Do we look longer at more beautiful art? Exploring the time spent looking at human-made and AI-generated artworks

Matúš Farkaš

S5221005

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group 22

Supervisor: prof. dr. R.F.A. (Ralf) Cox

Second evaluator: Dr. rer. nat. Stefanie Enriquez-Geppert

In collaboration with: Anna-Maria Dimitriou, Maria Bendykowska, Isa Nammensma, Wouter van Dam, and Yuhan Wang

July 4th, 2025

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.

Abstract

People's attitudes toward art created by artificial intelligence (AI) had been widely researched. There is a trend of art labelled as AI-generated to be rated less positively, even though the accuracy of people in distinguishing human-made and AI-generated art is low. In the present study, we explored whether looking time (LT) is a factor that plays a role in this bias. We conducted an online experiment, and presented participants (n = 199) with 12 images of artworks, half of them being made by humans, and half of them being generated with AI. Participants rated them on different evaluative scales, such as beauty and felt emotions. Also, LT for each viewing was recorded. We found a significantly higher LT for human-made paintings, and a significant correlation between LT and rated beauty. This correlation was present only for AI-generated artworks, however. No correlation was found between LT and emotional intensity. We speculate that these results may reflect a phenomenon specific to this experiment, rather than something generalizable to people's private artistic experiences.

Introduction

During the recent years, many of us have seen the rapid advancements of what artificial intelligence (AI) can do, and the speed at which it is improving. Now we have witnessed its development to the point, where it is involved in 'artistic creativity' – creating styles, paintings, songs, poetry, and other things (Millet et al., 2023). And AI has already developed rapidly even within this specific trained area. So consequently, questions and debates arose about the nature of an art creation, the role of a human in it, and whether AI can be really considered creative, or authentic (Cox et al., 2024).

When people appraise works that were generated with AI, and compare them to ones made by humans, it could be expected that different people may have different positions in these questions, and many would feel a certain lack in creations by AI. For example, missing intentions or emotional engagement (Demmer et al., 2023). And there might be another reason for people's lower likings and acceptance of AI-created art – 'anthropocentric bias', in this case the feeling of a central and unique place of humans in creation of an artwork. This activity is often seen as what makes us different to other species. And if this one, truly authentic, human area gets threatened by AI, it can manifest into more negative evaluation of AI-generated art (Millet et al., 2023). There are, however, also proponents of different approaches that highlight the collaborative nature of an artist and technology. In their views, different kinds of tools and technology have been used and enjoyed by artists for a much longer time now, and thus the AI-generated work could truly in some cases count as art (Cox et al., 2024; Zylinska, 2020).

Empirical research about evaluations of AI-generated art indicates that the anthropocentric attitudes are widely held among people. There is an important distinction in the methods of this research – in some cases participants were told about the art's origin (either truthfully, or with deception), and in other cases they were not. For example, Millet and his colleagues (2023) had people in one of their experiments listen to two music tracks generated

by AI, but participants were told that one is produced by a human, and one is AI-generated. They indicated experiencing more awe-feeling about the music they thought was human-created. The same results were replicated with paintings as well – participants rated the AI-labelled ones as less creative and awe-evoking. When an artwork – music, painting, or even a dance choreography – is labelled as AI-generated, people attribute less beauty, novelty, meaning, sensitivity, quality, and emotionality to it (Agudo et al., 2022; Darda & Cross, 2023).

Naturally, the question arises what happens when participants are not primed to believe in any specific origin. Do they rate the art generated by AI the same as human-made art? Can they distinguish these two? Firstly, it seems that people have a hard time assessing an artwork's true origin, and their accuracy often fluctuates around chance levels. In the study of Elgammal and his colleagues (2017), AI-generated and human-made visual art were rated similarly in visual structure, communication, inspiration, and intentionality. Participants did not distinguish well between the two types. An earlier study of Gangadharbatla (2002) brought about the same results. Along with the experiment of Chamberlain and her colleagues (2017), findings show another interesting theme apart from difficulty with guessing the true origin – people attribute abstract art more to AI, while representative/figurative art is more likely to be assigned to humans.

However, there is also potential evidence for a certain sensitivity of people to respond to man-made art. Demmer et al. (2023) conducted an extensive counter-balanced study with several intriguing results – the actual origin of artworks had been guessed correctly 63.8% of the time, and AI-generated images were generally rated slightly less positively than human-made ones – not merely as an effect of being labelled so. Furthermore, telling participants that the image is AI-generated led to lower likings especially if the true origin was indeed AI, but the disclosure of origin (whether truthful or false) on its own had no effect on peoples'

evaluation. Nevertheless, it is important to note that there was a big variety in these results between individual respondents, and also between different artworks.

Another possible hint of sensitivity toward human-made art is participants' accuracy in judging the artwork origin. People detect art by humans more correctly than art by AI. Chamberlain et al. (2017) found this difference to be 64.66% accuracy for human-art, and 40.31% for AI-art. In Zhou and Kawabata's study (2023) it was 68% accuracy for human-art, and 43% for AI-art. So, there is some inconsistency in findings, and the bias against AI-generated art needs to be researched more. People are not very good at detecting the origin of artworks, but they seem to prefer what is labelled – by them, or by the experimental priming – as human-made art. We go on to explore a potential factor in the relationship between artwork origin, and the preference for it.

Art and Looking Time

Being involved with art-related activities was shown to be associated with different types of well-being – greater mental, physical, and subjective health, or longer life – and Cotter et al. (2023) highlight that one of the mechanisms, through which art can improve life quality, is 'immersion' – "having attention captured by the art engagement, feeling flow during the engagement, or feeling "carried away"" (Cotter et al., 2023, p. 4). Different museums or galleries have been interested in maximizing the immersive experience of their visitors through a 'slow looking' approach, and while there seems to be beneficial outcomes to it, there is not much research about this specific movement (Igdalova & Chamberlain, 2023).

Even though immersion in art, and also the looking time (LT) spent on it, have been researched across many contexts (Smith et al., 2017), very few studies looked at the relationship between LT and art evaluation. In one of these studies (Heidenreich & Turano, 2011), four participants were wearing an eye-tracking device, and looked at artworks of different styles and periods, all human-made. Afterwards they were asked to evaluate the paintings on different

scales, such as Displeasing-Pleasing, Uninteresting-Interesting, Ugly-Beautiful. No correlation between LT and aesthetic evaluation was found. A different experiment (Smith et al., 2006) used the method of assigning people to conditions with different viewing times available – 1, 5, 30, or 60 seconds. Participants saw only human-made artworks, and were asked, again, to evaluate them on different scales. Again, no link was found between how long people looked at the works, and how they evaluated them. The previously mentioned study by Zhou and Kawabata (2023) suggests that participants looked longer at what they categorized as human-made art, compared to what they thought was AI-generated. To our knowledge, these are the only studies that looked for a link between LT at artworks, and subjective evaluations of them.

Given the small number of studies, and variety in their methodologies, it is difficult to draw any conclusions, but so far there is no evidence indicating that people look longer at the art that they like, or find more beautiful. This is surprising, as one could intuitively expect that longer LT may just mean a bigger reason for the viewer to look – for example, because they like the artwork, consider it beautiful, or at least engaging. And interestingly, the link between LT and evaluation can still be made indirectly. We have seen that people prefer art created by humans over AI-generated art, at least when it is labelled so (Agudo et al., 2022; Darda & Cross, 2023; Demmer et al., 2023; Millet et al., 2023), and they also tend to look longer at what they think is art by humans (Zhou & Kawabata, 2023).

The Current Study

The present research focused on expanding the limited literature about the link between subjective art experience, and time spent on the artworks. The study of Smith and her team (2006) set LT as an independent variable, assigning participants to conditions accordingly, therefore not allowing them limitless exposure. Heidenreich and Turano (2011) did a thorough study, but included only four participants. We aimed to expand on the knowledge about LT with our study, by attempting to improve these two factors.

We conducted an online experiment, in which participants were presented with two types of artwork images – some created by human artists, and some generated with AI. The origins of the artworks were truthfully revealed to participants through short textual descriptions that appeared right before the artworks. After seeing each image, we asked the participants for their perceived beauty, intentionality, creativity, and to select one emotion, and its intensity, that best captured their experience of the artwork. The time they spent looking at each painting was also measured, but without their knowledge.

The first research question was whether people's LTs for human-made paintings would be different to LTs for AI-generated paintings. It is hypothesized that we will replicate the results of Zhou and Kawabata (2023), that people will look longer, on average, at what is categorized as human-made artworks.

The second research question looked at two possible relationships – one between LT and perceived beautifulness, and the other one between LT and intensity of felt emotions. We aimed to put the results of Heidenreich and Turano (2011), and Smith et al. (2006), to a test. Therefore, the second hypothesis was that LT is be a weak/moderate predictor of both perceived beautifulness, and intensity of emotions, related to the artworks.

Lastly, we aimed to explore the idea of sensitivity toward human-art suggested by the results of Demmer et al. (2023), Chamberlain et al. (2017), and Zhou & Kawabata (2023). Our hypothesis was that the relationships between LT and beauty and emotional intensity have different strength within human-made artworks, compared to AI-generated ones. In other words, we tested whether artwork's origin (human or AI) moderates the relationship between LT and perceived beauty and emotional intensity.

Methods

Participants

In total, 258 people participated in the study. From these, 43 participants were excluded because of not completing the survey or giving consent. 16 participants were excluded for not passing the control question designed to check whether they are paying attention. The removal resulted in a final 199 participants included in the analysis (143 female; 47 male; two non-binary; seven preferred not to say/self-described). Most participants (n = 157; 79%) were in the age group 18-24 years old.

Participants took part either in exchange for course credits or voluntarily without reward. Convenience and snowball sampling was used – the link for this study had been sent around in online group-chats, to friends and acquaintances of the researchers, encouraging people to share the link further. It had also been posted on the SONA study system of the University of Groningen.

The research was approved and conducted in accordance with the ethical codes and regulations of the Faculty of Behavioural and Social Sciences at the University of Groningen (registration code: PSY-2425-S-0337).

Materials

The study ran on Qualtrics (https://www.qualtrics.com/) and was available in English and Dutch. Participants completed the survey on their own devices without hardware restrictions.

Artworks

From the artworks used, six were human-made, and six artificial intelligence (AI) – generated. See Appendix A for all artworks. Only abstract artworks were selected for reasons touched upon in the introduction – they are usually associated to AI more than to human artists (Chamberlain et al., 2017). For the purposes of the present study, it was more convenient to

research only one artwork type, and avoid the conflating factors resulting from including both abstract and representative artworks. Abstract was defined as no figurative elements present. Moreover, images were selected so that human-made and AI-generated artworks were pairwise similar in colour, composition or dynamic.

Artwork Descriptions

Every artwork was preceded by a short textual description that framed the work either neutrally, positively, or negatively, and mentioned the origin. The descriptions were created by the research team, specifically for this paper. There were two AI-generated positive, two AI-generated negative, two AI-generated negative, two human-made positive, two human-made negative and two human-made neutral descriptions. There was no deception on the origin of artworks. The framing was not based on actual facts about the artworks but was made-up. For a full list of all descriptions used in the framing manipulation, as well as their translations in Dutch, see Appendix B.

Intentionality, Beauty, and Creativity

Participants answered three evaluative statements about intentionality, creativity and beauty using sliders on a scale ranging from 0 to 100, based on Cox et al. (2024). For ‘beauty’ the statement “I find this work beautiful” was presented. A score of 0 represented “strongly disagree”, and a score of 100 represented “strongly agree”. For ‘intentionality’ and ‘creativity’ the statements “In my opinion, the level of intentionality involved in the creation of this work is…” and “In my opinion, the level of creativity involved in the creation of this work is…” were presented, respectively. A score of 0 indicated “very low” and a score of 100 indicated “very high”. For the translations of these statements, see Appendix C.

Geneva Emotion Wheel

Participants were then presented with the Geneva Emotion Wheel (GEW), a tool to label emotions and record emotional intensity and valence in surveys (Scherer, 2005). Participants

were asked to select the emotion-label that best captured their emotional response to the artwork, as well as indicate the intensity of that emotion, within one click. They were, for example, able to choose between emotions such as ‘Anger’, ‘Surprise’, ‘Fear’ or ‘Joy’. If no emotion was experienced, they could select the “None” option, and if their emotion was not represented in the wheel, they could select “Other”. Cronbach's α is not specified. An image of GEW from our study, along with all the emotions and their translations, can be seen in Appendix D.

Aesthetic Fluency Scale

The Aesthetic Fluency Scale was used as an approximate measure for art knowledge (Cotter et al., 2023b). Due to time constraints for the survey, the short version of the questionnaire was used. Participants' familiarity with 10 artists and art-related terms was assessed. Question 11 was a control question. The questionnaire presents a term, for example “Gouache” and gives three response options - “I don’t really know anything about this artist or term”, “I’m familiar with this artist or term”, and “I know a lot about this artist or term”. Cronbach's α was 0.84. For translations of these response options, see Appendix C.

General Attitudes toward Artificial Intelligence Scale

Participants completed an attitude scale assessing their attitudes of AI, consisting of five statements adapted from the General Attitudes toward Artificial Intelligence Scale (GAAIS, Schepman & Rodway, 2020). “Artificial Intelligence is exciting”, “I am impressed by what Artificial Intelligence can do”, “There are many beneficial applications of Artificial Intelligence”, “I am interested in using artificially intelligent systems in my daily life”, and “Artificial Intelligence can have positive impacts on people's wellbeing” were answered on sliders ranging from 0 to 100 in steps of 10, from ‘strongly disagree’ to ‘strongly agree’. Cronbach's α was 0.89. For the translations of response options, see Appendix C.

Procedure

After accessing the study environment, participants were informed about the study and its procedure, and then signed an informed consent. Next, participants were asked about age, having five options - 18-24, 25-34, 35-44, 45-54, or 55+ years old, and about gender, also having five options - male, female, non-binary, prefer to self describe, or prefer not to say.

The respondents were distributed equally among six experimental groups. Each group viewed the same 12 artworks and read the same descriptions, but the description–artwork pairings varied across groups, as shown in Table 1. The sequence of artwork-description presentations was randomised per participant.

First, participants evaluated the 12 artworks. Each artwork was preceded by one of the descriptions. There was no time limit to seeing the description. Participants had to click “next” to proceed to the next screen showing an artwork. On this screen descriptions were not visible anymore.

Each artwork appeared for a minimum of five seconds. After that the participant could choose freely when to continue, by clicking “next”. The time spent looking at each artwork, between first appearance of the artwork until clicking “next”, was measured for every participant, later used as the looking time variable.

In the next step the artwork was not visible anymore. Participants answered the three evaluative statements concerning intentionality, creativity and beauty using sliders on a scale ranging from 0 to 100 and continued. On the same site they used the GEW.

After all artworks had been evaluated, participants filled out the two questionnaires. First, the 11-item Aesthetic Fluency Scale was presented (Cotter et al., 2023b). On the next page, participants completed the General Attitudes toward Artificial Intelligence Scale (GAAIS, Schepman & Rodway, 2020).

Table 1

Artwork x Description Pairings per Group for Human-made and AI-generated artworks

Group	Positive Description		Negative Description		Neutral Description	
	D_1	D_2	D_3	D_4	D_5	D_6
1	1	2	3	4	5	6
2	6	1	2	3	4	5
3	5	6	1	2	3	4
4	4	5	6	1	2	3
5	3	4	5	6	1	2
6	2	3	4	5	6	1

Note: The numbers under description are representative of the artworks assigned in that group. For the artworks, see Appendix A. D_1 etc. are the human and AI descriptions, see Appendix B.

The study concluded with a debriefing screen that explained the manipulation of the framings, as well as the looking-time measurements, and informed all participants that they were not deceived of the true origin of each artwork. Participants were thanked for their participation and, if applicable, directed to collect their course credits through the university's SONA system.

Data analysis

The mean completion time of the study was 2122 seconds (35.4 minutes), while the median time was 726 seconds (12.1 minutes). This shows the skewed distribution of the duration times, most likely because of participants not completing the study in one go. Researchers made the decision to exclude participants, whose duration time was under 300 seconds (5 minutes). This duration had been determined by doing a trial run to see what would be the lowest time for a participation that still seems serious. We considered times under 300 seconds a non-serious engagement with the experiment for the sake of speed. This decision had

no consequences, as the lowest completion time was 302 seconds, and thus no participants were excluded for too low completion time.

As the focus of the current paper is not the effects of framing through textual descriptions, in the present data analysis we included only the trials with a neutral description accompanying the artwork. The cases where a positive or negative description appeared were excluded from the analysis.

To answer the first research question, the ‘looking time’ variable had been used, as well as the category of artworks’ origin. For the second research question, we have used variables ‘beauty’, ‘emotional intensity’, and ‘looking time’. Finally, for the third analysis all these three variables have been used, but now split into human and AI cases, so – ‘AI-beauty’, ‘human-beauty’, ‘AI-intensity’, ‘human-intensity’, ‘AI-looking time’, and ‘human-looking time’. The variables ‘intentionality’, ‘creativity’, ‘emotional valence’, and scores from the questionnaires Aesthetic Fluency Scale and GAAIS were not analysed in the current paper.

The ‘emotional intensity’ had been calculated as the Euclidean distance ($\sqrt{(x^2 + y^2)}$) from the point where the participant clicked on the GEW and the center of the GEW. In other words, it is the length of the line segment from that click point to the origin of the reference frame that passes through the center of the GEW. All distances shorter than 75 pixels are excluded, because they are in the center area where the options 'None' and 'Other' are located (for the image of GEW, see Appendix D). Those clicks did not refer to an emotions in the GEW, and are therefore not taken into account.

Results

Preliminary data considerations

The original data analysis plan was to conduct parametric tests: paired samples t-test for comparing looking time (LT) of human-made and AI-generated artworks (first hypothesis), linear regression for assessing the relationship between LT and perceived beauty, and emotion intensity (second and third hypotheses), and moderator analysis for comparing AI-generated and human-made artworks in the strength of this relationship (fourth hypothesis). However, the distributions of LT for both types of artworks were strongly skewed. Because of the insufficiency of logarithmic transformations to make the distributions at least semi-normal, non-parametric alternatives were selected for the analyses.

The median LT for human-made artworks was 7.45 seconds, while for AI-generated it was 7.24 seconds. For the beauty scores in human-made paintings, $M = 43.1$, and for the beauty scores in AI-generated, $M = 41.94$. Lastly, for the emotional intensity scores in human-made paintings, $M = 186.67$, and for the emotional intensity scores in AI-generated paintings, $M = 185.73$. See Table 2 for all the descriptives.

Main analyses

To answer the first research question, a Wilcoxon signed-rank test was conducted. This test showed a significant difference in LT between the two types of paintings, $W = 11857.5$, $z = 2.345$, $p = .019$. The median difference between them, also known as Hodges-Lehmann Estimate, turned out to be 0.47 seconds. Therefore, in terms of median, participants tended to look 0.47 seconds longer at human-made pictures than at AI-generated ones, which is what we hypothesized. The effect size, or rank-biserial correlation, was 0.192.

Regarding the second and third hypotheses, a viable alternative to regression analyses was to look at non-parametric correlation statistic Spearman's rho. It was used to correlate all possible pairs of the variables LT, perceived beauty, and intensity of felt emotions (in this case

Table 2

Descriptive statistics of variables human-made (Human), AI-generated (AI), beauty scores (Beauty), emotional intensity scores (Intensity), and looking time (LT, in seconds)

Statistic	Human Beauty	AI Beauty	Human Intensity	AI Intensity	Human LT	AI LT
Median	45.5	42	173.38	171.78	7.45	7.24
M	43.1	41.94	186.67	185.73	11.39	10.06
SD	20.69	22.16	64.5	64.95	14.37	10.06
Minimum	0	0	96.25	90.55	5.506	5.53
Maximum	90	99	407.45	392.6	154.43	82.68

not split for AI-/human- made artworks). There was a significant correlation of LT and perceived beauty, r_s (197) = .254, $p < .001$. This supports the second hypothesis but only partially – we hypothesized that LT would be a significant predictor of beauty, and in the present analysis it is only possible to establish a correlation, possibly non-linear, not prediction or explained variance. On the other hand, our third hypothesis is not supported, because the correlation of LT and emotional intensity did not turn out significant, r_s (96) = .081, p = .427. Note, that the difference in degrees of freedom between the test for beauty (197) and intensity (96) comes from the fact, that emotional intensity was not registered in the cases when participants chose 'None' or 'Other' on the Geneva Emotion Wheel, while LT and beauty had been documented in every single case.

To at least partially test the fourth hypothesis, we relied again on correlations, because the inability to perform regressions made it impossible to conduct an actual moderator analysis. As an alternative, the three variables LT, intensity, and beauty were split into human and AI cases. The resulting six variables had been correlated with one another just as in the previous test, and all outcomes are listed in Table 3. The two relevant results for this last hypothesis were

Table 3

Spearman's rho correlations between human-made (Human) and AI-generated (AI) artworks, perceived beauty scores (Beauty), intensity of felt emotions (Intensity), and looking time (LT)

Variable		Human Beauty	AI Beauty	Human Intensity	AI Intensity	Human LT
AI Beauty	*N*	199	-	-	-	-
	Spearman's rho	.359	-	-	-	-
	p-value	< .001	-	-	-	-
Human Intensity	*N*	130	130	-	-	-
	Spearman's rho	.318	.078	-	-	-
	p-value	< .001	.375	-	-	-
AI Intensity	*N*	132	132	98	-	-
	Spearman's rho	.257	.139	.622	-	-
	p-value	.003	.111	< .001	-	
Human LT	*N*	199	199	130	132	-
	Spearman's rho	.122	.116	-.025	.036	-
	p-value	.087	.102	.78	.681	-
AI LT	*N*	199	199	130	132	199
	Spearman's rho	.241	.216	-.034	.08	.419
	p-value	< .001	.002	.705	.36	< .001

that LT and beauty ratings in AI-generated artworks were significantly correlated, r_s (197) = .216, p = .002. Meanwhile, the correlation of LT and beauty ratings in human-made artworks was not significant, r_s (197) = .122, p = .087.

Exploratory analyses

Other interesting results from this analysis, although not directly related to our hypotheses, were that within all three main variables the values of human cases were significantly correlated to their AI counterparts. Beauty scores of human-made pictures were correlated to beauty scores of AI-generated pictures, r_s (197) = .359, $p < .001$, and it was the case for emotional intensity as well, r_s (96) = .622, $p < .001$, just as for LT, r_s (197) = .419, $p < .001$ (Table 3). This means that individual people were fairly consistent in how long they looked at the artworks, how beautiful they found them, and how intensely they felt about them, no matter the actual artwork origin.

Discussion

In this experiment we obtained comparable results to that of Zhou and Kawabata (2023). Our participants' median looking time (LT) was 0.47 seconds higher for human-made paintings, and even with different methodology and analyses, Zhou and Kawabata (2023) arrived at a surprisingly similar mean LT difference of 0.33 seconds. The notion that human-made artworks capture the viewers' attention for much longer is, therefore, not supported. It can still be hypothesized and analysed whether any cognitive processes behind LT, and the nature of assessment, are different for AI-generated artworks, compared to human-made ones. With such small differences in LT, we speculate that these results may not generalize in the same way to natural settings, where people casually look at artistic creations, like in a museum, and do not actively consider their origin. Rather, this difference may reflect something about the nature of our experimentation. Participants may spend a small part of their LT purely in the process of deciding whether the presented artwork counts as human-made, or AI-generated. And a higher LT within human-made artworks could reflect more factors being considered, for example because certain elements were noticed, which were not noticed within the AI-generated images. On the other hand, our effect size of 0.192 is considerable, and it may be interesting to see whether a similar effect will be found in other samples, or a more controlled settings.

To our knowledge, the current study is the first that empirically indicates a relationship between LT and aesthetic evaluation (see the results of Heidenreich & Turano, 2011, and Smith et al., 2006). We have found this link for the ratings of beauty, but not for emotional intensity. And what is more, this link persisted for the cases of AI-generated artworks, but disappeared for human-made artworks. Does this suggest that origin of a painting is a moderating variable for the LT-beauty relationship? The change in the strengths of relationship would be expected after introducing an actual moderator, so from a strictly technical point of view our results do indicate that. However, that the link was only present for the AI-generated artworks, but not for

the human-made ones was not expected. If some participants held any anthropocentric views, or negative attitudes towards AI that influenced their ratings, we would expect the link to be more complicated, and less robust, within the AI cases. Meanwhile, for the cases with human-made paintings, we could expect this relationship to be stronger and clearer – if people are used to view mainly human-made art, then the relationship between their LT and how beautiful they consider the art could be more straight-forward compared to a case when they are told that the origin of that art could be AI, which may initiate additional evaluative processes. And if we consider that previous studies failed to find any relationship at all, it could be argued that the present results reflect something that made our participants think, notice, feel, or evaluate differently within this specific experiment, rather than some kind of anthropocentric bias they held independently.

Limitations and Future Research

The current study, its motivation, and reasonings should serve further exploration of subjective artistic experience, and whether LT has a role in it. Future research could explore the relationships with other constructs, such as interest, engagement, meaning, reflection, and others. LT should be measured very precisely, and ideally in controlled settings, so that it reflects an actual engagement. Related to the previous point about a possibly more complex relationship within the AI-generated artworks, our study highlights how important it is to construct a rigorous methodology, with which the LT measurements are as clear as possible, and eliminate as many other factors as it can. Data that is not skewed, does not have strong outliers, and can reflect participants' immersion with higher sensitivity could reveal new findings. Different situations, for example participants assessing only about three to five artworks, may also result in different emerging themes – the nature of our experiment made it so that participants possibly decided not to spend too much on any of the artworks, since they knew that they would see 12 of them.

Another limiting side of the current study is the uncontrolled settings – the LT data very likely include other factors than just engagement with an artwork, for example personal circumstances and needs, factors of the environment, distractions, boredom, or fatigue. Further research needs to account for these, and look at LT in different contexts. The results from our study should be considered with reserve also due to the use of non-parametric tests that cannot produce as strong claims as parametric tests.

Conclusions

This was an exploratory study that built on the literature about bias against AI-generated art. From previous studies we could see that people mostly have a bias against the label 'AI-generated' art, and attribute lower quality to what they, or researchers, labelled so. At the same time, when art is not labelled, and people are to guess the actual origin, their accuracy usually stays around chance levels. They seem, however, to be able to recognize human-made artworks better than to detect AI-generated ones.

We wanted to explore the factors coming into play when viewing an artwork, and then consequently evaluating it. Our main focus was the LT. There are very few studies concerned with the relationship between LT and art evaluation, so we decided to expand the existing research – do people look longer at AI-generated or human-made artworks? Can looking time predict the beauty perceived while seeing an artwork? This was tested through an online experiment, in which people were presented with both human-made and AI-generated artworks, and asked to evaluate them on different scales.

Our results showed that participants looked longer at human-made artworks. We also found evidence for a link between LT and perceived beauty, something that past studies failed to find. The same link did not appear between LT and emotional intensity related to the artworks. The origin of an artwork (AI/human) could be something that changes the nature of

the link between LT and beauty, as our results indicate, however further research in more controlled settings is called for, to understand this relationship better.

References

Agudo, U., Arrese, M., Liberal, K. G., & Matute, H. (2022). Assessing Emotion and Sensitivity of AI Artwork. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.879088

Chamberlain, R., Wagemans, J., & Mullin, C. R. (2017). Putting the Art in Artificial: Aesthetic Responses to Computer-Generated Art. *Psychology of Aesthetics, Creativity, and the Arts.* https://doi.org/10.1037/aca0000136

Cotter, K. N., Harrouche, M., Rodriguez-Boerwinkle, R. M., Boerwinkle, M., Silvia, P. J., & Pawelski, J. O. (2023). Virtual art visits: Examining the effects of slow looking on well-being in an online environment. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000548

Cotter, K. N., Rodriguez-Boerwinkle, R. M., Christensen, A. P., Fekete, A., Smith, J. K., Smith, L. F., Tinio, P. P. L., & Silvia, P. J. (2023b). Updating the Aesthetic Fluency Scale: Revised long and short forms for research in the psychology of the arts. *PLoS ONE*, *18*(2), e0281547. https://doi.org/10.1371/journal.pone.0281547

Cox, R. F. A., Schino, G., van Klaveren, L., & Wittingslow, R. (2024, October 23). Exploring the Skill-Tech Space: Art-Making Awareness Influences Experiencing Visual Artworks. https://doi.org/10.31219/osf.io/jxnwc

Darda, K. M., & Cross, E. S. (2023). The computer, a choreographer? Aesthetic responses to randomly-generated dance choreography by a computer. *Heliyon, 9.* https://doi.org/10.1016/j.heliyon.2022.e12750

Demmer, T. R., Kühnapfel, C., Fingerhut, J., & Pelowski, M. (2023). Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI- versus human-created art and impact on aesthetic experience. *Computers in Human Behavior, 148.* https://doi.org/10.1016/j.chb.2023.107875

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. http://refhub.elsevier.com/S0747-5632(23)00226-1/sref17

Gangadharbatla, H. (2022). The role of AI attribution knowledge in the evaluation of artwork. *Empirical Studies of the Arts, 40*(2), 125–142. https://doi.org/10.1177/0276237421994697

Heidenreich, S. M., & Turano, K. A. (2011). Where does one look when viewing artwork in a museum? *Empirical studies of the arts, 29*(1), 51-72. https://doi.org/10.2190/EM.29.1.d

Igdalova, A., & Chamberlain, R. (2023). Slow looking at still art: The effect of manipulating audio context and image category on mood and engagement during an online slow looking exercise. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000546

Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior, 143*. https://doi.org/10.1016/j.chb.2023.107707

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, *1*, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Smith, L. F., Bousquet, S. G., Chang, G., & Smith, J. K. (2006). Effects of time and information on perception of art. *Empirical studies of the arts, 24*(2), 229-242. https://doi.org/10.2190/DJM0-QBDW-03V7-BLRM

Smith, L. F., Smith, J. K., & Tinio, P. P. L. (2017). Time Spent Viewing Art and Reading Labels. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 77-85. http://dx.doi.org/10.1037/aca0000049

Zhou, Y., & Kawabata, H. (2023). Eyes can tell: Assessment of implicit attitudes toward AI art. *i-Perception, 14*(5), 1-14. https://doi.org/10.1177/20416695231209846

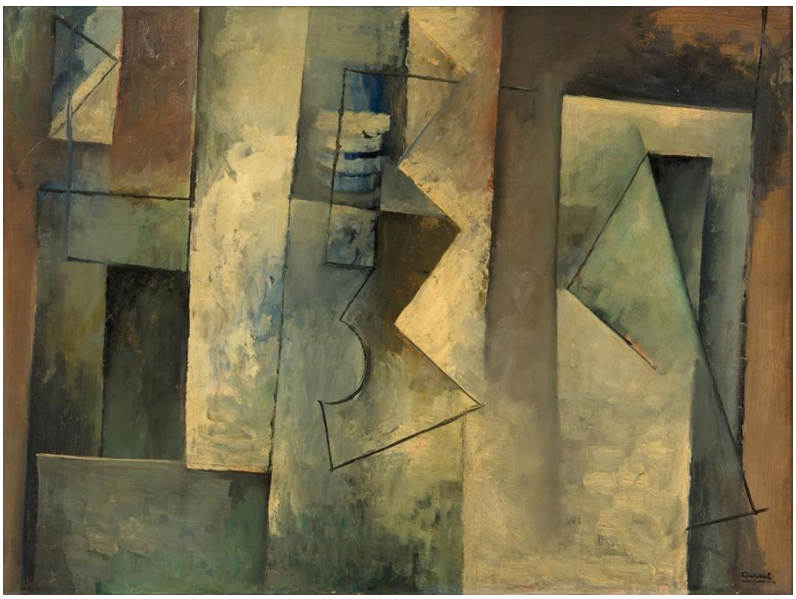
Zylinska, J. (2020). *AI art: Machine visions and warped dreams*. Open Humanities Press. http://www.openhumanitiespress.org/books/titles/ai-art/

Appendix A

Human-made pictures

Artwork 1

Sans titre (Composition brune II) by Youla Chapoval - Artvee. (o. D.). Artvee. https://artvee.com/dl/sans-titre-composition-brune-ii/#00



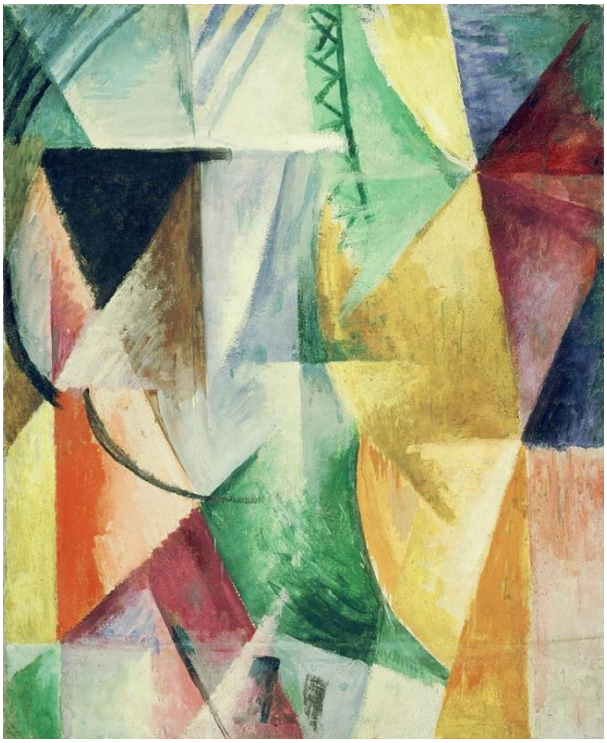
Artwork 2

abstract paintings - Abstract paintings Alessandro Tognin. (2023, 3. September). Abstract Paintings Alessandro Tognin. https://www.dreamsart.it/product/abstract-paintings/

Artwork 3

Une fenêtre by Robert Delaunay - Artvee. (o. D.). Artvee.

https://artvee.com/dl/une-fenetre/#00



Artwork 4

Figürliche Komposition by Adolf Hölzel - Artvee. (o. D.). Artvee.

https://artvee.com/dl/figurliche-komposition/#00

Artwork 5

Anitra by Anonymous - Artvee. (o. D.). Artvee.

https://artvee.com/dl/anitra/#00

Artwork 6

Komposition by Otto Freundlich - Artvee. (o. D.). Artvee.

https://artvee.com/dl/komposition-14/#0

AI-generated pictures

Artwork 1

Exploring Abstract Art with AI. (2024, 24. July).

https://deepdreamgenerator.com/blog/abstract-art-and-ai

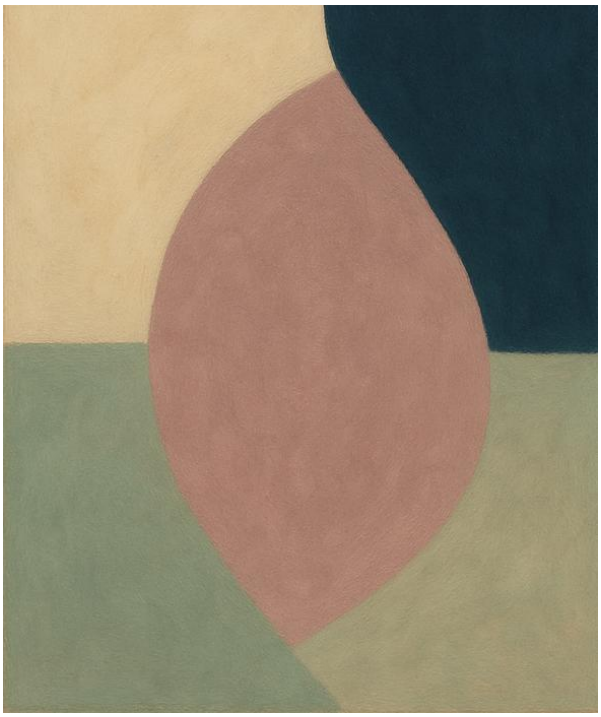
Artwork 2

Abstrakte Erdelemente auf Leinwand. (2025, 25. April). ChatGPT.

https://chatgpt.com/s/m_680b937d8b548191960f2c69fc085d2b

Artwork 3

Geometrische Abstraktion in Pastellfarben. (2025, 25. April). ChatGPT. https://chatgpt.com/s/m_680b9334cf0c819189fd5f1b73c92c39

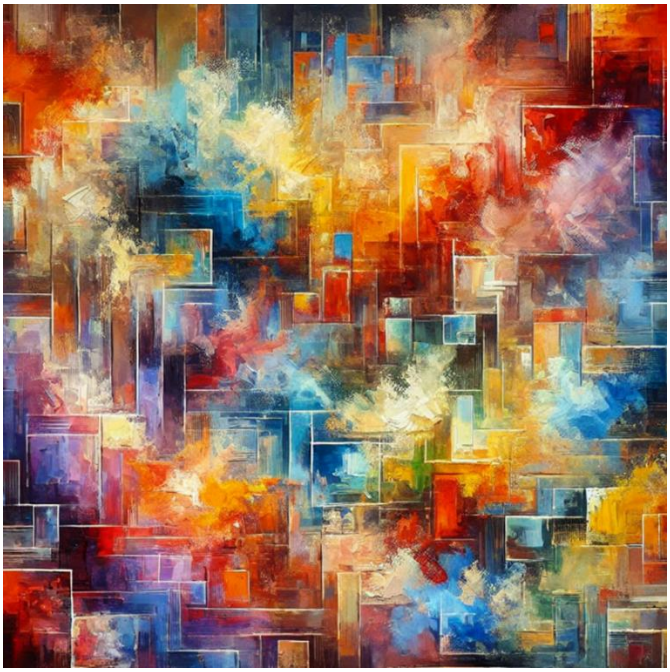


Artwork 4

Exploring Abstract Art with AI. (2024, 24. July). https://deepdreamgenerator.com/blog/abstract-art-and-ai

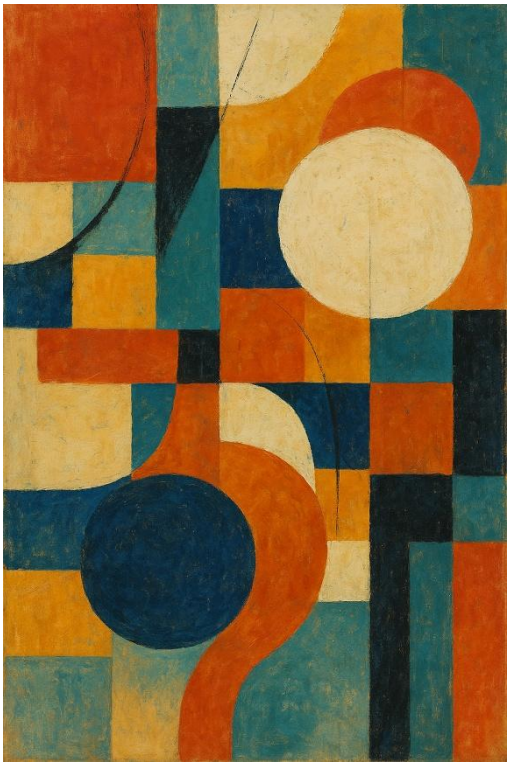
Artwork 5

AI-generated abstract painting inspired by the sensation of drinking three espressos,. (2025, 7. April). DALL·E via ChatGPT. https://chatgpt.com/



Artwork 6

Abstrakte geometrische Komposition mit Farben. (2025, 25. April). ChatGPT. https://chatgpt.com/s/m_680b926049d08191a564bb90dbfbf720



Appendix B

Artwork Descriptions

Positive Human Descriptions

HUM Description 1. "This artwork, created by a skilled artist, demonstrates mastery of technique and showcases years of experience and dedication to the craft."

"Dit kunstwerk, gemaakt door een getalenteerde kunstenaar, getuigt van zijn meesterschap in de techniek en van jarenlange ervaring en toewijding aan het vak."

HUM Description 2. "Each detail in this piece reflects the artist's refined expertise and distinctive approach, making it a truly unique expression of artistic vision."

"Elk detail in dit stuk weerspiegelt de verfijnde expertise en de unieke aanpak van de kunstenaar, wat het tot een werkelijk unieke uiting van artistieke visie maakt."

Negative Human Descriptions

HUM Description 3. "Despite being human-made, this artwork reveals the limitations of subjective interpretation, showing how even skilled artists can struggle with expression of their artistic vision."

"Hoewel dit kunstwerk door mensen is gemaakt, toont het de beperkingen van subjectieve interpretatie en laat het zien hoe zelfs getalenteerde kunstenaars moeite kunnen hebben met het uiten van hun artistieke visie."

HUM Description 4. "While created by hand, this piece reflects human biases and imperfections, highlighting how artistic vision is often constrained by personal and cultural influences."

"Hoewel dit kunstwerk met de hand is gemaakt, zijn er toch menselijke vooroordelen en onvolkomenheden in het werk te zien. Het laat zien hoe de artistieke visie vaak wordt beperkt door persoonlijke en culturele invloeden."

Neutral Human Descriptions

HUM Description 5 and 6. “This artwork is made by a human artist.”

”Dit kunstwerk is gemaakt door een menselijke kunstenaar.”

Positive AI Descriptions

AI Description 1. "This AI-generated piece demonstrates how technology is capable of artistic expression, creating intricate and thought-provoking visuals with precision and uniqueness."

"Dit door AI gegenereerde kunstwerk laat zien hoe technologie artistieke expressie mogelijk maakt door complexe en tot nadenken stemmende beelden te creëren met precisie en uniciteit."

AI Description 2. “Generated by advanced AI, this artwork pushes the boundaries, blending complex patterns and ideas beyond human imagination.”

"Dit kunstwerk is gemaakt met behulp van geavanceerde kunstmatige intelligentie (AI) en verlegt de grenzen door complexe patronen en ideeën te combineren die de menselijke verbeelding te boven gaan."

Negative AI Descriptions

AI Description 3. "This artwork, generated by AI, demonstrates that even the most advanced technology fails to inspire, revealing the mechanical nature of algorithms."

"Dit door AI gegenereerde kunstwerk laat zien dat zelfs de meest geavanceerde technologie niet kan inspireren en onthult de mechanische aard van algoritmes."

AI Description 4. "Despite being produced by advanced technology, this AI-generated piece highlights the absence of genuine human inspiration and artistic intent."

"Hoewel dit kunstwerk met behulp van geavanceerde technologie is gemaakt, benadrukt het de afwezigheid van echte menselijke inspiratie en artistieke intentie."

Neutral AI Description

AI Description 5 and 6. "This artwork is generated by AI."

"Dit kunstwerk is gegenereerd door AI."

Appendix C

Translations into Dutch

Beauty

"I find this work beautiful" - "Ik vind dit werk mooi"

A score of 0 = "strongly disagree" - "erg mee oneens"

A score of 100 = "strongly agree" - "erg mee eens"

Intentionality, Creativity

"In my opinion, the level of intentionality involved in the creation of this work is…" - "Naar mijn mening is het niveau van intentionaliteit dat betrokken is bij het maken van dit werk…"

"In my opinion, the level of creativity involved in the creation of this work is…" - "Naar mijn mening is het niveau van creativiteit dat betrokken is bij het maken van dit werk…"

A score of 0 = "very low" - "heel laag"

A score of 100 = "very high" - "heel hoog"

Aesthetic Fluency Scale

"I don't really know anything about this artist or term" - "Ik weet eigenlijk niets over deze kunstenaar of term"

"I'm familiar with this artist or term" - "Ik ben bekend met deze kunstenaar of term"

"I know a lot about this artist or term" - "Ik weet een hoop over deze kunstenaar of term"

General Attitudes toward Artificial Intelligence Scale

"Artificial Intelligence is exciting" - "Kunstmatige Intelligentie is uitdagend"

“I am impressed by what Artificial Intelligence can do” - “Ik ben onder de indruk van wat Kunstmatige Intelligentie kan doen”

“There are many beneficial applications of Artificial Intelligence” - “Er zijn veel nuttige toepassingen van Kunstmatige Intelligentie”

“I am interested in using artificially intelligent systems in my daily life” - “In mijn dagelijks leven ben ik geïntereseerd in het gebruik van Kunstmatige Intelligente systemen“

“Artificial Intelligence can have positive impacts on people's wellbeing” - “Kunstmatige Intelligentie kan een positieve impact hebben op het welzijn van mensen”

Appendix D

Geneva Emotion Wheel