# Automated Aesthetic Evaluation: Can AI Objectively Assess Creative Growth? - A Critical Analysis

## Dr Satyamangal Rege

Professor Dean, Vasantdada Patil Prathisthan's College of Engineering and Visual Arts, Mumbai

Abstract—An urgent topic facing the art world is whether artificial intelligence (AI) can objectively assess artistic development as AI becomes more and more integrated into creative workflows. The promise and constraints of AI-driven aesthetic judgment in monitoring and analyzing an artist's creative development are examined in this study. We investigate how AI systems evaluate style, composition, and visual development across artistic bodies of work by utilizing machine learning approaches like generative algorithms and convolutional neural networks.

This study assesses how AI may assist with portfolio building, exhibition curation, and self-reflective practices through practical applications and in-depth case studies of modern artists. AI is capable of identifying stylistic changes and visual patterns, but it is not as good at comprehending cultural context, emotional resonance, or philosophical depth-all of which are essential components of creative development. We provide a hybrid paradigm that offers more collaborative, nuanced assessments by fusing human criticism with algorithmic analysis. The purpose of this article is to facilitate an informed discussion among artists, curators, and educators regarding the moral and practical ramifications of letting machines contribute to the interpretation of creative progress.

*Index Terms*—AI in Art, Art Education Aesthetic Evaluation, Automated Critique, Artistic Development, Creative Growth, Human-AI Collaboration, Machine Learning, Portfolio Analysis, Visual Arts

#### I. INTRODUCTION

The development of art is rarely linear. It develops through time, via failure, exploration, and epiphanies. It encompasses deeper changes in narrative intent, emotional maturity, and cultural participation in addition to changes in technique and form. Conventional assessments of creative development depend on human intuition, conversation, and contextual awareness—elements that AI as it exists today finds difficult to mimic. However, AI-based aesthetic assessment is becoming more and more popular. Curators test algorithms for show planning, artists use platforms like Runway ML to evaluate their own portfolios, and educators investigate AI solutions to provide students with formative input. The purpose of this research is to examine whether these instruments can (or should) be trusted as impartial observers of creativity and to what degree they can significantly aid in assessing artistic progress.

The first goal of this research is to evaluate the present strengths and limitations of AI systems in the field of aesthetic evaluation. The second goal is to suggest a hybrid evaluative model that maintains the importance of human intuition while utilizing AI's analytical capabilities. Case studies of modern educators and artists who have used AI directly in their instructional or creative endeavors are at the heart of this investigation. We aim to give the art community a fair, useful, and morally sound viewpoint by looking at the possibilities as well as the philosophical conflicts that arise.

#### II. LITERATURE REVIEW

#### 2.1 Historical Foundations: Quantifying Aesthetics

Both philosophy and early science have a long history of attempting to measure beauty and artistic value. Early foundations for aesthetic theory were laid by philosophers such as Plato and Aristotle, who investigated the concepts of ideal forms and proportion. In order to achieve mathematical harmony in visual composition, Renaissance artists like Leonardo da Vinci and Albrecht Dürer used geometry and symmetry into their assessments of art (Eco. 2004). More formal theories of aesthetics were established in the 18th and 19th centuries, such as Edmund Burke's investigation of the sublime and Immanuel Kant's concept of "disinterested judgment"-concepts that are still difficult to explain algorithmically.

Researchers tried to use empirical models to organize aesthetic response in the 20th century. For example, George Birkhoff (1933) presented the formula M = O/C for aesthetic measure, where M stands for aesthetic measure, O for order, and C for complexity. Despite their influence, these attempts frequently fell short of capturing the emotional, cultural, and subjective elements that are essential to artistic merit. Nevertheless, modern computational aesthetics was made possible by these earlier endeavors.

#### 2.2 Emergence of AI in Aesthetic Evaluation

With the use of AI and machine learning, aesthetic judgment has advanced significantly in the twentyfirst century. Convolutional neural networks (CNNs) have made it possible for AI systems to evaluate visual data with remarkably high accuracy. To determine style, genre, and even emotional tone, these models are trained on enormous datasets of artwork (Tan et al., 2016). Generative adversarial networks (GANs) are used by programs like DeepArt and GANPaint to both produce and assess artworks by contrasting them with predetermined aesthetic standards.

More and more AI-powered systems provide input on elements like texture, colour harmony, composition, and stylistic coherence. Early algorithms to predict aesthetic scores from photographic images were created by researchers such as Datta et al. (2006). More current models use deep learning to evaluate the intricacy, symmetry, and balance of artworks (Murray et al., 2012). These days, digital painting, generative art, and illustration are all used with these tools in addition to photography.

By examining how an artist's visual language evolves over time, some systems even make an effort to quantify "creative growth." Elgammal et al. (2018), for instance, created an AI model that can trace stylistic changes in historical painting movements, suggesting that similar techniques may be used to contemporary painters.

#### 2.3 Limitations and Critiques from the Art World

Notwithstanding these developments, the art community's criticisms have brought to light important weaknesses in AI's ability to evaluate. Contextual blindness is one of the main problems; AI is unable to take into consideration the layers of culture, emotion, and personal history that are included in an artwork (Whitelaw, 2021). Life events, personal development, or political commentary are frequently the sources of artistic development—aspects that AI finds difficult to identify or understand in a meaningful way.

Additionally, art theorists contend that biases in training data cause AI to reinforce prevailing aesthetic norms (Galanter, 2019). Unconventional styles or culturally distinctive art forms that deviate from popular or Eurocentric visual patterns may be marginalized as a result. Furthermore, when AI blurs the boundaries between intention and interpretation, it becomes both a critic and a co-creator, raising issues of authorship and agency.

When it comes to incorporating AI into art instruction, educators have voiced both excitement and concern. Although AI can offer unbiased criticism on formal components, if used excessively, it could erode students' faith in their own distinctive creative style (McCormack et al., 2020). Because of this, a lot of educators support a hybrid approach in which AI serves as a tool, not a substitute, for human mentoring and criticism.

### 2.4 Toward a Hybrid Understanding

AI and human criticism may not always have to be mutually exclusive, according to recent research. Rather, scholars suggest hybrid frameworks that blend qualitative, human-centered interpretation with algorithmic analysis. By encouraging new kinds of artist-machine collaboration, these models seek to enhance rather than replace artistic judgment (Cetin et al., 2022). The idea is to employ AI as a mirror, reflecting patterns and opportunities that might otherwise go overlooked, rather than reducing creativity to code.

## III. RESEARCH DESIGN AND METHODOLOGY

### 3.1 Research Design

A comparative mixed-methods strategy is used in this study to determine whether AI can measure creative development in visual art in an objective manner. Using both quantitative metrics and qualitative observations, it compares expert assessments over time with AI-generated aesthetic scores. In order to evaluate trends in aesthetic evolution, the study focuses on longitudinal data gathered from a chosen group of visual artists.

#### 3.2 Participants

Three contemporary visual artists with distinct styles and career trajectories were selected through purposive sampling. Each artist provided a portfolio of works created over a 5 to 10-year span (minimum 20 pieces per artist), enabling chronological analysis of creative growth.

#### 3.3 Data Collection Instruments

AI Evaluation Tools: I employed DeepArt frameworks which analyze visual elements such as colour harmony, compositional balance, and novelty. Human Expert Panel: A group of five curators, art educators, and practicing artists rated the same artworks using a structured rubric (focusing on originality, technique, emotional impact, and conceptual depth).

#### 3.4 Procedure

To remove prejudice, each piece of art was anonymised and randomly selected. Using pretrained models, the AI technologies processed the artworks and produced an aesthetic quality score ranging from 0 to 1. The four components of the rubric were evaluated by the human experts using a Likert scale ranging from 1 to 10. Cronbach's alpha was used to calculate inter-rater reliability.

#### 3.5 Data Analysis Techniques

Quantitative analysis was conducted using:

Correlation analysis (Pearson's r) to compare AI scores with expert ratings.

Linear regression to examine trends in AI-assessed aesthetic growth.

ANOVA to assess differences in scoring variance between AI and human evaluations.

Qualitative analysis was applied to expert comments and AI system rationales (where available) to identify patterns in evaluative logic.

### 3.6 Quantitative Findings

Correlation Between AI and Expert Scores Overall correlation between AI scores and average expert ratings was moderate to high (r = 0.68) across the dataset. The correlation was strongest in technical and compositional dimensions and weakest in emotional and conceptual aspects.

#### Creative Growth Patterns

AI systems consistently identified aesthetic improvement over time in two of the three artists.

Linear regression revealed a positive trend (p < 0.05) in AI scores over time for Artists A and C, indicating recognition of growth. Artist B showed no significant upward trend, aligning with expert feedback citing stylistic stagnation.

### Variance Analysis

ANOVA results showed significantly lower variance in AI scoring compared to expert ratings (F(1, 119) = 4.87, p = 0.03), suggesting AI evaluations are more consistent but potentially less nuanced.

#### 3.7 Qualitative Findings

Experts cited symbolic depth and cultural context as key reasons for their higher ratings in some cases where AI gave low scores.

In contrast, AI systems praised symmetry, novel textures, and colour contrast, even in works deemed less conceptually evolved by experts.

In post-evaluation interviews, artists noted that AI feedback lacked subjective understanding but sometimes highlighted formal patterns they had not consciously recognized.

#### IV. DISCUSSION

The study's findings show that although AI-based aesthetic evaluation tools show a discernible degree of agreement with human expert judgment, their capacity to evaluate creative development is still complex and somewhat constrained. AI may be able to capture some formal and technical elements that professionals also value, according to the moderate to good overall correlation (r = 0.68). This result lends credence to the idea that AI could be used as an additional tool for assessing visual art, particularly in settings that prioritize compositional harmony, colour dynamics, and structure.

The somewhat weaker correlations in the emotional impact (r = 0.53) and originality (r = 0.61) dimensions, however, suggest that AI finds it difficult to take into consideration the more contextual and subjective elements that frequently influence human aesthetic judgment. These elements

are particularly important when analyzing the richness and development of an artist's work over time, as changes in thematic maturity and emotional resonance are significant factors.

According to the results of linear regression, AI tools identified two out of the three case-study artists as having consistently grown, which is compatible with expert assessments in those instances. Experts and AI both saw minimal development for Artist B, indicating that when macro-level progression is present, AI may be rather sensitive to it. However, AI tools' dependence on formal visual features may restrict their capacity to identify conceptual progress, which is frequently less associated with surface aspects.

The fact that AI rankings have far less volatility than expert ratings from humans is an important finding. This implies that AI offers assessments that are more consistent but less diversified, which may be a reflection of both its algorithmic stability and its limited interpretive flexibility. Human assessors, on the other hand, showed more variation, most likely as a result of subjective and context-driven perceptions, which are crucial to art critique and appreciation.

This disparity is highlighted by the qualitative findings. Experts often mentioned emotional, symbolic, and cultural content—aspects that are mostly missing from AI assessments. AI feedback, on the other hand, focused on texture, colour contrast, and symmetry—all of which are significant but fall short of capturing the entire aesthetic experience. Notably, artists acknowledged the AI's incapacity to discern intention, narrative, or emotional aspects, but they also noted that some of its judgments were astute.

Overall, AI lacks the interpretive depth necessary to completely evaluate creative advancement in the holistic sense that the art world understands, even though it shows promise in providing objective, consistent assessments of some aesthetic traits. Therefore, its application might be best suited to enhancing human critique rather than taking its place, especially in experimental or educational contexts where quantifiable input is advantageous.

#### V. CONCLUSION

By contrasting AI-generated aesthetic assessments with human expert opinions across the changing portfolios of modern artists, this study aimed to investigate whether artificial intelligence can objectively evaluate creative growth in visual art. According to the results, artificial intelligence (AI) tools can detect surface-level progression and identify formal aesthetic features with reasonable accuracy, but they are unable to fully capture the complexity of artistic development, especially when it comes to emotional resonance, symbolic depth, and conceptual innovation.

The relationship between AI and human assessments shows some agreement, especially when it comes to elements like composition and technique. Differences in the emotional and interpretive aspects, however, draw attention to the shortcomings of AI's present assessment frameworks, which are blind to psychological, historical, and cultural factors. Furthermore, while AI grading is helpful in standardizing input, it does not replace the nuanced view that comes with human evaluation.

Crucially, the findings highlight that while AI can be a useful supplementary tool, it is currently unable to assess creative growth on its own. It provides reproducible, data-driven insights that could be particularly helpful in large-scale portfolio assessments where scalability is required, iterative creative processes, or educational settings.

AI's function in evaluating art may grow as it develops further, incorporating cross-modal comprehension and contextual awareness. But for the time being, human knowledge is still indispensable for deciphering the complex, individualized process of artistic development. To better assist artists and educators in evaluating creative development, future studies should investigate hybrid models that blend computer analysis with curatorial reasoning.

#### REFERENCES

- [1] Berlyne, D. E. (1971). Aesthetics and psychobiology. Appleton-Century-Crofts.
- [2] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms. Proceedings of the 8th International

Conference on Computational Creativity (ICCC), 96–103.

[3] Franke, K. (2016). Computational aesthetics: Finding meaning in images. Leonardo, 49(3), 218–219.

https://doi.org/10.1162/LEON\_a\_01141

- [4] Greenfield, G. (2005). On the origins of generative art. Leonardo, 38(1), 74–75. https://doi.org/10.1162/0024094053083703
- [5] McCormack, J., Gifford, T., & Hutchings, P. (2019). Autonomy, authenticity, authorship and intention in computer generated art. Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems, 89–115.
- [6] Northedge, R. (2020). Evaluating art with algorithms: Can AI judge creativity? Art Monthly, 440, 1–5.
- [7] Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. Frontiers in Human Neuroscience, 9, 218. https://doi.org/10.3389/fnhum.2015.00218
- [8] Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. Creativity Research Journal, 24(1), 92–96. https://doi.org/10.1080/10400419.2012.650092
- [9] Schwarz, J., & Winkielman, P. (2020). Affective response to art: From architecture to literature. Current Opinion in Behavioral Sciences, 35, 109–114.

https://doi.org/10.1016/j.cobeha.2020.08.006

[10] Yanardag, P., Chatterjee, P., & Kann, M. (2021). ArtPI: Aesthetic evaluation and interpretation through perceptual indicators. International Journal of Artificial Intelligence in the Arts, 2(1), 34–52.