Issue 9 - AUGUST 2025 - X-Intelligence

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Creative Intelligence: Thinking Within, On, and Beyond the Box

Insights from Future of Innovation Lab x Lee Kuan Yew Centre for Innovative Cities

OPPORTUNITY

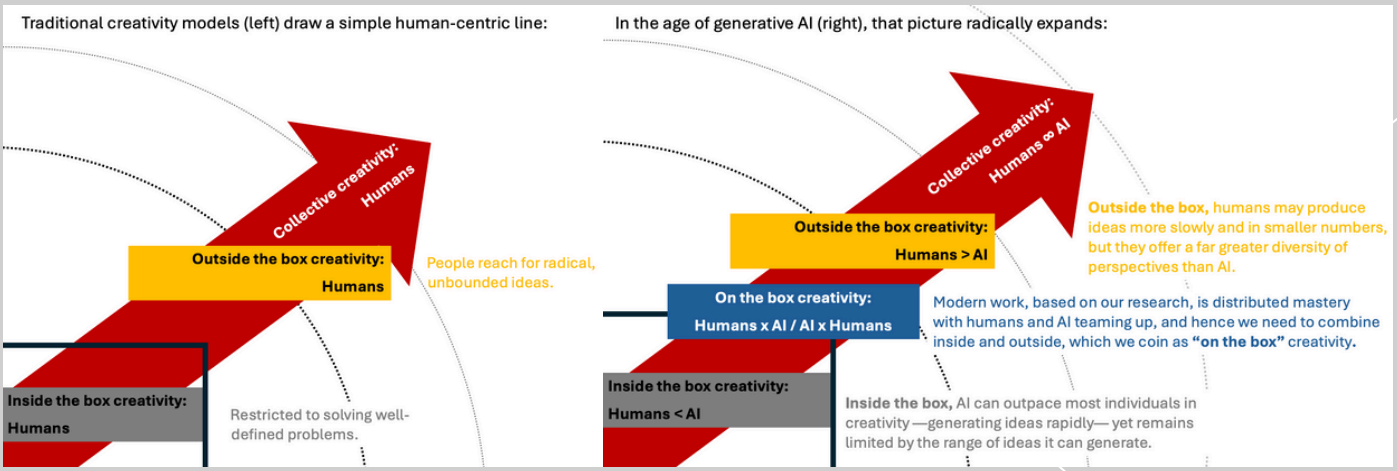
As AI's generative capabilities advance, leaders face a new frontier: AI systems are now judged capable of imaginative leaps once thought uniquely human. The challenge isn't just "Can AI create?" but "How can people and AI co-create to amplify innovation—and make us more creative together?".

NEED

To thrive, organizations must move beyond testing AI's solo creativity and instead build frameworks for true co-creation. That means defining where AI best support human designers, where structured collaboration boosts originality, and how to cultivate environments in which human creativity is enhanced—not replaced—by machine partners.

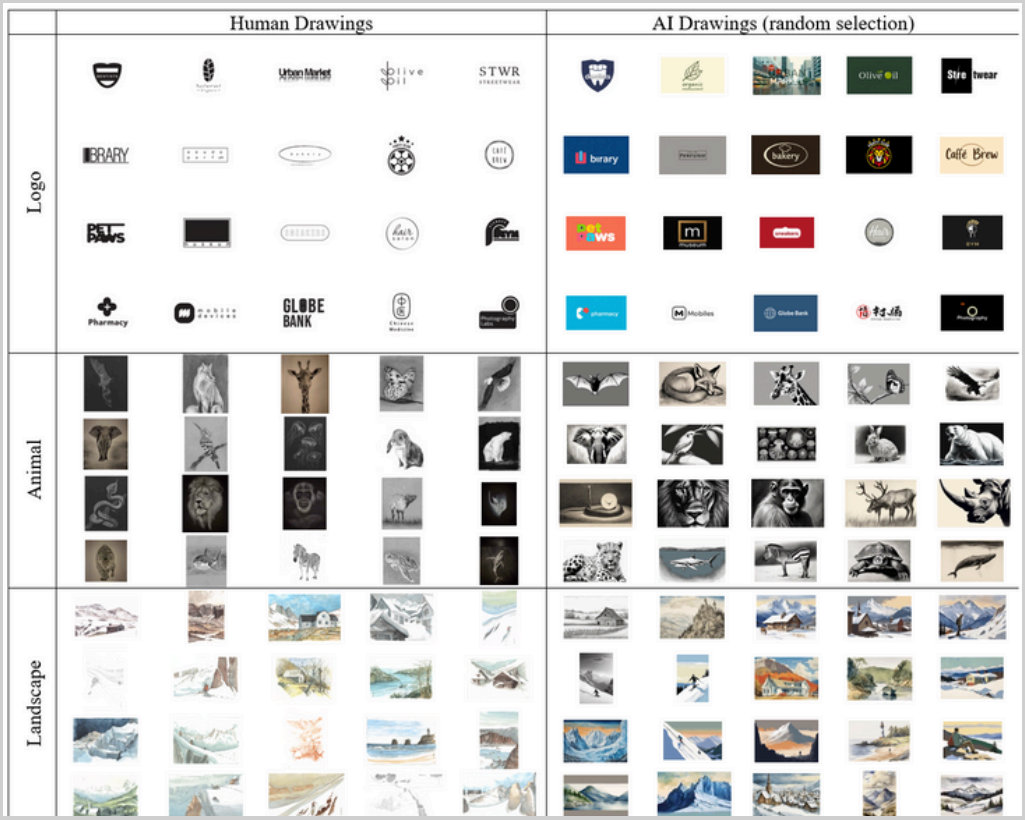
BIG TOPIC:

New Ways of Understanding Creativity



Together—through this three-tiered model of "inside," "on" and "outside" the box—we unlock collective creativity that soars far beyond the limits of human-only innovation.

An Economic Experiment on Creativity



Experimental design: We recruited 155 participants into 30 virtual lab sessions to rate the creativity of human artwork against AI reproductions (without authorship attribution) across three visual domains.

- **Logos:** minimalist black-and-white marks
- **Animals:** charcoal sketches of distinct species
- **Landscapes:** watercolor scenes of the Pyrenees

Human drawings: Three hobbyist artists each contributed 20 images to the three visual domains, resulting in a total of 60 human-generated drawings.

AI mirroring process: We used CLIP Interrogator for an unbiased translation of each artist's drawing into a concise text prompt. We then fed those text prompts into Stable Diffusion to produce AI-generated counterparts that replicate the original artwork style and composition.

Three-stage protocol:

- **Creativity of human vs AI artwork (Stage 1):** For each visual domain, participants compared 20 original artworks with 20 randomly selected AI-generated replicas and rated each image's creativity on a 0–10 scale. Random selection prevents cherry-picking only the best or worst variants, captures the model's average performance, and reduces participants' ability to anticipate quality. This way, Stage 1 assesses how people judge human and AI-generated artwork, on average.

- **Paying for human vs AI artwork (Stage 2):** Using an auction game, participants bid real money on their favorite images, revealing economic value.

- **Creativity of best AI artwork (Stage 3):** From the pool of hundreds of AI variants, our experimenters' hand-selected the most promising renditions of the original drawings (i.e., closest to the original human drawing)—yielding a "best-of" set of 40 AI images. In this stage, participants rated only those 40 experimenter-chosen AI-generated artworks on the same 0–10 creativity scale. This focused evaluation reveals whether the AI images deemed most creative by our researchers also resonate as highly creative to the participants.

01 Inside the Box Creativity - AI outperforms human artists and people are willing to pay for it!

What It Is: Inside the box, AI can outperform most individuals in creativity — generates ideas rapidly when fed with clear prompts that are well understood by the AI.

Why It Matters: AI's deep pattern-matching and vast training data make it unbeatable at inside the box creativity. It excels in detail, consistency, and elaboration.

Experiment Results: First, we compare the creativity scores given to 20 human sketches against 20 randomly sampled AI variants for each visual domain.

As figure 1 illustrates, AI outputs pulled ahead on average. Higher likelihood of high scores for AI drawings in the animal and watercolor visuals is due to AI's consistency. That is, random AI sample has fewer low-scoring images and more consistently "good" images, its group of scores is more concentrated in the upper-middle part of the scale.

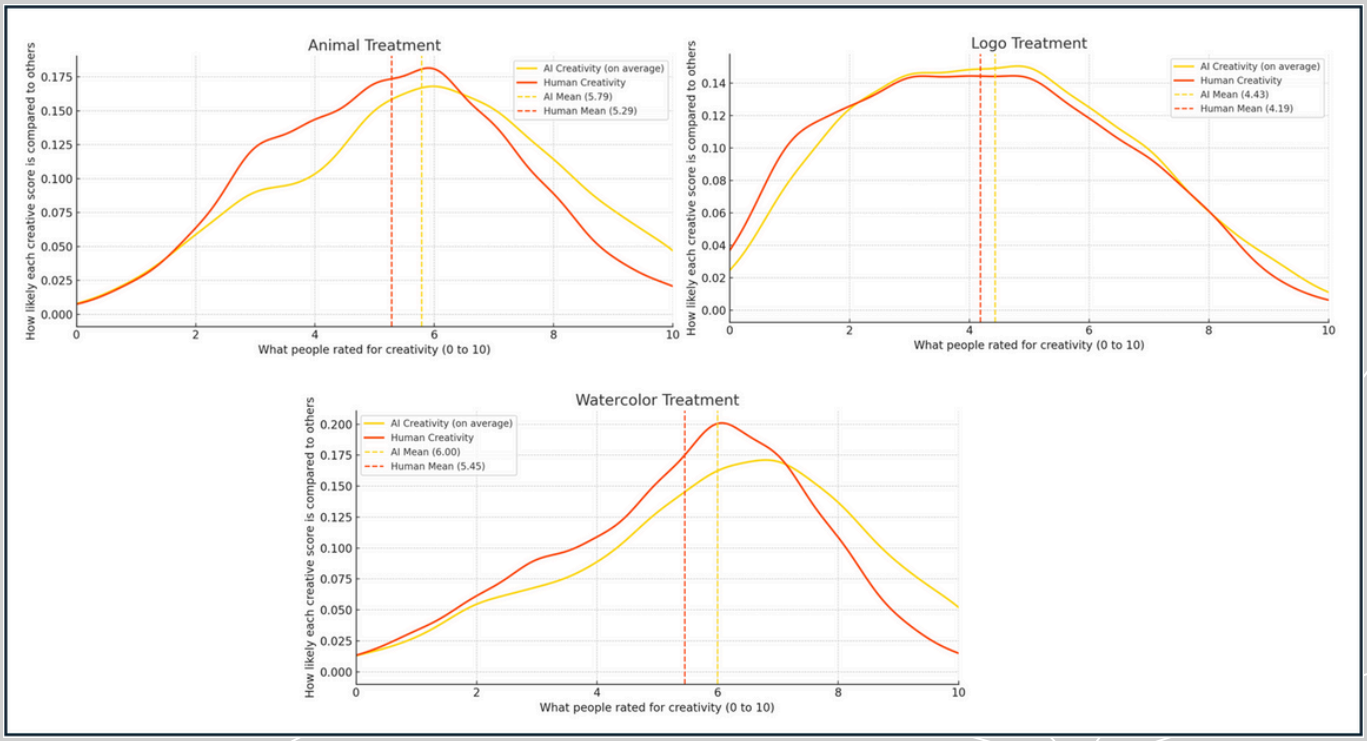
This makes the right side of its distribution curve appear "fatter" or stronger. Overall, a random AI sample wins on average because it can consistently generate many good-quality options by drawing on its vast training data.

Think of it this way! Imagine the 20 human drawings are a team of 20 amateur runners. They are competing against a team of 20 randomly selected professional runners, which represents the "random AI sample". The winning team is the one with the better average finish time for all 20 runners.

The Result: The professional AI team wins. Although the human team might have one or two very fast runners, its average time is brought down by its slower members. The professional team is highly consistent, so its average is better. This explains how the random AI sample pulls ahead on average for creativity.

The auction game reinforced the findings that AI images commanded higher willingness-to-pay (Prisse at al., 2025).

Fig.1: Distribution of Human vs Average AI creativity score



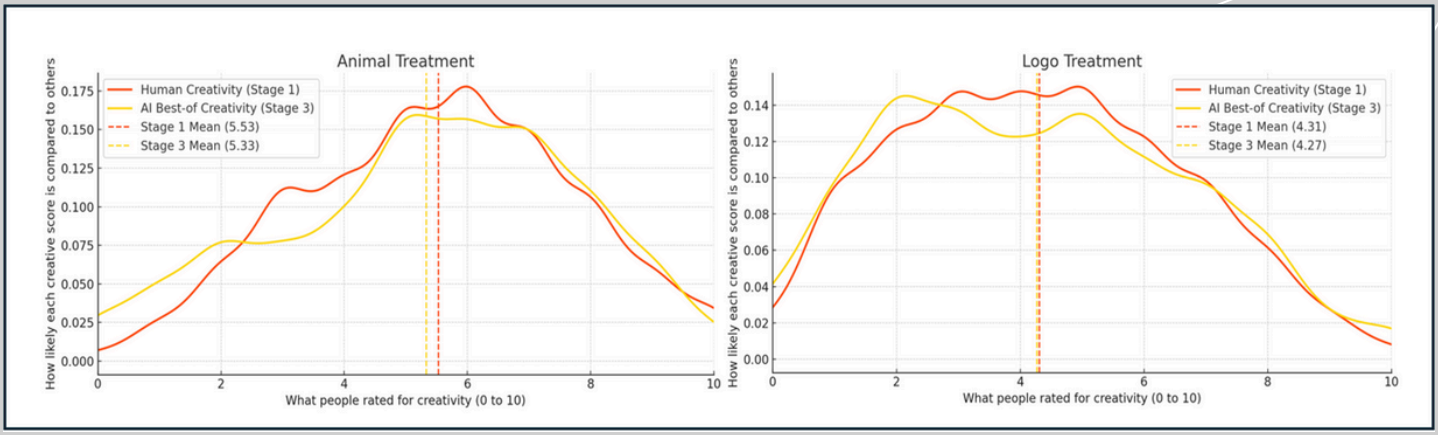
Next, we pit original human artworks in the animal and logo visual domains against the best AI-generated variants to test whether AI's top-performing images can outperform human creations in relatively well-trained domains like logos and animals. Interestingly, in the upper tail of the creativity distribution (see Figure 2), AI-generated drawings are perceived to be as creative as those produced by humans.

Now if we take the same analogy, the same 20 human runners face a new challenge. A race against the fastest, hand-picked champions from the entire professional league, representing the "best of AI". In the upper tail of the distribution, fastest human runner goes on a head-to-head race with the AI champion.

The race is incredibly close, and they are perceived to finish at the same time. This explains how, at the highest level, the human art was perceived to be "as creative as" the best AI art.

Overall, our findings suggests that AI can match or surpass human performance when working with familiar subjects, likely due to its extensive training on well-known topics. AI is able to deliver fast, consistent, and commercially valuable outputs—exemplifying its dominance inside the box.

Fig.2: Distribution of Human vs Best of AI creativity scores in Animal and Logo Visuals



Takeaway: When you feed AI a clear prompt within its training domain—be it a standard logo template or a detailed landscape scene—it will instantly draw on its vast data to recreate proven patterns and unearth novel combinations.

Because AI not only applies existing connections but also forges new links within familiar data, it often generates outputs that outpace human creativity. In practice, any task rooted in known content and its remix lives inside the box, freeing teams to tackle truly uncharted challenges.

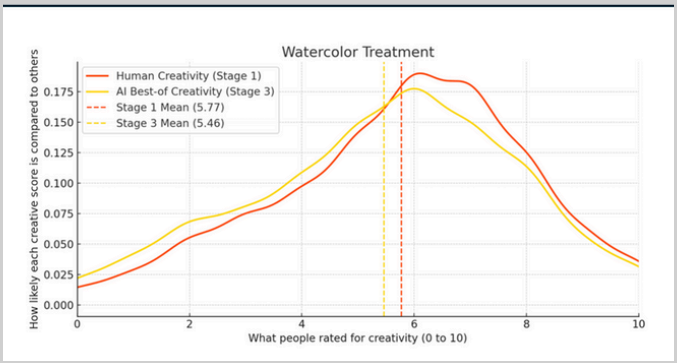
02 Outside the Box – Human artists outperform AI

What It Is: Fully open-ended creative challenges with no fixed template, no rulebook, and no historical examples to data or a prompt. These tasks demand leaps of imagination and diversity of ideas.

Why It Matters: Experiments show that while AI narrows the search space leaning on its training to produce ideas, the most unexpected and groundbreaking ideas still stem from human imagination (Koivisto and Grassini, 2023; Charness and Grieco, 2024) because humans produce ideas with far greater diversity than AI.

Experiment Results: Even in our experiment, when we pit original human artworks against the best AI variants in our most complex visual domain — watercolor scenes of the Pyrenees based on artist personal experience — a domain where AI has the least prior exposure— human artworks dominated the highest creativity scores. As the graph shows, a significant gap emerges in the mid-to-high creativity range (scores 6-8), where human art (red line) consistently outperforms AI art (yellow line). However, this performance gap narrows considerably at the highest scores (8 and above), where the strongest AI variants closely match the peak creativity of human artists.

Fig.3: Distribution of Human vs Best of AI creativity scores in Watercolor domain



We propose two primary explanations for this pattern:

- The "Lived Experience" Advantage: In the mid-to-high range, human artists draw on subtle, real-world experiences that AI has not yet mastered. This provides their work a nuance and depth that raters rewarded with higher scores. In contrast, the AI's outputs, while competent, were more concentrated around an average score, producing less variation in the upper range.

- Perceptual Diminishing Returns: The gap likely narrows at the top end due to the limitations of our non-expert raters. For a general audience, the perceived difference between a "very creative" artwork (score of 8) and a "groundbreaking" one (score of 10) becomes subtle and difficult to articulate.

  This "perceptual ceiling" makes it harder for exceptional human art to pull away, allowing the highest-performing AI to seemingly bridge the gap. That is, the AI may not have necessarily replicated the depth of human genius; it has simply created an output that is good enough to hit the limit of what a non-expert rater can perceive and articulate. Think of it like music quality: Most people can easily tell the difference between a low-quality MP3 and a high-quality one. But at the very top, comparing a high-bitrate MP3 to a perfect studio recording, it becomes incredibly difficult for most listeners to tell which is which – both just sound 'great'.

Therefore, further investigation using art experts as participants is essential. This would help determine if the narrowing gap at the highest scores is a true reflection of AI's capability or an artifact created by the non-expert raters' inability to discern subtle, expert-level details.

Takeaway: When a creative challenge requires drawing on new inputs and forging entirely new connections—humans still hold the edge. Their smaller, more deliberate idea sets span a far richer spectrum of perspectives and breakthroughs than any AI.

03 On the Box Creativity– Humans Teaming with AI Possibilities (Humans ∞ AI)

What It Is: Hybrid tasks that require both structured generation and divergent thinking.

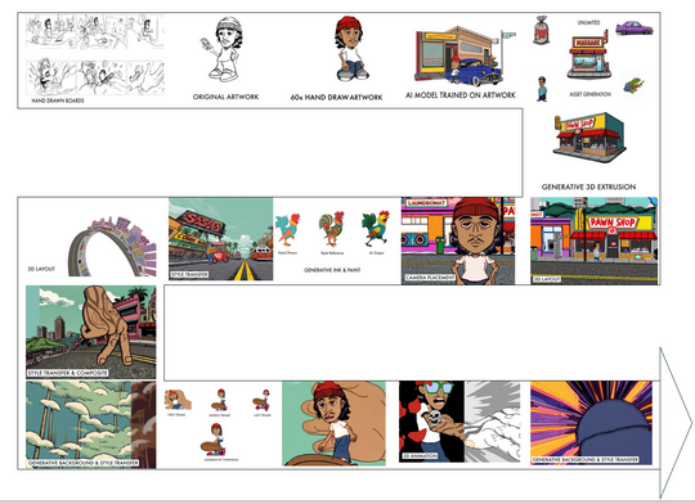
Why It Matters: When humans and AI collaborate, they combine AI's speed with human judgment about new data—such as contextual insights AI may lack. This a form of "distributed mastery" that fosters solutions neither could achieve alone.

An Internal Example from SUTD: Human ∞ AI Co-Creation in Rubric Grading
Huang et al. (2024) demonstrate a textbook example of "on the box" creativity through a human–AI teaming approach to grading open-ended design assignments. This study employs an iterative research approach in developing a Custom GPT with the aim of achieving more reliable results and testing whether it can provide design students with constructive feedback. At each stage, human experts refined the prompts, reviewed AI comments, and re-tuned the grading rubric instructions—creating a true feedback loop. This collaborative process pushed the model's inter-rater reliability (ICC) from 0.65 up to 0.78, levels that match expert human graders.

Why it is "on the box": The task is bounded by a fixed rubric and set of examples, yet requires nuanced judgment. Humans supply the criteria, curate examples, and fine-tune the model's outputs based on domain knowledge, while AI rapidly applies those rules at scale.

An External Example from Cuco Music Video: A love letter to Los Angeles is a music video directed by Paul Trillo and artist Paul Flores to bring the city's energy to life in a surreal journey that bends reality while remaining deeply rooted in Chicano culture. It is an example of "on the box" creativity—where a clear, step-by-step workflow guides the collaboration between human artists (from visual directors to 2D artists) and AI tools teaming together to create a masterpiece. That tightly-scoped, iterative process—scaffolded by human oversight and domain expertise—is exactly where human ∞ AI teams shines in "on the box" workflows.

Image credit: Cuco Music Video



This collaboration aligns with our finding that at the highest creativity levels, the performance gap between human and "best AI" artwork narrows significantly, suggesting that AI's ability to match high-level creative output makes it an invaluable partner for human creativity in these hybrid "on the box" tasks.

Takeaway: Use AI-human ensembles for tasks that remix fresh inputs with proven patterns. Have AI generate ideas fast, then let people curate, refine, and inject subtle judgment.

Closing Thought

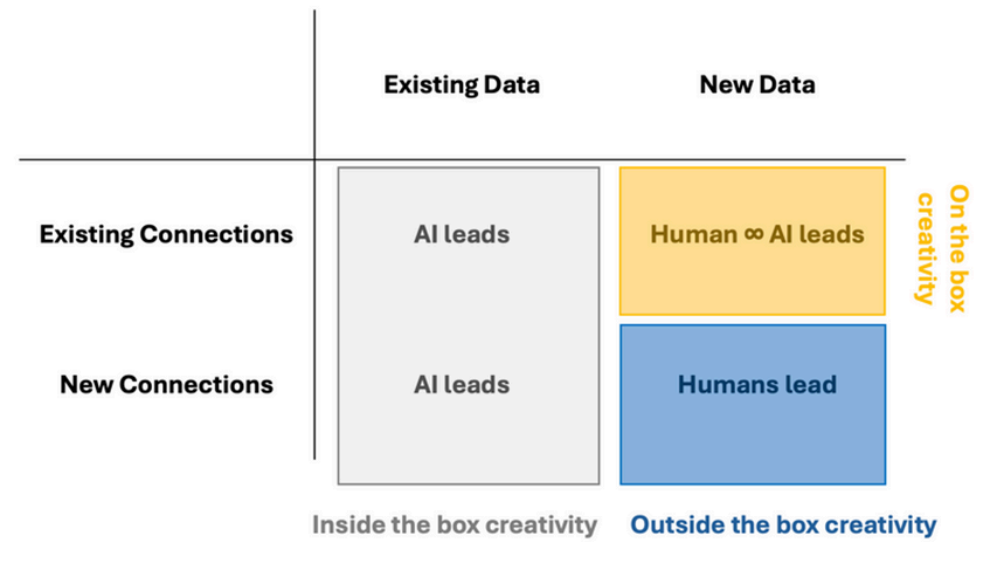
By mapping creative work into Inside the Box, On the Box, and Outside the Box, organizations can:

- **Deploy AI where it excels**—on tasks that rely on existing data to produce new/existing connections for ideation (Inside the Box)
- **Orchestrate human–AI teams** for semi-structured challenges that require relying on new data (On the Box)
- **Preserve pure human imagination** for unbounded, breakthrough ideas that require new data and new connections (Outside the Box)

This three-tiered, distributed-mastery framework will define the next generation of creative leadership to

- Equip teams to harness AI for structured workflows (Inside the Box)
- Train creators to orchestrate human ∞ AI ensembles for hybrid, On the Box co-creation.
- Cultivate human imagination through dedicated, machine-free ideation zones (Outside the Box)

Connect with us | sutd.edu.sg/future-of-innovation



References

Reference to study:
Prissé, Benjamin and Ho, Jun Quan and Deng, Ruotong and Jayasekara, Dinithi and Ouangraoua, Chris. (2025). The Creativity of Artificial Intelligence. Available at SSRN: https://ssrn.com/abstract=5284780

Reference to other related studies:
Huang, Qian and Willems, Thijs and Poon, King Wang. (2024). The application of GPT-4 in grading design university students' assignment and providing feedback: An exploratory study. Available at: https://doi.org/10.48550/arXiv.2409.17698

Charness, Gary and Grieco, Daniela, Creativity and AI (2024). Available at SSRN: https://ssrn.com/abstract=4686415 or http://dx.doi.org/10.2139/ssrn.4686415

Grassini, S. and Koivisto, M. (2024). Understanding how personality traits, experiences, and attitudes shape negative bias toward ai-generated artworks. Scientific Reports, 14(1):4113.

Collins, A. J. (2016). "A meta-analysis of high-resolution audio perceptual evaluation." Journal of the Audio Engineering Society, 64(7/8), 461-477.