

Measuring Value in AI Generated Visual Art

A structured methodological review of valuation measures

Liyuantan^{1*}, lichnagtan²

¹ Independent Researcher, Kunming, Yunnan, China

² Student, Guangzhou Medical University, China

* Correspondence: tanliyuan87@gmail.com

Author information: Liyuantan is the founder of Whispart, <https://whispart.com>

Review article • Preprint • 20 October 2026

Abstract

Studies of AI-generated visual art use value to describe aesthetic ratings, price judgments and acquisition decisions. A synthesis based on these labels alone can combine responses with different meanings. This structured methodological review examines the instruments behind those labels in 12 selected empirical reports from 2018–2026. Targeted searches, reference checking and examination of public methods, supplements and a codebook yielded 25 study units and 46 outcome-family records. Classification distinguished question content, response format and financial consequences, with separate records for source attribution and the object evaluated. Seven study units included economic wording answered without currency; five elicited sale-price judgments, five personal monetary willingness to pay, four purchase intentions or relative buying preferences, and two consequential auction bids. Categories overlap, including within mixed composites. Close reading identified scaled payment questions, a selling-price prediction described as willingness to pay, and an aesthetic-value composite containing economic items. Both auction samples came from one report. The contribution is a classification linked to source passages and a set of reporting fields that preserve these distinctions. For synthesis, a change in rated worth, a recommended or predicted selling price and a consequential bid should be interpreted as different outcomes. For new experiments, the question, response unit and acquisition rule need to be specified together.

Keywords: AI-generated art; empirical aesthetics; valuation; willingness to pay; purchase intention; measurement

1 Introduction

A viewer can admire an image and decline to buy it. A judgment of beauty concerns the viewing experience; a purchase decision also involves an object, a price and competing uses of money. Predicting what the image would sell for introduces a different perspective again. Research on AI-generated visual art uses all of these tasks. Reading their results under the common heading of value can obscure what participants actually evaluated.

The problem is visible within individual reports. Horton et al. (2023) used rating scales in early experiments and monetary questions in later ones. In Magni et al. (2024), a supplementary outcome called willingness to pay asked for an estimated selling price. Chamberlain et al. (2018) combined aesthetic and economic wording in a measure of aesthetic value. A review therefore needs to inspect the question and response procedure, rather than assign an interpretation from the outcome name.

This approach connects two established methodological concerns. Flake and Fried (2020) describe how inadequate measurement transparency obstructs evaluation of what an instrument measures. Research on willingness to pay also distinguishes hypothetical statements from decisions with payment obligations; Schmidt and Bijmolt (2020) show that the discrepancy depends on elicitation method and study context. The present review applies these concerns to visual art involving actual or attributed AI production. It asks how the instruments used in this literature support claims about appraisal, monetary value and acquisition.

Existing reviews establish the wider context. Oksanen et al. (2023) surveyed empirical research across fine-art domains. Bianchi et al. (2025) examined creativity, aesthetic evaluation and the importance of production context. The narrower contribution here is to connect study-level questions and response formats to their economic interpretation, including cases where one report uses several instruments or a composite crosses categories.

Source attribution is a further part of that interpretation. Assigning human and AI labels to the same image estimates a response to information about authorship. Comparing different sets of human and AI images also incorporates the generation and selection of those sets. The review records this distinction alongside the valuation measure. Its questions are which outcomes are elicited, which decisions carry financial consequences, and how the comparison and the offered object shape the meaning of a result.

2 Methods

2.1 Scope and selection

The review uses purposive selection to examine variation in measurement across empirical reports. It covers aesthetic responses, economic rating scales, monetary judgments, buying intentions and bids. The design is a structured methodological review, without a preregistered protocol or exhaustive database search. Classification rules were developed through preliminary reading and refined during extraction.

Eligible reports presented original human-participant research evaluating visual artworks, art images or art reproductions involving actual or attributed machine production. Public full text, publisher HTML or supplementary documentation had to support classification of the relevant task. Human-made paintings labeled as AI were eligible because they test attributed authorship. Earlier computer-generated and robotic art provided historical comparisons; their production methods were recorded separately from contemporary generative-image systems.

Studies of music, writing, startup ideas and advertising design were excluded, including in reports that also contained eligible visual-art studies. Pilots and stimulus pretests were excluded from study counts. Technical image-quality benchmarks without a human valuation task were outside scope. Selection concerned the measurement and domain, irrespective of whether a reported result favored human or AI art.

2.2 Search and source verification

Search and access work took place on 20 October 2026. Targeted web searches combined terms for AI-generated art and source attribution with valuation, willingness to pay, purchase intention and auctions. Exact-title searches and reference checking located publisher versions and public institutional copies. Supplementary Section S3 records retained queries and access decisions. Because the search was iterative, its output is a documented corpus rather than a fixed database export with a screening denominator.

Detailed classification used methods sections, relevant tables and supplements. An institutional author manuscript supplied Chamberlain et al. (2018). The publisher's supplementary Appendices 3 and 4 supplied the robot-drawing questionnaire and factor loadings. The September 2026 Creative PEC version supplied Lane et al. (2026); its earlier SSRN record was treated as another version of the same report. The public OSF codebook linked by Neef et al. (2025) established the response categories for attractiveness and buying intention. Bibliographic details were checked against publication records and available Crossref metadata. Issue years are used for Neef et al. and Magni et al., with earlier online years retained in the evidence record.

Follow-up retrieval targeted two transaction reports and unresolved measurement details in the retained corpus. The query and access log records the routes checked and the resulting decisions. This focused extension did not constitute a new exhaustive literature search. Reports whose complete relevant methods remained unavailable were retained in an access ledger. Abstracts were used to identify those gaps and describe their relevance, not to assign a detailed payment mechanism. This affected recent work by Depecik et al. (2026) and Prissé et al. (2026), among others.

2.3 Extraction units and classification rules

Extraction distinguishes reports, study units and outcome-family records. A study unit is an eligible main experiment or study within a report. An outcome-family record identifies a kind of evaluative response in that unit. Multiple liking or beauty items can share one record; sale-price judgments and personal payment amounts receive separate records. When a composite combines different content, each relevant family is recorded and the aggregation is noted. These records describe measurement content, including embedded items, and are not a count of independent dependent variables.

The evidence files record reported sample size, recruitment setting, actual production source, attributed source, object evaluated, response format, financial consequences and a source location. Design notes retain features that affect interpretation, such as artist reputation bundled with source information or a comparison without a no-purchase option. Conflicting sample sizes and incomplete endpoint information are preserved in specific fields. No total participant count or overall study-quality score was calculated.

Classification starts with the response. A question about worth, expensiveness or personal payment answered on a rating scale is coded as a noncurrency economic rating. A currency answer is then classified by perspective: predicting or recommending a seller's price is a sale-price judgment, with predictive and normative wording retained separately. Naming the respondent's own hypothetical

payment is personal stated willingness to pay (WTP). A likelihood or comparative preference for acquisition is a purchase-intention record. A bid is classified as consequential when the source documents a payment and allocation mechanism. Survey participation payments alone do not meet that criterion. Figure 1 presents the coding sequence. If the currency unit, aggregation or settlement rule cannot be established, the missing attribute remains unverified; an outcome name alone does not resolve it. A consequential purchase outside an auction would require a separate category.

The six families in Table 1 are practical combinations of content and procedure, not mutually exclusive psychological constructs. Aesthetic or artistic evaluation groups several forms of appraisal for mapping purposes. Response format and financial consequence remain separate fields so that the classification can accommodate a mixed composite or a future instrument with a different settlement rule. Missing detail is recorded as unverified rather than filled from a nearby scale or inferred from an outcome's name.

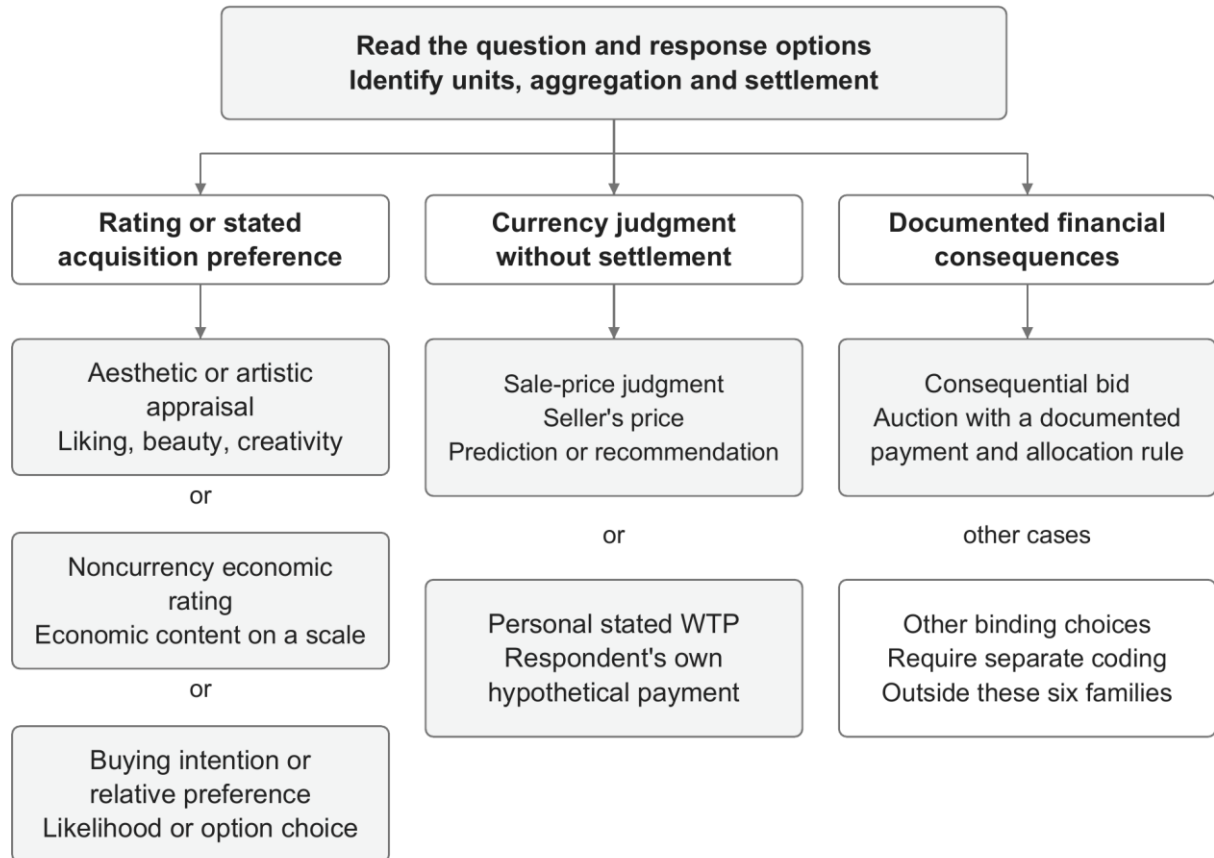
2.4 Synthesis and checks

The synthesis compares instruments and their interpretation. Counts are reported at study and report levels to show concentration within research reports. An exploratory coverage check removes each report in turn and recalculates category counts. A further check removes both the historical computer/robot-art report and the small five-pair valuation study. These checks describe the dependence of the measurement map on inclusion decisions; they do not estimate the stability of an effect size. No pooled effect or tally of statistically significant findings was calculated.

OpenAI Codex assisted literature discovery, retrieval, extraction, classification, writing and document preparation. Source passages were revisited within the same workflow, and scripts checked identifiers and counts. A targeted claim audit records the sources, locations, decisions and residual uncertainty for the central classification examples and auction results (Supplementary Section S8). The workflow included substantive AI assistance and no independent duplicate human extraction. Both authors have read the manuscript and agreed to its submission. An independent human audit of the source-level extraction was not performed.

Table 1 Measurement families and corpus coverage

Family	Classification rule	Studies	Reports
Aesthetic or artistic evaluation	Appraisal content such as liking, beauty or creativity	23	11
Noncurrency economic rating	Economic wording answered without a currency unit	7	4
Sale-price judgment	Currency prediction or recommendation of a selling price	5	2
Personal stated WTP	Respondent's own hypothetical monetary payment	5	2
Purchase intention or relative preference	Likelihood or relative preference for acquisition	4	3
Consequential bid	Bid with a documented payment and allocation mechanism	2	1

Figure 1 Classifying the recorded response

Keep each item's content when a composite crosses families.
Record missing units or settlement details as unverified.

Worked examples from the reviewed reports

Horton et al. 2023 Experiment 2

Payment wording answered on a 1-7 scale: noncurrency economic rating.

Magni et al. 2024 Study 4

Forecast of a selling price in USD: sale-price judgment, predictive subtype.

Lane et al. 2026 Visual-art auctions

Bid with a random binding round and second-price payment: consequential bid.

Apply the sequence to the documented task, retaining each item in a mixed composite. The three worked examples show how the response format, perspective and settlement rule change classification. These six families describe the selected corpus; other consequential choices require separate coding.

3 Results

3.1 Measurement coverage and report concentration

The corpus contains 12 reports, 25 study units and 46 outcome-family records. Twenty-three study units include aesthetic or artistic evaluation. Seven contain noncurrency economic ratings, five sale-price judgments, five personal stated monetary WTP, four buying intentions or relative preferences, and two consequential bids. Table 2 identifies the reports and eligible study units. The categories overlap.

Counting reports changes the apparent spread of the economic evidence. The seven noncurrency units come from four reports. Sale-price judgment and personal stated WTP each appear in two reports, while buying intention or relative preference appears in three. Both auction units belong to Lane et al. (2026). Thus, several experimental samples sometimes represent variations within one research program.

The coverage checks make this concentration explicit. Removing Horton et al. leaves one sale-price study and one personal monetary-WTP study in the coded corpus. Removing Lane et al. leaves no consequential auction unit. Removing Chamberlain et al. and Hall and Schofield together retains all six families across ten reports and 22 study units. The category structure is therefore broader than either historical robotics or the smallest valuation sample, while the acquisition evidence remains concentrated. Full recalculations are provided in Supplementary Table S3.

Table 2 Reports and eligible study units

Report	Study units and reported sample sizes	Coded measures
Bellaiche et al. (2023)	S1: 149; S2: 148	Appraisal; economic rating
Horton et al. (2023)	E1–E6: 119, 415, 405, 789, 709/710, 527	Appraisal; economic rating; sale-price judgment; personal monetary WTP
Gu and Li (2022)	S1: 106; S2: 299 (143 experts and 156 nonexperts)	Appraisal; buying intention/preference
Neef et al. (2025; online 2024)	Single study: 952	Appraisal; buying intention/preference
Millet et al. (2023)	S2: 298; S3: 404; S4: 800 after 17 exclusions	Appraisal; buying intention/preference
Magni et al. (2024; online 2023)	S1: 131; S4: 808	Appraisal; sale-price judgment
Chamberlain et al. (2018)	S1: 65; S2: 349 across groups reported as 145, 97 and 107	Appraisal; economic rating
Lane et al. (2026)	Lab: 224; Online: 525	consequential bid
Theirs et al. (2026)	S1: 207; S2: 209 reported	Appraisal; economic rating
Chiarella et al. (2022)	Single study: 110	Appraisal
Hall and Schofield (2025)	Single study: 27	Appraisal; personal monetary WTP
Cunningham et al. (2025)	Single study: 92	Appraisal

3.2 Economic language appears in rating scales and mixed composites

Bellaiche et al. (2023) measured monetary worth with a 1–5 response scale. Horton et al. (2023) used 1–7 ratings of expensiveness in Experiment 1 and included scaled WTP in Experiment 2. Theirs et al. (2026) used a 0–4 response for monetary value alongside aesthetic judgments. These instruments measure reported economic evaluations in ordered categories. They supply neither a currency amount nor a quantity purchased.

Chamberlain et al.’s (2018) second study exposes a related issue at the composite level. The methods describe rating questions about robot drawings; Table 4 includes items concerning money and willingness to pay. Eleven items were averaged into the variable called perceived aesthetic value. The present extraction records the economic content as well as the aesthetic content, with both linked to the same composite. The publisher’s Appendix 3 shows five agreement categories from strong disagreement to strong agreement for the drawing statements. Its separate semantic-differential questions concern the robots. Appendix 4 confirms that both economic items enter the composite and that the hand-crafted item is omitted. A composite bearing an aesthetic label can therefore contain acquisition-related judgments, just as an economically worded item can remain a noncurrency rating.

These examples also involve different interventions. Theirs et al. varied reported or experienced generation time, while Bellaiche et al. varied attributed authorship. A shared rating format makes the response units recognizable, but the experimental question still differs. Synthesis needs both pieces of information.

3.3 Currency questions separate sale-price judgments from personal offers

In Horton et al.’s (2023) Experiments 3–6, the gallery scenario introduced prices of \$50–220, with most works said to sell for \$150. Participants judged what the gallery should charge and what they would personally pay under the stated scenario. The seller question asks for a recommended price, whereas the buyer question asks for a personal hypothetical payment. The stated gallery range supplies context, not a verified response bound.

Magni et al.’s (2024) supplementary Appendix 5 labels an outcome WTP while asking participants in Study 4 to estimate a painting’s selling price between \$100 and \$10,000. This is a prediction of a sale price. The classification groups it with Horton et al.’s seller-price recommendation for descriptive coverage, while preserving the difference between what a seller should charge and what a painting would sell for. Neither question elicits the respondent’s own purchase commitment.

Hall and Schofield (2025) asked 27 participants for hypothetical payment amounts for five human–AI image pairs without disclosing the production source. Human images had higher mean amounts in three pairs, one pair had equal means, and one had a higher AI mean. The task supplies personal monetary judgments for those images. Because source and visual content vary together, the comparison does not isolate the effect of an AI label.

Price ranges, reference amounts and ownership descriptions form part of these instruments. Their presence should travel with the result when it is summarized. The reviewed comparisons do not identify the causal effect of an anchor unless the pricing information itself varies in the relevant contrast.

3.4 Buying intentions depend on the choice offered

Gu and Li (2022) used a 0–100 liking slider and separate 1–7 purchase and collection measures. Their distinction preserves the difference between appreciating a picture and expressing an inclination to acquire it. Neef et al. (2025) likewise used separate attractiveness and buying items. The linked codebook confirms seven agreement categories for both, with additional codes for inability to judge and nonresponse. Those special codes are not endpoints of the agreement scale.

Neef et al. varied whether source categories appeared separately or together. In the separate-source comparison, the buying-intention difference was small and nonsignificant ($d = 0.08$, $p = .363$); an equivalence test was not reported for that contrast. This result concerns the specified presentation context and single-item response.

Millet et al.'s (2023) Study 4 used a seven-point preference between two posters for an office, attributed to a drawing robot and a human painter. Participants were not offered a price or a choice of neither poster. The measure captures relative buying preference within that pair. An absolute purchase probability would require information about the option to abstain and the terms of acquisition.

3.5 Consequential bids concern a defined exchange

Lane et al. (2026) used second-price auctions for physical prints. Groups of four submitted bids over ten rounds, with one round selected for settlement within each group. The laboratory instructions explicitly permit £0–12 bids in £0.01 increments. The winner paid the second-highest bid from a £12 endowment and received the print. The visual-art samples comprised 224 laboratory participants and 525 online participants.

The source-disclosed laboratory means were £2.14 for human images and £2.22 for AI images ($p = .841$). Online means were £5.68 and £5.29 respectively ($p = .191$). In the online analysis, adjustment for baseline appeal and other controls yielded a statistically significant AI coefficient of –£0.82 in the fully adjusted Tobit model ($p < .05$; 523 respondents in that specification). That coefficient refers to the model's latent response, rather than the observed mean difference.

The exchange rule gives the bids a financial consequence. The paid price is nevertheless determined by another bidder, and the inference concerns the selected prints, experimental budget and recruited samples. Presenting raw and adjusted results together shows how the conclusion depends on the comparison and analysis.

The coded auctions are not the entire available literature on transactions. Depecik et al. (2026) report purchase behavior in a pop-up art market in their abstract. Prissé et al. (2026) describe an auction and an AI advantage for landscape drawings. Their complete relevant methods remained unavailable for coding. These reports establish a coverage gap, including a result in the opposite direction, and warrant full extraction in an expanded review.

3.6 Source attribution and the acquisition object

Bellaiche et al. (2023) assigned human and AI labels to AI-generated images; Gu and Li (2022) assigned these labels to human-made paintings. In both cases, attribution varies while image content can be held constant. Additional source information matters: the human description in Gu and Li includes famous-artist status, so the information contrast extends beyond a bare human label.

Chamberlain et al.'s (2018) first study instead asked viewers to classify the image source themselves. Classification can depend on image appearance, making its association with liking different from a randomized label effect. Their second study compared robot drawings across settings with different

opportunities to encounter the robots. Venue, participant composition and process visibility varied together.

Physical presentation also needs to be separated from acquisition. Chiarella et al. (2022) recorded liking and physiological responses to two actual paintings at an art fair. Cunningham et al. (2025) recorded eye movements during viewing and a subsequent collection-level impressions questionnaire. These outcomes concern encounters with art. Whether a work is physically present, whether it is attributed to AI, and whether a decision triggers payment are separate design features.

4 Discussion

4.1 The contribution is an instrument level distinction

The reviewed instruments show three ways that outcome names can obscure the response elicited. Horton et al. move from rating scales to monetary questions across experiments. Magni et al. label a selling-price prediction as WTP. Chamberlain et al. combine aesthetic and economic items within one score. The extraction records connect these cases to the documented questions, response formats and aggregation rules, making the classification open to checking and reuse.

The resulting six families help organize evidence by the question each measure answers. Beauty ratings can address reported beauty. Sale-price judgments may concern a market prediction or a recommendation about what a seller should charge; those perspectives remain distinct within the descriptive category. A consequential bid provides a different kind of evidence because an allocation rule connects it to an exchange. Measures should be matched to the proposed inference. The present classification does not establish their psychometric validity or imply that transactions exhaust what matters about art.

This distinction is consistent with Flake and Fried's (2020) emphasis on explicit measurement decisions. It also places the AI-art literature within research on WTP elicitation. Schmidt and Bijmolt (2020) document variation in hypothetical bias across consumer-goods studies. That literature motivates attention to payment obligations here; its pooled estimates should not be imported as a numerical correction for these art studies.

4.2 Interpreting an AI penalty requires a specified comparison

A reported AI penalty should identify the outcome and comparison: for example, a lower worth rating for the same image under a different label, or a lower bid for one selected collection of prints than for another. Those results can differ without being contradictory. They concern different interventions, objects and decisions.

The distinction also affects explanations based on prejudice. A label effect demonstrates sensitivity to authorship information. Whether that sensitivity constitutes an error depends on what the observer values and on the information conveyed. Human participation, authenticity and production history can be attributes of an acquisition. A useful test of a proposed mechanism therefore specifies the belief or preference being changed and measures the relevant consequence.

The coverage analysis adds a separate caution about generality. Four of the five personal monetary-WTP units belong to Horton et al.; both auction units belong to Lane et al. More experiments within these reports provide useful variation, but the number of units alone overstates the breadth of independent research programs. Future evidence should expand the objects, settings and study populations represented in those categories.

4.3 Experiments that connect appraisal to acquisition

A productive design would measure appraisal and an acquisition decision for the same described object. For a print, that description should include dimensions, material, framing and delivery. Participants could then rate the image and decide whether to obtain that print at a specified cost. Retaining the option to decline separates preference among images from willingness to make a purchase.

The design should also make its source comparison explicit. Randomized labels on fixed images target attribution. Comparisons between actual human and AI production target selected outputs and can include an undisclosed-source condition. Generation, selection and matching procedures should be retained with the stimuli. Where participants rate several works, analysis should account for repeated responses and image sampling.

For an auction, reports should distinguish the submitted bid, the winning price and the completed acquisition. The settlement rule, number of binding decisions and source of the purchasing budget determine the economic task. These details allow a reader to assess how closely the response corresponds to the decision that a prospective buyer would face.

4.4 Implications for evidence synthesis and reporting

Future synthesis can use the evidence fields to group comparable questions before extracting effect sizes. Studies of label effects on rating scales form one possible group; studies of monetary offers under a common acquisition rule form another. Agreement on response units alone is insufficient when the perspective or intervention differs.

Table 3 specifies information that would have resolved the main ambiguities encountered in this review. Exact questions and endpoints distinguish a scale from a monetary response. Aggregation details reveal whether a composite mixes appraisal and acquisition content. Payment and delivery information establish what follows from a choice. Reporting these fields makes later interpretation less dependent on the terminology chosen in an abstract.

For artists, galleries and studios, the practical consequence is to investigate the object actually offered. A digital image rating can inform presentation research; pricing a framed print requires evidence about decisions involving that print and its terms. The reviewed instruments provide a useful starting point for designing such work.

Table 3 Minimum information for interpreting a valuation result

Field	Information to report
Object	Image, physical work, print or edition; dimensions, medium and what ownership includes
Source	Actual production, selection and editing process; attributed source and exact disclosure wording
Question and response	Full wording, endpoints, units, bounds, comparative format and composition of any aggregated score
Decision context	Prices, anchors, budget, alternatives and whether a no-purchase response is available
Consequences	Payment rule, selection of binding decisions, delivery and observed completion
Design and sample	Recruitment, randomization unit, stimulus sampling and repeated-measure structure

Field	Information to report
Analysis	Prespecified estimand, raw and adjusted results, uncertainty, exclusions and missing data

5 Limitations

The purposive corpus supports an analysis of measurement practices, with incomplete coverage of the wider literature. Some relevant methods remained inaccessible, including recent transaction studies. The counts describe this corpus and cannot estimate how common a method or preference is across the field. The leave-one-report-out checks expose dependence within the selected evidence; they cannot correct missing evidence. In particular, the concentration of coded auction evidence in one report describes the accessible corpus, not the number of transaction studies that exist.

Extraction and source rechecks were performed within one AI-assisted workflow, without independent duplicate extraction. Reading the manuscript did not provide an independent human audit of every extraction record. Source locators and machine-readable records support further checking. Chamberlain et al.’s response categories were established from the publisher’s questionnaire, but inconsistent sample-size reporting in that and several other sources remains unresolved. Studies also share stimuli and differ in technology, population and object, so the records should not be treated as independent replications.

The classification describes what was elicited. It is not a validated scale of artistic value, and it does not estimate an overall human or AI advantage. The next step for a comprehensive synthesis would be a predefined search and extraction protocol, fuller access to transaction research and independent checking of the measurement records.

6 Conclusion

AI-art valuation studies elicit several kinds of response under overlapping names. The distinctions become clear at the instrument level: a rating of worth, a recommended or predicted selling price, a personal monetary offer and a consequential bid answer different questions. Mixed composites require attention to their constituent items as well. Evidence synthesis should retain the question, response format, aggregation and acquisition rule before grouping results. New studies can use the reporting fields in Table 3 to make those distinctions explicit.

Declarations

Data availability. Supplementary Methods S1–S8, evidence.json, evidence.csv, evidence-codebook.txt and reproduce_counts.py accompany this manuscript. The supplementary archive also contains the saved counts, coverage checks, claim audit and follow-up search log. These files contain classifications of published reports; no participant-level data were collected or reanalyzed. Source publications and the external codebook are cited through their public addresses.

Ethics. This review analyzes published research and does not recruit participants or collect personal data.

AI assistance. Section 2.4 describes the substantive use of OpenAI Codex for literature discovery, extraction, classification, writing and document preparation, together with the checks performed and their limitations.

Funding. The costs of preparing this review were covered by Whispart. Liyuantan is its founder and initiated the project.

Author contributions. Liyuantan initiated the project. Liyuantan and lichnagtan read the manuscript and agreed to its submission. Section 2.4 describes the substantive AI assistance used in discovery, extraction, classification and drafting.

Conflicts of Interest

Liyuantan is the founder of Whispart, an art business with an interest in AI-generated art and its commercial presentation. lichnagtan is a member of the Whispart team. These relationships are disclosed as potential competing interests. The review does not evaluate Whispart products.

References

- Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, 8, 42. <https://doi.org/10.1186/s41235-023-00499-6>
- Bianchi, I., Branchini, E., Uricchio, T., & Bongelli, R. (2025). Creativity and aesthetic evaluation of AI-generated artworks: bridging problems and methods from psychology to AI. *Frontiers in Psychology*, 16, 1648480. <https://doi.org/10.3389/fpsyg.2025.1648480>
- Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts*, 12(2), 177–192. <https://doi.org/10.1037/aca0000136>
- Chiarella, S. G., Torromino, G., Gagliardi, D. M., Rossi, D., Babiloni, F., & Cartocci, G. (2022). Investigating the negative bias towards artificial intelligence: Effects of prior assignment of AI-authorship on the aesthetic appreciation of abstract paintings. *Computers in Human Behavior*, 137, 107406. <https://doi.org/10.1016/j.chb.2022.107406>
- Cunningham, C. V., Radvansky, G. A., & Brockmole, J. R. (2025). Human creativity versus artificial intelligence: source attribution, observer attitudes, and eye movements while viewing visual art. *Frontiers in Psychology*, 16, 1509974. <https://doi.org/10.3389/fpsyg.2025.1509974>
- Depecik, B. E., Ozgen, E., Cosguner, K., & Inman, J. J. (2026). Art as revelation: Why AI art moves us less. *International Journal of Research in Marketing*. Advance online publication. <https://doi.org/10.1016/j.ijresmar.2026.09.002>
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465. <https://doi.org/10.1177/2515245920952393>
- Gu, L., & Li, Y. (2022). Who made the paintings: Artists or artificial intelligence? The effects of identity on liking and purchase intention. *Frontiers in Psychology*, 13, 941163. <https://doi.org/10.3389/fpsyg.2022.941163>
- Hall, J., & Schofield, D. (2025). The value of creativity: Human produced art vs. AI-generated art. *Art and Design Review*, 13(1), 65–88. <https://doi.org/10.4236/adr.2025.131005>
- Horton, C. B., Jr., White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports*, 13, 19001. <https://doi.org/10.1038/s41598-023-45202-3>
- Lane, T., Pickard, H., Walker, M. J., & Heinrich, T. (2026). A silver lining? Consumer preferences over human-made and AI-generated production in two domains. *Creative PEC Discussion Paper Series*, 2026/1. September version; manuscript dated July 17, 2026. <https://doi.org/10.5281/zenodo.22111802>

- Magni, F., Park, J., & Chao, M. M. (2024). Humans as creativity gatekeepers: Are we biased against AI creativity? *Journal of Business and Psychology*, 39, 643–656. <https://doi.org/10.1007/s10869-023-09910-x>
- 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, 107707. <https://doi.org/10.1016/j.chb.2023.107707>
- Neef, N. E., Zabel, S., Papoli, M., & Otto, S. (2025). Drawing the full picture on diverging findings: adjusting the view on the perception of art created by artificial intelligence. *AI & Society*, 40, 2859–2879. <https://doi.org/10.1007/s00146-024-02020-z>
- Oksanen, A., Cvetkovic, A., Akin, N., Latikka, R., Bergdahl, J., Chen, Y., & Savela, N. (2023). Artificial intelligence in fine arts: A systematic review of empirical research. *Computers in Human Behavior: Artificial Humans*, 1(2), 100004. <https://doi.org/10.1016/j.chbah.2023.100004>
- Prissé, B., Deng, R., Ho, J. Q., Jayasekara, D., & Ouangraoua, C. (2026). The creativity of artificial intelligence. *Journal of Behavioral and Experimental Economics*, 121, 102528. <https://doi.org/10.1016/j.socec.2026.102528>
- Schmidt, J., & Bijmolt, T. H. A. (2020). Accurately measuring willingness to pay for consumer goods: A meta-analysis of the hypothetical bias. *Journal of the Academy of Marketing Science*, 48, 499–518. <https://doi.org/10.1007/s11747-019-00666-6>
- Theirs, C., Villar, S., Martínez Marti, M. L., & Corradi, G. (2026). Can artificial intelligence exert effort? Effects of generation time on effort attribution and aesthetic evaluation of AI-generated art. *Collabra: Psychology*, 12(1), 164122. <https://doi.org/10.1525/collabra.164122>