Bias in Perceived Creativity of AI Artwork

UGS 303 Being Creative | Final Research Project

Student: Connor McKelvey
Instructor: Professor Dempster
November 13, 2025

Table Of Contents

Introduction

Following the rapid advancement of neural networks approaching the “revered” level of general intelligence, as noted by Dr. Bruce Porter, former Chair of Computer Science at the University of Texas at Austin, questions relating to the creativity of generative artificial intelligence have become urgent. After the sudden release of OpenAI’s premier large language model, generative artificial intelligence (AI) systems demonstrate the ability to produce artwork, music, and writing that rivals human output. This raises a fundamental question: Do people perceive human-made creative works as inherently more creative than AI-generated works when the source is revealed? Understanding this is critical, because if humans systematically favor human creators, AI-generated works may be undervalued despite their technical sophistication or aesthetic appeal. If this bias is found to be true, society might be hurt by neglecting to acknowledge incredible acts dreamed up by artificial intelligence. In this study, creativity is judged by human perception and is defined as requiring both originality and effectiveness. This is consistent with Runco and Jaeger’s widely cited definition across creativity science: “Creativity requires both originality and effectiveness” (Runco & Jaeger, 2012, p. 92).

The discussion of audience bias or hostility toward generated art has become increasingly important with the rise of models such as DALL·E and other artificial intelligence systems that are often accused of “stealing” or remixing original works (OpenAI, DALL·E: Creating Images from Text). While studies such as “AI Art is Theft: Labour, Extraction, and Exploitation” by Trystan S. Goetze concluded that AI image generators “involve an unethical kind of labour theft” (Goetze), which leads to a prejudice against these tools, many fail to address the deeper question of whether this response stems from an underlying discomfort with, or resistance to, machine creativity itself. This project hopes to approach the innovation in AI-generated art from the perspective of creativity science, focusing on the perception of artwork and whether humans hold some resentment or hesitation in rating AI generated outputs as extremely creative.

This research topic is not only globally applicable and timely during this age of innovation but also personally relevant to my own studies and goals. As an Electrical and Computer Engineering student at the University of Texas in the Robotics Honors Program, I am concerned that humans might reject creative solutions that robots using artificial intelligence might put forth. If this study reveals a negative bias towards generative AI, we could use this result as context for future research or real situations that relate to human-robot interaction.

Primary research question

The primary research question of this study is: **Does knowledge of the creator (human vs AI) affect perceived creativity of visual images?** To deduce a conclusion about this question we will test whether revealing the source of an art source changes people’s judgments of creativity. One of the main focuses of our methodology is to standardize the presentation of the art sources so that we can make meaningful conclusions about the creativity of the sources without any bias of style, color, complexity, etc. The methodology used to accomplish our primary research question and study goal is detailed in the Methodology section below.

Methodology

Overview & Procedure

The research study is set up to help answer the primary research question by focusing on stimulus standardization and presentation design. The data collection is done through an online Qualtrics form where participants are presented with two images side by side and asked to rate their creativity on a 1-7 Likert Scale, which is a "very popular psychometric item scoring schemes for attempting to quantify people's opinions, interests, or perceived efficacy of an intervention" (Bishop). The form was split into two blocks with a set of demographic questions in between to reduce any carry over effects. The first block presented artwork with no sources, while the second block correctly labeled artwork as "Human Made" or "AI Made." To prevent participants from remembering the images from the first block, the stimuli was split up into two sets (A and B) of five pairs with equal complexity. The form has two versions that are selected from random. In version one the participant is presented with set A without source labels and then presented with set B with source labels. In version two this order is reversed (unlabeled set B, followed by labeled set A). This study design allows us to measure if the addition of source labels has any effect on the perception of the art pieces' creativity.

Participant population information and demographics

The Qualtrics survey is completely anonymous and was mainly sent out to students at the university and to other groups. In the middle of the survey they will be asked their age, gender, level of education, prior experience with art, and prior experience with artificial intelligence. These basic demographics will help us understand any trends that might be seen in the responses and rule out other possible factors.

Stimulus design

This is the most important aspect of this research survey to help guarantee we get results based on the creativity of each image, not the complexity, style, medium, etc. Designing the stimuli also was by far the most time consuming process in preparing the survey as it utilized programming and simulations to help standardize the artworks. To achieve this standardization, semi-abstract monochromatic line art was used, as AI generated works would be near impossible to tell apart from human made works without labeling. The technology is not yet good enough and it is harder to compare the complexity of multi-colored art pieces. The human made stimuli was sourced from Dribbble.com (Dribble), a platform where artists share their work with brief descriptions. The AI generated lineart was made on Recraft.ai (Recraft) which has an incredibly powerful vector graphic generator from text. I would prompt the model with a slightly altered version of the brief description attached to the human made art, adding lines like "focus on abstractness and creativity". Once I had ten pairs of art, I began the important process of standardization. The original goal was to draw all stimuli using a two jointed drawing robot so that all images were presented on the exact same medium drawn in an algorithmic way. This further solidifies that participants' perceived creativity is not due to differences in presentation. However, given the time restraints and the scale for this project I settled for simulating this process. First, both images were converted to .SVGs (Scalable Vector Graphics) which store graphics as a list of path objects. Secondly, the stimuli was converted into sets of (X, Y) coordinates using Python. These coordinates were scaled to fit into the range of the drawing arm and simplified using the Ramer-Douglas-Peucker algorithm (Douglas & Peucker) using an ϵ = 5.0 (where ϵ is the maximum allowable distance between the original and simplified curves). This simplification allowed us to quantify the complexity of each drawing, see **Fig 3**. Finally, the sets of (X, Y) coordinates were converted into instructions that could be sent to a two jointed drawing robot. This was done using inverse kinematics equations as seen in **Fig 1**.

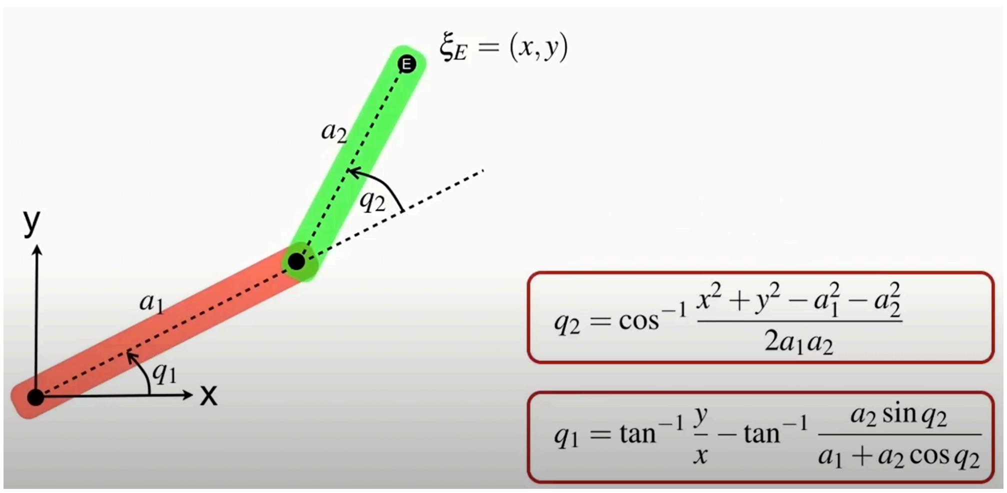


Fig 1. Inverse kinematic equations and robot arm visualization

These instructions could be “START”, “PEN UP”, “PEN DOWN”, “END”, or “(shoulder θ_1, elbow θ_2)”. The command instruction files can be found in the appendix. The final stimuli in the form presented to the participants were exported from the website simulation of the robot drawing arm (**Fig 2.**).

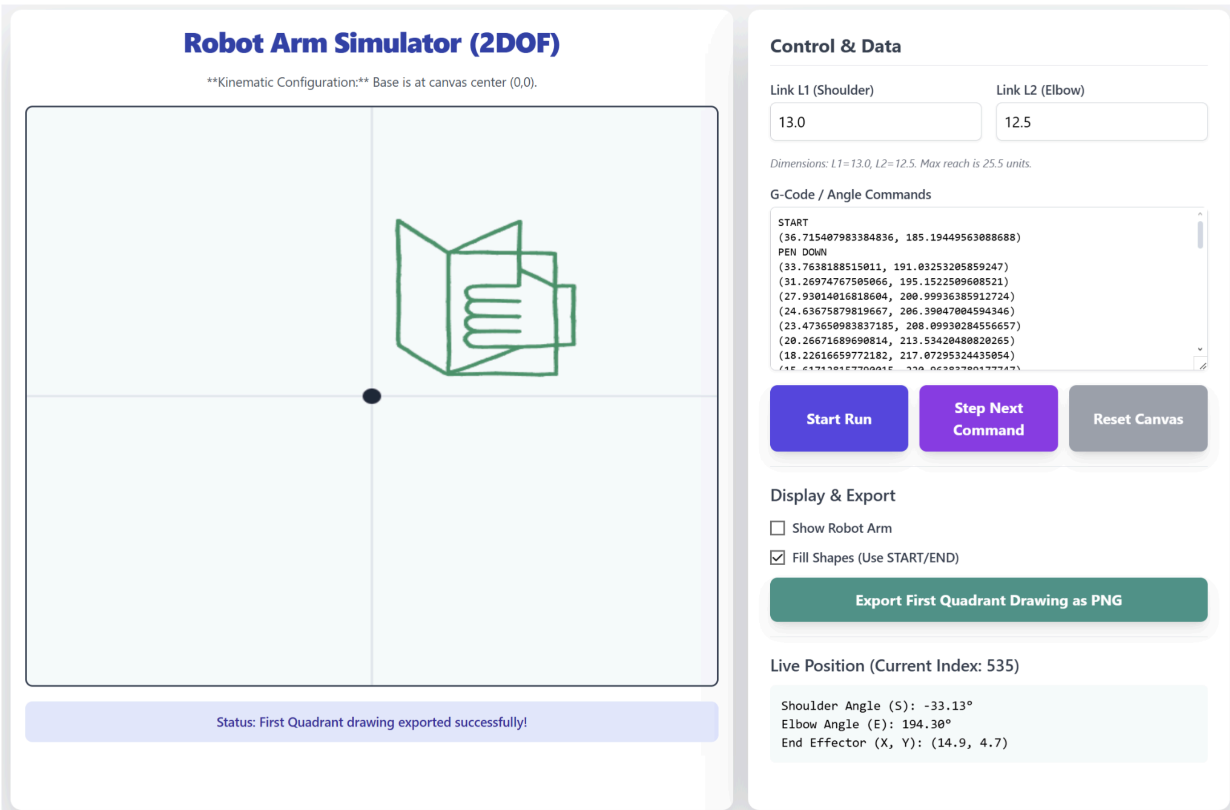


Fig 2. Drawing of "the_readers_touch.svg" made using my website which you can try out via the link in Appendix B

Image	Stroke Count
on_the_peak	8
on_the_peak_AI	6
take_it_slow	8
take_it_slow_AI	15
coffee_mind	8
coffee_mind_AI	10
geometric_dog	6
geometric_dog_AI	10
lost_in_a_page	10
lost_in_a_page_AI	6
bottle_and_glass	8
bottle_and_glass_AI	7
resting_mind	5
resting_mind_AI	9
rolled_into_a_paper_roll	7
rolled_into_a_paper_roll_AI	12
russian_ballerina	25
russian_ballerina_AI	12
the_readers_touch	8
the_readers_touch_AI	3

Set A	Set B
russian_ballerina	coffee_mind
lost_in_a_page	on_the_peak
geometric_dog	take_it_slow
resting_mind	rolled_into_a_paper_roll
bottle_and_glass	the_readers_touch

Fig 3. Image complexity and complexity balanced sets (~10 strokes with some outliers due to retracing and curves)

Summary of Findings

Statistical analysis

Before making meaningful conclusions the reliability of the data gathered by the survey needs to be verified using statistical analysis. The mean rating and standard deviation of the ratings of unlabeled human artworks, labeled human artworks, unlabeled AI artworks, and labeled AI generated images were calculated. Of the 30 survey responses received, we analyzed 27, removing three for large portions of missing question answers. The analysis is shown below in Fig 4. These averages were compared to see if there are any statistically significant differences. If the unlabeled and labeled AI generated averages have a significant difference we can reject our null hypothesis stated below.

H_0 = Labeling an AI generated line artwork has no effect of the perceived creativity of the piece

H_1 = Labeling an AI generated line artwork with the source decreases the perceived creativity of the piece

	Unlabeled Human	Unlabeled AI	Labeled Human	Labeled AI
number of ratings	134	135	134	133
mean rating	3.8	4.5	3.75	4.02
standard deviation	1.39	1.44	1.31	1.43

	Unlabeled Human vs AI	Labeled AI vs Unlabeled AI	Unlabeled Human vs Labeled Human
t - stat	-4.049	2.705	0.317
p - value	0.0001	0.0073	0.7517
significant	yes	yes	no

Fig 4. Statistical analysis of the survey results after removing majorly incomplete responses

this analysis was generated by `survey_data_analysis.py` in the appendix

Discussion

These results are incredibly exciting. The difference between the creativity ratings of labeled and unlabeled AI made line art is statistically significant. This means the null hypothesis is rejected and the alternative hypothesis can be accepted. Our participants, who were mainly individuals who use generative AI on the daily, hold a negative bias against AI creativity. This aligns with current studies with similar methodologies like “Defending humankind: Anthropocentric bias in the appreciation of AI art” (Kim and Kim) which also came to a similar conclusion. Despite younger generations adopting artificial intelligence as a tool used in daily life they still struggle to accept that generative AI can possess human-like creativity. As a computer engineering student excited about the future of this technology. I’m thrilled that my results were significant and interested in continuing this study on a greater scale with more participants and a greater variation in works. However, I'm cautious about what this perception bias could mean for my field. Studies like these are reinforcing the importance of the question: Is it good that humans are skeptical of machine learning and the intelligence of neural networks, or will AI creativity be written off?

Disclosures

Non-reviewed preliminary findings

This research study was performed for the final project of my class UGS 303 Being Creative taught by Professor Dempster. However, this project is only the preliminary findings which suggest this topic should be explored further. None of the results or information in the study has yet to be reviewed and approved by Dr. Dempster or similar peers.

AI usage

Artificial intelligence has been used to assist in the research project. I have used both Chatgpt.com and Google Gemini (Google). These tools have been mainly used for assistance in programming the inverse kinematics, simulator website, and statistical analysis program. However, they also provided incredible insight on learning about research methods, helping improve the grammar, and paper structure/presentation. I am proficient in python and have spent time reviewing code written by artificial intelligence. The physical portion of this project which was not used in answering the primary research question is a drawing arm pictured in the appendix, artificial intelligence was used to help design and code the buffer system implemented by the Arduino script.

Appendix

A: STATISTICAL ANALYSIS EQUATIONS

$$\bar{x} = \frac{\Sigma x_i}{n}$$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}$ *is the mean creativity score*

x_i *represents each individual creativity score*

n *is the total number of scores in the group*

s *is the sample standard deviation.*

t *is the value used to determine statistical significance*

p *value is calculated by the below python function which performs the Welch's Independent T Test,*

chosen because it does not assume that the two groups being compared have equal variances

```
scipy.stats.ttest_ind(group1, group2, equal_var = False)
```

B: DATA

Link to Qualtrics survey which randomly shows user version A or B.

Link to main Github repository which contains all data/code/information related to this project

all links below can be found in the repository and are only added for convenience

Link to most recent dataset used for analysis

Link to Stimuli Info Sheet which includes statistical analysis

Link to Drawing Robot Simulator Website

Link to general “data” folder which contains SVGs, PNGs, and command files sent to simulator

Link to “src” folder which contains all source code used for image processing and analysis

C: CODEBASE

The codebase is located in the Github repository (shared in Appendix B). It includes two main files, a survey analysis file, and four files containing helper functions to keep things organized.

`main.cpp`

This file is the code running on the Arduino microcontroller for the physical robot detailed in Appendix D. Not used in the research study survey.

`main.py`

The main python file that is run, implements the helper files and organizes them into easy to understand functions for the user. Put any of these functions in the main if statement to run.

```
visualize_xy_file()
    - reads the first xy file in /data/xy_files/ and visualizes it in a window
move_file_into_cmd_files(src_path)
    - moves a file at src_path into the /data/command_files/ directory
generate_robot_command_from_xy(xy_filename, l1, l2)
    - expects an xy file in /data/xy_files/xy_filename
    - l1 and l2 are the lengths of the two arm segments
    - generates a command file at /data/command_file_storage/commands_xy_filename.txt
generate_robot_command_from_svg(svg_filename, l1, l2)
    - expects an svg file in /data/svg_files/svg_filename
    - l1 and l2 are the lengths of the two arm segments
    - generates a command file at /data/command_file_storage/commands_svg_filename.txt
    - might error if the svg has points out of reach of the arm
main()
    - connects to the Arduino and sends commands from the command file
```

`survey_data_analysis.py`

Reads from the dataset specified and prints out the statistical analysis

`xy_drawing_tester.py`

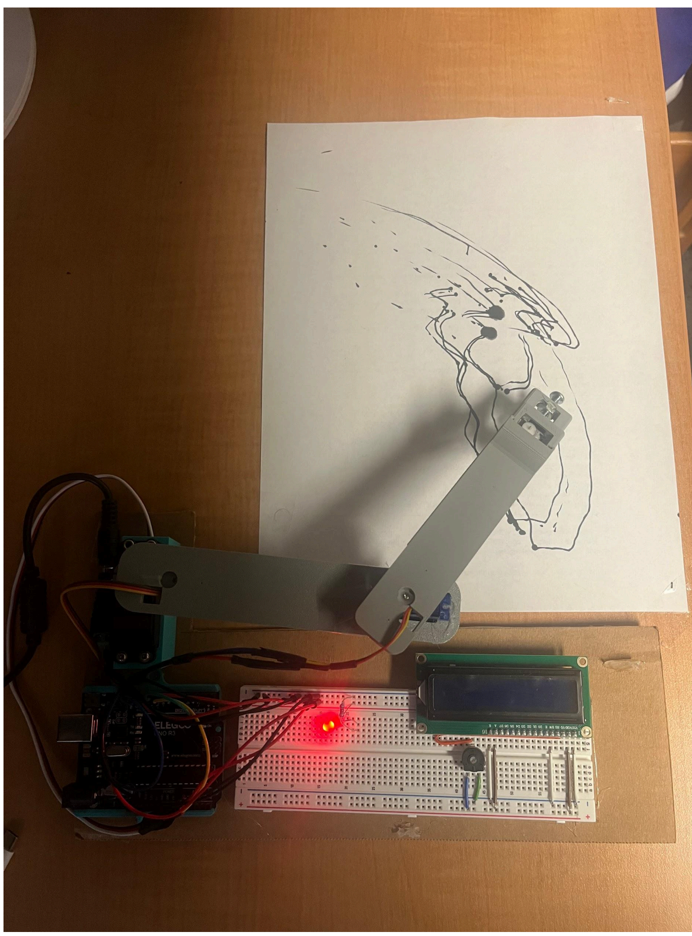
`svg_to_xy.py`

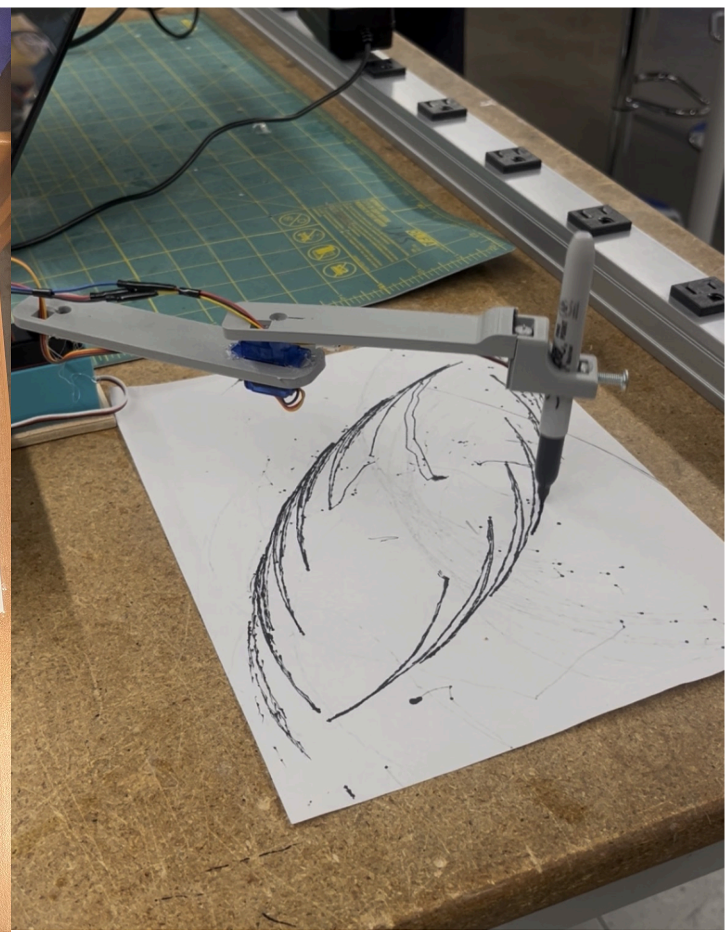
`xy_to_angles_inverse_kinematics.py`

`command_generator.py`

D: Physical robot arm design (not used in study due to time constraints)

I loved working on this project and plan to complete it by the end of term. I've got it to the point where it can be plugged into a computer and sent commands in the form of angles for the shoulder and elbow. It receives these commands and moves to the correct locations, my current problems I am facing is the range of the elbow shoulder is only 180 degrees. This prevents it from being able to reach points near the base of the arm. To solve this problem I will simply shift all the drawing away from the base. It was built using 3D printed parts from Texas Invention Works and electronic components I either already had or ordered online. The arm is controlled by an Arduino Uno microcontroller. Below are some images of the robot





Works Cited

Bishop PA, Herron RL. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. Int J Exerc Sci. 2015 Jul 1; https://pmc.ncbi.nlm.nih.gov/articles/PMC4833473

Douglas, David H. and Thomas K. Peucker. “ALGORITHMS FOR THE REDUCTION OF THE NUMBER OF POINTS REQUIRED TO REPRESENT A DIGITIZED LINE OR ITS CARICATURE.” Cartographica: The International Journal for Geographic Information and Geovisualization 10 (1973): 112-122.

Kim, Minji, and You-jung Kim. "Defending Humankind: Anthropocentric Bias in the Appreciation of AI Art." Computers in Human Behavior, vol. 143, June 2023, p. 107693. ScienceDirect, https://www.sciencedirect.com/science/article/pii/S0747563223000584.

Dribbble. Dribbble Holdings Ltd., dribbble.com. Accessed 10 Nov. 2025.

Recraft. recraft.ai. Accessed Nov. 2025.

Google. *Google AI Studio*. aistudio.google.com. Accessed Nov. 2025.

OpenAI. "DALL·E: Creating Images from Text." *OpenAI*, openai.com/blog/dall-e. Accessed Nov. 2025.

Runco, Mark A., and Garrett J. Jaeger. "The Standard Definition of Creativity." *Creativity Research Journal*, vol. 24, no. 1, 2012, pp. 92-96.

Goetze, Trystan S. "AI Art is Theft: Labour, Extraction, and Exploitation, or, On the Dangers of Stochastic Pollocks." arXiv, 20 Feb. 2024, https://arxiv.org/pdf/2401.06178.

Texas Invention Works. Cockrell School of Engineering, University of Texas at Austin, inventionworks.engr.utexas.edu.

Zherdetskii, Artur. Artur Zherdetskii Dribbble Profile. Dribbble, www.dribbble.com/ArturZherdetskii.com Accessed 10 Nov. 2025.

Connor McKelvey

Electrical and Computer Engineering

connormckelvey@utexas.edu