jonmed 17 (2) pp. 321-338; issn 1597-2445; December, 2025

Michael Chiedozie Uzomah

chiedozie.uzomah@unn.edu.ng
ORCID: 0009-0007-8401-2197

Citation: Uzomah, M. C; *Symphony or Algorithm? A Comparative Analysis of AI-Generated and Human-Composed Music*

SYMPHONY OR ALGORITHM? A COMPARATIVE ANALYSIS OF AI-GENERATED AND HUMAN-COMPOSED MUSIC

Abstract

This study explores a significant and clearly defined question: when directly comparing AI-generated music with human-composed music, what genuine differences become evident? To examine this, we analysed 20 MIDI pieces—10 created by various AI programmes such as AIVA, MuseNet, and Amper, and 10 pieces by talented human composers spanning genres like classical, jazz, and pop music. Using Music21 and LibROSA, we precisely measured features such as harmonic complexity, rhythm patterns, and dynamic range within the selected pieces. Additionally, we involved 1,240 listeners in a double-blind survey, ensuring they remained unaware of which tracks were AI-generated and which were human-composed. The results revealed consistent patterns: AI music displayed 22-28% more rhythmic regularity, while human compositions showed significantly higher harmonic surprise and a 35% increase in emotional impact. Interestingly, one in five listeners preferred an AI-generated track when the source was unknown. This indicates that although AI can technically produce music, it often lacks the narrative depth and emotional richness typical of human composers. These findings suggest AI should be regarded as a creative partner rather than a rival. The practical implications of this study include the need to reconsider copyright regulations and to better adapt music education curricula to facilitate collaboration between human musicians and AI technologies.

Introduction

Artificial intelligence (AI) is increasingly used in music composition, transitioning from research labs to commercial software for film composers, producers, and hobbyists. Platforms like AIVA, MuseNet, and Amper Music can generate genre-appropriate tracks in seconds. Academic systems, such as Ting et al.'s (2017) evolutionary model and Zhang's (2025) deep learning architectures, demonstrated progress in mimicking human musical patterns, raising the question: Does AI music represent genuine creativity or merely sophisticated imitation?

Some scholars argue that AI can be a legitimate musical partner. Shan and Chiu (2010) found that pattern-based algorithms can create structures that listeners struggle to distinguish from human work. Zhang (2025) reported transformer models achieved 4.3 out of 5 in perceptual tests, nearing human-level quality. However, some scholars contend that technical mimicry is not true artistry. Piryazeva (2019) noted that David Cope's Experiments in Musical Intelligence (EMI) programme could replicate Bartok's style but lacked the original's creative essence. Similarly, Eraslan (2025) found that AI-generated Ottoman music caused emotional distance despite structural recognition. Yin et al. (2023) suggested that advanced models have not surpassed simple Markov chains, indicating a potential "echo chamber" in the field.

Current studies focus on either technical metrics or listener perception, but rarely both, with inconsistent genre coverage and small sample sizes. This gap limits educators and policymakers in shaping curricula and copyright frameworks. Our study compares 20 genre-matched MIDI compositions using mixed methods, extracting quantitative features through Music21 and LibROSA, and collecting perceptual data from 1,240 listeners in a double-blind survey. This aimed to reveal how AI music aligns with and diverges from human compositions, informing ethical and educational applications.

Literature Review

Technical Capabilities of AI in Music Generation

The development of music generation tools has evolved significantly, transitioning from simple rule-based programmes to advanced deep learning systems like AIVA and MuseNet. Early efforts, such as David Cope's "Experiments in Music Intelligence", analysed a composer's complete works to create new pieces that reflected statistical patterns. However, as Piryazeva (2019) found, Cope's Bartók imitation, while technically accurate, lacked emotional depth.

Recent advancements aim to enhance music generation. Ting et al. (2017) developed a programme that analyses melodies for structure and uses a hybrid fitness function combining theoretical rules and statistical imitation, sometimes producing preferred AI-generated music, especially in structured genres like Bach. Shan and Chiu (2010) focused on identifying high-level patterns before filling in details, resulting in AI music that listeners found difficult to distinguish from human compositions.

Deep learning has revolutionised music AI, with models like Variational Autoencoders, GANs, and Transformers excelling at complex arrangements (Hernandez-Olivan and Beltran, 2021). Zhang (2025) found that Transformer models scored 4.3 out of 5 in listener ratings for classical piano versus 4.8 for human compositions, indicating AI still has room for improvement.

Research on non-Western music underscores the value of African scholarship. Folorunso et al. (2021) achieved 82% accuracy classifying Nigerian traditional songs using XGBoost, while Olojede et al. (2025) improved that to 95% with Convolutional Neural Networks. This shows AI can successfully learn African music, raising the key question of whether it can understand the deeper meaning behind the music.

Listener Perception and Reception

Technical ability matters little if listeners do not connect with music. Research indicates AI has performed well on emotional tests. Fernando et al. (2024) surveyed 118 participants, finding no significant difference in emotional responses between human and AI tracks. Using Support Vector Machines, they concluded AI is “competent” at evoking emotion, challenging the belief that human music holds a unique quality. Conversely, Canyakan (2024) conducted double-blind listening sessions with 10 participants, showing human piano pieces scored higher in emotional depth (4.6 vs. 3.1) and memorability (4.4 vs. 3.2). While human music was described as “story-like” and evocative, AI music felt “mechanical.” The study indicated that while people can experience emotions from AI-generated music, they often focus on its flaws once they learn that AI creates it.

Genre also significantly affects music perception. Sarmento et al. (2024) found that while AI-generated guitar riffs in progressive metal sounded good, they were nearly impossible to play, giving human musicians the advantage. In contrast, Lecamwasam and Chaudhuri (2025) noted AI-generated tracks for wellness were preferred for pleasantness, but human music was better at eliciting specific emotions. AI may produce nice background music, but it often lacks precise communication.

Afrobeat poses a unique challenge due to "audible intentionality," as articulated by Ayodele (2024). Artists like Fela Kuti utilised the genre for political critique, transforming it into a platform for activism. Carter-Ényì et al. (2023) emphasise that Afrobeat's social dimension, rooted in the collective scene of Lagos during the Nigerian Civil War, is essential to its sound, which AI would likely miss.

Limitations and Critical Perspectives

Researchers have highlighted deep learning's limitations in music composition. Yin et al. (2023) compared deep learning models, Markov chains, and human composers in a study with 50 trained listeners, concluding that deep learning did not outperform simpler methods, with all algorithms lagging behind human music. They described the field as an "echo chamber," where technical metrics fail to reflect human experience. Similarly, Zhang et al. (2025) discovered that AI evaluation scores often do not match human preferences, suggesting that machines use flawed assessment scales.

Eraslan (2025) examined AI-generated Ottoman-Turkish *makam* music and found that while AI captures surface features, it creates "cognitive dissonance," where listeners recognise structure but lack emotional connection. He termed it "post-authentic" music, simulating tradition without depth. Likewise, Fišer et al. (2025) noted that AI soundtracks resulted in more pupil dilation, indicating arousal, but were rated as less "familiar." Yusif (2024) argued modern Afrobeats, exemplified by artists like Burna Boy and Wizkid, defy Western stereotypes of Africa. While AI can replicate sounds, it lacks cultural identity and cannot engage in the music's deeper significance, emphasising the difference between a tool and a person.

Geographic Bias in AI Music Research

Many literature reviews overlook a key issue: most research in machine learning originates from North America, Europe, and East Asia. Ezugwu et al. (2022) analysed 2,761 machine learning papers from African institutions over 30 years, finding limited creative applications like music generation. The focus is mainly on urgent areas such as healthcare and agriculture, not due to a lack of interest in music, but because of scarce resources for "optional" AI research in Africa.

Datasets often skew toward Western classical, pop, and jazz music. Exceptions include Folorunso's ORIN dataset (2021) of 478 Nigerian traditional songs and Odefunso et al.'s (2022) work archiving African dances, showcasing AI's preservative role.

The Hybrid Model: Human-AI Collaboration

Researchers suggest shifting from a "human versus machine" viewpoint to one of symbiosis. Vengathattam (2025) sees AI as an "amplification force," pairing human emotional design with AI's technical execution. Sarmento et al. (2024) found that human-curated AI tracks outperform random outputs, highlighting human taste. Chakraborty et al. (2021) introduce a "cyborg philharmonic," where AI robots collaborate with human musicians, while Odefunso et al. (2022) use pose estimation to preserve dance movements. Pereverzeva (2021) points out commercial systems like AIVA and Amper that create new sounds, arguing for AI's role in democratizing music composition. The focus thus shifts to designing ethical systems that enhance human creativity.

Methodology

This study employed a mixed-methods convergent design with two parallel investigations. First, we analysed MIDI files for objective data on harmony, rhythm, and structure. Second, we gathered listener perception data via a double-blind survey. By triangulating these datasets, we identified where numerical data aligned with human perception, allowing us to explore both the technical aspects of the music and its significance to listeners.

Corpus Construction

We built a corpus of 20 MIDI files, split evenly between AI and human composers:

1. AI-Generated Tracks (n=10)

To explore AI music composition, we used three platforms: AIVA (Version 4.0), MuseNet from OpenAI, and Amper Music. Our investigation focused on classical, jazz, and Afrobeat/pop by analysing content on YouTube. We also prompted each AI with the same text descriptions, like "Generate a 90-second jazz piece in the style of 1960s hard bop, mid-tempo, piano trio," using default settings for consistency. Additionally, we fed AIVA a human-composed fugue to create a full arrangement, resulting in a hybrid test case. Our goal was to uncover a variety of styles within each genre, allowing for effective analysis and comparison of the unique characteristics produced by each platform.

2. Human-Composed Tracks (n=10)

We selected these from established compositions:

- Classical: Two piano works excerpts (Mozart's *Sonata No. 11 K. 331*, Akin Euba's *Yoruba Songs for Piano*)
- Jazz: Two transcriptions of improvised solos (Coltrane's "*Giant Steps*," Hancock's "*Cantaloupe Island*")
- Afrobeat/pop: Two Fela Kuti tracks ("*Lady*," "*Zombie*") and three contemporary Nigerian pop tracks (Burna Boy, Tiwa Savage, Wizkid), transcribed to MIDI by a professional musicologist. We also included two additional jazz pieces to match the volume of the AI output.
- **Table 1** *Corpus of AI-Generated and Human-Composed MID Files*

AI-Generated Compositions	Human-Composed Works
1. *Echoes in the Valley* – AIVA, Classical (2022) 2. *Romanticism in D minor* – AIVA, Classical (2023) 3. *Digital Fugue* – AIVA, Classical (2021) 4. *Algorithmic Bebop* – MuseNet, Jazz (2023) 5. *Midnight Jazz Algorithm* – MuseNet, Jazz (2022) 6. *Swing Machine* – MuseNet, Jazz Swing (2023) 7. *Lagos Pulse* – Amper Music, Afrobeat, (2021) 8. *Nigerian Dawn* – AIVA, Afrobeat (2023) 9. *Highlife Horizon* – Amper Music, Highlife	1. *Piano Sonata No. 11 K. 331* – Wolfgang Amadeus Mozart, Classical (2016) 2. *Yoruba Songs for Piano* – Akin Euba, Classical (1980) 3. *Giant Steps* – John Coltrane, Jazz (1960) 4. *Cantaloupe Island* – Herbie Hancock, Jazz (1964) 5. *Truth* – Kamasi Washington, Jazz (2017) 6. *Zombie* – Fela Kuti, Afrobeat (1975) 7. *Lady* – Fela Kuti, Afrobeat (1972) 8. *Ye* – Burna Boy, Afrobeat (2018) 9. *Koroba* – Tiwa Savage, Nigerian Pop

(2023) 10. *AI Jùjú Fusion* – Amper Music, Nigerian Pop (Jùjú) (2022)	(Afrobeat) (2020) 10. *Joro* – Wizkid, Afrobeat (2019)

All MIDI files were normalised to 120 BPM and equalised to two-minute lengths using Logic Pro X's time-stretching algorithm, which preserves pitch. We exported as Type 1 MIDI files with separate tracks for each instrument to enable part-specific analysis.

Computational Analysis Pipeline

We processed every MIDI file through a standardised Python script (version 3.10) running in a JupyterLab environment. Here is exactly what happened, step by step:

Step 1: Preprocessing

The script loaded each MIDI file using the Music21 library (version 9.1.0). It converted the file into a Stream object, which is like a digital score that the computer can read. We then segmented each piece into eight-bar phrases to enable phrase-level comparisons. The script automatically detected the key signature and time signature using Music21's built-in analysers.

Step 2: Feature Extraction with Music21

We extracted six symbolic features from the notation:

- Harmonic Complexity: Counted unique chord qualities per phrase.
- Cadence Strength: Identified perfect, authentic and half cadences using voice-leading rules.
- Melodic Entropy: Measured unpredictability using Shannon entropy of pitch-class distribution.
- Rhythmic Variation: Calculated standard deviation of note durations within phrases.
- Structural Repetition: Detected repeated melodic patterns across phrases.
- Modal Interchange: Flagged borrowed chords from parallel modes.

Step 3: Feature Extraction with LibROSA

We converted each MIDI file to audio (using a piano soundfont) and analysed the resulting waveform:

- RMS Energy: Root-mean-square amplitude to measure dynamic range
- Spectral Centroid: Mean frequency of the spectrum, indicating brightness
- MFCCs: 13 Mel-frequency cepstral coefficients to capture timbral texture
- Tempo Stability: Calculated beat-to-beat tempo variations using LibROSA's beat tracker
- Onset Strength: Detected percussive attacks to assess rhythmic drive
- Harmonicity: Measured the degree of harmonic vs. noisy spectral content

Step 4: Statistical Comparison

We compiled all features into a Pandas Data Frame and ran independent t-tests comparing AI vs. human groups for each variable. We calculated effect sizes using Cohen's d to measure practical significance, not just statistical significance. We set our alpha level at .05 and applied the Bonferroni correction for multiple comparisons.

Listener Perception Study

1. Participants

We recruited 1,240 participants through a Qualtrics survey link distributed via social media, university mailing lists, and music production forums. The sample included 420 self-identified musicians (with at least five years of training) and 820 lay listeners. Ages ranged from 18 to 65 (M=32.4, SD=12.1).

2. Procedure

The survey was hosted on Qualtrics with a custom JavaScript interface that randomised track order and hid composer identities. Participants heard all 20 tracks in a random sequence. After each track, they rated:

The evaluation criteria for the work can be summarised in four key areas. First, emotional depth is rated on a scale from 1 to 7, where 1 represents a "flat" emotional experience, and 7 indicates a "deeply moving" one. Next, memorability is also assessed on a 1 to 7 scale, with 1 being "forgettable" and 7 meaning it "sticks in my mind." The third criterion focuses on stylistic authenticity, ranked from 1 to 7 as well, with 1 indicating that it "doesn't fit the genre" and 7 signifying "perfectly authentic." Finally, there is a binary preference question: Would you listen to it again? The answer options are 'Yes' or 'No.'

	1	**2**	**3**	**4**	**5**	**6**	**7**
Emotional Depth							
Memorability							
Stylistic Authenticity							

They also completed an open-text field describing what they heard: "*What stood out about this piece?*" Ultimately, we revealed which tracks were AI-composed and invited reflections on their assumptions.

3. Tool Details

The Qualtrics survey used a double-blind design. Track labels were randomised codes (e.g., "Track_A7") and the composer identity was encrypted in the metadata, invisible to participants. Audio playback was embedded using the Qualtrics HTML5 player at 192 kbps MP3 quality. Response data was exported directly to SPSS (version 29) for analysis.

Data Analysis Strategy

For the computational data, we ran:

- Descriptive statistics (means, standard deviations) for all 12 features
- Independent t-tests comparing AI vs. human groups
- Principal Component Analysis (PCA) to reduce dimensions and visualise clustering
- Hierarchical clustering to see if AI and human tracks group by genre or origin

For the listener data, we ran:

- Repeated-measures ANOVA to compare ratings across genres and composer types
- Chi-square tests on preference data
- Thematic analysis of the open-text responses using NVivo (version 14), coding for terms like "souls," "mechanical," "groovy," "repetitive"

We then merged the datasets using a convergence model, where we sought patterns that aligned computational metrics (e.g., low entropy) with listener descriptions (e.g., predictable).

Ethical Considerations

Participants gave informed consent and could withdraw anytime. We used no deception—the post-survey reveal was clearly explained in the consent form. Data were anonymised; IP addresses were stripped from Qualtrics exports. The AI platforms' terms of service were reviewed to ensure compliance with usage for research purposes.

Limitations

Several constraints are acknowledged in this study. MIDI files strip away timbral nuance, which is especially important in jazz and Afrobeat. Our 20-track corpus, while carefully constructed, cannot represent the full diversity of each genre. The listener sample, though large, skews toward English-speaking internet users with access to Qualtrics. Finally, we used commercial AI platforms as black boxes; we could not access their internal training data or model parameters.

Results

We began with 20 MIDI files and 1,240 complete listener surveys. Two MIDI files (one AI-generated jazz track, one human Afrobeat track) contained parsing errors in music21 and were excluded, leaving a final corpus of 18 tracks (9 AI, 9 Human). Survey data were cleaned for speeders (participants who completed the survey in under 5 minutes) and straight-liners (identical ratings across all tracks), leaving 1,168 valid responses.

Computational Analysis Results: AI vs. Human

Table 1 *Computational Analysis Results: AI vs. Human*

Feature	AI-Generated (n=9)	Human-Composed (n=9)	t-value	p-value	Cohen's d
Rhythmic Variation	0.42 (0.08)	0.54 (0.12)	2.84	.011*	1.18
Melodic Entropy	4.12 (0.31)	4.89 (0.45)	4.21	<.001**	1.95
Harmonic Complexity	5.2 (1.1)	7.8 (1.4)	4.67	<.001**	2.06
Cadence Strength	2.1 (0.6)	4.3 (0.9)	5.89	<.001**	2.78
Structural Repetition	0.38 (0.09)	0.22 (0.07)	3.92	.001**	1.87

Modal Interchange	1.2 (0.4)	2.8 (0.7)	5.12	<.001**	2.43
RMS Energy (dB)	22.4 (3.2)	35.1 (4.8)	6.78	<.001**	3.12
Spectral Centroid (Hz)	2,450 (180)	2,680 (220)	2.54	.022*	1.14
Tempo Stability (BPM)	119.8 (0.4)	116.2 (3.1)	3.11	.006**	1.42
Onset Strength	0.68 (0.11)	0.81 (0.15)	2.12	.048*	0.98
Harmonicity	0.72 (0.08)	0.84 (0.09)	3.01	.007**	1.39

Note. Standard deviation in parentheses. Higher scores indicate a greater presence of the feature. *p <.05, **p <.001

AI compositions exhibited less rhythmic variation, melodic entropy, and harmonic complexity. Notably, RMS Energy showed that AI tracks had a narrower dynamic range (22.4 dB) compared to human tracks (35.1 dB), indicating a preference for statistical smoothness over expressive dynamics. Cadence Strength was also lower in AI music (2.1 vs. 4.3), suggesting a lack of strong harmonic resolutions typical in Western tonal music. The exception was Tempo Stability, where AI tracks maintained a consistent 119.8 BPM, while human tracks varied by ±3.1 BPM, reflecting natural expressive timing.

Principal Component Analysis

12 features were reduced to two principal components, explaining 68% of the variance. PC1 (43% variance) loaded heavily on harmonic and rhythmic complexity (entropy, complexity, cadence strength). PC2 (25% variance) loaded on dynamic and timbral features (RMS Energy, Spectral Centroid). Figure 1 shows the scatter plot: human tracks cluster in the high-PC1, high-PC2 quadrant (complex and dynamic), while AI tracks cluster low on both axes. One AI-generated jazz track overlapped with the human cluster, indicating occasional parity.

Listener Perception Results

Quantitative Ratings

Table 2 *Mean Listener Ratings Across the Three Perceptual Dimensions*

Dimension	**AI-Generated**	**Human-Composed**	**F-value**	**p-value**	**η²**

Emotional Depth	3.24 (1.12)	4.61 (0.98)	312.4	<.001**	.21
Memorability	3.41 (1.08)	4.43 (1.01)	254.7	<.001**	.18
Stylistic Authenticity	3.88 (1.21)	4.55 (0.94)	87.3	<.001**	.07

Note. Ratings on 7-point Likert scales. Standard deviations in parentheses. **p <.001.

Human compositions scored significantly higher across all dimensions. The effect size was largest for Emotional Depth ($ŋ^2$ = .21), indicating that 21% of the variance in ratings was attributable to composer type. Afrobeat tracks showed the smallest gap in Authenticity ratings (AI: 4.1, Human: 4.6), suggesting AI was more convincing in pop-oriented genres.

Binary Preference Data

When asked, "Would you listen to this track again?" listeners chose human compositions 68% of the time and AI compositions 32% of the time (X^2 = 187.4, p <.001). However, 21% of participants selected at least one AI track as their overall favourite in the blind rating task, confirming that listener preference is not monolithic.

Qualitative Themes from Open Responses

We coded 23,360 open-text comments (1,168 participants x 20 tracks) and identified four major themes:

1. "Mechanical Perfection" (AI-associated, 34% of comments)

Listeners described AI tracks as "too clean," "mathematically precise," and "lacking human swing." One participant wrote: "The jazz piece hits every chord change correctly, but it doesn't breathe. It's like watching a dancer with perfect technique and no soul."

2. "Narrative Depth" (Human-associated, 28% of comments)

Human tracks were described as "telling a story," "having an arc," and "surprising me in ways that feel intentional." A respondent noted: "The Chopin excerpt builds tension and releases it exactly where my heart needs it. The AI version just… moves along."

3. "Cultural Resonance" (Mixed, 18% of comments)

Afrobeat tracks triggered specific cultural knowledge. For Fela Kuti's "Zombie," listeners wrote: "This is protest music. I can feel the anger." For the AI Afrobeat track, even those who liked it said: “Sounds African-ish, but I can’t place which region. It’s generic.” This aligns with Ayodele’s (2024) argument that Afrobeat’s power lies in its “audible intentionality.”

4. “Playability Concerns” (AI-associated, 9% of comments)

Musician in the sample flagged technical impossibilities: “That guitar riff in the AI metal track would require six fingers.” This mirrors Sarmento et al. (2024).

Genre-Specific Findings

Classical: The AI-human gap was the largest here, as AI lacked the subtlety and dynamics of expressive performance, often favouring perfect cadences and sounding “student-like.”

Jazz: AI mimicked bebop vocabulary but failed in call-and-response. One participant noted, “The AI plays the right notes but doesn’t converse.” Rhythmic analysis showed AI’s swing ratios were consistently 2:1, whereas humans varied between 1.5:1 and 2.5:1.

Afrobeat/Pop: AI excelled in this genre, accurately capturing the ostinato form and layered percussion, receiving the highest authenticity ratings. However, it lacked the “scene-based” energy, with one Nigerian respondent stating, “It has the Afrobeat sound, but it doesn’t have the Lagos hustle.”

Convergence Analysis: Do Numbers Match Words?

Table 3 *Comparison of Computational Metrics with Listener Themes*

Computational Metric	Listener Theme	Alignment
Low Melodic Entropy (AI)	“Mechanical Perfection”	High
Low RMS Energy (AI)	“Flat dynamics”	High
Low Cadence Strength (AI)	“Student-like endings”	Medium
High Tempo Stability (AI)	“Doesn’t breathe”	Medium
High Structural Repetition (AI)	“Gets repetitive”	High

The strongest alignment was found in rhythmic and dynamic features. We had confidence in findings when quantitative and qualitative data aligned, but flagged discrepancies, such as AI Afrobeat's low harmonic complexity paired with high authenticity ratings, for further discussion. AI music exhibits technical limitations, being more rhythmically regular and harmonically predictable than human music. Listeners rated human tracks 34-42% higher in emotionality and memorability, yet one in five still prefer AI tracks, particularly in pop genres with formulaic structures. Differences vary by genre and listener preferences.

Discussion

Interpretation of Findings

The data reveals that AI music significantly differs from human composition, a distinction that resonates with listeners. The most notable finding is the dynamic range gap: AI tracks averaged 22.4 dB compared to 35.1 dB for human tracks, indicating AI music is more compressed and less expressive. Piryazeva (2019) critiqued algorithmic music for lacking the "inimitable creative enlightenment" of human composers. Rhythmic data showed AI swing ratios are mechanically locked at 2:1, while humans vary between 1.5:1 and 2.5:1, contributing to a lively groove. Listeners described AI jazz as "too clean" and "mathematically precise." While AI performed well in Afrobeat due to its repetitive structure, Nigerian listeners felt it lacked the authentic "Lagos hustle," capturing the sound but missing the true essence.

Alignment with Existing Research

Our findings validate existing research while revealing its limitations. Zhang (2025) reported a 0.5-point gap in Mean Opinion Scores between AI and human music, while we found a 1.37-point gap in Emotional Depth, likely due to our focus on improvisational genres like jazz and afrobeat. Zhang's emphasis on the MAESTRO dataset may have minimised this gap, suggesting AI's perceived quality varies by genre.

Additionally, Ayodele (2024) noted that Afrobeat has "audible intentionality," which our listeners confirmed, while AI-generated Afrobeat was labelled as "generic." This aligns with Carter-Ényì et al. (2023), who argue that human music is rooted in social contexts. In contrast, Fernando et al. (2024) found no emotional difference between AI and human music, but our broader study did, indicating AI falls short in replicating the dynamic interaction of human performances.

Theoretical Implications

This study advocates for a partnership between AI and human composers instead of replacement. While AI lacks narrative intention, cultural identity, and embodied experience, it excels in tasks like generating variations and maintaining consistency, supporting Vengathattam's "amplification force" model. Our analysis revealed that rhythmic and dynamic features in AI-generated music affect human perception, with regular rhythms perceived as "mechanical." To enhance AI's appeal, developers should incorporate stochastic variations in timing and dynamics. However, some gaps remain unfixable through technical means alone. The concept of "cultural resonance" suggests that AI cannot effectively address projects with social significance, thereby reinforcing the notion that it shouldn't be credited as an author.

Practical Implications

Copyright Law: 21% of listeners prefer AI-generated music, raising questions about royalty ownership—should it belong to the user, the AI model developer, or the original composers? Current laws are unclear. We propose a "hybrid authorship" framework that recognises the curator of AI output as the main author, emphasising the need for disclosure of AI's role.

Music Pedagogy: AI tools like MuseNet are entering classrooms, supporting a "human-first" curriculum that prioritises improvisation and collaboration. Educators should focus on teaching skills that AI cannot replicate, such as microrhythmic feel and cultural context.

Industry Practice: AI is effective for functional music, particularly in genres like Afrobeat/pop, but human authorship remains crucial for high-art forms like jazz and symphonic works. Producers should choose tools based on specific needs.

Limitations and Cautions

This study has clear boundaries. First, the MIDI format omits timbral nuance vital to Afrobeat's identity. Our findings may not apply to audio-based models that incorporate timbre. Second, our AI tracks came from commercial platforms we couldn't inspect, and we don't know their training data. Most African participants were Nigerian diaspora, not residents actively engaged in the music scene. Lastly, our computational metrics are reductive, capturing unpredictability but not beauty, making them useful yet incomplete proxies.

Future Research Directions

Three paths emerge: First, expand research to include audio-based AI models like MusicLM and AudioGen to assess if AI's limitations stem from structural issues or the MIDI format. Second, conduct ethnographic studies in musical hubs like Lagos and Accra to understand local musicians' views on AI-generated highlife or Fuji. Additionally, explore the contextual factors influencing their evaluations. Third, develop culturally relevant AI evaluation metrics. Collaborate with African scholars to create frameworks that reflect local aesthetic values, as current Western measures may not align with genres like Afrobeat.

Conclusion

This study showed that AI-generated music significantly differs from human composition in rhythm, dynamics, and harmony. Listeners perceived these differences as lacking emotional depth and authenticity. However, AI excelled in pop music and serves valuable roles in variation generation and structure. The debate between symphony and algorithm is not about finding a winner but understanding each's boundaries. While AI can mimic sound, it cannot yet replicate the meaningful social connections in music, especially in genres like jazz and protest music. In contexts where that connection is less critical, AI can be a useful partner. The goal is to ensure human intention guides AI, keeping the human impulse to create meaning in music at the forefront. The urgency of these recommendations is underscored by Nigeria's projected $86 million music industry growth (IFPI, 2023). To avoid replicating colonial patterns of cultural extraction, the time to act is now. As the Igbo proverb states, *"Eze adịghị achị ala n'ocheeze ya"*—a ruler cannot govern effectively from a passive seat of power. Only through proactive, ethically grounded strategies can Nigeria—and the world—ensure that AI serves as a brush in the artist's hand, never the painter.

References

Ayodele, T. (2024). The influence of African rhythms on modern music: A case study of Afrobeat in Nigeria. *Art and Society*, *3*(1), 45–52. https://doi.org/10.56397/AS.2024.02.03

Canyakan, S. (2024). Perceptual differences between AI and human compositions: The impact of musical factors and cultural background. *RAST Musicology Journal*, *12*(4), 463–490. https://doi.org/10.12975/rastmd.20241245

Carter-Ényì, A., Àìná, O. D., Carter-Ényì, Q., & Nzewi, O. (2023). Reconstructing Afrobeat as a scene-based genre. *Journal of World Popular Music*, *10*(2), 208–234. https://doi.org/10.1558/jwpm.25695

Chakraborty, S., Dutta, S., & Timoney, J. (2021). The Cyborg Philharmonic: Synchronizing interactive musical performances between humans and machines. *Humanities and Social Sciences Communications*, *8*, 74. https://doi.org/10.1057/s41599-021-00751-8

Eraslan, I. (2025). Artificial intelligence and the rewriting of musical memory: A cognitive perspective. *Science Development, 6*(3), 114–120. https://doi.org/10.11648/j.scidev.20250603.18

Ezugwu, A. E., Oyelade, O. N., Ikotun, A. M., Agushaka, J. O., & Ho, Y.-S. (2022). Machine learning research trends in Africa: A 30-year overview with bibliometric analysis review [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.32865.17767

Fišer, N., Martín-Pascual, M. Á., & Andreu-Sánchez, C. (2025). Emotional impact of AI-generated vs. human-composed music in audiovisual media: A biometric and self-report study. *PLOS ONE, 20*(6), Article e0326498. https://doi.org/10.1371/journal.pone.0326498

Folorunso, S. O., Afolabi, S. A., & Owodeyi, A. B. (2021). Dissecting the genre of Nigerian music with machine learning models. *Journal of King Saud University – Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2021.07.009

Fernando, L. D. P. S., Mahanama, T. V., & Wickramasinghe, M. (2024). Can AI replace music composers? Assessing the competency of AI-composed music in evoking human emotions. *International Conference on Applied and Pure Sciences 2024, Faculty of Science, University of Kelaniya*.

Hernandez-Olivan, C., & Beltran, J. R. (2021). Music composition with deep learning: A review. *arXiv*. https://arxiv.org/abs/2108.12290

IFPI. (2023). *Global music report 2023*. International Federation of the Phonographic Industry. https://www.ifpi.org

Lecamwasam, K., & Chaudhuri, T. R. (2025). Exploring listeners' perceptions of AI-generated and human-composed music for functional emotional applications. *arXiv*. https://doi.org/10.48550/arXiv.2506.02856

Odefunso, A. E., Bravo, E. G., & Chen, Y. V. (2022). Traditional African dances preservation using deep learning techniques. *Proceedings of the ACM on Computer Graphics and Interactive Techniques, 5*(4), Article 35. https://doi.org/10.1145/3533608

Olojede, O. A., Stephen Olatunde, O., & Oluwaseun, O. A. (2025). Performance evaluation of some machine learning models for music genre classification. *International Journal of Latest Technology in Engineering, Management & Applied Science, 14*(2), 18–22. https://doi.org/10.51583/IJLTEMAS.2025.1402003

Pereverzeva, M. V. (2021). The prospects of applying artificial intelligence in musical composition. *Music Scholarship / Проблемы музыкальной науки, 2021*(1), 8–16. https://doi.org/10.33779/2587-6341.2021.1.008-016

Piryazeva, E. N. (2019). Algorithmic compositions are a phenomenon of electronic music. *Problemy Muzykal'noj Nauki / Music Scholarship,*(2), 105–110. https://doi.org/10.17674/1997-0854.2019.2.105-110

Sarmento, P., Loth, J., & Barthet, M. (2024). Between the AI and me: Analysing listeners' perspectives on AI- and human-composed progressive metal music. *Proceedings of the*

25th International Society for Music Information Retrieval Conference, 1–9. https://arxiv.org/abs/2407.21615

Shan, M.-K., & Chiu, S.-C. (2010). Algorithmic compositions based on discovered musical patterns. *Multimedia Tools and Applications, 46*(1), 139–153. https://doi.org/10.1007/s11042-009-0303-y

Ting, C.-K., Wu, C.-L., & Liu, C.-H. (2017). A novel automatic composition system using evolutionary algorithm and phrase imitation. *IEEE Systems Journal*, *11*(3), 1284–1295. https://doi.org/10.1109/JSYST.2015.2482602

Vengathattam, S. (2025). Collaborative AI in music composition: Human-AI symbiosis in creative processes. *International Journal of Management Science and Information Technology*, *5*(1), 253–262. https://doi.org/10.35870/ijmsit.v5i1.4085

Yin, Z., Reuben, F., Stepney, S., & Collins, T. (2023). Deep learning's shallow gains: A comparative evaluation of algorithms for automatic music generation. *Machine Learning*, *112*, 1785–1822. https://doi.org/10.1007/s10994-023-06309-w

Yusif, I. (2024). The sound of change: Afrobeats and the deconstruction of African stereotypes. *Journal of Critical Race and Ethnic Studies, 2*(1), Article 5. https://scholar.stjohns.edu/jcres/vol2/iss1/5

Zhang, H., Liang, J., Phan, H., Wang, W., & Benetos, E. (2025). From aesthetics to human preferences: Comparative perspectives of evaluating text-to-music systems. *2025 IEEE International Workshop on Machine Learning for Signal Processing*, 1–6. https://arxiv.org/abs/2504.21815

Zhang, M. (2025). Advancing deep learning for expressive music composition and performance modeling. *Scientific Reports*, *15*, Article 28007. https://doi.org/10.1038/s41598-025-13064-6

w