Proceedings of ICEIPI 2025 Symposium: AI Am Ready: Artificial Intelligence as Pedagogical Scaffold
DOI: 10.54254/2753-7048/2025.CB23288

A Study of Generative Artificial Intelligence-Based Applications in Evaluating the Sketch Works of Undergraduate Art Students: A Comparison of ChatGPT and Claude

Chuhan Shangguan

School of Design and Art, Tianjin Academy of Fine Arts, Tianjin, China
2211120224@stu.tjarts.edu.cn

Abstract: The general field of Artificial Intelligence (AI) in art education is different from other applications, but there is not much research that can be used as a benchmark for the reasons that AI is effective in artwork evaluation. The researcher studied the effectiveness of ChatGPT-4o and Claude-3.5-Sonnet AI language models for evaluating the artwork sketches created by undergraduate art students. This study helps to realise the impacts of Artificial Intelligence on education through its ability to generate tailored feedback and real-time teaching in art education. Evaluation results are adversely affected as they are evaluated using different methods. Though both techniques bring direct strengths to assessment, they must be blended in such a way as to provide creative and technical feedback into complete evaluation results for an AI assessment system. The development of the student's artistic practice would be better in the event of the receipt of combined advantages from AIs. Based on this, the research proposes to enhance the AI model to come up with innovative motivation for enhancing the technical support until an advanced AI assessment system is developed that will improve the quality of evaluation and artistic skills in pedagogical art, as well as coming up with innovative advancements.

Keywords: AI-Assisted Art Evaluation, Chatgpt-4o, Claude 3.5-Sonnet, Analysis of Sketch Works, Art Education

1. Introduction

Reports show how Artificial Intelligence (AI) and educational methods in mass higher education systems have a long–term impact on the world. In the past ten years, the implementation of artificial intelligence education has grown rapidly and naturally increased research interest in incorporating AI technology into a college curriculum that is not in computer science [1]. AI has been viewed by the educational community for a long time as a system that efficiently performs student assignment assessment and grading through a personalised feedback system [2,3]. In art education, the evaluation of artworks necessitates sophisticated visual perception, creative analysis, and technical judgment. Introducing AI in this domain offers various methods of helping teacher grading, specifically by creating personalised critiques that help students improve their creative work [4]. Furthermore, AI provides teachers with the opportunity to delegate their maintenance work and other basic classroom duties to teach them to build those connections with their students [5]. The validity of AI assessment

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.CB23288

systems to evaluate originality and technical and expressive capabilities that now dominate new research in this area depends on inconsistent validity principles of art education.

Alongside these, it is the expansion of these digital artwork open source code repositories in conjunction with the quick rate of technological advancements of Artificial Intelligence systems in themselves and Deep Learning technologies that have made it possible to achieve initial success in the analysis of artwork. Therefore, this allows a predictive AI programme to analyse artwork features to detect which emotional dimensions the programme interprets as being present within the content [6,7]. Yinan Zhang created a modern art design system to extract features to improve artwork recognition through deep learning algorithms [8]. Eva and James proposed that artificial intelligence technology can autonomously analyse the visual characteristics of artworks and perform functions such as classification, target detection, similarity retrieval, multi-modal representation, and computational aesthetics, helping researchers better comprehend the content, style, and emotions of artworks, thereby offering references for art collection and investment [9].To further advance the use of deep learning techniques in evaluating art emotions, Panos et al. created the ArtEmis dataset and neural speaker models trained on the dataset. Researchers develop new models that make meaningful readings and emotional expressions [10]. Gregory et al. have introduced a novel system through bi-modal deep networks combining computer vision and natural language processing to find the corresponding meanings in artwork objects, and a new approach to art technology to further interpret artwork meaning and semantics is established [11].

However, before AI in art education can be applied, there are the following three vacancies, and it is necessary to deepen the research on AI in art education from an empirical viewpoint. One is that the profound impact of AI on higher education has not been widely studied in university art education [2, 12]. The second reason is that although existing studies mainly develop the technology, cross-model comparisons in the application for a given dimension (such as creativity and technicality) are rarely made.

In this paper it studied the performance of ChatGPT-4o and Claude-3.5-Sonnet was studied concerning assessing sketches using both comparative analysis and in-depth interviews. Then, based on six aspects, i.e. composition, proportion, line, light and dark, detail performance and creativity, comparisons are presented in terms of scoring style and feedback content of the two, and the influence of potential bias brought by AI assessment in the decision making of instruction and experience of the student is analysed. This study provides essential reference for the application of AI in art education evaluation and empirical evidence for developing a more intelligent AI grading system. Through the deep analysis of AI performance in evaluating sketching works, this study extends the research boundary of AI in art education and provides a theoretical basis for designing a more accurate artwork evaluation system.

2. Research Methods

Research measured ChatGPT-4o and Claude-3.5-Sonnet together using quantitative and qualitative methods while looking at drawings in both comparative methods and talking to experts. First, the research started by reviewing score consistency between the models, after that it analysed differential feedback about technical and artistic elements to discuss the potential of AI scoring in art education.

2.1. Comparative Analysis

2.1.1. AI Models Selection

OpenAI and Anthropic, respectively, developed the two cutting-edge natural language processing models ChatGPT-4o and Claude-3.5-Sonnet, therefore taking the research to choose to analyze. In May 2024, OpenAI debuted the deep learning model ChatGPT-4.0, which was built on ChatGPT-

DOI: 10.54254/2753-7048/2025.CB23288

4.0's construction and has better skills in comprehension and language creation. Text development and visual understanding, in conjunction with multidimensional data processing, help the users. Claude-3.5-Sonnet is the artificial intelligence system intended to improve the robustness of language processing compared to the previous Claude-3.0 and to incorporate secure and stable AI operations over many domains.

2.1.2. Data Collection

Twenty-five quarter-size sketches were obtained from comprehensive university art students to compare the evaluative performance of Claude-3.5-Sonnet and ChatGPT-4o in art critique tasks. The artworks spanned various skill levels and thus were built in a way that maintained data diversity. All pieces originated from professional art courses and were initially vetted by experienced instructors to meet standardised sketch assessment criteria. Standardised work photographing was done, and all works were taken at the same fixed camera angles with the same uniform lighting to achieve uniform brightness of the images, scoring consistency, and data quality improvement. To ensure that the student's confidentiality was protected and to eliminate scoring prejudice, the researchers digitised the works at 300 dpi resolution and eliminated all personal naming details from the works.

2.1.3. Experimentation

The experimental process contained three successive phases. First, preprocessing operations were applied to 25 sketches before the procedures to establish data security and normalise their formats. Next, the study used existing scoring criteria that Chinese universities employ when evaluating sketches through five core evaluation dimensions: line quality together with light/dark contrast, composition/spatial awareness, detail delineation, and creative expression. The research conducted its assessment using the identical instruction framework (ICIO) through the POE platform with ChatGPT-4o and Claude-3.5-Sonnet for scoring, along with detailed feedback schemes. Five evaluative dimensions enable a complete assessment of AI model artwork analysis by measuring how well it draws lines and how accurately it represents shapes, as well as its spatial organisation, capacity for visual detail representation, and capability to convey personal and distinctive artistic elements. In the third stage, comparison analysis using scoring results evaluated the performance of both AI models through scoring consistency measures and feedback variations and identified the primary aspects of each AI evaluation.

2.2. In-Depth Interview

The evaluation of AI grading's effectiveness and rationality required this study to conduct interviews with ten students of experimental sketches as well as three experienced drawing teachers. The main topic during these interviews focused on AI-generated feedback and discussion, which included evaluation process transparency, feedback accuracy, and availability of precise interpretations of technical and creative sketch elements. The interviews also looked into how teachers and students perceive the feasibility, applicability, and possible pedagogical value of AI grading systems in art schools. The research aims to offer more comprehensive insights into the future implementation of AI in art instructing by gathering teachers' and students' opinions on the quality of AI grading, the consistency of the criteria, and whether the feedback helps support learning and enhancing skills, as well as to help explore where AI meets traditional educational methods and to provide theoretical support for pedagogical reform and technological innovation.

DOI: 10.54254/2753-7048/2025.CB23288

3. Results

3.1. Overall Validity and Reliability of AI Scores

Overall, both ChatGPT-4o and Claude-3.5-Sonnet showed high reliability in evaluating the artwork's five key qualities (composition, spatiality, line, proportion, and light and dark treatment), guiding students in analysing and comprehending their work's strengths and faults. The majority of participants in interviews found both AI technologies to be reliable assessment tools during the evaluation process. The assessment by ChatGPT-4o was positive and supportive because it focused on motivating artistic expression through constructive feedback. Respondents largely believed that such comments had a favourable impact on students, particularly beginners, in the creative process and may motivate them to continue exploring and becoming creative. In the case of Sketch Teacher A, "ChatGPT-4o focuses on the overall composition and the integration of light and shadow of the completed piece, and this positive evaluation helps students build conviction and motivates them to make further creations." The technical assessment of Claude-3.5-Sonnet performs detailed analysis methods. It enables a more in-depth analysis of the work's lines, proportions, and light-dark contrasts, as well as the ability to discover and highlight significant technical issues in students' work. Students received helpful feedback, according to the majority of survey respondents, which allowed them to find and remedy technical mistakes in their artwork so they could enhance their artistic ability.

3.2. Discrepancies in AI Scores and Influences

The results of the research revealed considerable disparities in the grading styles of the two different AI models, as well as the impact of these variances on students' creativity. The grading style of ChatGPT-4o was relatively lenient, emphasising the overall effect and inventiveness of the work rather than the technical specifics. Many students noted that the comments on the ChatGPT-4o made them feel like their work was "on point", which enhanced their creative confidence. For example, the student in Sample 3 stated, "The feedback from ChatGPT made me feel like my work was on point, and although I know there is still work to be done, it made me feel like my creations were worthwhile." However, all three sketching instructors noted that ChatGPT-4o lacked depth in evaluating detailed representation, particularly in its study of local material expression, and failed to comprehensively guide students to improve their techniques.

In contrast, Claude-3.5-Sonnet provided precise instruction in art techniques alongside methods for recognising and improving basic artistic elements like lines and proportion, and chiaroscuro shadows. The student in Sample 2 stated: "Claude pointed out problems with the handling of line and chiaroscuro in my work, which made me acknowledge my shortcomings and motivated me to improve on them in my next creation." Students expressed frustration when Claude gave lower marks to certain aspects of their work even though, overall, their work was meritorious. For example, the student in Sample 5 commented, "Despite the overall harmony and beauty of my work, Claude gave a lower grade because of certain minor flaws, which was a little frustrating." Some students believed that excessive attention to technical issues would undermine their creative confidence and even dampen their enthusiasm for art creation. B, the drawing instructor, said, "Claude's high demand for details sometimes neglects the overall artistic expression of the work and may make students feel that their work is never perfect."

In general, most art teachers and students have positive impressions regarding the addition of AI scoring systems in their educational spaces because these tools deliver precise and objective feedback that focuses on technical areas that students and teachers view as essential. However, all three teachers agreed that AI scoring cannot wholly replace manual grading, particularly when it comes to creative expression and sentimentality in artworks, and that AI tends to be overly technical, ignoring the

DOI: 10.54254/2753-7048/2025.CB23288

artistry of the piece. Therefore, future research should focus on balancing technical evaluation with creative expression and designing a more comprehensive and adaptable AI-assisted assessment system that can better meet the demands of multiple learners and educational contexts.

4. Discussions

According to the findings of the interviews and studies, while the use of an AI grading system has certain effectiveness and benefits, it also highlights several issues that must be avoided and optimised in practical implementations. The diversity of the AI evaluation indicates that different scoring techniques have different impacts on students' artwork. ChatGPT -4o is more likely to stimulate creative expression and focus on the highlights of the creative process, which is a good incentive for novices and helps students gain faith in their paintings. However, this loose grading criterion may ignore the works' technical elements, particularly faults in material expression, line precision, and light-dark contrast.Claude-3.5-Sonnet, on the other hand, assists students in finding detailed problems in their pieces through rigours technical evaluations, particularly in composition, proportion, and line treatment, which can help students to improve their techniques in greater depth; however, it may also lead to excessive attention to details, which may hurt the overall visual and expressive quality of the creation. Technical assessment consumes an excessive amount of attention when it comes to students being forced to lose their freedom to imagine and to search for perfect details in a never-ending search that is preventing them from creating innovative and expressive works.

However, as it turns out, many experts are worried about biased behaviour and the lack of adaptability to various cultures, which is often demonstrated by AI evaluation [13,14]. Claude-3.5-Sonnet was composed of technical quality standards used to evaluate the quality of art, which conformed to traditional Chinese art education values. However, their bias partly prevented them from working with various educational methodologies. Given that the communication among the different countries is increasing and that it is necessary for this development to be rapid, AI systems for evaluation of artistic creativity must define examination standards of artistic considerations of globalised culture and local educational needs in different cultural and educational environments, where the evaluation will be done according to current educational objectives. For example, the Western educational system follows creative freedom and values open-ended inquiry; therefore, the AI scoring algorithm must be flexible enough to allow regional idiosyncrasies. Therefore, future AI assessment systems will have to find ways to combine technology-based approaches with dynamic assessment objectives to prevent damage to innovative student responses.

Several solutions are recommended to resolve these challenges. To enhance the creativity of students as well as aid them in finding technical errors in their work, teachers should use artificial intelligence characteristics and their educational requirements to develop instructional approaches. While teachers and students need to know more about AI systems, they also need to exercise cognitive evaluation in AI grading and the processing of AI feedback using reason, and then reduce their use of technical evaluations. Finally, the AI grading system should be flexible enough to change the criteria based on diverse educational backgrounds and students' demands, ensuring that technical assessment and imaginative expression are merged to boost students' overall development.

5. Conclusion

The research investigates the assessment capabilities of ChatGPT-4o and Claude-3.5-Sonnet on comprehensive university art student sketches. Both AI models demonstrate abilities to help students, yet their assessment techniques exhibit major differences according to the study results. These contrasts illustrate potential biases and limitations in AI assessment systems, especially as Claude-3.5-Sonnet's strictly technical criteria may cause it to disregard the work's inventiveness and general

DOI: 10.54254/2753-7048/2025.CB23288

quality, restricting students' artistic autonomy. On the other hand, while ChatGPT-4o might provide imaginative trust and encouragement, it could miss work specifics and technical concerns, hindering students' technical advancement. The performance potential of AI review systems depends on their ability to properly integrate aspects of creativity with technological capacities and holistic perspectives with specific details to create improved assessment approaches. Research needs to focus on developing methods that unite the advantages of multiple AI models into an improved artwork grading system that gives educators and students comprehensive artistic understanding.

References

[1] *Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2021). Evaluation of an Artificial Intelligence Literacy Course for University Students with Diverse Study Backgrounds. Computers and Education: Artificial Intelligence, 2, 100026.*

[2] *Luckin, R., Cukurova, M., Kent, Carmel and du Boulay, B. (2022). Empowering Educators to be AI-Ready. Computers & Education, 3, 100076.*

[3] *Watters, A. (2021). Teaching Machines: The History of Personalized Learning. The MIT Press.*

[4] *Lyuzhaozhao, Y., Runkun, P., Rui, Z., Suwan, G., Tianyi, Z., & Heng, X. (2024). Exploring the Feasibility of AI's Auxiliary Functions in the Field of Children's Painting, Cambridge Explorations in Arts and Sciences, 2.1*

[5] *Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education–Where are the Educators?. International Journal of Educational Technology in Higher Education, 16(1), 1-27.*

[6] *Zhuomin, Z., Jia, L., David G., S., Elizabeth, M., John, R., Catherine, A., & James Z., W. (2022). Reducing Bias in AI-based Analysis of Visual Artworks, IEEE BITS the Information Theory Magazine: 1-16.*

[7] *Xi, C., Yuebin, L., & Wei, Y. (2024). Generative AI in Higher Art Education, 2024 6th International Conference on Computer Science and Technologies in Education (CSTE), 135-140.*

[8] *Yinan, Z. (2022). Modern Art Design System Based on the Deep Learning Algorithm. Journal of Interconnection Networks, 22.*

[9] *Eva, C., & James, S. (2022). Understanding and Creating Art with AI: Review and Outlook, ACM Transactions on Multimedia Computing, Communications, and Applications, 18(2), 1-22.*

[10] *Panos, A., Maks, O., Kilichbek, H., Mohamed, E., & Leonidas, G. (2021). ArtEmis: Affective Language for Visual Art. Computing Research Repository, 11564-11574.*

[11] *Gregory, K., Ryan-Rhys, G., Anthony, B., & David G., S. (2022). Extracting Associations and Meanings of Objects Depicted in Artworks Through Bi-Modal Deep Networks, Electronic Imaging, 34(13), 170-14.*

[12] *Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial Intelligence Literacy in Higher and Adult Education: A Scoping Literature Review. Computers and Education: Artificial Intelligence, 100101.*

[13] *Henriikka, V., & Matti, T. (2023) Using Artificial Intelligence in Craft Education: Crafting with Text-to-image Generative Models, Digital Creativity, 34(1), 1-21.*

[14] *Zied, B., Chiraz, A., Vian, A., & Andrew, Z. (2023). Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings Through Bibliometric and Content Analysis, Sustainability, 15, 17.*