



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

**Using machine learning to learn machines: a cross-cultural study of users'
responses to machine-generated artworks**

Xu, Kun ; Liu, Fanjue ; Mou, Yi ; Wu, Yuheng ; Zeng, Jing ; Schäfer, Mike S

DOI: <https://doi.org/10.1080/08838151.2020.1835136>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-199078>

Journal Article

Accepted Version

Originally published at:

Xu, Kun; Liu, Fanjue; Mou, Yi; Wu, Yuheng; Zeng, Jing; Schäfer, Mike S (2020). Using machine learning to learn machines: a cross-cultural study of users' responses to machine-generated artworks. *Journal of Broadcasting and Electronic Media*, 64(4):566-591.

DOI: <https://doi.org/10.1080/08838151.2020.1835136>

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

Abstract

Drawing upon prior literature on machine-generated news, this study expands the research scope to machine-generated art works in a cross-cultural context. This study combines machine learning approaches with online experiments and investigates how different genres of art works 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 art works 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.

Keywords: Cross-cultural comparisons, machine heuristic, AI-generated content, machine agency, automated text analysis, artificial intelligence

“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 are precipitating the public discussion about the promises and perils of these technologies. One of the 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, Edwards, Spence, & Shelton, 2014). The difficulty to differentiate information sources may render users susceptible to fake news and consequently make 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, Liu, Elhoseiny, and Mazzone (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 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 art works, especially poems and paintings. If AI-generated art works can impose 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 the AI technologies could challenge the existing criteria for aesthetic evaluation of the art works, 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 affection. The results may inform scholars of both the benefits and risks of using AI to generate art works.

Unlike prior research on users' 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, Rozza, & Arvidsson, 2014). The combination of these two methods in this study would feature the following advantages. First, this study allows participants to produce open-ended responses after viewing AI-generated art works and enables researchers to find the latent topics from users' responses. Compared to closed-ended questions, open-ended written responses could provide a

richer understanding of users' psychological reactions, linguistic patterns, and attitudes toward machines. With the techniques of machine learning being more mature and accessible than before, analyzing large collections of digital texts is becoming feasible and practical (Gerrish & Blei, 2012). Second, although machine learning has been used to analyze the discussion on social media, rarely have researchers taken into consideration the covariate information or experiment conditions in the automated textual analyses (Roberts et al., 2014). To fill the gap, this study adopts one of the unsupervised machine learning methods to scrutinize how topic prevalence and topic content varies in the effects of authorship cues and the genres of the art works.

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: US, Germany, and China. As viewers' attitudes toward machines are situated within their traditional value, 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 the 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 experiment treatment. On the theory level, this study tests the explanatory mechanism of machine heuristic (Sundar, 2008) and examines whether users' understanding of machines is contingent upon the nature of tasks and sociocultural factors.

Literature Review

Machine-Generated Works

While algorithm-generated content is not fully-fledged, it reflects a new stage of technology evolution and a tendency to reshape and transform the news industry (Zheng, Zhong, & Yang, 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 the traditional journalism and how audiences perceive the 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 on a technological level, algorithm is 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 news industry (Thurman, Dörr, & Kunert, 2017).

As the convergence of computing and news production has begun to alter the 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 its credibility. Consistent with Clerwall's (2014) findings, van der Kaa and Kraemer (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 advance 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 underscore 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 the abovementioned research has focused on how machine-generated news affects audiences' perception and the news industry, little research has concentrated on art works 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 impact on viewers' perception of machines. Rau, Li, and Li (2009) compared Chinese and German participants' attitudes toward robots and found that Chinese participants perceived a robot to be more likable and trustworthy and were more likely to accept the implicit recommendations from the robot. German participants, however, were more anxious about the negative impact of the social robots (Li, Rau, & Li, 2010). Bartneck, Suzuki, Kanda, and Nomura (2007) also looked at the intercultural differences and found that among US, China, and Germany, US 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 users' 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. Comparatively, 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), technology 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 that they affiliate with. Turja and Oksanen (2019) applied SCOT to compare people's robot acceptance in 27 EU members. They found that when testing enjoyment, anthropomorphism, and perceived behavior control, individuals' perception of social robots varies with cultures even if they all come from Europe. Above all, 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 art works.

Machine Agency

To comprehend how users' perception of art works may be decided by the authorship cues, the MAIN model is applied as the theoretical framework in this study (Sundar, 2008). In this model, technology affordances serve as cues that affect how users evaluate technology interfaces (Sundar et al., 2015). Four types of affordances on user interfaces can cue the cognitive heuristics: modality, agency, interactivity, and navigability. 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, Sundar (2008) proposed the machine heuristic, which reflects the mental shortcut wherein people perceive a technology to be neutral and objective in its content presentation when the technology interface demonstrates machine agency cues. As some examples, researchers found that news selected by a computer was perceived as higher in quality than that selected by a human editor (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 facing a machine agent than a human agent (Sundar & Kim, 2019). While machine heuristic has been applied in users' attitudes toward news production, limited research has been conducted on whether and how the approach can be used to explain users' interpretation of creative works. When building the measures for machine heuristic, Yang and Sundar (2020) distinguished mechanical tasks and human tasks and suggested that users' interpretation of machines as efficient, objective, and unemotional could reinforce the heuristic that machines are worse than humans in performing human tasks. As art works can be considered as a form of human creativity, users' evaluation of machine-created art versus human-created art should demonstrate a certain level of disparities.

The effects of machine heuristic may vary in individuals' expectations and interpretations of automation. When Waddell (2018) looked into 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, Edwards, Edwards, and Lin (2019) found that users' low suspicion of machine-selected information was a strong predictor of the 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 that the effect of the machine heuristic is dependent upon individual attitudes, social norms, and news types, it is speculated that the genre of artworks and sociocultural differences may 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, McAninch, Slone, & Moore, 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 topic modeling (STM). STM allows researchers to examine participants' open-ended responses alongside their demographic information, predispositions, and the effects of experimental manipulation (Roberts et al., 2014). In the current context, there are a few advantages of using STM over the 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, Ng, & Jordan, 2003) and correlated topic models (Blei & Lafferty, 2007). According to Roberts 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 deliver the structured changes in topical prevalence and topic content.

As this study seeks to understand users' interpretations of machine-generated versus human-generated content in an experiment setting, STM is a tenable approach to capturing the

effects of different genres of art works 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 US, China, and Germany?

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

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

RQ4: Based on the open-ended remarks, how do participants respond to machine-generated art works vs. human-generated art works differently across US, 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 experiment treatment on users' attitudes toward machine-generated works, predictive validation is used. We selected the variables of perceived quality of art works, perceived imaginativeness, engagement, and spatial presence, as these variables could indirectly reflect how participants understood the art works 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 art works across US, China, and Germany?

RQ6: Based on the closed-ended responses, how do different authorship cues affect participants' perception of the art works across US, China, and Germany?

Method

Participants

In US, MTurk was used to recruit participants. A total of 422 US 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 was used to recruit participants. A total of 486 participants enrolled in the study. After removing the 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.

Experiment Stimulus and Procedures

An online experiment was conducted in US, China, and Germany. A 2 X 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 of

the 2017 Robot Art Competition (Robohub, 2017). The other one, titled “Pleased,” was created by Rob Boss of HEARTalioN at Halmstad University, which received the 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 native language. To minimize the confounding effects, we chose the 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 “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 some demographic questions. Then they were asked to read the poems or view the paintings. After they viewed each piece of art works, they were asked to describe their feelings about the works. Then they were answered closed-ended questions including perceived quality of the works, perceived imaginativeness of the works, their sense of engagement, and experiences 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 closed-ended questions.

Measures

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

The measure of perceived imaginativeness (US: $M = 3.36$, $SD = 1.02$, $\alpha = .93$; China: $M = 3.72$, $SD = .68$, $\alpha = .78$; Germany: $M = 2.60$, $SD = 1.05$, $\alpha = .88$) 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 (US: $M = 3.20$, $SD = .99$, $\alpha = .88$; China: $M = 3.10$, $SD = .76$, $\alpha = .83$; Germany: $M = 1.78$, $SD = .87$, $\alpha = .89$) was adapted from Lombard, Ditton, and Weinstein’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 (US: $M = 2.79$, $SD = 1.21$, $\alpha = .95$; China: $M = 2.99$, $SD = .79$, $\alpha = .79$; Germany: $M = 1.60$, $SD = .77$, $\alpha = .92$) 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 art works (i.e., each document). To answer RQ1 to RQ4, STM was conducted using *R*. Data analyses followed Roberts and colleagues’ (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 lowercased. We also used stemming to remove suffixes and prefixes to retrieve the word root so that words such as “paintings” and “paint” 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 the top features from the matrices to answer RQ1.

Based on the 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, Gerrish, Wang, Boyd-Graber, & Blei, 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. In this iterative process, we were able to identify the model that has the 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 et al., 2014).

As 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 different countries as independent variables were conducted to confirm the effects of authorship cues and content genres on participants' perception of the art works (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 US 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 was shown in Figure 1.

To answer what the main topics were among participants' responses to the poems and the paintings across US, China, and Germany (RQ2), five topics were identified among US participants' responses. These topics were identified as relaxation (topic 1), senses (topic 2), affective appraisals (topic 3), colored paintings (topic 4), and positive evaluations (topic 5). Three topics were identified among Chinese participants' responses. They were abstraction and vitality (topic 1), affective appraisals (topic 2), and maternal love (topic 3). Four topics were identified among 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 were shown in Figure 2 and Figure 3.

To answer how participants perceived poems and paintings differently across US, Germany, and China (RQ3), results suggested that among US 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 US 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 about 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 were shown in Figure 4. The effects of covariates, authorship cues, and genres on participants' discussion topics were shown in Table 1-3.

To answer how participants responded to machine-generated works vs. human-generated works differently across US, Germany, and China (RQ4), results from STM suggested authorship cues significantly predicted US participants' rate of word use within the topic about

relaxation (topic 1), $B = .08$, $p = .034$. However, authorship cues did not predict Chinese participants' nor German participants' rate of word use within any topics.

As validation, to answer how authorship cues and genres of content affect participants' perception of the creative works across US, 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 US participants ($M = 3.07$, $SD = .94$), $p < .001$. US participants scored lower than Chinese participants ($M = 3.33$, $SD = .58$), $p < .001$ (Figure 5).

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 US participants ($M = 3.53$, $SD = .99$), $p < .001$. US 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, $F(2, 822) = 215.02$, $p < .001$, where German participants ($M = 1.78$, $SD = .87$) felt less engaged than Chinese participants ($M = 3.10$, $SD = .76$), $p < .001$ or US participants ($M = 3.13$, $SD = 1.01$), $p < .001$. However, Chinese participants did not differ from US 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, $F(2, 822) = 190.98, p < .001$, where German participants ($M = 1.60, SD = .77$) scored lower than US participants ($M = 2.79, SD = 1.21$), $p < .001$. US participants scored significantly lower than Chinese participants ($M = 2.99, SD = .79$), $p = .026$ (Figure 8). Consistent with STM, authorship cues did not affect users perceived quality, imaginativeness, engagement, and spatial presence. Results of three-way ANOVAs were shown in Table 4.

Discussion

This study seeks to understand whether and how users perceive machine-generated art works and human-generated art works differently in a cross-cultural context. By combining experimental studies and machine learning methods, this study suggests that people developed different feelings and form different discussion topics when viewing creative works. Based on the different genres of the art works, participants assigned various topics and used different linguistic elements to comment on the works.

Based on the top features and the main topics related to participants' word use, both US 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 US 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. US participants were generally more positive about the art works. They put emphasis on how the art works were relaxing, well-written, colorful, and soothing. Chinese participants used more concrete imageries 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 art works were calming, overall their comments tended to be more negative. They mentioned that the art works 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 over 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 of the works.

The differences in their word use and topics could also be attributed to sociocultural factors. Specifically, German participants’ ratings of the art works were significantly lower than US or Chinese participants on all these dimensions. Based on a European survey (Technik Radar, 2019), it may be because Germans were 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 the society (Bernotat & Eyssel, 2018). Such preferences may be associated with German consumers being more sensitive to data privacy compared to other cultures such as US and Japan (Krasnova & Veltri, 2010; Nitto, Taniyama, & Inagaki, 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 where they found that genre played a role in users' evaluation of machine agency, the findings suggested that genre had main effects on participants' discussion of the creative works. US participants described the paintings to be more colorful and relaxing and used more emotional words such as "lost" or "depressed." They used 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 were related to a eulogy to the 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 art works. Paintings were perceived to be in 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 US participants used more words such as calm, enjoy, and relax when viewing human-authored paintings than when viewing machine-authored ones, meaning that they felt more relaxed at human-authored paintings. Overall, although participants from cross-cultural backgrounds assigned different

discussion topics to these art works, their responses were independent of the agency cues. Considering that people initially reported lower acceptance of computers taking interpretative roles (e.g., editorial writers, novelists) than the ones taking 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 taking interpretive roles like artists and composers has accrued. 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 augmented. 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 three-way ANOVAs.

According to the machine learning results, participants' prior AI use frequency influenced their discussion topics. For both US 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 art works. The results can serve as additional evidence to prior research findings that technology use experiences affected users' psychological responses to computing technologies (Johnson et al., 2004).

This study has theoretical, methodological, and practical implications. It applies the machine heuristic to understanding the role of authorship cues in users' evaluation of the art works. Prior research on machine heuristic has primarily focused on users' perception of the machine performances on 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 art works in a relatively objective and efficient way (Grimmer & Stewart, 2013). First, as participants may develop complex and extended feelings toward art works, STM allows them to express feelings without being limited by close-ended questions. Second, STM incorporates different experiment 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 not only learn the patterns and themes from large amount of digital data, but also use quantitative methods to validate and complement the findings. Future research could continue to combine computational methods with traditional social science methods to fathom 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 argued 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 one hand, machine-created works will significantly reduce the cost of aesthetic evaluation. Considering that social norms and culture backgrounds play a significant role in shaping users' attitudes toward machine-generated art works, the presentation of machine-created art works 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 value and social norms. Although we are facing the 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, this study examines whether there exist semantic commonalities and disparities in participants' attitudes toward machine-generated art and human-generated art. This study suggests that while genres and cultures affected participants' discussion topics and word use, the differences between participants' responses to machine-generated art works and human-generated ones were not evident.

The 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 experiment 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 experiment 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 the machine agencies 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 variances of the messages, the sampling of the creative works was still limited. Future research could use more experiment stimuli to study the effects of machine-generated art works on participants' attitudes. Second, unlike the assessment of news which could be measured by its credibility and authors' writing competency, the criteria for evaluating art works 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 the mental shortcuts (see Bellur & Sundar, 2014).

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, 217-230.
- 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, 116-137.
- 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. In *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.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 3, 416-431.
- 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* (pp. 288-296).
- Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice*, 8, 519-531.

- 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*, 317-332.
- Deahl, D. (2018 August 31). How AI-generated music is changing the way hits are made. *The Verge*. Retrieved from <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*, 168-189.
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics, 41*, 570-606.
- Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism, 4*, 700-722.
- 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.
- 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*,
- Gerrish, S., & Blei, D. M. (2012). How they vote: Issue-adjusted models of legislative behavior. In *Advances in Neural Information Processing Systems* (pp. 2753-2761).
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism, 19*, 595-610.

- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*, 267-297.
- 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.
- Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review, 52*(1), 273-292.
- Krasnova, H., & Veltri, N. F. (2010, January). Privacy calculus on social networking sites: Explorative evidence from Germany and USA. In *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*. London: 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*, 175-186.
- Liang, H. (2014). Coevolution of political discussion and common ground in web discussion forum. *Social Science Computer Review, 32*, 155-169.
- Lichtenegger, F. (2018 May 9). Dieser Bot schreibt tiefgründige Gedichte auf Instagram. *Vice*. Retrieved from <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*, 635-657.

- Lombard, M., Ditton, T. B., & Weinstein, L. (2009). Measuring presence: The temple presence inventory. Paper presented at the *Proceedings of the 12th Annual International Workshop on Presence*, 1-15.
- Merriman, C. (2018 August 15). Microsoft's xiaoice chatbot has written a book of poetry. *The Inquirer*. Retrieved from <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*. Retrieved from <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, 61-73.
- Nass, C. I., Lombard, M., Henriksen, L., & Steuer, J. (1995). Anthropocentrism and computers. *Behaviour & Information Technology*, 14, 229-238.
- 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.
- Peng, Y. (2018). Sa me candidates, different faces: Uncovering media bias in visual portrayals of presidential candidates with computer vision. *Journal of Communication*, 68, 920-941.
- 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*, 1-18.

- 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, 587-595.
- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York, NY, US: Cambridge University Press.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). Stm: R package for structural topic models. *Journal of Statistical Software*, 10, 1-40.
- 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, 1064-1082.
- Robohub. (2017 March 19). 2017 robot art competition winners announced. *Robohub*. Retrieved from <https://robohub.org/2017-robot-art-competition-winners-announced/>
- Šabanović, S. (2014). Inventing Japan's 'robotics culture': The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science*, 44, 342-367.
- Spence, P. R., Edwards, C., Edwards, A., & Lin, X. (2019). Testing the machine heuristic: Robots and suspicion in news broadcasts. Paper presented at the *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 568-569.

- 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, 480-499.
- 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. Cambridge, MA: The MIT Press.
- 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). UK: 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).
- Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1), 52-72.
- Technik Radar (2019). Was die Deutschen über Technik denken. Retrieved from https://www.koerber-stiftung.de/fileadmin/user_upload/koerber-stiftung/redaktion/technikradar/pdf/2019/Technikradar-2019_Booklet.pdf
- Trentini, Y. A. (2017, March 21). Computer generated poetry will knock your socks off. *Medium*. Retrieved from <https://medium.com/@Yisela/computer-generated-poetry-will-knock-your-socks-off-763c815a1b52>

- 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, 1240-1259.
- Turja, T., & Oksanen, A. (2019). Robot acceptance at work: A multilevel analysis based on 27 EU countries. *International Journal of Social Robotics*, 11, 679-689.
- Van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, 6, 648-658.
- Van der Kaa, H., & Krahmer, E. (2014). Journalist versus news consumer: The perceived credibility of machine written news. Paper presented at the *Proceedings of the Computation Journalism Conference, Columbia University, New York, 24-25*.
- Waddell, T. F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, 6, 236-255.
- Weiser, M. (1991). The computer for the 21st Century. *Scientific American*, 265, 94-105.
- Wiedemann, G. (2019). Proportional classification revisited: Automatic content analysis of political manifestos using active learning. *Social Science Computer Review*, 37, 135-159.
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121-136.
- Yang, H., & Sundar, S. (2020, May). Machine heuristic: A concept explication and development of a scale. Paper presented at the annual conference of the International Communication Association.
- 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.

Figure 1
Text plots of the top features: US, China, and Germany

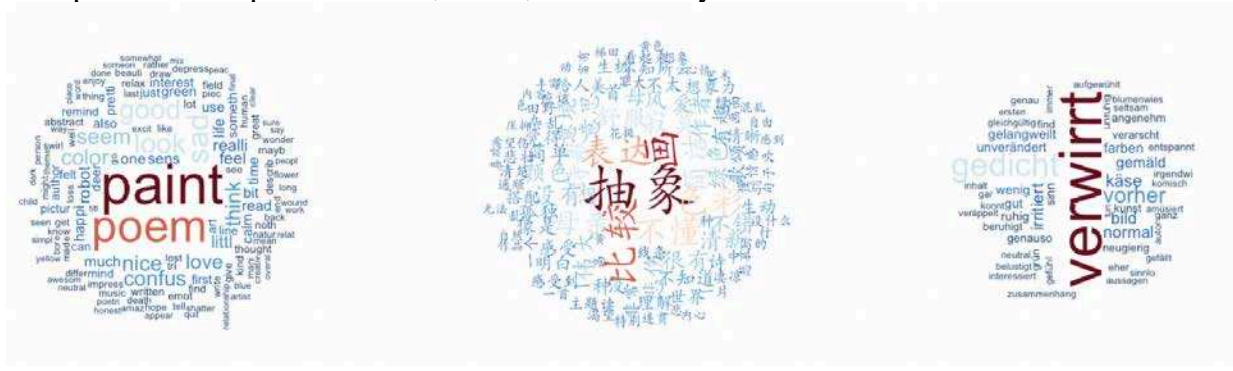


Figure 2
Document-topic proportions: US, China, and Germany

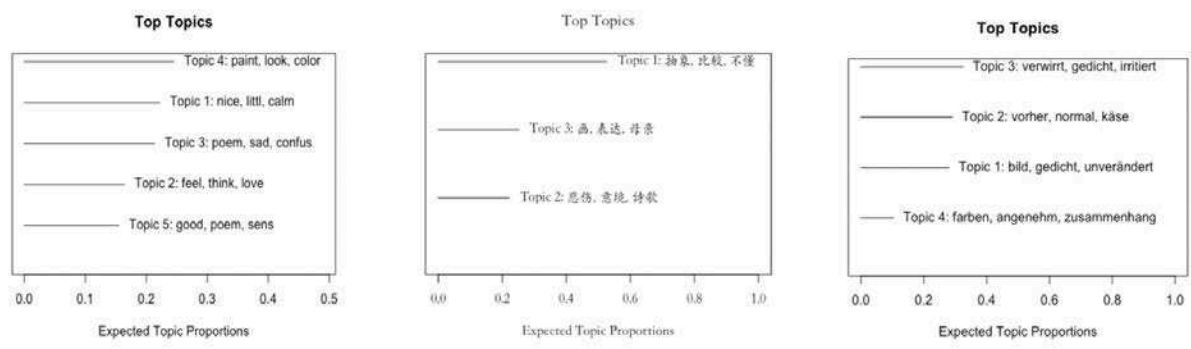


Figure 3
Sample word use based on highest probability and exclusivity within topics

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, 悄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 werdenso einen Unsinn kann man doch nicht Gedicht nennen	Ganz angenehm, die Wellen und die Farben beruhigen.	

Figure 4
Effects of genres on topic discussion: US, China, and Germany

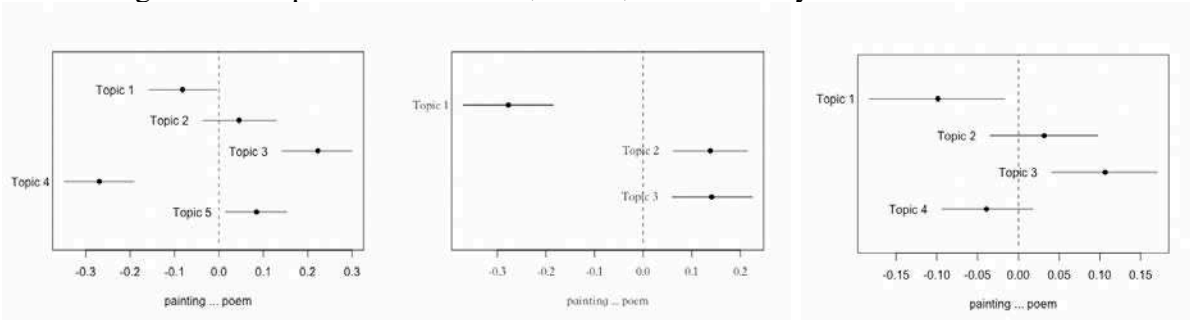


Figure 5.
Three-way interaction on perceived quality of the works

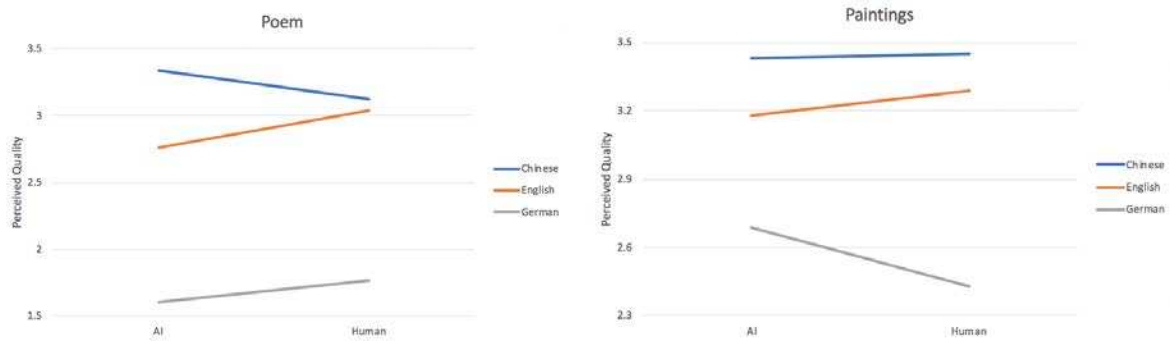


Figure 6.
Three-way interaction on perceived imaginativeness of the works

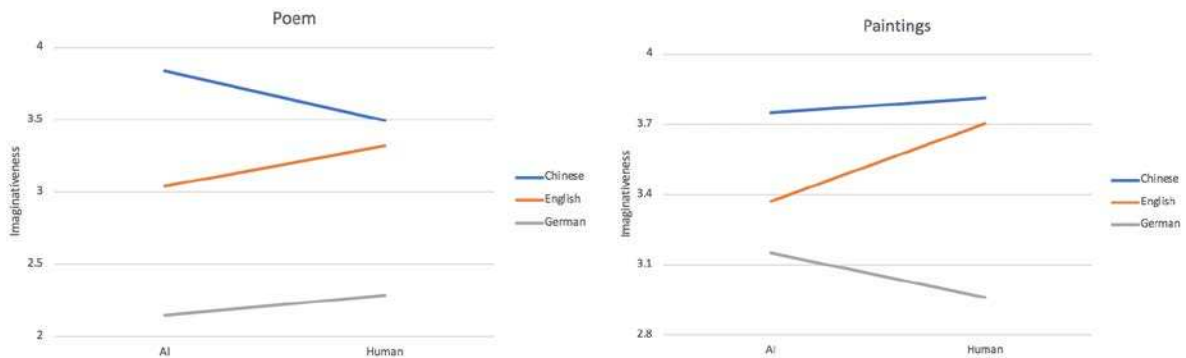


Figure 7.
Three-way interaction on engagement of the works

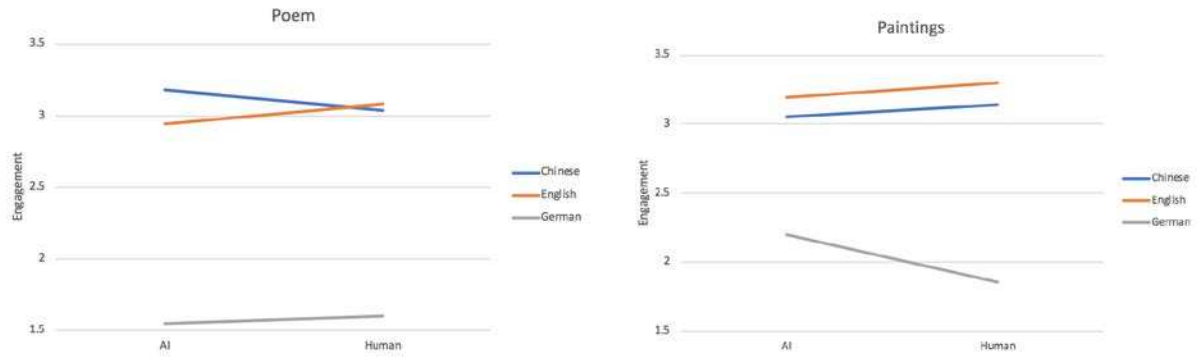


Figure 8.
Three-way interaction on spatial presence

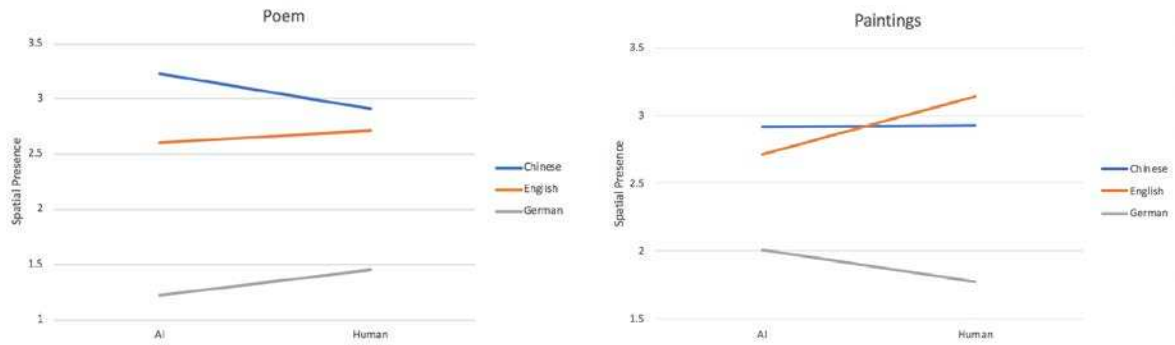


Table 1.

The effects of covariates, authorship cues, and genres on US participants' topics

Variables	Relaxation	Sensing	Affective	Colored	Positive
	Topic 1	Topic 2	appraisals	paintings	evaluations
	<i>B (SE)</i>	<i>B (SE)</i>	<i>B (SE)</i>	<i>B (SE)</i>	<i>B (SE)</i>
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)*

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

Table 2

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

Variables	Abstraction/vitality	Affective appraisals	Maternal love
	Topic 1	Topic 2	Topic 3
	<i>B (SE)</i>	<i>B (SE)</i>	<i>B (SE)</i>
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)***

Note: * $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3.

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

Variables	Invariableness	Normality	Negative evaluations	Relaxation
	Topic 1	Topic 2	Topic 3	Topic 4
	<i>B (SE)</i>	<i>B (SE)</i>	<i>B (SE)</i>	<i>B (SE)</i>
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)

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4

Results of Three-way ANOVAs

Variables	Perceived quality		Imaginativeness		Engagement		Spatial presence	
	<i>F</i>	ω^2	<i>F</i>	ω^2	<i>F</i>	ω^2	<i>F</i>	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

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