Universal Library of Arts and Humanities
Research Article

ISSN: 3064-9943 | Volume 2, Issue 4

Open Access | PP: 17-22

DOI: https://doi.org/10.70315/uloap.ulahu.2025.0204003

Writing the Future: The Impact of Artificial Intelligence and Knowledge Graphs on the Music Industry

Popova Anastasiia

Moscow State Tchaikovsky Conservatory, Specialist Diploma in Piano Performance (Concert Performer, Teacher, Accompanist), Argentina, Buenos Aires.

Abstract

This article investigates the synergistic potential of integrating knowledge graphs and state-of-the-art artificial intelligence models for the fundamental transformation of musical creativity processes. Against the backdrop of the growing generative music market, a critical lack of deep structural control and music-theoretical grounding in existing AI systems becomes apparent, inevitably leading to the production of compositions with limited long-term coherence. The objective of the study is to conceptualize and justify a hybrid model that unites the deterministic precision of knowledge graph structures with the timbral and textural richness of neural-network generative approaches. The methodological foundation comprised the analysis and synthesis of contemporary music-generation techniques, including Transformer architectures (Music Transformer, Museformer), diffusion models (GETMusic), text-to-music systems (MusicGen), as well as empirical experience of applying graph structures within the Teragraph hackathon. The results demonstrate that the hybrid approach effectively overcomes the limitations of both symbolic and exclusively neural methods. In the architecture the knowledge graph establishes a high-level compositional framework which the AI model realizes in the form of a finished audio artifact, providing an unprecedented level of control over the generation process. The conclusion of the study is that the further development of creative music technologies will proceed along the path of human–machine co-creation based on hybrid systems of this nature. The work will be of interest to researchers in the field of computer music, developers of AI systems, as well as composers and producers seeking to master new creative tools.

Keywords: *Artificial Intelligence, Knowledge Graphs, Music Generation, Deep Learning, Transformers, Diffusion Models, Symbolic Music, Computer Composition, Music Industry, Hybrid Models.*

INTRODUCTION

In the context of accelerated digital transformation, artificial intelligence (AI) is no longer confined to the role of an analytical and optimization tool, gradually mastering domains traditionally considered the prerogative of humans, in particular the creative sphere. The music industry is undergoing profound transformations under the influence of generative AI technologies: according to forecasts, it will increase from $3.62 billion in 2024 to $4.48 billion in 2025, with a compound annual growth rate (CAGR) of 23.7 percent. It is also expected that in the coming years AI in the music market will demonstrate exponential growth, reaching $10.43 billion in 2029 with a CAGR of 23.5 percent [1].

Nevertheless, despite the fact that modern neural network models can reproduce locally plausible musical fragments, they face a fundamental problem—a lack of reliable mechanisms for constructing the long-term structure of a composition, its form, and dramaturgy. The generated fragments often represent a set of successfully produced episodes but fail to maintain integrity within large-scale musical works. This gap in research is associated with a shortage of methodologies capable of combining the heuristic flexibility of deep learning with the formalized knowledge of music theory.

The objective of the study is to conceptualize and substantiate a hybrid model that integrates the deterministic clarity of knowledge graph structures with the timbral and textural richness of neural network generative approaches.

The study's scientific innovation lies in the development and justification of a hybrid human–machine methodology for music generation that overcomes the considerable limitations of fully automated systems. It is based on the integration of three complementary components: strict structural control implemented via knowledge graphs that formalize the

Citation: Popova Anastasiia, “Writing the Future: The Impact of Artificial Intelligence and Knowledge Graphs on the Music Industry”, Universal Library of Arts and Humanities, 2025; 2(4): 17-22. DOI: https://doi.org/10.70315/uloap.ulahu.2025.0204003.

compositional logic and architecture of a musical work; the ability of deep neural networks to render a sonic palette rich in texture and nuance, ensuring high-quality audio output; and the essential involvement of a human expert (composer or music theorist), whose professional knowledge is encoded in the rules of the graph-based model and steers the creative intent at every stage of the process.

The synergy of the formal graph-oriented compositional framework, the neural-network–driven sound synthesis engine, and the expert curatorial function inaugurates a new paradigm in computer music. Rather than aimlessly generating sound fragments, this method provides controlled co-authorship in which technology and creative consciousness collaborate closely, achieving a qualitatively elevated level of musical expression.

The author's hypothesis is that the synergy of these methods ensures a higher level of coherence, controllability, and semantic richness in the created musical works compared to the application of each approach individually.

MATERIALS AND METHODS

In recent years research at the intersection of artificial intelligence (AI) and knowledge graph models in the music industry has demonstrated a wide variety of approaches — from large market reports to highly specialized algorithms for melody generation and the construction of semantic connections between musical entities. According to the report by The Business Research Company, the global market for AI solutions in music is estimated at tens of billions of dollars with steady annual growth: key segments include software, cloud and on-premises services, as well as applications in composition, streaming platform recommendations and track mastering [1]. The review by Civit M. et al. [2] emphasizes that the main research activity is concentrated on algorithmic generation of musical content, whereas applied aspects — integration into existing production and producer pipelines — remain insufficiently studied.

In a number of studies traditional deep learning methods are supplemented by attention architectures and flexible mechanisms for controlling the stylistic features of a work. Thus, Pricop T. C., Iftene A. [4] investigate the application of deep neural networks to the creation of melodies and harmonies demonstrating that the combination of convolutional and recurrent layers enables more humanlike phrases by modeling local and long-term dependencies in musical text. In the work of Yu B. et al. [5] the Museformer model is proposed — a transformer with two-level attention capable of simultaneously accounting for rough macrostructural development of a composition and fine local changes in the melodic line.

Copet J. et al. [6] propose a methodology for simple and controlled creation of musical fragments, where the user adds high-level parameters (mood, tempo, genre) to a sample and the model adapts to these control vectors without significant degradation of quality. Lam M. W. Y. et al. [7] focus on accelerating the generation process itself: they develop optimized algorithms for transformers that reduce computational load while preserving professional-level sound, enabling real-time AI compositions on mobile devices.

Alongside neural network approaches graph models are gaining increasing popularity, allowing complex semantic and temporal connections within musical works to be taken into account. Türker I., Aksu S. [3] introduce the concept of Connectogram — a graph-based representation of sounds reflecting changes in spectral components over time, which opens new possibilities for analyzing rhythm and timbre in composition and remixing of tracks. In the work of Saravanou A. et al. [9] a multitask approach to learning inductive representations of music based on graphs is presented, where nodes correspond to fragments or attributes (genre, artist, key) and edges represent influence or similarity relationships; this enables effective solutions for classification, recommendation and search tasks based on semantic matching of musical objects.

Despite the successes of autonomous systems a number of researchers emphasize the need to include a human in the loop to improve the quality and relevance of generated content. Wu X. et al. [8] in their review of human-in-the-loop methods show that active involvement of musicians and listeners at the stages of annotation, hyperparameter tuning and interactive control not only improves subjective perception but also accelerates model adaptation to new styles and tasks.

In reference [10], a modular microservices approach to constructing a graph cloud (cloud teragraph) is described, wherein each service is responsible for a distinct stage—from audio data acquisition and preprocessing to visualization of a de Bruijn graph of a musical work in a web interface. The authors examine hardware requirements (GPU nodes for computation acceleration, distributed storage in NoSQL clusters) and software architecture (service containerization via Docker, orchestration with Kubernetes, RESTful APIs for module interaction). The provided visualization example highlights the necessity of optimizing the rendering of large-scale graphs containing millions of edges, proposing a level-of-detail (LOD) strategy and incremental data loading in the client application [10].

Source [11] presents a complete processing pipeline for a musical fragment: from extraction of spectrograms and Mel-frequency cepstral coefficient (MFCC) features to construction of a transition graph between acoustic templates, where nodes represent stable musical patterns and edges denote probabilistic transitions in the melody. Convolutional and recurrent neural networks within the PyTorch framework are employed for model training, and the authors demonstrate how the generated graph can form the foundation for recommendation systems and automatic harmonization algorithms [11].

The analyzed literature demonstrates a significant gap between commercial forecasts and academic results: market reports [1] predict widespread AI adoption across all segments of the music industry, whereas scientific works predominantly focus on algorithmic generation methods and rarely address aspects of production integration, monetization or legal frameworks. Moreover, although graph-based approaches promise deeper semantic understanding of music, they have not yet achieved large-scale application and are subject to limited empirical validation. Issues of evaluating AI-generated music quality in light of user preferences and ethical problems (copyright, manipulation of emotions) are significantly underestimated. Finally, despite general recognition of the value of human-in-the-loop, in most experiments the user's role is reduced to passive annotation, whereas genuine participation in the creative process remains poorly studied. Thus, promising directions remain the development of hybrid systems combining the power of graph models and neural networks, as well as in-depth investigation of the social and ethical consequences of widespread AI implementation in the musical domain.

RESULTS AND DISCUSSION

The conducted research enables the presentation of key findings, the central of which is the development and substantiation of a hybrid music generation model that combines the capabilities of knowledge graph representations with state-of-the-art neural network architectures. It is believed that such synergy can overcome the current limitations of computer composition methods and provide a qualitatively new level of precision in controlling the process of musical content creation.

The core of controlled generation of musical material is the formalized encoding of musical concepts. An example of this approach is the methodology tested during the Teragraph hackathon, where a hierarchical model of musical forms lies at its center [4]. In this system, the simplest elements are organized into more complex units, thereby reproducing the classical principles of compositional structure. This hierarchy can be visually represented as follows (see Figure 1).

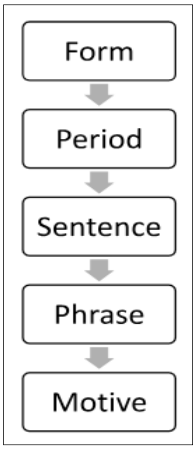

Fig.1. Hierarchical model of musical form based on the graph approach [2, 4]

Algorithmic generation of each musical motif is carried out by means of a stochastic traversal of two interrelated graph representations: melodic and rhythmic. The vertices of the first graph encode discrete pitch values, and the edges define transition probabilities between them taking into account tonal context, modal gravitation (for example, an increased probability of motion from the dominant to the tonic) and the direction of musical movement (ascending or descending). The rhythmic graph, in turn, includes vertices corresponding to various metric durations (whole, half, quarter, etc.), and its edges establish a probabilistic model of rhythmic pattern transitions. Thus the composition process constitutes a controlled walk through these networks, combining variability with strict adherence to the embedded musical norms. This approach ensures the formation of complex and coherent melodic lines possessing a clear internal logic of development, in contrast to banal random sequences of notes [6, 8]. Its key advantage — transparency: by modifying edge weights or the initial traversal conditions the composer gains a direct mechanism for regulating stylistic characteristics, complexity and emotional coloring of the generated music.

Within the Teragraph hackathon, the practical effectiveness of a graph-based approach to the analysis and visualization of musical works was demonstrated: a system was developed that employs De Bruijn graphs to represent structural information about music. The consultant functioned as a bridge between abstract mathematical models and the aesthetic-technical principles of the musical compositional process.

First, raw musical data in MIDI format underwent detailed syntactic parsing: note_on and note_off events—specifying pitch, dynamics and duration—were extracted from each track, after which all instruments (excluding percussion on channel 10) were merged into a single sequence to ensure the uniformity of the material. The algorithm then identified simultaneous combinations of notes forming chords and, using a Bayesian classifier based on the Temperley–Kostka–Payne method, determined the tonal context of each fragment (for example, C major or A minor), thereby allowing the harmonic functions of chords to be appropriately related to the established modal framework.

From the resulting sequence of chord events, a De Bruijn graph was constructed: each vertex represented a unique subsequence of L adjacent chords, and an edge connected two vertices if their subsequences occurred consecutively in the original composition. The consultant's expertise lay in developing criteria for selecting functionally significant chords and filtering out passing (non-harmonic) tones, as well as formalizing the notions of consonance and dissonance within the software structures, thus ensuring that the graph accurately reflected the genuine harmonic texture of the work.

In the final stage, network clustering using the Newman method revealed "communities"—groups of vertices with dense internal connections that, from a musicological standpoint, correspond to principal themes and recurring sections of the composition. The outcome was a clear visualization in which the identified thematic clusters were highlighted in different colors, providing an intuitive map of the harmonic and formal organization of the musical piece [10, 11].

Next, Figure 2 shows the process of identifying communities (themes of a piece of music) and visualizing the dEBRUYNE graph. J.S.Bach, Sonata in D minor BWV565.

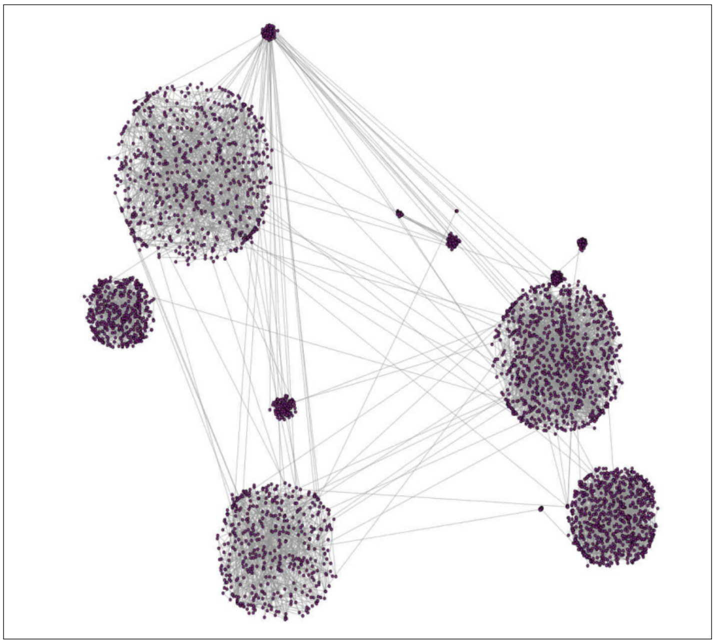

Fig.2. Highlighting communities (themes of a piece of music) and visualization of the dEBRUYNE graph. J.S.Bach, Sonata in D minor BWV565 [10].

Parallel to the advancement of symbolic methods, deep neural network models have achieved impressive results in the domain of audio generation. Contemporary architectures such as Transformer and diffusion networks are capable of producing musical fragments virtually indistinguishable in sound quality from professional studio recordings. The principal differences between these approaches and their impact on the final sonic outcome can be clearly compared in Table 1.

Table 1. Comparative characteristics of modern AI models for music generation [6–9]

Model	Main architecture	Input type	Output type	Key advantage
Music Transformer	Transformer	Symbolic sequence (MIDI-like)	Symbolic sequence	Consideration of long-term dependencies in the structure of the composition
Museformer	Transformer (hierarchical)	Symbolic sequence	Symbolic sequence	Simultaneous control over local details and global form
MusicGen	Transformer (language model) + EnCodec	Text description	Audio file (waveform)	Direct audio generation from text, high flexibility in stylization
GETMusic	Diffusion model	Audio tracks (e.g., drums, bass)	Audio tracks	Inpainting of arrangement, creation of accompaniment to existing parts

Despite technical achievements, modern generative models face a fundamental limitation in relative control over compositional-structural elements. For instance, although MusicGen enables adjustment of the overall stylistic character (for example, melancholic jazz piano theme), it does not allow detailed specification of harmonic progression, execution of modulation in a specific bar or integration of a predefined melodic motif during development. Similarly, solutions such as GETMusic demonstrate high effectiveness in arrangement but operate only on existing fragments (for example, a pre-made drum part). Consequently, a gap emerges between the initial creative intention at the abstract level and its final sonic embodiment [3, 4].

To eliminate this discrepancy, a hybrid architectural scheme is proposed, combining two approaches in a unified process. This model is constructed according to the principle of a two-level hierarchy, where each level is responsible for addressing a specialized task: the first is focused on structural planning of the composition, the second on detailed elaboration of timbral and sound characteristics (see Figure 3).

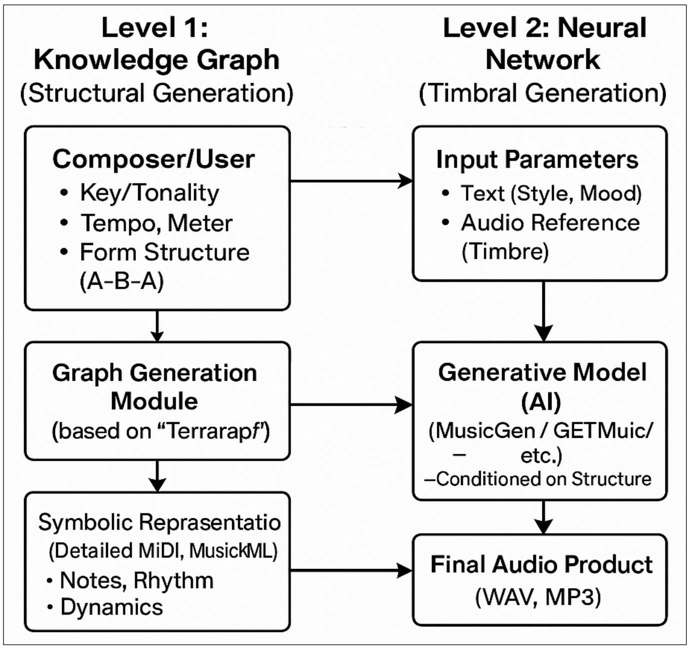

Fig.3. Conceptual diagram of a hybrid architecture for music generation [2, 5, 7]

As can be seen from the data shown in Figure 2, the first level comprises structural generation. At the initial phase the composer sets the global parameters of the work: key, meter, tempo, duration and form (for example, a sonata exposition or a three-part A–B–A). Relying on an ontological knowledge graph similar to Teragraph, the system produces a complete score in a symbolic format — whether extended MIDI or MusicXML. The resulting file records not only pitches and durations but also harmonic arrangement, dynamic articulations, intonational curves and boundaries of formal sections. The outcome is fully deterministic, transparent for analysis and amenable to precise editing.

Level 2 represents timbral generation. The generated symbolic score serves not as an end goal but as a high-precision prompt for a deep neural network model. Instead of a vague instruction such as a swift violin passage, the modified MusicGen or GETMusic accepts a concrete sequence of notes, rhythms and chords, and when necessary also verbal descriptors (for example, a string quartet in a hall with concert acoustics). The artificial intelligence renders the structure into an audio stream in which the original notated text is scrupulously preserved, while natural performance nuances, plausible instrument timbres and spatial acoustics — features that cannot be represented in a symbolic format — are simultaneously introduced.

This hybrid approach combines the strengths of both worlds: the logical coherence of the graph-based model and the sonic richness of modern generative networks, enabling the achievement of the highest possible musicality rating for the resulting track.

A new class of compositional tools thus emerges: the author works at the conceptual level — manipulating forms and structures — and immediately obtains a fully realized, professionally sounding audio recording. This radically accelerates the creative cycle and opens possibilities for experimentation whose realization would previously have required lengthy studio sessions.

In conclusion it should be emphasized that the two-level scheme is not an abstract utopia but a natural stage in the evolution of existing technologies. It systematizes the disparate achievements of symbolic and neural methods, unifying them into a single controllable ecosystem. Yes, practical implementation will require standardized

CONCLUSION

As a result of the conducted study, an analysis of current practices in the application of artificial intelligence technologies and knowledge ontologies in music generation was carried out, the main limitations of existing methodologies were identified, and the concept of a hybrid architecture for their overcoming was formulated. It was demonstrated that symbolic models based on knowledge graphs and subsymbolic deep-learning approaches possess complementary strengths and weaknesses: the former guarantee rigid structural control and clear interpretability but encounter difficulties in synthesizing expressive and authentic timbres, whereas the latter are capable of generating rich timbral textures but often lose the integrity and logical coherence of musical forms when extended temporal spans are considered.

The study's scientific innovation lies in the development and justification of a hybrid human–machine methodology for music generation that overcomes the considerable limitations of fully automated systems. It is based on the integration of three complementary components: strict structural control implemented via knowledge graphs that formalize the compositional logic and architecture of a musical work; the ability of deep neural networks to render a sonic palette rich in texture and nuance, ensuring high-quality audio output; and the essential involvement of a human expert (composer or music theorist), whose professional knowledge is encoded in the rules of the graph-based model and steers the creative intent at every stage of the process.

The synergy of the formal graph-oriented compositional framework, the neural-network–driven sound synthesis engine, and the expert curatorial function inaugurates a new paradigm in computer music. Rather than aimlessly generating sound fragments, this method provides controlled co-authorship in which technology and creative consciousness collaborate closely, achieving a qualitatively elevated level of musical expression.

Further work should be focused on prototyping the proposed architecture, formulating reliable protocols for interaction between the symbolic and neural-network levels, and conducting large-scale psychoacoustic investigations for the empirical validation of the hybrid approach's effectiveness.

REFERENCES

1. The Business Research Company. (2025). Artificial intelligence (AI) in music global market report 2025: By component (software, services), by deployment (cloud, on-premises), by application (music composition and generation, music streaming recommendation, music production and mastering): Market size, trends, and global forecast 2025–2034. Retrieved from https://www.thebusinessresearchcompany.com/report/artificial-intelligence-ai-in-music-global-market-report (date accessed: 05.05.2025).

2. Civit, M., et al. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. Expert Systems with Applications, 209. https://doi.org/10.1016/j.eswa.2022.118190

3. Türker, I., & Aksu, S. (2022). Connectogram: A graph-based time-dependent representation for sounds. Applied Acoustics, 191,. https://doi.org/10.1016/j.apacoust.2022.108660

4. Pricop, T. C., & Iftene, A. (2024). Music generation with machine learning and deep neural networks. Procedia Computer Science, 246, 1855–1864. https://doi.org/10.1016/j.procs.2024.09.692

5. Yu, B., et al. (2022). Museformer: Transformer with fine- and coarse-grained attention for music generation. Advances in Neural Information Processing Systems, 35, 1376-1388.

6. Copet, J., et al. (2023). Simple and controllable music generation. Advances in Neural Information Processing Systems, 36, 47704-47720.

7. Lam, M. W. Y., et al. (2023). Efficient neural music generation. Advances in Neural Information Processing Systems, 36, 17450-17463.

8. Wu, X., et al. (2022). A survey of human-in-the-loop for machine learning. Future Generation Computer Systems, 135, 364–381.https://doi.org/10.1016/j.future.2022.05.014

9. Saravanou, A., et al. (2021). Multi-task learning of graph-based inductive representations of music content. In Proceedings of the International Society for Music Information Retrieval Conference, 602-609.

10. The Cloud teragraph. Software and hardware architecture. Retrieved from https://alexbmstu.github.io/2023/#:~:text=4.6.2.-,Пример%20визуализации%20графа%20деБрюйна%20музыкального%20произведения,-Данный%20пример%20использует (date accessed: 08.07.2025).

11. Processing of a piece of music. Retrieved from https://latex.bmstu.ru/gitlab/hackathon/ex8/-/blob/main/lab6.ipynb?ref_type=heads (date accessed: 08.07.2025).