Computers in Human Behavior 104 (2020) 106186

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Investigating American and Chinese Subjects' explicit and implicit perceptions of AI-Generated artistic work

Yuheng Wu [a], Yi Mou [a,*], Zhipeng Li [a], Kun Xu [b]

[a] Shanghai Jiao Tong University, School of Media & Communication, 800 Dongchuan Road, Minhang District, Shanghai, 200240, China
[b] University of Florida, Department of Telecommunication, College of Journalism and Communications, United States

ARTICLE INFO

Keywords:
Artificial intelligence
AI-Generated artistic work
Explicit perception
Implicit perception
Experimental survey

ABSTRACT

As the prevalence of AI-generated content increases, examining viewers' perceptions of the content is crucial to understanding the human-machine relationship and further facilitating efficient human-machine collaboration. Prior literature has accumulated mixed findings regarding subjects' attitudes toward and perceptions of news and tweets written by natural language generation (NLG) algorithms. To resolve this inconsistency and expand our understanding beyond NLG, this study investigated the explicit and implicit perceptions of AI-generated poetry and painting held by subjects from two societies. An experimental survey was conducted to examine the subjects' explicit and implicit perceptions of AI-generated content in the U.S. and China. As the U.S. and China fiercely compete to lead the development of AI technology, their citizens exhibit divergent attitudes toward AI's performance in artistic work. The U.S. subjects were more critical of the AI- than the human-generated content, both explicitly and implicitly. Although the Chinese subjects were overtly positive about the AI-generated content, they appreciated less this content than the human-authored content. The findings enrich our understanding in the domain of AI generation. Theoretical and practical implications are discussed.

1. Introduction

With the rapid development of artificial intelligence (AI), natural language generation (NLG) technology has been widely applied across the globe (Caswell & Dörr, 2018). Take news as an example, mainstream newspapers such as *The New York Times* and news agencies such as The Associated Press have adopted NLG technology to write news reports (Haim & Graefe, 2017; Waddell, 2018). In 2014 alone, one of the leading NLG companies, Automated Insights, produced more than 1 billion news articles and wrote up to 2000 news stories per second (Pressman, 2017). Scholars have started to investigate the social and psychological effects of AI-generated content, but mainly focused on news, as this is probably the most mature genre that AI can generate.

Meanwhile, the impact of AI has become tangible in the fine arts and literature (Falcon, 2018; Schaub, 2016). For instance, a *Harry Potter* sequel story generated by a Botnik Studio's robot met with applause and amusement (Liao, 2017). Creative writing has also been automated in languages other than English, such as Russian, Japanese, and Chinese (Geng, 2018). Microsoft's AI bot XiaoIce had a Chinese poetry collection published in 2017 and a painting exhibition in 2019 (iFeng News, 2017; Xinhua News, 2019). Relatedly, the human chess champion Gary Kasparov's prediction has been contravened by the recent evolution of computing technologies. Following his defeat by IBM's Deep Blue in a game of chess, Kasparov once noted, "there is a frontier that they [machines] must not cross," referring to areas such as art, literature, and music (Kasparov 1996 quoted in Hofstadter, 2001).

While existent literature on evaluating AI-generated content has mainly focused on NLG or large-scale classifications of pictures (Saleh & Elgammal, 2015), research on people's understanding of AI-generated visual art has been under-examined. Relevant literature on AI art is still largely philosophical or meta-physical on questions such as "Is machine art acceptable in the artworld?" (Ch'ng, 2019, p. 1) or whether creativity is a unique human characteristic (Gayford, 2016). In this study, poetry and painting were selected to represent different creative genres. Moreover, empirical studies have used exclusively self-reported measures to assess subjects' attitudes toward and perceptions of AI-generated content. However, the longstanding love-hate relationship between humans and machines (with technology zealots and Luddites at opposite ends of the spectrum) may have complicated our feelings toward AI and AI-generated content (Tegmark, 2017). Hence, subtleties in the appraisal of AI-generated content may not be fully captured using this simplistic method.

* Corresponding author.
E-mail addresses: wu.yu.heng@qq.com (Y. Wu), yimou@sjtu.edu.cn (Y. Mou), ZhipengLi_Knute@outlook.com (Z. Li), kun.xu@ufl.edu (K. Xu).

https://doi.org/10.1016/j.chb.2019.106186
Received 22 February 2019; Received in revised form 30 October 2019; Accepted 30 October 2019
Available online 4 November 2019

To resolve the inconsistencies in previous empirical findings and expand our understanding beyond NLG, an experimental survey was conducted to examine the public's explicit and implicit perceptions of AI-generated poetry and painting. As the general opinion of AI technology varies from society to society, data were collected in the U.S. and China as two societies representing relatively pessimistic and optimistic attitudes toward AI, respectively, to provide a fuller picture of the issue under investigation.

2. Literature review

2.1. Perceptions of AI-generated content

As NLG algorithms are currently reshaping global journalism, the implications of this advancement have clearly gone beyond the technological (Van Dalen, 2012). Skeptics and proponents continue to debate whether AI-generated content will ultimately meet the benchmark of content produced by human writers (Latar, 2015). However, a question more relevant to readers is how audiences perceive AI-generated content. As previous studies have tended to focus on AI-generated news, credibility and readability have been frequently assessed. Previous comparisons of human and AI writers have led to the following three main conclusions.

First, consistent with Sundar. (2008) concept of machine heuristics, according to which nonhuman agents tend to be perceived as more objective than their human counterparts, Graefe, Haim, Haarmann, and Brosius (2018) observed that machine-written content was rated as more credible and demonstrating more expertise than human-authored content. More specifically, they developed a 12-item measure to capture subjects' perceptions of credibility (the degree to which the content was considered accurate, trustworthy, fair, and reliable), readability (entertaining, interesting, vivid, and well written), and journalistic expertise (coherent, concise, comprehensive, and descriptive). Clerwall (2014) explored subjects' perceptions of news credibility (the degree to which the content was considered informative, trustworthy, objective, and descriptive) and readability (pleasant to read, clear, well written, coherent, and not boring). His findings showed that without authorship disclosure, there was a slight difference in perceptions of the credibility and readability of text. AI-generated news was perceived as somewhat descriptive and objective but less interesting than articles written by human journalists.

Second, in contrast to the above findings, Waddell (2018) reported that machine authorship negatively affected perceived credibility, as news attributed to a machine was perceived as less credible than news attributed to a human journalist. Following Appelman and Sundar (2016), Waddell (2018) gauged message credibility using items such as "accurate," "authentic," "believable," "high quality," "newsworthy," and "representative."

Third, no significant differences in credibility or other relevant variables have been identified in other studies. van der Kaa and Krahmer (2014) conducted an experiment with a 2 (author: computer or journalist) X 2 (story topic: sport or finance) between-subject factorial design, and found that ordinary news consumers attributed the same levels of trustworthiness and expertise to a computer writer and a journalist. However, the sampled journalists perceived the computer to be less trustworthy but to demonstrate more expertise than the human journalist. Edwards, Edwards, Spence, and Shelton (2014) designed two mock Center for Disease Control and Prevention (CDC) Twitter pages on the topic of sexually transmitted infections, one authored by a CDC Twitterbot and the other by a human scientist. Similarly, no significant differences were found in the subjects' perceptions of source credibility and communication competence. However, the Twitterbot was rated as less attractive than the human Twitter agent.

These inconsistent findings call for closer examination of perceptions of the quality of AI-generated content. Almost all of the measures of appraisal used in the above studies were drawn from self-reported credibility scales whose effectiveness had been demonstrated in traditional news credibility research (Sundar, 1999). However, subjects involved in news credibility studies typically feel little pressure due to social desirability bias or the "spiral of silence" effect, as a majority opinion is rarely detected (Noelle-Neumann, 1991). Their perceptions of the content can thus be accurately and straightforwardly gauged from their self-reported answers. In a situation that involves mixed feelings about AI as a challenging out-group, however, a simplistic self-reported measure may not suffice.

2.2. Explicit and implicit perceptions

Traditionally, perception has been regarded as a conscious act, as we rely on awareness to describe our experiences to others and respond to stimuli surrounding us (Kihlstrom, Barnhardt, & Tataryn, 1992, pp. 17–54). In the last few decades, however, considerable scholarly attention has been paid to the process of nonconscious perception—or perception without awareness—by social psychologists and cognitive scientists (Bornstein & Pittman, 1992). They have advocated differentiating perception with awareness from that without to capture "true" perception. The terms "explicit perception" and "implicit perception" were coined to reflect the discrepancy between these two types of perception (Kihlstrom et al., 1992, pp. 17–54). Explicit perception refers to a person's "conscious perception of some object or event in the current stimulus environment," while implicit perception is reflected in "any change in experience, thought, or action that is attributable to some event in the current stimulus field, even in the absence of conscious perception of that event" (Kihlstrom et al., 1992, pp. 4–5). The former involves detecting, identifying, and describing sensations and experiences, while the latter does not require the subject to detect any object at all.

Normally, individuals' explicit perceptions overlap with their implicit perceptions, due to the "fundamental multiplicity of measures of a—presumably—unitary construct" (Maass, Castelli, & Arcuri, 2000, p. 96). However, when it is necessary to hide one's true perception, a discrepancy may appear. Indeed, the difference between explicit and implicit perceptions has usually been investigated in prejudice- or stereotype-related contexts (Greenwald & Banaji, 1995; Greenwald, McGhee, & Schwartz, 1998). As stereotype use meets with social disapproval and is negatively sanctioned, those who use stereotypes normally encounter overt rejection (Monteith, 1993). However, psychologists have also revealed subtle gestures of conformity with in-group members who use stereotypes (Castelli, Vanzetto, Sherman, & Arcuri, 2001). Conformity denotes the tendency to structure an ambiguous context congruently with others' suggestions (Asch, 1956; Cialdini & Trost, 1998), and reflects an implicit measure of perception. For instance, people who label Barack Obama as black implicitly perceive race as more categorical than those who label Obama as multiracial (Malahy, Sedlins, Plaks, & Shoda, 2010).

Sloman's (1996) model of two systems of reasoning has also been used to understand the formation of explicit and implicit perceptions and attitudes. The fast-learning reasoning system relies on logical, verbal, and symbolic representations at a high level of cognitive processing. In parallel, the slow-learning reasoning system pairs similar and contiguous associations. Conscious and verbally based explicit perceptions and attitudes form through fast-learning reasoning, while nonconscious and association-based implicit perceptions and attitudes form through slow-learning reasoning (Olson & Fazio, 2001; Petty & Wegener, 1998; Rydell & McConnell, 2006). Rydell, McConnell, Mackie, and Strain (2006) further indicated that implicit perceptions of an object reflect the valence of subliminal primes, whereas explicit perceptions reflect the valence of verbally presented behaviors.

Accordingly, a divergence between explicit and implicit perceptions may emerge in situations not related to prejudice or stereotyping, as long as two reasoning systems coincide. On confronting AI-generated content, a verbally based appraisal forms in the fast-learning system and can be reported with control. Concurrently, the encounter may activate nonconscious appraisal, to which people do not initially have conscious access. Explicit and implicit appraisals — as subcategories of explicit and implicit perceptions — of AI-generated content thus form independently.

2.3. Impact of general social opinion on AI

Since the emergence of AI in the 1950s, people's attitudes toward AI have fluctuated with the flow and ebb of technological development (Markoff, 2016). Notably, public opinion toward AI varies from society to society. Depictions of the impact of AI range from a rosy future to a catastrophic apocalypse (Tegmark, 2017). For instance, 73% of European Union (EU) citizens expressed the fear of losing jobs to robots in a Eurobarometer study (2012). Indeed, 40%–60% of jobs in EU countries were estimated to be at risk due to roboticization (Moniz & Krings, 2016). Similarly, it was estimated that 51% of economic activities in the U.S. could be automated using current technologies, and almost every occupation has at least some potential to be automated (Manyika et al., 2017). In addition to the fear of massive unemployment, the fear of being overtaken by AI is prevalent in the U.S. and elsewhere, as evinced by an open letter signed by Stephen Hawking, Elon Musk, and other AI experts (Future of Life Institute, 2015). This is why citizens of Western nations are rather cautious about developing AI technologies.

In sharp contrast, AI technology is widely applauded and praised in China, where its threats and pitfalls have not been fully expressed and discussed (Xinhua News, 2018). In July 2017, the Chinese government set the advancement of AI as a national development strategy, which further fueled the public's zeal for AI (Kania, 2018). Indeed, a recent survey found that 70% of Chinese workers believe that AI will have or already has had a positive impact on their working lives (McNeice, 2018); by comparison, 73% of American workers believe the introduction of AI will result in a net loss of jobs (Northeastern University & Gallup, 2018).

Therefore, we expected to observe the "bandwagon effect" in AI-related attitudes in Chinese society. The bandwagon effect is the phenomenon whereby people affiliate themselves with the position that they perceive to represent the majority view or to be dominant in society (Schmitt-Beck, 2015). For instance, Sundar, Xu, and Oeldorf-Hirsch (2008) found that in the context of Web browsing, when product ratings were based on the inputs of a large number of users, users' perceptions of their peers' opinions were affected. Individuals' impressions of the attitudes, beliefs, and behaviors of groups of others may shape their own attitudes, beliefs, and behaviors (Mutz, 1998). Therefore, support for AI technology may breed further support as long as it appears to be the majority opinion. Consequently, positive attitudes toward and beliefs about AI are generally reported by the Chinese public (Xinhua News, 2018).

In the Chinese context, but not necessarily the U.S. context, the bandwagon effect may motivate people to reduce their negative view of — but not necessarily their prejudice against — AI. Social psychologists have argued that implicit and explicit perceptions are unlikely to correlate in the presence of such a motivation, but they may substantially coincide when people are unmotivated to conceal their perceptions (Maass et al., 2000). Therefore, the U.S. and China provide two typical settings in which researchers could examine the public's implicit and explicit perceptions of AI.

2.4. Present study

The continuum of inhibition potential proposed by Maass et al. (2000) suggests a multiplicity of methods for measuring perceptions based on the level of difficulty subjects experience in controlling their responses. Using traditional self-reported measures — such as Likert scales, semantic differential scales, and feeling thermometers — is probably the most straightforward way to assess what people think (Cunningham, Preacher, & Banaji, 2001). However, as people can easily control their responses and provide socially desirable answers, this method may not yield sincere responses in normative contexts, especially when compliance with certain social norms is necessary (Lambert, Cronen, Chasteen, & Lickeel, 1996). Speech analysis is better able to capture unconscious responses. Although individuals can largely manipulate their overt rhetorical strategies (Schmid & Fiedler, 1996), they are generally unaware of their subtle variations in linguistic strategies and unlikely to reflect consciously on their language use (von Hippel, Sekaquaptewa, & Vargas, 1997). For instance, Maass, Salvi, Arcuri, and Semin (1989) detected a linguistic intergroup bias by coding language use in intergroup contexts. Specifically, people encode and communicate desirable in-group and undesirable out-group behaviors more abstractly than undesirable in-group and desirable out-group behaviors, as predicted by Semin and Fiedler's (1988) linguistic category model.

Drawing on the literature measuring the perceptions of in- and out-groups, the present study used two approaches to identify explicit and implicit appraisals of AI-generated content, respectively. A traditional self-reported questionnaire assessed the subjects' explicit evaluations of AI- and human-generated content. We gathered the subjects' verbal responses by asking them to type their first thoughts after viewing the content. Content analysis of these responses was conducted to detect the subjects' subtle mindless appraisals. This procedure followed Greenwald and Banaji's (1995) suggestion that "An implicit attitude toward B may be indirectly indicated by a (direct) measure of evaluation of A, when A and B have some relation that predisposes the implicit influence" (p. 8). Here, the subjects' appreciation based on the content analysis was assessed as A, while B was the AI-generated content. Compared with the commonly used Implicit Association Test, analyzing phrased responses better suited the context of this study, which focused on individuals' perceptions rather than attitudes (Maass et al., 2000).

Prior research has suggested that people do not distinguish the type authenticity or accuracy when it comes to specific categories such as autonomously improving the safety features of a car, but they do perceive algorithmic work as less morally authentic than human work (Jago, 2019). Here, type authenticity concerns "whether an entity is true to its associate type (or category or genre)", whereas moral authenticity focuses on "whether the decisions behind the enactment and operation of an entity reflect sincere choices rather than socially scripted responses" (Carroll & Wheaton, 2009, p. 255). Compared to news that is more fact-based and evaluated as more objective, poems and paintings were expected to elicit subtle emotional appraisals and affective reactions. As much affection is usually added to the work when composing poems or drawing a painting, individuals' perception on moral authenticity may play an underlying role in evaluating poems and paintings. As current AI has no consciousness or feelings, even though it is able to generate literature and visual art, the work reflects pre-scripted computing codes rather than human authors' sincere choices. Even with the agency (the capacity to act and do), the lack of experience (the capacity to feel and sense) of AI may make people feel unnerving (Gray & Wegner, 2012). Therefore, we postulate that subjects would report more positive appraisals of human work than identical AI work. However, as discussed in Section 2.3., the overwhelmingly favorable social atmosphere toward AI in current Chinese society may lead people to provide socially desirable responses in evaluating AI work. In that case, Chinese respondents may explicitly report more positive evaluation of AI work than human work.

Drawing on the literature reviewed above, the following hypotheses have been postulated:

H1. American subjects will appraise a human-generated (a) poem and (b) painting more favorably than AI-generated ones explicitly.

H2. American subjects will appraise a human-generated (a) poem and (b) painting more favorably than AI-generated ones implicitly.

H3. Chinese subjects will appraise an AI-generated (a) poem and (b) painting more favorably than human-generated ones explicitly.

H4. Chinese subjects will appraise a human-generated (a) poem and (b) painting more favorably than AI-generated ones implicitly.

3. Method

3.1. Procedure and stimuli

This study was part of a large-scale cross-national study. In the U.S. and China, an experimental survey with a 2 (author: AI vs. human) X 2 (genre: poem vs. painting) factorial design was conducted. The subjects were recruited and randomly assigned to one of four conditions. After reading a poem or viewing a painting, they were immediately asked to type their first thoughts, and then to answer a questionnaire. Authentic AI-generated content with similar levels of sophistication in English and Chinese was employed as stimuli. To avoid the complications associated with language translation, we chose two poems written by AI: one in English, and one in Chinese. The poem in English, "Long Years Have Passed," was written by Ray Kurzweil's Cybernetic Poet, and the poem in Chinese, "Window" was written by Microsoft's XiaoIce bot and published in a poetry collection in 2017 (iFeng News, 2017). Both poems describe nostalgic complex. The selected painting was drawn by PIX18, a robot invented by the Creative Machines Lab at Columbia University. With its clear abstract style, this painting won the top prize in the Robot Art 2017 competition (Dabai, 2017). As there was no language barrier to the appreciation of this painting, it was used in both the U.S. and the Chinese context. To exclude any possible confounding variables associated with existing perceptions of technology brands, a fictitious AI product "Starbot" was named as the AI poet or painter. Similarly, fictitious human names, either English or Chinese, were given to the human poets and painter. The experimental materials were exactly the same across the genre conditions, with the exception of their designated authorship. Two coders coded the subjects' phrased responses independently once the inter-coder reliability had reached a desirable level.

3.2. Sample

Four hundred and twenty-two U.S. participants were recruited through Amazon Mechanical Turk (MTurk) and randomly assigned to one of the four conditions on Qualtrics. All of the participants were located in the U.S. and aged 18 or above. Every participant was compensated with 0.75 USD upon completing the experiment. Responses that failed the manipulation check were removed, leaving 251 valid responses for data analysis. Slightly more than half ($n = 134$, 53.4%) of the participants were males and 117 (46.6%) were females. The average age of the 251 participants was 37.06 years old ($SD = 13.28$). The average yearly income level was between 45,001 and 55,000 USD ($M = 3.01$, $SD = 1.41$). All of the participants had high school qualifications or higher: 25.5% received high school education, 53.0% had Bachelor's or other college degrees, and 21.5% had post-graduate degrees or higher.

Three hundred and forty-four Chinese participants were recruited through a commercial online survey service Survey Star. This platform was contracted to send a recruitment announcement to its national sampling pool of adults. All of the participants were compensated with bonus points, which could be accumulated and exchanged for cash or consumer products. After a manipulation check, 293 valid responses were subjected to data analysis. Females made up 62.8% ($n = 184$) of the participants, and 37.2% ($n = 109$) were males. On average, the Chinese participants were younger than their U.S. counterparts ($M = 32.10$ years old, $SD = 7.65$). The average monthly income was between 6001 and 9000 yuan ($M = 3.05$, $SD = 1.12$). The majority of the participants ($n = 252$, 86.0%) had a Bachelor's or higher college degree.

3.3. Measures

The appraisal of poem/painting was operationalized into five related but distinct constructs: perceived quality, imaginativeness, spatial presence, empathy, and competence. Those constructs were selected as they are commonly gauged in content evaluation studies and pertinent to poem and painting appreciation (e.g., Lüdtke, Meyer-Sickendieck, & Jacobs, 2014; Yeung, 2015). The original measure scales were in English. Two bilingual researchers back-translated the English instruments into Chinese. The measurement items were identical across the two languages, with the exception of certain demographic measures included to fit the culture of each society. Unless otherwise specified, all items were measured on a 5-point Likert scale from "strongly disagree" to "strongly agree" (see Table 1).

Perceived quality was gauged using Sundar's (2000) 9-item scale, including "well written" and "interesting." The reliability coefficient Cronbach's α was 0.92 for the U.S. participants and .82 for their Chinese counterparts.

Perceived imaginativeness was assessed using a self-constructed scale. The three items were "imaginative," "creative," and "innovative." Cronbach's α was 0.89 for the U.S. participants and .72 for their Chinese counterparts.

Table 1
Measurement scale items for key variables.

Variable (Source)	Measurement Items	α for American	α for Chinese
Perceived quality (Sundar, 2000)	1) Enjoyable 2) Clear 3) Coherent 4) Well-written 5) Lively 6) Interesting 7) Concise 8) Comprehensive 9) Informative	.92	.82
Perceived imaginativeness (Self-constructed)	1) Imaginative 2) Creative 3) Innovative	.89	.72
Spatial presence (Lombard et al., 2009)	1) How much did it seem as if the objects or people in the poem/painting had come to the place you were? 2) How much did it seem as if you could reach out and touch the objects or people in the poem/painting? 3) To what extent did you experience a sense of being there inside the poem/painting?	.92	.74
Empathy (Jin, 2011)	1) I could relate to the author of the poem/painter. 2) I felt close to the author of the poem/painter. 3) I felt empathetic towards the author of the poem/painter. 4) I could identify with the author of the poem/painter.	.94	.81
Writer/painter's competence (van der Kaa & Krahmer, 2014)	1) Knowledgeable 2) Expert 3) Intelligent 4) Gifted 5) Authoritative	.91	.81
Attitude toward AI (Ajzen, 1991)	1) Bad … Good 2) With ill will … With good will 3) Not Beneficial … Beneficial 4) Silly … Intelligent 5) Not helpful … Helpful	.89	.57

Spatial presence was measured using the modified Temple Presence Inventory (Lombard, Ditton, & Weinstein, 2009) after removing some unsuitable items. Cronbach's α was 0.92 for the U.S. participants and .74 for their Chinese counterparts.

Empathy was gauged using Jin's (2011) 4-item scale. An example was "I felt empathetic toward the author of the poem." The reliability coefficient, Cronbach's α, was 0.94 for the U.S. participants and .81 for their Chinese counterparts.

Writer/painter's competence was assessed using van der Kaa and Krahmer's (2014) 5-item scale, with items such as "expert" and "authoritative." Cronbach's α was 0.91 for the U.S. participants and .81 for their Chinese counterparts.

Attitude toward AI was measured using a version of Ajzen's (1991) semantic differential scale, with items such as "bad–good" and "not helpful–helpful." The reliability coefficient Cronbach's α was 0.89 for the U.S. participants and .57 for their Chinese counterparts.

3.4. Attention and manipulation check

Both online survey platforms, Qualtrics and Survey Star, provide basic attention check and remove the responses of those participants who complete questions too quickly. Moreover, to ensure that the responses were based on assigned authorship, the participants were asked to recall the name of each poet or painter and answer a multiple-choice question. If they gave an incorrect answer or chose "can't remember," their responses were counted as invalid and removed from the data analysis.

3.5. Coding scheme

A coding scheme was developed to analyze the participants' phrased responses (see details in Table 2). The unit of analysis was the individual response. To extract the subtle perceptions from the subjects' phrased responses, we specifically coded tone, meaning making, emotion evocation, awareness of imagery, and expression of empathy in the phrased responses. We based our coding on three assumptions. The first was that people who were asked to immediately phrase comments would be unlikely to strategically choose words to conceal their true appraisals. The second was that responses in a non-neutral (vs. neutral) tone and exhibiting more meaning making, emotion evocation, awareness of imagery, and expression of empathy would be associated with higher levels of appreciation, as comprehensive aesthetic appreciation of literature and art needs viewers' imagination, knowledge, mental state, and the text (Shu, 2018). Third, as people generally appreciate high-quality rather than low-quality materials (Douglas & Hargadon, 2000), we assumed that the level of the subjects' appreciation level would reflect their implicit appraisal of the content.

Table 2
Coding scheme and examples of implicit perceptions in content analysis.

Coding Scheme Variables	Definition	Attribution
Tone of response	General feeling toward the content.	1 = positive, 2 = negative, 3 = mixed, 4 = neutral
Meaning making	Whether a participant believed the content was meaningful, which was the premise of appreciation.	1 = yes, 2 = no
Emotion evocation	Emotional response to the content.	1 = yes, 2 = no
Awareness of imagery	Imagery allows readers "to bring their associations to understand and truly experience a new emotion" (Kao & Jurafsky, 2012, p. 11). Following Kao and Jurafsky's (2012) computational aesthetics coding procedure, we coded this measure based on whether a response contained concrete details rather than the abstractions and generalizations described in the poem or painting.	1 = yes, 2 = no
Expression of empathy	Whether a participant expressed empathy with the author or painter, since poetry and art should engage viewers to mentally simulate and affectively resonate with the depicted state of affairs (Lüdtke et al., 2014).	1 = yes, 2 = no

Examples

Phrase Response	Coding Results	
(American) "its not bad."	TR = 1 EE = 2 EM = 2	MM = 1 AI = 2
(American) "My first thought was green goo. Then Van Gogh popped into my mind. It has a certain ""boldness"" but I am not a fan of modern, impressionist art. It certainly requires no talent to smear green, yellow and black paint on a canvas. IMO … I am a devotee of 19th century neoclassicism. William-Adolphe Bouguereau would be an example …"	TR = 3 EE = 1 EM = 2	MM = 1 AI = 2
(Chinese) "[I] felt the author's praise to mothers, great mothers."	TR = 1 EE = 1 EM = 1	MM = 1 AI = 1
(Chinese) "It should be misty poetry. I've read many in college, and my classmates also wrote them. To be honest, I don't quite understand them back then; and I feel the same way now. It uses approaches of metaphors. You may sense some smell, some appeal, some imagery beauty, and some hope. It's about dream and chase of soul."	TR = 4 EE = 1 EM = 2	MM = 1 AI = 1

Note: "TR" represents "Tone of response", "MM" represents "Meaning making", "EE" represents "Emotion evocation", "AI" represents "Awareness of imagery", and "EM" represents "Expression of empathy".

3.6. Coder training and inter-coder reliability

Two coders were trained to perform the content analysis. After three training sessions based on 10% of the sample each time, the inter-coder reliability coefficients had reached an acceptable level of 0.70 (Lombard, Snyder-Duch, & Bracken, 2002). The inter-coder reliability coefficients for all of the items were as follows: 1) tone, kappa = .88; 2) meaning making, kappa = .94; 3) emotion evocation, kappa = .88; 4) awareness of imagery, kappa = .91; and 5) expression of empathy, kappa = .74.

3.7. Data analysis

To test H1 and H3, a series of t-tests were conducted to compare the perceived a) quality, b) imaginativeness, c) spatial presence, d) empathy, and e) writer's competence in the AI- and human-generated content conditions. To test H2 and H4, two-way chi-square tests were conducted to compare b) the tone of the comments and the levels of c) meaning making, d) emotion evocation, e) awareness of imagery, and f) expression of empathy in the two conditions. Data analyses were run on the Chinese and U.S. data separately to detect potentially different patterns in the two sets of responses. To provide a baseline for comparison, we also ran a series of t-tests to compare the U.S. and Chinese responses regarding a) quality, b) imaginativeness, c) spatial presence, d) empathy, and e) writer's competence. As different poems were used, we caution readers that a direct comparison of the poems' conditions may not be meaningful.

4. Results

4.1. American vs. Chinese subjects' appraisals

To provide a baseline, we first compared American and Chinese subjects' appraisals. Overall, the U.S. participants' attitudes toward AI were less positive (M = 3.87, SD = 0.77) than those of their Chinese counterparts (M = 4.07, SD = 0.50). The appraisals of the AI-authored painting were similar across the two subsets of subjects, but American subjects felt significantly more spatial presence and empathy with human-authored one than their Chinese counterparts (see Table 3).

A series of independent-sample t-tests indicated American participants and Chinese participants evaluated the poems similarly. Except perceived imaginativeness ($M_{Am} = 3.42$, SD = 1.10; $M_{Ch} = 3.68$, SD = 0.84; t(462.24) = 2.99, $p < .01$), there existed no significant difference between the evaluations in perceived quality ($M_{Am} = 3.17$, SD = 0.97; $M_{Ch} = 3.18$, SD = 0.72; t(455.90) = 0.20, p = .84), spatial presence ($M_{Am} = 2.90$, SD = 1.30; $M_{Ch} = 2.82$, SD = 0.96; t (452.84) = 0.81, p = .42), empathy ($M_{Am} = 3.04$, SD = 1.26; $M_{Ch} = 2.89$, SD = 0.94; t(455.35) = 1.55, p = .12), and writer's competence ($M_{Am} = 3.30$, SD = 0.99; $M_{Ch} = 3.32$, SD = 0.78; t(473.48) = 0.25, p = .80) from American and Chinese participants(See Figs. 1 and 2).

4.2. American subjects' explicit vs. implicit appraisals

Across the two genres of content, the American subjects rated the perceived imaginativeness (M = 3.29, SD = 1.12), empathy (M = 2.80, SD = 1.29), and competence (M = 3.12, SD = 1.03) of the AI-generated content significantly lower than those of the human-generated content ($M_{imaginativeness} = 3.57$, SD = 0.93; $M_{empathy} = 3.31$, SD = 1.18; $M_{competence} = 3.49$, SD = 0.90). They took a more non-neutral tone, either positive or negative, and a less mixed tone toward the human-generated content than the AI-generated content: $\chi^2(3) = 11.49$, Cramer's V = 0.22, $p < .01$. More empathy was expressed after viewing the human-generated content than the AI-generated content: $\chi^2(1) = 4.71$, Cramer's V = 0.14, $p < .05$.

For the American subjects who read poem, the only difference between the poems lay in the poet's competence (see Table 3). Those who read the human-authored poem (M = 3.34, SD = 0.90) rated the poet's competence higher than those who read the AI-authored poem (M = 2.91, SD = 1.05), t(119) = 2.38, $p < .05$. In parallel, based on their phrased responses (see Table 4a), compared with those who read the human-authored poem, those who read the AI-authored poem tended to take a more mixed tone (χ^2 (3) = 9.30, Cramer's V = 0.28, $p < .05$) with less evocation of emotions (χ^2 (1) = 4.49, Cramer's V = 0.19, $p < .05$) and less expression of empathy (χ^2 (1) = 4.82, Cramer's V = 0.20, $p < .05$).

Of the American subjects in the painting viewing groups, those in the human-authored group rated spatial presence (t(128) = 2.09, $p < .05$), empathy (t(128) = 3.23, $p < .05$), and painter's competence (t (128) = 1.98, $p < .05$) significantly higher than those in the AI-authored group (see Table 3). Yet based on their phrased responses, the levels of c) sense making, d) emotion evocation, e) awareness of imagery, and f) expression of empathy did not differ significantly between the two conditions (see Table 4b). Therefore, H1 and H2(a) were partially supported, but H2(b) was not.

4.3. Chinese subjects' explicit vs. implicit appraisals

Across two genres of content (see Table 3), the Chinese participants rated the perceived imaginativeness (M = 3.79, SD = 0.74) of the AI-generated content significantly higher than that of the human-generated content (M = 3.57, SD = 0.90). They noticed more imagery in the human-generated content than in the AI-generated content: χ^2 (1) = 9.95, Cramer's V = 0.19, $p < .01$.

For the Chinese subjects who read the poem (see Table 3), those in the AI-authored group rated perceived a) quality (t(140) = 2.08, $p < .05$), b) imaginativeness (t(140) = 3.17, $p < .01$), e) empathy (t (140) = 2.55, $p < .05$), and e) writer's competence (t(140) = 2.22, $p < .05$) significantly higher than those in the human-authored group. In parallel, based on their phrased responses (see Table 5a), compared with those who read the AI-authored poem, those who read the human-authored one noticed more imagery (χ^2 (1) = 5.58, Cramer's V = 0.20, $p < .05$).

The Chinese subjects in the painting viewing groups reported no significant differences in perceived a) quality, b) imaginativeness, c) spatial presence, d) empathy, or e) painter's competence (see Table 3). However, based on their phrased responses (see Table 5b), compared with those who viewed the AI-drawn painting, those who viewed the human-generated one tended to notice more imagery (χ^2 (1) = 4.45, Cramer's V = 0.17, $p < .05$). Therefore, H3(a) and H4 were supported, but H3(b) was not.

5. Discussion

5.1. Summary of findings

In response to the inconsistent findings in previous research regarding perceptions of AI-generated content, this study conducted an experimental survey to compare subjects' explicit and implicit perceptions of poems and paintings generated by AI versus human authors. Overall, this study has shown that American subjects appraised human-generated poem and painting more favorably than AI-generated ones, both explicitly and implicitly. In particular, their implicit attitudes toward AI suggested that they had mixed feelings about the AI authorship and developed more empathy with human-authored poems. When it comes to paintings, American participants assigned higher ratings to

Table 3
The Means and Standard Deviations (in parentheses) of Key Dependent Variables.

	American						Chinese					
	Overall		Poem		Painting		Overall		Poem		Painting	
	AI (n = 133)	Human (n = 118)	AI (n = 64)	Human (n = 57)	AI (n = 69)	Human (n = 61)	AI (n = 140)	Human (n = 153)	AI (n = 68)	Human (n = 75)	AI (n = 72)	Human (n = 78)
Quality	3.07^{a}	3.28^{a}	2.87^{a}	3.15^{a}	3.26^{a}	3.40^{a}	3.23^{a}	3.14^{a}	3.22^{b}	2.95^{a}	3.25^{a}	3.32^{a}
	(.99)	(.93)	(1.04)	(.98)	(.92)	(.88)	(.77)	(.67)	(.83)	(.69)	(.73)	(.60)
Imaginativeness	3.29^{a}	3.57^{b}	3.05^{a}	3.32^{a}	3.52^{a}	3.80^{a}	3.79^{c}	3.57^{b}	3.78^{b}	3.36^{a}	3.81^{a}	3.77^{a}
	(1.12)	(1.06)	(1.16	(1.10)	(1.05)	(.97)	(.74)	(.90)	(.76)	(.83)	(.73)	(.93)
Presence	2.77^{a}	3.04^{a}	2.73^{a}	2.79^{a}	2.81^{a}	3.27^{b}	2.84^{a}	2.79^{a}	3.04^{a}	2.75^{a}	2.65^{a}	2.84^{a}
	(1.32)	(1.26)	(1.31)	(1.29)	(1.34)	(1.19)	(.99)	(.92)	(.97)	(.96)	(.98)	(.89)
Empathy	2.80^{a}	3.31^{b}	2.91^{a}	3.24^{a}	2.69^{a}	3.39^{b}	2.95^{a}	2.83^{a}	3.15^{a}	2.72^{b}	2.77^{a}	2.94^{a}
	(1.29)	(1.18)	(1.31)	(1.18)	(1.27)	(1.18)	(.97)	(.91)	(1.01)	(1.00)	(.90)	(.80)
Competence	3.12^{a}	3.49^{b}	2.91^{a}	3.34^{b}	3.31^{a}	3.64^{b}	3.37^{c}	3.27^{c}	3.31^{c}	2.99^{d}	3.42^{ab}	3.54^{ab}
	(1.03)	(.90)	(1.05)	(.90)	(.98)	(.89)	(.78)	(.78)	(.89)	(.81)	(.67)	(.67)

Notes: In each row, within the same genre, comparisons have been made between AI-author for American, human-author for American, AI-author for Chinese, and human-author for Chinese. The letters of "a," "b," "c," and "d" in the superscripts refer to significantly different means; and "ab" means no significant difference with either "a" or "b".

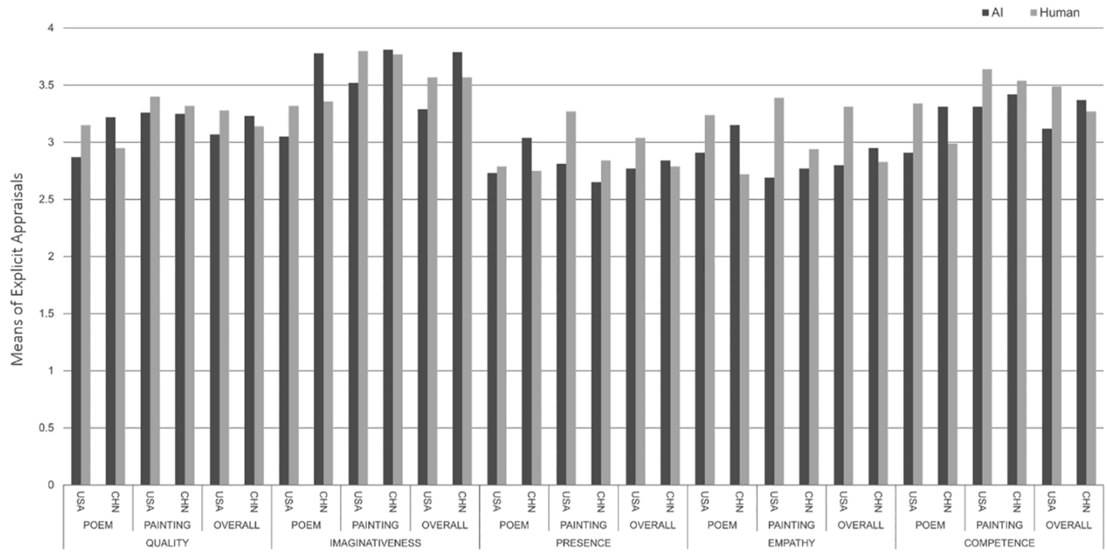

Fig. 1. American and Chinese subjects' explicit appraisals.

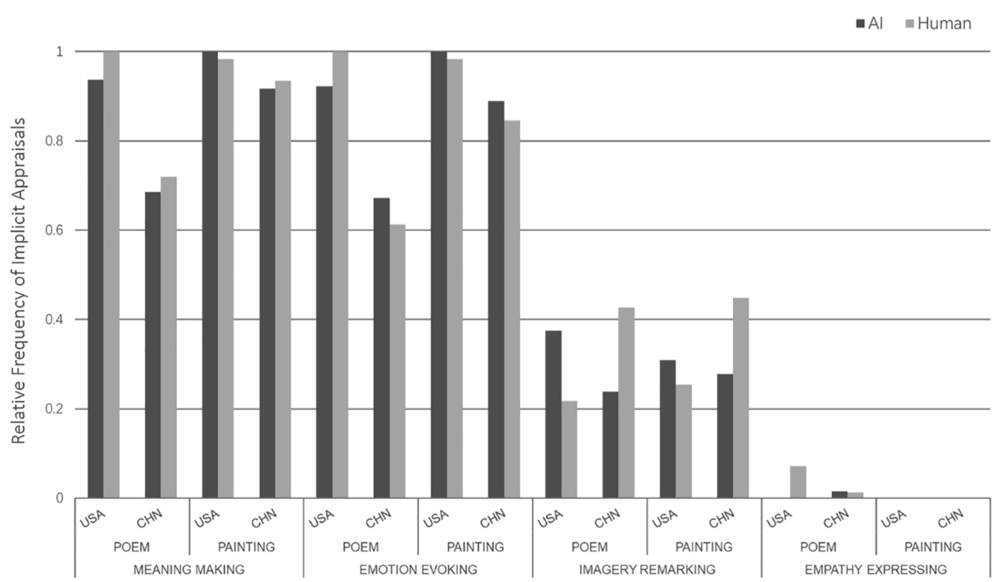

Fig. 2. American and Chinese Subjects' Implicit Appraisals.
Note: y axis. = relative frequency of implicit appraisals $= \frac{\text{the number of the ``Yes'' responses}}{\text{the number of the total valid responses}}$

Table 4a
Cross tabulation between authorship and coding variables in American responses in poem conditions.

		Tone				Meaning making		Emotion evoking		Imagery remarking		Empathy expressing		Total
		Pos	Neg	Mixed	Neutral	Yes	No	Yes	No	Yes	No	Yes	No	
Authorship	AI	13	35	12	4	60	4	59	5	24	40	0	64	64
	Human	11	37	1	6	55	0	55	0	12	43	4	51	55
Total		24	72	13	10	115	4	114	5	36	83	4	115	119
χ^2		χ^2 (3) = 9.30, Cramer's V = .28, p < .05				χ^2 (1) = 3.56, Cramer's V = .17, p = .06		χ^2 (1) = 4.49, Cramer's V = .19, p < .05		χ^2 (1) = 3.45, Cramer's V = .17, p = .06		χ^2 (1) = 4.82, Cramer's V = .20, p < .05		

human-authored paintings on spatial presence, empathy, and competence than to AI-generated paintings. Their implicit attitudes, nevertheless, indicated no significant differences in tones, meaning making, emotion evoking, imagery remarking, and empathy with the authorship when evaluating the AI- vs. human-authored paintings.

These findings imply that despite the similar quality of AI-generated and human-generated poems and their concomitant spatial presence experiences, American participants still believed that human authors were more capable and skilled in writing poems than AI. Their implicit perception is congruent with their explicit evaluation, in which they could more easily identify with human authors. The findings were in line with the results of a general poll regarding AI and automation in the U. S., which underscored US citizens' uncertainty and fear surrounding AI (Manyika et al., 2017). US participants' concern about AI was more

Table 4b
Cross tabulation between authorship and coding variables in American responses in painting conditions.

		Tone				Meaning making		Emotion evoking		Imagery remarking		Empathy expressing		Total
		Pos	Neg	Mixed	Neutral	Yes	No	Yes	No	Yes	No	Yes	No	
Authorship	AI	40	10	3	15	68	0	68	0	21	47	0	59	68
	Human	43	7	0	9	58	1	58	1	15	44	0	68	59
Total		83	17	3	24	126	1	126	1	36	91	0	127	127
χ^2		χ^2 (3) = 4.52, Cramer's V = .19, NS				χ^2 (1) = 1.16, Cramer's V = .10, NS		χ^2 (1) = 1.16, Cramer's V = .10, NS		χ^2 (1) = .46, Cramer's V = .06, NS		–		

Table 5a
Cross tabulation between authorship and coding variables in Chinese responses in poem conditions.

		Tone				Meaning making		Emotion evoking		Imagery remarking		Empathy expressing		Total
		Pos	Neg	Mixed	Neutral	Yes	No	Yes	No	Yes	No	Yes	No	
Authorship	AI	32	23	1	11	46	21	45	22	16	51	1	66	67
	Human	33	21	2	19	54	21	46	29	32	43	1	74	75
Total		65	44	3	30	100	42	91	51	48	94	2	140	142
χ^2		χ^2 (3) = 2.13, Cramer's V = .12, NS				χ^2 (1) = .19, Cramer's V = .04, NS		χ^2 (1) = .52, Cramer's V = .06, NS		χ^2 (1) = 5.58, Cramer's V = .20, $p < .05$		χ^2 (1) = .01, Cramer's V = .01, NS		

Table 5b
Cross tabulation between authorship and coding variables in Chinese responses in painting conditions.

		Tone				Meaning making		Emotion evoking		Imagery remarking		Empathy expressing		Total
		Pos	Neg	Mixed	Neutral	Yes	No	Yes	No	Yes	No	Yes	No	
Authorship	AI	34	11	4	22	66	5	64	7	20	51	0	72	72
	Human	38	12	5	23	73	5	66	12	35	43	0	78	78
Total		72	23	9	45	139	10	130	19	55	94	0	150	150
χ^2		χ^2 (3) = .07, Cramer's V = .02, NS				χ^2 (1) = .02, Cramer's V = .01, NS		χ^2 (1) = 1.02, Cramer's V = .08, NS		χ^2 (1) = 4.45, Cramer's V = .17, $p < .05$		–		

obviously highlighted in their attitudes toward machine-produced paintings. They explicitly expressed that they were immersed in a different space and shared feelings with human authors. However, their implicit responses might have disclosed their long-term attitudes, in which AI would one day have the ability to achieve humans' accomplishments even in the areas of arts.

Furthermore, one would normally expect a human-authored poem to make more sense than an AI-authored one, as machines would have (as yet) no awareness of what they write. However, meaning making was found to be the same in AI-generated poems as in human-generated ones. US participants' implicit attitudes toward the emotional aspects and the imagery created by AI and humans in both poems and paintings were also similar. A hint of possible reason was that many US participants were surprised at how well AI could paint. Their recognition of AI's talent in their phrased responses might have led them to appreciate the AI paintings as much as the human's works. All of these responses suggest that today's AI technology has achieved at least some degree of success in mimicking human's creative writing and art work.

However, the patterns of the findings flipped in another culture context. Compared to US participants, Chinese participants perceived the AI-generated poem to have higher quality and imaginativeness. They also reported higher empathy with the AI authors and considered machines as more competent. But the analyses of their implicit attitudes toward AI suggested that Chinese participants attributed more imagery to human-generated poems than AI-generated ones. When evaluating the human-vs. AI-authored paintings, Chinese participants did not report significantly different attitudes toward them. But similar to their assessment of the poems, they assigned more imagery to the human-generated paintings.

Despite the substantial differences, it would not be hard to acknowledge that cultural contexts have direct and powerful influence on users' perception of AI technologies. The wide adoption of AI technologies in China including Microsoft chatbot XiaoIce, facial recognition, and mobile payment may have allowed for Chinese users to have more positive use experiences of AI. Considering that it is not rare to be exposed to AI technologies that include but are not limited to beauty cameras, real-time on-site food ordering systems, and automatic pianos in shopping malls, participants' expectations for AI-generated art works may have been elevated to a higher level. This pattern was more salient in the poem conditions than in the painting ones, probably due to users' different appreciation processing. Poems could be more provocative in eliciting readers' affective responses. However, paintings, especially abstract paintings adopted in the current experiment, may take more time to interpret and digest. For those who do not have much art appreciation experiences, their responses to paintings may not differ much between two similar painting styles.

While seemingly Chinese participants would appreciate the artificially intelligent devices that bring about convenience, efficiency, and comfort, their implicit responses have provided more insights into their relationship with machines. Chinese participants have described more imagery in both human-authored poems and paintings. The discrepancy between their explicit and implicit perceptions may have been caused by the bandwagon effect or the spiral of silence mechanism. That is, people may identify more with the majority positions and views in the society (Schmitt-Beck, 2015). As public opinions in China are largely guided by the authorities, the optimism about AI technology promoted by the government has become the majority opinion (Kania, 2018). Hence, despite their distrust and unease of machines (e.g., Bartneck, Suzuki, Kanda, & Nomura, 2007), individuals in China openly favor this technology trend, but simultaneously hold a hesitant perspective of AI, as the implicit attitudes cannot be easily changed or formed after a single message exposure (Bekker, Fischer, Tobi, & van Trijp, 2017). The findings are also consistent with Hetts, Sakuma, Pelham. (1999) argument that individuals tend to endorse explicit self-evaluations that are aligned with their current cultural context, whereas their implicit self-evaluations bear the mark of their long-term interpretations and beliefs.

5.2. Theoretical and practical implications

Theoretically, this study enriches our understanding of implicit and explicit cognition in a context without prejudice. Prejudice is not a requirement for inconsistencies between implicit and explicit perceptions. Rather, such inconsistencies are widely observed. As Nosek (2007) posited, implicit and explicit attitudes express the distinction between mental processing and mental experience. The former is how the mind operates, and the latter is the subjective experience emerging from the former. Correlations between implicit and explicit attitudes in various domains range from weakly positive (below 0.20, e.g., attitudes toward Caucasians and Asian Americans) to strongly positive (above 0.75, e.g., pro-choice vs. pro-life beliefs) (Nosek, 2005; Nosek & Smyth, 2007). In any case, explicit and implicit measures "seem to tap different constructs" (Castelli et al., 2001, p. 424).

The findings of this study seem inconsistent with the Computers Are Social Actors paradigm (Nass, Fogg, & Moon, 1996; Nass, Steuer, & Tauber, 1994), according to which people treat computers and televisions like real people. For instance, a mock-up company's human and computer representatives would elicit similar reactions toward the representative and the organization (Shank, 2013). However, as AI technology advances, people start to ascribe both agency and experience to machines (Gray & Wegner, 2012). Obviously, the moral authenticity of AI work is generally questioned, even though the type authenticity can be achieved (Jago, 2019). van der Kaa and Krahmer (2014) suggested that readers' initial expectations regarding the quality of viewed content may affect their perceptions of quality. If they have low expectations and are positively surprised by the quality, they may assign higher ratings. However, if their expectations are not fulfilled, they may assign lower ratings to the content. Without any knowledge of the subjects' initial expectations prior to their encounter with the AI-generated content, it would be difficult to recognize whether participants treated machines and humans equally. Thus, future research on human-AI interaction should continue to explore how individuals' expectations shape their evaluations of the AI-generated works.

The practical implications of this study are also salient. Prior literature has indicated that human-AI interaction is affected by humans' implicit, not their explicit, attitudes toward robots (Mirnig, Strasser, Weiss, & Tscheligi, 2012; Strasser, Weiss, & Tscheligi, 2012). As the divergence between those two types of cognition may be very large, as indicated in this study, policy makers should base their decisions on more accurately assessed measures of public opinion of AI on this matter. After all, public acceptance has been identified as key to the successful implementation of AI technology (Heerink, Kröse, Evers, & Wielinga, 2010). In particular, caution should be raised about an AI "craze" generated through media hype in Chinese society, as previously documented in Japan. For instance, MacDorman, Vasudevan, and Ho (2009) used both implicit and explicit measures to compare Japanese and U.S. faculty's attitudes toward robots. Despite the stereotypical portrayal of Japan's "robot mania," the implicit measures showed that both the Japanese and the U.S. faculty had more pleasant associations with humans than with robots, and associated weapons more strongly with robots than with humans.

5.3. Limitations

This study is not without limitations. The first lies in the selection of experimental stimuli. Stimulus materials are highly subject to researchers' idiosyncratic preferences—a common drawback of experimental designs. Although these selected works have been published or awarded an industry prize, a larger sample of these AI works are still needed to serve as experimental stimuli (Reeves, Yeykelis, & Cummings, 2016). Second, although we tried to select similar sub-samples of U.S. and Chinese subjects, their demographic differences were still salient. For instance, the U.S. subjects were on average 5 years older than their Chinese counterparts. This may partially explain their more conservative attitude toward AI, as younger people are found to be more likely to adopt new technology (Rogers, 2003). Edwards, Edwards, Stoll, Lin, and Massey (2019) also found that age could influence users' trust in AI technologies. Yet, the percentage of females was larger in the Chinese sub-sample than the American one. It might counterbalance the skewness brought by the age difference, as females are found to be less open to new technology (Rogers, 2003). Third, in this study we used a genderless name for AI, but a male name for human poet/painter. Subjects may form different impressions based on the gender itself, as gender stereotypes are salient in both humans and robots (e.g., McCauley & Thangavelu, 1991; Tay, Jung, & Park, 2014).

6. Conclusion

As the prevalence of AI-generated content increases, examining viewers' perceptions of this content is crucial to understand the human-machine relationship and further facilitate efficient human-machine collaboration. Prior literature has accumulated mixed findings regarding subjects' attitudes toward and perceptions of news and tweets written by NLG algorithms. To resolve this inconsistency, this study investigated the explicit and implicit perceptions of AI-generated poetry and painting held by subjects from two different cultural backgrounds.

This study is one of the first attempts to probe humans' true appraisal of AI's performance on literature and art work. As the U.S. and China fiercely compete to lead the development of AI technology, their citizens exhibit divergent attitudes toward AI's performance in NLG and the creation of art. The U.S. subjects were more critical of the AI- than the human-generated content, both explicitly and implicitly. Although the Chinese subjects were overtly affirmative about the AI-generated content, they implicitly accepted less AI-generated content than human-generated content. The findings can enrich our understanding in the domain of AI generation. As researchers have predicted that human's future is entangled with AI development (Tegmark, 2017), an accurate assessment of human's cognition of AI is in order. Future research should continue to probe the antecedents and outcomes of people's attitudes toward and perceptions of AI and AI-generated content.

Acknowledgement

This work was supported by the National Social Science Fund of China under Grant No. 18BXW046.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly, 93*(1), 59–79.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1.

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society, 21*, 217–230.

Bekker, G. A., Fischer, A. R. H., Tobi, H., & van Trijp, H. C. M. (2017). Explicit and implicit attitude toward an emerging food technology: The case of cultured meat. *Appetite, 108*, 245–254.

Bornstein, R. F., & Pittman, T. S. (1992). Perception without awareness: Cognitive, clinical, and social perspectives. In *Perception without aware-ness: Cognitive, clinical and social perspectives*. Gettysburg, PA: Guilford Press.

Carroll, G. R., & Wheaton, D. R. (2009). The organizational construction of authenticity: An examination of contemporary food and dining in the U.S. *Research in Organizational Behavior, 29*, 255–282.

Castelli, L., Vanzetto, K., Sherman, S. J., & Arcuri, L. (2001). The explicit and implicit perception of in-group members who use stereotypes: Blatant rejection but subtle conformity. *Journal of Experimental Social Psychology, 37*(5), 419–426.
Caswell, D., & Dörr, K. (2018). Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories. *Journalism Practice, 12*(4), 477–496.
Ch'ng, E. (2019). Art by computing machinery: Is machine art acceptable in the artworld? *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 15*(2s), 59.
Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 151–192). Boston: McGraw-Hill.
Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice, 8*(5), 519–531.
Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*(2), 163–170.
Dabai. (2017, May 18). *RobotArt 2017 | These paintings are drawn by robots, will you collect them?*. Retrieved from https://www.zhinengl.com/2017/05/robotart-2017/.
Douglas, Y., & Hargadon, A. (2000). May). The pleasure principle: Immersion, engagement, flow. In *Proceedings of the eleventh ACM on Hypertext and hypermedia* (pp. 153–160). ACM.
Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on twitter. *Computers in Human Behavior, 33*, 372–376.
Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social identity theory in human-robot interactions. *Computers in Human Behavior, 90*, 357–362.
Eurobarometer. (2012). *Public attitudes towards robots. Special Eurobarometer 382*. European Commission. Retrieved from http://ec.europa.edu/public_opinion/archives/ebs/ebs_382_en.pdf.
Falcon, W. (2018). What Happens Now That An AI-Generated Painting Sold For $432,500?. Forbes. Retrieved from https://www.forbes.com/sites/williamfalcon/2018/10/25/what-happens-now-that-an-ai-generated-painting-sold-for-432500/. October 25.
Future of Life Institute. (2015). An Open Letter: Research priorities for robust and beneficial artificial intelligence. Retrieved from https://futureoflife.org/ai-open-letter/.
Gayford, M. (2016). *Robot art raises questions about human creativity*. MIT Technology Review. Retrieved from https://www.technologyreview.com/s/600762/robot-art-raises-questions-about-human-creativity/.
Geng, L. (2018, January 22). Status and Prospect of Robot-written news [in Chinese]. Retrieved from http://zgbx.people.com.cn/n1/2018/0122/c415418-29778770.html.
Graefe, A., Haim, M., Haarmann, B., & Brosius, H. B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism, 19*(5), 595–610.
Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*, 125–130.
Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464.
Haim, M., & Graefe, A. (2017). Automated News: Better than expected? *Digital Journalism, 5*(8), 1044–1059.
Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: the almere model. *International Journal of Social Robotics, 2*(4), 361–375.
Hetts, J. J., Sakuma, M., & Pelham, B. W. (1999). Two roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology, 35*, 512–559.
Hofstadter, D. R. (2001). Staring Emmy straight in the eye – And doing my best not to flinch. In D. Cope (Ed.), *Virtual music: Computer synthesis of musical style* (pp. 33–82). Cambridge, MA: MIT Press.
Jago, A. S. (2019). Algorithms and authenticity. *Academy of Management Discoveries, 5*(1), 38–56.
Jin, S. A. (2011). "My avatar behaves well and this feels right": Ideal and ought selves in video gaming. *Social Behavior and Personality: An International Journal, 39*(9), 1175–1182.
Kania, E. (2018). China's AI Agenda Advances. The Diplomat. Retrieved from https://thediplomat.com/2018/02/chinas-ai-agenda-advances/. February 14.
Kao, J., & Jurafsky, D. (2012). A computational analysis of style, affect, and imagery in contemporary poetry. In *Proceedings of the NAACL-HLT 2012 workshop on computational linguistics for literature* (pp. 8–17).
Kihlstrom, J. F., Barnhardt, T. M., & Tataryn, D. J. (1992). *Implicit perception. Perception without awareness: Cognitive, clinical, and social perspectives*.
Lambert, A. J., Cronen, S., Chasteen, A. L., & Lickel, B. (1996). Private vs public expressions of racial prejudice. *Journal of Experimental Social Psychology, 32*(5), 437–459.
Latar, N. L. (2015). The robot journalist in the age of social physics: The end of human journalism?. In *The new world of transitioned media* (pp. 65–80). Cham: Springer.
Liao, S. (2017, December 12). *This Harry Potter AI-generated fanfiction is remarkably good*. The Verge. Retrieved from https://www.theverge.com/2017/12/12/16768582/harry-potter-ai-fanfiction.
Lombard, M., Ditton, T. B., & Weinstein, L. (2009, October). Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence* (pp. 1–15).
Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*, 587–604.
Lüdtke, J., Meyer-Sickendieck, B., & Jacobs, A. M. (2014). Immersing in the stillness of an early morning: Testing the mood empathy hypothesis of poetry reception. *Psychology of Aesthetics, Creativity, and the Arts, 8*(3), 363.
Maass, A., Castelli, L., & Arcuri, L. (2000). Measuring prejudice: Implicit versus explicit techniques. In *Social identity processes: Trends in theory and research* (pp. 96–116).
Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*(6), 981.
MacDorman, K. F., Vasudevan, S. K., & Ho, C. C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society, 23*(4), 485–510.
Malahy, L. W., Sedlins, M., Plaks, J., & Shoda, Y. (2010). Black and White, or shades of gray? Racial labeling of Barack Obama predicts implicit race perception. *Analyses of Social Issues and Public Policy, 10*(1), 207–222.
Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., et al. (2017). *Harnessing automation for a future that works*. McKinsey Global Institute.
Markoff, J. (2016). *Machines of loving grace: The quest for common ground between humans and robots*. Harper Collins Publishers.
McCauley, C., & Thangavelu, K. (1991). Individual differences in sex stereotyping of occupations and personality traits. *Social Psychology Quarterly, 54*(3), 267–279.
McNeice, A. (2018, June, 22). *China opens hearts and minds to artificial intelligence at work*. China Daily. Retrieved from http://www.chinadaily.com.cn/a/201806/22/WS5b2bd9b1a3103349141dda0d.html.
Mirnig, N., Strasser, E., Weiss, A., & Tscheligi, M. (2012, October). Studies in public places as a means to positively influence people's attitude towards robots. In *International conference on social robotics* (pp. 209–218). Berlin, Heidelberg: Springer.
Moniz, A. B., & Krings, B. J. (2016). Robots working with humans or humans working with robots? Searching for social dimensions in new human-robot interaction in industry. *Societies, 6*(3), 23.
Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65*(3), 469.
Mutz, D. C. (1998). *Impersonal influence: How perceptions of mass collectives affect political attitudes*. Cambridge University Press.
Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*, 669e678.
Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Human Factors in Computing Systems, 94*, 72e78.
iFeng News. (2017, May 20). Microsoft XiaoIce published its poetry collection Sunlight Lost the Window Glass [in Chinese]. Retrieved from http://tech.ifeng.com/a/20170520/44616613_0.shtml.
Noelle-Neumann, E. (1991). The theory of public opinion: The concept of the spiral of silence. *Annals of the International Communication Association, 14*(1), 256–287.
Northeastern University, & Gallup. (2018). *Optimism and anxiety: Views on the impact of artificial intelligence and higher education's response*. Retrieved from https://perma.cc/57NW-XCQN.
Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*(4), 565.
Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science, 16*(2), 65–69.
Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test. *Experimental Psychology, 54*(1), 14–29.
Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*(5), 413–417.
Petty, R. E., & Wegener, D. T. (1998). Matching versus mismatching attitude functions: Implications for scrutiny of persuasive messages. *Personality and Social Psychology Bulletin, 24*(3), 227–240.
Pressman, L. (2017, February 28). *The automated future of journalism*. Retrieved from https://automatedinsights.com/blog/the-automated-future-of-journalism/.
Reeves, B., Yeykelis, L., & Cummings, J. (2016). The use of media in media psychology. *Media Psychology, 19*, 49–71.
Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*(6), 995.
Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science, 17*(11), 954–958.
Saleh, B., & Elgammal, A. (2015). *Large-scale classification of fine-art paintings: Learning the right metric on the right feature*. arXiv preprint arXiv: 1505.00855.
Schaub, M. (2016, March 22). *Is the future award-winning novelist a writing robot?* Los Angeles Times. Retrieved from https://www.latimes.com/books/jacketcopy/la-et-jc-novel-computer-writing-japan-20160322-story.html.
Schmid, J., & Fiedler, K. (1996). Language and implicit attributions in the nuremberg trials analyzing prosecutors' and defense attorneys' closing speeches. *Human Communication Research, 22*(3), 371–398.
Schmitt-Beck, R. (2015). *Bandwagon effect. The international encyclopedia of political communication* (pp. 1–5).

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54*(4), 558.

Shank, D. B. (2013). Are computers good or bad for business? How mediated customer-computer interaction alters emotions, impressions, and patronage toward organizations. *Computers in Human Behavior, 29*, 715–725.

Shu, L. (2018). Exploring "fabula" and "sjuzhet" in classical Chinese poetry from the perspective of cognitive poetics. *Neohelicon, 45*(2), 505–516.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.

Strasser, E., Weiss, A., & Tscheligi, M. (2012, March). Affect misattribution procedure: an implicit technique to measure user experience in hri. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 243–244). ACM.

Sundar, S. S. (1999). Exploring receivers' criteria for perception of print and online news. *Journalism & Mass Communication Quarterly, 76*(2), 373–386.

Sundar, S. S. (2000). Multimedia effects on processing and perception of online news: A study of picture, audio, and video downloads. *Journalism & Mass Communication Quarterly, 77*(3), 480–499.

Sundar, S. S. (2008). *The main model: A heuristic approach to understanding technology effects on credibility*. Digital media, youth, and credibility, 73100.

Sundar, S. S., Oeldorf-Hirsch, A., & Xu, Q. (2008, April). The bandwagon effect of collaborative filtering technology. In *CHI'08 extended abstracts on Human factors in computing systems* (pp. 3453–3458). ACM.

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior, 38*, 75–84.

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Knopf.

Van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice, 6*(5–6), 648–658.

Van der Kaa, H., & Krahmer, E. (2014). Journalist versus news consumer: The perceived credibility of machine written news. In *Proceedings of the computation + journalism conference* (Vol. 24, p. 25). New York: Columbia University.

Von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology, 33*(5), 490–509.

Waddell, T. F. (2018). A Robot Wrote This? How perceived machine authorship affects news credibility. *Digital Journalism, 6*(2), 236–255.

Xinhua News. (2018, February, 8). *Chinese people most optimistic about AI impact on job market: Survey*. Retrieved from http://www.xinhuanet.com/english/2018-02/08/c_136959614.htm.

Xinhua News. (2019, July 16). *AI robot holds solo art show in Beijing*. Retrieved from http://zjlieyan.com/english/2019-07/16/c_138231273.htm.

Yeung, H. H. (2015). *Spatial engagement with poetry*. New York: Palgrave MacMillan.