Creativity Rivalry: Human, Artificial Intelligence, and Co-Design

Hoa Vo
Eskenazi School of Art, Architecture + Design
Indiana University
Bloomington, IN 47405 USA
hoavo@iu.edu

Abstract

This study explores the creativity rivalry between AI, human designers, and co-design efforts in the 2024 Robert Bruce Thompson Lighting Design Competition, which tasked participants with designing a pediatric waiting room light fixture. A total of 120 Amazon Mechanical Turk (MTurk) participants evaluated design solutions across three stages: sketches (S), three-dimensional renderings (3D), and virtual room models (VR). Using the Creative Product Semantic Scale (CPSS), they assessed novelty (originality and surprise), resolution (logic and utility), and style (craftsmanship and elegance). Results showed that AI excelled in style at the 3D stage, while the human designer led in novelty during S and VR stages. Unexpectedly, co-design underperformed across criteria. These findings challenge assumptions about AI's creative limitations and highlight its strengths in visualization and refinement. The study offers practical insights for designers and educators on strategically integrating AI while maintaining human originality in the creative process.

Introduction

Is creativity becoming a battleground between humans and Artificial Intelligence (AI)? As AI advances beyond automating repetitive tasks into the realm of creative production, it has ignited legal and ethical debates (Rattay et al., 2022). The five-month Writers Guild of America strike[1] successfully defended the rights of human scriptwriters, yet artists and illustrators contesting the appropriation of their styles by generative AI platforms saw their claims dismissed[2]. Meanwhile, AI-generated artwork has won prestigious competitions, such as the Colorado State Fair[3], and real estate AI platforms are now challenging the expertise of interior designers (qbiq, 2024; TestFit, 2024).

These tensions raise pressing questions: "Is human creativity under threat, or can AI and humans coexist as creative collaborators?" The classic Turing Test (1950) suggests that AI's intelligence is determined by whether a human evaluator can distinguish its responses from those of a person. Similarly, in design, AI's creative ability is assessed by how closely its outputs align with human standards. This study explores the *creativity rivalry* between humans, AI, and co-design by empirically evaluating their design solutions in a real-world competition. Through this lens, we examine whether AI is a true creative rival or a transformative partner in design.

Aim and Significance

This study aims to empirically examine differences in creativity among design solutions generated by AI, human-AI co-design, and a human designer. To achieve this, three hypotheses were tested within the context of the Robert Bruce Thompson Lighting Design Competition[4]:

- **H1:** AI-generated designs will exhibit lower creativity than those of a human designer.
- **H2:** Co-design between humans and AI will produce more creative outcomes than AI alone.
- **H3:** Co-design between humans and AI will be more creative than human-designed solutions alone.

This research advances AI studies in design by employing a rigorous creativity assessment framework, using the well-established Creative Product Semantic Scale or CPSS (Besemer, 2006) to ensure objective evaluation. To mitigate bias, blind conditions concealed the origin of each design, preventing inherent assumptions that AI-generated work is less creative than those of human-designed (Horton Jr et al., 2023). By providing empirical insights, the study also offers practical guidance for designers and educators

[1] Merchant, B. (2023). The writers' strike was the first workplace battle between humans and AI. The humans won. *Los Angeles Times*. Retrieved from https://www.latimes.com/projects/writers-strike-deal-ai/

[2] Cho, W. (2023). Artists lose first round of copyright infringement case against AI art generators. *The Hollywood Reporter*. Retrieved from https://www.hollywoodreporter.com/business/business-news/artists-copyright-infringement-case-ai-art-generators-1235632929/

[3] Roose, K. (2022). AI-generated art won a prize. Artists aren't happy. *The New York Times*. Retrieved from https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

[4] https://rbtcompetition.org/

on strategically integrating AI into design. Specifically, it highlights how AI can break fixation patterns, introduce novel perspectives, and optimize creativity at key stages of the design process.

Method

This study employed a multi-stage design process to compare the creativity of AI-generated, human-designed, and co-designed solutions in the context of the 2024 Robert Bruce Thompson Lighting Design Competition (Figure 1). This annual competition presents a unique theme each year, requiring creativity within defined parameters. In 2024, the challenge was to design a mounted light fixture for a pediatric waiting room that is both comfortable and engaging for children. This theme introduced an additional layer of complexity, as the designs needed to balance functional, aesthetic, and emotional considerations (Norman, 2005). Specifically, the competition aligns with established creativity assessments, like Besemer's CPSS (2006) with three criteria: *novelty* (freshness of ideas), *style* (completeness via visual representations), and *resolution* (functions). In addition, *resolution* expands to emotional responses that a design can evoke (Horn & Salvendy, 2009), in this case, comfortable and engaging as the brief required.

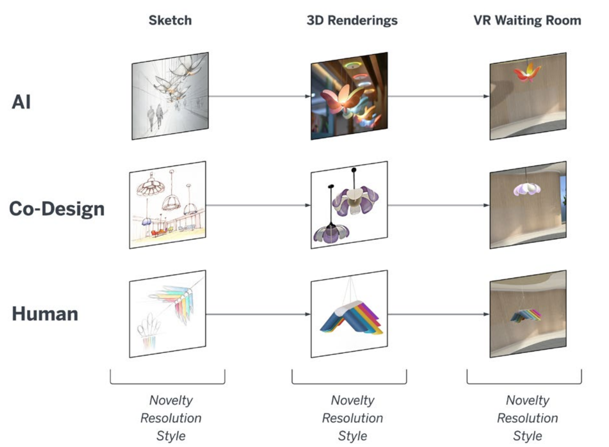

Figure 1. Three conditions and three stages of the design process measured on CPSS criteria for creativity.

Design assessments occurred at three stages: ideation (2D sketches), development (3D renderings), and final solution (VR models in a virtual waiting room). The use of VR enhanced spatial exploration and deepened design understanding, supporting emotional response assessments (Portman et al., 2015). The study was conducted under the Institutional Review Board (IRB) approval #22214.

Participants

A total of 120 Amazon Mechanical Turk (MTurk) workers participated, meeting inclusion criteria of being U.S.-based, aged 18 or older, and holding Master Qualifications to ensure high-quality responses. Exclusion criteria prevented multiple attempts and prior participation in similar studies. Blinded conditions concealed design origins to mitigate bias against AI-generated work. Participants reviewed 2D sketches, 3D renderings, and VR models and completed a Qualtrics questionnaire measuring creativity perceptions and emotional responses.

Design Conditions

Three conditions were tested:
• **AI Condition:** The author used ChatGPT-4, Midjourney, and Kaedim AI to generate design solutions. ChatGPT-4 produced 73 text prompts, which Midjourney used to generate 292 concept sketches. The author screened and refined these into 19 viable sketches, which Midjourney further developed into 3D renderings. A selected rendering was processed through Kaedim AI for final 3D modeling, requiring five hours of computation. The author's role was limited to screening for common or erroneous designs.
• **Co-Design Condition:** A junior interior design student worked on the competition as a class project. The student used Midjourney AI for initial ideas, enhanced outputs through hand sketches and manual refinements, and reimported them into Midjourney for refinement. The final three designs were developed in 3Ds Max, with one selected as the final model. This process spanned five weeks, ensuring active human-AI collaboration while allowing student agency in prompting, selecting, and refining the design.
• **Human Condition:** A freelance designer from Upwork with proficiency in design heuristics and CAD tools—but without design expertise in lighting—created the design solutions independently. Given the absence of domain-specific expertise, the freelancer did not possess a clear advantage over the junior interior design student, allowing for a meaningful and balanced comparison across conditions. Over one month, the designer submitted a 2D sketch, a 3D rendering, and a final model, responding freely to the competition brief. The author monitored deliverables to maintain experimental integrity but did not intervene in the creative process.

Data Collection

Participants completed a Qualtrics questionnaire after providing informed consent. The survey included demographic questions (e.g., age, occupation, education) and commitment and attention check to ensure thoughtful responses and engagement with the virtual waiting room. To assess creativity, the study employed the CPSS (Besemer, 2006; Wei et al., 2015), which evaluated novelty, resolution, and style (see Table 1) across three design stages: 2D sketches, 3D renderings, and VR models.

To measure emotional responses—an extension of *resolution*, the third criterion of creativity specified in CPSS—participants rated their feelings using two Self-Assessment Manikin (SAM) items (Bradley & Lang, 1994), capturing calmness vs. excitement and unhappiness vs. happiness. Participants reviewed 2D sketches and 3D renderings directly within the questionnaire, while the VR models were accessed via an external link, allowing interactive exploration before completing the survey.

Criterion	Binary
Novelty	Over Used – Fresh
	Predictable – Novel
	Usual – Unusual
	Ordinary – Unique
	Conventional – Original
Resolution	Worthless – Valuable
	Unimportant – Important
	Insignificant – Significant
	Inessential – Essential
	Illogical – Logical
Style	Disordered – Ordered
	Disarranged – Arranged
	Formless – Formed
	Awkward – Graceful
	Plain – Ornate

Table 1. CPSS with sample binaries for each criterion. For a full list of binaries, see Besemer (2006) and Wei et al. (2015).

Data Analysis

Descriptive statistics (mean—M, standard deviation—SD, and percentage—%) summarized demographics, self-reported emotions, and VR effectiveness. The author conducted one-way ANOVA to compare CPSS ratings across conditions (AI, Co-Design, and Human) for novelty, resolution, and style. Shapiro-Wilk and Levene's tests checked normality and variance assumptions. When violated, Kruskal-Wallis tests served as nonparametric alternatives (Lock et al., 2013; Yazici & Yolacan, 2007). Statistical analyses were performed in RStudio Desktop (RStudioTeam, 2023), with significance set at $p < 0.05$.

Results

The overall sample ($n = 120$) showed similar age distributions across conditions but exhibited some demographic disparities. Notably, the AI condition had a male-dominant gender distribution (80% male, 20% female) and a high concentration of participants aged 25–44 years old (82.5%).

A large proportion (35%) held management positions, followed by sales (17.5%), technical (17.5%), freelance (15%), and other roles (12.5%). In contrast, the Co-Design condition had a more balanced gender distribution (50% male, 45% female, 5% prefer not to disclose) with a slightly younger sample (67.5% aged 25–44 years old).

A significant portion (42.5%) worked in sales, followed by management (32.5%), freelance (12%), and other roles (13%). The Human condition also had a more even gender split (55% male, 45% female) with 75% aged 25–44 years old. Management (25%) and freelance (25%) were the most common occupations, alongside sales (17.5%), manual labor (10%), and other roles (22.5%).

Self-reported Emotions

Participants rated their emotions on a 9-point Likert scale, with 1 representing calm/unhappy, 9 representing excited/happy, and 5 as the neutral midpoint. Across all conditions, participants generally reported feeling happy, though their levels of calmness versus excitement varied.

• **AI Condition:** Participants reported a high level of happiness ($M = 6.90$, $SD = 1.77$). However, their calmness-excitement scores leaned toward neutrality ($M = 5.35$) with greater variability ($SD = 2.35$).

• **Co-Design Condition:** Participants reported high happiness ($M = 6.83$, $SD = 1.52$). They felt somewhat calmer ($M = 4.63$) but with wide individual differences ($SD = 2.36$).

• **Human Condition:** Happiness levels were consistent ($M = 6.83$, $SD = 1.57$). Participants also reported a calmer state ($M = 4.78$), though variability remained high ($SD = 2.38$).

Overall, happiness levels remained consistently high across conditions, while calmness versus excitement varied among individuals. These findings suggest that, regardless of design origin, participants responded positively to the light fixture designs, with emotional reactions influenced more by individual differences than by condition.

H1: AI vs. Human Creativity

Statistical analysis from the MTurk sample ($n = 120$, $N = 40$ per condition) provided no significant evidence to support the hypothesis that AI is less creative than a human designer. Due to violations of homogeneity of variances, particularly in the 2D sketch (S) and 3D rendering (3D) stages ($p < .001$), Kruskal-Wallis tests were used instead of one-way ANOVA.

In the S stage, a Kruskal-Wallis test for novelty indicated significant differences between conditions ($X^2 = 18.21$, $p < .001$), but post-hoc tests showed no statistical difference between AI and human novelty scores ($p = .06$). Resolution and style showed no significant differences across conditions ($p = .35$ and $p = .12$, respectively).

In the 3D stage, AI outperformed the human designer in style, with a significant Kruskal-Wallis test ($X^2 = 6.87$, $p = .03$). A follow-up Mann-Whitney U test ($p = .02$) confirmed that AI scored significantly higher than the human designer in style (see Figure 2) with a medium effect size ($d = .60$). Novelty and resolution differences were not significant ($p > .10$).

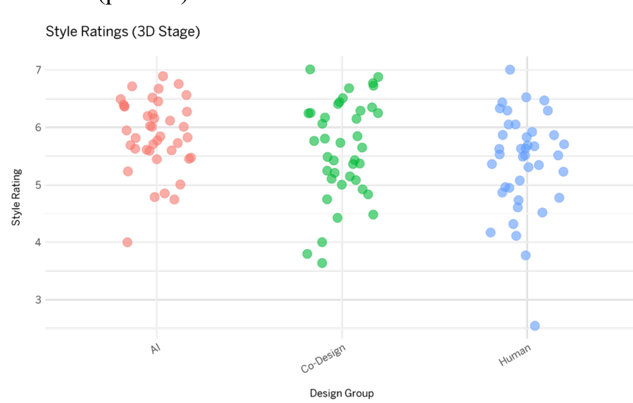

Figure 2. Mean-differences in style across three conditions.

In the VR stage, all Kruskal-Wallis tests were insignificant ($p > .10$), indicating no meaningful differences between conditions in the final evaluation phase.

Overall, these findings contradict the assumption that AI is less creative than a human designer. AI performed comparably to human designers across most criteria and even surpassed them in style during the 3D stage.

H2: Co-Design vs. AI

Statistical analysis rejected H_2 and provided no evidence that human-AI co-design was more creative than AI alone. Instead, findings partially contradicted expectations, as co-design underperformed in novelty during the ideation (S) stage.

In the S stage, a Kruskal-Wallis test for novelty was significant, and a follow-up Mann-Whitney U test with Bonferroni adjustment showed that AI novelty scores (AI_N_M) were significantly higher than co-design (Co_N_M) ($W = 19.25$, $p < .001$, $d = 1.09$), indicating a large effect size (see Figure 3). Resolution and style showed no significant differences across conditions.

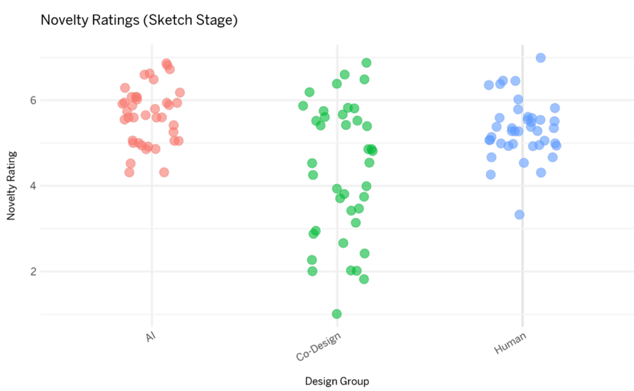

Figure 3. Mean-differences in novelty across three conditions.

In the 3D stage, a Kruskal-Wallis test for style was significant, but the follow-up test ($p = .6$) found no clear advantage for co-design over AI. All other CPSS ratings in 3D and VR stages were statistically insignificant, suggesting no meaningful differences between conditions.

Overall, co-design did not surpass AI in creativity and even scored lower in novelty during the ideation stage, rejecting the hypothesis that human-AI collaboration would yield more creative outcomes than AI alone.

H3: Co-Design vs. Human Designer

Statistical analysis rejected H_3 and provided no evidence that human-AI co-design was more creative than a human designer alone. Instead, findings partially contradicted expectations, as the human designer outperformed co-design efforts in novelty during the ideation (S) stage.

In the S stage, a Kruskal-Wallis test for novelty was significant, and a Mann-Whitney U test with Bonferroni adjustment showed that the human designer (Hu_N_M) scored significantly higher than co-design (Co_N_M) ($W = 1.34$, $p = .04$, $d = .8$), indicating a large effect size.

Across all other CPSS criteria and design stages (3D and VR), Kruskal-Wallis tests were insignificant, meaning there was no statistical advantage for co-design over human designers.

Overall, these results do not support the hypothesis that co-design is more creative than human effort alone. Instead, the human designer demonstrated stronger novelty in the ideation stage, further challenging assumptions about the benefits of AI-assisted collaboration in design.

Discussion

This study provides empirical insights into the evolving creativity rivalry between AI, human designers, and co-design efforts. Findings challenge the assumption that AI is inherently less creative than humans, revealing that AI-generated solutions performed comparably across multiple design criteria and even excelled in style during the 3D stage. Conversely, co-design efforts did not surpass AI or human designers, particularly struggling with novelty in the ideation stage. These results highlight AI's potential beyond simply generating variations—demonstrating its capability to rapidly visualize preliminary design concepts.

A key limitation of this study is the reliance on crowd-sourced CPSS ratings, without comparison to expert evaluations. Although non-parametric tests controlled for demographic disparities, the findings contrast with studies using expert judges (Zhou et al., 2023), raising questions about whether creativity assessments should integrate both crowd-sourced and expert perspectives. While prior research supports the convergence of large-scale crowd and expert ratings (Yuan et al., 2016; Foong et al., 2017), these discrepancies suggest that creativity evaluation methods must evolve—especially when assessing AI, co-design, and human-generated solutions.

Beyond methodological implications, this study underscores the uncertainty surrounding AI's creative capacity in design research. The results provide empirical support for AI's ability to independently generate novel outputs, challenging the perception that creativity is an exclusively human trait. More importantly, these findings call for a reconsideration of creativity itself: traditionally understood as a social construct (Amabile, 1983; Csikszentmihalyi, 1988) and assessed through established criteria (Horn & Salvendy, 2006; O'Quin & Besemer, 2006), creativity may now be shifting toward a hybrid concept—one that encompasses human-technology collaboration. As AI continues to shape creative disciplines, future research must explore how designers and educators can strategically integrate AI while preserving the originality and intentionality that define human creativity.

Conclusion

The insights from this research provide practical and pedagogical guidance for designers and educators seeking to integrate AI into the design process. Designers can leverage AI as a tool for rapid visualization, using its generative capabilities to explore diverse renderings and stylistic options in the early stages of concept development. However, they must recognize that AI's novelty may plateau, necessitating human intervention for creative breakthroughs in later stages.

For educators, the challenge lies in teaching students to critically engage with AI-generated outputs, ensuring that AI serves as a catalyst for creativity rather than a crutch for repetitive ideation. Encouraging a balance between AI-assisted and manual ideation can help designers avoid fixation and refine their originality. Ultimately, creativity is a dynamic and evolving construct, shaped by technological advancements and peer evaluations of their novelty and relevance. By offering empirical evidence, this study positions AI as both novel and appropriate within the design process and the broader design community, reinforcing its role as a transformative yet complementary creative tool.

Acknowledgements

This research was supported by the Eskenazi Technology Innovation Lab (ETIL) via the 2024 Eskenazi Technology Innovation Grant.

The author conceptualized, designed, and conducted the study, and wrote the manuscript. Chat GPT-4 was utilized for refining grammar and wording.

References

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357-376. https://doi.org/https://doi.org/10.1037/0022-3514.45.2.357

Besemer, S. P. (2006). *Creating products in the age of design: How to improve your new product ideas!* New Forums Press.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Csikszentmihalyi, M. (1988). The Systems Model of Creativity. https://doi.org/10.1007/978-94-017-9085-7

Horn, D., & Salvendy, G. (2006). Consumer-based assessment of product creativity: A review and reappraisal. *Human factors and ergonomics in manufacturing & service industries*, *16*(2), 155-175.

Horn, D., & Salvendy, G. (2009). Measuring consumer perception of product creativity: Impact on satisfaction and purchasability. *Human factors and ergonomics in manufacturing & service industries*, *19*(3), 223-240. https://doi.org/https://doi.org/10.1002/hfm.20150

Horton Jr, C. B., White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports*, *13*(1), 19001.

Lock, R. H., Lock, P. F., Morgan, K. L., Lock, E. F., & Lock, D. F. (2013). *Statistics: Unlocking the power of data*. Wiley ; John Wiley [distributor].

Norman, D. A. (2005). *Emotional design: why we love (or hate) everyday things*. Basic Books. https://research.ebsco.com/linkprocessor/plink?id=7d4460c6-a3f2-3454-99f4-364e6e98a923

O'Quin, K., & Besemer, S. P. (2006). Using the creative product semantic scale as a metric for results-oriented business. *Creativity and Innovation Management*, *15*(1), 34-44.

Portman, M. E., Natapov, A., & Fisher-Gewirtzman, D. (2015). To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. *Computers, environment and urban systems*, *54*, 376-384. https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2015.05.001

qbiq. (2024). *Generative AI for Real Estate Layout Design & Virtual Tours*. https://qbiq.ai/

Rattay, S., Yurrita, M., Garnham, I., & Browne, J. T. (2022). Prototyping tensions: How to talk to your colleagues about AI. In D. Lockton, S. Lenzi, P. Hekkert, A. Oak, J. Sádaba, & P. Lloyd (Eds.), *DRS2022: Bilbao*, 27 June - 3 July, Bilbao, Spain. Design Research Society. https://doi.org/10.21606/drs.2022.909.

RStudioTeam. (2023). *RStudio: Integrated Development for R*. In RStudio, PBC. https://posit.co/download/rstudio-desktop/

TestFit, I. (2024). *A New Real Estate Feasibility Workflow*. TestFit. Retrieved 06/10 from https://www.testfit.io/

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *49*, 433-460.

Wei, X., Weng, D., Liu, Y., & Wang, Y. (2015). Teaching based on augmented reality for a technical creative design course. *Computers & Education*, *81*, 221-234. https://doi.org/10.1016/j.compedu.2014.10.017

Yazici, B., & Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, *77*(2), 175-183. https://doi.org/ 10.1080/10629360600678310

Zhou, C., Zhang, X., & Yu, C. (2023). How does AI promote design iteration? The optimal time to integrate AI into the design process. *Journal of Engineering Design*, 1-28. https://doi.org/10.1080/09544828.2023.2290915