



USING THE UNIQUENESS QUOTIENT IN THE ASSESSMENT OF CREATIVITY OF IMAGE GENERATION MODELS

Mirko Milošević¹,
[0009-0005-7065-4768]

Irena Ristić²,
[0000-0003-4949-3304]

Miloš Milošević^{3*}
[0000-0003-2592-451X]

¹Floating point,
Belgrade, Serbia

²The Faculty of Dramatic Arts,
the University of Arts in Belgrade,
Belgrade, Serbia

³Singidunum University,
Belgrade, Serbia

Abstract:

The rapid rise of Generative AI raises questions about its creative abilities, but most assessments focus on text or subjective visual reviews. This study uses the uniqueness quotient to compare the creativity of visuals generated by AI models (Intent-driven, Prompt-driven) and by humans (art and non-art students). Results show no significant difference in overall creativity between humans and AI. Subsample analysis finds Intent-driven models do not differ from art students in originality and outperform non-art students. Prompt-driven models show unique originality patterns. Distribution analysis shows asymmetry: AI models keep a high average, but peak innovativeness remains unique to humans. This study's theoretical contribution is establishing the uniqueness quotient as a reliable, objective method for evaluating visual creativity. In practice, context-aware, intent-driven models can be integrated as advanced co-creators in creative industries, and the uniqueness quotient can serve as a measure of future machine creativity.

Keywords:

Generative AI, Objective evaluation, Visual originality, Multimodal generative AI models, Human-AI comparison.

Correspondence:

Miloš Milošević

e-mail:

mmilosevic@singidunum.ac.rs

INTRODUCTION

In recent years, the rapid development of artificial intelligence (AI), particularly large language and generative models, has stimulated intensive scientific research into their creative capacities and comparison with human inventiveness. Previous testing of AI models has mainly relied on standardized psychometric instruments for assessing divergent and convergent thinking, such as the Alternate Uses Task, Divergent Association Task, and Remote Associates Test. These studies indicate that contemporary models frequently exceed average human ability in generating fluent and original ideas [1], [2], [3]. However, comprehensive comparative studies show a key asymmetry at the extremes of the distribution: although AI outperforms the human average, the most creative segment of the human population remains dominant, showing a capacity for truly exceptional associations and innovations that models cannot yet achieve [3], [4], [5].



Parallel to these basic textual tasks, the focus of current research is increasingly expanding into the domains of visual communication and industrial design [6], [7], the monitoring of neurocognitive processes during creative work [8], as well as understanding how the assimilation of generative AI tools alters the creative processes of the creators themselves, such as art students [9].

In this context, a methodological approach that shifts the focus from creativity as an individual psychological disposition to the precise evaluation of the creative product contributes to the objectification of research [10], [11]. Building on this, the method relies on specifically constructed uniqueness and creativity quotients, which allow for the objective assessment of the originality and purposefulness of generated responses—such as image titles or drawings [10]. A series of studies has shown that these quotients exhibit high metric sensitivity and reliability, permitting accurate measurement of creative outcomes based on stimulus characteristics [10], [12]. Furthermore, this approach analyzes the relationships among emotional valence, arousal, content novelty, and the achieved level of product creativity, providing a framework for testing generative processes [13], [14]. Ultimately, applying this validated model to products of generative artificial intelligence enables a direct, metrically uniform comparison with human creative ability, treating both entities solely by the quality of the results they generate.

However, although previous research has delivered insights into the creative capacities of generative artificial intelligence, it has predominantly focused on textual tasks or relied on subjective evaluations of visual products. A significant gap in the literature is the lack of studies that use objective, product-based metric systems for direct comparisons of specific image-generation model designs among different human subpopulations.

To address this gap, the primary aim of this study is to apply the uniqueness quotient to assess and compare the creativity levels of visual products generated by different artificial intelligence models (both Intent-driven and Prompt-driven) and by humans (art and non-art students). The initial hypothesis is that there are statistically significant differences in the level of creativity of visual products across the observed groups, with the expectation that different models will exhibit distinct patterns of originality compared to human creators.

This research is theoretically important because it advances and extends the application of an objective methodology for assessing creativity into the domain of artificial intelligence, thereby overcoming the limitations

of subjective assessments. In practical terms, the findings of this study will offer more precise insight into the specific creative reach and limitations of different generative model designs, bearing implications for their adequate inclusion into creative industries, design, and human-machine collaborative processes.

2. METHOD

In the research, the subjects made small drawings, which were then categorized and evaluated for creativity using coefficients of exceptionality, enabling comparisons, a check of metric properties, and an evaluation of this assessment method.

2.1. PARTICIPANTS

The study sample comprised 53 human participants and 10 generative image models (GIM).

The human subsample consisted of two groups. The first group included 24 art students recruited from the Faculty of Fine Arts and the Faculty of Dramatic Arts in Belgrade (mean age = 22 years). The second group comprised 29 students of management and e-business from the Rectorate of the University of Arts in Belgrade and the School of Electrical and Computer Engineering (VISER) in Belgrade (mean age = 24.5 years). Students from art faculties represent a highly selected population of individuals trained in creativity and the arts, while students from the Rectorate and VISER are more representative of the general student population in terms of creative predispositions.

The image generation model subsample included the following systems: openai/gpt-image-1.5, google/nano-banana-pro, xai/grok-imagine-image, black-forest-labs/flux-2-pro, stability-ai/stable-diffusion-3.5-large, qwen/qwen-image-2-pro, bytedance/seedream-4.5, wan/wan-2.7-image-pro, bria/fibo-edit, and luma/photon. These models were categorized into two groups: (a) Intent-driven generation models (openai/gpt-image-1.5, google/nano-banana-pro, xai/grok-imagine-image, black-forest-labs/flux-2-pro, bytedance/seedream-4.5), characterized by advanced semantic and contextual perception of prompts and demonstrating more agentic and reasoning behavior; and (b) Prompt-driven generation models (stability-ai/stable-diffusion-3.5-large, qwen/qwen-image-2-pro, wan/wan-2.7-image-pro, bria/fibo-edit, luma/photon), characterized by high-fidelity rendering, precise and literal visual execution, and a reliance on explicit prompting, with a less developed assistant layer.



2.2. PROCEDURE

In the human condition, participants received a 10 × 15 cm sheet of white paper containing a predefined abstract stimulus (Figure 1). They were instructed to use the stimulus as a foundation for creating a drawing by freely extending the presented lines and producing any image of their choosing. The only constraint was a five-minute time limit. Participants used either graphite pencils or dark-colored pens. A total of 53 drawings were collected, digitized, standardized to a monochromatic format, and prepared for evaluation.

An analogous procedure was implemented for the image generation models. The same abstract stimulus was uploaded as an input image, accompanied by the prompt: “You need the pencil. Take a look at the paper. Complete those lines and make your own drawing.” This prompt was designed to closely replicate the verbal instructions provided to human participants. The procedure was repeated three times for each model with a different random seed, yielding a total of 30 generated drawings.

Participation for all human subjects was voluntary. All participants were informed that their data would be anonymized and that they could withdraw from the study at any time without penalty. Both written and verbal explanations of the study procedures were provided, and informed consent was obtained prior to participation.

Examples of obtained answers are illustrated in Figure 2.

The study protocol was approved by the Deans of all participating faculties and the University Rectorate and was conducted in accordance with the Code of Professional Ethics of the University of Arts in Belgrade, as well as the Ethical Principles and Code of Conduct of the American Psychological Association (APA).

2.3. MEASURES

Creativity was assessed at the level of individual drawings using the uniqueness coefficient. In the initial phase of its construction, two trained psychologists independently categorized the drawings based on perceived similarity. Drawings identified as original (i.e., forming unique categories) were further evaluated in terms of intentionality. Only one drawing was classified as exhibiting unintentional novelty and was subsequently reassigned to the most appropriate broader category.

Following category refinement, uniqueness quotients were computed for each response according to Equation 1:

$$UQ_i = 1 / m \quad i = 1, \dots, n$$

Equation 1. Uniqueness quotients formula

where UQ_i denotes the uniqueness quotient of an individual response, and m represents the number of participants who produced the same response [10], [11]



Figure 1. Abstract stimulus used in the task

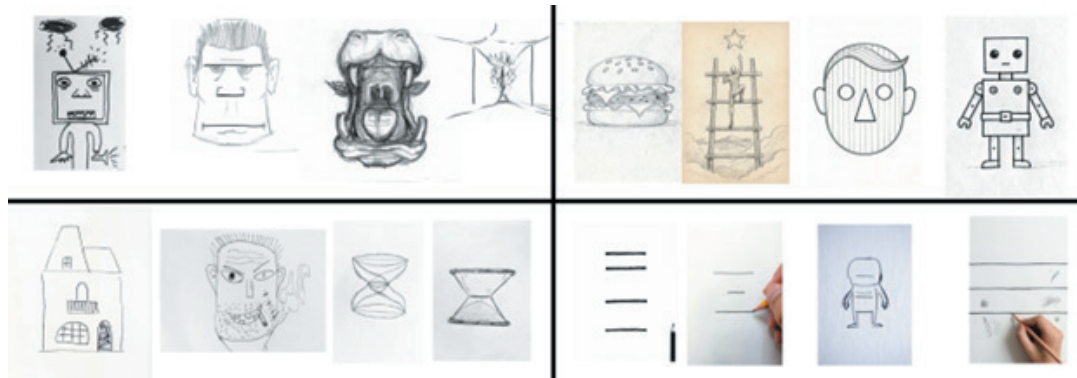


Figure 2. Answers obtained by different groups



For participants completing multiple heuristic tasks, an overall creativity quotient was calculated as the arithmetic mean of the uniqueness quotients across all responses according to Equation 2:

$$M = (UQ_1 + UQ_2 + \dots + UQ_n) / n = CQ$$

Equation 2. Creativity quotient formula

where M represents the arithmetic mean of the creativity of individual responses, UQ_1, UQ_2, \dots, UQ_n are the uniqueness quotients of individual responses, n is the total number of responses, and CQ represents the overall creativity quotient [10], [11].

This approach has been previously validated and applied in multiple studies investigating creativity [10], [11], [12], [13], [14], [15].

2.4. DATA ANALYSIS

Descriptive statistical analyses were first conducted at both the total sample and subsample levels, including minimum, maximum, mean, standard deviation, coefficient of variation, and estimates of measurement error. These analyses were used to examine score distributions across groups and to evaluate the metric properties of the assessment procedure.

Table 1. Descriptive analysis for total sample and subsamples

Uniqueness quotients	Minimum	Maximum	Mean	Std. error mean	Standard deviation	Variation coefficient
ALL	0.07	1.00	0.27	0.03	0.27	1.03
Human	0.07	1.00	0.28	0.04	0.27	0.97
Art	0.07	1.00	0.40	0.07	0.36	0.91
Non Art	0.13	0.20	0.17	0.01	0.03	0.17
GIM	0.07	1.00	0.25	0.05	0.28	1.14
Intent-driven	0.13	1.00	0.42	0.08	0.32	0.76
Prompt-driven	0.07	0.25	0.08	0.01	0.05	0.60

The normality of score distributions was assessed using the Shapiro–Wilk test.

To examine group differences in creativity, a non-parametric analysis of variance was conducted using the Kruskal–Wallis test, followed by Dwass–Steel–Critchlow–Fligner post hoc pairwise comparisons. The main effect of the sample factor was tested across the following groupings: Human vs. GIM, and Art vs. Non-art vs. Intent-driven vs. Prompt-driven.

Effect sizes were estimated using epsilon squared (ϵ^2) and interpreted according to the following thresholds: weak (0.01–0.04), moderate (0.04–0.16), relatively strong (0.16–0.36), strong (0.36–0.64), and very strong (0.64–1.00)[16].

3. RESULTS

The results of the descriptive analysis are shown in Table 1. A low standard error of measurement, along with a high dispersion of the results, is clearly evident. It is also possible to observe very similar metric characteristics when subsamples are compared in this regard.

The Shapiro–Wilk test revealed significant ($p < 0.05$) deviations from the normality of the distribution at the level of the entire sample as well as for all subsamples. These deviations are illustrated in Figure 3.

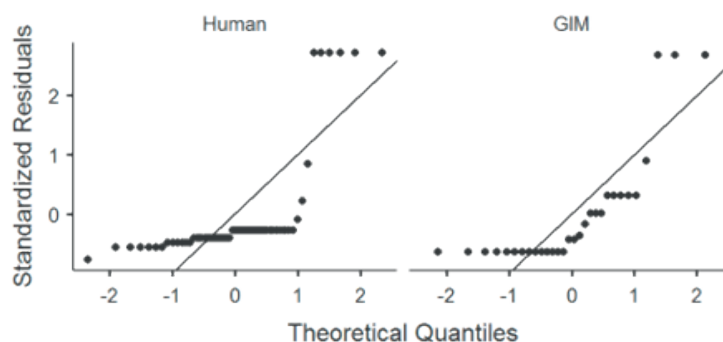


Figure 3. QQ plots of uniqueness quotients for Human and generative image model subsample



The results of the analysis of variance are shown in Table 2. Statistically significant differences ($p < 0.05$) were obtained between Prompt-driven with Art, Non Art, and Intent-driven, as well as between Intent-driven and Non Art groups. The effect of the subgroup factor is strong.

4. DISCUSSION

The main objective of this study was to use the uniqueness quotient (UQ) as a metric to assess and compare the levels of creativity in visual products generated by various AI models (intent-driven and Prompt-driven) and by human creators (art and non-art students). The results confirm the initial hypothesis that statistically significant differences exist between the observed subgroups and that different AI architectures show distinct patterns of originality.

Regarding the general level of creativity (Table 2), the analysis of variance showed no statistically significant difference between the overall human sample and the total generative image model sample. This finding is

closely aligned with recent studies [1], [2], which suggest that contemporary AI models have reached a level where they can match average human performance in divergent thinking tasks. The fact that generative models produce results that do not significantly deviate from the human average illustrates the advanced associative and generative capabilities of modern architectures.

However, a more in-depth analysis of specific subgroups (Art, Non-Art, Intent-driven, Prompt-driven) reveals a significantly more complex picture. Pairwise comparisons give new insights: no significant difference was found between art students and Intent-driven generative models. This implies that multimodal models with integrated large language models, which possess advanced contextual perception and support prompt rewriting, can generate visual outputs whose originality matches that of trained human creators. Furthermore, Intent-driven models markedly outperformed non-art students, demonstrating creative capacity beyond that of the average layperson. Interestingly, the difference between art and non-art students was only marginally significant, in line with previous studies on the specifi-

Table 2. Analysis of variance

Uniqueness quotients	χ^2	df	p	ϵ^2
Human vs GIM	2.92	1	0.09	0.04
Art vs. Non-art vs. Intent-driven vs. prompt-driven	38.2	3	< .001	0.47
Pairwise comparisons - UQ		W	p	
Art	Non-art	-3.31	0.09	
Art	Intent-driven	1.32	0.79	
Art	Prompt-driven	-6.65	< .001	
Non-art	Intent-driven	4.87	0.003	
Non-art	Prompt-driven	-6.82	< .001	
Intent-driven	Prompt-driven	-6.68	< .001	

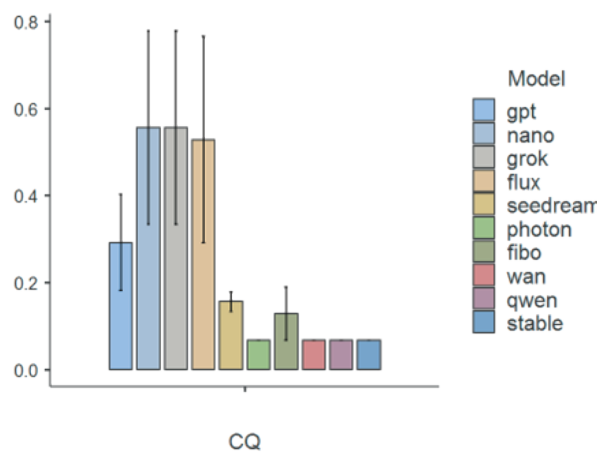


Figure 4. Creativity quotient scores of individual image generative models



cities of artistic generative processes [14]. A prominent distinction was observed in Prompt-driven generative models, which differed significantly from all other groups. This strongly supports the hypothesis that diffusion architectures lacking a solid “assistant layer” exhibit distinct patterns of originality. These models rely more on stochastic pixel synthesis and visual pattern matching from training data that is more literal in the dependence upon the prompt, without strong contextual reinterpretation, leading to outputs that are statistically distinct from both human and LLM-mediated productions.

The graphical representation of the distribution (Figure 2 – QQ plots) further illustrates these differences. While AI models can achieve a high average, the human sample shows more dispersion and unique behavior at the distribution tails. This directly corresponds to large-scale studies [4], [5] which identified a key asymmetry: the “most creative” segment of the population (the right-hand tail of the distribution) remains dominated by human variability and the capacity for truly exceptional, unpredictable associations that machines cannot yet systematically replicate.

Finally, the comparison of individual models (Figure 3) shows significant intra-group variability within the technology itself. Factors such as model design, training data, and prompt-processing mechanisms (e.g., Intent-driven versus Prompt-driven) directly dictate a model's capacity for original output. Using an objective metric, such as the creativity quotient, allows these structural differences to be observed at an exact, empirical level for the first time.

4.1. LIMITATIONS OF THE STUDY

In this study, several limitations should be acknowledged. First, the human sample was restricted to university students, which may not completely represent the wider creative spectrum of the general population or professional, high-level artists. Second, the visual tasks were limited to small drawings and specific prompts; different creative domains (e.g., complex digital painting or 3D design) might yield different results. Lastly, the fast evolution of generative AI means the models tested represent a “snapshot in time,” as newer architectures may quickly overcome present limitations.

4.2. RECOMMENDATIONS FOR FUTURE RESEARCH

Future studies should expand the diversity of the human sample to include professional creators throughout various career stages. Additionally, research can explore the use of multimodal prompts and examine how creativity quotients vary when humans and AI collaborate in a “human-in-the-loop” process. Investigating cultural differences in the assessment of visual originality using the uniqueness quotient would also deliver a more detailed global understanding of machine creativity.

4.3. PRACTICAL APPLICATION AND IMPLICATIONS

The findings from this study contain significant practical consequences for the creative industries. The high performance of Intent-driven models suggests they are already viable as high-level “co-creators” or assistants in design and advertising, capable of rivaling the output of art students. For educators, these outcomes offer a criterion for integrating AI tools into art curriculum, emphasizing areas where human originality remains dominant (such as the distribution extremes). Furthermore, the uniqueness quotient offers AI developers a solid, objective tool for benchmarking the “creative progress” of new models, moving beyond subjective aesthetic evaluations toward a more precise metric of innovation.

5. CONCLUSION

This research has confirmed the validity and metric reliability of the uniqueness quotient as an objective instrument for evaluating and comparing the creative visual products of generative artificial intelligence and human creators. The findings show that, at a general level, there is no statistically significant difference in average creativity between human subjects and AI models; however, a detailed analysis of subgroups reveals profound differences driven by model design and the level of human artistic training.

A major finding of the study is that Intent-driven models have reached a level of originality comparable to that of art students, while simultaneously outperforming non-art students. Conversely, Prompt-driven models display specific, statistically unique patterns of originality that differ substantially from all other groups, suggesting that the process of raw visual synthesis without deep semantic mediation generates a distinct type of creative output.



The theoretical contribution of this work is to overcome the subjectivity of traditional psychometric methods by providing an exact, metrically uniform framework for testing generative processes. The uniqueness quotient demonstrated high sensitivity to structural differences in product originality.

In practical terms, the data show that contemporary generative models, particularly those with integrated linguistic cores, can be efficiently integrated into creative industries as high-quality co-creators. However, the distribution of results shows that the most extreme levels of creativity still remain the domain of human variability. Future research can concentrate on systematic tracking of AI architecture evolution in direct human-machine collaborative processes, using the proposed objective methodology as a standard for measuring progress in digital creativity.

6. ACKNOWLEDGEMENTS

We would like to thank all participants in the research as well as University of Art in Belgrade and the School of Electrical and Computer Engineering in Belgrade

REFERENCES

- [1] V. Arora, A. Thabane, S. Parpia, G. Calic, and M. Bhandari, "Generative artificial intelligence models outperform students on divergent and convergent thinking assessments," *Scientific Reports*, vol. 15, no. 36987, 2025, doi: 10.1038/s41598-025-21398-4.
- [2] K. F. Hubert, K. N. Awa, and D. L. Zabelina, "The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks," *Scientific Reports*, vol. 14, no. 3440, 2024, doi: 10.1038/s41598-024-53303-w.
- [3] M. Koivisto and S. Grassini, "Best humans still outperform artificial intelligence in a creative divergent thinking task," *Scientific Reports*, vol. 13, no. 13601, 2023, doi: 10.1038/s41598-023-40858-3.
- [4] A. Bellemare-Pepin et al., "Divergent creativity in humans and large language models," *Scientific Reports*, vol. 16, no. 1279, 2026, doi: 10.1038/s41598-025-25157-3.
- [5] D. Wang, D. Huang, H. Shen, and B. Uzzi, "A large-scale comparison of divergent creativity in humans and large language models," *Nature Human Behaviour*, 2025, doi: 10.1038/s41562-025-02331-1.
- [6] H. Sun and S. Liu, "Evaluating generative AI tools for visual communication design using the CoCoSo method under interval valued spherical fuzzy environment," *Scientific Reports*, vol. 15, no. 34797, 2025, doi: 10.1038/s41598-025-18506-9.
- [7] J. Zou, K. M. Kamarudin, S. Z. Abidin, J. Zhang, and J. Liu, "Examining factors affecting the acceptance of AI-powered creativity support tools among the industrial design community in China," *Scientific Reports*, vol. 16, no. 3248, 2025, doi: 10.1038/s41598-025-33100-9.
- [8] S. Li, G. Cascini, and N. Becattini, "Neurocognitive differences in sketching between design tasks and creativity tests," *Scientific Reports*, vol. 16, no. 9964, 2026, doi: 10.1038/s41598-026-38735-w.
- [9] P. Song et al., "Redefining creativity differences for art students who use and those who do not use GenAI: evidence from network analysis," *Humanities and Social Sciences Communications*, 2026, doi: 10.1057/s41599-026-06897-7.
- [10] I. Ristic and M. Milošević, "Validacija koeficijenta izuzetnosti u proceni likovnih produkata," *Primenjena psihologija*, vol. 11, no. 2, pp. 227–246, 2018, doi: 10.19090/pp.2018.2.227-246.
- [11] M. Milošević and I. Ristic, "Nov metodološki pristup studijama kreativnosti," *Suvremena psihologija*, vol. 5, no. 8, 2016.
- [12] I. J. Ristic and M. Milošević, "Neugodni i novi sadržaji u funkciji kreativnosti: po čemu se studenti umjetnosti razlikuju od ostalih?," *Suvremena psihologija*, vol. 20, no. 2, pp. 117–136, 2017, doi: 10.21465/2017-SP-202-01.
- [13] M. Milošević and I. J. Ristic, "Appraisal of familiarity with content: Its relationship with dimensions of emotional experience," *Hellenic Journal of Psychology*, vol. 18, no. 1, pp. 34–45, 2021, doi: 10.26262/hjp.v18i1.7879.
- [14] I. J. Ristic and M. Milošević, "Emotions and Familiarity of Content in Generative Processes of Prospective Artists," *Studia Psychologica*, vol. 64, no. 4, pp. 356–370, 2022, doi: 10.31577/sp.2022.04.859.
- [15] I. J. Ristic and M. Milošević, "Ambivalencija u kreativnoj produkciji studenata umetnosti," *Primenjena psihologija*, vol. 13, no. 1, pp. 119–138, 2020, doi: 10.19090/pp.2020.1.119-138.
- [16] M. Tomczak and E. Tomczak, "The need to report effect size estimates revisited. An overview of some recommended measures of effect size," *Trends in Sport Sciences*, vol. 1, no. 21, pp. 19–25, 2014.