

SUNDAY AKINFOLARIN MAKINDE

folarinmak@yahoo.com

ORCID; 0009-0007-8197-8607

Citation: Makinde, S. A; *Artificial Intelligence (AI) in Musical Performance Assessment: A Preliminary Inquiry into Objectivity and Pedagogical Implications*

ARTIFICIAL INTELLIGENCE (AI) IN MUSICAL PERFORMANCE ASSESSMENT: A PRELIMINARY INQUIRY INTO OBJECTIVITY AND PEDAGOGICAL IMPLICATIONS

Abstract

Human judgment continues to be the primary method of assessing musical performance in higher education. Although experience-based, this practice is typically inconsistent and mostly dependent on human subjectivity. The concern, therefore, is regarding fairness, standardization, and the balance between technical precision and expressive interpretation in performance assessment. In light of these challenges, this paper explores a preliminary pathway to establishing Artificial Intelligence (AI) as a complementary aid in musical performance assessment. It pays particular attention to Nigerian musical traditions, where the act of performance is inseparable from personal and communal identity. Instead of encouraging absolute implementation, it reflects on ways in which AI could facilitate structured objectivity via the assessment of quantifiable presentation parameters like pitch accuracy, timing, and dynamics. It also examines the pedagogic assumptions of such integration by questioning how digital tools may risk compromising empathy, cultural literacy, and interpretative agency. This paper argues that while AI may enhance consistency, it cannot replicate the human capacity for emotional and contextual understanding. The use of AI for assessing music performance in higher education should be in a supportive capacity, rather than displacing the role of human judgment.

Keywords: Artificial Intelligence, Musical Performance, Education, Assessment, Tertiary Institutions

Introduction

The assessment of musical performance within tertiary institutions has long been rooted in human judgment, through years of experience, training, and cultural understanding. For generations, music educators have relied on careful listening, interpretative sensitivity, and pedagogical wisdom to evaluate students' performances (Okunbor & Alordiah, 2025; Nzewi, 2025). This approach, though deeply valuable, is not without its challenges, as differences in opinion among assessors, variations in individual expectations, and the inherently subjective nature of musical expression often lead to inconsistencies in grading and evaluation outcomes (Wesolowski, Wind & Engelhard Jr, 2016). Such inconsistencies are difficult to address without first establishing a clear understanding of what assessment itself is meant to achieve. Payne et al. (2019) define assessment as "the collection, analysis, interpretation, and applied response to information about student performance...in order to make educational decisions resulting in continual improvement." (p.43). What this definition makes plain is that assessment is not a singular act of judgment but a purposeful, multi-stage process directed toward growth. Therefore, the inconsistencies observed in musical performance evaluation can lead to a potential failure of the very process that is meant to support student development.

Traditionally, musical performance in Nigeria is not merely a technical display but a lived and cultural experience. It embodies emotion, identity, and communal meaning (Okechukwu, 2025; Okujeni, 2024; Abiodun, 2018). Assessment of musical performance, particularly within the Nigerian tertiary institution context, goes beyond correctness of notes and rhythm; it involves recognizing intention, effort, stylistic awareness, musical creativity, and expressive depth. These elements are not easily quantifiable, yet they are central to meaningful musical interpretation. Consequently, the reliance on human assessors, despite its limitations, has remained the cornerstone of performance evaluation in tertiary institutions.

However, new developments in technology have begun to challenge this long-standing tradition. Artificial Intelligence (AI), particularly within the field of Music Information Retrieval (MIR), has demonstrated the capacity to analyze musical data with remarkable precision (Dwivedi, Ganguly & Haragopal, 2023). Through computational models, AI systems measure pitch

accuracy, timing consistency, dynamic variation, and other performance-related features. These capabilities present a compelling possibility, showing that AI could assist in introducing a degree of objectivity into performance assessment, thereby addressing concerns of inconsistency and bias (Bai, 2025).

The aforementioned possibility must be approached with caution as the introduction of AI into musical assessment raises important pedagogical questions. This is because, while machines may offer consistency, they lack the human qualities that define musical understanding, such as empathy, cultural awareness, and emotional intelligence (Hossain, 2025). For instance, a student's performance is often shaped by context. There is a case of nervousness during an examination, gradual improvement over time, or a sincere attempt to convey meaning beyond technical perfection. A human assessor can perceive these subtleties and respond with encouragement or constructive interpretation. AI, by contrast, operates within predefined parameters and lacks the capacity to interpret intention or emotional perceptions.

The central concern of this paper is focused on the contention between objectivity and human sensibility forms. On one hand, there is a clear need to address inconsistencies in assessment practices. On the other hand, there is a risk that an overemphasis on measurable criteria may diminish the expressive and humanistic dimensions of music education. In attempting to resolve one problem, there is the danger of creating another by reducing music to only what can be calculated, rather than what can be felt and understood. It is within this context that the present study positions itself. Rather than advocating for the replacement of traditional assessment methods, it offers a preliminary reflection on the role AI might play as a complementary tool. The aim is not to promote technological dominance, but to explore whether a careful and limited integration of AI could support existing practices without undermining their foundational values.

This inquiry is guided by a balanced perspective that acknowledges both the promise and the limitations of Artificial Intelligence. While innovation should not be resisted outright, it must be approached with discernment, ensuring that the adoption of new tools does not come at the expense of the very qualities that give music its meaning. In the context of tertiary education, where the goal is not only to assess but also to nurture and develop musicologists, the human element must remain central.

Methodology

This study adopts a qualitative, conceptual approach to examine the potential role of Artificial Intelligence (AI) as a complementary tool in musical performance assessment. Rather than relying on empirical data collection or experimental design, the paper is grounded in critical reflection, drawing upon existing literature in music education, Music Information Retrieval (MIR), and performance studies. The methodology is informed by a tradition-oriented pedagogical perspective, shaped by long-standing practices in musical training and assessment within tertiary institutions. The author draws on over three decades of experiential knowledge of music teaching and evaluation within the Nigerian tertiary context, combined with a synthesis of scholarly sources. Such an approach is consistent with conceptual and position-based research, where the aim is to interrogate emerging ideas, identify gaps, and propose informed viewpoints rather than to test hypotheses.

Key themes, such as objectivity, subjectivity, empathy, and cultural context, are examined through a synthesis of related academic works and contextual understanding of Nigerian musical traditions. This allows for a balanced consideration of both technological possibilities and pedagogical realities. It is important to note that the scope of this study is preliminary. The intention is not to provide definitive conclusions or implementable models, but to open a scholarly conversation on the implications of integrating AI into musical performance assessment. As such, the findings should be understood as exploratory and interpretative, offering a foundation for future empirical research.

Conceptual Framework

The discussion of Artificial Intelligence (AI) in musical performance assessment requires a careful examination of five interrelated areas: traditional approaches to performance evaluation, the development of Music Information Retrieval (MIR), machine learning and performance analysis, the conceptual understanding of objectivity and subjectivity in pedagogy, and the pedagogical implications of AI integration. This section situates the present inquiry within these areas, offering a critical and reflective synthesis rather than a purely technical review.

Traditional Approaches to Musical Performance Assessment

Historically, musical performance assessment has been grounded in expert listening and interpretative judgment. Educators assess not only technical accuracy but also expressive qualities such as phrasing, articulation, tone production, and emotional communication (Hossain, 2025). These elements are often evaluated holistically, with assessors drawing upon years of training and performance experience. In many tertiary institution settings, assessment practices involve a panel of music lecturers, student recitals, and continuous evaluation throughout the semester, with multiple assessors contributing to a final grade. While this system allows for a diversity of perspectives, it also introduces variability. Studies have shown that inter-rater reliability in music performance assessment can be inconsistent, particularly when evaluators prioritize different aspects of performance (Wesolowski, 2016). Sometimes evaluators exhibit bias based on their areas of specialization or personal preferences.

These challenges of subjectivity and bias are further compounded in culturally specific contexts, where assessment criteria extend beyond Western technical conventions. In Nigeria, music is not solely an artistic product but a social practice embedded in ritual, storytelling, and communal participation (Authority, 2025; Olajide, O. A. & Nweke, 2025; Nwokolo & Efurhievwe, 2025). Evaluation, therefore, often considers authenticity, cultural appropriateness, and the performer's connection to tradition during performance. Such criteria often appear subjective and therefore resist standardization to highlight the importance of contextual understanding.

Emergence of Music Information Retrieval (MIR)

The field of Music Information Retrieval (MIR) focuses on extracting meaningful information from audio signals, enabling tasks such as pitch detection, tempo estimation, and timbral analysis (Peeters et al., 2025). Techniques such as Mel-Frequency Cepstral Coefficients (MFCCs), chroma features, and spectral analysis have become standard tools for representing musical data in machine-readable form. These representations allow AI systems to process and analyze large volumes of audio data with efficiency and consistency. In the context of performance analysis, MIR enables the quantification of aspects that were previously assessed qualitatively. For example, deviations in timing can be measured with millisecond precision, and pitch accuracy can be evaluated against reference frequencies.

Indeed, Peeters et al. (2025, p.2) confirm that MIR systems are designed to target "the estimation of attributes related to music notation: pitch, dominant-melody, multi-pitch, transcription, sequence of chords, key/mode, positions of beats/downbeats, tempo/meter, global structure (verse/chorus positions) or lyrics transcription" demonstrating the field's capacity to operationalise what has historically been left to human judgment. However, it is important to recognize that MIR primarily captures measurable features. While these features are valuable, they do not fully represent the expressive and interpretative dimensions of performance. This limitation becomes particularly significant when considering the role of AI in educational contexts.

Machine Learning and Performance Analysis

Machine learning techniques have further expanded the capabilities of MIR systems. Supervised learning models can be trained on labeled datasets to classify performances according to style, performer identity, or quality (Li, 2024). Deep learning approaches, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated strong performance in modeling complex musical patterns. More recently, transformer-based architectures, such as those underpinning models like wav2vec 2.0 and MusicBERT, have further advanced the field. As Zaman et al. (2025) note, transformers "have been successfully adapted for audio detection tasks in recent years, showing promise for tasks such as speech recognition, audio event classification, and environmental sound analysis" (p.4), achieving state-of-the-art results in audio classification and automatic music transcription tasks. These developments are central to any current appraisal of AI capabilities in music analysis. These models enable end-to-end analysis, where raw audio input can be transformed into evaluative outputs without extensive manual feature engineering. Such developments have led to the emergence of automated systems capable of providing feedback on performance accuracy and consistency.

Despite these advances, the application of machine learning in performance assessment remains limited by data availability and representational bias. Many datasets are derived from Western classical music, raising concerns about the generalizability of models to other musical traditions. This limitation is of great concern in the Nigerian context, where musical structures and performance practices may differ significantly from those represented in training data.

Objectivity and Subjectivity in Assessment

The concept of objectivity in assessment is often associated with consistency, fairness, and transparency. In tertiary institutions, objective measures are valued for their ability to reduce bias and ensure equitable evaluation. AI systems, with their capacity for standardized analysis, appear to align with these goals. However, in the context of musical performance, the pursuit of objectivity must be carefully examined. Music is inherently expressive, and its evaluation involves interpretative judgment; therefore, subjectivity is not actually a limitation but an essential component of musical understanding. The challenge, therefore, is not to eliminate subjectivity but to manage it responsibly. Human assessors bring empathy, contextual awareness, and pedagogical sensitivity to the evaluation process. These qualities enable them to recognize growth, encourage creativity, and adapt their judgments to individual circumstances. AI, in contrast, operates within predefined parameters and lacks the capacity for such adaptive interpretation. While it can enhance consistency, it may also impose rigid criteria that overlook meaningful variations in performance. This contention emphasizes the importance of viewing AI as a complementary tool rather than a replacement for human judgment.

A Conceptual Overview of Pedagogical Implications

The integration of AI into musical performance assessment raises important pedagogical questions at the conceptual level. On one hand, AI can provide immediate and detailed feedback, supporting student learning and practice. On the other hand, an over reliance on machine-generated evaluation may shift the focus of education toward measurable outcomes at the expense of expressive development. From a tradition-informed perspective, any introduction of AI must be approached with caution and intentionality, serving to enhance rather than replace the humanistic values that underpin music education.

Complementarity suggests a relationship in which two systems, human and artificial, function together while retaining their distinct strengths. In this context, AI may be positioned as a technical assistant capable of providing consistent and measurable feedback, while the human assessor remains the primary interpreter of musical meaning. The consideration of Artificial Intelligence (AI) as a complementary tool in musical performance assessment must begin with a clear philosophical position: AI is not to replace the human assessor but to assist, support, and

possibly refine certain aspects of the evaluation process. This distinction is crucial, particularly within tertiary institutions where assessment is not merely about grading but about guiding artistic development.

AI systems can accurately measure pitch deviation, tempo consistency, and dynamic range. These are areas where human perception, though refined, may vary due to fatigue or bias. By providing a stable analytical baseline, AI could support assessors in making more informed decisions. However, such assistance must remain secondary. The final evaluative authority should reside with the human educator, whose judgment is shaped by experience, cultural understanding, and pedagogical responsibility.

One of the motivations for introducing AI into assessment is the perceived need to reduce subjectivity. In many institutions, students often question grading fairness, particularly when different assessors produce divergent results. AI appears to offer a solution through standardization. Yet, it must be emphasized that subjectivity in music is not entirely undesirable. As Malloch and Trevarthen (2018) argue, music is a form of human action that involves interpretation, intention, and expression. To remove subjectivity entirely would be to misunderstand the nature of musical practice. What AI can offer, therefore, is not the elimination of subjectivity but its moderation. By providing objective reference points, it may help to balance human judgment rather than replace it. In this way, assessment becomes a dialogue between measurable accuracy and interpretative insight.

A major limitation of AI lies in its inability to perceive emotional depth and cultural meaning. A performance may be technically imperfect yet profoundly expressive and culturally authentic. Human assessors, especially those grounded in local traditions, can recognize these qualities. They understand when a performer is communicating effectively within a cultural framework, even if certain technical elements are not perfectly executed. AI, however, lacks this sensitivity. It evaluates based on patterns and parameters, not lived experience. As such, there is a risk that AI-assisted systems may undervalue performances that deviate from standardized norms but are culturally meaningful.

The introduction of AI into assessment practices may also influence how music is taught and learned. If students become aware that their performances are being evaluated by systems that

prioritize measurable accuracy, they may begin to focus excessively on technical perfection at the expense of expressive development. This shift could have long-term consequences for musical creativity and individuality. Music education, particularly at the tertiary level, is not only about producing technically competent performers but also about nurturing artistic voices. The risk of a narrowing curriculum, one that privileges what can be measured over what can be felt, must be actively resisted by educators who adopt AI tools.

Educators must therefore consider how AI tools shape student behavior. If used carelessly, such tools may encourage conformity and discourage experimentation. If used wisely, however, they may support disciplined practice while leaving room for creative exploration. The distinction lies not in the technology itself but in the intentionality of its deployment within the pedagogical relationship.

A Tradition-Informed Approach to Integration

The adoption of AI from a tradition-informed perspective, must be approached with restraint and discernment. Long-standing pedagogical practices should not be discarded in favor of technological efficiency. Rather, AI should be carefully integrated in ways that respect existing values. This may involve, using AI for practice and self-assessment, rather than final grading, combining AI feedback with human commentary, and ensuring that cultural context remains central to evaluation criteria. Such an approach acknowledges the potential benefits of AI while safeguarding the human essence of musical education.

Conclusion

Preliminary inquiry into the possibility of introducing Artificial Intelligence (AI) as a complementary tool in the assessment of musical performance within tertiary institutions is the focus of this paper. Beginning from the recognition of persistent subjectivity and inconsistency in traditional assessment practices, the discussion has explored whether AI offers structured support without undermining the humanistic foundations of music education. The central finding of this inquiry is not simply that AI has limitations, though those limitations are real and significant, but that the terms on which AI is introduced into musical education will determine whether it enriches or impoverishes the discipline. AI holds genuine potential in enhancing the consistency and transparency of technical assessment, particularly in the measurement of pitch

accuracy, timing, and dynamics. Yet these gains are meaningful only within a framework that preserves human interpretive authority. The paper thus argues for a model of complementarity, in which, AI provides an analytical baseline while the human assessor retains final evaluative responsibility.

Within Nigerian and broader African contexts, this argument is sustainable. Music in these traditions is not merely a technical artefact but a vehicle for cultural identity, communal memory, and expressive meaning. Any assessment framework that reduces performance to what can be computationally measured risks severing music from the very values that animate it. Human assessors, shaped by tradition and cultural knowledge, remain indispensable custodians of this dimension of musical meaning. From a pedagogical standpoint, the integration of AI must therefore be approached with caution. While it may support learning and provide valuable technical feedback, it should not assume assertive control over evaluation processes. Instead, AI should function as a supplementary tool, one that enhances, rather than replaces, human judgment.

The future of musical performance assessment lies not in choosing between tradition and technology, but in finding a principled balance between them. As institutions consider the adoption of AI, it is essential to preserve the central role of empathy, cultural awareness, and interpretative insight in music education. Future research should move beyond conceptual inquiry toward empirical validation: pilot studies involving AI-assisted feedback in Nigerian and other African tertiary music programmes, comparative analyses of assessor agreement with and without AI support, and the development of culturally inclusive evaluation frameworks would each constitute meaningful and necessary contributions to this emerging field. Only through such evidence-based, culturally grounded innovation can technology be embraced without compromising the soul of musical practice.

References

- Abiodun, F. S. (2018). Teaching Art Music in Nigerian Schools: The academic, technical and social implications. *African Journal of Teacher Education*, 7(1).
<https://doi.org/10.21083/ajote.v7i1.3582>

- Authority, A. O. (2025). From roots to innovation: Thematic metamorphosis and transformation of Nigeria's modern art music scene. *Awka Journal of Research in Music and the Arts (AJRMA)*, 18(1). 106-113
- Bai, Y. (2025). Exploring the role and impact of artificial intelligence in personalized foreign language teaching. *Discover Artificial Intelligence*, 5(1), 1-21. <https://doi.org/10.1007/s44163-025-00546-9>
- Dwivedi, D., Ganguly, A. & Haragopal, V. V. (2023). Contrast between simple and complex classification algorithms. In *Statistical Modeling in Machine Learning* (pp. 93-110). Academic Press.
- Hossain, A. (2025). Emotional and Artificial Intelligence: Narrative of human and machine relation. *Transactions on Engineering and Computing Sciences*, 13(05). <https://doi.org/10.14738/tmlai.1305.19324>
- Li, T. (2024). Optimizing the configuration of deep learning models for music genre classification. *Heliyon*, 10(2). <https://doi.org/10.1016/j.heliyon.2024.e24892>
- Malloch, S., & Trevarthen, C. (2018). The Human Nature of Music. *Frontiers in psychology*, 9, 1680. <https://doi.org/10.3389/fpsyg.2018.01680>
- Nwokolo, T. & Efurhievwe, M. A. (2025). Voices of tradition: A study of Ekobe music, Enuani culture, the art of masterful singing. *Abuja Journal of Humanities*, 6(1), 152-162.
- Nzewi, O. D. (2025). *African performance epistemology: fundamental indigenous creative theories and philosophies for application in contemporary music education* (Doctoral dissertation, University of Limerick).
- Okechukwu, C. V. (2025). The intersection of music and education: Promoting cultural identity in Nigeria. *ISA Journal of Engineering and Technology (ISAJET)*, 2 (1), 3049-1843.
- Okujeni, R., Zamani, Y., Akpe, S. & Okocha, D. O. (2024). Reclaiming music as a communicative instrument of social change in Nigeria. *Nasarawa Journal of Multimedia and Communication Studies*, 6(1), 40–55.
- Okunbor, D. & Alordiah, C. O. (2025). Toward a Critical Pedagogy of Music Assessment in Nigeria: Unpacking the Hegemonic Structures of Music Education and their Implications for Assessment. *NIU Journal of Humanities*, 10(2), 215-226. <https://doi.org/10.58709/niujhu.v10i2.2221>
- Olajide, O. A. & Nweke, F. (2025). Exploring an interdisciplinary framework for integrating indigenous musical traditions in Nigerian secondary schools. *Awka Journal of Research in Music and the Arts (AJRMA)*, 18(1). 39-44.
- Payne, P. D., Burrack, F., Parkes, K. A. & Wesolowski, B. (2019). An emerging process of assessment in music education. *Music Educators Journal*, 105(3), 36-44. <https://doi.org/10.1177/0027432118818880>

- Peeters, G., Rafii, Z., Fuentes, M., Duan, Z., Benetos, E., Nam, J. & Mitsufuji, Y. (2025). *Twenty-five years of MIR research: Achievements, practices, evaluations, and future challenges*. arXiv. <https://doi.org/10.48550/arXiv.2511.07205>
- Wesolowski, B. C., Wind, S. A., & Engelhard Jr, G. (2016). Examining rater precision in music performance assessment: An analysis of rating scale structure using the multifaceted Rasch partial credit model. *Music Perception: An Interdisciplinary Journal*, 33(5), 662-678. <https://doi.org/10.1525/mp.2016.33.5.662>
- Zaman, K., Li, K., Sah, M., Direkoglu, C., Okada, S. & Unoki, M. (2025). Transformers and audio detection tasks: An overview. *Digital Signal Processing*, 158, 104956. <https://doi.org/10.1016/j.dsp.2024.104956>