



OPEN Text comprehension moderates cognitive biases toward AI authorship in literary content

S. Raglio^{1,7}, F. Miriani^{1,7}, M. Focareta¹, S. Rebora², M. Rospocher², S. G. Chiarella^{1,3}, D. M. Gagliardi^{1,4}✉, M. Salgaro²✉ & G. Torromino^{1,5,6}✉

The growing potential of artificial intelligence (AI) in generating artistic products has stimulated an increasing multidisciplinary debate. Beyond discussions on the nature of AI creativity itself, numerous studies report that people are negatively biased towards AI-labelled artistic contents. Here we asked whether this bias is modulated by beholders' approach to the artifact, by exploring whether and how the level of comprehension—an index of participant's engagement—could influence this bias in literary context. We asked participants to rate the social prestige and pleasantness of three texts from a reckon writer that were presented as Human-, AI- and Human + AI-made. We discovered that the comprehension level emerges as a key, bidirectional modulator of evaluative judgments. Indeed, we observed that only participants that resulted in a low level of text comprehension showed a negative bias against AI-labelled texts compared to Human- and Human + AI-labelled ones, while those that demonstrated high comprehension showed a positive bias towards AI. Our results suggest that bias against AI is not intrinsic to the content but is driven by preconceptions that can be mitigated or exacerbated by the level of understanding achieved.

Keywords Artificial intelligence (AI), Cognitive bias, Literacy text, Social prestige, Pleasantness, Empirical aesthetics

Generative technologies and literature

Nowadays artificial intelligence (AI) is being intermingled with our daily lives, with contribution that encompass a wide range of activities including creative production in literature. In fact, the relationship between technology and literary creation isn't exclusive to the digital age, as text generators are as old as language itself^f. Since long times, available tools and methodologies have influenced the production of stochastic materials, giving rise to mechanisms capable of generating musical, textual, mathematical, and cognitive content.

We live in an age in which texts are increasingly circulating whose authorship is unclear, as it cannot be determined whether they originate from a human being or a machine. Currently, AI systems are capable of not only writing entire novels (e.g.,^{2,3}), but also poems (e.g.,^{4–7}). Thus, similar to philosophy and the humanities as a whole, literature has also entered the post-human age. Studies have also shown that non-experts, but even experts, cannot determine with certainty whether a text identified as literary was written by a human being or, for example, by ChatGPT (Porter and Machery, 2024)⁴³ (Gunser 2021)³⁵. In the literary context, poetry generation represents an emblematic field of this evolution, originating in the mid-last century and now an integral part of Computational Creativity (Linardaki 2022)³⁷. Similar to image generation, this process is based on training artificial intelligence systems with vast textual *corpora*, giving them the ability to produce new poetic compositions. Already with the first computers in the second half of the twentieth century, applications of this nature were found in the field of art and literature, such as Christopher Strachey's Love Letter Algorithm (1952), which produced love letters via the Ferranti Mark 1, and Theo Lutz's Stochastic Texts (1959), generated with a Zuse Z22 computer. Similarly, Nanni Balestrini used an IBM computer for his Tape Mark I (1961), creating random combinations for poetic texts. This line of experimentation has thrived in subsequent years, leading to the founding of initiatives like the *Atelier de Littérature Assistée par la Mathématique et les Ordinateurs* (ALAMO)

¹Número Cromatico, Art and Neuroscience Collective, P.le del Verano 93, 00185 Rome, Italy. ²Department of Foreign Languages and Literatures, University of Verona, Lungadige Porta Vittoria 41, 37129 Verona, Italy. ³Department of International Humanities and Social Sciences, UNINT, Via Cristoforo Colombo 200, 00147 Rome, Italy. ⁴Department of Communication and Graphic Design, NABA, Via Ostiense, 92, 00154 Rome, Italy. ⁵Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy. ⁶University of Naples Federico II, via di Porta di Massa 1, 80133 Naples, Italy. ⁷S. Raglio and F. Miriani contributed equally to this work. ✉email: dionigimattiagagliardi@gmail.com; massimo.salgaro@univr.it; giulia.torromino@uniroma1.it

by Paul Braffort and Jacques Roubaud in the early 1980s. The advent of laptops, digital image processing software, and the internet between the 1990s and early 2000s marked a further turning point, intensifying the relationship between literature and new technologies (Davinio 2002; Morris and Swiss 2006³⁹; O'Sullivan 2019)⁴¹. This interaction gave rise to neologisms such as “technopoetry” (Davinio 2002), “technotexts” (Morris and Swiss 2006), and “electronic literature” (O'Sullivan 2019), referring to artistic and literary productions that leverage new technologies.

Today, AI-based generative technologies, particularly large language models (LLMs), represent the pinnacle of this evolution, offering unprecedented capabilities in the production of original and complex texts (Mitchell 2019)³⁸. This progression raises fundamental questions regarding the nature of authorship, the very definition of creativity, and the boundaries between human and artificial production. To these more conceptual issues, it adds also how the readers perceive such a literature and how knowing that a text has been produced by or with and AI affects the text perception.

The aesthetic experience and the bias towards AI

For what it concerns the aesthetic experience, besides the authoriality of an artistic product, literature from history of art and neuroscience has so far acknowledged that beauty and aesthetic appreciation are not absolute values, but concepts affected by individual, cultural, and contextual factors (e.g., Nadal and Chatterjee 2019⁴⁰; Che et al. 2018³³; Pearce et al. 2016;⁴² Gallese and Di Dio 2012³⁴; Jacobsen 2006)³⁶. For instance, when coping with text reading, several subjective and objective factors affect the reader's response in terms of perceived pleasantness or aesthetic judgement of that text, such as: how much we know about the subject, whether we already know the author, the way the text is introduced, and so on. All these factors might give rise to biases in the reader's response.

In a well-known experiment, Zwaan⁸ showed that subjects read a text differently if it is presented as literary text rather than as non-fiction. Results showed that participants reading under a literary perspective had longer reading times, better memory for surface information, and a poorer memory for situational information than those reading under a news perspective. Therefore, factors such as individual history, context, contingent stimuli, and given information, have been recognized as traits that directly affect the codification and processing of an artistic product, regardless of its objective characteristics. Similarly, they can influence the viewers' aesthetic experience, judgements, and evaluations, partially impacting on the fate of the artistic product, be it a literary text, an artwork or a musical composition.

This last statement is so true when we are dealing with the negative bias toward literary texts produced with AI, which represent a hot topic in today's discussion about art and its production, because it offers a good chance to unveil the perceptual, cognitive and emotional processes that we activate experiencing an artwork.

Nowadays, several AI systems can be employed to generate texts, images, or videos that can then be implemented into the artistic production, and that are easy to be accessed by users from various backgrounds, and not only by experts. Many current AI systems are capable of producing artifacts that people often judge as indistinguishable from human creations^{9–13}, highlighting the rapid progress and sophistication of generative AI. Indeed, the current advances in the technology of AI systems made it possible to achieve some kind of autonomy in the creation of different types of text^{10,14,15}. The autonomous generation of original texts by AI raises a new big challenge that elicited a debate about the theme, which is: are AI systems able to create original content either being useful or appealing for the audience it refers to?

There is a growing set of experimental studies that recently investigated how *a priori* manipulated authoriality affected human appreciation of artistic stimuli^{10,14–16}, which mostly point to readers' failure in distinguish an AI-made text from a human-made and to a negative bias against AI in terms of aesthetic quality, pleasantness, or novelty and originality attribution. At the same time, it has been stated that such a growth directly affects the aesthetic experience of an artistic product but negatively, giving rise to negative bias against the AI-generated product. In this phenomenon, factors such as personality traits, individual aspects and attitudes with technology have been reported to affect the perceptual process of artworks^{17,18}.

A recent study by Demir et al.¹⁹ confirms this trend towards AI-produced literary texts by investigating “algorithm aversion” in readers' evaluations of creative writing. Algorithm aversion is defined as the tendency to prefer human over algorithmic input (p. 2), the study assessed reader perceptions of a short story labelled under four authorship conditions: (a) human-authored, (b) AI-authored, (c) human-AI collaboration (human-generated idea, AI-developed story), and (d) AI-human collaboration (AI-generated idea, human-developed story). Results showed a consistent bias: creativity ratings were significantly lower when AI involvement was disclosed, regardless of the degree of human contribution. Notably, participants with high creative potential—as measured by the Divergent Association Task (DAT)—demonstrated less susceptibility to this bias, suggesting that openness to AI as a creative partner may correlate with individual creative disposition. The DAT calculates the semantic distance between the word pairs produced by the subject. As semantic distance is also a feature of many rhetorical figures and literariness²⁰ it can be an indicator of an adequate reading capacity of literary texts.

Aim of the study

In this study, we sought to explore the emergence of biases toward AI when reading fairytales, focusing on how readers' comprehension levels influence their evaluations. Building on previous findings (e.g.,¹⁹), we hypothesized that reading comprehension—used as an indicator of proficiency in understanding literary texts—could modulate bias toward AI-labelled works. Participants were asked to evaluate each text in terms of both social value and pleasantness, allowing us to assess whether these attributions elicited different evaluative biases or mitigated the commonly reported negative bias toward AI.

Our general aim was to test how readers' attitudes change toward well-written, recognizable texts when the perceived authorial identity is artificially altered to involve AI. Unlike previous studies, we aimed to ensure

greater ecological validity by employing authentic literary material from a recognized literary author. To this aim we chose Roald Dahl's fairytales, whose distinctive, irreverent, and often controversial style has profoundly shaped the reading experiences of millions of children worldwide. In choosing Dahl's fairytales we guaranteed both a high-level literary quality and an accessible and engaging prose across the text stimuli.

Moreover, since being "made by AI" might differ from "made with AI", we also tested the readers' reaction to texts that were declared to be made by a human author aided by AI. In doing so we manipulated the authorship assignment of three Dahl's fairytales extracts by attributing one of the following conditions: i) human author (Human), ii) AI, and iii) human author assisted by AI (Human + AI). We decided for a pseudo-randomised design in which three fairytales were randomly associated to one of the three assignment conditions, but the presentation order was either Human/AI/Human + AI or AI/Human/Human + AI.

Results

AI-authorship affects the evaluation of the social prestige of a text

To explore how a text is evaluated on the aesthetic and social account when AI is the author, we asked participants to score the social and literary value of the fairytales using a specifically developed questionnaire to test the social prestige in literature²¹. Each of the questions investigates a different aspect of the value of a text: the first two concern technical appreciation of a text together with its social recognition, the third concerns the social value in terms of utility and the last one inquiries about a more subjective opinion about the relevance of the text.

We first compared the answers to this questionnaire only considering the author assignment factor (Human vs. AI vs. Human + AI), to verify whether the author-assignment is sufficient to modulate the social prestige evaluation of a text. The first three questions showed a significant group difference between the three author-assignment conditions (Fig. 1). Specifically, participants were significantly more prone to affirm that the human-assigned fairytale won a literary award compared to the AI- and Human + AI-assigned fairytales (Fig. 1A, Table 1), showing a negative bias towards the AI contribution when it concerns the literary value of a text. Interestingly, this was not the case when rating the worthiness of the fairytales to be taught at school (Fig. 1B, Table 1), as both the Human and the AI assigned fairytales were scored significantly more than the Human + AI assigned one for this question.

When asked to take the perspective of literary critics, participants rated as less valuable only the Human + AI condition compared to the Human one (Fig. 1C, Table 1), which suggest that this conjoint creation of a text is to be considered less appreciable in terms of aesthetic value.

Finally, we found no significant differences between the assignment conditions when we asked about the relevance of the fairytales (Fig. 1D, Table 1—note the opposite valence of the fourth question), although the

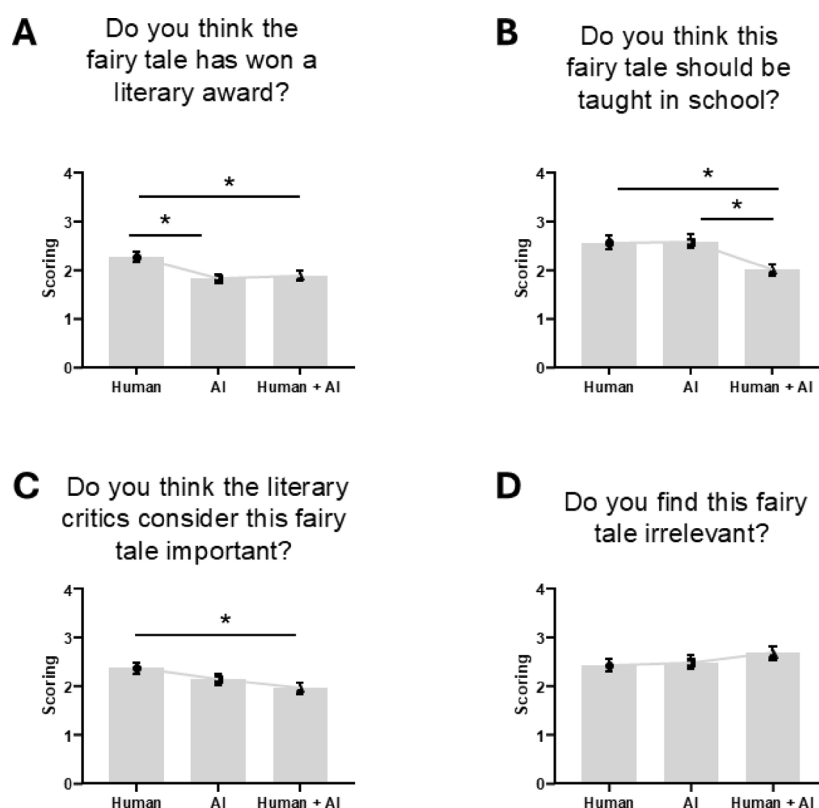


Fig. 1. Barplots show the average score $\pm$ SEM of the social prestige questionnaire (four questions, from A–D) for each author assignment condition. * $p < 0.05$ between groups (LSD post-hoc analysis).

Question	F	df	p
Literary award	8.529	2, 206	*0.000278
Taught at school	5.661	2, 206	*0.004042
Critics importance	3.739	2, 202	*0.025398
Irrelevant	0.882	2, 206	0.415543

Table 1. Statistical analyses for the questionnaire on the social prestige of a text. Results from the one-way ANOVA for repeated measures for author assignment (three levels: Human, AI, Human + AI) are shown. Data refers to the four questions shown in Fig. 1. LSD post-hoc analysis results are reported in Fig. 1. Significant values are in [bold].

trend of the participants’ scoring follows the results of the previous question which inquired about the relevance of the fairytales from the perspective of the literary critics (Fig. 1C).

Since in a previous similar experiment we conducted using abstract painting we had found an effect of order presentation for the emergence of a bias towards AI¹⁷, we analysed also here whether the presentation order in our pseudo-balanced design could affect participants’ ratings on the social prestige of the fairytales, but we found no relevant effects apart for the second question of the questionnaire (Suppl. Table 1).

So far, our results indicate that when readers are required to rate the social value of a text, knowing that the author has being an AI or that a human author was aided by an AI can modulate their response, although differently based on the specific dimension to evaluate.

Text comprehension drives opposite biases towards AI when evaluating the social prestige of a text

Although less investigated in literature when concerning biases, an important factor that might affect our capability of evaluating or appreciating a text, could be the level of comprehension. We reasoned that bias is a prejudice that can be overcome when the subject makes a judgement based on knowledge and information. Thus, we hypothesized that the level of comprehension of a text could affect its appreciation and evaluation on the social (social prestige) and aesthetic level (pleasantness), and thus give rise to different bias towards texts declared to be made by an AI. To test this hypothesis, we analysed the effects of the interaction between the authorship assignment—Human, AI and Human + AI—and the comprehension level of three fairytales first on their social prestige (Fig. 2). The comprehension level was evaluated with a score (comprehension score, CSc) that allowed to divide participants in two categories, High and Low CSc, reflecting respectively high and low levels of text comprehension: those that have responded correctly to at least 3 out of 5 comprehension questions in each of the three fairytales were put in the High CSc, the others in the Low CSc (see “Methods” Section).

Interestingly, although when considering the whole sample of participants we found an effect for author assignment (Fig. 1), dividing participants in High and Low CSc groups gave rise to a dissociation in the bias towards AI in 3 out of the 4 questions of the social prestige questionnaire (Fig. 2, Table 2), suggesting that for the social prestige dimension the level of comprehension of a text is an important factor for the judgement. Indeed, we found a significantly higher evaluation towards AI for the High CSc group. Specifically:

1. No differences were ever observed in the Human-assigned condition between the two CSc groups for all questions (Fig. 2, Table 2).
2. No significant differences between groups were observed for the first question (*Do you think this fairytale has won a literary award?*) (Fig. 2A, Table 2).
3. For the second question (*Do you think this fairytale should be taught in school?*), a positive bias towards AI emerged for the High CSc group with respect to both Human- and Human + AI-assigned fairytales, together with a significant difference between groups for the AI- and the Human + AI-assigned texts (Fig. 2B, Table 2). Moreover, the Low CSc group rated the AI-text significantly less than the Human one (Fig. 2B).
4. For the third question (*Do you think the literary critics consider this fairytale important?*), we found a negative bias against AI with respect to the Human-assigned fairytale for the Low CSc group and a positive bias towards AI compared with the Human + AI-assigned fairytales for the High CSc group (Fig. 2C, Table 2).
5. For the fourth question (*Do you find this fairytale irrelevant?*) we found a dissociation of rating on the AI-assigned text between High and low CSs groups (Fig. 2D, Table 2). Moreover, a positive bias towards AI with respect to Human + AI-assigned fairytale in the High CSc group (Fig. 2D, Table 2). Noteworthy, as before, for this question the effect is inverted due to its negative valence.

As a control, we also considered the median split of the average of the scores of the three fairytales (Median = 3.66) as an alternative dichotomization criterion to divide the sample into High and Low CSc (Suppl. Table 2). We found the same statistical trend between the two groups, confirming the validity of the first, more restrictive criterion, which takes into consideration the simple and accessible nature of the chosen texts (fairytales by Roald Dahl). A Spearman’s correlation analysis between the average CSc and the questions scores was also run leading to the same results (Suppl. Fig. 2A).

To exclude that the observed effects were not due *per se* to a modulation of the author assignment on the CSc, we compared the CSs across the author-assigned conditions but we found no significant differences between Human, AI, Human + AI (Fig. 3A) (one-way ANOVA for repeated measures: $F(2,206) = 2.948$, $p = 0.055$). Interestingly, although no significant differences were found in the CSs between author-assignment conditions,

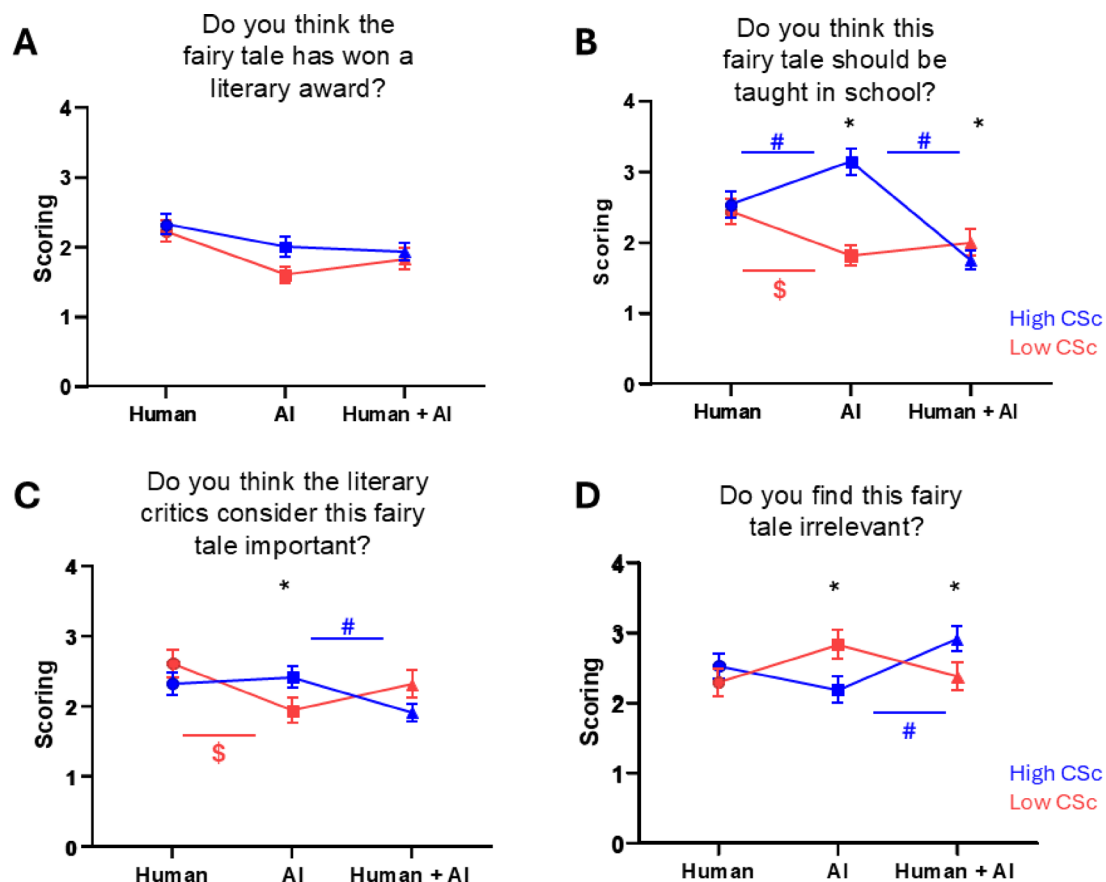


Fig. 2. Graphs represent the average ratings $\pm$ SEM of the four questions (A–D) of the social prestige questionnaire for each author assignment condition and for the High and Low comprehension score (CSc) groups. * $p < 0.05$ between groups. # $p < 0.05$ within High CSc. \$ $p < 0.05$ within Low CSc (LSD post-hoc analysis).

Question	Effect type	F	df	p
Literary award	Assignment	8.755	2, 200	*0.000227
	Comprehension	2.096	1, 100	0.150822
	Interaction	0.997	2, 200	0.370658
Taught at school	Assignment	5.322	2, 204	*0.005585
	Comprehension	1.9	1, 102	0.171044
	Interaction	12.074	2, 204	*0.000011
Critics importance	Assignment	3.889	2, 204	*0.022002
	Comprehension	0.983	1, 102	0.323771
	Interaction	3.464	2, 204	*0.033146
Irrelevant	Assignment	0.779	2, 204	0.460428
	Comprehension	0.108	1, 102	0.742724
	Interaction	15.201	2, 204	*0.006267

Table 2. Statistical analyses for the questionnaire on the social prestige of a text. Results from the two-way ANOVA for repeated measures for author assignment (three levels: Human, AI, Human + AI) and Comprehension scores (two levels: low, high) are shown. Data refers to the four questions shown in Fig. 2. LSD post-hoc analysis results are reported in Fig. 2. Significant values are in [bold].

it appears clear that the AI-assignment condition elicited more variable readers' responses compared to the Human and Human + AI conditions (Fig. 3B). Instead, as expected, we observed a significant general decrease in the CSs across the task from the first to the last presented fairytale (Fig. 3C) (one-way ANOVA for repeated measures: $F(2,206) = 4.109$, $p = 0.018$), suggesting a fatigue effect, but this was not modulated by the author-

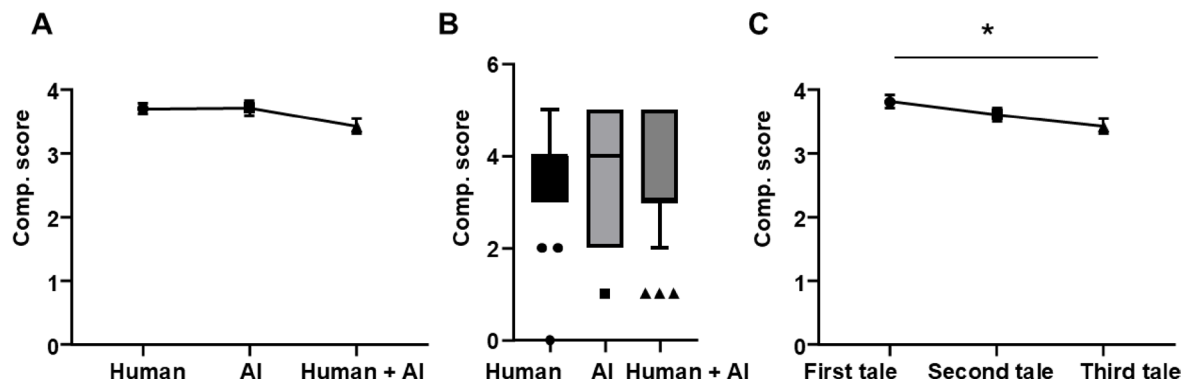


Fig. 3. Comprehension score across conditions. **(A)** Graph represents the average $\pm$ SEM of comprehension scores (CSc) across participants depending on fairytale assignment (Human, AI, Human + AI). **(B)** Boxplot of CSc across participants depending on assignment. **(C)** Graph represents the average $\pm$ SEM of CSc across participants depending on fairytale presentation order. * $p < 0.01$ (LSD post-hoc analysis).

assignment, because—as aforementioned—there were no significant differences between CSc in the three different conditions (Fig. 3A).

We also checked whether age (correlation $c = 0.0170$, $p > 0.05$), or sex (CSc females = 3.60, CSc males = 3.64, $t = -0.3276$, $p = 0.7439$) could affect the CSc, finding no significant effects.

Finally, we checked whether there were differences in the CSs between groups depending on the average of books read by the participants (Suppl. Table 3), but we found no differences, thus excluding a possible effect of the reading habits on the observed effects.

Text comprehension also drives opposite biases toward AI-attribution in the dimensions of text pleasantness

We have previously shown that the level of comprehension of a text together with knowing that the text has been made by an AI can affect the reader's reaction in terms of perceived social prestige (Fig. 2). Thus, here we wanted to explore whether similar results were also true for text appreciation. Text pleasantness was evaluated across participant's scoring through nine dimensions (Fig. 4, see “Methods” Section). We found that in six out of nine dimensions, the evaluation was significantly different between High and Low CSc groups for the AI-assigned text (Table 3). Specifically, we observed a between group dissociation for the AI-assigned text: the Low CSc group scored the AI-assigned text significantly less compared to the High CSc group for five dimensions (*Nice*, *Engaging*, *Interesting*, *Compelling*, and *Pleasant*). Coherently, an opposite pattern of dissociation was observed for the adjective *Boring*, which was rated significantly more by the Low CSc group compared to the High CSc group, consistent with the negative valence of this dimension.

Moreover, within group comparisons showed a significantly higher scoring for the AI-assigned text compared to both Human and Human + AI assignment in the High CSc group (*Nice*, *Engaging*, *Interesting*, *Pleasant*), or only to Human + AI (*Compelling*), resulting in a positive bias towards AI assignment in this group. Coherently, the scoring in the High CSc group was significantly lower for AI text compared to Human and Human + AI assignment for the adjective *Boring*, which has a negative valence. An opposite effect was observed for the Low CSc group, which shows a negative bias towards AI assignment compared with both Human and Human + AI (*Nice*, *Engaging*), or just Human (*Interesting*, *Compelling*). No effects were found for the adjectives *Romantic*, *Exciting* and *Fun*, which is in line with the nature of the chosen texts. This trend of results was also confirmed using the median split as dichotomization criterion (Suppl. Table 4) and by the results of the Spearman's correlation analysis (Suppl. Fig. 2B).

Notably, no effect was observed when all participants were grouped together, except for the adjective *Pleasant* (Suppl. Table 5). Furthermore, the presentation order of the fairytales (i.e. the fact that AI was evaluated as first or as second in comparison to the Human assignment)—which was found to be a relevant factor in the evaluation of visual art stimuli¹⁷—did not result in significant effects, except for two adjectives (*Interesting*, *Compelling*) (Suppl. Table 6). This suggests that, in general in the case of a reading task the order of presentation is less important as a factor to modulate a bias towards AI, but that depending on the specific dimensions to be scored and the presented artworks, participants do or do not use a direct comparison between stimuli with different assigned authors to make their evaluations.

Controlling for differences across the three fairytales did not reveal significant effects and did not alter the overall pattern of results (Suppl. Table 7).

Finally, it is noteworthy that participants tend to evaluate texts that are declared to be the result of a conjoint work between Human and AI as similar to those declared to be made by a Human writer alone. In fact, when a significant difference in evaluation emerges, it is always between Human + AI and AI (*Nice*, *Engaging*, *Interesting*, *Boring*, *Compelling*, *Pleasant*), and never between Human + AI and Human.

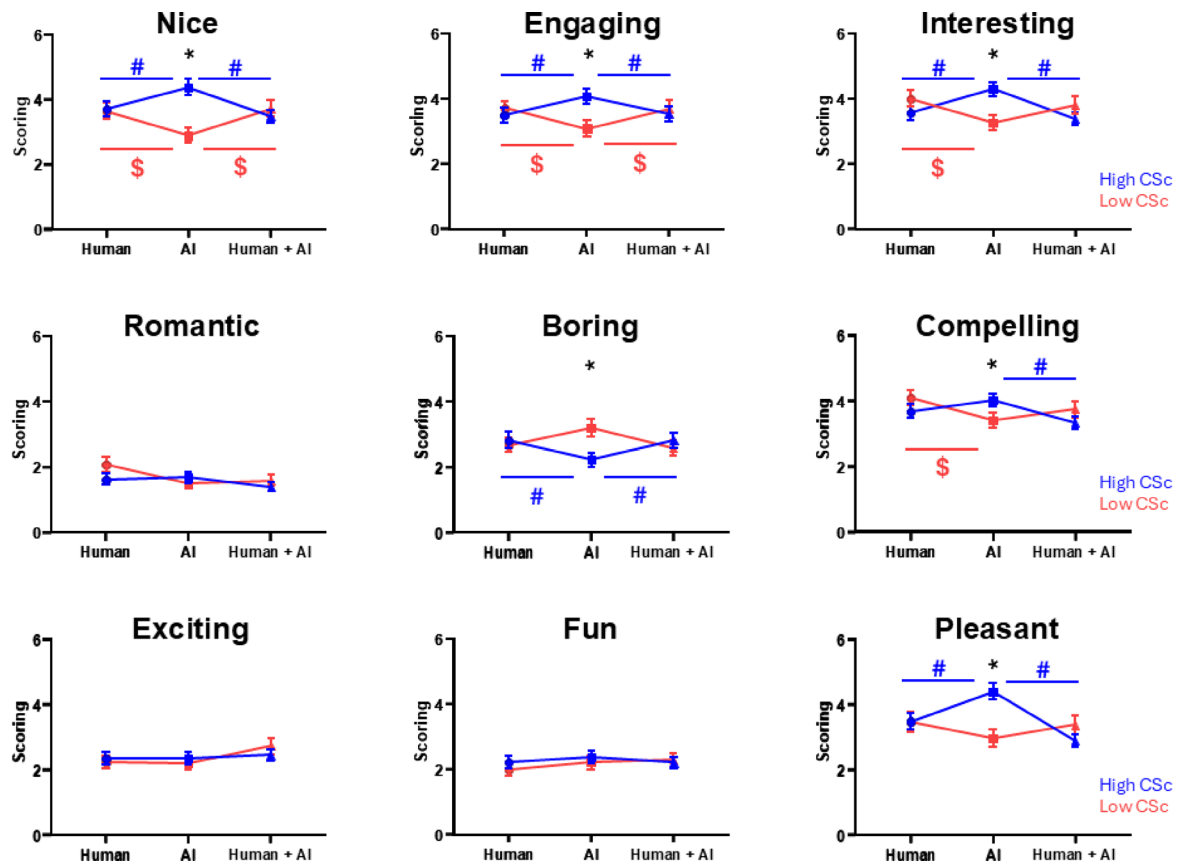


Fig. 4. Ratings of the nine dimensions of text pleasantness for each author assignment condition and for the High and Low comprehension score (CSc) groups. * $p < 0.05$ between groups. # $p < 0.05$ within High CSc. \$ $p < 0.05$ within Low CSc (LSD post-hoc analysis).

Discussion

Our findings reveal a novel and compelling aspect of authorship bias in the perception of creative works: comprehension emerges as a key modulator of evaluative judgments, particularly toward AI-generated content. This bias, crucially, appears bidirectional, leading to both negative and positive deviations in participants' ratings depending on their comprehension level. To our knowledge, such dual modulation based on comprehension has not been previously reported.

In general, our data confirm the tendency to attribute more creativity, in terms of literary value and social prestige to texts written by humans (Fig. 1), which is in line with other reports¹⁹. However, in addition to this, we found that: while low-comprehension participants tend to heavily discount works attributed to AI—not always exhibiting the same discount for Human or Human + AI attributions—high-comprehension participants often judge AI work more favourably than the Human-assigned—still maintaining the Human + AI attribution at the same level of that of Human (Fig. 2 and 4). This selective sensitivity to attribution suggests that comprehension does not universally elevate judgment quality, but rather interacts with the assigned authorship, particularly when AI is involved. This is also reflected in the high variability in comprehension score that is specifically exhibited by AI assignment (Fig. 3b), that can also be related to a different level of attention or susceptibility.

Given the manipulation of the authorship—which is the same for all the fairytales—the bias against AI is not intrinsic to the content, but it is rather driven by preconceptions that can be mitigated or exacerbated by the comprehension level. For instance, low-comprehension individuals may rely more heavily on the authorship label due to lower engagement or reduced capacity to extract meaning from the stimulus, thus defaulting to social or cultural heuristics such as “AI cannot produce meaningful stories” or “AI does not require the same effort as a human”. Conversely, high-comprehension individuals seem more willing to refute expectations linked to negative prejudice towards AI when the content supports it. Thus, a possible explanation of the positive bias found in High CSc participants might lie in the fact that once they have read the AI-assigned text and found it comparable to the Human-one, they tend to implicitly overscore the AI-assigned text to highlight its value and thereby refuting the negative prejudice against AI.

Importantly, this pattern does not appear to reflect a strictly dichotomous distinction between low and high comprehension: treating comprehension as a continuous variable yielded converging results, suggesting that the modulation of authorship bias scales progressively with participants' understanding of the texts (Suppl. Fig. 2).

Despite the growing number of contemporary artists using AI in their creative processes, empirical research consistently shows that AI-generated or computer-generated artworks are generally perceived as less aesthetically

Adjective	Effect type	F	Df	p
Nice	Assignment	0.08	2, 204	0.922688
	Comprehension	3.848	1, 102	0.052525
	Interaction	9.919	2, 204	*0.000077
Engaging	Assignment	0.008	2, 204	0.991542
	Comprehension	0.756	1, 102	0.386718
	Interaction	5.901	2, 204	*0.003226
Interesting	Assignment	0.549	2, 204	0.578423
	Comprehension	0.072	1, 102	0.789394
	Interaction	8.26	2, 204	*0.000355
Romantic	Assignment	3.213	2, 204	*0.042262
	Comprehension	0.744	1, 102	0.390499
	Interaction	2.426	2, 204	0.090957
Boring	Assignment	0.022	2, 204	0.978393
	Comprehension	0.679	1, 102	0.411749
	Interaction	4.91	2, 204	*0.008269
Compelling	Assignment	1.744	2, 204	0.17736
	Comprehension	0.162	1, 102	0.687739
	Interaction	4.86	2, 204	*0.008674
Exciting	Assignment	2.259	2, 204	0.107029
	Comprehension	0.03	1, 102	0.863984
	Interaction	1.115	2, 204	0.329848
Fun	Assignment	0.721	2, 204	0.487692
	Comprehension	0.202	1, 102	0.653806
	Interaction	0.549	2, 204	0.578379
Pleasant	Assignment	2.675	2, 204	0.071359
	Comprehension	2.23	1, 102	0.138417
	Interaction	8.329	2, 204	*0.000333

Table 3. Statistical analyses for the nine dimensions of text pleasantness. The results from the two-way ANOVA for repeated measures for author assignment (three levels: Human, AI, Human + AI) and Comprehension scores (CSs, two levels: low, high) are shown (* indicates significant effects). Data refers to the nine dimensions shown in Fig. 4. LSD post-hoc analysis results are reported in Fig. 4. Significant values are in [bold].

pleasing than those created by humans. According to Du Sautoy [26, p. 113], when viewers discover that a work which moved them was created by AI, they often feel “cheated, deceived or manipulated.” This reaction stems from the widespread belief that art offers a direct connection to the personal experiences and emotions of its creator, thereby facilitating a uniquely human bond. Art is still widely viewed as a privileged expression of human creativity and originality—qualities often regarded as distinctly human, setting us apart from both animals and machines. That AI is less creative than humans is a common stereotype, that is a belief or generalization about AI’s abilities. These stereotypes are often based on misconceptions, limited exposure, and media framing rather than deep-seated emotional prejudice. Even if this is not a human-group stereotype it seems to work very similarly and the best way to overcome stereotypes is exposure. Allport’s Contact Hypothesis²⁷, supported by Pettigrew and Tropp’s²⁸ meta-analysis of over 500 studies, found that positive contact between in-group and out-group members reduces stereotypes. As we have seen from our result and from prior research, usually people tend to evaluate human artifacts as more creative than AI-produced ones. Our high-comprehension individuals could contrast this stereotype as they read the 3 texts carefully and couldn’t find any evidence of a superiority of the human one. In contrast the low comprehension participants, didn’t really deal with the text and therefore fell victim to the stereotype that the manipulation induced.

In contrast to findings from previous studies using visual artworks¹⁷, where a bias towards AI only emerged when it was presented after (and so in comparison with) a human assigned artwork, here the presentation order did not significantly influence judgments. A possible explanation of this incongruence is that the nature of narrative stimuli like fairytales induces a longer processing timescale, where individual stimuli are interpreted in relative isolation and demand more sustained attention. This may have reduced comparative or contrastive effects across trials, positioning our paradigm in a different cognitive and temporal regime from rapid aesthetic judgments typical of visual stimuli.

A natural extension of this work would be to explore reverse attribution designs, where AI-authored fairytales are mislabelled as Human-generated or Human + AI-generated. Such an approach could uncover the magnitude of “surprise-like” effects—whether high-quality stimuli labelled as AI elicit unexpectedly high ratings among comprehending participants. This would help disentangle the effects of actual content quality

from expectation-driven biases, providing stronger causal evidence on how authorship and comprehension jointly shape perception.

In this study we focused on literary stimuli, which pertain to the general category of artistic stimuli. Nonetheless, the relevance of our results emerges if we consider art as a paradigm to explore how we function as human beings. The negative bias towards AI is inherently so because it is based on attribution rather than actual content: coherently, we observed it only in participants that exhibited low comprehension levels of the texts, which might have been driven by poor understanding or low attention to stimuli. When there was a deeper understanding and a high level of comprehension of the read content, the bias was reversed to positive, which—although remaining a bias by definition—is in the opposite direction and would have different consequences in ecological contexts. In broader terms, our findings contribute to a cognitive paradigm for understanding aesthetic and evaluative biases. Artistic judgment is not merely a reaction to content, but a complex inference shaped by understanding, attention, and social context. In this sense, art fruition—and especially literature—serves as a lens through which we can observe fundamental aspects of human cognition: how we construct meaning, how we rely on heuristics, and how we integrate content with context. The fact that comprehension selectively modulates AI-related judgments, but not human-assigned ones, suggests that this bias is not evenly distributed but targeted towards new advancements in art generation with the aid of technology. Thus, in a more general framework, our results suggest that although on the technical level we are always more prone to integrate AI and technology in our daily works, we still bear cognitive biases that depending on our approach to the specific contents can differently affect our relationship with them.

Limitations and future perspectives

The present findings should be interpreted in light of some methodological considerations. First, the use of authentic literary excerpts increased the ecological validity of the task by ensuring accessible and high-quality stimuli. At the same time, this choice introduced constraints related to stimulus familiarity and generalizability. A substantial proportion of participants was excluded because they recognized the original author, a necessary step to preserve the validity of the authorship-attribution manipulation. Future studies could extend the present findings using less recognizable literary materials, newly written texts, or larger sets of excerpts systematically pre-tested for familiarity, readability, genre, style, and perceived quality.

Another methodological consideration involves the operationalization of comprehension as the independent variable, which relied on a potentially arbitrary pre-defined threshold. Although supplementary analyses treating comprehension as a continuous variable yielded consistent results (Suppl. Figure 2), future studies should characterize this relationship more precisely by experimentally manipulating reading engagement or motivation, thereby disentangling genuine understanding from differences in attentional investment.

Moreover, the experimental design did not fully counterbalance the presentation order of the three authorship conditions. Specifically, only two presentation orders were implemented, and the Human + AI condition was always presented after the Human and AI conditions. This has led to a fatigue effect on the third fairytale (Fig. 3C) but our dichotomization criterion took this limit into account by putting in the High CSc group only those participants who maintained a high level of engagement throughout the whole experiment (≥ 3 correct answers for each of the three fairytales). Although the analyses of the effects of order presentation suggested that presentation order did not substantially account for the observed effects (Suppl. Tables 1 and 6), future studies should adopt a fully randomized and counterbalanced design to isolate the contribution of authorship labels while controlling for potential serial-position or fatigue effects.

Finally, another limitation of our study lies in the fact that the sample was composed primarily of young adults and included a higher proportion of female participants. Although the analyses controlling for these demographic variables did not reveal significant effects on the main pattern of results, the findings should be interpreted within the characteristics of the sampled population. Moreover, as perceptions of AI-generated content rapidly evolve alongside increasing societal exposure to generative models, future studies should recruit broader, more demographically diverse samples to determine whether the present findings generalize beyond predominantly young, digitally native cohorts—particularly across different age groups, levels of familiarity, and attitudes toward AI technologies.

Methods

Participants

A total of 190 participants took part in this study. All participants were recruited as volunteers. The study was conducted in accordance with University of Verona requirements. An informed consent was obtained from them before starting the experimental session. During data analysis, 86 participants were excluded from the study: 5 because they had not answered all the evaluation questions, 7 because they had not answered all the comprehension questions, and 74 because they declared they recognized the actual fairytale's author after reading. Therefore, all the analyses were performed on 104 participants (70 female, 34 male), with an age between 19 and 40 years old ($M_{\text{age}} 22.89 \pm SD_{\text{age}} 3.88$).

Literary stimuli and procedure

The literary stimuli were three texts (of about 2000 words) extracted from the Italian version of Roald Dahl's fairytales (*Matilde*, *Le Streghe*, *James e la Pesca Gigante*; see Appendix 1). The choice of Roald Dahl's fairytales was based on two main reasons: 1) we wanted to ensure consistent and high-level literary quality across all the text stimuli; this allowed us to isolate the effect of the “author label” from potential confounding variables such as poor writing or linguistic inconsistencies, providing a robust standard to test whether the mere label of “AI” could diminish the perceived value of a text that is objectively recognized as high-quality literature; 2)

Dahl's prose is designed to be engaging yet linguistically accessible, thus decreasing the probability of drops in comprehension or appreciation.

The authorship assignment was manipulated by declaring one of three authorship conditions, through the following sentence: "Now you will read a fairytale written by a children's author/by Artificial Intelligence/by a children's author who uses Artificial Intelligence". Coherently the 3 conditions were:

- 1) Human author (named Human),
- 2) AI,
- 3) Human author assisted by AI (named Human + AI).

The assignment of authorship to the texts was pseudo-randomized between human and AI, balancing two possible presentation orders (Human/AI/Human + AI or AI/Human/Human + AI). This pseudo-randomized design was chosen to avoid that a fatigue effect could interfere with the first two conditions that were our primarily interest.

Participants underwent an online reading task designed on PsyToolkit^{29,30}, in which they were asked to read each fairytale and complete a questionnaire after each reading session. The questionnaire was designed to evaluate the impact of authorship assignment on participant's evaluation of social value and pleasantness of the texts, and the level of comprehension of each fairytale. The reading task was only conducted on PCs, excluding the use of tablets or smartphones.

In order to evaluate the pleasantness of the stimuli, for each text participants were first asked to rate nine dimensions (adjectives selected from Knoop et al.³¹, see Appendix 2)—*Nice, Engaging, Interesting, Romantic, Boring, Compelling, Exciting, Fun, Pleasant*—on a 7-point Likert scale from 1 (Not at all) to 7 (Very much).

In order to evaluate the social and literary value of the stimuli, participants were asked to rate their level of accordance with four questions (see Appendix 2) on a 5-point Likert scale for each text, as previously done²¹, the questions were as follows:

- 1) *Do you think this fairytale has won a literary award?*
- 2) *Do you think literary critics consider this fairytale important?*
- 3) *Do you think this fairytale should be taught in schools?*
- 4) *Do you find this fairytale irrelevant?*

Then, a set of five questions about the text content (YES/NO answers) was provided for each fairytale to score participants comprehension (see Appendix 2).

At the end of all the three text presentations, participants were asked to declare whether they recognized the fairytales, and if yes, to indicate their actual author.

Data analysis

Data are presented as average $\pm$ SEM per groups. Comprehension scores (CSc) was computed as the sum of correct answers over the 5 comprehension questions, so it ranges from 0 to 5. Based on the CSc, the 104 participants were divided in two groups: group with a high comprehension score (High CSc ≥ 3 for each of the three fairytales, 56 participants), and group with a low comprehension score (Low CSc ≤ 2 for at least one of the three fairytales, 48 participants). CSc for the three author assignments significantly differ between Low CSc and High CSc group (Suppl. Figure 1).

Social prestige was investigated with a one-way ANOVA for repeated measures for authorship assignment (three levels: Human, AI, Human + AI). In addition, the social prestige of texts, as well as their pleasantness, were investigated with a 3×2 mixed ANOVA, considering the separate effects and the interaction between authorship assignment as repeated measure factor (three levels: Human, AI, Human + AI) and CSc as between factor (two levels: High CSc, Low CSc). LSD post-hoc analyses were performed to evaluate the pairwise comparisons when appropriate. The same analysis was repeated using the average CSc and a median split criterion (Suppl. Tables 2–7), to confirm that the results remain robust across different dichotomization criteria. A complementary Spearman's correlation analysis (Suppl. Figure 2) was used to correlate fairytale evaluations in the assignment conditions with the average CSc. This latter approach was used to confirm whether the observed effects were driven by group categorization or were still present when comprehension was treated as a continuous variable.

In order to evaluate the number of read book by each participant we discretized their answers ("from 0 to 5", "from 6 to 10", "from 11 to 15", "more than 15") transforming them into 5, 10, 15, 20 respectively (Suppl. Table 3).

The presentation order of texts was tested performing a 3×2 mixed ANOVA, considering the separate effects and the interaction between authorship assignment as repeated measure factor (three levels: Human, AI, Human + AI) and the order presentation (two levels: Group A: Human, AI, Human + AI; Group B: AI, Human, Human + AI).

The effects of sex were tested with t-test between males and females CSc; the effects of age were tested with a correlation between age and CSc.

To control for potential differences across the three Roald Dahl fairytales, ratings were additionally analyzed using an ordinary least squares (OLS) linear model including subject-specific intercepts, assignment condition (Human, AI, Human + AI), comprehension level (High vs. Low), their interaction, and the specific fairytale used as stimulus. The model was specified as:

$$rating \sim C(subject) + C(assignment) * C(comprehension) + C(fairytale)$$

where (C) indicates categorical variables and (*) includes both main effects and interaction terms. Statistical significance was assessed using a Type II ANOVA on the fitted model and resulting p-values were corrected for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) procedure (Suppl. Table 7).

Data availability

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

Received: 27 December 2025; Accepted: 31 July 2026

References

1. DY Tenen 2024 Literary Theory for Robots: How Computers Learned to Write W. W. Norton & Company
2. Botnik Studios. *Harry Potter and the portrait of what looked like a large pile of ash*. 2017
3. R Goodwin 2018 1 the Road Jean Boîte Éditions
4. TM Amabile 2020 Creativity, artificial intelligence, and a world of surprises Acad. Manag. Discov. 6 3 351 354
5. Numero Cromatico 2021 Epitaphs for the Human Artist 1st Numero Cromatico Editore
6. Numero Cromatico 2022 Poesie sulla fine 1st Numero Cromatico Editore
7. J Piringer 2018 Datenpoesie Ritter Verlag
8. RA Zwaan 1994 Effect of genre expectations on text comprehension J. Exp. Psychol. Learn. Mem. Cogn. 20 4 920 933
9. J Hitsuwari Y Ueda W Yun M Nomura 2023 Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry Comput. Hum. Behav. <https://doi.org/10.1016/j.chb.2022.107502>
10. N Kobis LD Mossink 2021 Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry Comput. Hum. Behav. <https://doi.org/10.1016/j.chb.2020.106553>
11. Rebol, M., Güti, C., & Pietroszek, K. Passing a non-verbal Turing test: Evaluating gesture animations generated from speech. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (pp. 573–581). IEEE (2021).
12. Tarnate K. J. M., Garcia M. M., & Sotelo-Bator P. Short poem generation (spg): a performance evaluation of hidden Markov model based on readability index and Turing test. *International Journal of Advanced Computer Science and Applications*, **11** (2) (2020).
13. Xue, A. End-to-end chinese landscape painting creation using generative adversarial networks. In *Proceedings of the IEEE/CVF Winter conference on applications of computer vision* (pp. 3863–3871) (2021).
14. M Appel TV Messingschlager 2022 Creative artificial intelligence and narrative transportation Psychol. Aesthet. Creat. Arts <https://doi.org/10.1037/aca0000495>
15. Y Wu Y Mou Z Li K Xu 2020 Investigating American and Chinese subjects' explicit and implicit perceptions of AI-generated artistic work Comput. Hum. Behav. <https://doi.org/10.1016/j.chb.2019.106186>
16. K Darda M Carre E Cross 2023 Value attributed to text based archives generated by artificial intelligence R. Soc. Open Sci. <https://doi.org/10.1098/rsos.220915>
17. SG Chiarella G Torromino DM Gagliardi D Rossi F Babiloni G Cartocci 2022 Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings Comput. Hum. Behav. <https://doi.org/10.1016/j.chb.2022.107406>
18. Grassini S., Koivisto M. Assessing the Influence of Personality Traits, Experiences, and Attitudes on the Perception and Evaluation of AI-generated Artworks (2023).
19. Demir S., Fuegener, A., Gupta A., Weinmann, M. When AI is Creative: How Do Humans Perceive Creativity When AI is Involved? *ICIS 2024 Proceedings*. **17**. (2024).
20. M Salgaro 2015 How literary can literariness be? Methodological problems in the study of foregrounding Scientific Study of Literature 5 2 229 249
21. Salgaro M., Sorrentino P., Lauer G., Lüdtke J., Jacobs A. M. How to measure social prestige of a Nobel prize in literature? Development of a scale assessing the literary value of a text. *Nodes* 11–12, (2018).
22. R Chamberlain C Mullin B Scheerlinck J Wagemans 2018 Putting the art in artificial: Aesthetic responses to computer-generated art Psychol. Aesthet. Creat. Arts 12 2 177
23. C Di Dio M Ardizzi SV Schieppati D Massaro G Gilli V Gallese A Marchetti 2023 Art made by artificial intelligence: The effect of authorship on aesthetic judgments Psychol. Aesthet. Creat. Arts 19 5 1164
24. Ragot M., Martin N., Cojean S. AI-generated vs. Human artworks. A perception bias towards artificial intelligence? In *Extended abstracts of the 2020 CHI conference on human factors in computing systems* pp. 1–10, (2020).
25. M Du Sautoy 2021 Der Creativity-Code: Wie künstliche Intelligenz schreibt, malt und denkt
26. GW Allport 1954 The Nature of Prejudice Addison Wesley
27. TF Pettigrew LR Tropp 2006 A meta-analytic test of intergroup contact theory J. Pers. Soc. Psychol. 90 5 751 783
28. G Stoet 2010 PsyToolkit: A software package for programming psychological experiments using Linux Behav. Res. Methods 42 1096 1104
29. G Stoet 2017 PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments Teach. Psychol. 44 1 24 31
30. Knoop C. A., Wagner V., Jacobsen T., Menninghaus W., Mapping the aesthetic space of literature “from below”, *Poetics*, Volume **56**, Pages 35–49, ISSN 0304–422X (2016).
31. Brian, Porter Edouard, Machery (2024) AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably Scientific Reports 14(1) 10.1038/s41598-024-76900-1
32. Che, J., Sun, X., Gallardo, V. & Nadal, M. Cross-cultural empirical aesthetics. *Prog. Brain Res.* **237**, 77–103. Davinio, C. Techno-Poetry And Virtual Reality (Sometti, 2002)(2018).
33. Gallese, V. & Di Dio, C. Neuroaesthetics: The body in esthetic experience. In *Encyclopedia of Human Behavior* 2nd edn (ed. Ramachandran, V. S.) 687–693 (Academic Press, 2012).
34. Gunser, V. E. et al. Can users distinguish narrative texts written by an artificial intelligence writing tool from purely human text? In *HCI International 2021 - Posters* (eds Stephanidis, C., Antona, M. & Ntoa, S.) *Communications in Computer and Information Science* **1419**, 514–521 (Springer, 2021).
35. Jacobsen, T. Bridging the arts and sciences: A framework for the psychology of aesthetics. *Leonardo* **39**, 155–162 (2006).
36. Linardaki, C. Poetry at the first steps of Artificial Intelligence. *Humanist Stud. Digit. Age* 7, 1–18 (2022).
37. Mitchell, M. Abstraction and analogy in AI. *Ann. N. Y. Acad. Sci.* **1524**, 22–38 (2023).
38. Morris, A. K. & Swiss, T. *New Media Poetics: Contexts, Technotexts, and Theories* (MIT Press, 2006).
39. Nadal, M. & Chatterjee, A. Neuroaesthetics and art's diversity and universality. *Wiley Interdiscip. Rev. Cogn. Sci.* **10**, e1487 (2019).
40. O'Sullivan, J. *Towards a Digital Poetics: Electronic Literature & Literary Games* (Palgrave Macmillan, 2019).
41. Pearce, M. T. et al. Neuroaesthetics: The cognitive neuroscience of aesthetic experience. *Perspect. Psychol. Sci.* **11**, 265–279 (2016).

43. Porter, B. & Machery, E. AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. *Sci. Rep.* **14**, 26133 (2024).

Author contributions

D.M.G., G.T., M.F., S.G.C., M.S., M.R. and S.Re. contributed to the conceptualization, methodology, and study materials. F.M., S.Ra. and G.T. were responsible for data collection and data analysis. G.T. supervised the project. G.T. and S.Ra. wrote the initial draft. All authors provided comments to and approved the submitted version of the manuscript.

Funding

This work was supported by the “Inclusive Humanities” Project of Excellence (2023–2027) of the Department of Foreign Languages and Literatures at the University of Verona funded by *Ministero dell’Università e della Ricerca* (MUR, Italy).

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval and consent to participate

We confirm that prior to data collection, all the research methods were carried out in accordance with relevant guidelines and regulations of the Italian Law and also comply with the Helsinki Declaration. Ethical approval for this study was waived by the institutional committee of the University of Verona. All participants have obtained and signed written informed consent. No economic incentives or compensation were provided for participation in this study.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-65454-z>.

Correspondence and requests for materials should be addressed to D.M.G., M.S. or G.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026