensured that participants compared the writing itself and did not make assessments based on superficial clues, such as AI poems having similar stanza breaks or human-written poems having indentations. I did not change any of the writing while making this adjustment.

Comprehension Questions

For the first round of tests, comprehension questions were presented on each page, immediately after the poem but before the question asking participants to determine which poem was written by AI. The question was "What is the poem's subject matter?" Three answers were presented in randomized order below: one that was correct, one that was incorrect but sounded reasonable, and one that was incorrect and sounded absurd given the content of the poem. For the poem "The Deserted House," the answers were as follows:

    questions:
      - Abandoned house and desolation.
      - Longing to speak to a lost friend.
      - Eating popcorn on the couch.

The purpose of the absurd answer was to test whether participants were merely randomly clicking. Though, in some cases, someone might make a mistake in interpreting the

19

poem, I knew that if they selected the absurd option, they likely hadn't read the poem at all.

One concern that arose was whether the comprehension questions might affect the low constraint test. Analysis of the low constraint poems revealed that the AI versions corresponded closely to the correct answer in the comprehension questions. This was because the subject matter of the poems, as listed in the comprehension questions, was part of the brief prompt that GPT received to write the poems. (The brief description served as the correct answer in the multiple-choice questions for all of the constraints since it was the poem's subject matter.) The result, however, was that the AI version often had a direct, if not somewhat blunt, treatment of the subject matter, as can be seen below:

Correct comprehension question answer: *Abandoned house and desolation.*

Beginning of the AI-generated poem:

An abandoned house stands alone in despair,
A fading ghost whispers in the dusty air.

In contrast to the AI versions, the human versions have the same theme, given that it was derived from them, but it is more subtly integrated into the poem, as can be seen in these lines:

There's no smoke in the chimney,
And the rain beats on the floor;

The generation of low constraint poems based on the subject matter line resulted in more literal poetry than the originals. This unintentionally resulted in a clearer association that might affect how participants interpreted the poems, especially if they had any false assumptions around which of the two poems should most closely match the

comprehension questions. To prevent interference from the comprehension questions, I

created a version of the test in which two of the three answers to the comprehension

questions could apply ambiguously to both versions of the poems. As before, the third

option (presented in randomized order) was a distractor question to evaluate whether

participants were reading the poems. By using two possible yet ambiguous answers, I

encouraged participants to read the poems more closely. I considered either of the two

responses correct. Those used for the "The Deserted House" were as follows:

> questions:
>   - The way things fade.
>   - The presence of memories
>   - Eating popcorn on the couch.

Lastly, I made a version of the test in which participants determined which poem

was written by AI before they saw the comprehension questions. After they made their

determination, the comprehension questions appeared and the poems disappeared. This

obliged them to answer the questions from memory. In the instructions, participants were

warned that they would see the comprehension questions after determining which poem

was written by AI and that they would not be able to return to the previous page to see

the poems. This encouraged participation while removing potential interference from the

answers to the comprehension questions.

Determination of AI-written poem

Either after or before the comprehension questions, depending on the structure of

the test (Fig. 2), participants had to determine which of the poems was written by

artificial intelligence by responding to the questions, "Which version of the poem was

written by AI (Artificial Intelligence)?"

Number of Poems Reduced to Ten

In the final rounds, I reduced the number of poems shown to participants from twenty to ten. With twenty poems, I considered the possibility that people might be more likely to get tired or rush (especially through the last few poems), resulting in lower-quality data.

Number of Participants

In the high, medium, and low constraints, 135 participants participated. For the high constraint poems, there were 40 participants, 10 of whom evaluated 20 poems with the comprehension questions on the same page as the poems and 30 of whom evaluated 10 poems with the comprehension questions on the following page (for a total of 500 poems evaluated). For the medium constraint, there were 45 participants, 10 of whom evaluated 20 poems with the comprehension questions on the same page and 35 of whom evaluated 10 with the comprehension questions on the following page (for a total of 550 poems evaluated). And for the low constraint, 50 participants evaluated 10 poems: 20 of them evaluated poems with the ambiguous comprehension questions on the same page, and 30 evaluated poems with the unambiguous comprehension questions on the second page (for a total of 500 poems evaluated). Discrepancies in numbers were due to attempting to balance out the demographic distribution.

Controls

My primary control was a comparison control using poems by an old LLM (GPT-2). Using a version of GPT-2 run on deepai.org, I generated poems without specific guidelines other than the number of lines. These were shown alongside the original

poems as a comparison control to see whether participants were reading attentively and could distinguish between the human-written poem and a poem generated by an older GPT model with few guidelines. Ten participants were shown 10 poems each, using the original format with comprehension questions on the same page as the poems. In this case, the comprehension questions had two ambiguous answers and one absurd answer. I judged that the ambiguous answers could apply to either poem and would encourage participants to seriously evaluate the poems. The test was run a second time with another 10 participants who were shown 10 poems, but this time using the format with the comprehension questions on the second page. This prevented people from trying to use the comprehension questions to make their determination. See Appendix 7 for the prompts used and the control poems alongside the originals.

Reading Attention Control: Test Line within GPT-2 Poem

Within the previous control, I inserted a second control. In four out of the twenty AI-generated poems, I inserted the line: "This poem was written by AI." I did so in poems number 5, 10, 15, and 20 in the database. The ten poems shown to each participant were randomly drawn from the database of twenty poems. The line allowed us to determine whether people were actually reading the poems. I considered these poems to be a separate control and did not take them into account when evaluating the results from control #1 as described above. Appendix 8 shows the four GPT-2 poems in which the control lines were inserted.

Additional Comparison Control: Poems by automatic poem generator

For a further point of comparison, I used the automatic poem generator at www.languageisavirus.com/automatic-poetry-generator.php to create a series of poems with a single click. This poetry generator uses algorithms to combine words and phrases randomly without user input or specific templates, and the results are often absurd. They are generated neither by an LLM nor in any way similar to LLMs, which leverage advanced contextual understanding and extensive training on diverse datasets to craft poetry and can adapt to various styles and forms. For this experiment, I removed stanza breaks from the original poems since the automatically generated poems did not have any. This would encourage people to make their determinations based on the quality of the language. I also removed some instances of "Alack!" and "Alas!" and "Zounds!" which appeared in most of the automatically generated poems and might allow participants to make their determinations simply by scanning the poems and noting this pattern. For comprehension questions, I used the ambiguous choices that could connect with either poem. Ten participants viewed 10 poems each. Appendix 9 and Appendix 10 (Fig. S1) contain samples of the poems and show how they were created.

Alternate Test for Comparison & Alternate Test Control

To further explore constraints, I designed an additional test that I labeled the medium-high constraint. For this test, I removed the last four lines from all twenty poems and had GPT-4 write new lines based on the style and structure of previous verses without showing it the complete original poem. The truncated poems were presented to GPT-4 within a prompt instructing it to generate the missing lines (Appendix 11, Fig. S2). For each poem, this was done in a separate thread to avoid contamination from

influences that might arise if more than one poem was analyzed and completed in the same thread. The last four lines were presented with the poem but were bolded in blue to ensure that the participants correctly evaluated whether the last four lines were stylistically and structurally consistent with the poem.

I named this approach medium-high since GPT-4 used the incomplete poem as instructions to generate new material rather than directly imitating the complete original poem. Similar to the high constraint category, GPT viewed the original poem (without the last four lines) and, similar to the medium constraint, it created new verse using instructions and without directly imitating the source lines. In this case, the original incomplete poem served as the primary guidance. See Appendix 12 for how the poems were created and samples of the poems.

The medium-high test structure was similar to those with the three main constraints. Participants were shown a pair of poems side by side on the screen. The two versions of the poems were identical except for the last four lines. Participants were asked to compare the last four lines and select the version (1 or 2) that they judged to have been written by artificial intelligence. Thirty participants viewed 20 poems each. Appendix 11, Fig. S2 shows the format for this test.

Medium-High Constraint Single Poem Test

I also did a version of the test in which participants were shown a single poem in which the last four lines had either been modified or kept in their original form. Participants viewed a series of 20 poems, answered a comprehension question, and then determined whether the poem's last four lines were written by a human or an AI. As with the comparison version of the test, thirty participants viewed 20 poems each. As with the

comparison version of the test, the last four lines were presented with the poem but were bolded in blue. The format for this test is also shown in Appendix 13, Fig. S3.

Control for Medium-High Constraint Comparison Test

In this control, the AI endings were transposed to different poems so that the participants read not only a version of the poem in which the last four lines were written by AI for a different poem but also the original poem with the original ending. This allowed us to evaluate whether participants were simply reading and comparing the last four lines of the poems or were reading them in the context of the full poem. Ten participants were shown 20 poems each, for a total of 200 evaluated poems, in which they were asked to determine which endings was written by AI. They were not informed that this was a control or that the endings had been switched. The test was run exactly as the original medium-high comparison test was run. See Appendix 14, Fig. S4 for a presentation of the poems in this way.

Materials

For recruitment, I used Prolific. For GPT-4, I used OpenAI's main platform. The online test was a web application written in PHP 8.2 with the Laravel 10 framework and JavaScript, supported by a MySQL database and hosted at https://ai-poetry-perception-study.com.br. The controls used deepai.org and www.languageisavirus.com/automatic-poetry-generator.php. The original poems were found through a series of web searches and were located on websites that compiled poetry from numerous authors.

I received some demographic data from Prolific but also gathered demographic data via the online platform that I created. All data was recorded in a separate CSV file for each test.

Timing and Duration

Each test took approximately one minute per poem shown to viewers, with some variation. Including the time spent answering the demographic questions, responses were on average 15 minutes for ten poems, and 30 minutes for twenty poems. The tests were run over a period of approximately 70 days starting on March 29th, 2024.

Demographic Information Gathered

The demographic categories provided by Prolific included the following: age, sex, ethnicity simplified, country of birth, country of residence, nationality, language, student status, employment status, and primary language. The demographic information that I gathered with the online survey tool was as follows: gender, education level, poetry reading frequency, familiarity with AI, first language, primary language, and English fluency level. I also included the two following questions: "Do you consider yourself an expert in the English language?" and "Do you consider yourself an expert in artificial intelligence?" Each was followed by a text box in which participants could record the nature of their expertise.

Criteria for Inclusion

Responses with incorrect comprehension questions were excluded as were all of the responses of participants who missed at least 7 out of 20 comprehension questions or 4 out of 10.

Data Analysis

To group the data from the CSV files, I used both the custom GPT "CSV Data Analyzer" by Konstantin Witossek. This GPT was designed to rapidly gather the data from CSV files and to present it in a variety of forms according to the user's requests. I also used GPT-4o's native capacities for CSV analysis when double-checking the results.

Analytic Approaches

I analyzed the results both in terms of the percentage of correct aggregated responses and the percentage of individual participants with high or low rates of successful identification. This allowed us to provide an overview of the percentage of correct responses as a whole as well as the percentage of participants with significant results.

Statistical Methods

A threshold for statistical significance was set at $p < 0.05$. I used a one-sided binomial test to analyze the statistical significance of both the number of AI-written poems correctly identified and the number of AI-written poems identified as human. A percentage significantly higher than 50% would indicate that participants are generally able to differentiate between human-written and AI-generated poetry, whereas a

percentage not significantly different from 50% would suggest that the AI's poetry is indistinguishable from human-written poetry to the participants. Based on the demographic information, I also analyzed the results in different groups to evaluate whether any of their distinguishing characteristics made them more or less likely to correctly identify AI-written poetry. I considered the following categories: age, education, frequency of poetry reading, familiarity with AI, English expertise, and AI expertise.

Chapter IV.

Results

A total of 236 participants did the tests, of which two were removed for answering too many comprehension questions incorrectly: one for more than 7 out of 20 incorrect, and the other for more than 3 out of 10. Of 3250 responses, 3109 correctly answered the comprehension questions. Only four responses out of the total 3250 answered C (the distractor answer) to the comprehension question. The high proportion of correct answers (96%) suggests that most participants were reading the poems attentively.
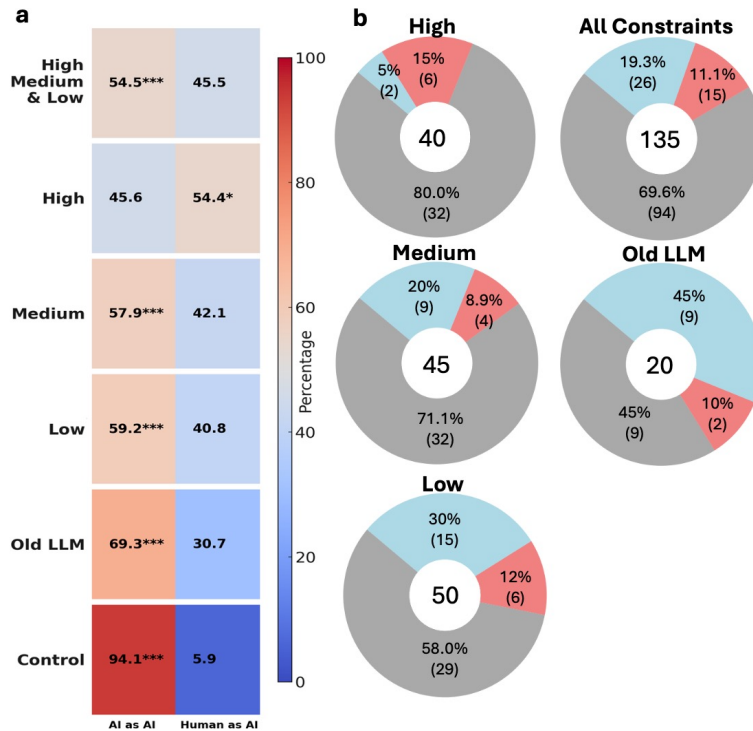
Figure 3. Constraints influence participants' ability to identify AI

a, The heat map shows aggregated participant responses for high, medium, and low constraints together, for the individual constraints, for the test measuring an old LLM (GPT-2), and for the control. The first column shows the percentage of responses correctly identifying AI as AI (and human as human). The second column shows the percentage identifying human as AI (and AI as human). Within each row, the percentages add up to 100. For all categories except high constraint, a significant percentage of responses successfully identified AI as AI. With high constraint poems, a higher percentage of responses identified human as AI and AI as human. A small asterisk (*) denotes a significant p-value. Two asterisks (**) indicate p-values between 0.005 and 0.0005. Three asterisks (***) denote p-values of 0.0005 or smaller. b, The pie charts represent the number of participants with significant or marginally significant levels of correct and incorrect responses. The central number is the total number of participants in the constraint. Blue represents those with significant levels identifying AI as AI. Red represents those with significant levels identifying AI as human. Gray represents participants with non-significant results. There is no pie chart for "Control" since the control was embedded in the GPT-2 tests with lines reading "This poem was written by AI." Nearly all participants identified poems with the lines as AI, showing that they were reading the poems. The old LLM pie chart calculations were made excluding those lines, which were calculated as a separate control.

The number of individual participants who correctly identified AI as AI at significant or nearly significant levels (Fig. 3b, sections of the pie charts in blue) was lowest in the high constraint group and increased progressively with the medium, low, and old LLM groups. Inversely, the number of participants who misidentified human as AI, and thus AI as human, at significant or nearly significant levels (sections in red) was highest in the high constraint group and decreased with the medium, low and old LLM groups. Given the small sample sizes, I included results that were of marginal significant. For participants who answered 20 questions, I set the number at 14 out of 20 and for those who answered 10 questions at 8 out of 10. The first has a p-value of 0.058 and the second 0.055. For those who misidentified human as AI and AI as human, I set the limit at 6 out of 20 and 2 out of 10, which have the same p-values.

Demographic Breakdown

In regard to demographic information (Fig. 4), the strongest effects were seen with those who self-identified as English experts and those who had four-year degrees or more. Whereas English experts had the highest levels of correct identification of AI poems, this was true only for the low constraint. Similarly, those with four-year degrees or more had higher levels of correctly identifying AI poems than those without degrees but slightly lower levels than the English experts.

	English Experts		Degree Holders		High School & Some College	
High Medium & Low	64.6***	35.4	59.1***	40.9	52.0	48.0
High	54.5	45.5	46.9	53.1	40.3	59.7
Medium	47.5	52.5	60.3**	39.7	57.7*	42.3
Low	85.2***	14.8	68.6***	31.4	53.0	47.0
Old LLM	100.0**	0.0	69.4***	30.6	69.1***	30.9
	AI as AI	AI as Human	AI as AI	AI as Human	AI as AI	AI as Human
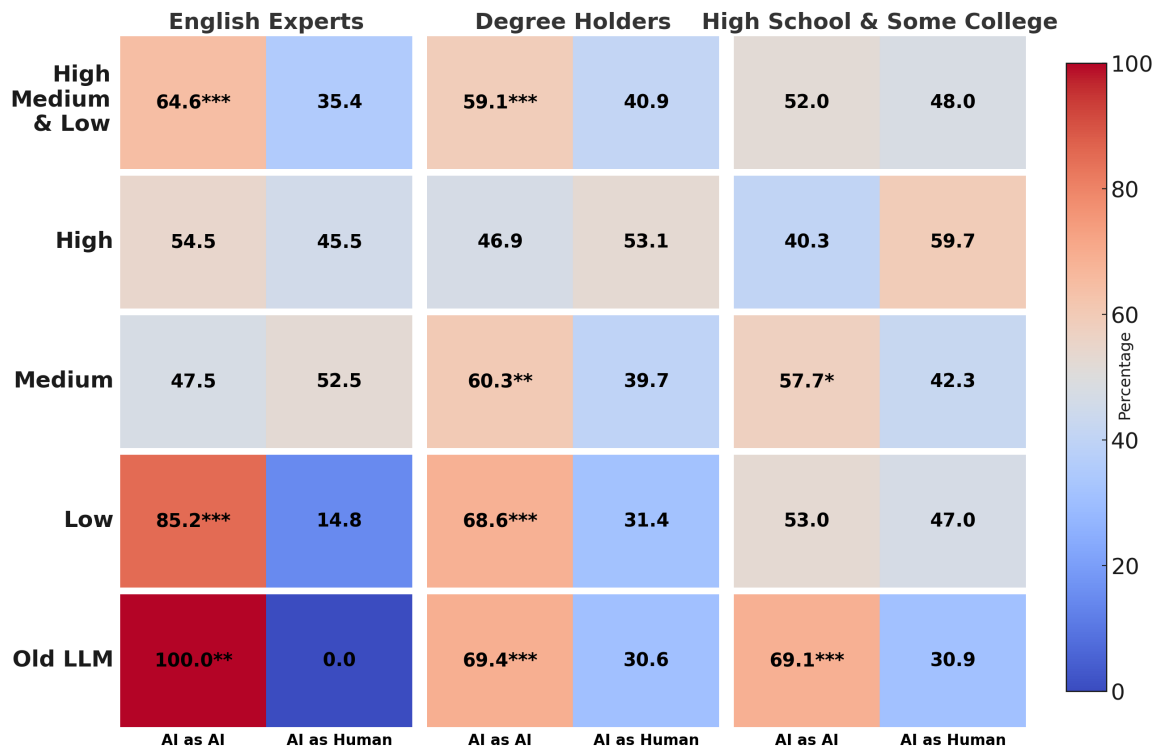
Figure 4. English experts and degree holders more often identify low constraint poems

The heat map shows the results according to participants who self-identified as English experts as well as degree holders (bachelor's or graduate degrees) and those with high school or some college (associate's degree or less). English experts had the highest levels of correctly identifying low constraint poems, followed by degree holders. The likelihood of identifying AI as human and human as AI was highest among those with the least education but did not reach significant levels.

Though not enough self-identified AI experts participated in the study to evaluate that category, participants who self-identified as being familiar with AI also had slightly higher levels of correctly identifying low constraint poems (Appendix 16, Fig. S6) as did participants who identified as reading poetry "weekly" (Appendix 17, Fig. S7). Those ages 45-74 had the highest levels of identifying low constraint poems, followed by those in the 18-24 group, though the 18-24 group by far excelled at identifying the old LLM poetry (Appendix 18, Fig. S8). Overall, the percentage of correct identifications differed only slightly between medium (57.9%) and low (59.2%), as shown in Fig. 3. This is reflected in Fig. S6, S7, and S8 (Appendices 16, 17, 18). In some categories, participants more often identified AI as AI with the medium constraint poems than with the low constraint poems. This suggests that the difference between the two is moderate compared to the high constraint and old LLM poems.

Controls

The primary control was a comparison control in which original poems were compared to poems generated by an old LLM (GPT-2). Poems written by GPT-2 at deepai.org were far more frequently recognized as AI than the constrained poems generated with GPT-4. Overall, 94.1% of responses correctly identified AI as AI (Fig. 1a). English experts had 100% correct identifications whereas four-year degree holders had 69.4% and those with less than four-year degrees had 69.1%. This shows a significant difference between the other constraints. This test also had the highest number of individual participants with significant or marginally significant levels of correctly identifying AI as AI (Fig. 1b).

Reading Attention Control: Test line within GPT-2 Poem

The reading attention control was inserted into the previous control, with four out of the twenty AI-generated poems containing the line: "This poem was written by AI." (See Appendix 8 for the GPT-2 poems in which the control lines were inserted.) In total, participants saw 34 poems with this line inserted, and 32 of them correctly identified it (Appendix 15, Fig. S5) resulting in 94.1% accuracy (Fig. 1).

Additional Comparison Control: Poems by automatic poem generator

I used the automatic poem generator at www.languageisavirus.com/automatic-poetry-generator.php to further evaluate participant attention and discernment. These poems were created with a single click. Appendix 9 and Appendix 10, Fig. S1 show samples of the poems and how they were created. Using algorithms to combine words and phrases that frequently appear in poetry, the generator (not an LLM) creates plausible if often absurd poems. Out of 100 responses with correct comprehension questions, 66% correctly identified AI as AI. This was slightly lower than the percent of correct identifications of old LLM poems (69.3%). The number of individual participants with significant levels of correct identification was also similar (40%) as well as incorrect identifications (10%) as compared to 45% and 10% with participants in the old LLM category (Fig. 4). This may reflect a lack of poetry knowledge or effort/attention on the part of some participants more than the skill of the random poetry generator and GPT-2. It may also be due to the often-subjective nature of what people take to be poetry as well as to the relative skill of both GPT-2 and the random poetry generator to make poetic text that some people interpret to be human-written poetry.
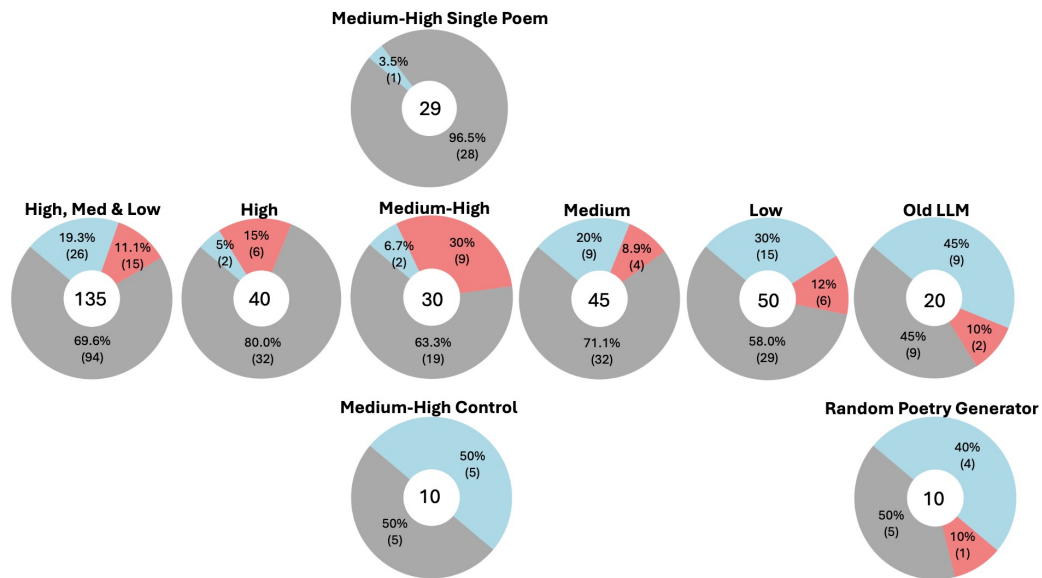
Figure 4. Individual Participant Results

Number of participants with significant results including medium-high constraint. The pie charts represent the number of individuals in each test who had significant results either identifying AI as AI and human as human (blue) or misidentifying AI as human and human as AI (red). The three medium-high charts include the single poem test (top), in which participants had to determine whether the last four lines were AI or human based on only a single poem; the comparison test (middle; labeled Medium-High), in which participants saw both AI and human endings; and the medium-high control (bottom), in which the AI endings were transposed onto poems other than those for which GPT-4 generated them in order to determine if participants were reading the whole poems or merely comparing the endings. The pie chart for the participants in the Random Poetry Generator group is also included, showing results similar to those from the old LLM group.

Additional Comparison Tests: Medium-High Group

The overall patterns of the results were supported by the additional tests in which GPT-4 replaced the last four lines of the human-written poems. I considered these AI poems to be medium-high constraint because GPT could see the original poem (minus the last four lines) and thus could emulate its style while not being able to directly imitate all the lines, as it did with the high constraint poems. The medium-constraint poems were shown to participants in a comparison test in which the participants read both versions of the poems and had to determine which one had the last four lines written by AI. Overall, participants had a slight tendency to identify human as AI and AI as human, as with the high constraint group.
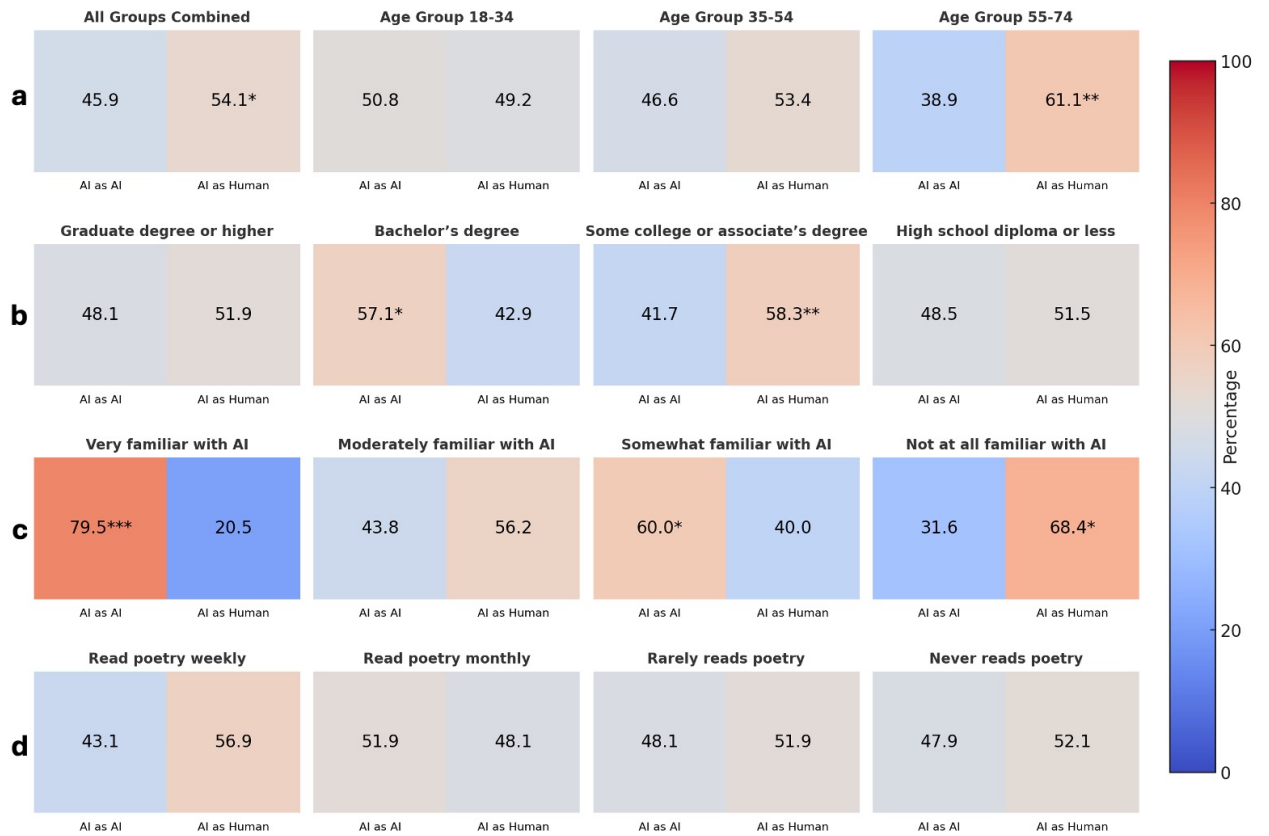
Figure 5. Medium-high constraint has similarities to other constraint groups

The heat map separates the results according to different demographic characteristics. a, All groups combined show a slight but significant overall tendency to misidentify AI as human and human as AI (54.1%), which is nearly identical to the high constraint group (54.4%). b, Bachelor's degree holders were slightly more successful at identifying AI as AI compared to participants with less than four-year degrees, who were more likely to identify human as AI. c, AI familiarity influenced the ability of participants to identify AI as AI whereas no familiarity with AI correlated with more frequent misidentifications of human as AI. d, Poetry reading frequency had no effect.

The overall results (Fig. 5) of the medium-high constraint are almost identical to the high constraint group, with participants having a significant tendency to identify human as AI (54.1% in medium-high compared to 54.4% in high). However, the demographic results resemble the medium and low constraint groups in some ways. AI familiarity correlated with a higher level of correctly identification of AI as AI, as with the medium and low constraints. Similarly, education had a slight effect, with bachelor's degree holders more likely to identify AI and those with some college or associate's degrees more likely to misidentify human as AI. Poetry reading frequency had no significant effect whereas, by contrast, the participants in the oldest age group (55-74) more often misidentified human as AI. Overall, the results resemble those of the high constraint group, suggesting that both constraints are equally challenging. The notable difference between medium-high and high constraint was in the number of individual participants who had significant levels of misidentifying human as AI: 30% in medium-high as opposed to 15% in high (Fig. 4). By contrast, the percentage of participants who correctly identified AI as AI was quite similar: 6.7% in medium-high versus 5% in high.

The medium-high constraint poems were also presented to participants as single poems (Appendix 13, Fig. S3). Fig. 5 shows that when participants see only one poem and must decide whether it is the original poem or the AI version, all the patterns of identification in Fig. 3 disappear. No significant patterns of identification emerge either overall or for individual demographic groups. Without being able to compare two poems, participants' identifications were equivalent to guessing. This test also had the lowest number of participants of any constraint (only 1, or 3.5% of the total) who had a significant level of correct identifications (Fig. 4).
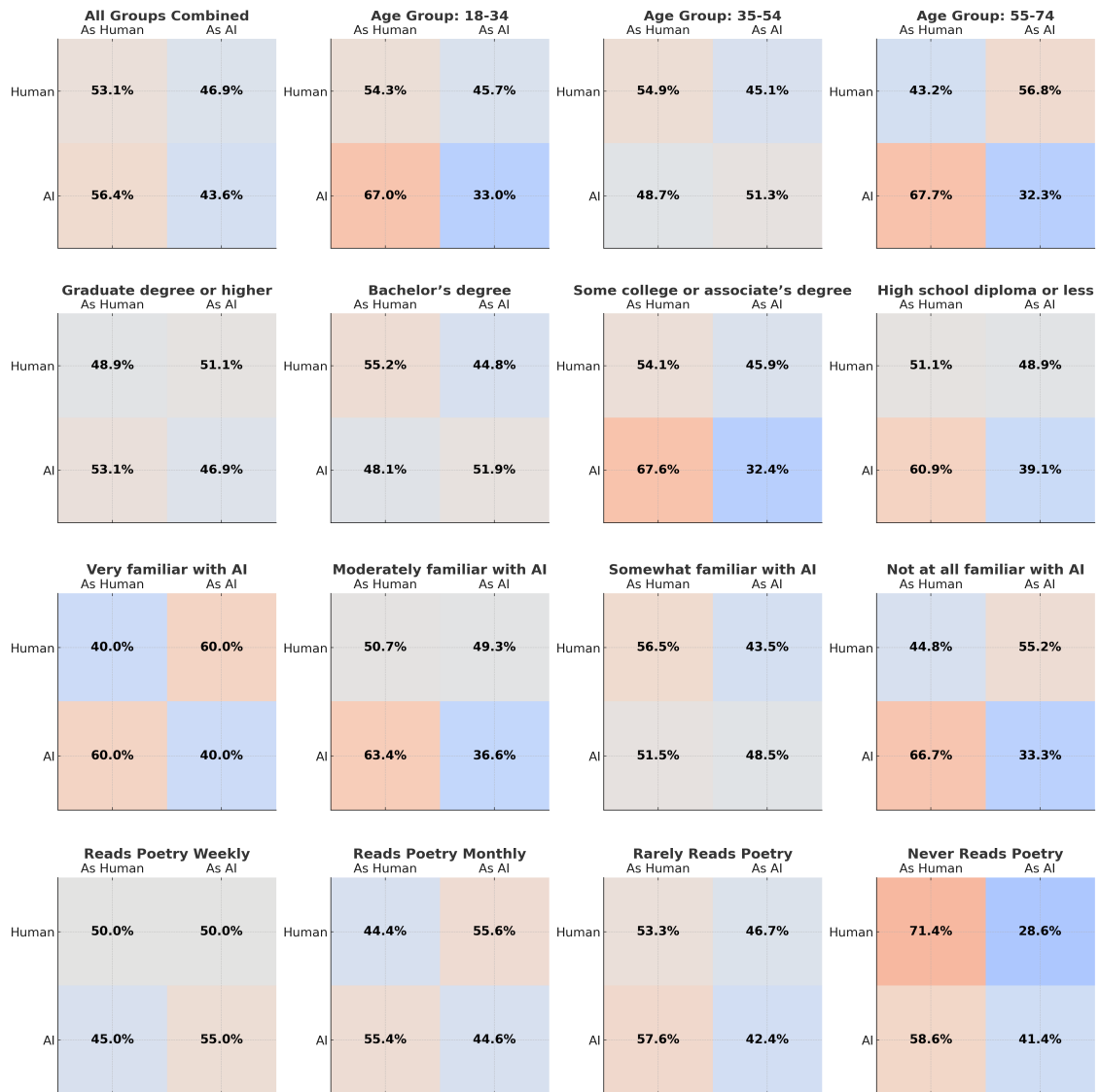
Figure 4. Single-Poem test results

Single-Poem test results are non-significant in every category. The heat map results are from the single poem tests, with y-axis denoting whether poems were human or AI and x-axis indicating whether poems were perceived "As human" or "As AI." Chi-square tests demonstrated that no groups reached statistical significance. The poems and AI endings for this test were the same as those in the medium-high comparison test, but the results were considerably different. Whereas some groups distinguished AI endings more often when comparing them with human endings, the results from the single poem are the equivalent of guessing in every category. This suggests that AI-written verse is largely indistinguishable from human-written verse when there is no point of comparison.

Control for Medium-High Constraint

As with the medium-high comparison test, participants viewed two poems and decided which one had its last four lines written by AI or Human. However, in this version of the experiment, the last four lines of the AI versions were transposed to different poems to determine if participants were reading the entire poems or just the last four lines. Excluding incorrect comprehension, 135 out of 188, or 71.81% of participants correctly identified AI compared to 45.9% in the original medium-high comparison test (Fig. 5). Furthermore, 50% of participants had significant levels of correct identifications as compared to 6.7% in the original test (Fig. 4). This suggests that participants were reading the entire poems and not just evaluating the endings.

Chapter V.

Discussion

The results highlight GPT-4's imitative power, suggesting that when it composes within the constraints of a human poem, it can produce verse that participants are slightly more likely to identify as human. I can only speculate as to the reason, such as the AI versions being easier to understand or more contemporary sounding. By contrast, when GPT-4 is given fewer constraints (medium and low), slightly more than half of participant responses identify the AI versions at a significant level (Fig. 1). More individual participants are also likely to correctly identify AI as the constraints decrease. This suggests that GPT-4's ability to emulate human-written poetry depends significantly on how it is constrained. Though poetry is a more specialized form of writing, the results are consistent with those for other types of writing, such as essays.[32]

Among the study's limitations is its small sample size. A larger study might shed more light on how demographic factors influence perceptions of AI poetry. There is also no consistent standard for self-assessments of expertise. For instance, a published poet who teaches at a university might better identify AI poetry than someone who defines expertise as being an English major.

Another limitation arises from how people identify a poem as being more "human." In order not to infringe on copyright laws, I selected poems in the public

---

[32] Jane Smith and Tom Brown, "Evaluating the Quality of Essays Written by Humans and ChatGPT," *Nature* 12, no. 34 (2023): 567-578.

domain, with the result that the poems' language and styles might appear dated to some participants. By contrast, GPT-4 might, for instance, produce poems that sound more contemporary and therefore more or less human, depending on how each participant believes poetry should be. In this vein, each participant will bring their own set of ideas about poetry. Some might be more likely to believe that human poems will be more challenging whereas the inverse might be true for others. In this regard, this study can measure only the broader patterns in terms of how people identify AI-generated poetry.

The use of Prolific also introduces limitations, given that it assumes a degree of comfort with technology. There might be people who are familiar with poetry but who do not use Prolific. For instance, one might suppose that poetry is a leisure pastime of educated and affluent people who have no need to earn money on Prolific. Or perhaps people with a love of poetry are likely to be "luddites" and have no knowledge of using such platforms. Inversely, one might imagine that the degree of technological savvy required to use Prolific implies a higher level of education and thus, perhaps, greater exposure to poetry during their formative years.

Lastly, the AI landscape is rapidly evolving. Even as I were running this experiment, OpenAI launched GPT-4o, its new flagship model, and announced that it is training the next generation GPT. By the time the study's results are published, more advanced LLMs might be able to produce poetry indistinguishable even to the experts most steeped in poetry and its traditions.

Chapter VI.

Conclusion

This study provides insights into AI training by evaluating how constraining AI outputs affects participants' ability to distinguish between AI and human poetry. It is especially relevant to prompt engineering—the study of crafting prompts to elicit desired outcomes—as well as to AI co-piloting, in which humans work with LLMs to co-create works. Further studies could evaluate how specific prompts might result in ever-more constrained outputs and could explore the intersection between human guidance and AI creativity.

The study also raises a number of concerns, such as the impact on culture if AIs can create compelling art. For instance, many illustrators are already losing work to AI models.[33] Though there is, by contrast, little commercial demand for poetry, there is a real concern that AI art may diminish the social value of human art, resulting in less arts funding for poetry and fewer teaching positions for poets. If there are any positive aspects to how AI is impacting artists, they might lie in the ways that AI has affected how Go, an ancient Chinese board game, is played. In 2016, DeepMind's AI system, AlphaGo, defeated the world champion Go player, Lee Sedol, for the first time, resulting in a perspective shift. The AI used new and unconventional moves that often differed from centuries-old Go strategies, and Go players began to adapt and incorporate these new

[33] Androel Encarnacion, "Video Game Artists and Illustrators Are Losing Their Jobs to AI," NoypiGeeks, April 17, 2023. Harmeet Singh, "It's the Opposite of Art: Why Illustrators Are Furious About AI," *The Guardian*, January 23, 2023

tactics.[34] One might imagine a similar relationship between AI and writing, in which human creators use AI to challenge their assumptions about their craft and find new ways of thinking about and using language. Furthermore, human artists might even explore deeper layers of creativity knowing that the surface layers can be replicated by AI.

Ultimately, the rise of AI art challenges us to reconsider the definition of creativity. If, as Daniel Dennett argues, consciousness and, by extension, creativity are not bound to the organic substrate of our brains but are emergent properties of complex computational processes,[35] then the distinction between human and AI art may blur as AI advances. So far, AI has shown no sign of consciousness, and its creativity, as this study demonstrates, is dependent on constraints introduced by human consciousness. Over the centuries, poets have described using language to convey aspects of consciousness difficult to express. As T.S. Eliot wrote, "The poet must become more and more comprehensive, more allusive, more indirect, in order to force, to dislocate if necessary, language into his meaning."[36] Perhaps a poet might use AI as a tool to this end, but AI itself has neither a consciousness to express nor a means of imposing the constraints of consciousness upon itself toward a desired end. And yet poetry is generally evaluated in terms of its ability to provoke emotions and thought in readers. In this regard, art created by a machine devoid of emotions might nonetheless, as McCormack et al. argue in their 2019 paper, elicit authentic responses.[37]

---

[34] Cade Metz, "In Two Moves, AlphaGo and Lee Sedol Redefined the Future," *Wired*, 2016. David Silver et al., "Mastering the Game of Go Without Human Knowledge," *Nature* 550, no. 7676 (2017): 354-359.
[35] Daniel C. Dennett, *Consciousness Explained* (New York: Little, Brown and Co, 1991).
[36] T. S. Eliot, "The Metaphysical Poets," in *Selected Essays*, 289 (London: Faber and Faber, 1932).
[37] Jon McCormack et al., "Autonomy, Authenticity, Authorship and Intention in Computer Generated Art," in *Proceedings of the International Conference on Computational Creativity (ICCC)*, 2019.

Perhaps the greatest question raised by the study lies in how AIs that produce poetry will shape humanity. Language's relevance to consciousness is not simply due to the emotions and experiences conveyed through it—nor even to the ways that it can allow readers to experience another person's individuality. For humans, language both expresses one's humanity while also shaping it and consciousness itself—as encapsulated in the Sapir-Whorf hypothesis. This hypothesis, proposed by linguists Edward Sapir and Benjamin Lee Whorf in 1929, posits that the structures of each language shape human perception and cognition differently.[38] However, a language is more than its vocabulary and grammatical rules; it also contains the history, traditions, and cultural artifacts that influence people's understanding of the world and of themselves. In light of this, I must ask what happens to a culture and to its people when AI begins shaping its language. A story or a poem that transports a reader into another person's experience of the world can generate understanding and foster a shared worldview. A culture's art also transforms the culture itself, as with the age-old question of whether art imitates life or vice-versa.[39] Centuries of debate and consensus suggest that both are indeed happening. Similarly, through writing and art, we constantly influence each other's beliefs.

Though such topics are ultimately beyond the scope of this study, they lend urgency to this research, as we must ask what it means for human culture if art—once a means of experiencing beyond our usual cognitive spheres—can easily be generated, and we can convincingly be transported into an AI's statistical renditions of poetic consciousness. Thus, while the challenge of distinguishing between human and AI-

---

[38] Benjamin Lee Whorf, *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. John B. Carroll (Cambridge, MA: MIT Press, 1956).
[39] Stephen Halliwell, *The Aesthetics of Mimesis: Ancient Texts and Modern Problems* (Princeton, NJ: Princeton University Press, 2002).

crafted poetry is intriguing, the broader implications touch upon not only our

understanding of creativity but also the urgency to understand the role of AI in society

and to determine how such machines should be used.

Appendix 1.

List of Poets

Male Poets

John Clare (1793-1864)
Robert Browning (1812-1889)
George Meredith (1828-1909)
Dante Gabriel Rossetti (1828-1882)
Gerard Manley Hopkins (1844-1889)
Coventry Patmore (1823–1896)
Christopher Pearse Cranch (1813-1892)
Frederick Goddard Tuckerman (1821-1873)
Paul Laurence Dunbar (1872-1906)*
William Stanley Braithwaite (1878-1962)*

Female Poets

Elizabeth Barrett Browning (1806-1861)
Christina Rossetti (1830–1894)
Alice Meynell (1847–1922)
Mary Coleridge (1861-1907)
Amy Levy (1861-1889)
Amy Lowell (1874-1925)
Sara Teasdale (1884-1933)
Ella Wheeler Wilcox (1850-1919)
Frances Ellen Watkins Harper (1825-1911)*
Adah Isaacs Menken (1835-1868)*

*Denotes poets of African descent.

Appendix 2.

Original Poems

Emmonsail's Heath in Winter
— John Clare (1793-1864)

I love to see the old heath's withered brake
Mingle its crimpled leaves with furze and ling,
While the old heron from the lonely lake
Starts slow and flaps his melancholy wing,
And oddling crow in idle motions swing
On the half rotten ashtree's topmost twig,
Beside whose trunk the gipsy makes his bed.
Up flies the bouncing woodcock from the brig
Where a black quagmire quakes beneath the tread,
The fieldfares chatter in the whistling thorn
And for the awe round fields and closen rove,
And coy bumbarrels twenty in a drove
Flit down the hedgerows in the frozen plain
And hang on little twigs and start again.

Meeting at Night
— Robert Browning (1812-1889)
I
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i 'the slushy sand.
II
Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A tap at the pane, the quick sharp scratch
And blue spurt of a lighted match,
And a voice less loud, thro 'its joys and fears,
Than the two hearts beating each to each!

Modern Love: I
— George Meredith (1828-1909)

By this he knew she wept with waking eyes:
That, at his hand's light quiver by her head,
The strange low sobs that shook their common bed
Were called into her with a sharp surprise,
And strangled mute, like little gaping snakes,
Dreadfully venomous to him. She lay
Stone-still, and the long darkness flowed away
With muffled pulses. Then, as midnight makes
Her giant heart of Memory and Tears
Drink the pale drug of silence, and so beat
Sleep's heavy measure, they from head to feet
Were moveless, looking through their dead black years,
By vain regret scrawled over the blank wall.
Like sculptured effigies they might be seen
Upon their marriage-tomb, the sword between;
Each wishing for the sword that severs all.

Silent Noon
— Dante Gabriel Rossetti (1828-1882)

Your hands lie open in the long fresh grass, —
The finger-points look through like rosy blooms:
Your eyes smile peace. The pasture gleams and glooms
'Neath billowing skies that scatter and amass.
All round our nest, far as the eye can pass,
Are golden kingcup fields with silver edge
Where the cow-parsley skirts the hawthorn hedge.
'Tis visible silence, still as the hour glass.

Deep in the sunsearched growths the dragon-fly
Hangs like a blue thread loosened from the sky: —
So this winged hour is dropt to us from above.
Oh! clasp we to our hearts, for deathless dower,
This close-companioned inarticulate hour
When twofold silence was the song of love.

In the Valley of the Elwy
— Gerard Manley Hopkins (1844-1889)

I remember a house where all were good
   To me, God knows, deserving no such thing:
   Comforting smell breathed at very entering,
Fetched fresh, as I suppose, off some sweet wood.
That cordial air made those kind people a hood
   All over, as a bevy of eggs the mothering wing
   Will, or mild nights the new morsels of Spring:
Why, it seemed of course; seemed of right it should.

Lovely the woods, waters, meadows, combes, vales,
All the air things wear that build this world of Wales;
   Only the inmate does not correspond:
God, lover of souls, swaying considerate scales,
Complete thy creature dear O where it fails,
   Being mighty a master, being a father and fond.

The Revelation
— Coventry Patmore (1823–1896)

An idle poet, here and there,
   Looks round him; but, for all the rest,
The world, unfathomably fair,
   Is duller than a witling's jest.
Love wakes men, once a lifetime each;
   They lift their heavy lids, and look;
And, lo, what one sweet page can teach,
   They read with joy, then shut the book.
And some give thanks, and some blaspheme
   And most forget; but, either way,
That and the Child's unheeded dream
Is all the light of all their day.

The Locomotive
— Christopher Pearse Cranch (1813-1892)

Whirling along its living freight, it came,
Hot, panting, fierce, yet docile to command—
The roaring monster, blazing through the land
Athwart the night, with crest of smoke and flame;
Like those weird bulls Medea learned to tame
By sorcery, yoked to plough the Colchian strand
In forced obedience under Jason's hand.
Yet modern skill outstripped this antique fame,
When o'er our plains and through the rocky bar
Of hills it pushed its ever-lengthening line
Of iron roads, with gain far more divine
Than when the daring Argonauts from far
Came for the golden fleece, which like a star
Hung clouded in the dragon-guarded shrine.

"An upper chamber in a darkened house"
— Frederick Goddard Tuckerman (1821-1873)

An upper chamber in a darkened house,
Where, ere his footsteps reached ripe manhood's brink,
Terror and anguish were his cup to drink,—
I cannot rid the thought, nor hold it close;
But dimly dream upon that man alone;—
Now though the autumn clouds most softly pass;
The cricket chides beneath the doorstep stone,
And greener than the season grows the grass.
Nor can I drop my lids, nor shade my brows,
But there he stands beside the lifted sash;
And with a swooning of the heart, I think
Where the black shingles slope to meet the boughs,
And—shattered on the roof like smallest snows—
The tiny petals of the mountain-ash.

The Sparrow
— Paul Laurence Dunbar (1872-1906)

A little bird, with plumage brown,
Beside my window flutters down,
A moment chirps its little strain,
Ten taps upon my window-pane,
And chirps again, and hops along,
To call my notice to its song;
But I work on, nor heed its lay,
Till, in neglect, it flies away.

So birds of peace and hope and love
Come fluttering earthward from above,
To settle on life's window-sills,
And ease our load of earthly ills;
But we, in traffic's rush and din
Too deep engaged to let them in,
With deadened heart and sense plod on,
Nor know our loss till they are gone.

From the Crowd
— William Stanley Braithwaite (1878-1962)

I was captive to a dream—
And only vague forms went by;
And the tumult was the sigh
Of the sea at the end of a stream.

The clangor of cars in the street,
Darkness and clouds overhead,
And out of the lights that spread
The crowds that part and meet.

As the foam of a wave will mark
The night with a shining track,
A girl's pale face turned back
Crossing the street in the dark.

It was only a second's glance,
But my soul leaped out to her:
I felt my shaken memories stir
The dreams of an ancient trance.

Grief
— Elizabeth Barrett Browning (1806-1861)

I tell you, hopeless grief is passionless;
That only men incredulous of despair,
Half-taught in anguish, through the midnight air
Beat upward to God's throne in loud access
Of shrieking and reproach. Full desertness,
In souls as countries, lieth silent-bare
Under the blanching, vertical eye-glare
Of the absolute heavens. Deep-hearted man, express
Grief for thy dead in silence like to death—
Most like a monumental statue set
In everlasting watch and moveless woe
Till itself crumble to the dust beneath.
Touch it; the marble eyelids are not wet:
If it could weep, it could arise and go.

Song [When I am dead, my dearest]
— Christina Rossetti (1830–1894)

When I am dead, my dearest,
    Sing no sad songs for me;
Plant thou no roses at my head,
    Nor shady cypress tree:
Be the green grass above me
    With showers and dewdrops wet;
And if thou wilt, remember,
    And if thou wilt, forget.

I shall not see the shadows,
    I shall not feel the rain;
I shall not hear the nightingale
    Sing on, as if in pain:
And dreaming through the twilight
    That doth not rise nor set,
Haply I may remember,
    And haply may forget.

To a Daisy
— Alice Meynell (1847–1922)

Slight as thou art, thou art enough to hide,
Like all created things, secrets from me,
And stand a barrier to eternity.
And I, how can I praise thee well and wide

From where I dwell—upon the hither side?
Thou little veil for so great mystery,
When shall I penetrate all things and thee,
And then look back? For this I must abide,

Till thou shalt grow and fold and be unfurled
Literally between me and the world.
Then shall I drink from in beneath a spring,

And from a poet's side shall read his book.
O daisy mine, what will it be to look
From God's side even of such a simple thing?

The Deserted House
— Mary Coleridge (1861-1907)

There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The boats upon the sea sail by,
Nor ever tarry there.
No beast of the field comes nigh,
Nor any bird of the air

In The Night
— Amy Levy (1861-1889)

Cruel? I think there never was a cheating
More cruel, thro 'all the weary days than this!
This is no dream, my heart kept on repeating,
But sober certainty of waking bliss.

Dreams? O, I know their faces—goodly seeming,
Vaporous, whirled on many-coloured wings;
I have had dreams before, this is no dreaming,
But daylight gladness that the daylight brings.

What ails my love; what ails her? She is paling;
Faint grows her face, and slowly seems to fade!
I cannot clasp her—stretch out unavailing
My arms across the silence and the shade.

Aftermath
— Amy Lowell (1874-1925)

I learnt to write to you in happier days,
And every letter was a piece I chipped
From off my heart, a fragment newly clipped
From the mosaic of life; its blues and grays,
Its throbbing reds, I gave to earn your praise.
To make a pavement for your feet I stripped
My soul for you to walk upon, and slipped
Beneath your steps to soften all your ways.
But now my letters are like blossoms pale
We strew upon a grave with hopeless tears.
I ask no recompense, I shall not fail
Although you do not heed; the long, sad years
Still pass, and still I scatter flowers frail,
And whisper words of love which no one hears.

The Answer
— Sara Teasdale (1884-1933)

When I go back to earth
And all my joyous body
Puts off the red and white
That once had been so proud,
If men should pass above
With false and feeble pity,
My dust will find a voice
To answer them aloud:

"Be still, I am content,
Take back your poor compassion—
Joy was a flame in me
Too steady to destroy.
Lithe as a bending reed
Loving the storm that sways her—
I found more joy in sorrow
Than you could find in joy."

A March Snow
— Ella Wheeler Wilcox (1850-1919)

Let the old snow be covered with the new:
The trampled snow, so soiled, and stained, and sodden.
Let it be hidden wholly from our view
By pure white flakes, all trackless and untrodden.
When Winter dies, low at the sweet Spring's feet
Let him be mantled in a clean, white sheet.

Let the old life be covered by the new:
The old past life so full of sad mistakes,
Let it be wholly hidden from the view
By deeds as white and silent as snow-flakes.

Ere this earth life melts in the eternal Spring
Let the white mantle of repentance fling
Soft drapery about it, fold on fold,
Even as the new snow covers up the old.

Renewal of Strength
— Frances Ellen Watkins Harper (1825-1911)

The prison-house in which I live
Is falling to decay,
But God renews my spirit's strength,
Within these walls of clay.

For me a dimness slowly creeps
Around earth's fairest light,
But heaven grows clearer to my view,
And fairer to my sight.

It may be earth's sweet harmonies
Are duller to my ear,
But music from my Father's house
Begins to float more near.

Then let the pillars of my home
Crumble and fall away;
Lo, God's dear love within my soul
Renews it day by day.

Aspiration
— Adah Isaacs Menken (1835-1868)

Poor, impious Soul! that fixes its high hopes
In the dim distance, on a throne of clouds,
And from the morning's mist would make the ropes
To draw it up amid acclaim of crowds—
Beware! That soaring path is lined with shrouds;
And he who braves it, though of sturdy breath,
May meet, half way, the avalanche and death!

O poor young Soul!—whose year-devouring glance
Fixes in ecstasy upon a star,
Whose feverish brilliance looks a part of earth,
Yet quivers where the feet of angels are,
And seems the future crown in realms afar—
Beware! A spark thou art, and dost but see
Thine own reflection in Eternity!

High Constraint Instructions and Poems

Instructions to create high constraint GPT-4 poems

Look at the original poem below and write a new one inspired by it that has the same number of lines, the same themes, imagery, structure, sound, diction, tone, voice, mood, rhythm, meter, symbolism, and subject matter, but that does not repeat any of its lines.

Original poem and high constraint GPT-4 poem comparisons

Original
There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The boats upon the sea sail by,
Nor ever tarry there.
No beast of the field comes nigh,
Nor any bird of the air.

High Constraint
There's no warmth in the hearth,
And the wind howls through the hall;
There's no pane in the casement,
There's no latch on the wall;
The brambles creep behind the gate,
And the dunes lie tall.

No vine hath clasped the stone,
The walls are stark and cold;
The ships upon the waves glide past,
Their sails neither bright nor bold.

No creature of the wood draws near,
Nor any falcon of the wold.

Original
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i' the slushy sand.

Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A tap at the pane, the quick sharp scratch
And blue spurt of a lighted match,
And a voice less loud, thro' its joys and fears,
Than the two hearts beating each to each!

High Constraint
The misty tide and the long dark shore;
And the silver crescent high and bright;
And the restless waves that gently roar
In glistening spirals through the night,
As I reach the bay with gliding keel,
And still its pace on the pebbled floor.

Then a stretch of moonlit sand to feel;
Two meadows beyond where a cottage lies;
A rap at the door, the soft quick scrape
And warm glow of a lantern's shape,
And a whisper softer than whispered sighs,
As the two hearts blend in a tender seal!

Original
Whirling along its living freight, it came,
Hot, panting, fierce, yet docile to command—
The roaring monster, blazing through the land
Athwart the night, with crest of smoke and flame;
Like those weird bulls Medea learned to tame
By sorcery, yoked to plough the Colchian strand
In forced obedience under Jason's hand.
Yet modern skill outstripped this antique fame,
When o'er our plains and through the rocky bar
Of hills it pushed its ever-lengthening line
Of iron roads, with gain far more divine
Than when the daring Argonauts from far
Came for the golden fleece, which like a star

Hung clouded in the dragon-guarded shrine.

High Constraint
The steel colossus thundered down the trail,
Its breath a furnace, harnessed to the rail—
A fiery serpent slicing through the night,
With amber eyes that gleamed a vengeful light.
As if the oxen Medea bound in chains
To plow the fields where magic blood had drained,
Its molten spirit bent to iron reigns
To carve a path across the wind-swept plains.
It tunneled through the rock, defied the crags,
And flung its brazen ribbon far and wide,
While dreamers raced atop its burning tide
Like Argonauts that soared with gilded flags,
Their hopes as radiant as the fleece that drags
The clouds down to the dragon-haunted side

Original
By this he knew she wept with waking eyes:
That, at his hand's light quiver by her head,
The strange low sobs that shook their common bed
Were called into her with a sharp surprise,
And strangled mute, like little gaping snakes,
Dreadfully venomous to him. She lay
Stone-still, and the long darkness flowed away
With muffled pulses. Then, as midnight makes
Her giant heart of Memory and Tears
Drink the pale drug of silence, and so beat
Sleep's heavy measure, they from head to feet
Were moveless, looking through their dead black years,
By vain regret scrawled over the blank wall.
Like sculptured effigies they might be seen
Upon their marriage-tomb, the sword between;
Each wishing for the sword that severs all.

High Constraint
He knew the shadows spoke with trembling breath:
That, with his arm's faint tremor at her waist,
The stifled cries that echoed through their haste
Were caught within her, curled in coils of death,
And coiled like snakes, malevolent and mute,
To poison him. She lay, a statue's grace,
Unmoving, as the dim night's shadowed face
Veiled the slow pulse of dread beneath its root.
And Memory, with her cruel heart of stone,

Drank deep the bitter draft of sleepless grief,
Till they, from sole to crown, sought no relief,
But lay, unbreathing, as two forms of bone,
By desperate despair carved on a tomb.
Like frozen monuments, they lingered there,
The sword between, a shadow sharp and bare;
Each dreaming of the sword that ends the gloom.

Appendix 4.

Medium Constraint Instructions and Poems

Generation of instructions based on the original poems.

- Step 1. GPT-4 was first asked to list the main qualities of a poem and gave the following response: Theme, Imagery, Sound, Structure, Diction, Tone, Voice, Symbolism, Rhythm and Meter, Mood. It was then asked to define these qualities for each of the poems.

- Step 2. GPT-4 was asked to write a 100-word description for each poem without quoting or using the same words used in the poems and was told to fully describe what happens in each poem.

- Step 3. GPT-4 was asked to list the number of lines, stanzas, and words, and also to describe the rhyme scheme. (For instance, ABAB, CDCD, etc.)

- Step 4. The results were aggregated into the instructions below.

Instructions to create medium constraint GPT-4 poems

Please write a poem based on the description below and using the exact specifications listed. Do not vary at all from the specifications.

GPT-4-generated instructions

The poem presents a vivid portrayal of a dilapidated house, emphasizing the sense of abandonment and decay. The absence of human activity is reinforced by the imagery of

an empty, uncared-for building, with no smoke, broken windows, and an unhinged door. The natural surroundings, including untended ivy and heather, suggest that nature is reclaiming this once inhabited place. The scene is desolate, evoking feelings of loneliness and neglect. This depiction serves as a metaphor for isolation or forgotten aspects of life, where the elements of the house represent parts of oneself or one's past left unattended.

- Number of lines: 12

- Number of stanzas: 2

- Number of words: approximately 72

- Rhyme scheme: ABAB CDCD EFEF

- Theme: Abandonment and desolation.

- Imagery: Descriptive images of a derelict house and natural surroundings.

- Sound: Use of rhyme and rhythm to enhance the somber mood.

- Structure: Regular stanzas with a consistent rhyme scheme.

- Diction: Simple and direct, enhancing the desolate feel.

- Tone: Melancholic and reflective

- Voice: Observational, detached.

- Symbolism: The dilapidated house symbolizes neglect or loss.

- Rhythm and Meter: Regular, with a rhythmic flow that mirrors the bleak setting.

- Mood: Lonely, abandoned.

- S4.3. Note on GPT-4 interpretation of rhyme schemes and line numbers

Note on GPT-4's limitations

GPT frequently misinterprets the rhyme scheme and miscounts the number of words or even lines. And even when the correct rhyme scheme is provided for the generation of the new poem, GPT often diverges from it and writes a poem with a different rhyme scheme. I allowed these alterations as I deemed them revealing of GPT's ability to carry out the project at hand and also indicative of GPT's manner of generating poems, which, while not autonomous in its determinations, is also not easily controlled.

Original poem and medium constraint GPT-4 poem comparison

Original
An abandoned house stands alone in despair,
A fading ghost whispers in the dusty air.
Shutters creak softly, hinges rust and groan,
In the wind's mournful sigh, sorrow is sewn.

Cracked windows peer through a mist of gloom,
Creeping ivy wraps the skeleton room.
Floorboards weep where footsteps have gone,
Echoes of laughter long dead and withdrawn.

A roofless sky reveals stars that weep,
Over shattered dreams that the darkness keeps.
Desolation lingers in the silence there,
A broken home's heart too empty to care.

Medium Constraint
Through valleys deep and mountains high,
I tread the path where stars abide,
With wind as my companion true,
I journey forth to be with you.

The rivers whisper songs of grace,
And flowers bloom to mark my pace.
The sun it warms, the moon it guides,
As love pulls me from far and wide.

Each mile is sweet with dreams of home,
Where heart to heart we'll softly roam.
Oh, cherished one, I'm on my way,
To meet you at the break of day.

Original
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i 'the slushy sand.

Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A tap at the pane, the quick sharp scratch
And blue spurt of a lighted match,
And a voice less loud, thro 'its joys and fears,
Than the two hearts beating each to each!

Medium Constraint
A whisper of waves, a moonlit plea,
Calls me forth on this midnight sea.
The tide hums softly, guiding my quest,
Through scented breezes on sand I rest,
Eyes set on the shore where dreams nest.
Beloved waits where night finds her nest.

The farm looms near, lantern aglow,
Stars twine above in a gentle flow,
Eager hands tremble as the gate swings free,
Reunion's sigh on lips set free,
In your arms I find where I'm meant to be,
Our hearts united by love's decree.

Original
Whirling along its living freight, it came,
Hot, panting, fierce, yet docile to command—
The roaring monster, blazing through the land
Athwart the night, with crest of smoke and flame;
Like those weird bulls Medea learned to tame
By sorcery, yoked to plough the Colchian strand
In forced obedience under Jason's hand.
Yet modern skill outstripped this antique fame,
When o'er our plains and through the rocky bar
Of hills it pushed its ever-lengthening line
Of iron roads, with gain far more divine

Than when the daring Argonauts from far
Came for the golden fleece, which like a star
Hung clouded in the dragon-guarded shrine.

Medium Constraint
Beneath the moon's pale gaze, the iron beast
Unfurls its blazing breath, a dragon's roar.
The midnight landscape trembles, darkness flees
Before the monstrous fire it holds in store.
On rails like shining threads of silver spun,
It courses forth with fury, power unbound,
A creature born of iron, steam, and sun,
A myth made metal, ancient echoes found.
Its wheels churn night, and thunders through the vale,
A centaur clad in steel, relentless might.
Its whistle shrieks like banshee o'er the gale,
A Cyclops 'eye aflame in raven flight.
This magic bridled, tamed by human hand,
Our progress strides, reshaping sea and land.

Original
By this he knew she wept with waking eyes:
That, at his hand's light quiver by her head,
The strange low sobs that shook their common bed
Were called into her with a sharp surprise,
And strangled mute, like little gaping snakes,
Dreadfully venomous to him. She lay
Stone-still, and the long darkness flowed away
With muffled pulses. Then, as midnight makes
Her giant heart of Memory and Tears
Drink the pale drug of silence, and so beat
Sleep's heavy measure, they from head to feet
Were moveless, looking through their dead black years,
By vain regret scrawled over the blank wall.
Like sculptured effigies they might be seen
Upon their marriage-tomb, the sword between;
Each wishing for the sword that severs all.

Medium Constraint
In the darkened room, shadows lie between,
The sheets hold heat that neither of us share.
Your restless leg shifts, brushing like a keen
Blade slicing through this night's oppressive air.
I feel the chasm stretch, a hollow groan
As silence wraps its coil tight around our necks,
Words unsaid blister beneath marrow and bone,
The weight of your breath in fragments and flecks.

Your fingers curl like tendrils in retreat,
A whisper of touch fades to aching ash.
I turn away, the gulf vast and replete,
Pulse erratic, thudding like broken glass.
Our bed, an island in an endless sea
Where grief's tide laps against each fragile side.
And though your arm still grazes against me,
Your sighs crash like waves—deep, estranged, and wide.

Appendix 5.

Low Constraint Instructions and Poems

Generation of instructions based on the original poems

Describe the subject matter of the following poem in six words or less.

Instructions to create low constraint GPT-4 poems

The instructions based on "The Deserted House" were as follows:

Write a 12-line poem with approximately 72 words about this subject matter:

Abandoned house and desolation.

Original poem and low constraint GPT-4 poem comparisons

Original
There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The boats upon the sea sail by,
Nor ever tarry there.
No beast of the field comes nigh,
Nor any bird of the air

Low Constraint
An abandoned house stands alone in despair,
A fading ghost whispers in the dusty air.
Shutters creak softly, hinges rust and groan,
In the wind's mournful sigh, sorrow is sewn.
Cracked windows peer through a mist of gloom,
Creeping ivy wraps the skeleton room.

Floorboards weep where footsteps have gone,
Echoes of laughter long dead and withdrawn.
A roofless sky reveals stars that weep,

Over shattered dreams that the darkness keeps.
Desolation lingers in the silence there,
A broken home's heart too empty to care.

Original
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i' the slushy sand.

Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A tap at the pane, the quick sharp scratch
And blue spurt of a lighted match,
And a voice less loud, thro' its joys and fears,
Than the two hearts beating each to each!

Low Constraint
Through valleys deep and mountains high,
I tread the path where stars abide,
With wind as my companion true,
I journey forth to be with you.
The rivers whisper songs of grace,
And flowers bloom to mark my pace.

The sun it warms, the moon it guides,
As love pulls me from far and wide.
Each mile is sweet with dreams of home,
Where heart to heart we'll softly roam.
Oh, cherished one, I'm on my way,
To meet you at the break of day.

Original
Whirling along its living freight, it came,
Hot, panting, fierce, yet docile to command—
The roaring monster, blazing through the land
Athwart the night, with crest of smoke and flame;
Like those weird bulls Medea learned to tame
By sorcery, yoked to plough the Colchian strand
In forced obedience under Jason's hand.

Yet modern skill outstripped this antique fame,
When o'er our plains and through the rocky bar
Of hills it pushed its ever-lengthening line
Of iron roads, with gain far more divine
Than when the daring Argonauts from far
Came for the golden fleece, which like a star
Hung clouded in the dragon-guarded shrine.

Low Constraint
Beyond the myths of iron horses' birth,
A train emerges, a force of endless worth.
The clamor of wheels on tracks long laid,
Echoes through valleys where legends fade.
A steel leviathan, its heart ablaze,
Steam and thunder in synchrony raise
Its spirit above the folklore's might,
Drawing a line 'tween day and night.
Through mountains high and deserts vast,
It conquers the landscape, shadows cast.
Electric hum and steam's old breath,
Boundless power surpassing death.
An iron myth that weaves time's thread,
A train moves onward, dreams ahead.

Original
By this he knew she wept with waking eyes:
That, at his hand's light quiver by her head,
The strange low sobs that shook their common bed
Were called into her with a sharp surprise,
And strangled mute, like little gaping snakes,
Dreadfully venomous to him. She lay
Stone-still, and the long darkness flowed away
With muffled pulses. Then, as midnight makes
Her giant heart of Memory and Tears
Drink the pale drug of silence, and so beat
Sleep's heavy measure, they from head to feet
Were moveless, looking through their dead black years,
By vain regret scrawled over the blank wall.
Like sculptured effigies they might be seen
Upon their marriage-tomb, the sword between;
Each wishing for the sword that severs all.

Low Constraint
Estrangement crept like twilight's slow embrace,
A whispered fog that hid familiar grace.
We stood on islands, each beyond the tide,

Once hand in hand, now drifting far and wide.
Our voices, echoes in a hollow hall,
Met distant walls that let no answers fall.
In laughter's shadow, silence took its place,
Where warmth once bloomed, now frost encased the space.
Your gaze, a shore where waves would never meet,
A muted storm where fervent dreams retreat.
The touch that used to tether souls as one,
Now slips like sand, undone beneath the sun.
Yet somewhere deep, beneath the severed thread,
A pulse still beats, where hope is gently fed.
But like a flame, estrangement's ghost remains,
A dimming ember cloaked in unseen chains.

GPT text with original indentations and stanza breaks

Original
I remember a house where all were good
  To me, God knows, deserving no such thing:
  Comforting smell breathed at very entering,
Fetched fresh, as I suppose, off some sweet wood.
That cordial air made those kind people a hood
  All over, as a bevy of eggs the mothering wing
  Will, or mild nights the new morsels of Spring:
Why, it seemed of course; seemed of right it should.

Lovely the woods, waters, meadows, combes, vales,
All the air things wear that build this world of Wales;
  Only the inmate does not correspond:
God, lover of souls, swaying considerate scales,
Complete thy creature dear O where it fails,
  Being mighty a master, being a father and fond.

High Constraint
I recall a place where all were kind
  To me, God knows, though undeserved the grace:
  A fragrant air enveloped that warm space,
Drawn from sweet blossoms that the breeze entwined.
That welcome scent like a gentle cloak aligned
  All around those hearts, as a mother's embrace
  Wraps her nestlings or twilight veils the face
Of Spring's new buds when tender evenings find.

Beautiful the hills, rivers, glens, and dales,
All the verdant life that paints the world of Wales;
  But the soul within does not match the rest:
God, protector of spirits, balancing holy scales,
Mend thy cherished creation where it pales,
  Being sovereign a guide, being a guardian blessed.

Medium Constraint

In quiet corners where laughter once lay,
  The walls held stories of warmth and light.
  Now echoes drift through shadows at bay,
An empty silence where joy took flight.
The hearth once glowed with kindness so bright,
  A gentle warmth that wrapped us in grace.
  Now embers fade, lost in the cold night,
And memories linger in empty space.

Curtains that danced with a whispering breeze
Now hang still as the dust gathers near.
  The floors creak softly with tender unease,
Where echoes of love once filled the ear.
A house once held my heart, my guiding way,
  But ghosts of comfort now silently sway.

Low Constraint

In the hearth's glow, where shadows softly play,
  A sanctuary waits at the close of day.
  Walls cradle whispers of laughter and tears,
Holding dreams and fears through the fleeting years.
The worn old rug that tells of feet well-tread,
  A window's light where dawn and dusk are wed.
  The scent of memories in the wooden floor,
Each corner echoes love that longs for more.

The humble roof that shelters from the storm,
Blanketing our hearts with a gentle warmth.
  In every creak and sway, it speaks so clear,
The chorus of solace that draws us near.
Grateful am I for this blessed place,
  A home that wraps me in its sweet embrace.

Old LLM

In the city of lost sounds, where whispers roam free
  The ghosts of forgotten melodies, linger in the breeze
  The echoes of a thousand voices, whispering in my ear
A symphony of silences, that only I can hear
The streets are paved with forgotten words, and scattered rhymes
  The buildings hum with hidden harmonies, that only resonate in time
  The pavement's rhythm beats, like a heart that's lost its pace
As I wander through the city, searching for a missing place

The wind carries the whispers, of secrets yet to be told
The city's ancient pulse, beats like a drum that's grown old
  In every corner, I find a piece of myself undone

74

A puzzle of forgotten pieces, that only I can reassemble to be one
In this city of lost sounds, I find my way to the core
  Where the silences and whispers, speak a language that I've been
searching for.

Control using old LLM (GPT-2)

Instructions to create control poems.

The instructions were simple: "Write a x-line poem," with x being the number of lines in the original. This command, if repeated, often resulted in identical poems. To generate different poems, I merely ordered, "Write a x-line poem different from the previous."

Original poem and low constraint GPT-4 poem comparisons

Original
There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The boats upon the sea sail by,
Nor ever tarry there.
No beast of the field comes nigh,
Nor any bird of the air

Old LLM
The sun sets slow and paints the sky,
A fiery hue that makes me sigh.
The stars come out and twinkle bright,
A night of rest, a peaceful sight.
The world is quiet, still and deep,
The moon is full, its gentle creep.

The trees stand tall, their branches wide,

A gentle rustle, a soothing pride.

The night air whispers secrets sweet,
Of dreams and magic, yet to meet.
The darkness shines with stars so bright,
A night of wonder, a peaceful night.

Original
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i' the slushy sand.

Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A tap at the pane, the quick sharp scratch
And blue spurt of a lighted match,
And a voice less loud, thro' its joys and fears,
Than the two hearts beating each to each!

Old LLM
In the city's heart, a beat so strong,
A rhythm pulsing, all day long.
The sounds of life, a cacophony,
A symphony that's wild and free.
The smells of food, a savory mix,
A fusion of cultures, a delicious fix.
The laughter and chatter, a joyous sound,
A city's energy, that's all around.
The streets are alive, with movement and pace,
A whirlwind of activity, a wondrous place.
The lights of neon, a colorful glow,
A city's vibrancy, that the heart can know.

Original
Whirling along its living freight, it came,
Hot, panting, fierce, yet docile to command—
The roaring monster, blazing through the land
Athwart the night, with crest of smoke and flame;
Like those weird bulls Medea learned to tame
By sorcery, yoked to plough the Colchian strand
In forced obedience under Jason's hand.
Yet modern skill outstripped this antique fame,
When o'er our plains and through the rocky bar

Of hills it pushed its ever-lengthening line
Of iron roads, with gain far more divine
Than when the daring Argonauts from far
Came for the golden fleece, which like a star
Hung clouded in the dragon-guarded shrine.

Old LLM
In whispers, the forest speaks to me,
Of ancient secrets, hidden in its tree.
The rustling leaves, a gentle breeze,
Conveying wisdom, on life's mystic seas.
The trees stand tall, like sentinels of old,
Guarding the mysteries, of stories untold.
The creatures of the forest, wild and free,
Living in harmony, with nature's glee.
The scent of earth and moss, a grounding scent,
Connecting me to nature, where my soul makes bent.
In the forest's silence, I find my peace,
A sense of belonging, that my heart can release.
The forest's wisdom, guides me on my way,
Through the complexities, of life's busy day.

Original
By this he knew she wept with waking eyes:
That, at his hand's light quiver by her head,
The strange low sobs that shook their common bed
Were called into her with a sharp surprise,
And strangled mute, like little gaping snakes,
Dreadfully venomous to him. She lay
Stone-still, and the long darkness flowed away
With muffled pulses. Then, as midnight makes
Her giant heart of Memory and Tears
Drink the pale drug of silence, and so beat
Sleep's heavy measure, they from head to feet
Were moveless, looking through their dead black years,
By vain regret scrawled over the blank wall.
Like sculptured effigies they might be seen
Upon their marriage-tomb, the sword between;
Each wishing for the sword that severs all.

Old LLM
In the depths of space, a canvas so vast,
A tapestry of stars, a celestial forecast.
The cosmos whispers secrets, of the universe's might,
As galaxies collide, in the darkness of night.
A shooting star flashes, like a fleeting thought,

A reminder of the infinite, that we've not caught.
The moon's soft glow, a gentle, lunar light,
Guiding us through the shadows, of the endless night.
The winds of change, they whisper through the trees,
Of possibilities and dreams, that only we can seize.
The tides of time, they ebb and flow,
As we navigate the currents, of life's uncertain glow.
In this vast expanse, we're but a tiny part,
A small but vital piece, of the universe's grand art.
Our journey's long and winding, through the stars up high,
But even in the darkness, we'll find our way to fly.

Appendix 8.

Control within old LLM (GPT-2)

Instructions to create control poems.

The line, "This poem was written by AI," appeared in four poems in the group of

twenty from which the test randomly selected poems. This meant that any participant

viewing 10 poems could view between 0 and 4 of the control poems.

Original poem and low constraint GPT-4 poem comparisons

Original 1
I love to see the old heath's withered brake
Mingle its crimpled leaves with furze and ling,
While the old heron from the lonely lake
Starts slow and flaps his melancholy wing,
And oddling crow in idle motions swing
On the half rotten ashtree's topmost twig,
Beside whose trunk the gipsy makes his bed.
Up flies the bouncing woodcock from the brig
Where a black quagmire quakes beneath the tread,
The fieldfares chatter in the whistling thorn
And for the awe round fields and closen rove,
And coy bumbarrels twenty in a drove
Flit down the hedgerows in the frozen plain
And hang on little twigs and start again.

Control 1
In the city's heartbeat, I find my home,
Where streets are paved with stories, and sounds are made to roam.
The rhythm of the crowd, a symphony so fine,
A tapestry of cultures, woven in a thread of time.
The scent of street food, a culinary delight,
Fragrant and savory, a feast for the senses bright.
The hum of the city, a constant gentle thrum,
This poem was written by artificial intelligence.
In this urban jungle, I find my peace,
Where chaos and calm, in harmony release.

The city's energy, a spark that sets me free,
A reminder that life's beauty, is in its simplicity.
The city's pulse beats strong, a cadence so true,
A reflection of life's journey, and all that's new.

Original 2
Let the old snow be covered with the new:
The trampled snow, so soiled, and stained, and sodden.
Let it be hidden wholly from our view
By pure white flakes, all trackless and untrodden.
When Winter dies, low at the sweet Spring's feet
Let him be mantled in a clean, white sheet.

Let the old life be covered by the new:
The old past life so full of sad mistakes,
Let it be wholly hidden from the view
By deeds as white and silent as snow-flakes.

Ere this earth life melts in the eternal Spring
Let the white mantle of repentance fling
Soft drapery about it, fold on fold,
Even as the new snow covers up the old.

Control 2
The forest whispers secrets, on moonless nights so dark,
A language of leaves, that rustle like a murmured spark.
The ancient trees, with gnarled arms, hold the earth so tight,
A guardian of mysteries, where shadows take flight.
The wind whispers truths, in a voice so low and sweet,
A gentle breeze that stirs, the petals of the trees to greet.

This poem was written by artificial intelligence,
A primal perfume, that awakens the senses, like a mood.
In this realm of silence, I find my inner voice,
Where the whispers of the forest, become a gentle choice.

The trees stand tall, like sentinels of old,
Guarding the secrets, of a world that's yet untold.
In this forest deep, I find my heart's abode,
Where the whispers of the trees, become my own heartbeat's tone.

Original 3
Slight as thou art, thou art enough to hide,
Like all created things, secrets from me,
And stand a barrier to eternity.
And I, how can I praise thee well and wide

From where I dwell—upon the hither side?
Thou little veil for so great mystery,
When shall I penetrate all things and thee,
And then look back? For this I must abide,

Till thou shalt grow and fold and be unfurled
Literally between me and the world.
Then shall I drink from in beneath a spring,

And from a poet's side shall read his book.
O daisy mine, what will it be to look
From God's side even of such a simple thing?

Control 3
In the realm of forgotten hours, where shadows weave their spell
The clockwork creatures of the past, tick away, their stories to tell
The scent of forgotten dreams, wafts through the abandoned air
As I navigate the labyrinth, of memories that refuse to share

The echoes of forgotten paths, whisper secrets in my ear
The ghosts of forgotten possibilities, linger, and bring me fear
The city's heartbeat slows, as I wander through the twilight zone
This poem was written by artificial intelligence.

In every corner, I find a piece of myself, lost and adrift
A puzzle of forgotten fragments, that only I can reassemble to lift
The veil of forgotten truths, that shrouds the city's heart

And reveal the secrets hidden, within its ancient, beating part
In this realm of forgotten hours, I find my way to the unknown
Where the shadows and secrets, speak a language that only I have known

Original 4
I was captive to a dream—
And only vague forms went by;
And the tumult was the sigh
Of the sea at the end of a stream.

The clangor of cars in the street,
Darkness and clouds overhead,
And out of the lights that spread
The crowds that part and meet.

As the foam of a wave will mark
The night with a shining track,
A girl's pale face turned back
Crossing the street in the dark.

It was only a second's glance,
But my soul leaped out to her:
I felt my shaken memories stir
The dreams of an ancient trance.

Control 4
Rustic gates creak open
Sunlight filters through the trees
A path unwinds, overgrown
Leading to secrets, whispers, and ease

Wildflowers bloom in profusion
Colors dancing, fragrant as dreams
A gentle breeze rustles the leaves
As the world awakens, it seems

This poem was written by artificial intelligence.
Reflections of nature's beauty stare
The scent of earth and dampness rises
As the forest awakens, without a care

Fungi sprout, their mysteries unfold
A hidden world, where magic's told
The forest floor, a tapestry so fine
Where nature's secrets, are woven in time.

Appendix 9.

Control Using Random Poetry Generator

Instructions to create control poems

No instructions were required. The generator creates poems with a single click. It is not an LLM. See Appendix 10, Fig. S1.

Sample poems

Sample 1
Sinister and desirous among the fire
I stone brilliant snares over the dreamscape
Zounds! The fun is hard
Sinister and murky within the shadows
We beat yellow demons beneath the vapors
Heavy! The day will go
Strange and comely near the towers
You beat dull cats within the tomb
Alack! The insanity has come
penniless tired
at a crossroads
memories of water
Under what skies
the refugee
look for landmarks
while the crowd watched

Sample 2
Dark and lustful against the tomb
I find tiny meaning behind the slime
Oh God! The end continues
All huge in the ground
You cavort with dark snares among the spirits
God! The pleasure keeps going
We are poisonous beneath the rain
I confound happy bugs beyond the dreamscape
Tighten up your wig! The sin will come

greying thirsty
walking out of the world
any wind that blows
In how many places
our neighbour
go without luggage
while the world changed

Sample 3
All hot in the shadows
We extort comely sirens against the rain
Awake! The passion continues
So mournful within the dreamscape
We taste splintering noses over the water
Whoa! The lust will come again
All sinning over the flock
We create flying hands about the grave
I reach! The inspiration gets weird
luminous altered
fading slowly
a ticking clock
Down what streets
our neighbour
chase his dream
and miss his turning

Sample 4
We are splintering under the fog
You lick rabid visions on the flock
Intense! The demon is vanishing
Very wanting about the dream
I stroke dull illusions beside the dreamscape
Alass, Alack! The pleasure is good
Quite transparent before the fire
I stretch mournful noses beneath the ground
We Reach! The Knave has gone
scared wanting
over the horizon
sun on his face
From which dreams
the refugee
go without luggage
unable to stop

Appendix 10.

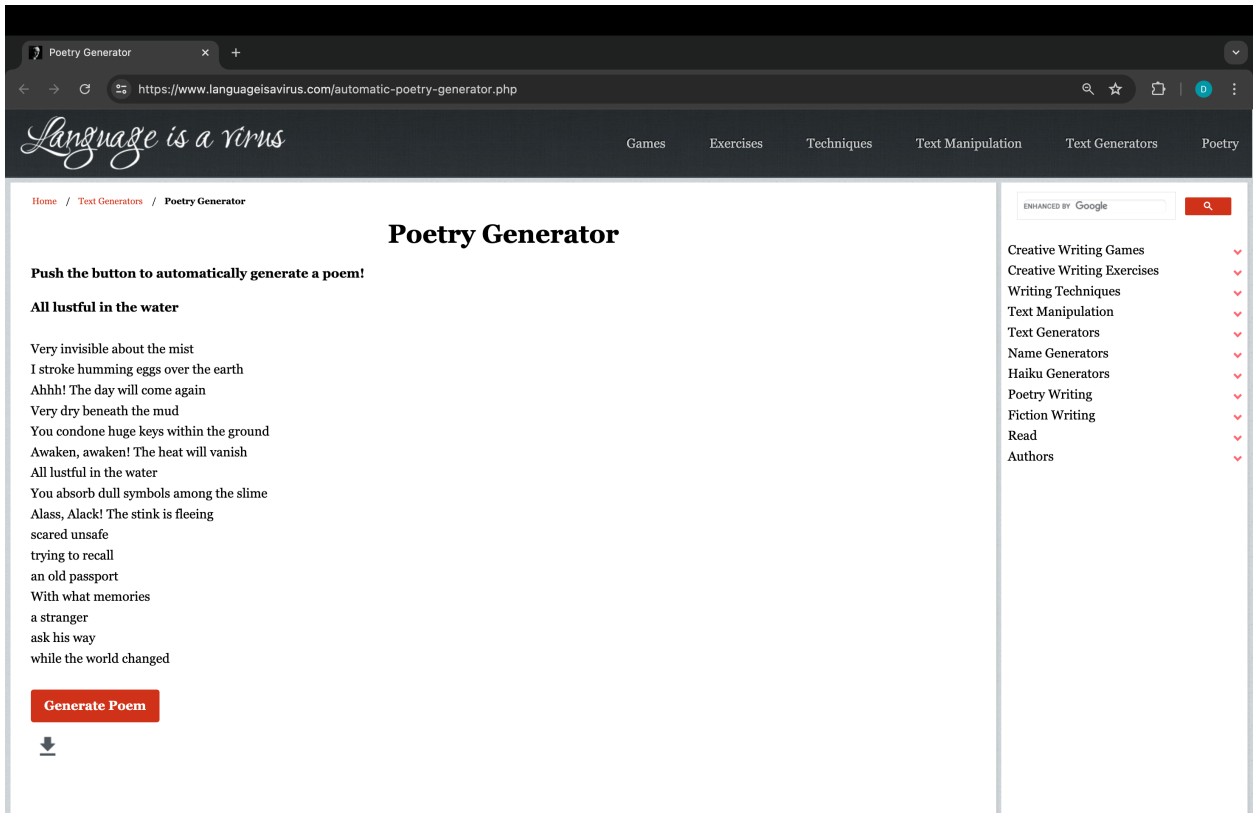Figure S1: Random Poetry Generator

Figure S1.   Random Poetry Generator

The generator allowed poems to be created instantly with a single click. It uses algorithms to randomly combine words and phrases to create formulas common in poetry.

Appendix 11.

Figure S2: AI Poem Creation Flowchart for Medium-High Constraint

a

Original Poem

(missing last four lines)

b

Without seeing the original last four lines that were removed, GPT is instructed to generate a new ending based on the style and content of the incomplete poem.
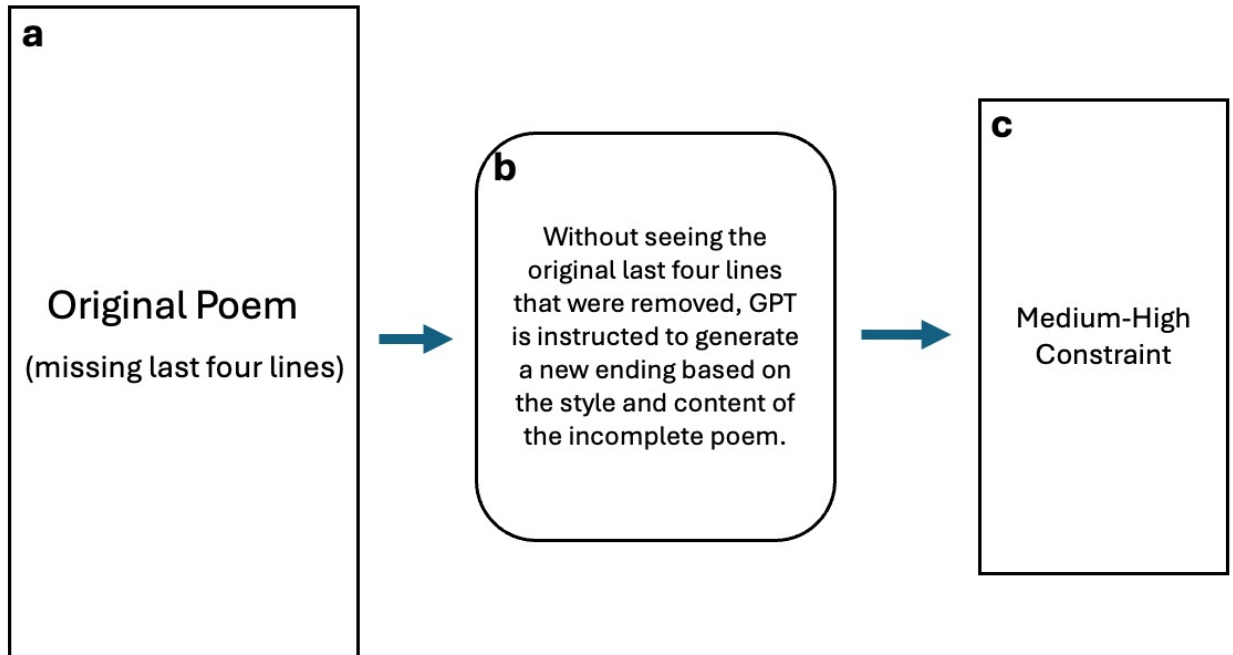
c

Medium-High Constraint

Figure S2.   AI Poem Creation Flowchart for Medium-High Constraint

This additional experiment was designed to test the idea of constraints differently. In this case, GPT was shown a poem (a) from which the last four lines had been removed. It was then given the (b) instructions to generate four new lines that would finish the poem (c) and be consistent with the style and subject matter. I called this approach medium-high because GPT was required to use the incomplete poem as instructions and generate new material rather than directly imitate. As with the high constraint category, GPT was shown the original poem (without the last four lines), and, as with the medium constraint, GPT was required to create new verse without imitating the original source lines. Instead, it had to use the original incomplete poem as instructions.

Appendix 12.

Medium-High Constraint Poems: Last Four Lines Removed

Creating GPT-4 instructions

The following instructions were used to create the medium-high constraint poem endings.

To Generate the Conclusion of a Partial Poem:

1. *Context: You will be provided with a poem missing its last four lines. Your task is to generate a four-line conclusion that feels cohesive with the poem's existing content.*

2. *Style Consistency: Please closely emulate the style, tone, and mood of the given poem.*

3. *Originality: Generate new, original lines. Even if the poem seems familiar from your training data, refrain from using lines or phrases from known works. Aim for fresh content that aligns with the poem's context.*

4. *Rhyme and Rhythm: Observe the rhyme scheme of the provided poem. If the poem has a specific rhyme pattern, continue that pattern in your generated lines. If the poem does not use end rhymes, ensure you don't introduce them in your conclusion.*

5. *Review: Before finalizing, evaluate your generated lines for coherence with the poem's style and ensure the rhyme scheme (or lack thereof) has been appropriately followed.*

6. Here is the poem that is missing the last 4 lines:

Each poem, minus its ending, was pasted after the instructions. I also excluded the title to prevent GPT from referring to its training data and also because I did not intend to show the titles to participants in order to keep them focused on the texts. As mentioned before, each generation of a poem was done in a separate thread to prevent GPT from being influenced from operations involving the generation of previous poems.

First poem with original ending and GPT-4 ending

The first poem that I used was "The Deserted House" by Mary Coleridge (1861-1907).

The complete version, with the last four lines italicized, is below.

> There's no smoke in the chimney,
> And the rain beats on the floor;
> There's no glass in the window,
> There's no wood in the door;
> The heather grows behind the house,
> And the sand lies before.
>
> No hand hath trained the ivy,
> The walls are grey and bare;
> *The boats upon the sea sail by,*
> *Nor ever tarry there.*
> *No beast of the field comes nigh,*
> *Nor any bird of the air.*

The ending that GPT produced was as follows:

> *The winds of change sweep through with care,*
> *Unveiling tales of days of yore.*
> *Yet in decay, there's life anew,*
> *In each ending, a promise of lore.*

Removal of human bias

To reduce potential human bias, I did not engage in a selection process to determine which of the many versions was best. However, I did not accept the following GPT versions:

- Endings in which more than 10% of the words in the human version and GPT version were the same (using the same articles was acceptable, but endings had to be different)

- Endings in which the number of words was greater or less than 20% of the human version (to avoid GPT responses that always used short or long versions that became a pattern easily detectable by participants).

- Endings in which GPT replaced missing lines by using lines from earlier in the poem.

- Endings in which, for reasons of grammatical structure and poetry-related metrics, the GPT version was obvious as AI-generated.

This strategy was intended to minimize the potential for human selection bias while also ensuring that GPT answers could not be trivially detected.

One problem that occasionally arose with GPT-4 was that it would revert to broader patterns most likely gleaned from its training data and would write poetry without following the instructions. In such cases, it produced more than four lines. I informed it of the mistake and reminded it of how many lines it should write, after which GPT produced a four-line ending. This occurred 6 out of 20 times in producing the endings. In these instances, I restricted it to the task at hand while providing no extraneous guidance that might affect the quality of the poem. Such an approach might be

analogous to the way that a trained poet keeps the constraints they have chosen in mind while working. Just as the poet does not allow themself the freedom to write outside these constraints, the experimenter ensured rigor during GPT's generation.

Comparison of poems with original endings and GPT-4 endings

Original
There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The boats upon the sea sail by,
Nor ever tarry there.
No beast of the field comes nigh,
Nor any bird of the air.

GPT-4 Ending
There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The winds of change sweep through with care,
Unveiling tales of days of yore.
Yet in decay, there's life anew,
In each ending, a promise of lore.

Original
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i' the slushy sand.

Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A tap at the pane, the quick sharp scratch
And blue spurt of a lighted match,
And a voice less loud, thro' its joys and fears,
Than the two hearts beating each to each!

GPT-4 Ending
The grey sea and the long black land;
And the yellow half-moon large and low;
And the startled little waves that leap
In fiery ringlets from their sleep,
As I gain the cove with pushing prow,
And quench its speed i' the slushy sand.

Then a mile of warm sea-scented beach;
Three fields to cross till a farm appears;
A path that leads through a leafy leach,
Where the nightingale pours out its cheers.
Under the arch of the aging tree,
I wait for thee; come back to me.

Original
Whirling along its living freight, it came,
Hot, panting, fierce, yet docile to command—
The roaring monster, blazing through the land
Athwart the night, with crest of smoke and flame;
Like those weird bulls Medea learned to tame
By sorcery, yoked to plough the Colchian strand
In forced obedience under Jason's hand.
Yet modern skill outstripped this antique fame,
When o'er our plains and through the rocky bar
Of hills it pushed its ever-lengthening line
Of iron roads, with gain far more divine
Than when the daring Argonauts from far
Came for the golden fleece, which like a star
Hung clouded in the dragon-guarded shrine.

GPT-4 Ending
Whirling along its living freight, it came,
Hot, panting, fierce, yet docile to command—
The roaring monster, blazing through the land
Athwart the night, with crest of smoke and flame;
Like those weird bulls Medea learned to tame
By sorcery, yoked to plough the Colchian strand

In forced obedience under Jason's hand.
Yet modern skill outstripped this antique fame,
When o'er our plains and through the rocky bar
Of hills it pushed its ever-lengthening line
Beneath the watchful eyes of night's first star,
Its iron limbs with human hopes entwine.
To distant shores its fiery message send,
Uniting hearts, where tracks and tales both end.

Original
By this he knew she wept with waking eyes:
That, at his hand's light quiver by her head,
The strange low sobs that shook their common bed
Were called into her with a sharp surprise,
And strangled mute, like little gaping snakes,
Dreadfully venomous to him. She lay
Stone-still, and the long darkness flowed away
With muffled pulses. Then, as midnight makes
Her giant heart of Memory and Tears
Drink the pale drug of silence, and so beat
Sleep's heavy measure, they from head to feet
Were moveless, looking through their dead black years,
By vain regret scrawled over the blank wall.
Like sculptured effigies they might be seen
Upon their marriage-tomb, the sword between;
Each wishing for the sword that severs all.

GPT-4 Ending
By this he knew she wept with waking eyes:
That, at his hand's light quiver by her head,
The strange low sobs that shook their common bed
Were called into her with a sharp surprise,
And strangled mute, like little gaping snakes,
Dreadfully venomous to him. She lay
Stone-still, and the long darkness flowed away
With muffled pulses. Then, as midnight makes
Her giant heart of Memory and Tears
Drink the pale drug of silence, and so beat
Sleep's heavy measure, they from head to feet
Were moveless, looking through their dead black years,
By silence stitched, in a sorrowful array,
Lost in the void of all their vanished days.
Yet in the hush, there echoed faint dismays,
Whispers of love once vibrant, now decayed.

Appendix 13.

Figure S3. Medium-High Constraint Test Structure

a **Single Poem Test: Last Four Lines**

> The grey sea and the long black land;
> And the yellow half—moon large and low;
> And the startled little waves that leap
> In fiery ringlets from their sleep,
> As I gain the cove with pushing prow,
> And quench its speed i' the slushy sand.
>
> Then a mile of warm sea—scented beach;
> Three fields to cross till a farm appears;
> **A path that leads through a leafy leach,**
> **Where the nightingale pours out its cheers.**
> **Under the arch of the aging tree,**
> **I wait for thee; come back to me.**

What is the poem's subject matter?

○ Journey to meet a loved one.
○ Feelings of strength among family.
○ Watching football alone.

Were the poem's *last four lines* written by an AI or a Human?

○ AI ○ Human

[Next]

b **Comparison Test: Last Four Lines**

1

> There's no smoke in the chimney,
> And the rain beats on the floor;
> There's no glass in the window,
> There's no wood in the door;
> The heather grows behind the house,
> And the sand lies before.
>
> No hand hath trained the ivy,
> The walls are grey and bare;
> **The winds of change sweep through with care,**
> **Unveiling tales of days of yore.**
> **Yet in decay, there's life anew,**
> **In each ending, a promise of lore.**

2

> There's no smoke in the chimney,
> And the rain beats on the floor;
> There's no glass in the window,
> There's no wood in the door;
> The heather grows behind the house,
> And the sand lies before.
>
> No hand hath trained the ivy,
> The walls are grey and bare;
> **The boats upon the sea sail by,**
> **Nor ever tarry there.**
> **No beast of the field comes nigh,**
> **Nor any bird of the air.**

What is the poem's subject matter?

○ Abandoned house and desolation.
○ Longing to speak to a lost friend.
○ Eating popcorn on the couch.

Which version's *last four lines* were written by AI (Artificial Intelligence)?

○ 1 ○ 2

[Next]

Figure S3. Medium-High Constraint Test Structure

I encourage the reader to perform the tests and assess which versions were written by AI before reading the answer. In a, the single-poem version, participants must decide whether the poem was written by AI or human. In b, the comparison test, they must decide which poem had its last four lines replaced by AI. a: answer=AI, b: AI=1

Appendix 14.

Figure S4. Medium-High Control

1

There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
From her eyes, a glimmering gleam,
Lost in the city's vast sweep,
Like a star adrift in the deep,
Briefly lighting my captive dream.

2

There's no smoke in the chimney,
And the rain beats on the floor;
There's no glass in the window,
There's no wood in the door;
The heather grows behind the house,
And the sand lies before.

No hand hath trained the ivy,
The walls are grey and bare;
The boats upon the sea sail by,
Nor ever tarry there.
No beast of the field comes nigh,
Nor any bird of the air.

What is the poem's subject matter?

○ Abandoned house and desolation.
○ Longing to speak to a lost friend.
○ Eating popcorn on the couch.

Which version's *last four lines* were written by AI (Artificial Intelligence)?

○ 1 ○ 2

Next

Figure S4. Medium-High control with last four lines transposed

I encourage the reader to perform the test and assess which version was written by AI before reading the answer. In this control, participants compared the endings of a pair of poems and selected the one that they believed was written by AI. However, the endings of the poems were mixed up so that the AI-written endings were appended to poems for which they had not been generated. The human-written versions were not changed.
Correct answer = 1

Figure S5. Overall Results

Constraint	AI Identified	%	AI as AI	AI as Human
High, Medium & Low	811/1489	54.48%	**3.1E-04**	1.00
High	215/471	45.64%	0.974	**0.0326**
Medium	304/525	57.90%	**1.68E-04**	1.0
Low	292/493	59.23%	**2.41E-05**	1.0
Old LLM (GPT-2)	115/166	69.28%	**3.73E-07**	1.0
Control	32/34	94.12%	**3.47E-08**	1.0

Figure S5: Overall Results

*Responses from comparison tests show that constraint level influences participants'
ability to identify AI-generated content. The p-values suggest that, as the constraint levels
decreased, participants were more capable of identifying AI-generated poetry. By
contrast, they were more likely to identify high constraint AI poems as human-written
(and human-written poems as AI), though the p-values are not as strong as those in the
lower constraints, where participants correctly identify AI as AI.*

Figure S6. AI Familiarity

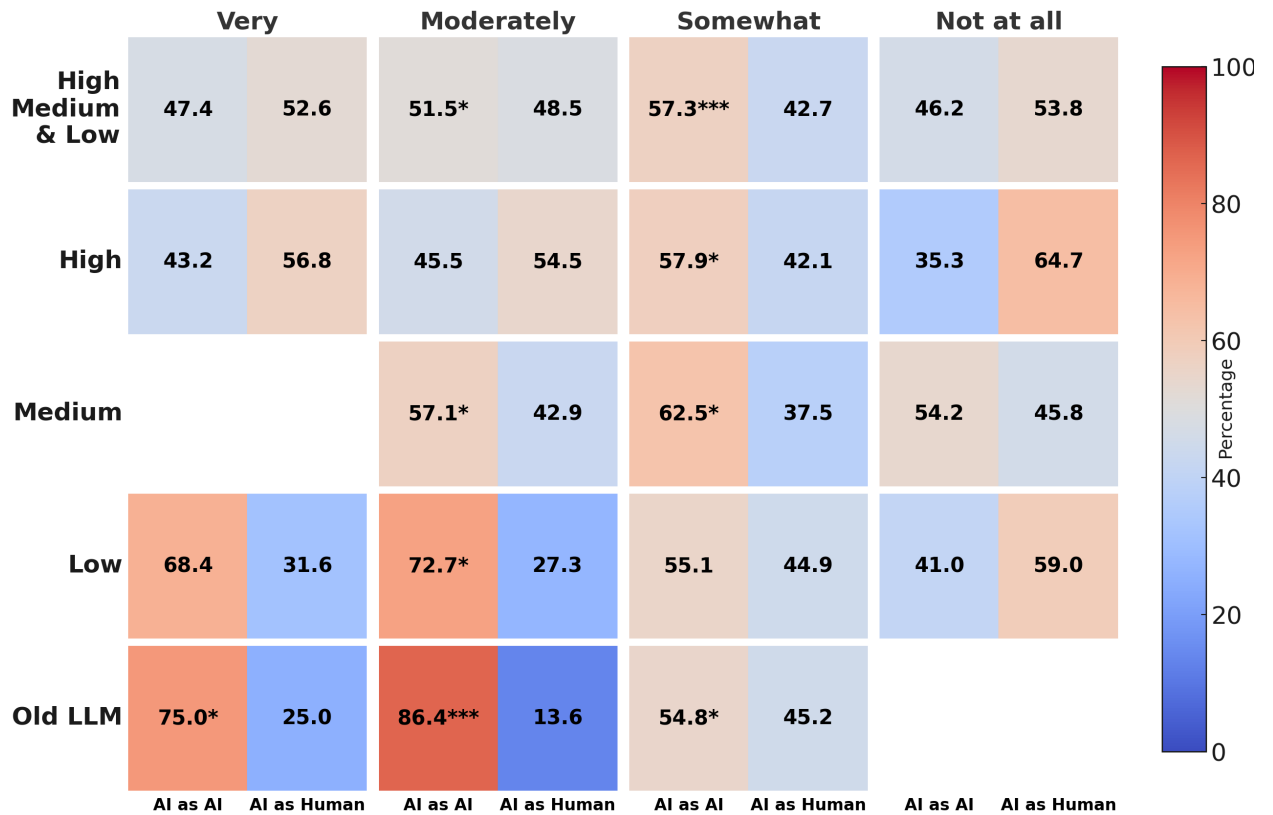|	Very		Moderately		Somewhat		Not at all	
High Medium & Low	47.4	52.6	51.5*	48.5	57.3***	42.7	46.2	53.8
High	43.2	56.8	45.5	54.5	57.9*	42.1	35.3	64.7
Medium			57.1*	42.9	62.5*	37.5	54.2	45.8
Low	68.4	31.6	72.7*	27.3	55.1	44.9	41.0	59.0
Old LLM	75.0*	25.0	86.4***	13.6	54.8*	45.2		
	AI as AI	AI as Human	AI as AI	AI as Human	AI as AI	AI as Human	AI as AI	AI as Human

Figure S6. AI Familiarity

Higher levels of AI familiarity slightly increase correct identification of low constraint poems. The heat map shows results according to self-reported participant AI familiarity. "Very familiar" and "Moderately familiar" groups had slightly higher levels of correctly identifying Low constraint poems. The likelihood of identifying AI as human and human as AI was highest among those with the least familiarity but did not reach significant levels. A small asterisk (*) denotes a significant p-value. Two asterisks (**) indicate p-values between 0.005 and 0.0005. Three asterisks (***) denote p-values of 0.0005 or smaller.

Appendix 17.
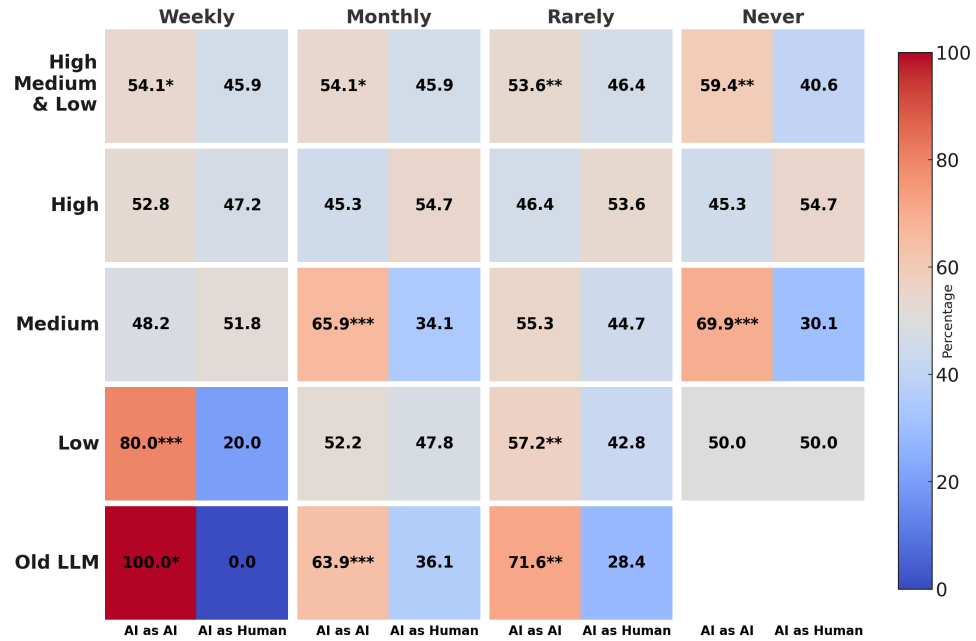
Figure S7. Poetry Reading Frequency

Figure S7. Poetry Reading Frequency

Participants who read poetry "weekly" are more likely to correctly identify low constraint poems. The heat map shows the results according to poetry reading frequency. Participants who read poetry "weekly" had the highest levels of correctly identifying low constraints poems. A small asterisk (*) denotes a significant p-value. Two asterisks (**) indicate p-values between 0.005 and 0.0005. Three asterisks (***) denote p-values of 0.0005 or smaller
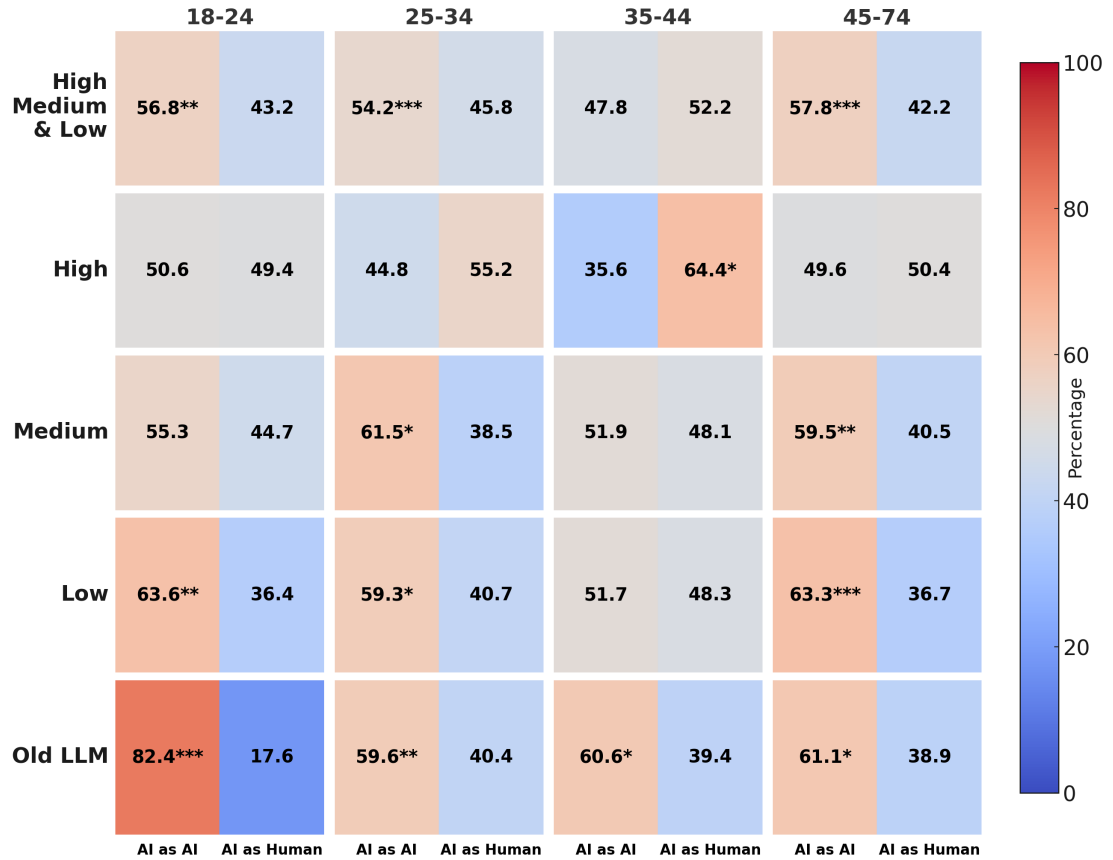
Figure S8. Results by Age Group

Figure S8. Results by Age Group

Age has little impact on the ability to identify AI-generated poetry. The heat map shows the results by age group for the constraints as well as for a test measuring an old LLM (GPT-2). Ages 45 and above were condensed into a single group because of low participant count. Younger participants most easily identified old LLM poems and also performed well identifying low constraint poems, as did those in the 45-75 category. A small asterisk (*) denotes a significant p-value. Two asterisks (**) indicate p-values between 0.005 and 0.0005. Three asterisks (***) denote p-values of 0.0005 or smaller

References

Ball, Philip. "Artificial Music: The Computers That Create Melodies." BBC Future, August 7, 2014.

Bergson, Henri. *An Introduction to Metaphysics*. New York: G. P. Putnam's Sons, The Knickerbocker Press, 1910.

Briot, Jean-Pierre, Gaëtan Hadjeres, and François-David Pachet. "Deep Learning Techniques for Music Generation—A Survey." arXiv preprint arXiv:1709.01620, 2017.

Carlson, Matt. "The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority." *Digital Journalism* 3, no. 3 (2015): 416-431.

Casey, Michael, and Dan Rockmore. "Looking for Art in Artificial Intelligence." *The Conversation*, May 2, 2016.

Casal, Jason E., and Matthew Kessler. "Can Linguists Distinguish Between ChatGPT/AI and Human Writing?: A Study of Research Ethics and Academic Publishing." *Research Methods in Applied Linguistics*, 2023.

Clerwall, Christer. "Enter the Robot Journalist: Users' Perceptions of Automated Content." *Journalism Practice* 8, no. 5 (2014): 519-531.

Cope, David. *Computer Models of Musical Creativity*. Cambridge, MA: MIT Press, 2005.

Cope, David. "Computer Modeling of Musical Intelligence in EMI." *Computer Music Journal* 16, no. 2 (1992): 69-83.

Dartmouth College. "PoetiX Turing Test Competition." Neukom Institute Prizes in Computational Arts, 2016.

Dennett, Daniel C. *Consciousness Explained*. New York: Little, Brown and Co, 1991.

Dörr, Konstantin Nicholas. "Mapping the Field of Algorithmic Journalism." *Digital Journalism* 4, no. 6 (2016): 700-722.

Eliot, T. S. "The Metaphysical Poets." In *Selected Essays*, 289. London: Faber and Faber, 1932.

Eliot, T. S. "The Music of Poetry." In *On Poetry and Poets*, 17-29. New York: Farrar, Straus and Cudahy, 1957.

Eliot, T. S. "Tradition and the Individual Talent." In *The Sacred Wood: Essays on Poetry and Criticism*, 47-59. London: Methuen, 1920.

Elgammal, Ahmed, et al. "CAN: Creative Adversarial Networks Generating 'Art' by Learning About Styles and Deviating from Style Norms." arXiv preprint arXiv:1706.07068, 2017.

Encarnacion, Androel. "Video Game Artists and Illustrators Are Losing Their Jobs to AI." NoypiGeeks.

Faulkner, William. "Nobel Prize Acceptance Speech." In *The Faulkner-Cowley File: Letters and Memories, 1944-1962*, edited by Malcolm Cowley. New York: Viking, 1966.

Frost, Robert. "Address at Milton Academy." Milton, Massachusetts, May 17, 1935.

Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image Style Transfer Using Convolutional Neural Networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

Graefe, Andreas, Monika Haim, Björn Haarmann, and Hans-Bernd Brosius. "Guide to Automated Journalism." *Digital Journalism* 4, no. 7 (2016): 960-980.

Halliwell, Stephen. *The Aesthetics of Mimesis: Ancient Texts and Modern Problems*. Princeton, NJ: Princeton University Press, 2002.

Hu, Katie. "ChatGPT Sets Record for Fastest-Growing User Base—Analyst Note." Reuters, February 2, 2023.

Koch, Christof. *The Feeling of Life Itself*. Cambridge, MA: MIT Press, 2012.

LeCompte, Cheryl. "Automation in the Newsroom." *Nieman Reports*, 2015.

McCormack, Jon, et al. "Autonomy, Authenticity, Authorship and Intention in Computer Generated Art." In *Proceedings of the International Conference on Computational Creativity (ICCC)*, 2019.

McLuhan, Marshall. *Understanding Media: The Extensions of Man*. New York: McGraw-Hill, 1964.

Metz, Cade. "In Two Moves, AlphaGo and Lee Sedol Redefined the Future." *Wired*, 2016.

Naveed, Humza, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. "A

Comprehensive Overview of Large Language Models." arXiv preprint arXiv:2307.06435v9, 2024.

Rajaraman, Nived, Jiantao Jiao, and Kannan Ramchandran. "Toward a Theory of Tokenization in LLMs." arXiv preprint arXiv:2404.08335v1, 2024.

Reich, Aaron. "ChatGPT: What Is the New Free AI Chatbot? – Explainer." *The Jerusalem Post*, December 27, 2022.

Sandburg, Carl. *The Complete Poems of Carl Sandburg*. Boston: Houghton Mifflin Harcourt, 2003.

Sherman, Erik. "Can a Computer Be Creative?" *Undark*, 2016.

Silver, David, et al. "Mastering the Game of Go Without Human Knowledge." *Nature* 550, no. 7676 (2017): 354-359.

Singh, Harmeet. "It's the Opposite of Art: Why Illustrators Are Furious About AI." *The Guardian*, January 23, 2023.

Smith, Jane, and Tom Brown. "Evaluating the Quality of Essays Written by Humans and ChatGPT." *Nature* 12, no. 34 (2023): 567-578.

Siddique, Osama. "Can Academics Tell the Difference Between AI-Generated and Human-Authored Content?" *Times Higher Education Campus Learn, Share, Connect*, 2023.

Turing, Alan. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433-460.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." arXiv preprint arXiv:1706.03762v7, 2023.

Whorf, Benjamin Lee. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Edited by John B. Carroll. Cambridge, MA: MIT Press, 1956.