

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Judgements of AI-Art: New insights into both aesthetics and human reaction to AI (#111260)

Created: 10/31/2022 07:22 AM (PT)

This is an anonymized copy (without author names) of the pre-registration. It was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) should be made available by the authors when the work it supports is made public.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Do people differentially evaluate art that is labelled as Artificial Intelligence (AI)-made or Human-made, and are these evaluations affected by aspects of the rater's personality and attitudes? In addition, to what degree do these ratings depend on the type of painting (abstract vs representational)?

H0: People do not view art labeled as AI-made differently from art labeled as Human-made

HA: People do view art labeled as AI-made differently from art labeled as Human-made

H0: People's ratings do not depend on the type of painting.

HA: People's ratings do depend on the type of painting.

H0: Ratings on the type of painting do not depend on people's level of openness.

HA: Ratings on the type of painting do depend on people's level of openness

3) Describe the key dependent variable(s) specifying how they will be measured.

We will be measuring raters' evaluations of art that is labelled as AI- or Human-made, but in reality is all AI-made. These evaluations consist of 9 criteria: likability, perceived profundity, perceived worth, perceived beauty, induced emotion, ability to imagine a story, personal meaning, perceived effort, and estimation of time to create the work [free-response for each : Day, Hour, Minute]. Each, except Time, is measured on a Likert scale of 1-5 [Not at all... Extremely].

4) How many and which conditions will participants be assigned to?

This design is within-subjects (i.e., one condition). Participants will rate 30 paintings, and a randomized half of the 30 (~15) will be labelled as AI-made, and the remainder as Human-made. A pre-determined half of the paintings are also abstract, and the other half representational.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will run a Bayesian linear mixed model, using pilot data to inform our prior distributions for the effects of AI/human-made and painting type (abstract/representational), on each of the 9 evaluation criteria (i.e., each one being a DV), with fixed effects being the label (AI- or human-made), the painting type (abstract vs. representational), and an interaction of the two. Based on pilot data, we suspect that the interaction will not be significant, but we will attempt to replicate this finding regardless. Participants and paintings will both be modeled as random effects. We will include a random slope of AI/human-made for paintings (i.e., the slopes are expected to be different depending on the maker). We suspect that each of the 9 DVs will reflect a human-made bias, such that ratings for human-made art will be higher (both in Likert scales and the free-response Time question).

In addition, we will assess how empathy, as measured by the Toronto Empathy Questionnaire, affects these ratings. In one set of moderation analyses, as determined by interactions in our linear mixed model, we will be assessing empathy as a moderator for difference scores of AI-Human ratings per participant. In another set of analyses, we will assess empathy as a moderator of AI-labelled ratings, and empathy as a moderator of Human-labelled ratings separately.

We will run the same analysis above on empathy for the following as well: Openness (from the NEO-FFI-3), perceptions of human creativity (Creative Mindset Scale), and attitudes towards artificial intelligence (The General Attitudes towards Artificial Intelligence Scale). We specifically expect Openness to interact with ratings of abstract vs representational paintings, such that those more open rate abstract art higher.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Two attention checks will be administered at the beginning of the experiment. If participants fail to provide a correct response to both attention checks, as dictated in our IRB, they will immediately be omitted from the experiment. If participants fail to fully complete the experiment, their data will be omitted from analyses. In addition, we will ask participants their (a) birthyear at the beginning of the study and (b) age at the end of the study (Kennedy et al., 2020). Those that do not match will be excluded from data analyses.

Outliers will be defined as participants whose ratings are +/- 3 standard deviations above/below the mean for each of the DVs, and their data will be excluded from analyses.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, 8(4), 614–629. <https://doi.org/10.1017/psrm.2020.6>

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

150 participants will complete the study.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will also be asking the following questions:

"How much do you like AI-created art?", "How much do you like human-created art?", "To what extent do you care about the purely perceptual properties of art?", and "To what extent do you care about the communicative aspect of art?". These scales are 5-point Likerts: [Not at all...Very much]

Answers to these are suspected to be correlated in some way with the 9 ratings.

We will provide a general free-response question asking, "What aspects or properties do you value in a creative product like art?" Participants will be asked to "Write as much as possible in as much detail as possible." This data will be analyzed with Natural Language Processing pipelines on huggingface.co to determine key shared words among participants who like human-made or AI-made art better. We particularly expect shared words such as: "meaningful", "positive", "negative", "wonder", "beautiful", "profound", and/or "pleasing", in line with our ratings and questions.

We will correlate the following with one another and the other DVs previously mentioned:

-□Cognitive Reflection Test scores

-□Demographics (e.g., age, political orientation)

-□Single-item questions assessing attitudes towards AI. These items are:

[("How frequently do you use "smart" technology, like home-assistants (e.g., Siri, Amazon Alexa) or other artificial intelligence products?"), 1-5 Likert ranging from "Never" to "Extremely frequently, several times a day"; ("How valuable do you find "smart technology" to be in your personal life?"), 1-5 Likert ranging from "Not valuable at all" to "Extremely valuable"; ("Deep down inside, AIs have a true self that is morally good."); "Deep down inside, humans have a true self that is morally good."; "Artwork produced by an AI reflects their true self."; "Artwork produced by a human reflects their true self."; "AIs make decisions, problem-solve, and learn."; "AIs produce original ideas and think abstractly.", "AIs think, feel, and act like human beings do."; "AIs create artwork like human beings do."), 1-5 Likert: "Completely disagree" to "Completely agree"; ("How long do you think it will take for AIs to develop consciousness?"), Slider: 1950s (advent of computer) to very far future. Option for "never"; ("How familiar are you with AI?"), 1-5 Likert ranging from [Not familiar at all – Very familiar]; ("What is your relationship with AI like?"), [Negative – Neutral – Positive]; ("Before this survey, have you ever used AI to create artwork before? If you answered 'yes' above, please explain."), [Yes / No] – [if yes, free response]

We will check normality assumptions across ratings: if $|skew| > 3$ and $kurtosis > 4$ (Kline, 1998), we will apply a rank-base inverse normal transformation.

Kline, R. B. (1998). Principles and practice of structural equation modeling. New York, NY: Guilford Press.