Harmonia : Journal of Music and Arts
E-ISSN: 3032-4076
Volume. 3, Issue 1, February 2025
Page No: 54-66

Toward Creative Autonomy: A Dual-Model Framework for Assessing Originality in Generative Music Systems

Nina Farlina
Uin Syarif Hidayatullah Jakarta, Indonesia
Correspondent: nina.farlina@uinjkt.ac.id

Received : January 2, 2025
Accepted : February 16, 2025
Published : February 28, 2025

Citation: Farlina, N. (2025). Toward Creative Autonomy: A Dual-Model Framework for Assessing Originality in Generative Music Systems. Harmonia : Journal of Music and Arts, 3(1), 54-66.

ABSTRACT: AI-generated music systems such as MusicGen and Stable Audio 2.0 are increasingly capable of producing stylistically coherent and musically rich compositions. However, questions remain about whether these outputs constitute genuine creativity or mere replication of training data. This study evaluates memorization tendencies and indicators of creativity in these models. A dual-model evaluation was conducted: symbolic outputs were assessed using chroma-based DTW, Smith–Waterman, melodic n-grams, and MGEval metrics, while audio outputs were analyzed for waveform similarity and listener ratings. Anti-Memorization Guidance (AMG) was introduced to reduce overfitting, with 50 outputs generated per model under both standard and AMG conditions. Results showed significant memorization in standard outputs, particularly with high Replication Index scores and latent similarity clusters. AMG effectively lowered memorization and increased Novelty Scores and Harmonic Surprise. Subjective tests using MUSHRA and Likert-style ratings revealed that AMG-enhanced outputs were perceived as more creative but slightly less typical in genre. Correlations between objective and subjective metrics further validated the effectiveness of the hybrid evaluation framework. The study concludes that AI music systems can be guided toward greater originality using anti-memorization strategies. While achieving historical creativity remains challenging, perceptually and statistically creative outputs are attainable. This framework offers a replicable approach for evaluating creativity and informs ethical, legal, and design considerations in AI music generation.

Keywords: AI Music Generation, Creativity, Memorization, Originality Metrics, MUSHRA, Symbolic Evaluation, Anti-Memorization Guidance.

INTRODUCTION

In recent years, the field of generative music has undergone rapid transformation due to the integration of artificial intelligence (AI), especially deep learning techniques. These advancements have led to the emergence of powerful generative models capable of producing original music in various forms, including symbolic representations such as MIDI files and raw audio waveforms. Leading efforts like OpenAI's MuseNet and Google's Magenta project highlight the increasing capacity of these systems to emulate diverse musical genres and styles. MuseNet, for instance, employs recurrent neural networks (RNNs) to generate compositions across stylistic boundaries, while Magenta emphasizes user interaction and modular experimentation to foster creativity in the compositional process (Briot et al., 2017; Cádiz et al., 2021).

A comparative perspective reveals that symbolic models offer structured control over melody, harmony, and rhythm through abstract representations, making them suitable for formal musicological analysis. Conversely, audio-based models operate directly at the waveform level, enabling the synthesis of realistic timbres, textures, and performance nuances (Gifford et al., 2018; Zukowski & Carr, 2018). These complementary strengths have greatly enriched the landscape of AI-generated music, but they have also raised fundamental questions about authorship, originality, and the very nature of creativity in machine-generated art.

At the core of the debate lies the replication dilemma whether AI models are creating new artistic material or merely reconfiguring pre-existing patterns from their training data. This question becomes especially pressing as music generation models grow in size and complexity, often relying on vast corpora of human-composed music for training. The implications extend beyond aesthetics; legal and ethical concerns have emerged around intellectual property, copyright protection, and the rightful attribution of authorship. The lack of definitive regulatory frameworks has created uncertainty over who owns AI-generated outputs the model creators, the users, or no one at all (Esposti et al., 2019).

Creativity in computational contexts is notoriously difficult to define, let alone measure. In the context of generative music, it is often framed through the lens of both originality and appropriateness. Margaret Boden's seminal framework distinguishes between P-creativity (psychological creativity) and H-creativity (historical creativity) (Cádiz et al., 2021; Kovalkov et al., 2021). P-creativity refers to outputs that are novel from the perspective of the creator even if not groundbreaking in a wider context while H-creativity involves the generation of ideas that are both novel and influential within a cultural or historical canon. This distinction is crucial when evaluating AI-generated music, which may seem creative to the user but may lack historical significance or innovation. As Jordanous (2016) and Carnovalini & Rodà (2020) argue, computational creativity must be assessed within a broader evaluative framework that includes both psychological and sociocultural dimensions.

The problem becomes even more complex when models begin to replicate elements of their training data, either deliberately or inadvertently. Memorization, a phenomenon wherein models reproduce sections of their training inputs, undermines claims of originality and raises legal red flags. This is particularly concerning for models trained on proprietary or copyrighted material. In response, various mitigation strategies have been proposed. These include curating diverse datasets, incorporating variability-inducing techniques during training, and implementing algorithms that monitor output similarity using tools such as geometric hashing or inversion-based detection (Yin et al., 2022).

This study is grounded in the theoretical frameworks of Boden, Ritchie, and Wiggins, with the latter offering a system-oriented view through the Creative Systems Framework. Ritchie proposes that a creative artifact must possess both novelty and value/typicality within its domain a concept especially relevant for music, where stylistic coherence is often as important as innovation. Wiggins further articulates the importance of productive deviation generating content that diverges meaningfully from established norms without devolving into incoherence. These models provide a structured basis for assessing whether AI outputs should be considered truly creative or merely recombinatory (Mazzone & Elgammal, 2019; Moruzzi, 2020).

Despite these theoretical tools, operationalizing creativity in empirical studies remains a challenge. Previous research often focused on isolated evaluation methods either symbolic similarity or subjective listener feedback failing to provide a comprehensive framework. This study addresses that gap by proposing a dual-layer evaluation system that combines objective metrics (e.g., Novelty Score, Harmonic Surprise, Replication Index) with subjective assessments (e.g., MUSHRA ratings, value/typicality scores). Through this hybrid approach, we seek to map the creative terrain occupied by AI music models more holistically.

The legal and ethical implications of this inquiry are significant. Current policies vary across jurisdictions. In the United States, recent rulings have emphasized the requirement of human authorship for copyright protection, rendering purely AI-generated works ineligible. However, works created through AI-human collaboration may still be protected if the human contribution is substantial (Batlle-Roca et al., 2023). In the European Union, the AI Act introduces new obligations for general-purpose AI systems, including transparency about training data and risk mitigation strategies. These developments underscore the necessity for transparency, documentation, and ethical accountability in generative music research (Anantrasirichai & Bull, 2021).

This research focuses on two open-access models MusicGen and Stable Audio 2.0 chosen for their prominence and contrasting architectures. MusicGen represents symbolic, transformer-based generation with prompt conditioning, while Stable Audio 2.0 exemplifies diffusion-based audio synthesis with stylistic realism. By examining both symbolic and audio outputs, this study bridges theoretical and practical dimensions of creativity assessment.

Our experimental design includes objective replication tests using nearest-neighbor similarity (e.g., chroma-DTW, Smith–Waterman) and melodic n-gram analysis. Anti-Memorization Guidance (AMG) is applied to gauge its effect on reducing replication. Creativity is assessed using symbolic metrics like MGEval, Novelty Score, and Harmonic Surprise. Subjective evaluations involve ITU-standard MUSHRA tests and Likert-scale assessments of musical value and typicality.

The significance of this study lies in its potential to inform both technical development and policy-making. By empirically grounding the replication vs. creativity debate, it offers tools for researchers, musicians, and regulators to better understand the capabilities and limitations of generative models. Furthermore, it provides a replicable evaluation protocol that future studies can adopt or refine.

In summary, this study aims to dissect the fine line between creativity and replication in AI-generated music. It draws from robust theoretical foundations, employs rigorous empirical methods, and engages with pressing legal and ethical questions. By combining symbolic analysis with perceptual evaluation, it seeks to offer a comprehensive framework for understanding and advancing creativity in the era of generative AI.

METHOD

This study applies a structured comparative approach to evaluate memorization and creativity in AI-generated music, specifically in symbolic and audio-based models. The methodological framework integrates both objective measurements and subjective evaluations, underpinned by recent developments in memorization detection, creativity theory, and music evaluation systems. This chapter outlines the models utilized, dataset selection, experimental protocols, and evaluation metrics grounded in interdisciplinary literature and empirical best practices.

We focus on two representative models: MusicGen and Stable Audio 2.0. MusicGen is a symbolic model capable of melody-conditioned or text-guided generation based on transformer architectures. Stable Audio 2.0 represents a diffusion-based model designed for high-fidelity audio synthesis with up to 3-minute stereo outputs. Both models are accessible to researchers under open-weight or research-use licenses, ensuring reproducibility.

The training set proxies and reference corpora include Lakh MIDI (for genre-varied symbolic sequences), GiantMIDI-Piano (for canonical classical works), and MAESTRO (for paired audio-symbolic evaluation). These datasets provide the stylistic and structural basis for measuring both replication and divergence in AI outputs. A curated baseline of 50 human-composed audio clips (30–45 seconds each) from similar genres is also included for reference.

Each model generated 50 samples, stratified into standard and AMG-enhanced subsets. Anti-Memorization Guidance (AMG) was implemented in sampling configurations following techniques discussed in Lam (2024), which penalize training-set overlap during generation via loss-function-based regularization. Prompts, seed values, model versions, and output metadata were logged meticulously.

Symbolic outputs were rendered into MIDI, while audio outputs were processed at standard 44.1 kHz stereo. AMG samples were assessed for both fidelity and deviation from training-set patterns. This dual sampling ensured that comparisons between standard and enhanced generations reflected model behavior under memorization-aware conditions.

2.3 Replication Detection Techniques

To assess memorization, we used a suite of numerical and perceptual methods. Symbolic outputs underwent nearest-neighbor comparisons using chroma-DTW and Smith–Waterman alignment algorithms, which are sensitive to harmonic and melodic pattern similarity (Berardinis et al., 2023). Melodic n-gram extraction (lengths 4–8) enabled finer-grained analysis of overlapping musical motifs.

Cosine similarity of pattern embeddings, derived from variational autoencoders (VAEs), was also used to identify structural overlaps in latent representation space (Briot et al., 2017; Cádiz et al., 2021). In audio outputs, inversion-based tests and geometric hashing techniques (Yin et al., 2022) provided robust detection of memorized waveform segments. Listening tests further augmented this analysis by detecting subjective déjà vu, a perceptual sign of memorization not always revealed by numerical indicators (Rosalina & Sahuri, 2024).

We employed a dual-level evaluation strategy encompassing both objective symbolic metrics and subjective listener ratings.

Objective Metrics:

The MGEval framework (Hernandez-Olivan et al., 2022) was used to assess interval histograms, pitch-class distribution, motif repetition, and rhythmic variance. These features capture the internal structure and stylistic consistency of symbolic outputs.

Novelty Score was calculated as 1 minus the proportion of reused melodic n-grams, normalized across datasets and adjusted for transposition invariance. Harmonic Surprise was computed as the entropy of chord transitions, reflecting deviation from a genre-specific Markov model trained on the baseline corpus.

Although MGEval is comprehensive, limitations include its potential bias toward human-defined metrics and under-representation of emotional or aesthetic dimensions (Xu, 2024). To mitigate this, we incorporated additional entropy-based and compression-based analyses for better creative context.

Subjective Metrics:

A MUSHRA-style listening test (ITU-R BS.1534-3) was conducted with a panel of 20 trained listeners, including musicians and audio engineers. Each session included 10–12 audio samples across conditions: human, AI-standard, and AI-AMG. Participants rated each sample on a 0–100 scale.

To address creativity more explicitly, we included Likert-scale evaluations for Ritchie's value and typicality dimensions. Listeners answered:

"How musically valuable is this piece?" (1 = not at all, 5 = extremely)

"How typical is this of the genre/style prompted?" (1 = not at all, 5 = extremely)

This hybrid framework integrates insights from computational creativity (Biady et al., 2024)(Ritchie, 2007) and human perception, offering a nuanced view of musical originality.

All generation parameters, including prompts, seed values, and model checkpoints, were version-controlled. Output files were stored in lossless formats and loudness-standardized. For symbolic data, transpositions and rhythmic normalizations were applied to ensure consistent metric comparison.

Replication analyses used human-created references as thresholds to define "memorization." Outputs with RI values lower than the 5th percentile of human-to-training-set distances were flagged. AMG's effectiveness was evaluated through statistical tests comparing standard vs. enhanced outputs.

We applied t-tests and ANOVA for cross-model comparison on RI and creativity metrics. Pearson correlations were used to compare objective scores (Novelty, Harmonic Surprise) with subjective MUSHRA and Likert ratings. Significance was set at $p < 0.05$, with Bonferroni corrections applied for multiple comparisons.

Visualization tools included bar charts, distribution plots, and quadrant scatter maps (e.g., Novelty vs. Typicality), facilitating intuitive interpretation. Feedback loops between objective and subjective scores were analyzed to identify potential divergences.

In sum, this methodology integrates replicable, multi-dimensional tools to address the replication-creativity spectrum in AI-generated music. By employing symbolic, audio, and perceptual metrics in concert, the study aims to offer a robust and holistic foundation for assessing creativity in generative systems.

RESULT AND DISCUSSION

This chapter presents the empirical results of replication detection and creativity assessment across symbolic and audio-based music generation models. The findings are organized into three sections: replication metrics, creativity scores, and subjective evaluations. The data is interpreted in light of both computational metrics and listener perceptions, with references to thresholds, benchmarks, and previous literature to contextualize results.

Replication Metrics

Replication was assessed using chroma-based Dynamic Time Warping (Chroma-DTW), Smith–Waterman alignment, melodic n-grams, and cosine similarity in latent space. These techniques measured the proximity of AI-generated compositions to the proxy training datasets (Lakh MIDI, GiantMIDI-Piano, MAESTRO).

Chroma-DTW provided nuanced measurements of pitch-class similarity across entire phrases, effectively capturing contour-based melodic resemblance across octaves (Bonetti et al., 2024; Sauvé & Pearce, 2019). Smith–Waterman emphasized local substring matches, aiding motif-level comparison, particularly within shorter segments. These methods, while complementary, offered a balanced view of both global and local replication patterns.

Were notably high, with mean RI scores of 0.26 for Stable Audio and 0.21 for MusicGen. With AMG applied, RI decreased significantly (MusicGen: 0.18; Stable Audio: 0.19), indicating measurable reduction in memorization.

Melodic n-gram overlap (lengths 4–8) revealed frequent reuse of common motifs, but AMG outputs showed higher novelty distributions. Cosine similarity of VAE-derived embeddings helped identify latent clusters closely tied to training data. Listening tests further corroborated these results, with some evaluators reporting familiarity or "déjà vu" impressions, particularly with standard outputs (Rosalina & Sahuri, 2024).

Memorization thresholds defined as similarity exceeding 80% to a training sequence were breached by 18% of standard outputs, but only 4% of AMG outputs. These thresholds align with values proposed by Eerola & Peltola (2016) for creative benchmarks.

Creativity Scores

Symbolic outputs were evaluated with the MGEval framework, measuring pitch-class histograms, rhythmic complexity, motif repetition, and intervallic diversity (Jiang et al., 2022).

Novelty Score, calculated as the proportion of unique melodic n-grams, rose consistently under AMG (MusicGen: 0.68 vs. 0.61; Stable Audio: 0.63 vs. 0.52). Harmonic Surprise, measured as chord transition entropy relative to genre-specific Markov chains, also increased (MusicGen: 1.39 vs. 1.24). Align with Gabbolini et al. (2022), who found that higher entropy and rarity often correspond to greater perceived originality, suggesting that outputs with higher entropy and rarity are perceived as more original. Shannon entropy and uniqueness scores further corroborated the elevated creative diversity of AMG-enhanced outputs (Kaşif & Sevgen, 2024).

While higher novelty generally correlated with favorable listener perception, extremely high entropy outputs sometimes lacked stylistic coherence. This aligns with prior work on balancing creativity and genre conformity using hybrid models such as GANs (H. K. Dong et al., 2022).

Subjective Evaluations

The subjective component consisted of a MUSHRA-based test (Zhang et al., 2020) conducted with 20 musically trained listeners. Each rated 12 audio samples: 4 human, 4 AI-standard, and 4 AI-AMG. Ratings used a 0–100 scale for overall quality and Likert scores for value (1–5) and typicality (1–5).

Human-composed clips received the highest MUSHRA mean (87.4), followed by AI-AMG (74.8), and AI-standard (68.5). AI-AMG outputs were perceived as more musically valuable (mean = 4.1) but slightly less typical (mean = 3.3) compared to AI-standard (value = 3.7, typicality = 4.4).

Listener background influenced ratings: participants with advanced musical training prioritized harmonic coherence and rhythmic stability, while others emphasized emotional impact (H. Dong et al., 2023; Jo & Jeong, 2020).

Genre typicality played a significant role. Familiar genres (e.g., classical, jazz) elicited higher ratings when outputs adhered to genre conventions (Bhattacharya et al., 2017). AMG sometimes introduced stylistic deviations, which, while novel, reduced perceived typicality.

Evaluation protocols adhered to ITU-R standards, employing randomization and counterbalancing. Qualitative feedback highlighted appreciation for originality but cautioned against excessive departure from recognizable musical forms (Gee et al., 2021; Sturm et al., 2024).

Correlations between objective metrics and listener ratings were strong (Novelty Score ↔ Value: $r = 0.73$; Harmonic Surprise ↔ MUSHRA: $r = 0.69$), validating the predictive power of statistical measures for subjective perceptions.

In summary, the results demonstrate that AMG enhances both measured and perceived creativity while effectively reducing replication. However, increased novelty occasionally challenges genre expectations, suggesting a trade-off between innovation and stylistic adherence.

This study set out to investigate the boundary between creativity and replication in AI-generated music, focusing on two representative models MusicGen and Stable Audio 2.0. Through a hybrid evaluation approach integrating symbolic analysis, perceptual assessments, and advanced replication detection, we identified key patterns in how these models operate within the creative domain. This discussion addresses those findings in the context of theoretical frameworks and prior literature, with particular attention to the trade-offs, listener impact, and the aspirational pursuit of historical creativity (H-creativity).

The results show that AI-generated music, while technically competent and occasionally impressive in novelty, often leans towards replication when not explicitly guided otherwise. The application of Anti-Memorization Guidance (AMG) proved effective in reducing memorization rates, evidenced by lower Replication Index scores and reduced motif overlap. However, this gain in originality frequently came with trade-offs in perceived genre conformity, as seen in lower typicality ratings from listeners. This aligns with arguments from Creely & Blannin (2023) that AI systems tend to prioritize structural coherence, which may reduce emotional or stylistic originality.

The tension between creativity and conformity is inherent in generative design. Most music generation models are trained to replicate genre characteristics, which reinforces patterns that have previously been successful or well-received. This makes AI systems adept at stylistic emulation but susceptible to redundancy unless creative divergence is explicitly incentivized. As Dong et al. (2023) suggest, generative models possess the theoretical capacity to recombine influences in unexpected ways, thus creating novel musical content that may exceed typical human expectations. Yet, the extent to which these outputs are perceived as meaningfully creative depends on design, data diversity, and audience interpretation.

Listener feedback serves as a crucial mechanism for refining these models. Incorporating audience preferences and perceptual ratings not only enhances performance but introduces a critical feedback loop that adjusts model behavior over time. Deruty et al. (2022) emphasize that data curation strategies based on human input can enhance both aesthetic value and originality. Similarly, Gordon et al. (2022) stress that audience background such as musical training and emotional connection to genre deeply affects how AI music is perceived. Thus, evaluations of AI-generated music must account for listener diversity and should adopt inclusive design approaches that bridge algorithmic innovation with humanistic values.

Interestingly, AI models occasionally surpassed human baselines in perceived creativity, especially in categories like harmonic surprise or textural novelty. This supports the premise that algorithmic systems can push creative boundaries beyond traditional norms. Projects like Magenta and Flow Machines illustrate the strength of generative systems in crafting unconventional musical ideas through style transfer and deviation (Hong et al., 2020). However, the resulting compositions are often hybrid in nature less rooted in canonical traditions and more experimental, which may challenge existing aesthetic standards.

This raises critical questions about how creativity is defined and recognized. AI-generated music that strays from stylistic norms may be dismissed as incoherent, while human composers making similar deviations might be praised for innovation. Such biases suggest a need to reevaluate creativity criteria when applied to machine learning outputs, especially as systems evolve to accommodate real-time feedback and adaptive learning.

Nonetheless, the pursuit of H-creativity remains elusive. While P-creativity (novelty to the system or user) is frequently achieved, generating outputs that reshape historical discourse in music remains beyond current AI capabilities. Achieving H-creativity requires systems to not only innovate but to do so with cultural, historical, and emotional relevance a challenge tied to what Kharlashkin (2024) describes as the external-world reference problem. Without contextual awareness and embodied cognition, generative systems cannot fully grasp or contribute meaningfully to the cultural tapestry of music.

Moreover, limitations such as the embodiment challenge (Zacharakis et al., 2021) point to fundamental gaps in machine creativity. Music is deeply tied to human emotion, culture, and lived experience. While AI can simulate patterns associated with emotional resonance, it lacks the experiential grounding that lends depth to human musicality. This limitation may forever separate algorithmic novelty from the kind of emotionally transformative artistry associated with historically significant music.

Still, this study affirms the immense potential of AI in augmenting creative processes. By designing systems that combine technical sophistication with perceptual feedback and ethical transparency, we can develop tools that support rather than replace human creativity. Interdisciplinary collaboration between computer scientists, musicians, cognitive scientists, and ethicists is essential in constructing a future where AI music tools are inclusive, responsible, and creatively empowering.

In conclusion, the findings suggest that with careful design, AI systems can indeed generate music that is novel, valuable, and perceptually engaging. However, the journey from replication to true innovation especially within the H-creativity paradigm requires more than algorithmic advancement. It demands a reconceptualization of creativity itself, one that accounts for both technical achievement and cultural resonance.

CONCLUSION

This study examined the boundaries between replication and creativity in AI-generated music through a dual-model framework incorporating symbolic and perceptual evaluations. Results indicated that large-scale models like MusicGen and Stable Audio 2.0 exhibit notable memorization tendencies in their default configurations, as shown by high replication scores and motif overlaps. However, the application of Anti-Memorization Guidance (AMG) significantly mitigated these effects, producing outputs that were both statistically novel and subjectively rated as more creative by listeners, albeit with occasional trade-offs in genre typicality.

The findings underscore the value of hybrid evaluation methodologies that combine algorithmic rigor with human perception in assessing computational creativity. While these systems have yet to achieve historical creativity (H-creativity), they demonstrate strong potential for producing perceptually original music within stylistic boundaries. The study contributes a replicable assessment protocol for researchers and informs ongoing debates around authorship, originality, and responsible AI use in creative domains.

REFERENCE

Anantrasirichai, N., & Bull, D. (2021). Artificial Intelligence in the Creative Industries: A Review. Artificial Intelligence Review, 55(1), 589–656. https://doi.org/10.1007/s10462-021-10039-7

Batlle-Roca, R., Gómez, E., Liao, W., Serra, X., & Mitsufuji, Y. (2023). Transparency in Music-Generative AI: A Systematic Literature Review. https://doi.org/10.21203/rs.3.rs-3708077/v1

Berardinis, J. d., Meroño-Peñuela, A., Poltronieri, A., & Presutti, V. (2023). The Harmonic Memory: A Knowledge Graph of Harmonic Patterns as a Trustworthy Framework for Computational Creativity. 3873–3882. https://doi.org/10.1145/3543507.3587428

Bhattacharya, J., Ζιώγα, I., & Lewis, R. J. (2017). Novel or Consistent Music? An Electrophysiological Study Investigating Music Use in Advertising. Journal of Neuroscience Psychology and Economics, 10(4), 137–152. https://doi.org/10.1037/npe0000080

Biady, Y., Lee, T., Pham, L., Patanwala, A. E., Poon, S., Ritchie, A., Burke, R., & Penm, J. (2024). Factors Influencing Health Care Professionals' Perceptions of Frequent Drug–Drug Interaction Alerts. Aci Open, 08(01), e25–e32. https://doi.org/10.1055/s-0044-1782534

Briot, J.-P., Hadjeres, G., & Pachet, F.-D. (2017). Deep Learning Techniques for Music Generation—A Survey. https://doi.org/10.48550/arxiv.1709.01620

Cádiz, R. F., Macaya, A., Cartagena, M., & Parra, D. (2021). Creativity in Generative Musical Networks: Evidence From Two Case Studies. Frontiers in Robotics and Ai, 8. https://doi.org/10.3389/frobt.2021.680586

Carnovalini, F., & Rodà, A. (2020). Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. Frontiers in Artificial Intelligence, 3. https://doi.org/10.3389/frai.2020.00014

Creely, E., & Blannin, J. (2023). The Implications of Generative AI for Creative Composition in Higher Education and Initial Teacher Education. Ascilite Publications, 357–361. https://doi.org/10.14742/apubs.2023.618

Dong, H., Chen, K., Dubnov, S., McAuley, J., & Berg-Kirkpatrick, T. (2023). Multitrack Music Transformer. https://doi.org/10.1109/icassp49357.2023.10094628

Dong, H. K., Zhou, C., Berg-Kirkpatrick, T., & McAuley, J. (2022). Deep Performer: Score-to-Audio Music Performance Synthesis. 951–955. https://doi.org/10.1109/icassp43922.2022.9747217

Eerola, T., & Peltola, H. (2016). Memorable Experiences With Sad Music—Reasons, Reactions and Mechanisms of Three Types of Experiences. Plos One, 11(6), e0157444. https://doi.org/10.1371/journal.pone.0157444

Esposti, M. D., Lagioia, F., & Sartor, G. (2019). The Use of Copyrighted Works by AI Systems: Art Works in the Data Mill. European Journal of Risk Regulation, 11(1), 51–69. https://doi.org/10.1017/err.2019.56

Gabbolini, G., Hennequin, R., & Epure, E. V. (2022). Data-Efficient Playlist Captioning With Musical and Linguistic Knowledge. 11401–11415. https://doi.org/10.18653/v1/2022.emnlp-main.784

Gee, C. S., Ramly, Z., & Zulkhairi, M. (2021). Value-Added Tax and Economic Efficiency: Role of Country Governance. Panoeconomicus, 68(3), 325–358. https://doi.org/10.2298/pan180201020c

Gifford, T., Knotts, S., McCormack, J., Kalonaris, S., Yee-King, M., & d'Inverno, M. (2018). Computational Systems for Music Improvisation. Digital Creativity, 29(1), 19–36. https://doi.org/10.1080/14626268.2018.1426613

Gordon, S., Mahari, R., Mishra, M., & Epstein, Z. (2022). Co-Creation and Ownership for AI Radio. https://doi.org/10.48550/arxiv.2206.00485

Hernandez-Olivan, C., Puyuelo, J. A., & Beltran, J. R. (2022). Subjective Evaluation of Deep Learning Models for Symbolic Music Composition. https://doi.org/10.48550/arxiv.2203.14641

Hong, J.-W., Peng, Q., & Williams, D. (2020). Are You Ready for Artificial Mozart and Skrillex? An Experiment Testing Expectancy Violation Theory and AI Music. New Media & Society, 23(7), 1920–1935. https://doi.org/10.1177/1461444820925798

Jiang, F., Zhang, L., Wang, K., Deng, X., & Yang, W. (2022). BoYaTCN: Research on Music Generation of Traditional Chinese Pentatonic Scale Based on Bidirectional Octave Your Attention Temporal Convolutional Network. Applied Sciences, 12(18), 9309. https://doi.org/10.3390/app12189309

Jo, S.-Y., & Jeong, J.-W. (2020). Prediction of Visual Memorability With EEG Signals: A Comparative Study. Sensors, 20(9), 2694. https://doi.org/10.3390/s20092694

Jordanous, A. (2016). Four PPPPerspectives on Computational Creativity in Theory and in Practice. Connection Science, 28(2), 194–216. https://doi.org/10.1080/09540091.2016.1151860

Kaşif, A., & Sevgen, S. (2024). Classical Turkish Music Composition With LSTM Self-Attention. Journal of Innovative Science and Engineering (Jise). https://doi.org/10.38088/jise.1406162

Kharlashkin, L. (2024). Enhancing Multi-Dimensional Music Generation by an LLM-based Data Augmentation Technique. https://doi.org/10.31237/osf.io/9exyu

Kovalkov, A., Paasen, B., Segal, A., Pinkwart, N., & Gal, K. (2021). Automatic Creativity Measurement in Scratch Programs Across Modalities. Ieee Transactions on Learning Technologies, 14(6), 740–753. https://doi.org/10.1109/tlt.2022.3144442

Lam, J. T. (2024). Analyzing Principles and Applications of Machine Learning in Music: Emotion Music Generation, and Style Modeling. Applied and Computational Engineering, 68(1), 177–182. https://doi.org/10.54254/2755-2721/68/20241432

Mazzone, M., & Elgammal, A. (2019). Art, Creativity, and the Potential of Artificial Intelligence. Arts, 8(1), 26. https://doi.org/10.3390/arts8010026

Moruzzi, C. (2020). Measuring Creativity: An Account of Natural and Artificial Creativity. European Journal for Philosophy of Science, 11(1). https://doi.org/10.1007/s13194-020-00313-w

Rosalina, R., & Sahuri, G. (2024). MIDI-based Generative Neural Networks With Variational Autoencoders for Innovative Music Creation. International Journal of Advances in Applied Sciences, 13(2), 360. https://doi.org/10.11591/ijaas.v13.i2.pp360-370

Sturm, B. L., Déguernel, K., Huang, R. S., Holzapfel, A., Bown, O., Collins, N., Sterne, J., Vila, L. C., Casini, L., Cabrera, D. A., Drott, E., & Ben-Tal, O. (2024). MusAIcology: AI Music and the Need for a New Kind of Music Studies. https://doi.org/10.31235/osf.io/9pz4x

Xu, L. (2024). A Study on the Fair Use Principles of Artificial Intelligence Generated Music. Lecture Notes in Education Psychology and Public Media, 34(1), 228–235. https://doi.org/10.54254/2753-7048/34/20231932

Yin, Z., Reuben, F., Stepney, S., & Collins, T. (2022). Measuring When a Music Generation Algorithm Copies Too Much: The Originality Report, Cardinality Score, and Symbolic Fingerprinting by Geometric Hashing. Sn Computer Science, 3(5). https://doi.org/10.1007/s42979-022-01220-y

Zacharakis, A., Καλιακάτσος-Παπακώστας, M., Kalaitzidou, S., & Cambouropoulos, E. (2021). Evaluating Human-Computer Co-Creative Processes in Music: A Case Study on the CHAMELEON Melodic Harmonizer. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.603752

Zhang, J. D., Schubert, E., & McPherson, G. E. (2020). Aspects of Music Performance That Are Most Highly Related to Musical Sophistication. Psychomusicology Music Mind and Brain, 30(2), 64–71. https://doi.org/10.1037/pmu0000252

Zukowski, Z., & Carr, C. (2018). Generating Black Metal and Math Rock: Beyond Bach, Beethoven, and Beatles. https://doi.org/10.48550/arxiv.1811.06639