ARCHAIA JOURNAL OF SCIENCE AND TECHNOLOGY
VOL. 3, NO. 4, November 2026
https://doi.org/10.48161/ajst.v3n4a101

Artificial Intelligence in Visual Arts: Transforming Creativity, Design, and Artistic Production

Zaripova Mukaddas Djumayozovna[1*], M.S. Sureshkumar[2]

[1] Computer and Software Engineering, College of Information Technologies, Information Technologies, Termez State University, Termez, Uzbekistan;
[2] Department of Polymer Technology and Engineering, Harare Institute of Technology, Harare, Zimbabwe;

*** Corresponding author:** zaripovamuqaddas0407@gmail.com.

ABSTRACT: Generative image systems are increasingly used in visual-art and design workflows, but controlled evidence rarely separates support for ideation from support for image generation and editing. This multicentre 2×2 counterbalanced crossover experiment enrolled 240 artists and advanced art/design students across six centers; 229 completed all four periods and 949 of 960 randomized participant artworks were submitted. Participants completed equivalent briefs under human-only, AI-assisted ideation, AI-assisted production, and full human–AI co-creation conditions, while 240 AI-only works formed a non-randomized benchmark. Four masked experts rated each artwork for originality, aesthetic quality, conceptual coherence, technical execution, and brief fulfillment. Rating-level cross-classified mixed models accounted for artwork, artist, and judge dependencies and adjusted for brief, center, period, sequence, and carryover, with Holm control across ten confirmatory main-effect tests. AI ideation improved originality (+0.34, 95% CI 0.16–0.52), aesthetic quality (+0.18, 0.01–0.35), conceptual coherence (+0.28, 0.10–0.46), and brief fulfillment (+0.24, 0.07–0.41), but not technical execution. AI production improved aesthetic quality (+0.47, 0.29–0.65), technical execution (+0.61, 0.43–0.79), and brief fulfillment (+0.38, 0.20–0.56), while reducing median production time by 31%; effects on originality and conceptual coherence were inconclusive. Full co-creation produced the fastest workflow and high technical quality, but secondary interaction estimates indicated no general synergy for originality. Production assistance reduced perceived agency, authorship, ownership, and collective visual diversity. In a randomized audience-label experiment, AI-assisted disclosure reduced authenticity and purchase intention, although it did not explain all expert–audience differences. Interviews linked benefit to reversible control, selective rejection, and prompt literacy. AI assistance therefore altered creative work differently by stage, improving selected output and efficiency dimensions while introducing agency, convergence, and provenance trade-offs.

Keywords: artificial intelligence; visual arts; human–AI co-creativity; creative agency; artistic production.

I. INTRODUCTION

Generative artificial intelligence has moved rapidly from specialist computational practice into studio art, graphic design, illustration, advertising, concept development, and arts education. Text-to-image and image-to-image systems can generate technically polished alternatives within seconds, but their artistic meaning remains dependent on how creators frame problems, select outputs, reject defaults, revise images, and communicate authorship. Earlier scholarship distinguished automation of craft from the social attribution of artistic intention [1]. A systematic review of empirical research in fine arts found rapid growth in studies of evaluation and acceptance but comparatively little controlled evidence on how artists use generative systems during actual production [2].

Evaluation is socially situated. Viewers often respond differently to works described as human-made or AI-assisted, particularly when judging authenticity, effort, emotional depth, and authorship [3]–[5]. Bias can persist even when visible features are comparable [6], while production-stage disclosure may reduce perceived authenticity more strongly than early ideation support [7]. Consequently, a single aesthetic rating cannot adequately represent creative achievement. Evaluation needs to separate visible product quality from beliefs about origin, human contribution, labor, and accountability, while distinguishing blinded judgments from responses after disclosure.

Field and experimental evidence indicates that generative tools may improve productivity, market value, individual idea generation, and creative confidence when users retain editorial control [8]–[10]. Productivity and creativity are nevertheless non-equivalent. Faster rendering can release time for exploration, but it can also encourage premature selection of fluent outputs. AI may widen the option set available to one artist while simultaneously drawing many artists toward similar visual conventions. This tension between individual enhancement and collective convergence matters especially in visual culture, where originality concerns differentiation across works, artists, and communities, as well as the quality of an isolated product.

Existing studies seldom combine stage-specific randomization, within-artist control of baseline skill, masked expert evaluation, complete workflow telemetry, computational diversity analysis, provenance auditing, audience-label manipulation, and longitudinal qualitative follow-up. Professional artists and advanced students are also often pooled without testing whether expertise changes the benefits and costs of assistance. The present study addressed these limitations through a multicentre counterbalanced crossover experiment that separated AI-assisted ideation from AI-assisted production. The confirmatory estimands were the main effects of ideation and production, averaged across five expert-rated domains. Interactions, agency and workflow outcomes, computational similarity, audience disclosure, moderators, and process associations were secondary or exploratory. We hypothesized that ideation support would primarily improve originality, conceptual coherence, and brief fulfillment; production support would primarily improve aesthetic quality, technical execution, and speed; and extensive production assistance could reduce authorship, control, and collective diversity even when technical quality improved.

II. RESEARCH PROBLEM, AIM, OBJECTIVES, AND QUESTIONS

1. RESEARCH PROBLEM

Current evidence does not establish whether generative AI improves visual creativity by expanding early conceptual search, accelerating execution, increasing technical fluency, or changing the standards by which work is judged. Researchers often compare human-only, AI-assisted, and AI-generated images across participants or with mismatched tasks. Such comparisons confound production condition with baseline skill, brief difficulty, prompt experience, exposure order, evaluator expectations, and selective curation. They also obscure when AI entered the workflow. An image generator used to create provisional references may affect ideation differently from the same system used to synthesize pixels in the final work. Without stage-specific randomization, reported advantages cannot be attributed clearly to exploration or execution.

A second problem concerns how we define and measure creative benefit. Higher aesthetic polish or faster completion does not necessarily indicate greater originality, stronger conceptual coherence, or a more agentic creative experience. Assistance may improve technical execution while weakening perceived authorship, psychological ownership, autonomy, and confidence in one's own decision-making. Conversely, AI suggestions may support divergent thinking when artists can reject or transform them. Studies therefore require multiple expert-rated creative domains, validated or transparently adapted psychological measures, complete records of prompts and revisions, and analyses that preserve the distinction between product quality, creative process, and human agency.

A third problem concerns collective and ethical consequences. Shared models and prompt conventions may increase within-artist exploration while reducing between-artist diversity. CLIP, LPIPS, and related embeddings can quantify selected semantic and perceptual distances, but they cannot determine cultural originality, artistic significance, or ethical legitimacy. Computational results therefore require triangulation with expert judgment, workflow evidence, provenance records, and artist accounts. Crossover studies of AI

use are also vulnerable to learning and carryover because prompting, curation, and error detection can improve across periods. A rigorous design must counterbalance briefs and conditions, model artwork-level and judge-level dependencies, examine period and sequence effects, test first-period and complete-sequence sensitivities, and interpret an AI-only benchmark descriptively because it is not human-randomized.

2. RESEARCH AIM AND OBJECTIVES

The study aimed to estimate how AI assistance at the ideation and production stages independently and jointly affected visual creative quality, workflow efficiency, human agency, diversity, and responsible production.

- Objective 1— To estimate the independent and combined effects of AI-assisted ideation and AI-assisted production on masked expert ratings of originality, aesthetic quality, conceptual coherence, technical execution, and brief fulfillment.
- Objective 2— To estimate effects on agency, authorship, psychological ownership, autonomy, creative self-efficacy, cognitive load, production time, exploration breadth, revision depth, satisfaction, and artistic risk-taking.
- Objective 3—To compare within-artist exploration, between-artist diversity, perceptual distance, provenance completeness, disclosure accuracy, audience responses, and the descriptive AI-only benchmark.
- Objective 4—To examine interactions, professional status, prior AI experience, task complexity, perceived control, and temporally ordered process associations, and integrate quantitative, computational, and qualitative evidence.

3. RESEARCH QUESTIONS

- RQ1. What were the independent and combined effects of AI-assisted ideation and AI-assisted production on masked expert-rated creative quality?
- RQ2. How did ideation and production assistance affect creative agency, authorship, ownership, autonomy, self-efficacy, cognitive load, production time, exploration, revision, satisfaction, and artistic risk-taking?
- RQ3. How did the randomized conditions differ in computational diversity, provenance compliance, and audience response, and how did AI-only benchmark works compare descriptively?
- RQ4. How did professional status, prior AI experience, brief complexity, perceived control, and artist accounts explain benefit, limited benefit, convergence, and ethical or cultural concerns?

The detailed objective–question–analysis alignment is provided in Supplementary Table S1.

III. LITERATURE REVIEW

1. STAGE-SPECIFIC HUMAN–AI CO-CREATIVITY

Human–AI co-creativity research increasingly treats generative systems as interactive sources of representations rather than passive software features. Co-creative drawing experiments show that visual suggestions can support conceptual shifts while also changing turn-taking, control, and perceived contribution [11]. However, much of this evidence comes from short laboratory tasks, proof-of-concept systems, or student samples. Evidence is therefore stronger for immediate interaction experience than for professional artistic production, stage-specific attribution, or sustained practice.

Text-to-image prompting is a cultivated practice involving descriptive language, modifiers, iteration, seed management, curation, and participation in online communities [12]. Image generators can stimulate unexpected associations but may also produce visually fluent solutions that neglect constraints or converge on familiar styles. These findings support separating ideation and production. Evaluate ideation support by the breadth and conceptual usefulness of alternatives, and evaluate production support by execution, speed, revision, and the extent of human transformation.

https://doi.org/10.48161/ajst.v3n4a101

2. MEASUREMENT OF CREATIVE QUALITY AND AGENCY

Authorship in computer-generated art depends on intention, intervention, selection, and responsibility rather than on visible form alone [13]. The Consensual Assessment Technique provides a well-established basis for judging creative products through independent evaluations by domain-relevant experts [14]. Its validity depends on appropriate expertise, independent rating, calibration, sufficient judges, and transparent reliability analysis. Aesthetic quality alone is insufficient because creativity judgments involve novelty, conceptual resolution, technical execution, and fit to task constraints. The present study therefore used five anchored domains and four masked ratings per artwork, with artwork and judge dependencies represented explicitly in the analysis.

Agency-related constructs require operational separation. Creative self-efficacy concerns confidence in producing creative outcomes [15]. Psychological ownership concerns the experienced sense that an object is "mine" [16], while sense of agency concerns perceived control over actions and outcomes [17]. Cognitive demand was assessed through the NASA Task Load Index [18]. Artistic risk-taking has no universally accepted visual-AI measure; therefore, we treated the adapted creative-risk items as exploratory and informed by the DOSPERT approach [19]. We measured task autonomy and perceived authorship separately so we did not treat control, ownership, and credit as interchangeable constructs.

3. DIVERSITY, DISCLOSURE, AND PRODUCTION GOVERNANCE

Evaluation bias can lower appreciation of AI-labelled art even when visible features are comparable [3], [5], [6]. Another study indicates that disclosure effects can differ depending on whether AI enters ideation or final production [7]. A blinded primary evaluation is therefore necessary, while a separate randomized disclosure experiment can estimate the effect of labels on authenticity, creativity, and purchase or display intentions. Such a label effect cannot explain all differences between expert and audience samples because the samples, tasks, and judgment contexts differ.

Computational image embeddings can supplement human evaluation by quantifying semantic and perceptual dispersion. CLIP cosine distance can represent selected conceptual differences [20], and LPIPS can provide a perceptual-distance sensitivity analysis [21]. These measures remain model-dependent proxies. We examined robustness using DINOv2 [22], palette entropy, and layout descriptors, with participant-clustered bootstrap uncertainty. Provenance manifests documented declared tools, edits, and model versions in accordance with the official C2PA Technical Specification, Version 2.1 [23]; they supported traceability and signed workflow history rather than proving artistic authorship, cultural originality, or truthfulness.

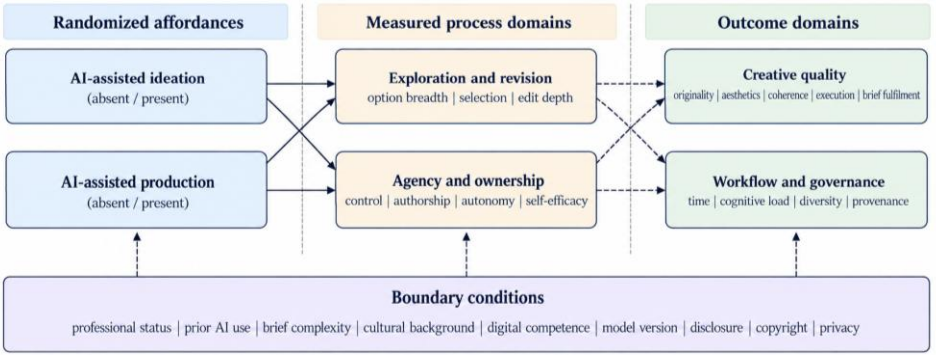
4. CRITICAL SYNTHESIS AND PRESPECIFIED EXPECTATIONS

The literature supports a stage-specific account rather than a single claim that AI enhances or diminishes creativity. Ideation support is most plausibly related to exploration breadth, conceptual alternatives, and self-efficacy. Production support is most plausibly related to technical execution, time, and revision options, but may also transfer expressive decisions to the system. Combined use may improve some cell means without producing a positive statistical interaction. The confirmatory analysis therefore focused on main effects for ideation and production, while interactions, mediation, moderators, computational diversity, and AI-only comparisons remained secondary or exploratory. Agency and ownership were treated as outcomes in their own right rather than inferred from product ratings.

IV. STAGE-SPECIFIC HUMAN–AI CO-CREATIVITY FRAMEWORK

The framework integrates three complementary perspectives. Distributed cognition explains how prompts, generated images, drafts, interfaces, and prior outputs change the representations available to the artist [24]. The dynamic componential model distinguishes domain expertise, creativity-relevant processes, motivation, and environmental affordances [25]. Self-determination theory distinguishes assistance that supports autonomy and competence from assistance that feels controlling or substitutive [26]. Creative self-efficacy remains a measurable process rather than a separate theoretical system [15].

Two randomized affordances—AI-assisted ideation and AI-assisted production—were linked to two measured process domains: exploration/revision and agency/ownership. Creative-quality outcomes were originality, aesthetic quality, conceptual coherence, technical execution, and brief fulfillment; workflow and



governance outcomes included time, cognitive load, diversity, and provenance. Professional status, prior AI use, brief complexity, cultural background, digital competence, model version, and disclosure were boundary conditions. Only the stage-specific AI factors were randomized. We interpreted associations involving exploration, self-efficacy, revision depth, control, or ownership as explanatory and model-dependent rather than as unique causal mechanisms.

FIGURE 1. Stage-Specific Human–AI Co-Creativity Framework. Solid arrows represent randomized ideation and production effects; dashed arrows represent exploratory process associations. Boundary conditions may influence processes and outcomes without being randomized.

V. METHODOLOGY

1. STUDY DESIGN, SETTING, REGISTRATION, AND REPORTING

A multicentre randomized 2×2 counterbalanced crossover experiment examined AI-assisted ideation (absent/present) and AI-assisted production (absent/present). The four conditions were human-only, AI ideation with human production, human ideation with AI generation/editing, and full human–AI co-creation. Each participant completed one equivalent brief in each condition. Four Williams Latin-square sequences balanced period and first-order carryover, and sessions were separated by at least seven days. Briefs differed in theme, visual genre, and audience to discourage direct reuse. Because prompting and curation skills can persist across periods, we defined the primary estimand as the average stage-specific effect under repeated exposure rather than a learning-free, single-use effect. The AI-only benchmark was generated separately and was excluded from randomized causal estimands.

Six university art schools and professional creative hubs participated. We recruited participants from March to June 2023, conducted experimental sessions from April to November 2023, and followed up for three months from July 2023 to February 2024. The database was locked on 18 March 2024. We prospectively archived the protocol, confirmatory outcomes, brief-matching procedure, analysis plan, and provenance rules before enrolment. Reporting followed the CONSORT extension for randomized crossover trials [27].

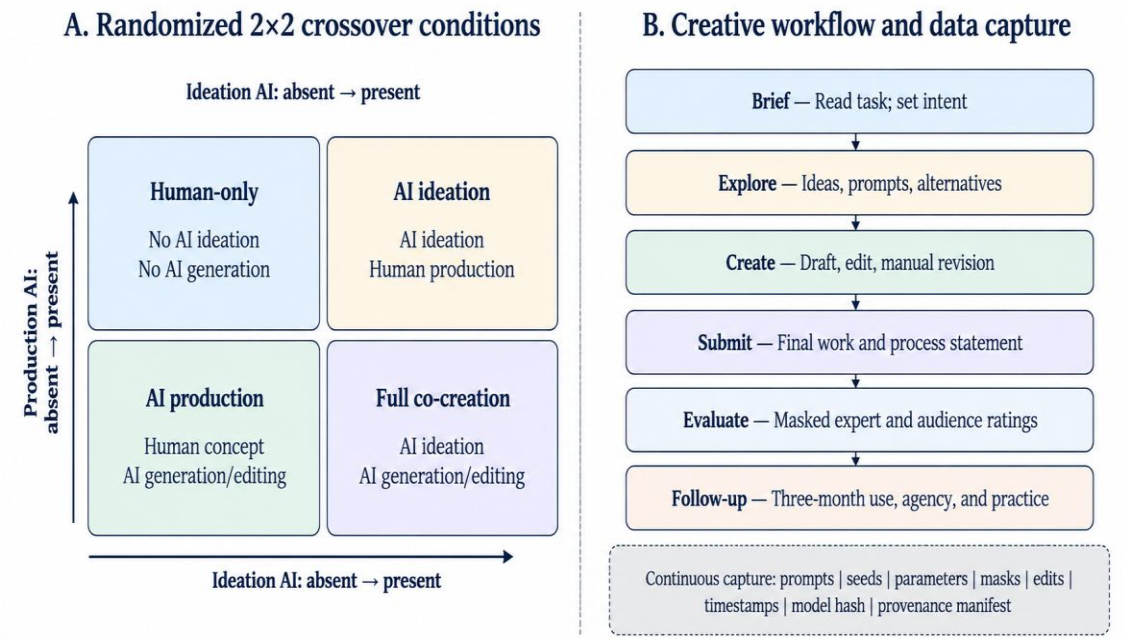
2. PARTICIPANTS, POWER, RANDOMIZATION, AND MASKING

The sample comprised 240 adults: 120 visual artists with at least three years of paid practice and 120 advanced students in the final two years of an accredited art or design program. Eligible disciplines included illustration, graphic design, digital painting, photography-based art, concept art, and mixed-media design. Participants needed normal or corrected vision and basic digital production skills. Exclusion criteria included employment by the AI-system provider, prior involvement in brief development, or inability to consent to recording the process.

The power analysis generated 10,000 datasets representing four ratings per artwork, participant, artwork, judge, brief, center, period, sequence, first-order carryover, a within-person correlation of 0.50, judge

reliability of 0.70, and 10% incomplete sequences. With Holm control across ten confirmatory main-effect tests and an initial alpha of .005, 240 participants provided 81.3% power for a standardized main effect of 0.20. Power was 52.6% for a 0.20 interaction and 79.8% for a 0.30 interaction; interactions, subgroup contrasts, and mediation were therefore secondary or exploratory.

An independent statistician generated sequence assignments within center and professional-status strata using permuted blocks of eight. Allocation was released after baseline assessment and familiarization. Participants could not be blinded to condition, but briefs were coded, and final files were stripped of workflow metadata before rating. Expert judges, initial audience raters, data abstractors, and the primary statistician remained masked until confirmatory models were finalized. Audience participants were assigned



1:1:1 to human-made, AI-assisted, or undisclosed labels after an initial unlabeled rating, using computer-generated blocks stratified by baseline attitude toward AI art.

FIGURE 2. Randomized 2×2 crossover architecture and creative workflow. We counterbalanced equivalent briefs across four periods and captured complete workflow telemetry and provenance records.

3. EXPERIMENTAL CONDITIONS, AI SYSTEM, AND CREATIVE BRIEFS

In the human-only condition, participants used conventional non-generative digital or physical tools and a fixed reference library. In AI-assisted ideation, participants could generate moodboards, compositional alternatives, and conceptual prompts, but they had to produce the final image without generative synthesis or AI editing. In AI-assisted production, participants developed the concept independently during a timed paper-based phase and then used generation, inpainting, outpainting, and variation tools. Full co-creation permitted AI at both stages. All conditions allowed ordinary color correction, cropping, and manual drawing. Direct prompting in the style of living artists, use of personal data, and upload of third-party copyrighted images were prohibited.

AI conditions used a locked research deployment of Stable Diffusion XL 1.0 with standardized safety filters and a fixed inference library. The system automatically captured prompts, negative prompts, seeds, generation parameters, masks, output identifiers, and manual edits. C2PA-compatible manifests linked final works to declared conditions, source files, edits, and model-version records in accordance with the C2PA Technical Specification, Version 2.1 [23]. The manifests documented signed workflow history and validation status but were not treated as independent proof of authorship, cultural originality, or artistic value.

Eight briefs covered editorial illustration, cultural-poster design, environmental concept art, narrative character design, album-cover design, public-service communication, speculative-product visualization, and abstract identity design. External pilot testing involved 32 artists who did not enter the experiment. Mean estimated completion time ranged from 147.2 to 152.6 minutes, complexity ratings from 5.18 to 5.43 on a seven-point scale, and semantic openness from 5.31 to 5.58; omnibus tests showed no material brief differences in time, $F(7,217)=0.84$, $p=.556$, complexity, $F(7,217)=0.91$, $p=.501$, or openness, $F(7,217)=1.02$, $p=.418$. Each session lasted 150 minutes: 15 minutes for reading and planning, 115 minutes for production, and 20 minutes for submission, process reflection, and post-task measures.

Table 1. Experimental conditions and permitted artificial-intelligence functions.

Condition	Ideation stage	Production stage	Permitted AI functions	Core interpretive contrast
Human-only	Manual sketching and fixed reference library	Conventional physical/digital tools	None	Baseline human workflow
AI-assisted ideation	AI moodboards, variants, conceptual alternatives	Human-only production	Prompted image generation for reference; no AI pixels in final	Effect on exploration
AI-assisted production	Timed human concept development	AI generation, inpainting, outpainting, variations	AI pixels permitted with manual revision	Effect on execution
Full human–AI co-creation	AI-supported concept exploration	AI-supported generation and editing	Full permitted workflow with provenance	Combined effect and interaction
AI-only benchmark	Standardized brief-to-prompt conversion	Locked model output; no artist revision	Fixed seeds and safety filters	Descriptive benchmark only

4. OUTCOMES, INSTRUMENTS, AND MEASUREMENT QUALITY

The confirmatory outcomes were expert ratings of originality, aesthetic quality, conceptual coherence, technical execution, and brief fulfillment on anchored 1–10 scales. The rubric followed the Consensual Assessment Technique [14]. Twenty-four judges with at least seven years of relevant professional or academic experience completed two calibration rounds, and every work received four independent ratings balanced across judges, centers, and briefs. Calibration ICC was .81; rating-model generalizability coefficients ranged from .74 for originality to .86 for technical execution.

Creative self-efficacy used the three-item Tierney–Farmer scale [15]. Psychological ownership used four task-adapted Van Dyne–Pierce items [16]. Task agency used five state-adapted items from the Sense of Agency Scale [17], and autonomy used a seven-item task-adapted perceived-choice measure grounded in self-determination theory [26]. We used the six-domain Raw NASA-TLX [18] for cognitive workload and treated it as a multidimensional workload index rather than a unidimensional trait. Perceived authorship was assessed with the four-item Perceived Authorship Attribution Scale (PAAS-4; 1–7, higher scores indicating stronger attribution of the final work to the participant), grounded in authorship and intention research [4], [13]. Artistic risk-taking was assessed with four task-specific Artistic Risk-Taking Items (ARTI-4; 1–7, higher scores indicating greater willingness to depart from familiar solutions), informed by DOSPERT [19]. We designated PAAS-4 and ARTI-4 as study-specific exploratory measures rather than established, validated scales. Satisfaction used three seven-point items.

Sample-specific omega coefficients were .88 for creative self-efficacy, .91 for psychological ownership, .86 for agency, .89 for autonomy, .80 for ARTI-4, .90 for PAAS-4, and .87 for satisfaction. Within-person reliability ranged from .72 to .81. Configural and metric invariance were supported across professional status and

conditions; scalar invariance was supported for self-efficacy, autonomy, and satisfaction and was partial for agency, ownership, and authorship after freeing three item intercepts. Sensitivity analyses using partial-invariance factor scores and item-level ordinal models reproduced the direction of the observed-score findings. PAAS-4 and ARTI-4 remained outside the confirmatory multiplicity family, and we interpreted their results as exploratory. Full wording, scoring, timing, factor structure, and psychometric results are reported in Supplementary Tables S2 and S6.

5. STATISTICAL ANALYSIS

The rating, rather than the participant or artwork mean, was the primary observational unit. For rating y of artwork a created by artist i under brief b, rated by judge j at center c in period p and sequence s, the confirmatory model included fixed effects for AI ideation, AI production, their interaction, period, sequence, brief, center, professional status, prior AI use, task complexity, and prespecified first-order carryover terms. Random intercepts represented artist, artwork nested within artist, and judge. Artist-level random slopes for ideation and production were retained for originality, aesthetic quality, and technical execution; singular fits for conceptual coherence and brief fulfillment led to the prespecified random-intercept specification. This structure prevented the four ratings of one artwork from being treated as independent replications [28].

The model was $y = \beta0 + \beta1(\text{Ideation}) + \beta2(\text{Production}) + \beta3(\text{Ideation}\times\text{Production}) + \beta4(\text{Period}) + \beta5(\text{Sequence}) + \beta6(\text{Brief}) + \beta7(\text{Centre}) + \beta8(\text{Carryover}) + u\text{Artist} + u\text{Artwork(Artist)} + u\text{Judge} + \varepsilon$. The five ideation and five production main-effect tests formed one ten-test Holm family. The five interactions formed a secondary Holm family. Agency/workflow, audience, computational, moderator, and process analyses were organized in separate false-discovery-rate families defined in Supplementary Table S3. AI-only contrasts remained descriptive and used Dunnett adjustment because the benchmark was not human-randomized.

Crossover diagnostics included previous-condition terms, factor-by-period interactions, sequence-specific estimates, first-period-only analyses, exclusion of period 4, and analyses restricted to complete sequences. Additional learning-adjusted models included the cumulative number of prior AI-assisted periods and its interactions with ideation and production; these terms assessed acquired prompting and curation skill separately from immediate carryover. Brief-by-condition interactions assessed residual mismatch among the eight briefs. We handled missing rating outcomes with maximum-likelihood mixed models; we multiply imputed missing questionnaire and workflow variables in 50 datasets using allocation, sequence, center, brief, professional status, repeated outcomes, prior AI use, and recorded missingness reasons. Sensitivity analyses used robust judge weighting, alternative covariance structures, exclusion of protocol deviations, and artwork-level aggregated ratings.

The audience-label experiment used mixed models with randomized label, actual production condition, label-by-condition interaction, baseline attitude toward AI art, and random intercepts for audience participant and artwork. Computational diversity estimates used participant-clustered bootstrap intervals from 5,000 resamples and were repeated using CLIP [20], LPIPS [21], DINOv2 embeddings [22], palette entropy, and layout descriptors. We interpreted computational metrics as similarity indicators rather than direct measures of originality or cultural value.

6. QUALITATIVE AND MIXED-METHODS EVALUATION

Maximum-variation sampling selected 48 participants after period 4: 24 professionals and 24 students spanning high and low quality change, strong and weak agency, extensive and minimal AI use, and varied cultural backgrounds. Forty completed a second interview three months later. The authors judged information power adequate because the sample was specific, the questions were focused, dialogue quality was high, and the analysis compared predefined trajectories across conditions.

Framework Analysis combined deductive domains from the conceptual model with inductively developed categories [29]. A visual-arts researcher and a human–computer interaction researcher independently charted the first 12 transcripts; disagreements were resolved in recorded meetings with a qualitative-methods auditor. Reflexive statements documented disciplinary position, prior AI use, and relationships with participants. The audit trail linked quotations to participant code, professional status,

center, condition history, quality trajectory, agency trajectory, and interview time point. Negative and disconfirming cases were retained. Joint displays compared experimental effects, workflow telemetry, computational metrics, provenance records, and participant explanations without treating quotations as verification of statistical mediation [30].

7. ETHICS, COPYRIGHT, PRIVACY, AND PROVENANCE

The Coordinating University Arts and Human Research Ethics Committee approved the study before recruitment (approval 114; 16 December 2022), and each center completed a data-governance review. Participants provided written informed consent for artwork creation, screen recording, prompt capture, questionnaires, interviews, controlled sharing of de-identified workflow data, and separate artwork publication. Participation was voluntary, withdrawal did not affect education or employment, and participants could request deletion of screen recordings before de-identification. We stored identifiers separately from encrypted, role-restricted research files. Copyright screening, privacy review, provenance verification, and cultural-harm checks preceded public display. C2PA manifests were validated and interpreted within the scope and limitations of the official specification [23].

VI. RESULTS

1. PARTICIPANT, ARTWORK, AND RATING FLOW

Of 312 individuals assessed, 240 were randomized: 120 professional artists and 120 advanced students. Eleven participants missed one period, producing 949 of 960 randomized participant artworks; 229 completed all four periods, and all 240 contributed to mixed models. Period completion was 240, 238, 236, and 235, respectively, corresponding to 2 missed sessions in period 2, 4 in period 3, and 5 in period 4. The four sequences contributed 237, 237, 238, and 237 artworks. The participant tracking log retained individual reasons for each missed period, and the supplementary flow record summarized them. Twenty-four expert judges produced 4,756 complete rating sets across 1,189 participant and benchmark artworks. The AI-only benchmark comprised 240 works, and 720 audience participants produced 14,400 ratings. Forty-eight participants completed an initial interview, and 40 completed the three-month interview.

Baseline characteristics were balanced across sequences. The mean age was 29.8 years (SD 7.4); 58.3% identified as women, 39.6% as men, and 2.1% used another description or preferred not to respond. 54.2% reported prior text-to-image use, and 17.9% reported weekly use. Eleven of 949 sessions (1.2%) had minor timing or logging deviations, and six (0.6%) had a prohibited prompt or imported asset detected before final submission. All remained in intention-to-treat analyses; exclusion of the six major deviations did not materially change estimates.

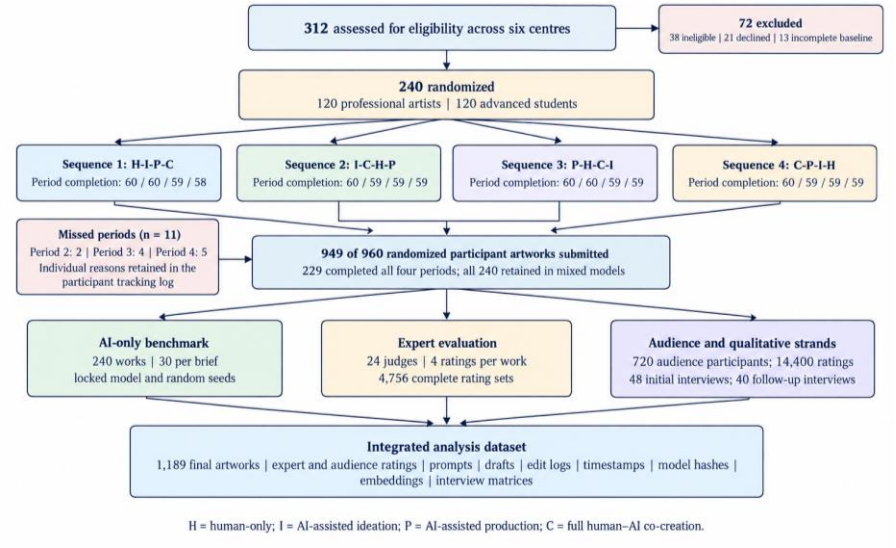


FIGURE 3. Participant, sequence, artwork, expert rating, audience rating, and interview flow through the study. Period-specific completion counts are shown within each randomized sequence.

Table 2. Selected baseline participant characteristics by professional status.

Characteristic	Artists (n=120)	Advanced students (n=120)	Total (n=240)
Age, mean (SD)	34.7 (6.5)	24.9 (3.4)	29.8 (7.4)
Women, n (%)	66 (55.0)	74 (61.7)	140 (58.3)
Years of advanced practice, median (IQR)	8.0 (5.0–12.0)	3.0 (2.0–4.0)	4.5 (3.0–8.0)
Weekly prior AI use, n (%)	25 (20.8)	18 (15.0)	43 (17.9)
Any prior text-to-image use, n (%)	68 (56.7)	62 (51.7)	130 (54.2)
Creative self-efficacy, mean (SD), 1–7	5.42 (0.78)	5.06 (0.82)	5.24 (0.82)
Digital competence, mean (SD), 1–7	5.66 (0.71)	5.51 (0.76)	5.59 (0.74)
Disciplines: illustration / design / digital art / other	39 / 31 / 35 / 15	34 / 38 / 32 / 16	73 / 69 / 67 / 31
Centers represented, n	6	6	6

2. CONFIRMATORY CREATIVE-QUALITY OUTCOMES

Adjusted cell means were 6.55 (human-only), 6.89 (AI ideation), 6.66 (AI production), and 6.80 (full co-creation) for originality. Corresponding means were 6.70, 6.88, 7.17, and 7.27 for aesthetic quality; 6.61, 6.89, 6.50, and 6.63 for conceptual coherence; 6.48, 6.56, 7.09, and 7.07 for technical execution; and 6.74, 6.98, 7.12, and 7.20 for brief fulfillment.

AI ideation increased originality by 0.34 points (95% CI 0.16–0.52; g=.24; Holm-adjusted p<.001), aesthetic quality by 0.18 (0.01–0.35; g=.13; p=.044), conceptual coherence by 0.28 (0.10–0.46; g=.20; p=.005), and brief fulfilment by 0.24 (0.07–0.41; g=.18; p=.013). The technical-execution estimate was 0.08 (−0.09 to 0.25; p=.341). AI production increased aesthetic quality by 0.47 (0.29–0.65; g=.34; p<.001), technical execution by 0.61 (0.43–0.79; g=.43; p<.001), and brief fulfillment by 0.38 (0.20–0.56; g=.28; p<.001). Originality and conceptual coherence were inconclusive. The originality interaction was −0.20 (−0.41 to 0.01; secondary Holm p=.083); the other interaction intervals included zero.

Variance attributable to artist ranged from 0.18 to 0.31, artwork from 0.34 to 0.46, judge from 0.10 to 0.21, and residual error from 0.68 to 0.82 across outcomes. Previous-condition coefficients did not exceed 0.09 points, and all adjusted p-values exceeded .28. Factor-by-period interactions were non-significant (omnibus p=.173–.684), and cumulative-exposure interactions were small and did not alter the direction or interpretation of the factorial main effects. First-period-only, complete-sequence, period-4 exclusion, robust-judge, and artwork-mean analyses reproduced the direction and approximate magnitude of the main effects. The largest brief-by-condition deviation was 0.14 points, with no significant omnibus interaction, p=.214. These analyses reduce, but do not eliminate, the possibility that acquired prompting or curation skill influenced later periods.

Table 3. The corresponding confidence-interval plot is provided as supplementary figure S1.

Outcome	Adjusted cell means H / I / P / C	Ideation effect (95% CI); Holm p	Production effect (95% CI); Holm p	Interaction (95% CI); Holm p
Originality	6.55 / 6.89 / 6.66 / 6.80	+0.34 (0.16, 0.52); <.001	+0.11 (−0.07, 0.29); .241	−0.20 (−0.41, 0.01); .083
Aesthetic quality	6.70 / 6.88 / 7.17 / 7.27	+0.18 (0.01, 0.35); .044	+0.47 (0.29, 0.65); <.001	−0.08 (−0.28, 0.12); .572

Outcome	Adjusted cell means H / I / P / C	Ideation effect (95% CI); Holm p	Production effect (95% CI); Holm p	Interaction (95% CI); Holm p
Conceptual coherence	6.61 / 6.89 / 6.50 / 6.63	+0.28 (0.10, 0.46); .005	−0.11 (−0.29, 0.07); .221	−0.15 (−0.35, 0.05); .238
Technical execution	6.48 / 6.56 / 7.09 / 7.07	+0.08 (−0.09, 0.25); .341	+0.61 (0.43, 0.79); <.001	−0.10 (−0.30, 0.10); .511
Brief fulfilment	6.74 / 6.98 / 7.12 / 7.20	+0.24 (0.07, 0.41); .013	+0.38 (0.20, 0.56); <.001	−0.16 (−0.36, 0.04); .196

3. *SECONDARY CREATIVE AGENCY AND WORKFLOW OUTCOMES*

AI production reduced median production time from 149 to 102 minutes (a 31% reduction), while AI ideation increased the mean number of distinct retained concepts from 4.7 to 6.5. Full co-creation was fastest (96 minutes) and generated the most machine alternatives, but the median proportion of the selected image receiving structural manual revision was lower than in AI production alone (31% versus 44%).

AI ideation increased creative self-efficacy (g=.21), satisfaction (g=.30), and artistic risk-taking (g=.17), and reduced cognitive load (g=−.18). AI production reduced task agency (g=−.41), perceived authorship (g=−.48), psychological ownership (g=−.31), and autonomy (g=−.27), despite higher satisfaction and lower time demand. Professionals showed a larger reduction in production-related authorship than students (interaction −0.19, 95% CI −0.34 to −0.04; FDR-adjusted p=.048), whereas the production benefit for technical execution was 0.22 points smaller among professionals (−0.40 to −0.04; p=.048). Prior weekly AI use moderated cognitive load (interaction −0.15 SD, −0.28 to −0.02; p=.041), but other subgroup interactions were imprecise.

Table 4. Creative-agency and workflow outcomes. Positive standardized effects indicate higher scores; production-time estimates are in minutes.

Outcome	Human-only	AI ideation effect (95% CI)	AI production effect (95% CI)
Creative agency, z	0.00	−0.08 (−0.21, 0.05)	−0.41 (−0.54, −0.28)
Perceived authorship, z	0.00	−0.05 (−0.18, 0.08)	−0.48 (−0.61, −0.35)
Psychological ownership, z	0.00	−0.03 (−0.16, 0.10)	−0.31 (−0.44, −0.18)
Autonomy, z	0.00	−0.02 (−0.15, 0.11)	−0.27 (−0.40, −0.14)
Creative self-efficacy, z	0.00	+0.21 (0.08, 0.34)	+0.05 (−0.08, 0.18)
Production time, median minutes	149	−7 (−14, 1)	−47 (−55, −39)
Distinct retained concepts, mean	4.7	+1.8 (1.3, 2.3)	+0.2 (−0.3, 0.7)
Cognitive load, z	0.00	−0.18 (−0.31, −0.05)	−0.12 (−0.25, 0.01)
Satisfaction, z	0.00	+0.30 (0.17, 0.43)	+0.22 (0.09, 0.35)
Artistic risk-taking, z	0.00	+0.17 (0.04, 0.30)	−0.01 (−0.14, 0.12)

The corresponding standardized-effect plot is provided as Supplementary Figure S2.

4. *COMPUTATIONAL DIVERSITY, AI-ONLY BENCHMARK, AND PROVENANCE*

AI ideation increased within-artist CLIP dispersion by 0.24 SD (95% bootstrap CI 0.15–0.33), indicating a broader personal option set. AI production reduced between-artist CLIP distance by 0.18 SD (−0.26 to −0.10) and LPIPS distance by 0.16 SD (−0.24 to −0.08); full co-creation reduced them by 0.24 and 0.22 SD. DINOv2

between-artist estimates were similar: −0.05 SD for ideation, −0.15 for production, −0.21 for full co-creation, and −0.33 for the AI-only benchmark. Palette entropy declined by 0.12 SD under production and 0.18 SD under full co-creation, while layout similarity increased by 0.14 and 0.22 SD, respectively. The convergent pattern supported a conclusion of feature-space convergence, not a comprehensive claim about cultural or artistic homogenization.

AI-only benchmark works achieved mean technical execution of 7.05 but lower conceptual coherence of 6.21. Provenance completeness was 96.2% (229/238) for AI ideation, 97.9% (232/237) for AI production, 99.1% (233/235) for full co-creation, and 100% (240/240) for the benchmark. First-audit disclosure accuracy was 95.0%, 97.0%, 97.9%, and 100% for the first-audit, second-audit, third-audit, and benchmark, respectively.

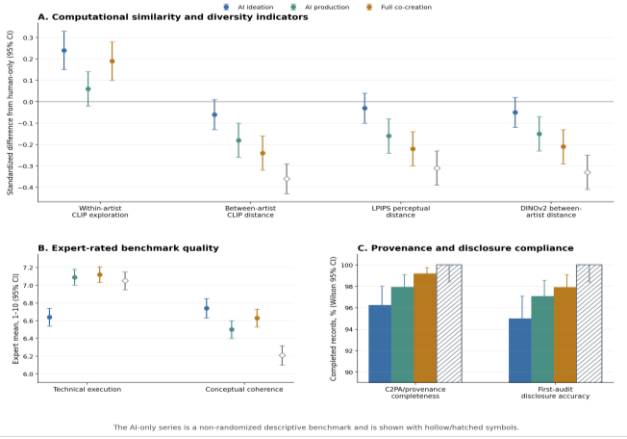


FIGURE 4. Computational diversity, expert-rated benchmark quality, and provenance outcomes displayed on separate scales. Panel A shows standardized differences with participant-clustered 95% bootstrap intervals; Panel B shows expert means with 95% confidence intervals; Panel C shows proportions with Wilson 95% confidence intervals. The AI-only series is a non-randomized descriptive benchmark, distinguished by hollow or hatched symbols.

Table 5. Computational diversity, benchmark quality, and provenance findings with explicit units, denominators, and uncertainty.

Outcome (unit / denominator)	AI ideation	AI production	Full co-creation	AI-only benchmark
Within-artist CLIP exploration, SD difference (95% CI)	+0.24 (0.15, 0.33)	+0.06 (−0.02, 0.14)	+0.19 (0.10, 0.28)	Not applicable
Between-artist CLIP distance, SD difference (95% CI)	−0.06 (−0.13, 0.01)	−0.18 (−0.26, −0.10)	−0.24 (−0.32, −0.16)	−0.36 (−0.43, −0.29)
LPIPS perceptual distance, SD difference (95% CI)	−0.03 (−0.10, 0.04)	−0.16 (−0.24, −0.08)	−0.22 (−0.30, −0.14)	−0.31 (−0.39, −0.23)
DINOv2 between-artist distance, SD difference (95% CI)	−0.05 (−0.12, 0.02)	−0.15 (−0.23, −0.07)	−0.21 (−0.29, −0.13)	−0.33 (−0.41, −0.25)
Mean expert technical execution, 1–10	6.64	7.09	7.12	7.05
Mean expert conceptual coherence, 1–10	6.74	6.50	6.63	6.21

Outcome (unit / denominator)	AI ideation	AI production	Full co-creation	AI-only benchmark
C2PA/provenance completeness	229/238 (96.2%; 95% CI 93.0–98.0)	232/237 (97.9%; 95% CI 95.2–99.1)	233/235 (99.1%; 95% CI 97.0–99.8)	240/240 (100%; 95% CI 98.4–100.0)
First-audit disclosure accuracy	226/238 (95.0%; 95% CI 91.4–97.1)	230/237 (97.0%; 95% CI 94.0–98.6)	230/235 (97.9%; 95% CI 95.1–99.1)	240/240 (100%; 95% CI 98.4–100.0)

5. AUDIENCE DISCLOSURE AND ETHICAL COMPLIANCE

Audience participants initially rated works without condition labels. Following 1:1:1 randomization, 240 participants received human-made labels, 240 AI-assisted labels, and 240 no label; each rated 20 works. Relative to no label, AI-assisted disclosure reduced authenticity by 0.32 SD (95% CI −0.40 to −0.24; p<.001), willingness to purchase by 0.21 SD (−0.29 to −0.13; p<.001), and perceived creativity by 0.10 SD (−0.18 to −0.02; p=.015), while the liking effect was −0.07 SD (−0.15 to 0.01; p=.082). The authenticity effect was 0.14 SD larger for works produced with generation/editing than for human-produced works (−0.24 to −0.04). Strongly negative baseline attitudes toward AI art amplified the label effect by 0.18 SD (−0.27 to −0.09). The manipulation altered audience reception but did not explain all differences between expert and audience judgments.

All AI-condition works received machine-readable provenance manifests. Ninety-seven percent of final disclosures matched captured workflow records on first review. Discrepancies reflected forgotten inpainting or imported AI-derived textures and were corrected before rating. Automated and manual screening flagged 4.2% of AI-assisted works for prohibited living-artist style references, identifiable faces, or copyrighted source material. Flagged content was replaced before submission. No privacy breach or unresolved authorship dispute occurred.

6. EXPLORATORY PROCESSES AND MODERATORS

Exploratory process models linked exploration breadth with part of the ideation–originality association (indirect estimate +0.08, 95% CI 0.03–0.14) and creative self-efficacy with +0.04 (0.01–0.09). Faster rendering and alternative generation were associated with the production effect on technical execution, while reduced perceived control was associated with lower authorship. Joint inclusion of process variables reduced the production coefficient for technical execution by 21% and for authorship by 27%. Estimates changed after adjustment for prior AI experience and prompt iteration and were interpreted as non-unique, model-dependent associations rather than confirmed causal mediation. Complete moderator estimates are provided in Supplementary Table S8.

7. QUALITATIVE FINDINGS AND MIXED-METHODS INTEGRATION

Framework Analysis produced five themes. First, AI ideation served as a divergence resource, allowing artists to reject, combine, or reinterpret suggestions. A professional illustrator described the system as “a wall of possible directions” while emphasizing that decisions about meaning remained human. Second, AI production shifted labor from rendering toward comparison, curation, defect detection, and revision. Third, agency depended on reversible control: participants retained ownership when they could mask, redraw, reject, and document consequential edits. Fourth, professionals and students framed risks differently: professional artists emphasized reputation, style dilution, and client disclosure, while students emphasized access to iteration and reduced execution barriers. Fifth, provenance became a practical creative boundary by making generated, inpainted, and manually edited regions more traceable.

Negative and contrasting cases qualified the positive themes. Some artists abandoned technically strong outputs because they redirected concepts toward generic cinematic conventions. Others described selection overload, fixation on early suggestions, culturally stereotyped motifs, or discomfort with living-artist style references. At three months, 74% had continued using generative tools, but reported use was selective rather than unconditional. Trajectory-linked analysis showed that participants who retained high agency generally

used AI for references, partial editing, or reversible variation, whereas those with sustained agency losses described accepting highly resolved outputs before the concept had stabilized.

Table 6. Mixed-methods joint display separating experimental effects, computational evidence, participant explanation, and interpretive limits.

Integrated finding	Quantitative / computational evidence	Qualitative explanation	Bounded interpretation
Ideation expanded creative search	Originality +0.34; CLIP exploration +0.24 SD	Artists treated AI as a provisional field of directions and rejected most outputs	Divergent benefit depended on active human selection
Production improved polish and speed	Technical execution +0.61; time −31%	Effort shifted from rendering to curation, defect detection, and revision	Efficiency gains did not imply stronger authorship
Full co-creation lacked general synergy	Originality interaction −0.20; lower between-artist distance	Highly resolved outputs constrained later conceptual change	More assistance was not automatically more original
Disclosure changed audience reception	Authenticity −0.32 SD after AI-assisted label	Viewers associated AI with reduced effort and personal expression	Labels affected audience ratings but not every expert–audience difference
Professional status altered selected costs	Authorship interaction −0.19; technical interaction −0.22	Professionals' artists protected style and reputation; students valued access and iteration	Expertise altered the balance of benefit and cost; subgroup evidence remained exploratory

VII. DISCUSSION

The stage-specific design showed that AI assistance did not have a single effect on visual creativity. Ideation support was most clearly associated with originality, conceptual coherence, and brief fulfillment, whereas production support was most clearly associated with aesthetic quality, technical execution, and speed. Effect sizes were small to moderate (g=.13–.43), which is consistent with recent experimental evidence showing that generative assistance can improve selected dimensions of individual output without uniformly improving creativity [8]–[10]. This distinction separates expanding the problem space from automating image realization and explains why full co-creation achieved high technical quality and rapid completion without a positive interaction for originality.

The ideation effects align with experiments describing generative systems as divergent prompts and interactive creative partners [9], [11]. Text-to-image research similarly emphasizes iterative prompting, selection, and curation [12]. Participants benefited when suggestions remained provisional and easy to reject. Production effects were larger for technical execution and time, but agency, ownership, and authorship declined. Product quality and creative experience therefore moved in different directions. This pattern supports treating authorship and control as substantive outcomes rather than attitudes inferred from aesthetic scores.

The computational results sharpened the distinction between individual and collective diversity. AI ideation broadened artists' option sets, while production assistance reduced distance across artists. Similarity measures do not establish cultural homogenization, but convergence across CLIP, LPIPS, DINOv2, palette and layout descriptors, qualitative reports of generic style, and the AI-only benchmark strengthens the interpretation of feature-space convergence. This finding resembles evidence that generative assistance can raise individual performance while reducing collective diversity [9].

Audience disclosure findings also require a bounded interpretation. Randomized AI-assisted labels reduced authenticity and purchase intention, supporting earlier evidence that origin and disclosure influence reception [5]–[7]. However, expert and audience samples differed in expertise, task, and context. The label manipulation therefore identified a social-evaluation effect rather than fully explaining expert–audience disagreement.

The framework was best supported at the level of stage-specific outcomes. Distributed cognition explained why generated representations changed the search space [24]. The componential model clarified why expertise and task complexity altered the usefulness of options [25]. Self-determination theory explained why technically improved outputs could coincide with lower control [26]. Process models and interview accounts refined this explanation but did not establish unique causal mediation because exploration, control, and early quality change were not randomized.

Methodological strengths included within-artist crossover control, counterbalanced conditions and briefs, masked expert rating, explicit artwork-level dependence, multiple judges per artwork, workflow telemetry, computational robustness analysis, randomized audience disclosure, provenance capture, and longitudinal qualitative follow-up. Practically, the findings favor stage-sensitive governance over universal acceptance or rejection of generative AI. Ideation tools may support exploration when outputs remain provisional, while production tools require stronger safeguards for human revision, disclosure, copyright, and control. Controlled co-creation may be appropriate where editorial control, transparent provenance, and accountable human transformation are maintained.

VIII. STRENGTHS AND LIMITATIONS

The study combined a preregistered factorial crossover with professional and student samples, multiple masked judges, workflow telemetry, computational image analysis, an audience-label experiment, provenance auditing, and explanatory qualitative follow-up. The artwork–artist–judge model represented the actual rating structure, and sensitivity analyses addressed sequence, period, carryover, brief mismatch, protocol deviation, and alternative similarity metrics.

Participants could not be blinded, and demand characteristics may have influenced self-reported agency, ownership, and satisfaction. The locked Stable Diffusion XL deployment improved reproducibility but limits generalization to other models, interfaces, and rapidly changing tool ecosystems. A seven-day interval reduced immediate contamination but could not erase acquired prompting, curation, error-detection, or expectation effects. The study therefore interpreted the main estimands as average stage-specific effects under repeated exposure and triangulated them with previous-condition, period, cumulative-exposure, first-period-only, complete-sequence, and period-exclusion analyses. Equivalent briefs were pilot-matched rather than identical, and the study was powered for main effects rather than small interactions, subgroup effects, or definitive mediation.

We adapted several agency-related measures to the task. PAAS-4 and ARTI-4 showed acceptable internal consistency, factor separation, and item-level sensitivity, but remained study-specific exploratory measures because no established instrument specifically measures AI-assisted visual production. Therefore, we did not include their findings in the confirmatory multiplicity family and recommend independent replication. CLIP, LPIPS, DINOv2, palette, and layout metrics quantify selected semantic and perceptual similarities but cannot determine cultural originality, artistic significance, ethical legitimacy, or style appropriation. The AI-only benchmark was not randomized, and the audience experiment did not replicate real art-market transactions. The digitally competent sample and three-month follow-up limit transfer to beginners, traditional-media practitioners, other cultural settings, and long-term artistic development.

IX. CONCLUSION

AI-assisted ideation and AI-assisted production altered visual creative work through different stage-specific profiles. Ideation support produced modest gains in originality, conceptual coherence, brief fulfillment, exploration breadth, and creative self-efficacy. Production support produced larger gains in aesthetic quality, technical execution, and speed but reduced perceived agency, ownership, and authorship

and was associated with lower collective visual diversity. Full co-creation achieved high quality and rapid completion, yet the interaction evidence did not support a general claim of synergy. Audience disclosure affected authenticity and purchase judgments, while qualitative findings showed that benefits depended on reversible control, critical rejection, task fit, and provenance. The results support context-dependent, stage-sensitive use of generative systems, with humans retaining responsibility for intention, selection, transformation, cultural meaning, disclosure, and accountability. These findings indicate that the value of generative AI in visual arts depends less on the amount of assistance provided than on where and how that assistance enters the creative process. Effective practice therefore requires transparent disclosure, documented provenance, opportunities for substantial human revision, and training that strengthens prompt literacy without weakening independent artistic judgment.

X. CONSENT FOR PUBLICATION AND SUPPLEMENTARY MATERIAL.

Artwork publication required separate written permission. We included no directly identifying personal information. A separate supplementary document accompanies this study and contains the objective–analysis map, measurement matrix, multiplicity plan, pilot brief-matching results, crossover diagnostics, complete psychometric outputs, computational robustness analyses, moderator and audience-label models, qualitative audit materials, provenance checklist, and supplementary figures.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1]. A. Hertzmann, “Can computers create art?,” Arts, vol. 7, no. 2, art. 18, 2018, doi: 10.3390/arts7020018.

[2]. A. Oksanen, A. Cvetkovic, N. Akin, R. Latikka, J. Bergdahl, Y. Chen, and N. Savela, “Artificial intelligence in fine arts: A systematic review of empirical research,” Computers in Human Behavior: Artificial Humans, vol. 1, no. 2, art. 100004, 2023, doi: 10.1016/j.chbah.2023.100004.

[3]. J. W. Hong and N. M. Curran, “Artificial intelligence, artists, and art: Attitudes toward artwork produced by humans vs. artificial intelligence,” ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 15, no. 2, art. 58, 2019, doi: 10.1145/3326337.

[4]. Z. Epstein, S. Levine, D. G. Rand, and I. Rahwan, “Who gets credit for AI-generated art?,” iScience, vol. 23, no. 9, art. 101515, 2020, doi: 10.1016/j.isci.2020.101515.

[5]. L. Bellaiche, R. Shahi, M. H. Turpin, et al., “Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork,” Cognitive Research: Principles and Implications, vol. 8, art. 42, 2023, doi: 10.1186/s41235-023-00499-6.

[6]. C. B. Horton Jr., M. W. White, and S. S. Iyengar, “Bias against AI art can enhance perceptions of human creativity,” Scientific Reports, vol. 13, art. 19001, 2023, doi: 10.1038/s41598-023-45202-3.

[7]. U. Messer, “Co-creating art with generative artificial intelligence: Implications for artworks and artists,” Computers in Human Behavior: Artificial Humans, vol. 2, art. 100056, 2024, doi: 10.1016/j.chbah.2024.100056.

[8]. E. Zhou and D. Lee, “Generative artificial intelligence, human creativity, and art,” PNAS Nexus, vol. 3, no. 3, art. pgae052, 2024, doi: 10.1093/pnasnexus/pgae052.

[9]. A. R. Doshi and O. P. Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," Science Advances, vol. 10, no. 28, art. eadn5290, 2024, doi: 10.1126/sciadv.adn5290.

[10]. J. McGuire, D. De Cremer, and T. Van de Cruys, "Establishing the importance of co-creation and self-efficacy in creative collaboration with artificial intelligence," Scientific Reports, vol. 14, art. 18525, 2024, doi: 10.1038/s41598-024-69423-2.

[11]. P. Karimi, J. Rezwana, S. Siddiqui, M. L. Maher, and N. Dehbozorgi, "Creative sketching partner: An analysis of human–AI co-creativity," in Proc. 25th Int. Conf. Intelligent User Interfaces, Cagliari, Italy, 2020, pp. 221–230, doi: 10.1145/3377325.3377522.

[12]. J. Oppenlaender, "The creativity of text-to-image generation," in Proc. 25th Int. Academic Mindtrek Conf., Tampere, Finland, 2022, pp. 192–202, doi: 10.1145/3569219.3569352.

[13]. J. McCormack, T. Gifford, and P. Hutchings, "Autonomy, authenticity, authorship and intention in computer generated art," in Computational Intelligence in Music, Sound, Art and Design, LNCS 11453, Cham, Switzerland: Springer, 2019, pp. 35–50, doi: 10.1007/978-3-030-16667-0_3.

[14]. T. M. Amabile, "Social psychology of creativity: A consensual assessment technique," Journal of Personality and Social Psychology, vol. 43, no. 5, pp. 997–1013, 1982, doi: 10.1037/0022-3514.43.5.997.

[15]. P. Tierney and S. M. Farmer, "Creative self-efficacy: Its potential antecedents and relationship to creative performance," Academy of Management Journal, vol. 45, no. 6, pp. 1137–1148, 2002, doi: 10.2307/3069429.

[16]. L. Van Dyne and J. L. Pierce, "Psychological ownership and feelings of possession: Three field studies predicting employee attitudes and organizational citizenship behavior," Journal of Organizational Behavior, vol. 25, no. 4, pp. 439–459, 2004, doi: 10.1002/job.249.

[17]. A. Tapal, E. Oren, R. Dar, and B. Eitam, "The Sense of Agency Scale: A measure of consciously perceived control over one's mind, body, and the immediate environment," Frontiers in Psychology, vol. 8, art. 1552, 2017, doi: 10.3389/fpsyg.2017.01552.

[18]. S. G. Hart, "NASA-Task Load Index (NASA-TLX): 20 years later," Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, no. 9, pp. 904–908, 2006, doi: 10.1177/154193120605000909.

[19]. A.-R. Blais and E. U. Weber, "A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations," Judgment and Decision Making, vol. 1, no. 1, pp. 33–47, 2006, doi: 10.1017/S1930297500000334.

[20]. A. Radford, J. W. Kim, C. Hallacy, et al., "Learning transferable visual models from natural language supervision," in Proc. 38th Int. Conf. Machine Learning, vol. 139, pp. 8748–8763, 2021, doi: 10.48550/arXiv.2103.00020.

[21]. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2018, pp. 586–595, doi: 10.1109/CVPR.2018.00068.

[22]. M. Oquab, T. Darcet, T. Moutakanni, et al., "DINOv2: Learning robust visual features without supervision," arXiv preprint arXiv:2304.07193, 2023, doi: 10.48550/arXiv.2304.07193.

[23]. Coalition for Content Provenance and Authenticity, "Content Credentials: C2PA Technical Specification," Version 2.1, Sep. 2024. [Online]. Available: https://spec.c2pa.org/specifications/specifications/2.1/specs/C2PA_Specification.html. DOI not assigned by the standards body.

[24]. J. Hollan, E. Hutchins, and D. Kirsh, "Distributed cognition: Toward a new foundation for human–computer interaction research," ACM Transactions on Computer-Human Interaction, vol. 7, no. 2, pp. 174–196, 2000, doi: 10.1145/353485.353487.

[25]. T. M. Amabile and M. G. Pratt, "The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning," Research in Organizational Behavior, vol. 36, pp. 157–183, 2016, doi: 10.1016/j.riob.2016.10.001.

[26]. R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions," Contemporary Educational Psychology, vol. 61, art. 101860, 2020, doi: 10.1016/j.cedpsych.2020.101860.

[27]. K. Dwan, T. Li, D. G. Altman, and D. Elbourne, "CONSORT 2010 statement: Extension to randomized crossover trials," BMJ, vol. 366, art. l4378, 2019, doi: 10.1136/bmj.l4378.

[28]. R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," Journal of Memory and Language, vol. 59, no. 4, pp. 390–412, 2008, doi: 10.1016/j.jml.2007.12.005.

[29]. N. K. Gale, G. Heath, E. Cameron, S. Rashid, and S. Redwood, "Using the Framework Method for the analysis of qualitative data in multi-disciplinary health research," BMC Medical Research Methodology, vol. 13, art. 117, 2013, doi: 10.1186/1471-2288-13-117.

[30]. T. C. Guetterman, M. D. Fetters, and J. W. Creswell, "Integrating quantitative and qualitative results in health science mixed methods research through joint displays," Annals of Family Medicine, vol. 13, no. 6, pp. 554–561, 2015, doi: 10.1370/afm.1865.