JOURNAL OF BROADCASTING & ELECTRONIC MEDIA
https://doi.org/10.1080/08838151.2020.1835136

Using Machine Learning to Learn Machines: A Cross-Cultural Study of Users' Responses to Machine-Generated Artworks

Kun Xu[a], Fanjue Liu[a], Yi Mou[b], Yuheng Wu[b], Jing Zeng[c], and Mike S. Schäfer[c]

[a]College of Journalism and Communications, University of Florida; [b]School of Media and Communication, Shanghai Jiao Tong University; [c]Department of Communication and Media Research, University of Zurich

ABSTRACT
Drawing from prior literature on machine-generated news, this study examines machine-generated artworks in a cross-cultural context. It combines machine learning approaches with online experiments and investigates how different genres of artworks and different authorship cues influence participants' open-ended responses to machine-generated works. Results suggest that while genres and cultures affected participants' discussion topics and word use, the differences between participants' responses to machine-generated artworks and human-generated ones were not evident. This study tests the explanatory power of machine heuristic and demonstrates the feasibility of integrating multiple methods in future AI-based media research.

> "Human beings have dreams. Even dogs have dreams, but not you, you are just a machine. An imitation of life. Can a robot write a symphony? Can a robot turn a canvas into a beautiful masterpiece?"

- Movie *I, Robot* (2004)

In 1991, Weiser proposed the idea of ubiquitous computing, where he envisioned that computing technologies would one day form a network that is seamless, natural, and omnipresent in our daily life. While Weiser's (1991) vision is yet to be fully realized, various artificial intelligence (AI) programs including AlphaGo and humanoid robot Sophia a are stimulating public discussion about the promises and perils of these technologies. One of the

CONTACT Yi Mou yimou@sjtu.edu.cn

areas that have increasingly adopted AI technology is journalism. In 2016, the Washington Post published over 850 articles that were produced by their AI programs (Moses, 2017). News agencies such as Reuters and the Associated Press have also used algorithms to compile news stories (Van Dalen, 2012). Despite the growing use of robot journalists, what is worrisome is that readers may not be able to tell machine-generated information from human-generated information, whether it is published in newspapers or social media (Clerwall, 2014; Edwards et al., 2014). The difficulty in differentiating information sources may render users susceptible to believing in fake news and consequently making misinformed decisions.

While machines have been utilized to produce news, people used to believe that art is the bulwark of human creativity (Gunkel, 2012). However, in the past few years, machines have been trained to compose poems, paintings, and even symphonies. Elgammal et al. (2017) used creative adversarial networks to generate art images and found that viewers rated machine-created art to be more innovative than human-created art. Machine-generated poems have received growing attention, as they have been collected on websites, literary databases, and published in books (Trentini, 2017). Furthermore, companies including Sony and Google have invested in various AI programs to create stylistic music and songs (Deahl, 2018).

While much research has been conducted to understand AI-generated news (Clerwall, 2014; Waddell, 2018), limited research has focused on readers' perception and responses to AI-generated artworks, especially poems, and paintings. If AI-generated artworks can have the same effects as human-generated ones, the cost of artistic appreciation would be substantially affected, as machines can efficiently generate art based on humans' preferences and simulate those works that are rated as valuable and creative. Meanwhile, evolution of AI technologies could challenge the existing criteria for esthetic evaluation of artworks, which may lead people to revisit and fathom the implications of human creativity and inspiration. Therefore, in this study, we take the first step and examine whether machine-generated creative works could exert the same influence as human-generated ones on viewers' perception, cognition, and affect. The results may inform scholars about both the benefits and risks of using AI to generate artworks.

Unlike prior research on user responses to AI-produced content that has mostly relied on survey experiments with closed-ended questions, this study aims at combining both experiments and machine learning techniques to solve the question. Machine learning refers to the development of algorithms that are used to recognize patterns of data (Colleoni et al., 2014). The combination of these two methods in this study offers the following advantages. First, this study allows participants to produce open-ended responses after viewing AI-generated artworks and enables researchers to find latent topics from user responses. Compared to closed-ended questions, open-ended written responses could provide a richer understanding of user psychological reactions, linguistic patterns, and attitudes toward the machines. With the techniques of machine learning being more mature and accessible than before, analyzing large collections of digital texts has become feasible and practical (Gerrish & Blei, 2012). Second, although machine learning has been used to analyze discussion on social media, rarely have researchers taken into consideration the covariate information or experimental conditions in automated textual analyses (Roberts, Steward, Tingley, Lucas, et al., 2014). To fill the gap, this study adopts an unsupervised machine learning method to scrutinize how topic prevalence and topic content varies as a function of authorship cues and genres of artworks.

Further, as users' understanding of machines could be socially constructed and culturally produced (Pinch & Bijker, 1987; Winner, 1980), it is important to expand the research focus from individual countries to cross-cultural contexts. This study thus compares viewers' responses from three different cultural backgrounds: U.S, Germany, and China. As viewers' attitudes toward machines are situated within their traditional value system, which can inversely be reconstructed through new technology materials and practices (Šabanović, 2014), it is important to understand how cultures may have shaped users' interpretation of machine-generated content.

Above all, this study seeks to make a methodological contribution to the current literature, as machine learning can not only proffer us the information about the semantic topics from users' discussion about machines but also demonstrate how the discussion may be affected by experimental treatment. At a theoretical level, this study tests the explanatory mechanism of machine heuristic (S. Sundar, 2008) and examines whether users' understanding of machines is contingent upon the nature of tasks and socio-cultural factors.

Literature Review

Machine-Generated Works

While algorithm-generated content is not full-fledged, it reflects a new stage of technology evolution and a tendency to reshape and transform the news industry (Zheng et al., 2018). In the area of news writing, algorithms have reconstructed newsrooms and affected the relationships between journalism and its publics (Pavlik, 1999). A series of concerns and questions have been raised regarding how the adoption of AI-generated news may potentially impact traditional journalism and how audiences perceive news coverage written by machines vs. humans.

Some researchers see automated journalism as a threat to human journalists. For instance, Carlson (2015) discussed the development of automated journalism in relation to labor and authority. He pointed out that automated journalism has raised journalists' concerns and fears of displacement. Dörr (2016) noted that at a technological level, algorithms are capable of performing standardized tasks of professional journalism, which was further supported by a follow-up study about the advantage of applying automated journalism in the news industry (Thurman et al., 2017).

As the convergence of computing and news production has begun to alter traditional journalism practices, researchers have attempted to compare the quality of news produced by machines vs. humans. Clerwall (2014) found that participants were not able to tell the differences between software-produced news and human journalist-produced news in terms of their credibility. Consistent with Clerwall's (2014) findings, Van der Kaa and Krahmer (2014) found that human-authored news and algorithm-generated news was almost indiscernible. The boundary between these two types of news became even blurrier as Graefe et al. (2018) in their experiment found that automated news was rated lower in readability but higher in credibility than human-authored news.

The advent of AI-generated news has led scholars to ponder what makes human-authored content unique. Van Dalen (2012) argued that human journalists feature analytical skills, personality, creativity, and the ability to write linguistically complex sentences, while machine-generated content underscores factuality, objectivity, simplification, and speed. Thurman et al. (2017) further pointed out that although automated content has the potential to benefit news organizations and consumers, such developments may not go beyond the fundamental limitations of automated journalism, which lie in their heavy reliance on isolated data streams, lack of contemporaneous contexts, and the difficulty of operating creatively with the data in the production process.

Whereas prior research, as discussed above, has focused on how machine-generated news affects audiences' perception and the news industry, little research has concentrated on artworks such as poems and paintings. Compared to news, these genres of works produced by AI could be even harder to distinguish, as they feature less factuality, less objectivity, and more novelty. This study expands the research foci to genres of poems and paintings and seeks to understand viewers' attitudes toward machine-generated creative works.

Cross-Cultural Attitudes Toward Machines

Prior research has presented some empirical evidence of cultural impacts on viewers' perception of machines. Rau et al. (2009) compared Chinese and

German participants' attitudes toward robots; they found that Chinese participants perceived a robot to be more likable and trustworthy in addition to being more more likely to accept the implicit recommendations from a robot. German participants, however, were more anxious about the negative impact of social robots (Li et al., 2010). Bartneck et al. (2007) also examined intercultural differences, and found that among the U.S, China, and Germany, U.S participants were most concerned with robots designed with emotions. German participants expressed most concerns with robots' negative influence on children, while Chinese participants felt most uncomfortable interacting with a social robot. Zheng et al. (2018) further identified the moderators in the relationship between cultural influence and user interactions with machines. They found that American participants did not report much difference in algorithm-authored news versus human-authored news, whether the news was presented online or via traditional media. However, Chinese participants reported the quality of AI-produced news to be higher than that of human-produced news when the news was presented online. When the news was presented via traditional media, the quality of AI-produced news was perceived to be lower than that of human-produced news.

Social constructivism of technology (SCOT) may explain why users from different cultural backgrounds form different attitudes toward machines. According to Pinch and Bijker (1987), technological artifacts should reflect interpretive flexibility, which means that people's knowledge about technologies is socially and culturally produced. Individuals' interpretation of technologies is contingent upon the social groups with which they are affiliated. Turja and Oksanen (2019) applied SCOT to compare people's robot acceptance in 27 EU member countries. They found that when testing enjoyment, anthropomorphism, and perceived behavior control, individuals' perceptions of social robots may vary with cultures, even if they all come from Europe. In this way, past research has shown that culture may shape users' perception of machines. This study situates users' appraisal of machine-generated content in a cross-cultural context and explores the effects of sociocultural factors on users' perception of artworks.

Machine Agency

To comprehend how users' perception of artworks may be decided by authorship cues, the MAIN model is applied as the theoretical framework in this study (S. Sundar, 2008). In this model, technological affordances serve as cues that affect how users evaluate media interfaces (S. S. Sundar et al., 2015). Four types of affordances on user interfaces – modality, agency, interactivity, and navigability – can cue cognitive heuristics or mental shortcuts. As this study examines how authorship cues may elicit different cognitive processing of creative works, we investigate how agency, the source of media content, influences users' judgment. In particular, we consider the machine heuristic proposed by S. Sundar (2008), which reflects the mental shortcut that machines are neutral and objective in content presentation. This heuristic is said to be triggered by cues on the interface that convey machine agency. As some examples, researchers have found that news selected by a computer is perceived as higher in quality than that selected by a human editor (S. S. Sundar & Nass, 2001). Liu and Wei (2019) found that machine-written news was perceived as more objective than human-written news, regardless of the news institutions. Also, users were more likely to reveal private information to a machine agent than a human agent (S. S. Sundar & Kim, 2019, May). While the machine heuristic has been applied to interpret user attitudes toward news production, limited research has been conducted on whether and how the approach can be used to explain user interpretation of creative works. When building measures for machine heuristic, Yang and Sundar (2020) distinguished between mechanical tasks and human tasks; they suggested that user interpretation of machines as efficient, objective, and unemotional could reinforce the heuristic that machines are worse than humans in performing human tasks. As artworks can be considered a form of human creativity, users' evaluation of machine-created art versus human-created art is likely to be different.

The effects of machine heuristic may vary in individuals' expectations and interpretations of automation. When Waddell (2018) examined the effects of machine agency on news credibility, he found that machine-authored news was perceived to be less credible and newsworthy than human-authored news. He explained that automation in news writing was still considered new and could violate readers' expectations. Likewise, Spence et al. (2019) found that users' low suspicion of machine-selected information was a strong predictor of machine heuristic. The genre of news also leveraged the association between machine authorship and news credibility. Liu and Wei (2019) found that for human-produced news, there was no difference in perceived credibility between spot news (i.e., hard news) and interpretive news (i.e., soft news), whereas for machine written news, interpretive news was regarded as more credible than spot news. As these studies show, the effect of the machine heuristic is dependent upon individual attitudes, social norms, and news types. Therefore, we can speculate that the genre of artworks and sociocultural differences may also affect the processing of machine heuristic.

Machine Learning in Open-Ended Survey Responses

Machine learning has been widely applied in speech recognition, computer vision, political discussion, and text mining (Colleoni et al., 2014; Liang, 2014; Peng, 2018). Supervised and unsupervised learning are considered as two major approaches in machine learning (Kadhim, 2019). Supervised learning refers to the process where an algorithm learns from the training dataset and replicates human coding tasks with a machine (Pilny et al., 2019; Wiedemann, 2019). Supervised learning is used when data or documents need to be classified into pre-determined categories. Comparatively, unsupervised learning refers to a set of methods that explore the patterns or features of data without relying on predetermined classification (Grimmer & Stewart, 2013). Unsupervised learning can be especially valuable when some patterns of data are understudied or overlooked in the research process (Mourtgos & Adams, 2019).

One of the approaches used in unsupervised machine learning is structural topical modeling (STM). STM allows researchers to examine participants' open-ended responses alongside their demographic information, predispositions, and the effects of experimental manipulation (Roberts, Steward, Tingley, Lucas, et al., 2014). In the current context, there are a few advantages of using STM over traditional topic modeling like the Latent Dirichlet Allocation (LDA) models. Whereas LDA provides topic proportions within documents and word prevalence within topics, and assumes that clusters of words that relate to each other reflect the same latent topic, STM is built on both probabilistic topic models (Blei et al., 2003) and correlated topic models (Blei & Lafferty, 2007). According to Roberts, Steward, Tingley, Lucas, et al. (2014), STM can provide four sets of information: 1) the relationship between a covariate and the probability that a document discusses each topic (i.e., topic prevalence covariate effects), 2) the relationship between a covariate and the probability of word use in a particular topic (i.e., topic content covariate effects), 3) the proportion of words in a given document about each topic (i.e., document topic proportions), and 4) the probability of observing a word exclusively under a particular topic (i.e., topic word proportions). That is, while the document using LDA is unstructured, STM incorporates additional information such as participants' dispositions, covariates, and experimental treatment in the analyses, which would identify structural changes in topical prevalence and topic content.

As this study seeks to understand users' interpretations of machine-generated versus human-generated content in an experimental setting, STM is a tenable approach to capturing the effects of different genres of artworks and authorship cues on users' discussion. Here, we propose the following RQs.

RQ1: What are the top features of participants' responses to poems and paintings across U.S, China, and Germany?

RQ2: What are the main topics discussed among participants' responses to poems and the paintings across U.S, China, and Germany?

RQ3: Based on the open-ended remarks, how do participants perceive poems vs. paintings differently across U.S, China, and Germany?

RQ4: Based on the open-ended remarks, how do participants respond to machine-generated artworks vs. human-generated artworks differently across U.S, China, and Germany?

While machine learning can reduce the cost and time of analyzing large collections of digital texts, the output of the computer-assisted analyses needs to be verified. Grimmer and Stewart (2013) described two ways of validation: semantic validation and predictive validation. Semantic validation tests the degree to which machine-identified topics are consistent with human input. Predictive validation refers to the idea that if the identified topics are valid, experiment manipulation or covariates should explain the variations regarding the topics. As this study seeks to confirm the effects of experimental treatment on users' attitudes toward machine-generated works, predictive validation is used. We selected the variables of perceived quality of artworks, perceived imaginativeness, engagement, and spatial presence, as these variables could indirectly reflect how participants understood the artworks and how involving their experience was. Thus, to validate the results from STM, we propose the following RQs.

RQ5: Based on the closed-ended responses, how do different genres affect participants' perception of the artworks across the U.S, China, and Germany?

RQ6: Based on the closed-ended responses, how do different authorship cues affect participants' perception of the artworks across the U.S, China, and Germany?

Method

Participants

In the U.S, MTurk was used to recruit participants. A total of 422 U.S participants enrolled in the study. After eliminating invalid responses (e.g., those that failed manipulation checks), 251 participants' responses were included in final analyses. Among these participants, 134 were females (53.4%) and 117 were males (46.6%). They were 37.06 years old on average (*SD* = 13.28). In China, Sojump was used to recruit participants. The survey platform uses a national sampling pool of 2.6 million adults from China. After removing invalid cases from 344 participants, 293 participants' responses were included. There were 184 females (62.8%) and 109 males (37.2%). The average age was 32.1 years old (*SD* = 1.12). In Germany, Respondi, an online survey platform a was used to recruit participants. A total of 486 participants enrolled in the study. After removing invalid cases, 293 participants' responses were included. Among them, 159 were females (54.3%) and 133 were males (45.4%). The average age was 45.88 years old (*SD* = 14.64). In sum, 837 participants' responses were included in the final data analyses.

Experimental Stimulus and Procedures

An online experiment was conducted in the U.S, China, and Germany. A 2 × 2 between-subjects factorial design was used. Participants in each country were randomly assigned to four conditions: poems written by AI author, poems written by human author, paintings created by AI author, and paintings created by human author.

Two paintings were selected. One of them was titled "Field," which was created by the PIX 18 Creative Machines Lab at Columbia University. The painting received the top award at the 2017 Robot Art Competition (Robohub, 2017). The other one, titled "Pleased," was created by Rob Boss of HEARTalion at Halmstad University, which received sixth place in the competition. Participants in both authorship conditions viewed the same two paintings. For those who were assigned to human author conditions, the authorship cues were human names (e.g., Stephan Morgan, William Garner), while in AI author conditions, the authorship cue was "Starbot by RisingStar Tech (An automated painting robot)." To avoid the confounding effects of the styles of paintings, both paintings featured impressionist styles.

Participants assigned to the poem conditions read the same two poems across authorship conditions. However, due to language barrier and the difficulty of translating poems and their cultural connotations, participants in each country read two poems in their native language. To minimize the confounding effects, we chose poems that had similar styles, length, and structure. The English poems were "Long years have passed" and "A wounded deer leaps highest," created by Ray Kurzweil's Cybernetic Poet and collected in the book *The Age of Spiritual Machines* (Kurzweil, 2000). The Chinese poems were "Windows" and "Prisoners at lonely nights," created by Microsoft AI Xiao Ice and collected in the AI-authored poetry book "The Sunlight that Lost the Glass Window," which includes 139 AI-generated poems (Merriman, 2018). The German poems were "The flower meadow" and "I wish you a cheese," selected from works generated by Fabian Navarro's poetry bot "Eloquentron3000" (Lichtenegger, 2018). The bot has composed more than 500 poems, since it was initially launched. Those in the human author conditions were exposed to human names (e.g., Tobias Keller, Julia Schwartz), while those in the AI author conditions were exposed to the fictional name “Starbot by RisingStar Tech (A poem-writing robot).”

During the experiment, participants were first asked to answer demographic questions. Then, they were asked to read the poems or view the paintings. After they viewed each, they were asked to describe their feelings about the work. Then, they answered closed-ended questions relating to its perceived quality, perceived imaginativeness, their sense of engagement, and experience of spatial presence. Last, they completed the measures of control variables, such as their AI use frequency and their attitudes toward AI. Manipulation checks were conducted by asking participants to recall the author names of the poems or the paintings after they filled out the open-ended questions, but prior to answering the closed-ended questions.

Measures

The measure of perceived quality (U.S: M = 3.07, SD = .94, α = .95; China: M = 3.33, SD = .58, α = .84; Germany: M = 2.09, SD = .85, α = .95) was adapted from the S. S. Sundar (2000) measure. Participants were asked to report on a nine-item Likert-type scale (1 = not at all, 5 = very much). An example item is “I think the poem I just read was well-written.”

The measure of perceived imaginativeness (U.S: M = 3.36, SD = 1.02, α = .93; China: M = 3.72, SD = .68, α = .78; Germany: M = 2.60, SD = 1.05, α = .88) was a self-constructed scale. Participants were asked to report on a three-item Likert-type scale (1 = not at all, 5 = very much). Examples of the items include “I think the poem I just read was imaginative” and “I think the poem was creative.”

The measure of engagement (U.S: M = 3.20, SD = .99, α = .88; China: M = 3.10, SD = .76, α = .83; Germany: M = 1.78, SD = .87, α = .89) was adapted from Lombard et al.’s (2009) Presence Inventory. Participants were asked to report on a three-item Likert-type scale (1 = not at all, 5 = very much). Examples of the items include “how involving was the reading experience?” and “to what extent did you feel mentally immersed in the reading experience?”

The measure of spatial presence (U.S: M = 2.79, SD = 1.21, α = .95; China: M = 2.99, SD = .79, α = .79; Germany: M = 1.60, SD = .77, α = .92) was also adapted from the Presence Inventory (Lombard et al., 2009). Participants were asked to report a three-item Likert-type scale (1 = not at all, 5 = very much). Examples of the items include “to what extent did you experience a sense of being there in the painting?” and “how much did it seem as if you could reach out and touch the objects or people in the painting?”

Data Analyses

Data analyses were split into two parts. The first part deals with machine learning techniques using STM. The unit of analysis was each participant's open-ended responses to the artworks (i.e., each document). To answer RQ1 to RQ4, STM was conducted using *R*. Data analyses followed Roberts, Steward, Tingley, Lucas, et al.'s. (2014) procedures of ingestion, preparation, and estimation, where estimation was further divided into model evaluation, interpretation, and visualization. The datasets were first converted to corpora. Then, text preprocessing was conducted (see Denny & Spirling, 2018). Besides the original stop words stored in the *Quanteda* package, we included additional stop words such as "feel," "feeling," "like," and "make" in all the three datasets. In English and German datasets, all letters were lowercase. We also used stemming to remove suffixes and prefixes to retrieve the word root, so that words, such as "paintings" and "paint" would share similar meanings in the document-feature matrix. Infrequent terms, numbers, and punctuations were also removed. We included both uni-gram and bi-grams in the analyses. After preprocessing, English text had 251 documents and 1300 features. Chinese text had 293 documents and 1046 features. German text had 293 documents and 693 features. Thus, a total of 837 documents with 3039 features were analyzed. We extracted top features from the matrices to answer RQ1.

Based on document-feature matrices, to answer RQ2 to RQ4, we evaluated the models using *stm*. We first searched the approximate number of topics that may fit the models. As an initial trial, we selected 2 to 20 topics to see the patterns of the results. We evaluated the models by plotting held-out likelihood, semantic coherence, and residual analyses. As a general rule of thumb, higher held-out likelihood, higher semantic coherence, and lower residual analyses would provide relative goodness of fit. Then, we narrowed down the range of topics. As English, Chinese, and German documents vary in their topics, selection of the range of topics was different. As an example, for the English dataset, we narrowed down the number of topics to the range of 4 to 8. As there is no statistical criterion to provide a definite answer to the best number of topics (DiMaggio et al., 2013), we relied on both machine learning and human interpretation to evaluate the models. With fewer number of topics, we were able to manually examine each model by looking into the semantic coherence, the exclusivity of words to a topic (i.e., FREX), and the highest probability words. We also relied on the method of "reading tea leaves" (see Chang et al., 2009) by looking for word intrusion and topic intrusion. We asked two research assistants to read the words and topics related to the documents and ensured that the words were coherent and shared similar meanings under a general topic. Via this iterative process, we were able to identify the model that had optimal goodness of fit. After that, we estimated the effects of experiment treatment on the topics. *estimateEffect* renders the relationships among experiment manipulation, covariates, and the rate of individual word use within a certain topic (Roberts, Stewart, Tingley, et al., 2014).

In the second part of the data analyses, traditional quantitative data analyses were used to validate the results from machine learning and to answer RQ5 and RQ6. Three-way ANOVAs with authorship cues, genres, and country as independent variables were conducted to confirm the effects of authorship cues and content genres on participants' perception of the artworks (RQ5 and RQ6).

Results

To find the top features among participants' responses to the poems and the paintings (RQ1), the most frequent features in the document-feature matrices were computed. Top seven features were found in each document-feature matrix. Among U.S participants' responses, paint, poem, sad, look, good, seem, and color were the top features in the documents. Among the responses from Chinese participants, 抽象 (abstract), 画 (paintings), 比较 (relatively), 表达 (express), 不懂 (confused), 色彩 (color), and 母亲 (mother) were the top features. German participants reported verwirrt (confused), gedicht (poem), vorher (before), käse (nonsense/cheese), irritiert (irritating), normal (normal), and bild (picture) as their top features in their responses. The text plot of the top features is shown in Figure 1.

As part of our effort to assess the main topics among participants' responses to the poems and the paintings across the U.S, China, and Germany (RQ2), we identified five topics among U.S participants' responses. These topics were relaxation (topic 1), senses (topic 2), affective appraisals (topic 3), colored paintings (topic 4), and positive evaluations (topic 5). We identified three topics in Chinese participants' responses. They were abstraction and vitality (topic 1), affective appraisals (topic 2), and maternal love

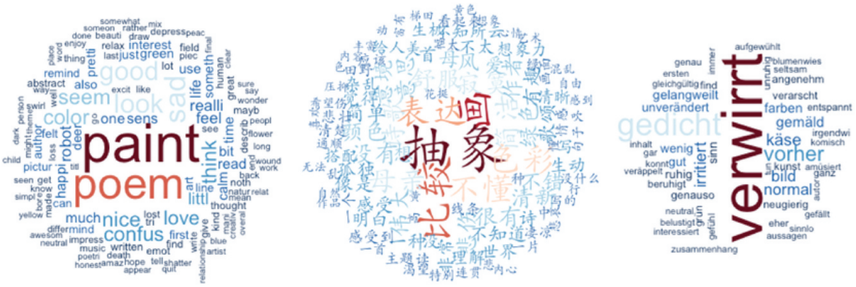

Figure 1. Text plots of the top features: U.S, China, and Germany.

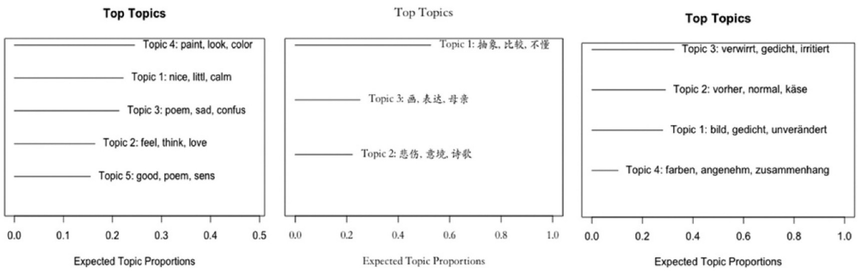

Figure 2. Document-topic proportions: U.S, China, and Germany.

(topic 3). We identified four topics in German participants' responses. They were invariableness (topic 1), normality (topic 2), negative evaluation (topic 3), and relaxation (topic 4). Document-topic proportions and word use within topics are shown in Figures 2 and 3.

US PARTICIPANTS					
Topics	Relaxation	Senses	Affective Appraisals	Colored Paintings	Positive Evaluations
	Probability: nice, littl, calm, relax Exclusivity: relax, bore, neutral, nice, calm, littl	Probability: feel, think, love Exclusivity: come, noth, feel, tri	Probability: sad, confus, seem Exclusivity: lost, depress, sad, confus, author	Probability: paint, look, color, green, art Exclusivity: abstract, paint, swirl, seen, art	Prob: sens, seem, line, think Exclusivity: line, good, lot, better, rest
Representative response	I felt very relaxed and somewhat connected to nature.	Thought about my past incidents, events. That makes me feel and enjoy makes me very stronger in mind.	I feel kind of sad and depressed after reading this poem. Sounds like the robot is going through some tough times.	It's a great painting. It does remind me of fields. Green and lush. I would hang that on my wall.	know this poem is not meant to be humorous, and it actually seems well written.

CHINESE PARTICIPANTS			
Topics	Abstraction and Vitality	Affective Appraisals	Maternal love
	Probability: 抽象,不懂, 色彩, 寂寞, 简洁 Exclusivity: 生机,一片, 开发	Probability: 悲伤, 意境, 诗歌, 感受 Exclusivity: 感受, 意境, 悲伤, 凄凉	Probability: 画, 表达, 母亲, 伟大, 作者, 母, 爱 Exclusivity 母亲, 母, 爱, 歌颂, 明白
Representative response	比较抽象,不懂内容比较轻松愉悦.	意境高，仿佛置身其中一种凄凉的感觉	母爱是无私的力量，消无声息却滋润着一颗颗生命的幼苗成长孤独

GERMAN PARTICIPANTS				
Topics	Invariableness	Normality	Negative Evaluation	Relaxation
	Probability: unverändert, immer, neutral Exclusivity: unverädert, egal, neutral	Probability: vorher, normal, genauso Exclusivity: normal, genauso, erst, vorher, irgendwi, genau	Probability: verwirrt, irritiert, verarscht Exclusivity: sinnlos, veräppelt, verwirrt, irritiert, verarscht	Probability: angenehm, zusammenhang, beruhigend, erinnert Exclusivity: beruhigend, angenehm, erkenn
Representative response	Unverändert. Das Gedivht reimte sich nicht, unterhalten	fühle mich normal, immer noch normal	Ich habe das Gefühl verarscht zu werden ....so einen Unsinn kann man doch nicht Gedicht nennen	Ganz angenehm, die Wellen und die Farben beruhigen.

Figure 3. Sample word use based on highest probability and exclusivity within topics.

In response to the question about how participants perceived poems and paintings differently across the U.S, Germany, and China (RQ3), results suggested that among U.S participants, the paintings were more likely to affect users' word use about relaxation (topic 1), $B = .08$, $p = .034$ and colors (topic 4), $B = .27$, $p < .001$, while the poems were more likely to affect users' word use about affective appraisals (topic 3), $B = -.22$, $p < .001$ and positive evaluations (topic 5), $B = -.08$, $p = .016$. It was also found that U.S participants' AI use frequency negatively predicted their use of words about affective appraisals (topic 3), $B = -.04$, $p = .021$.

Among Chinese participants, the paintings were more likely to affect users' word use about abstraction and vitality (topic 1), $B = .28$, $p < .001$, while poems were more likely to affect users' word use about affective appraisals (topic 2), $B = -.14$, $p < .001$, and maternal love (topic 3), $B = -.14$, $p < .001$. Chinese participants' AI use frequency negatively predicted the use of words about affective appraisals (topic 2), $B = -.04$, $p = .04$.

Among German participants, the paintings were more likely to affect participants' discussion of invariableness (topic 1), $B = .10$, $p = .020$. The poems were more likely to affect users' word use about negative evaluation (topic 3), $B = -.11$, $p = .002$. The effects of genres on participants' discussion topics are shown in Figure 4. The effects of covariates, authorship cues, and genres on participants' discussion topics are shown in Tables 1–3.

To answer how participants responded to machine-generated works vs. human-generated works differently across the U.S, Germany, and China (RQ4), STM analysis found authorship cues significantly predicting U.S

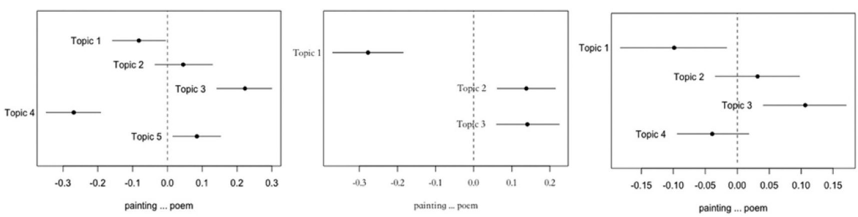

Figure 4. Effects of genres on topic discussion: U.S, China, and Germany.

Table 1. The effects of covariates, authorship cues, and genres on U.S participants' topics.

	Relaxation Topic 1	Sensing Topic 2	Affective appraisals Topic 3	Colored paintings Topic 4	Positive evaluations Topic 5
Variables	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
Gender	.01 (.04)	−.01 (.04)	.02 (.04)	−.00 (.04)	−.02 (.04)
Attitudes	−.02 (.03)	−.01 (.03)	.04 (.03)	.03 (.03)	−.04 (.03)
AI use	.00 (.01)	−.00 (.02)	−.04(.02)*	.02 (.02)	.01 (.01)
Authorship	.08 (.04)*	−.01 (.04)	−.04(.04)	−.02 (.04)	−.01 (.04)
Genre	.08 (.04)*	−.05 (.04)	−.22(.04)***	.27 (.04)***	−.08 (.03)*

*p < .05, **p < .01, ***p < .001. Author (1 = machine, 2 = human). Genre (1 = poem, 2 = painting).

Table 2. The effects of covariates, authorship cues, and genres on Chinese participants' topics.

	Abstraction/vitality Topic 1	Affective appraisals Topic 2	Maternal love Topic 3
Variables	*B (SE)*	*B (SE)*	*B (SE)*
Gender	.04 (.05)	−.04 (.04)	.01 (.04)
Attitudes	.01 (.04)	−.00 (.04)	−.01 (.00)
AI use	.04 (.02)	−.04 (.02)*	−.00 (.02)
Authorship	−.00 (.04)	−.01 (.04)	.01 (.00)
Genre	.28 (.05)***	−.14 (.04)***	−.14 (.04)***

$^{+}p < .1$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 3. The effects of covariates, authorship cues, and genres on German participants' topics.

	Invariableness Topic 1	Normality Topic 2	Negative evaluations Topic 3	Relaxation Topic 4
Variables	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
Gender	.04 (.04)	−.01 (.04)	.01 (.04)	−.04 (.03)
Attitudes	.00 (.03)	−.02 (.02)	−.00 (.03)	.02 (.02)
AI use	.02 (.02)	−.00 (.02)	−.01 (.02)	−.00 (.01)
Authorship	.02 (.04)	.02 (.04)	−.00 (.03)	−.03 (.03)
Genre	.10 (.04)*	−.03 (.03)	−.11 (.03)**	.04 (.03)

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

participants' rate of word use within the topic of relaxation (topic 1), $B = .08$, $p = .034$. However, authorship cues did not predict Chinese and German participants' rate of word use within any topic.

As validation, to answer how authorship cues and genres of content affect participants' perception of the creative works across U.S, Germany, and China (RQ5 & RQ6), results from three-way ANOVAs suggested that paintings ($M = 3.09$, $SD = .84$) were perceived to have higher quality than poems ($M = 2.56$, $SD = 1.00$), $F(1, 822) = 83.46$, $p < .001$. Culture was found to have main effects on perceived quality, $F(2, 822) = 211.28$, $p < .001$, where German participants ($M = 2.09$, $SD = .85$) rated the quality of the works to be lower than U.S participants ($M = 3.07$, $SD = .94$), $p < .001$. U.S participants scored lower than Chinese participants ($M = 3.33$, $SD = .58$), $p < .001$ (Figure 5).

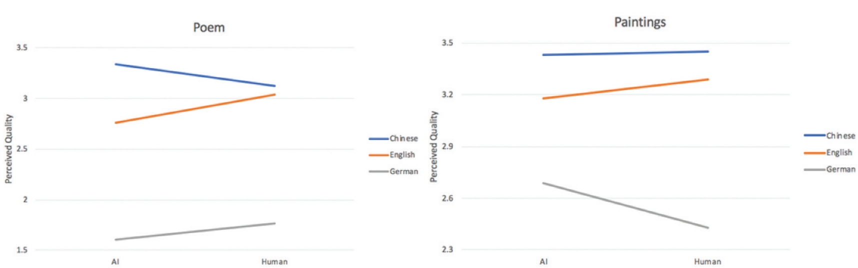

Figure 5. Three-way interaction on perceived quality of the works.

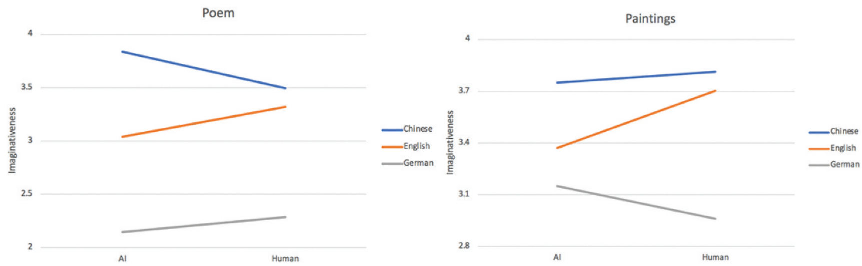

Figure 6. Three-way interaction on perceived imaginativeness of the works.

Paintings (M = 3.47, SD = .93) were also perceived to be more imaginative than poems (M = 2.98, SD = 1.09), $F(1, 822) = 50.81$, $p < .001$. Culture had main effects on perceived imaginativeness, $F(2, 822) = 112.41$, $p < .001$, where German participants (M = 3.05, SD = .96) assigned lower scores than U.S participants (M = 3.53, SD = .99), $p < .001$. U.S participants scored lower than Chinese participants (M = 3.78, SD = .66), $p < .001$ (Figure 6).

Additionally, paintings (M = 2.80, SD = 1.03) were perceived to be more engaging than poems (M = 2.51, SD = 1.12), $F(1, 822) = 13.54$, $p < .001$. Culture had main effects on participants' sense of engagement, $F(2, 822) = 215.02$, $p < .001$, with German participants (M = 1.78, SD = .87) feeling less engaged than Chinese participants (M = 3.10, SD = .76), $p < .001$ and U.S participants (M = 3.13, SD = 1.01), $p < .001$. However, Chinese participants did not differ from U.S participants in sense of engagement, $p = .93$ (Figure 7).

Participants experienced more spatial presence of the paintings (M = 2.59, SD = 1.08) than the poems (M = 2.30, SD = 1.15), $F(1, 822) = 12.95$, $p < .001$. Culture had main effects on spatial presence, $F(2, 822) = 190.98$, $p < .001$, such that German participants (M = 1.60, SD = .77) scored lower than U.S participants (M = 2.79, SD = 1.21),

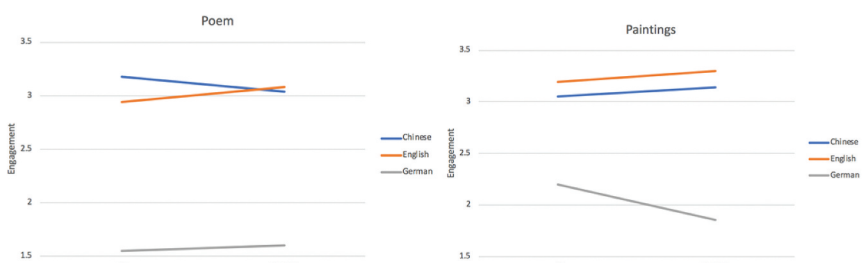

Figure 7. Three-way interaction on engagement of the works.

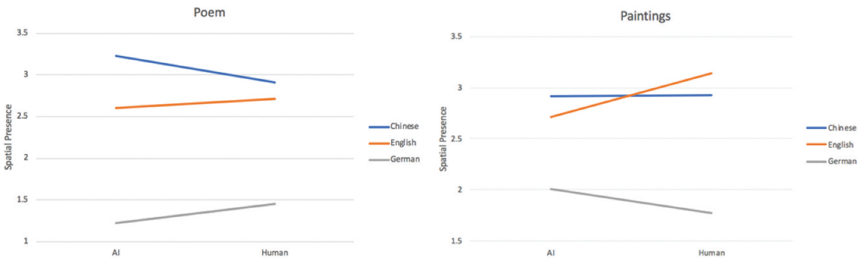

Figure 8. Three-way interaction on spatial presence.

Table 4. Results of three-way ANOVAs.

	Perceived quality		Imaginativeness		Engagement		Spatial presence	
Variables	*F*	ω^2	*F*	ω^2	*F*	ω^2	*F*	ω^2
Genre	83.46***	.06	50.81***	.04	13.54***	.01	12.95***	.01
Authorship	.12	.00	.51	.00	.08	.00	.33	.00
Nation	211.28***	.30	112.41***	.19	215.02***	.34	190.98***	.30
Genre*Author	1.35	.00	.10	.00	.25	.00	.21	.00
Genre*Nation	15.97***	.02	12.41***	.02	5.27**	.01	10.62***	.02
Author*Nation	3.00	.00	4.43*	.01	1.70	.00	3.66*	.00
Genre*Author*Nation	3.87*	.00	2.95	.00	2.36	.00	4.42*	.01

*$p < .05$, **$p < .01$, ***$p < .001$. ω^2 omega squared for effect size.

$p < .001$. U.S participants scored significantly lower than Chinese participants ($M = 2.99$, $SD = .79$), $p = .026$ (Figure 8). Consistent with STM, authorship cues failed to predict users' perceived quality, imaginativeness, engagement, and spatial presence. Results of three-way ANOVAs are shown in Table 4.

Discussion

This study seeks to understand whether and how users perceive machine-generated artworks and human-generated artworks differently in a cross-cultural context. By combining experimental design and machine learning methods, this study suggests that people developed different feelings and form different discussion topics when viewing creative works.

Based on the top features and the main topics related to participants' word use, both U.S participants and Chinese participants mentioned that the creative works elicited affective appraisals. For instance, they both used words such as "depressed," "sad," and "restless" in their open-ended remarks on poems. Also, both German participants and U.S participants found the creative works relaxing and calming. They used words like "angenehm" (pleasant) to describe their feelings. Yet, participants from each cultural background presented distinct word selection and discussion patterns. U.S participants were generally more positive about the artworks. They put emphasis on how the artworks were relaxing, well-written, colorful, and soothing. Chinese participants used more concrete imagery to describe their feelings. They used words such as “母爱” (maternal love) and “歌颂” (ode) to describe the poems. They also stated that the creative works seemed abstract but restful [e.g., “新的事物，生机勃勃简洁，舒服” (New things, vital and energetic, comfortable.)].

Although some German participants mentioned that the artworks were calming, overall their comments tended to be more negative. They mentioned that the artworks were “unverändert” (changeless). They stated that the works were “genervt” (annoying), “sinnlos” (pointless), and “irritiert” (irritating). German participants’ attitudes toward machine-created works were aligned with prior research findings where Bartneck et al. (2007) and Rau et al. (2009) found that Germans had more concerns about the negative influence of AI. According to Zheng et al. (2018), the difference between their responses and Chinese participants’ responses could be attributed to the high-context and low-context culture dimensions, where German participants may have relied more on explicit messages and Chinese participants may have attempted to attribute more implicit meanings to the works.

The differences in their word use and topics could also be attributed to sociocultural factors. Specifically, German participants’ ratings of the artworks were significantly lower than the U.S or Chinese participants on all these dimensions. Based on a European survey (Technik Radar, 2019), Germans were found to be more skeptical about the application of AI in society. German users expected robotic technologies to be used for simple tasks and play a safer role in society (Bernotat & Eyssel, 2018). Such preferences may be associated with German consumers being more sensitive to data privacy, compared to other cultures such as the U.S and Japan (Krasnova & Veltri, 2010; Nitto et al., 2017). By contrast, just as how robots have been framed as social agents that can help advance the modernity of the society in Japan (Šabanović, 2014), China has promoted the narratives about the positive use of AI in smart cities (e.g., better transportation management, Internet of Things). Despite people’s growing concerns about the invasion of privacy by technology, the convenience of mobile payment, facial recognition, and blockchain may have led Chinese participants to have more anticipations for AI technologies. These differences have reflected Pinch and Bijker’s (1987) viewpoint that people’s understanding of technologies is culture driven and socially constructed.

Consistent with Liu and Wei’s (2019) research, which found that genre played a role in users’ evaluation of machine agency, our findings show genre main effects on participants’ discussion of the creative works. U.S participants described the paintings to be more colorful and relaxing and used more emotional words, such as "lost" or "depressed." They used a more positive lens (e.g., "well-written") to analyze the poems. Similarly, Chinese participants reported the paintings to be abstract and the poems to be engaging and sentimental. Many of them mentioned that the poems seemed like a eulogy to maternal love. German participants' negative evaluation [e.g., "unruhig" (restless), "unbeeindruckt" (unimpressed)] was more related to the poems, while their feelings about lack of change were associated with their perception of the paintings. From the perspective of predictive validation, the results of the three-way ANOVAs corroborated that genres had main effects on users' responses to the artworks. Paintings were perceived to be higher quality, more imaginative, more engaging, and evoked stronger spatial presence than poems.

Results from machine learning further suggested that when participants revealed their feelings about the creative works, they did not use significantly different words between human-generated works and machine-generated works. The only exception was that U.S participants used more words such as calm, enjoy, and relax -- when viewing human-authored paintings than when viewing machine-authored ones--suggesting that they felt more relaxed with human-authored paintings. Overall, although participants from cross-cultural backgrounds assigned different discussion topics to these artworks, their responses were independent of agency cues. Considering that people initially reported lower acceptance of computers taking on interpretative roles (e.g., editorial writers, novelists) than the ones taking on routinized roles (e.g., ATM, automatic cashiers, mall guides) (Nass et al., 1995), the findings here imply that over time people's acceptance of machines assuming interpretive roles like artists and composers has increased. Even though sometimes machine-generated content could violate readers' expectations (Waddell, 2018), due to people's growing exposure to AI devices, their acceptance of machine-generated works may have grown as well. The finding that participants did not use different linguistic elements to describe machine-generated works and human-generated ones was validated by the results of our three-way ANOVAs.

According to the machine learning results, participants' prior AI use frequency influenced their discussion topics. For both U.S and Chinese participants, lower AI use frequency led to more affective appraisals, meaning that those who interacted with AI less often in their daily life were more willing to express their emotional feelings about the artworks. The results can serve as additional evidence to prior research findings that technology use experiences affect users' psychological responses to computing technologies (Johnson et al., 2004).

This study has theoretical, methodological, and practical implications. It applies the concept of machine heuristic to understanding the role of authorship cues in users' evaluation of the artworks. Prior research on machine heuristic has primarily focused on users' perception of machine performance in mechanical tasks, such as hard news production (Spence et al., 2019; Waddell, 2018). This study tests machine heuristic in a human task context and discovers that participants in each country not only initiated similar discussion topics when viewing machine-generated works and human-generated works but, also reported similar levels of perceived quality, imaginativeness, and sense of engagement in viewing these two types of works. While the findings seem to refute the effects of machine agency cues, it should be noted that according to Yang and Sundar (2020), for human tasks, perceiving machines to have expertise and accuracy would alleviate machine heuristic, meaning that users are less likely to feel the inferiority of machines to humans in human tasks. This study thus confirms the necessity to differentiate mechanical tasks and human tasks and apply different machine heuristic models to explain different scenarios.

Methodologically, to our knowledge, this study is the first to apply STM to understand users' evaluation of machine-generated art in a cross-cultural context. By using an unsupervised learning approach, this study retrieves the major patterns from users' comments on the artworks in a relatively objective and efficient way (Grimmer & Stewart, 2013). First, as participants may develop complex and extended feelings toward artworks, STM allows them to express feelings without being limited by close-ended questions. Second, STM incorporates different experimental conditions and individuals' AI use experiences into automated text analyses. This study demonstrates that computational methods and experiments can be, combined in that researchers can learn the patterns and themes from large amounts of digital data as well as use quantitative methods to validate and complement the findings. Future research could continue to combine computational methods with traditional social science methods to gauge users' responses in online settings.

Practically, this study shows that technology has elevated machine-generated works to a level where people react to them as human works. Although scholars argue that machines are better at presenting fact-based information than creativity or personalities (Van Dalen, 2012), the current findings suggest that machine-produced works can indeed be perceived as creative, innovative, and engaging. On the one hand, machine-created works will significantly reduce the cost of esthetic evaluation. Considering that social norms and cultural backgrounds play a significant role in shaping users' attitudes toward machine-generated artworks, the presentation of machine-created artworks may even be personalized to accommodate viewers' preferences and cultural values. On the other hand, such expanded use of machine works may become a threat to the traditional norms that people hold to assess poems and paintings. If machines could create works that are perceived to be imaginative and engaging, people may need to reconsider the value of human-created arts. We may need to ponder what we mean by "creative" and "innovative," and whether art is a reflection of human inspiration or a product of computation and algorithms (Gunkel, 2012). These questions may lead us to further contemplate whether we need laws to protect our intellectual property regarding AI content, what data algorithms we should use to create works free from biases or stereotypes, and how the use of these works could fit into people's cultural values and social norms. Although we are facing challenges from more machine-generated content (e.g., deepfake), it is simultaneously our opportunity to scrutinize what role machines should play in the society and what it means to be human in this digital age.

Conclusions and Limitations

This study first expands the prior research scope of machine-generated content from news and social media posts to poems and paintings. Then, combining computational methods and online experiments, it examines whether there exist semantic commonalities and disparities in participants' attitudes toward machine-generated art and human-generated art. Data suggest that while genres and cultures affected participants' discussion topics and word use, no differences were found between participants' responses to machine-generated artworks and human-generated ones.

The current study can guide future research in two ways. First, this study demonstrates the feasibility of combining experimental design with the machine learning techniques. As prior research on online experiments has often relied on closed-ended questions to analyze the relationships between experimental manipulation and participants' responses, online open-ended responses have been largely overlooked, partly due to the cost of manually analyzing large collections of textual data. The introduction of machine learning can efficiently help researchers make inferences from participants' open-ended answers about how the experimental treatments leverage their cognitive processes and responses. Second, this study implies that when applying machine heuristic to understand users' attitudes toward machines, researchers should distinguish mechanical tasks from human tasks (Yang & Sundar, 2020). Depending on the specific tasks and contexts, users may form different mental shortcuts to respond to machine agency or machine-generated content.

The study has some limitations. First, although we selected multiple poems and paintings for each participant to view in order to increase the variance in the stimuli, our sampling of the creative works was still limited. Future research could use more experimental stimuli to study the effects of machine-generated artworks on participants' attitudes. Second, unlike the assessment of news which could be measured by its credibility and authors' writing competency, the criteria for evaluating artworks can be diverse and complicated. People may assess art works through the lens of styles, colors, themes, time, and strokes. To validate participants' discussion topics, this study only measured perceived quality, creativity, engagement, and sense of being in the art works. Future research could incorporate more variables to validate users' experiences in viewing the art works. Third, as this study did not test any psychological constructs that link the cues to the machine heuristic, it is unknown whether it was the authorship cues that triggered the heuristic. Therefore, future research should treat heuristics as variables to verify the activation of mental shortcuts (see Bellur & Sundar, 2014).

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

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

Notes on contributors

Kun Xu (Ph.D., Temple University) is an assistant professor of emerging media at the College of Journalism and Communications, University of Florida. His research centers on the concept of presence in human-robot interaction, virtual/augmented reality, and other emerging technologies.

Fanjue Liu (M.A., University of Florida) is a doctoral student at the University of Florida. Her research interests include human-machine communication and computer-mediated communication. She seeks to investigate how and why people make sense of and interact with social bots.

Yi Mou (Ph.D., University of Connecticut) is a special researcher at the School of Media and Communication of Shanghai Jiao Tong University. Her research interests include new media effects, human-machine communication, and health communication.

Yuheng Wu (M.A., Shanghai Jiao Tong University) is a doctoral student at the School of Media and Communication of Shanghai Jiao Tong University. His research interests include new media effects, human-machine communication, and health communication.

Jing Zeng (Ph.D., Queensland University of Technology) is a senior research and teaching associate at the University of Zurich. Her research interests include science communication, online misinformation, and digital methods.

Mike S. Schäfer (Ph.D., Free University of Berlin) is full professor at the University of Zurich. His working topics are science communication, environmental and climate communication, and social media debates about science and technology.

ORCID

Kun Xu http://orcid.org/0000-0001-9044-821X
Jing Zeng http://orcid.org/0000-0001-5970-7172
Mike S. Schäfer http://orcid.org/0000-0002-0847-7503

References

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*(1–2), 217–230. https://doi.org/10.1007/s00146-006-0052-7

Bellur, S., & Sundar, S. S. (2014). How can we tell when a heuristic has been used? Design and analysis strategies for capturing the operation of heuristics. *Communication Methods and Measures*, *8*(2), 116–137. https://doi.org/10.1080/19312458.2014.903390

Bernotat, J., & Eyssel, F. (2018, August). Can ('t) wait to have a robot at home? Japanese and German users' attitudes toward service robots in smart homes. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 15–22). IEEE.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(1), 17–35. https://doi.org/10.1214/07-AOAS114

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. http://www.cse.cuhk.edu.hk/irwin.king/_media/presentations/latent_dirichlet_allocation.pdf.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*(3), 416–431. https://doi.org/10.1080/21670811.2014.976412

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In Advances in neural information processing systems. 22, 288–296

Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice*, *8*(5), 519–531. https://doi.org/10.1080/17512786.2014.883116

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication*, *64*(2), 317–332. https://doi.org/10.1111/jcom.12084

Deahl, D. (2018, August 31). *How AI-generated music is changing the way hits are made*. The Verge.https://www.theverge.com/2018/8/31/17777008/artificial-intelligence-taryn-southern-amper-music

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168–189. https://doi.org/10.1017/pan.2017.44

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S government arts funding. *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, *4*(6), 700–722. https://doi.org/10.1080/21670811.2015.1096748

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. https://doi.org/10.1016/j.chb.2013.08.013

Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv Preprint arXiv:1706.07068*, 1–22. https://arxiv.org/pdf/1706.07068.pdf

Gerrish, S., & Blei, D. M. (2012). How they vote: Issue-adjusted models of legislative behavior. Advances in neural information processing systems, 25, 2753–2761.

Graefe, A., Haim, M., Haarmann, B., & Brosius, H. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, *19* (5), 595–610. https://doi.org/10.1177/1464884916641269

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.

Johnson, D., Gardner, J., & Wiles, J. (2004). Experience as a moderator of the media equation: The impact of flattery and praise. *International Journal of Human-computer Studies*, *61*(3), 237–258. https://doi.org/10.1016/j.ijhcs.2003.12.008

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, *52*(1), 273–292. https://doi.org/10.1007/s10462-018-09677-1

Krasnova, H., & Veltri, N. F. (2010, January). Privacy calculus on social networking sites: Explorative evidence from Germany and USA. *2010 43rd Hawaii international conference on system sciences* (pp. 1–10). IEEE.

Kurzweil, R. (2000). *The age of spiritual machines: When computers exceed human intelligence*. Penguin Books.

Li, D., Rau, P. P., & Li, Y. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, *2*(2), 175–186. https://doi.org/10.1007/s12369-010-0056-9

Liang, H. (2014). Coevolution of political discussion and common ground in web discussion forum. *Social Science Computer Review*, *32*(2), 155–169. https://doi.org/10.1177/0894439313506844

Lichtenegger, F. (2018, May 9). *Dieser Bot schreibt tiefgründige Gedichte auf Instagram*. Vice. https://www.vice.com/de/article/wj7w74/eloquentron3000-dieser-bot-schreibt-tiefgrundige-gedichte-auf-instagram

Liu, B., & Wei, L. (2019). Machine authorship in situ: Effect of news organization and news genre on news credibility. *Digital Journalism*, *7*(5), 635–657. https://doi.org/10.1080/21670811.2018.1510740

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). *Measuring presence: The temple presence inventory* [Paper presentation]. Proceedings of the 12th Annual International Workshop on Presence (pp. 1–15), Los Angeles, CA.

Merriman, C. (2018, August 15). *Microsoft's xiaoice chatbot has written a book of poetry*. The Inquirer. https://www.theinquirer.net/inquirer/news/3061160/microsofts-xiaoice-chatbot-has-written-a-book-of-poetry

Moses, L. (2017). *The Washington Post's robot reporter has published 850 articles in the past year*. Digiday. https://digiday.com/media/washington-posts-robot-reporter-published-500-articles-last-year/

Mourtgos, S., & Adams, I. T. (2019). The rhetoric of de-policing: Evaluating open-ended survey responses from police officers with machine learning based structural topic modeling. *Journal of Criminal Justice*, *64*(C), 61–73. https://doi.org/10.1016/j.jcrimjus.2019.101627

Nass, C. I., Lombard, M., Henriksen, L., & Steuer, J. (1995). Anthropocentrism and computers. *Behaviour & Information Technology*, *14*(4), 229–238. https://doi.org/10.1080/01449299508914636

Nitto, H., Taniyama, D., & Inagaki, H. (2017). Social acceptance and impact of robots and artificial intelligence. *Nomura Research Institute Papers, 211*.

Pavlik, J. V. (1999). New media and news: Implications for the future of journalism. *New Media & Society*, *1*(1), 54–59. https://doi.org/10.1177/14614448990010010 09

Peng, Y. (2018). Sa me candidates, different faces: Uncovering media bias in visual portrayals of presidential candidates with computer vision. *Journal of Communication*, *68*(5), 920–941. https://doi.org/10.1093/joc/jqy041

Pilny, A., McAninch, K., Slone, A., & Moore, K. (2019). Using Supervised Machine Learning in Automated Content Analysis: An Example Using Relational Uncertainty. Communication Methods and Measures, *13*(4), 287–304.

Pinch, T. J., & Bijker, W. E. (1987). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, & T. J. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 17–50). MIT Press.

Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587–595. https://doi.org/10.1016/j.chb.2008.12.025

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). Stm: R package for structural topic models. *Journal of Statistical Software*, *10*(2), 1–40. doi:10.18637/jss.v091.i02.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Šabanović, S. (2014). Inventing Japan's 'robotics culture': The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science*, *44*(3), 342–367. https://doi.org/10.1177/0306312713509704

Robohub. (2017, March 19). *2017 robot art competition winners announced*. https://robohub.org/2017-robot-art-competition-winners-announced/

Spence, P. R., Edwards, C., Edwards, A., & Lin, X. (2019). *Testing the machine heuristic: Robots and suspicion in news broadcasts* [Paper presentation]. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 568–569), Daegu, Korea.

Sundar, S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). The MIT Press.

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. https://doi.org/10.1177/107769900007700302

Sundar, S. S., Jia, H., Waddell, F., & Huang, Y. (2015). Toward a theory of interactive media effects (TIME): Four models for explaining how interface features affect user psychology. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 47–86). John Wiley & Sons, Inc.

Sundar, S. S., & Kim, J. (2019, May). Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–9). doi:10.1145/3290605

Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, *51*(1), 52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

Technik Radar. (2019). *Was die Deutschen über Technik denken*. Körber-Stiftung. https://www.koerber-stiftung.de/fileadmin/user_upload/koerber-stiftung/redaktion/technikradar/pdf/2019/Technikradar-2019_Booklet.pdf

Thurman, N., Dörr, K., & Kunert, J. (2017). When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, *5*(10), 1240–1259. https://doi.org/10.1080/21670811.2017.1289819

Trentini, Y. A. (2017, March 21). *Computer generated poetry will knock your socks off*. Medium. https://medium.com/@Yisela/computer-generated-poetry-will-knock-your-socks-off-763c815a1b52

Turja, T., & Oksanen, A. (2019). Robot acceptance at work: A multilevel analysis based on 27 EU countries. *International Journal of Social Robotics*, *11*(4), 679–689. https://doi.org/10.1007/s12369-019-00526-x

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. https://doi.org/10.1080/17512786.2012.667268

Van der Kaa, H., & Krahmer, E. (2014). *Journalist versus news consumer: The perceived credibility of machine written news* [Paper presentation]. *Proceedings of the Computation Journalism Conference, Columbia University* (pp. 24–25), New York, NY.

Waddell, T. F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, *6*(2), 236–255. https://doi.org/10.1080/21670811.2017.1384319

Weiser, M. (1991). The computer for the 21st Century. *Scientific American*, *265*(3), 94–105. https://doi.org/10.1038/scientificamerican0991-94

Wiedemann, G. (2019). Proportional classification revisited: Automatic content analysis of political manifestos using active learning. *Social Science Computer Review*, *37*(2), 135–159. https://doi.org/10.1177/0894439318758389

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136. https://www.jstor.org/stable/20024652?seq=1#metadata_info_tab_contents.

Yang, H., & Sundar, S. (2020, May). *Machine heuristic: A concept explication and development of a scale* [Paper presentation]. Annual Conference of the International Communication Association, Gold Coast, Australia.

Zheng, Y., Zhong, B., & Yang, F. (2018). When algorithms meet journalism: The user perception to automated news in a cross-cultural context. *Computers in Human Behavior*, *86*, 266–275. https://doi.org/10.1016/j.chb.2018.04.046