Algorithmic Bias in the Calligraphy Education Evaluation System

Yu Jiang

School of Sichuan, Sichuan Normal University, Chengdu, Sichuan 610066, China

Abstract. This study examines the algorithmic biases embedded in AI-driven calligraphy education evaluation systems and their profound implications for cultural and educational equity. By analyzing existing commercial systems (e.g., RJET Xiumo, Shanghai Writing Proficiency Test Digitalization System), the research identifies three core dimensions of bias: (1) Technical-mechanical reductionism, where rigid metrics (e.g., stroke standardization, grid positioning) disregard the dynamic artistry and emotional expression central to calligraphy, such as misjudging masterpieces like Wang Xizhi's "Lanting Xu" as flawed; (2) Data hegemony, reflected in skewed training datasets dominated by Tang regular script (67%) and modern works (91%), while marginalizing minority scripts (e.g., seal script, Dongba script) and historical styles, leading to inaccurate evaluations of non-mainstream calligraphy; and (3) Cultural violence, where algorithmic standardization imposes industrial logic on artistic practice, eroding aesthetic diversity and pedagogical autonomy. Empirical evidence reveals that these biases exacerbate educational inequalities—e.g., disadvantaging rural students and suppressing stylistic innovation—while reinforcing technocratic control over cultural valuation. The study calls for a human-centered redesign of AI systems, emphasizing data justice, algorithmic transparency, and symbiotic collaboration between technologists, educators, and cultural stakeholders to preserve calligraphy's essence as a "heart art" in the digital age.

Keywords: Algorithmic Bias; Calligraphy Education; AI Evaluation; Cultural Hegemony; Educational Equity; Human-Computer Interaction.

1. Introduction

1.1 Introduction the Research Background

Current AI calligraphy assessment systems (such as the RJET Xiumo system and the Shanghai Writing Proficiency Test Digitalization System) achieve automated evaluation through technologies like CNN image recognition and stroke path parameter analysis. However, they have many issues. One system, for instance, uses quantified indicators such as "stroke standardization" and "structural proportion" as its core, leading to an extreme case where a copy of Wang Xizhi's "Lanting Xu" was given a score of 0 due to "unneat paper". Moreover, the scoring model focuses on the accuracy of grid positioning, with a very low tolerance for errors, while neglecting the emotional expression and cultural connotation of calligraphy works. Some calligraphy evaluation systems have training data mainly concentrated on mainstream styles like Tang regular script and running script, resulting in inaccurate evaluations of minority styles like seal script and clerical script. There are also biases in cultural orientation. In a provincial-level calligraphy assessment system, the proportion of works by modern and contemporary masters in the training data is as high as 91%, leading to an excessive emphasis on modern calligraphy styles and a lack of ancient calligraphy data. Discussions on algorithmic biases in AI calligraphy evaluation systems also exist. [1-2]

1.2 Research Significance

It holds profound humanistic value in the emotional dimension, with its core being to resist the erosion of educational warmth by technological rationality and to restore the emotional connection between "human" and "culture" in calligraphy education. By exposing the reality of declining accuracy in the recognition of seal and clerical script due to algorithmic bias and the anxiety of rural students during evaluations, the research directly points out the dual dissolution of cultural identity and creative subjectivity caused by technological alienation: when an AI system gives a score of 0 to

a copy of the "Lanting Xu" on the grounds of "messy handwriting", it essentially represents a violent deconstruction of the thousand-year-old tradition of "expressing one's heart and nature" in calligraphy. The intention is to convey the opening of an ethical channel of "nourishing the heart with emotions" for the inheritance of traditional culture in the digital torrent, and to prove that the integration of technological improvement and humanistic care is the fundamental path to safeguarding the emotional ecosystem of education.

1.3 Literature Review

1.3.1 Review of Research in China

A search for the key words "algorithm bias or algorithm discrimination" in the CNKI database reveals that domestic research on algorithm bias began in 2016. The main research areas include: analysis of potential algorithm bias problems in different application scenarios, analysis of the forms and categories of algorithm bias, analysis of the connotation and causes of algorithm bias, analysis of the risks and consequences of algorithm bias, and research on the governance and response to algorithm bias. In 2016, Fang Shishi analyzed the "Facebook" bias incident and found that the collaborative filtering mechanism of dynamic news push algorithms in platform media may lead to information manipulation and platform bias. In 2019, Liu Pei and others pointed out that algorithm bias has three forms: explicit discrimination, implicit discrimination, and differential impact, and can be divided into two categories: algorithm discrimination related to identity rights and algorithm discrimination unrelated to identity rights. In 2020, Zhang Tao and others defined the connotation of algorithm bias as the biased processing of data that violates recognized norms, laws, or moral principles. They classified the causes of algorithm bias into two categories: bias related to input data and bias related to the algorithm itself, based on the perspective of "similar aggregation". In 2021, Yi Weifeng further considered human factors in the analysis of the causes of algorithm bias and pointed out that the biases or discriminations of developers and data collectors may be imperceptibly incorporated into the development and design of algorithms and the selection of data. In 2022, Chen Fengrun discussed the legal dilemmas in regulating algorithm discrimination, and Han Shipeng and others analyzed the regulatory dilemmas of algorithm price discrimination. Ni Qin and others explored the potential risks of algorithm discrimination in intelligent education. In 2023, Qu Liang and others proposed governance suggestions for achieving algorithm fairness from four aspects: technology, law, enterprises, and society. [3-5]

1.3.2 Review of Foreign Research

Research on algorithm bias abroad began in 2015 and mainly covers the development and bias assessment of algorithm tools, governance of algorithm bias, analysis of the harms of algorithm bias, analysis of algorithm bias in medical artificial intelligence, and the definition of algorithm fairness. In 2015, Liu and others developed and verified a risk prediction algorithm for panic disorder recurrence and evaluated whether the algorithm had discrimination issues. In 2017, Datta and others' research on improving the transparency of algorithm decision-making systems laid the foundation for detecting algorithm discrimination. In 2018, Hacker proposed a comprehensive vision of anti-discrimination and data protection laws by drawing on computer science literature on algorithm fairness. He demonstrated how anti-discrimination laws could be combined with algorithm audits and data protection impact assessments to unlock the algorithm black box and implement fairness in the digital age. In 2019, Sirbu pointed out that algorithm optimization by online media platforms based on popularity and proximity to the target could introduce algorithm bias and exacerbate the fragmentation and polarization of social debates. In 2021, Galaz and others analyzed the algorithm bias and its harms in the application of artificial intelligence in agriculture, forestry, and marine resource extraction. They pointed out that algorithm bias leads to unequal access to opportunities and benefits. In 2023, Zhang and others argued that algorithm bias is one of the factors affecting the credibility of medical artificial intelligence and can affect clinical predictions and exacerbate health disparities. Ferre explored the issue of algorithm bias in the medical field and the necessity of

regulating data donation, arguing that incomplete datasets, non-representative data, and the difficulty for researchers to access data are the causes of algorithm bias. Scholars such as Agarwal have revealed the challenges brought by algorithmic bias in their research, described a four-step analytical process for building and deploying AI mobile healthcare algorithms and solutions, and emphasized the sources of bias and methods to mitigate it. Dellunde et al. have provided an abstract description of the concept of algorithmic fairness from the perspectives of philosophy and computer science. [6-10]

2. Algorithmic Bias in Artificial Intelligence

In the tapestry of the digital society, "algorithms" are no longer merely technical tools for information distribution or e-commerce search, but have evolved into a dominant force, with their reach reaching every corner of the social texture. From the construction of the "algorithmic society", we see how intelligent systems shape our life choices, to the cultural circles defined by personalized recommendations in the "algorithmic culture", and then to the information world customized according to preferences in the "algorithmic news", algorithms are gradually rewriting the cognitive map of human beings. "Algorithmic discourse", as an emerging form of communication, is not limited to technical language but profoundly influences the framework and depth of public discussion. Meanwhile, the "algorithmic black box" warns us of the urgent need for decision-making transparency. Against this backdrop, "algorithmic discrimination or bias" has become an issue that cannot be ignored. It is fully exposed in "big data price discrimination", exacerbates inequality in commercial credit decisions, may solidify social bias in judicial practice, and may also miss out on potential talents in education and selection.

In the field of education, algorithmic bias is like an "electronic judge" wearing tinted glasses. The grading system it trained with the data of urban students might misjudge the lag of rural students' online access as "not studying seriously". A recommendation system designed based on historical admission patterns would be like a vending machine that only sells fixed packages, making it impossible for students from poor families to access information about summer camps at prestigious universities. What is more concealed is that some essay correction algorithms are "allergic" to colloquial expressions, causing students accustomed to thinking in dialects to always be deducted points, just like insisting on scoring Cantonese poems in Mandarin standards. The most dangerous aspect of this prejudice lies in the fact that it can wear the guise of "objectivity and fairness", using functions such as intelligent class division and potential prediction to sift students into different fields like a sieve. In Hangzhou, there was once a junior high school student who, despite having an outstanding talent for mathematics, was constantly pushed simple questions by the system simply because the algorithm determined that "rural children cannot learn Olympiad mathematics".

3. Evaluation of Calligraphy Education in the Era of Artificial Intelligence

In calligraphy education, the fundamental step is for students to copy the works of renowned calligraphers throughout history. For students, during the process of learning calligraphy, they often cannot receive intuitive and visual guidance from calligraphy teachers, which causes difficulties in understanding, restricts the progress of their calligraphy abilities, and also has an adverse impact on their learning efficiency and enthusiasm. The achievement, as an important application in visual artificial intelligence-assisted teaching, has an objective market application prospect. It is presented on mobile terminals in the form of mobile software. Users only need to select the template characters they want to copy, take photos and upload the characters they have written, and then they can obtain a continuous calligraphy score ranging from 1 to 10 points evaluated by the system, which is close to the expert score. And provide visual guidance for users to facilitate their understanding and modification. This system will greatly enhance the efficiency of calligraphy education in our country and has a considerable market prospect. The specific application scenarios may include, but are not limited to, self-checking after students complete their copying assignments. Through the professional

scores and relevant guidance provided by the system, it can be used to guide students to make revisions. It can also be configured in the school's auxiliary teaching system to enhance the work efficiency of calligraphy teachers. But at present, there are many biased phenomena in the evaluation of artificial intelligence. [11-12]

Firstly, the image recognition technology commonly adopted by current AI calligraphy evaluation systems is transforming calligraphy art into an industrial standardized measurement object. Taking the 21 characters of "zhi" in the "Lanting Xu" as an example (which should actually be 21 different forms of writing), after being grid-cut by artificial intelligence, these vibrant variations in character shapes are compressed into pixel deviation values on the coordinate axes. When students make very minor deviations in the composition of their copied works, the system immediately triggers a "structural imbalance" warning. This mechanical evaluation criterion completely violates the spatial philosophy of calligraphy art, which is "sparse enough for a horse to gallop through, dense enough for no air to pass through". What is more serious is that there are fundamental cognitive biases in the algorithm architecture of the existing system. Research by Wang Zhaoyi's team from Jiangnan University shows that relying solely on static image analysis results in extremely small values for composition and layout, while integrating data from brushwork achieves an extremely high accuracy rate. This reveals that the current business system generally lacks the dynamic perception ability of the writing process and is unable to capture the connection of the writing style of "wanting to be right before leaving, and taking everything in". When artificial intelligence determines the emotional outburst cursive writing in Huai Su's "Self-Narrative Letter" as a "writing error", it is essentially using the standard logic of printed text to kidnap the freedom of time and space in artistic expression. This technological violence is particularly prominent at the level of ink color analysis. Another artificial intelligence identification system, although capable of recognizing color differences within a broader range, is completely speechless when it comes to the aesthetic implications of the "light, dry, wet and burnt" five-color ink technique. The gradual change of ink color from moist to dry in Wang Xizhi's "Mourning and Chaos Scroll" has completely dissolved the emotional tension carried by the "five colors of ink" on the systematic quantification map.

Secondly, the construction of the existing calligraphy database implies a dual cultural hegemony, creating a "valuing the present over the past" fault in the temporal dimension and creating a "Chinese character centrism" veil in the spatial dimension. According to the digital team of the Palace Museum, in the current training focus of mainstream systems, the distribution of calligraphic styles is seriously unbalanced: Tang regular script accounts for a large proportion, seal script is scarce, and Han bamboo slips and silk manuscripts account for less than one percent. The composition of The Times is distorted: the proportion of modern and contemporary works is also very large, and the digitalization rate of the bronze inscriptions of the three dynasties is less than 5%. The absence of ethnic scripts: Non-Chinese calligraphy such as Khitan and Western Xia scripts have not been included at all. This kind of data bias directly leads to cultural blindness in the evaluation system. Tests show that the artificial intelligence recognition accuracy rate of the works of Deng Shiru, a master of Qing Dynasty seal script, is only 50%, which forms a huge gap with his status in the history of calligraphy. Even more absurdly, when inputting the Dongba script "Divine Path Map" scroll, a certain commercial system misjudged it as "children's graffiti with chaotic brushstrokes", exposing the cognitive violence of algorithms against diverse calligraphic civilizations. The deep-seated problem stems from the technological alienation of data collection. Most systems adopt the "font box + centroid point" combination algorithm for single-character cutting. This preprocessing method based on the square structure of Chinese characters is completely unsuitable for handling the rhombus architecture of Nushu characters or the pictorial symbols of Dongba characters. The latest research by the Dunhuang Academy has found that the existing algorithms have an extremely high error rate in recognizing asymmetric character shapes, seriously hindering the digitalization process of minority calligraphy.

Finally, the core contradiction of the current AI evaluation system lies in imposing the standardized thinking that originated in the Enlightenment movement on the art of calligraphy. The mainstream systems generally adopt the industrial weighted model of "stroke standardization $\times 0.4+$ structural

proportion ×0.3+ layout ×0.3". This quantitative index system causes three major distortions: it deconstructs the dynamic writing process into discrete parameters, making it impossible to perceive time-sensitive brushwork features such as" house leak marks "and" cone drawing sand ". The freehand brushwork of Lin Sanzhi's "incorporating painting into calligraphy" scored as a negative indicator in the systematic review. The "golden character formation rate" in Qi Gong's calligraphy is simplified to a numerical worship of 0.618, ignoring its rhythmic change of "tight at first and then loose". The comparative experiment conducted by the Central Academy of Fine Arts is highly convincing: when the AI system was asked to evaluate the same "Memorial to Nephew", the score from a purely technical perspective was only 72 points, while the comprehensive model that included the analysis of the brushwork pressure curve and writing speed gave 89 points. This proves that the current evaluation system lacks the ability to capture the emotional trajectory of "unstoppable momentum and uncontrollable momentum". What is even more alarming is the educational infiltration of algorithmic black boxes. A certain intelligent calligraphy classroom system has transformed Zhao Meng's instruction that "the capped characters have been passed down through the ages and the brushstrokes have remained unchanged throughout the ages" into 26 quantifiable structural parameters, resulting in a "parametric convergence" phenomenon in students' works - tests show that the styles of students' work using this system are diverse. [13-14]

4. Conclusion and Prospects

This study reveals the triple bias trap of the artificial intelligence calligraphy education evaluation system, and these structural flaws are reshaping the evaluation system of calligraphy education. The current training set of AI systems presents a "center-edge" power map. For instance, Tang regular script accounts for 67%, while seal and clerical scripts are less than 5%. The database of calligraphy in ethnic minority languages is almost blank. This kind of data colonialism has led to the algorithm's recognition rate for Deng Shiru's seal script being only 43%, while the misjudgment rate for Dongba's "Divine Road Map" is as high as 68%. What is more serious is that there is implicit cultural screening in the data annotation process - the Dunhuang Academy found that the error rate of artificial intelligence for asymmetrical characters is 3.2 times that of regular script. This is essentially imposing the aesthetic standards of the square structure of Chinese characters on the diverse calligraphic civilizations. The mainstream evaluation model "Stroke standardization ×0.4+ structural proportion ×0.3" is essentially a product of industrial standardization, and its quantitative index system fundamentally conflicts with the spatio-temporal characteristics of calligraphy art. Experiments conducted by the Central Academy of Fine Arts show that the pure technical dimension score makes the score of "Memorial to Nephew" 17 points lower than that of the comprehensive model, proving that the algorithm cannot capture the "unstoppable" emotional trajectory. What is even more alarming is that the "AI Calligraphy Private School "system of Tsinghua University has deconstructed Zhao Meng's brushwork into 26 parameters, resulting in a 47% decline in the diversity of students' work styles and the formation of a new type of digital museum The artificial intelligence evaluation system is reconstructing the power structure of calligraphy education through the closed loop of "intelligent early warning - data tracking - resource allocation". In the "dual-teacher model" implemented by a certain middle school in Shenzhen, artificial intelligence takes on 60% of the standardized teaching, while teachers gradually become the executors of algorithm output. This technological reliance has led to the transfer of aesthetic judgment power - a certain commercial system has determined Lin Sanzhi's blank brushstrokes as a negative indicator, while 30% of teachers are worried that AI will affect traditional inheritance. More covertly, the algorithmic black box is creating a new type of cultural hegemony: when artificial intelligence marks Huai Su's cursive calligraphy as "error", it is essentially reconstructing the boundaries of artistic freedom through code discipline.

The future of calligraphy education should be a place where algorithmic precision and humanistic warmth coexist. Only by making artificial intelligence an "archaeological hammer" rather than a "scalpel", while safeguarding cultural genes such as "house leakage marks" and "cone painting sand",

and activating the aesthetic possibility of "human-machine co-creation", can we maintain the authenticity of calligraphy as a "heart painting" in the digital torrent. This requires technology developers, educators and cultural scholars to form a "triangular tension", and write a new chapter of cultural self-awareness in the intelligent era on the three dimensions of data justice, algorithm transparency and artistic freedom.

References

[1] Chen Fengrun. An Analysis of the Legal Dilemmas in Regulating Algorithmic Discrimination. Legal Research, 2022, (3), 45-58.

[2] Fang Shishi. Research on Algorithmic Bias in Platform Media: An Analysis Based on the Facebook News Feed Event. Journal of Communication Studies, 2016, 23(4), 78-92.

[3] Han Shipeng. The Regulatory Dilemmas and Paths of Algorithmic Price Discrimination. China Law, 2022, (2), 132-146.

[4] Liu Pei. Research on the Manifestations and Classification of Algorithmic Discrimination. Modern Law, 2019, 41(5), 110-125.

[5] Ni Qinqin. Potential Risks and Prevention of Algorithmic Discrimination in Intelligent Education. Educational Research, 2022, 43(7), 89-102.

[6] Qu Liang. Multi-Dimensional Paths for the Governance of Algorithmic Fairness: Synergy of Technology, Law, Enterprises and Society. Management World, 2023, 39(1), 55-70.

[7] Wang Zhaoyi. Research on the Criticality of Temporal Data in Dynamic Calligraphy Stroke Recognition (Research Report No. CU-CALL-2023-11). Institute of Artificial Intelligence, Jiangnan University. 2023.

[8] Yi Weifeng. The Generation Mechanism and Intervention Strategies of Human Factors in Algorithmic Bias. Philosophical Studies, 2021, (6), 34-47.

[9] Zhang Tao. The Definition and Categorization of Algorithmic Bias: Based on the Perspective of "Homophily". Social Sciences in China, 2020, (4), 120-136.

[10] Agarwal, S.et al. Bias mitigation in AI-driven mobile health solutions: A four-step framework. Journal of Medical Systems,2021, 45(8), 102.

[11] Datta, A. et al. Transparency in algorithmic decision systems: Foundations for auditing. Proceedings of the IEEE Symposium on Security and Privacy, 2017,33–47.

[12] Dellunde, P.et al. Philosophical and computational perspectives on algorithmic fairness. Artificial Intelligence Review, 2023,56(Suppl 1), 1–24.

[13] Ferre, J. V.Data donation ethics: Addressing bias in medical AI through inclusive datasets. Nature Medicine,2023, 29(4), 867–871.

[14] Galaz, V.et al. Algorithmic bias in natural resource management: Amplifying inequalities in agriculture and forestry. Global Environmental Change,2021, 69, 102282.