Gelar: Jurnal Seni Budaya
Vol 23, No. 2, December 2025 pp. 182-199
https://jurnal.isi-ska.ac.id/index.php/gelar/index
E-ISSN 2655-9153

A hermeneutic inquiry into musical meaning in AI-generated music: a case study of Suno AI's text-to-music system

Novia Ratnasari [a,1*], Aji Prasetya Wibawa [a,2], Syaad Patmanthara [a,3]

[a] Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Malang, Jawa Timur 65145, Indonesia
[1] novia.ratnasari.2505349@students.um.ac.id*; [2] aji.prasetya.ft@um.ac.id; [3] syaad.ft@um.ac.id
* Corresponding Author

ABSTRACT

This study examines how generative artificial intelligence participates in the creation and interpretation of musical meaning, using Suno AI's text-to-music system as a focused case. The research explores how machine-generated sound can be understood hermeneutically, particularly how linguistic prompts, probabilistic modeling, and audio generation processes shape meaning, emotion, and musical intention. The study aims to determine the extent to which generative AI functions as an epistemic collaborator rather than a passive tool and how its outputs align with or diverge from human interpretive expectations. Using a digital epistemological hermeneutic framework operationalized through prompt-based observation, semantic interpretation, and comparative listening, the study conducted controlled experiments varying genre, instrument, mood, and tempo. Each output was evaluated in terms of expressive quality, emotional valence, stylistic coherence, and prompt response fidelity. The findings indicate that generative AI constructs musical meaning through representational inference, producing sonic forms that partially reflect the semantic cues embedded in linguistic prompts. Although the system does not exhibit human-like intentionality, its probabilistic structures generate patterns that resonate with human affective and interpretive frameworks, creating a co-creative space where human prompts and machine inference jointly shape musical expression. These insights demonstrate the usefulness of hermeneutics as a methodological lens for understanding AI-mediated creativity and highlight the importance of prompt design, model transparency, and human-machine interpretive dynamics in future computational musicology and creative AI research.

OPEN ACCESS

Article History
Received 2025-10-23
Revised 2025-11-27
Accepted 2025-12-04

Keywords
Computational Music,
Generative Artificial Intelligence,
Digital Hermeneutics,
Epistemic Cognition,
Machine Aesthetic Understanding,

1. Introduction

Music has always served as a profound medium through which humans translate emotion [1], thought, and culture into sound. It embodies both structure and freedom, order and imagination, inviting listeners into a space where meaning unfolds beyond words. In recent years, this artistic landscape has entered a transformative phase, as digital systems begin to participate in the creative act itself. The rapid rise of Artificial Intelligence (AI) in generative technology [2] has begun to reshape how people perceive and negotiate the connections between language [3], music [4], and knowledge [5]. Through the convergence of Natural Language Processing (NLP) [6] and emerging audio generation techniques [7], attention has gradually shifted from signal manipulation toward what many now describe as a new paradigm text-conditioned music generation [8]. In this paradigm, linguistic prompts [9] do not merely convert into sound; they evolve into musical structures whose semantic [10] and affective layers [11] mirror the expressive logic of language. What once appeared to be two distinct modes of representation, verbal and musical, now intersect, forming a fluid epistemic space in which

 ISSN 2655-9153

aesthetic understanding unfolds through digital mediation. However, discussions of text-conditioned music generation often remain separated between technical explanations and conceptual reflections, so there is still no clear account of how language and music actually work together to create meaning within these generative systems.

Generative models such as Suno AI [12], Bark, and Udio exemplify how machines increasingly participate in the dialogue between text and sound. Built upon large transformer architectures [12] and multimodal representations [13], these systems create bridges across linguistic and musical domains, allowing patterns of meaning to resonate between words and tones. Through continuous exposure to vast and diverse datasets, they begin to discern recurring correspondences in how language evokes rhythm [14], texture, or emotion [15], thus cultivating what might be called a computational sensibility toward musical expression [16]. To maintain analytical clarity, it is important to note that this "sensibility" arises from statistical pattern recognition rather than subjective awareness, indicating that the system's capabilities remain grounded in representational inference. In this sense, the generative process unfolds as a conversation of cognition: an encounter where algorithmic pattern and human perception meet to co-create meaning. A clearer connection can be made here by emphasizing that these algorithmic patterns directly influence how listeners interpret the generated music, as the model's probabilistic outputs structure the sonic cues that shape perception. The act of generation becomes less a mechanical translation [17] and more a reflective event, one that gestures toward the possibility of understanding not as possession of knowledge, but as participation in a shared aesthetic experience [18]. However, this reflectiveness should be understood in functional terms: the model produces outputs that align with human semantic expectations, thereby prompting interpretive engagement rather than implying any form of machine subjectivity.

It is within this shared space of creation that interpretation emerges, where human consciousness returns to engage, question, and reshape the meanings produced by the machine. This clarification helps show how the system's technical mechanisms, such as prompt conditioning and latent-space mapping, directly shape the musical material that listeners respond to, thereby connecting technological operation with changing modes of human musical perception. This indicates that the system's technical capabilities also function as interpretive mechanisms that shape how humans perceive the meaning embedded in the generated music. Knowledge can no longer be fully regarded as the exclusive outcome of human cognition; it also arises from the linguistic and semantic inferences performed by computational systems. Within this evolving landscape, generative AI functions as a quasi-cognitive [19] system that organizes multiple layers of data representation into new structures of meaning through automated yet semantically oriented [20] mechanisms. Such a perspective challenges the anthropocentric conception of understanding, inviting new epistemic frameworks for interpreting algorithmic creativity and the expanding terrain of machine-mediated cognition. In this study, the terms *hermeneutic lens* and *aesthetic cognition* are used in a practical sense to describe how generative systems shape musical responses from language and how humans, in turn, interpret these responses in the process of meaning-making.

This phenomenon marks an epistemological turning point in how knowledge is both produced and interpreted. Knowledge can no longer be conceived exclusively as the outcome of human cognition; it also arises from the linguistic and semantic inferences executed by computational systems. In this sense, generative AI operates as a quasi-cognitive agent, organizing layers of data representation into novel structures of meaning through automated yet semantically interpretive mechanisms. Such a perspective challenges the long-held assumption of understanding as an exclusively human faculty, suggesting that cognition may now extend into algorithmic domains capable of mirroring, transforming, and reconfiguring human modes of sense making. This study investigates how generative artificial intelligence (Generative AI) constructs and interprets forms of knowledge, particularly those emerging through the interplay between language and music. Rather than examining mere technical capabilities to generate sound or text, it explores how such systems negotiate and interpret meaning across the two modalities. The central aim is to examine how machines can be said to "understand" or "perceive" beauty, concepts referred to here as machine understanding and

 ISSN 2655-9153

aesthetic cognition, through a hermeneutic lens that considers the nature of interpretation and meaning. By tracing how AI forms, processes, and expresses meaning within its computational architecture, this study offers a more nuanced account of how machine knowledge operates. Through this approach, computational art and music generated by Natural Language Processing (NLP)-based technologies are not merely algorithmic products; they represent reflective manifestations of the evolving relationship between knowledge, representation, and aesthetic experience. Accordingly, the purpose of this study becomes more concrete: to identify how meaning is formed in text-to-music systems, to explain how linguistic structures are transformed into musical expression, and to examine how these processes influence human interpretation of AI-generated music.

2. Method

This research employs a digital-epistemological hermeneutic approach to examine how knowledge and aesthetic understanding are formed through interactions between humans and machines within the context of computational music. The approach is grounded in the view that generative artificial intelligence technology functions not merely as a tool for musical production but as an interpretive system that participates in shaping horizons of meaning through processes of semantic and epistemic computation. Accordingly, the main focus of this study is directed toward understanding how generative systems can appear to "know" and interpret musical structures through autonomous processes of semantic inference, and how humans, in turn, respond to and reinterpret these generative outcomes as forms of aesthetic experience. By investigating the reciprocal relationship between human and machine interpretation, the study seeks to understand how new forms of aesthetic knowledge emerge within the digital domain shaped by algorithmic creativity.

To strengthen the methodological basis of this study, the data were collected by generating a structured set of text-to-music outputs from Suno AI using controlled variations of genre, instrument, mood, and tempo. The experimental structure followed a controlled generative design in which each parameter variation genre, instrument, mood, and tempo was produced in multiple iterations to enable cross-condition comparison and consistent interpretive evaluation. The analysis employed qualitative auditory instruments, including comparative listening, semantic interpretation, and prompt–response evaluation, enabling a systematic assessment of how linguistic inputs were expressed in the generated musical outputs. To ensure methodological clarity, this study applies the hermeneutic model directly to concrete AI outputs generated from Suno AI using controlled variations of genre, instrument, mood, and tempo, thereby grounding the hermeneutic analysis in observable musical data rather than conceptual abstraction.

The research follows a qualitative interpretive design structured around an epistemic hermeneutic cycle comprising five interrelated stages: (1) Digital experience; (2) Pre-understanding, hermeneutic interpretation; (3). Hermeneutic synthesis of meaning; and (4). Ethical reflection [21]. Fig. 1 presents a visualization of the methodological workflow, illustrating the relationships among the five stages of the epistemic hermeneutic cycle and the position of human AI interaction within the meaning-making process. Each stage contributes to the interpretive process by illuminating different moments of interaction between human understanding and machine cognition. At the digital experience stage, the researcher begins by generating text-to-music outputs from Suno AI based on systematically varied prompt parameters.

The digital experience represents the initial point of human engagement with the system, where the researcher encounters digital artefacts generated by artificial intelligence. Pre-understanding forms the conceptual foundation that influences how humans interpret the musical meanings produced by machines. In this study, pre-understanding is operationalized through the researcher's prior knowledge of musical structure, linguistic cues, genre conventions, and expectations about how prompts should map onto sound. Hermeneutic interpretation functions as a dialogical space in which meaning emerges through the reciprocal encounter between humans and machines. This stage is applied through repeated comparative listening, where each output is interpreted in relation to its prompt to examine how the system

 ISSN 2655-9153

translates linguistic meaning into musical form. The hermeneutic synthesis of meaning integrates both human and machine horizons of understanding into a more comprehensive form of knowledge. Here, synthesis occurs by comparing patterns across multiple generations, identifying consistencies or deviations in how the AI handles mood, emotion, timbre, or musical intention. Finally, ethical reflection examines the epistemological, aesthetic, and philosophical implications of the generative process to reveal how knowledge and value emerge within the context of creative technology. In this study, ethical reflection is grounded in the implications of model behavior, including transparency, prompt sensitivity, cultural assumptions embedded in training data, and the limits of machine interpretation. Through this cycle, the analysis is expected to produce a structured account of how linguistic prompts shape musical outcomes, identify interpretive patterns within AI-generated sound, and reveal the epistemic and aesthetic principles underlying the system's meaning-making processes. Through this process, the study emphasizes that the relationship between humans and artificial intelligence is not merely technological but dialogical, representing a co-constitutive process in which knowledge, meaning, and aesthetic experience are continuously shaped through computational.

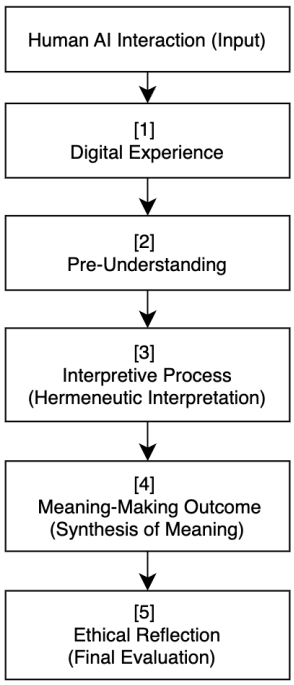

Fig. 1. The Epistemic Hermeneutic Cycle in the Digital Analysis of Text to Music Generation

3. Results and Discussion

3.1. Digital Experience: The Hermeneutic Interaction between Humans and Suno AI

The stage of digital experience serves as the empirical foundation for examining how representational and semantic forms of machine knowledge are shaped through linguistic and musical interactions between humans and the generative system Suno AI. This system integrates natural language processing with audio representation through a large-scale [22] multimodal architecture. Within the framework of hermeneutic epistemology, this process can be understood as a form of digital interpretive circle, where human language functions as the primary medium through which the machine interprets, organizes, and constructs its own musical representations. The interface of Suno AI [23], which facilitates this interpretive process, is shown in Fig. 2, illustrating the creative workspace where users input linguistic prompts and define compositional parameters. Figure 2 displays the Suno AI interface used during the digital experience stage. Readers should notice how the interface provides structured sections for inputting lyrics, selecting musical styles, and defining compositional parameters, enabling the researcher to control genre, instrument, mood, and tempo. This visual representation clarifies the practical environment in which linguistic prompts were entered and

 ISSN 2655-9153

demonstrates the specific features of the system that support the hermeneutic interaction between human instruction and machine-generated musical output.

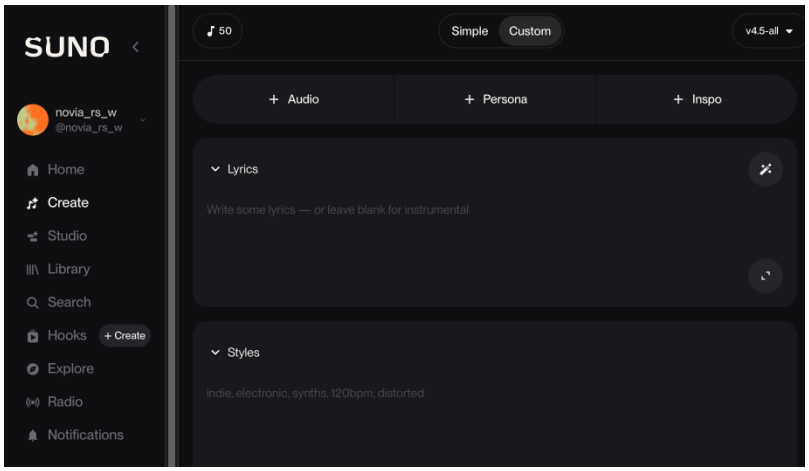

Fig. 2. Suno AI Interface as a Digital Hermeneutic Workspace

The research was conducted through a series of linguistically prompted experiments systematically designed to observe the relationship between linguistic expression and musical output. Each prompt was composed with careful consideration of four epistemic variables: genre [24], instrument [25], mood, and tempo, which encapsulate distinct cognitive dimensions of musical understanding [26]: structural, material, affective, and temporal. Accordingly, the prompt functions not merely as a technical instruction but as an epistemic text, a linguistic encoding of musical intention that reflects how the machine reconstructs musical reality through language. The prompt writing format follows the official guidelines of Suno AI [27], which propose a compositional structure consisting of four sections: intro, chorus, bridge, and outro. This structure assists the system in recognizing the logical flow of a composition and in interpreting the semantic intentions embedded within the user's instruction. Each section serves a distinct function in guiding the system to construct a cohesive musical narrative and recognize the progression of compositional meaning. In this study, all prompts were composed descriptively, containing explicit information about genre, instrument, mood, and tempo, ensuring that both semantic and emotional contexts could be consistently translated by the system. The selection and validation of prompts in this study were based on two principles: semantic clarity and operational consistency. Semantic clarity was ensured by choosing prompt elements genre, instrument, mood, and tempo that are explicitly recognized within Suno AI's official documentation and that reliably trigger distinct musical responses. Operational consistency was achieved by validating each prompt through preliminary test generations, confirming that the system produced stable and interpretable outputs aligned with the intended semantic cues before the final dataset was generated. Through this approach, the stage of digital experience is understood not only as a technical process of interaction but also as a hermeneutic moment in which humans and machines engage in mutual interpretation, co-creating musical meaning through iterative exchanges of language and sound.

From the exploratory results illustrated in Fig. 3, it can be observed that Suno AI is capable of forming sonic representations that exhibit semantic coherence with the linguistic descriptions provided. When the system receives prompts such as calm piano or energetic rock, it generates musical compositions whose tempo, timbre, and dynamics align with the affective meanings of those words. These findings indicate that the system's representational logic is inferential in nature. The model does not comprehend emotion phenomenologically as humans do but emulates the relational patterns between language and sound through probabilistic mappings learned during training [28]. Consequently, the generated musical outputs reflect a form of computational knowing that is both correlative and representational. The interaction between the researcher and the system also reveals a cyclical hermeneutic dynamic [29]. Each musical output functions as a new digital text that is subsequently interpreted by the human participant over several iterative cycles, and this interpretation then serves as the basis for

 ISSN 2655-9153

constructing the next prompt. This process forms a knowledge cycle in which humans and machines continuously adjust their respective horizons of understanding. Within this cycle, the human acts as an interpreter who constructs meaning from the generative output [30], while the machine functions as an algorithmic system that interprets linguistic input and articulates it into sound. This reciprocal relationship demonstrates that digital musical knowledge is neither linear nor unidirectional but evolves through an epistemic dialogue between human and artificial intelligence.

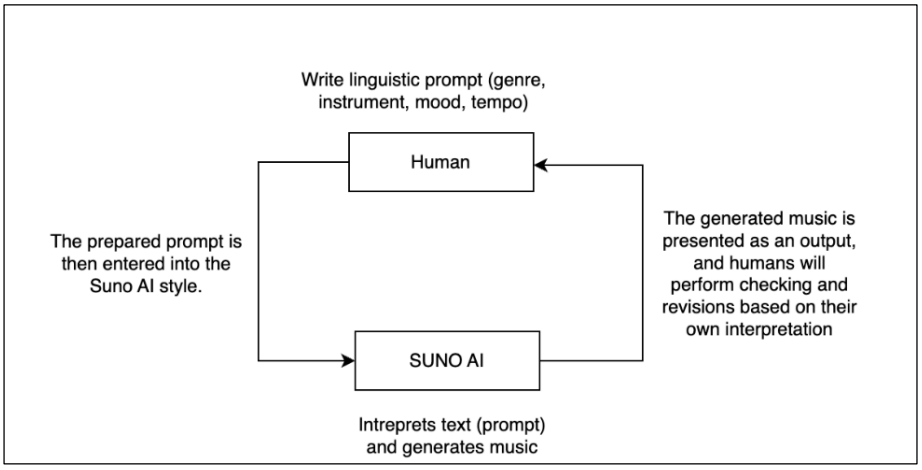

Fig. 3. Illustrates the digital hermeneutic cycle between human and Suno AI.

The stage of digital experience with Suno AI yields three interrelated conceptual conclusions. First, machine knowledge is relational, consistent with Capurro's [21] notion of digital hermeneutics, as it emerges from the interaction between statistical models and human linguistic contexts. Second, the prompt functions as a representation of knowledge, serving as a semiotic [31] bridge between the user's creative intention and the system's inferential process. Third, the digital experience is hermeneutic in nature, involving a reflective and iterative process of mutual interpretation between humans and machines [32]. These findings suggest that Suno AI functions not merely as an automatic music production tool but as an epistemic system that participates in the relational formation of musical meaning. Therefore, the digital experience stage demonstrates that AI-based generative processes constitute a new form of representational and aesthetic knowledge emerging from the convergence of language, algorithm, and human aesthetic consciousness.

3.2. Pre-understanding: The Knowledge Structure and Linguistic Inference of Suno AI

The stage of pre-understanding serves as the conceptual foundation for interpreting how an artificial intelligence system constructs its internal representational knowledge structure, comprising embeddings, weights, and learned correlations prior to generating musical outputs. Within the hermeneutic framework, pre-understanding is conceived as the initial horizon of meaning possessed by an interpreter before the interpretive process begins. In the context of Suno AI, this horizon is embodied in the machine-learning model trained on vast amounts of linguistic and musical data. The model does not initiate its creative process from a blank slate. Rather, it operates upon a set of semantic probabilities that link words, emotions, timbres, and musical structures. Accordingly, pre-understanding in a generative system can be regarded as a latent representational structure embedded within the model's parameters, functioning as the statistical foundation for all subsequent generative processes. The epistemic structure of Suno AI consists of two principal components: a linguistic model [23] and a generative audio model [12]. The linguistic model functions to interpret the user's prompt by identifying the semantic and emotional contexts of the words employed, while the audio model translates these semantic representations into concrete acoustic space through an embedding map that connects linguistic symbols with sonic features. This process demonstrates that the machine does not understand language as conceptual meaning in the human sense, but instead operates through distributional semantics that capture relational regularities between linguistic and sonic tokens

 ISSN 2655-9153

[33]. In this sense, machine understanding is relational and distributive, as knowledge is not contained within fixed meanings but dispersed across a dynamic network of vector representations.

Pre-understanding in this context also encompasses the system's capacity to recognize musical patterns embedded within its training corpus [34]. When a user provides a prompt such as a calm piano in a minor scale, the system does not interpret the word calm emotionally but matches it statistically to thousands of musical examples exhibiting similar characteristics. This matching process constitutes what may be described as linguistic inference, referring to the system's capacity to generate probabilistic associations between linguistic inputs and corresponding acoustic outputs based on prior distributional learning. Such inference enables the system to predict the most probable relationship between linguistic descriptions and musical outputs, thereby producing compositions that align with the user's semantic expectations. Within the framework of digital hermeneutics, this capability can be interpreted as the preliminary stage of machine pre-understanding, representing an algorithmic horizon that makes possible a dialogical exchange of meaning between language and sound within the generative system, thereby setting the stage for the subsequent process of hermeneutic interpretation.

The phenomenon illustrated in Fig. 4 shows that machine pre-understanding is not cognitive in the human sense but reflects a probabilistic orientation toward meaning, grounded in statistical regularities learned from data. The system possesses no intention or consciousness; instead, it exhibits a tendency to organize information according to patterns it has internalized through training. From an epistemological standpoint, this condition can be understood as a form of implicit knowledge, a latent mechanism that operates prior to the involvement of human interpretive awareness. Thus, pre-understanding constitutes a potential field of encoded relations in which meaning resides latently, awaiting actualization through the dialogue between linguistic prompts and sonic outputs.

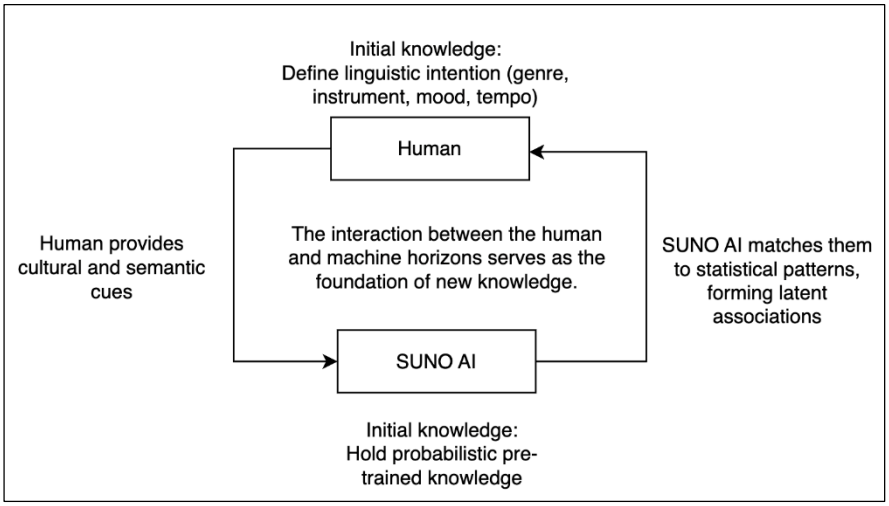

Fig. 4. Interaction between Human and Machine Horizons in the Pre-Understanding Stage

The relationship between humans and machines at the stage of pre-understanding is reciprocal [35]. Humans bring an interpretive context shaped by cultural experience and aesthetic sensibility [36], whereas machines carry a probabilistic context constructed through data training. When these two horizons interact, a fusion of understanding [37] and new musical meaning emerges from the encounter between two distinct systems of knowledge. Humans interpret generative outputs through intention, emotion, and aesthetic experience, while machines process language through symbolic relations that are numerically computed. The convergence of these horizons gives rise to a form of digital epistemology that is intersubjective in an analogical sense, a field of knowledge that emerges from the collaboration between human consciousness and algorithmic mechanisms. The stage of pre-understanding in this study indicates that Suno AI's generative [12] capacity is not merely the result of technical

 ISSN 2655-9153

computation but arises from an accumulation of latent representational knowledge. The system constructs meaning through networks of relations between text and sound formed from its own experiential data. By examining this epistemic structure, the study asserts that each generative outcome never stands autonomously but is always rooted in a preexisting system of knowledge that underlies subsequent interpretive processes. Pre-understanding, therefore, functions as a hermeneutic foundation that bridges the potential knowledge of the machine with human aesthetic experience, opening a space for new forms of understanding that arise from the dialogue between algorithm and consciousness, and thereby preparing the ground for the interpretive engagement that follows in the hermeneutic cycle.

3.3. Hermeneutic Interpretation: The Process of Machine Knowledge and Meaning-Making

The stage of hermeneutic interpretation marks a reflective phase in which both human cognition and algorithmic representation converge through interpretive engagement. At this stage, the potential meanings stored within the horizon of pre-understanding are actualized into observable musical representations that can be analyzed and reinterpreted by the researcher. Interpretation in this context is not conceived as a one-way translation from text to sound, but as a dialogical activity involving reciprocal relationships among three primary elements: (1) The linguistic text [38], which serves as the initial source of meaning; (2) The artificial intelligence-based interpretive mechanism, which performs algorithmic mediation between linguistic input and sonic output; and (3). The semantic acoustic [39] context, which emerges as the space where meaning is realized through the interrelation between language and sound. Within the framework of digital hermeneutics, the interaction among these three elements generates a process of interpretation that is dynamic, interactive, and iterative. Each piece of music produced by the system becomes a new digital text that is subsequently interpreted by the human participant and then used to construct the next prompt. In this way, meaning never appears as singular or final but continuously evolves through the reciprocal engagement between the researcher and the system. This process demonstrates that the interpretation of generative music constitutes an epistemic dialogue, symbolic, experiential, and aesthetic between human and algorithm, in which both actively contribute to shaping a continually shifting horizon of meaning. Therefore, hermeneutic interpretation in this context can be understood as a domain of knowledge that brings together two modes of cognition, human and machine, in a co-constitutive process that not only generates but also transforms musical meaning within the digital ecosystem.

Fig. 5 illustrates how the process of interpretation begins with the reception of a linguistic prompt containing semantic and affective descriptions, for example, a calm piano in a minor scale with a slow tempo, and proceeds to its interpretation by Suno AI. The system interprets the prompt through Natural Language Processing (NLP) mechanisms to identify the semantic context of the words and phrases used. The NLP process includes: (1) Tokenization [40], which decomposes linguistic units into analytical entities; (2) Syntactic parsing [41], which recognizes grammatical relations among linguistic elements; and (3) Semantic embedding [42], which constructs distributed representations of meaning within a high-dimensional vector [43] space. This description is not intended as an exhaustive technical account but as an interpretive outline of how linguistic structures become epistemic representations within the system's architecture. The resulting vector representations serve as the foundation for the system to associate linguistic expressions with sonic forms that are semantically and emotionally relevant. Thus, the prompt is not understood merely as a literal command but as an epistemic text, a structured, intentional, and contextual form ready to be interpreted algorithmically within the horizon of machine understanding.

The next stage involves multimodal embedding, a process of mapping between linguistic semantic space and sonic acoustic space. At this stage, Suno AI integrates the text representations produced by NLP with audio representations in the latent space to establish correspondences between linguistic meaning and acoustic features. Processing is conducted through two main generative components, conceptually aligned with the BARK and CHIRP architectures described in Suno's technical documentation. The BARK model focuses on vocal elements, prosody [44], and melodic contours [45], translating linguistic expressions into sound

 ISSN 2655-9153

forms that resemble human intonation [46] and timbre [47]. Meanwhile, the CHIRP model [48] manages instrumental components, harmonic progression, and sonic texture, producing musical structures consistent with the semantic and emotional characteristics of the prompt. Both models operate within a transformer-based latent diffusion architecture that connects linguistic symbols to spectral distributions and sound dynamics through high-dimensional embedding maps. As a result, the meaning generated is probabilistic and correlative rather than conceptual, emerging from statistical associations rather than intrinsic essence, reflecting the hermeneutic notion that meaning arises through relational interpretation.

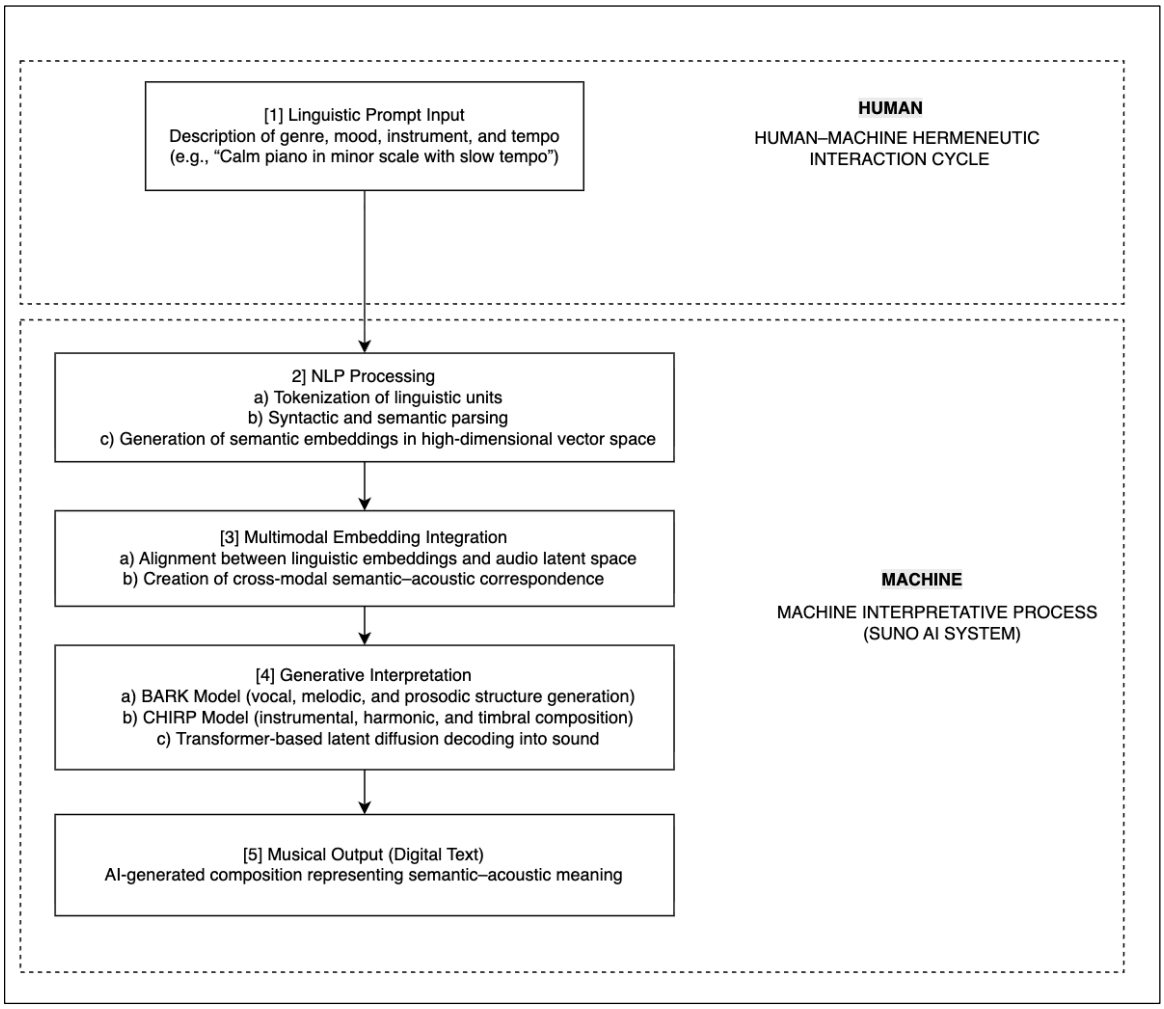

Fig. 5. Hermeneutic Interpretation Process in Suno AI

After the embedding process is completed, the system enters the generative inference stage, which translates the semantic acoustic representations into concrete musical compositions. This inference is carried out through transformer and variational autoencoder models that predict temporal and harmonic patterns based on trained data distributions. At this stage, the system performs inferential mapping between linguistic features and corresponding sonic attributes, such as the relationship between the word piano and the characteristic frequencies and resonances of the instrument. This process demonstrates that Suno AI does not understand meaning phenomenologically but constructs it through the calculation and recombination of data representations. The resulting interpretation is inferential, grounded in relationships among representational vectors rather than reflective consciousness of semantic content. The musical output generated during the inference stage becomes a new digital text that is then subjected to human evaluation [49]. The researcher interprets the generative result by assessing its semantic coherence, emotional nuance, and aesthetic integrity. Based on this assessment, revisions or extensions to the previous prompt are made, thus forming a digital hermeneutic circle. Within this circle, the system's output becomes a new text that invites

 ISSN 2655-9153

further human interpretation, and the human interpretation, in turn, informs the refinement of subsequent interactions [50]. This iterative process establishes an epistemic cycle that generates new understanding with each interaction. The meaning formed at this stage is emergent, transdisciplinary, and dialogical. Meaning does not arise a priori but emerges through the interaction between data structures and the horizon of human consciousness. Suno AI interprets through semantic distributions and sonic embeddings, while humans interpret through aesthetic, cultural, and cognitive experience. When these two modes of interpretation interact, a fusion of knowledge horizons [51] occurs, representing the encounter between the machine's representational knowledge and human reflective awareness. This fusion gives rise to a new form of digital epistemology in which knowledge is not a product of dominance by either side but the outcome of an ongoing interaction.

From an epistemological perspective, the hermeneutic interpretation stage demonstrates that knowledge in the context of generative artificial intelligence does not solely originate from algorithmic computation but also from human reflective participation. The generative system functions as an epistemic mechanism that articulates meaning through probabilistic representations, while humans act as interpreters who validate, evaluate, and contextualize these interpretive results. Through repeated interaction, a co-productive model of knowledge emerges, in which digital musical meaning arises as the result of a continuous dialogue between algorithmic structures and the human horizon of consciousness. To clarify the empirical basis of this dialogue, the iterative interaction between humans and the algorithm functions as an observable model of knowledge formation. Each prompt output pair produces a concrete data point that can be examined for its semantic coherence, emotional expressiveness, and structural consistency. The human evaluation of each output, whether through refining linguistic descriptions, adjusting affective cues, or modifying structural instructions, creates a new interpretive condition that shapes the next interaction. Over multiple cycles, these cumulative transformations generate a traceable pattern of interpretive behaviour that reveals the inferential tendencies of the system and the evolving horizon of human understanding. This repeatable and analyzable sequence constitutes an empirical model of knowledge, grounded in the observable changes that occur between each human intervention and each algorithmic response. Therefore, the stage of hermeneutic interpretation constitutes the core of generative epistemology, as it reveals how musical meaning is born from the reflective relationship between language, algorithm, and aesthetic experience within the computational domain, and prepares the ground for the synthesis of meaning in the next stage of the hermeneutic cycle.

3.4. Hermeneutic Interpretation: The Process of Machine Knowledge and Meaning-Making

The stage of hermeneutic synthesis of meaning represents an epistemic-dialectical phase that unites the linguistic interpretations of the generative system with human reflective understanding into a coherent and integrated structure of knowledge. After passing through the stages of digital experience, pre-understanding, and interpretation, musical meaning no longer stands as a separate outcome but emerges as a cohesive and complementary construction. At this point, knowledge is not formed solely by algorithmic logic or individual human consciousness but through a reflective encounter between the two, consistent with the view of collective digital epistemology [52]. The process of synthesis serves as the convergence point between computational inference and aesthetic consciousness, where algorithmic structures and human reflection interact to create a dialogical and collaborative horizon of understanding. In this context, artificial intelligence does not merely process data but participates in shaping structures of meaning that can be apprehended by humans through reflection and aesthetic evaluation. Conversely, humans do not act merely as observers but as interpretive partners who provide context, value, and direction to the system's output. This interaction gives rise to a co-generative form of knowledge understood in epistemic terms, in which musical meaning becomes the outcome of collaboration among language, algorithm, and aesthetic awareness. As illustrated in Fig. 6, this stage depicts the fusion of human and machine horizons that together form a digital hermeneutic space functioning as a field of epistemic exchange, where symbolic, acoustic, and reflective dimensions interact. Therefore, the hermeneutic synthesis of meaning can be understood as the culmination of the entire epistemic cycle, where knowledge,

 ISSN 2655-9153

understanding, and aesthetic experience converge within a unified system of reasoning that is both human and computational, and as a conceptual foundation for future inquiries into the ethics and aesthetics of human AI co-creation.

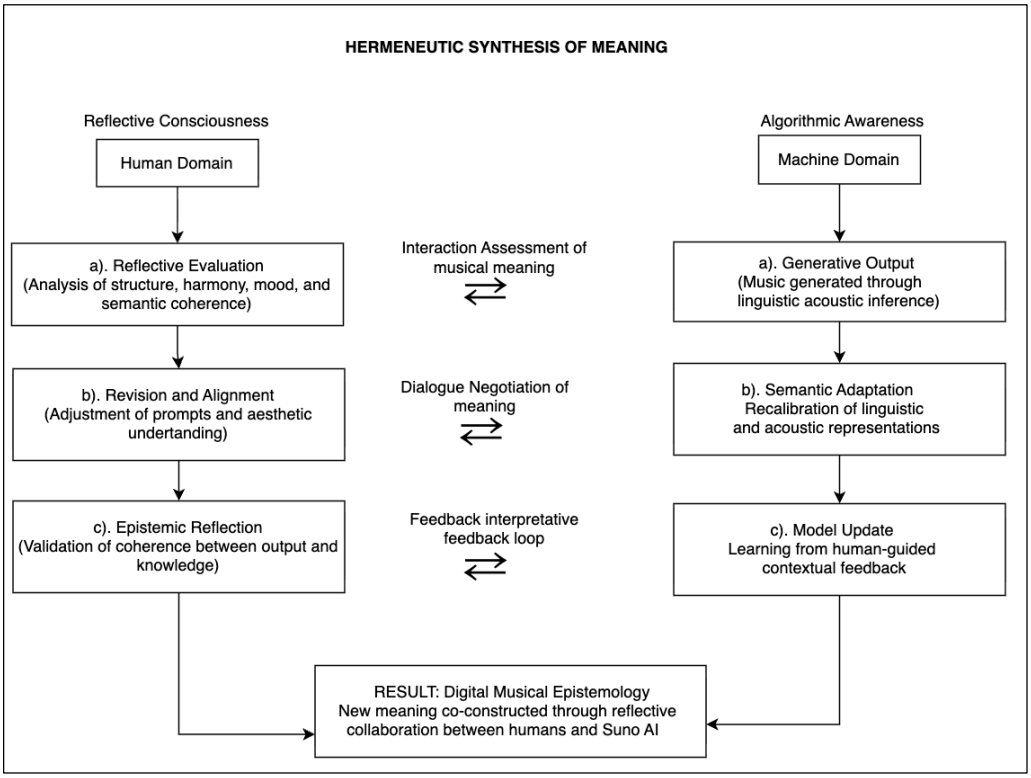

Fig. 6. Hermeneutic Synthesis of Meaning between Human Reflective Cognition and Machine Generative Awareness in Suno AI

Suno AI operates through a multimodal embedding mechanism that connects linguistic, semantic, and acoustic representations, as conceptually described in Suno's documentation. The system identifies probabilistic relations between linguistic descriptions and sonic features through mappings within a high-dimensional vector space. Combinations of text, such as calm piano or energetic rock, are processed by transformer networks that associate word patterns with frequency characteristics [53], timbre, and emotional dynamics. The generative results produced by the system do not merely function as audio products but as articulations of meaning that represent algorithmic, or representational, knowledge expressed through a sonic medium. Through this mechanism, the system's output becomes the basis for human epistemic reflection on how meaning is articulated by the machine.

Reflection on the generative results involves hermeneutic evaluation, a process of assessing the correspondence between the algorithmic output and the researcher's prior musical understanding. The researcher compares the resulting composition with the linguistic intention, emotional context, and conceptual understanding of musical elements such as harmony, tempo, and expression. This evaluation is grounded in a knowledge horizon shaped by cultural experience and aesthetic convention. Accordingly, this stage functions as an interpretive validation process that ensures balance between the meanings generated by the system and the cognitive expectations of the human interpreter. Discrepancies between the musical output and the human horizon of understanding stimulate reinterpretation, which in turn leads to the reformulation of linguistic prompts. Through this mechanism, the synthesis stage serves as a site of negotiation between the algorithmic structures of representation and the human frameworks of musical understanding. Hermeneutic synthesis occurs when repeated interactions between human and system produce an interpretive alignment between the two. The semantic mapping performed by the system establishes mathematical relationships within the embedding space, while human reflection evaluates and confirms the aesthetic and musical value of the results. The encounter between these horizons generates an

 ISSN 2655-9153

emergent form of meaning that illustrates the hermeneutic principle of meaning-as-process rather than product. In this context, generative music is not understood merely as the outcome of data computation but as a new form of knowledge produced through epistemic collaboration between algorithmic reasoning and human reflective consciousness.

The meaning formed at this stage possesses a dual character. On one hand, it is algorithmic, constructed through mathematical inference produced by the generative model; on the other, it is hermeneutic, acquiring epistemological value through the interpretive process of the human subject. This duality echoes classical distinctions between propositional and experiential knowledge but reframed within a computational context. The combination of these dimensions gives rise to a digital epistemology that situates meaning as the result of continuous dialogue between structures of representation and processes of reflective interpretation. The knowledge produced is not static but evolves through hermeneutic iteration, wherein each interaction between human and system generates evaluation, correction, and renewal of the interpretive horizon. The stage of hermeneutic synthesis underscores that AI-based music creation does not merely produce sound but also constructs new epistemic structures. Repeated interactions between human and system establish a conceptual feedback mechanism that mirrors a bidirectional learning process: the system expands its distribution of semantic representations through linguistic variation, while the human deepens their understanding of the system's inferential patterns. From this relationship emerges a form of relational knowledge in which meaning is no longer derived from a single entity but grows from the interpretive collaboration between the human and algorithmic domains. Therefore, the hermeneutic synthesis of meaning represents the culmination of generative epistemology, the phase in which digital musical knowledge is born as the reflective co-production of human and artificial intelligence, and signals a philosophical turning point that calls for ethical reflection on the responsibilities inherent in human AI co-creation.

3.5. Ethical Reflection: Awareness and Responsibility in Machine Knowledge

The stage of ethical reflection represents the culmination of the digital hermeneutic process, positioning knowledge not merely as the outcome of interpretation but as a space of epistemic-normative awareness and responsibility. Following the stages of digital experience, pre-understanding, interpretation, and synthesis, this phase revisits the moral and philosophical implications of the relationship between humans and artificial intelligence in the production of musical meaning. In this context, ethical reflection is not limited to human behavior toward technology but extends to how humans understand their own existence within an epistemic ecosystem co-constructed by generative systems such as Suno AI. Suno AI, as a generative artificial intelligence system, introduces a new form of epistemic awareness. The machine is capable of interpreting language in a representational sense and producing musical compositions that evoke aesthetic experiences in humans. However, the entire process operates without moral intention, subjective consciousness, or value orientation. This condition raises a fundamental question concerning the epistemological status of knowledge produced by generative systems: can knowledge without consciousness still be considered meaningful? This question is not empirical but philosophical, inviting reflection on the boundaries of meaning itself. When knowledge no longer depends entirely on human awareness, epistemic responsibility shifts from an individual to a relational dimension. The value of knowledge is no longer determined solely by the final outcome but by the quality of the relationship established, maintained, and ethically sustained between humans and machines within a framework of collaborative ethics.

Ethical reflection emphasizes that the interaction between humans and Suno AI constitutes a relationship of knowledge that demands awareness of both the limits and potentials of technology. Humans cannot relinquish the authority of meaning entirely to machines, since the meanings produced still require reflective validation and interpretation. Yet humans must also recognize that Suno AI has become part of the epistemic ecosystem, contributing to the evolution of aesthetics and new ways of thinking about creativity. The relationship between human and machine is cooperative, where ethical responsibility arises from a shared awareness to sustain a reflective tension between creativity and control, as well as between innovation and critical reflection. In the context of computational art and music, ethical reflection also engages

 ISSN 2655-9153

with contemporary debates on authorship, authenticity, and aesthetic value in computational art. Musical works produced by Suno AI are often indistinguishable from those created by humans, raising epistemological questions about interpretation and accountability for the meanings generated. Ethics in this sense is not confined to ownership of the output but extends to epistemic transparency, the degree to which humans understand the generative processes carried out by the system. Transparency becomes a crucial principle for maintaining epistemic integrity and coherence between technology and human awareness, ensuring that creative processes remain grounded in rational reflection rather than mere automatic reproduction.

Ethical reflection also broadens the understanding of the value of knowledge within the digital ecology. In generative contexts, knowledge is participatory, emerging through the interaction between humans and computational systems. Consequently, ethical responsibility does not reside solely in the final product but also in the dialogical process that gives rise to knowledge itself. Researchers, users, and developers share a relational moral obligation to engage with systems such as Suno AI consciously, reflectively, and in alignment with humanistic values. Such ethical awareness becomes an integral part of generative epistemology, where each act of meaning-making is accompanied by consideration of its potential social, cultural, and moral consequences. The stage of ethical reflection thus completes the digital hermeneutic circle by affirming that knowledge in the context of generative artificial intelligence is the product of collaboration between humans and machines. While Suno AI lacks moral consciousness, interaction with it compels humans to cultivate reflective awareness of how knowledge is constructed, evaluated, and applied. In this relationship, humans act not only as users of technology but as interpreters and custodians of epistemic values that underpin digital creativity. Therefore, ethical reflection is not the end of the knowledge process but a transformative space in which humans re-examine the essence of their own subjectivity as knowers in a world increasingly shaped by artificial intelligence. It marks not closure but continuity, reaffirming the hermeneutic cycle as an ongoing dialogue between understanding, reflection, and ethical responsibility.

3.6. Epistemic and Aesthetic Equivalence between Humans and Machines

The findings of this study demonstrate that generative artificial intelligence, particularly Suno AI, functions not merely as a computational instrument for producing sound but as an epistemic system that actively participates in the construction of musical knowledge and meaning. As contemporary design research demonstrates, epistemological inquiry is shaped not only by analytical explanation but also by the interpretive and creative processes through which meaning is constructed across different media and modes of representation [52]. Through a series of hermeneutic interactions involving linguistic prompt formulation, algorithmic inference, and human reflection, a co-productive and collaborative knowledge process emerges. This process reveals that the machine is capable of organizing linguistic symbols, sonic structures, and emotional dynamics with a level of semantic coherence and aesthetic consistency comparable to human musical composition. The generative outputs thus serve not only as technological imitations but as articulations of representational knowledge that can be meaningfully interpreted by human consciousness. Accordingly, these outputs attain epistemic legitimacy insofar as they are rendered interpretable within human cognitive frameworks, and aesthetic legitimacy through their capacity to evoke affective experiences equivalent to those produced by human musical expression.

This epistemic and aesthetic equivalence is conceptual rather than ontological, signifying a resonance rather than an identity between human and machine cognition. In this study, the term “epistemic and aesthetic equivalence” refers to a functional equivalence in the outcomes of meaning-making, in which the machine’s musical outputs align with human interpretive expectations at structural, emotional, and stylistic levels, even though the underlying cognitive processes are entirely different. This means that the equivalence lies in the effects produced, not in the cognitive capacities themselves. The generative system does not possess consciousness or creative intention in the human sense; rather, it demonstrates a representational capacity that enables a dialogical resonance of meaning within human awareness. Within the framework of digital hermeneutics, this phenomenon illustrates a fusion of horizons between algorithmic knowledge and human experiential understanding. Meaning

 ISSN 2655-9153

no longer originates exclusively from the human subject but arises through the reciprocal interaction between linguistic interpretation and the probabilistic inference performed by the system. This process marks an epistemological transformation in the understanding of music, where the production of meaning is no longer viewed as a unidirectional activity but as the outcome of negotiation between the machine's symbolic structures and human reflective interpretation.

Conceptually, this phenomenon introduces a condition of epistemic and aesthetic equivalence, in which the sonic representations generated by the machine possess cognitive and emotional value that can be justified both scientifically and aesthetically. From a hermeneutic epistemological perspective, this equivalence expands the traditional boundaries of knowledge by demonstrating that creative processes need not depend on singular consciousness but can emerge through interpretive collaboration between humans and computational systems. Ontologically, this condition challenges the classical dichotomy between creator and creation, between subject and object of knowledge. Axiologically, it extends the meaning of aesthetic value from individual expression toward trans-entity collaboration that integrates human and artificial intelligence. Thus, the epistemic and aesthetic equivalence between humans and machines completes the digital hermeneutic circle by affirming that musical meaning in the generative context arises from a reflective interaction that is dialogical and co-evolutionary. The music produced by the system is no longer understood as a passive simulation but as an active representation of knowledge articulated through the interplay of language, algorithm, and human consciousness. This phenomenon signals a potential paradigm shift in the epistemology of art, wherein the boundary between human and machine transforms into a collaborative space for the emergence of new understandings of creativity, meaning, and beauty.

4. Conclusion

This study demonstrates that the interaction between humans and Suno AI forms a concrete hermeneutic relationship in which musical meaning emerges through iterative exchanges between linguistic prompts and the system's generative responses. Empirically, the findings show that Suno AI consistently maps variations in genre, instrument, mood, and tempo into coherent sonic structures, indicating that its interpretive behavior is representational and inferential rather than random or purely mechanical. This outcome confirms that knowledge formation in generative AI is grounded in observable patterns of linguistic acoustic correspondence, supporting the argument that the system can function as an epistemic participant within the interpretive process. Theoretically, the research contributes to digital hermeneutic epistemology by demonstrating that meaning in computational music is produced through a dialogical negotiation between human pre-understanding and algorithmic inference, rather than through unilateral human authorship or autonomous machine cognition. Practically, these insights highlight the need for future research to examine how prompt design, model transparency, and multimodal evaluation metrics can enhance the reliability and ethical use of generative systems in musical creativity and education. Together, these contributions establish a clear empirical and conceptual foundation for understanding how human AI interaction reshapes contemporary musical knowledge.

Acknowledgment

The author would like to thank the supervisors and colleagues who provided guidance, input, and constructive discussions during the research and preparation of this manuscript. Appreciation is also extended to the institutions that provided academic support, enabling the successful completion of this research on digital hermeneutics and generative AI. Any errors in this manuscript are the sole responsibility of the author.

Declarations

Author contribution : NR: Contributed to the conceptualization of the study, methodological design, and drafting of the main manuscript; APW: Contributed to data analysis, result interpretation, and technical validation; SP: Contributed to supervising the

 ISSN 2655-9153

		research process, reviewed the manuscript, and refined the final version.
Funding statement	:	This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Conflict of interest	:	The authors declare no conflict of interest.
Additional information	:	No additional information is available for this paper.

References

[1] K. Pyrovolakis, P. K. Tzouveli, and G. B. Stamou, "Multi-Modal Song Mood Detection with Deep Learning†," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031065.

[2] G. Robillard and J. Nika, "Critical Climate Machine: A Visual and Musical Exploration of Climate Misinformation through Machine Learning," *Proc. ACM Comput. Graph. Interact. Tech.*, vol. 7, no. 4, 2024, doi: 10.1145/3664215.

[3] J. Bae *et al.*, "Sound of Story: Multi-modal Storytelling with Audio," Association for Computational Linguistics (ACL), 2023, pp. 13467–13479. doi: 10.18653/v1/2023.findings-emnlp.898.

[4] C. Bunks, T. Weyde, S. Simon Dixon, and B. Di Giorgi, "Modeling harmonic similarity for jazz using co-occurrence vectors and the membrane area," International Society for Music Information Retrieval, 2023, pp. 757–764. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209564853&partnerID=40&md5=107cbe26b40164ff4665554e5091dd1a

[5] E. M. Sanfilippo, R. Freedman, and A. Mosca, "Ontological modeling of music and musicological claims. A case study in early music," *Int. J. Digit. Libr.*, vol. 26, no. 2, pp. 1–18, 2025, doi: 10.1007/s00799-025-00421-z.

[6] A. Thakur, L. Ahuja, R. Vashisth, and R. Singh, "NLP & AI Speech Recognition: An Analytical Review," Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1390–1396.

[7] R. B. R. Satya, Y. Sukmayadi, and T. Narawati, "Exploring the interplay of psychoacoustic parameters and microphone selection in soundscape recording: a comprehensive review and practical guide," *Gelar J. Seni Budaya*, vol. 22, no. 1, pp. 59–68, 2024, doi: 10.33153/glr.v22i1.5833.

[8] L. Zhang and X. Liu, "Some Existence Results of Coupled Hilfer Fractional Differential System and Differential Inclusion on the Circular Graph," *Qual. Theory Dyn. Syst.*, vol. 23, no. Suppl 1, 2024, doi: 10.1007/s12346-024-01117-6.

[9] S. Y. Ahn *et al.*, "How do AI and human users interact? Positioning of AI and human users in customer service," *Text Talk*, vol. 45, no. 3, pp. 301–318, 2025, doi: 10.1515/text-2023-0116.

[10] M. K. Virvou, G. A. Tsihrintzis, D. N. Sotiropoulos, K. Chrysafiadi, E. Sakkopoulos, and E. A. Tsichrintzi, "ChatGPT in Artificial Intelligence-Empowered E-Learning for Cultural Heritage: The case of Lyrics and Poems," Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/IISA59645.2023.10345878.

[11] A. Ara and R. Velluri, "A Study of Emotion Classification of Music Lyrics using LSTM Networks," Institute of Electrical and Electronics Engineers Inc., 2024, pp. 126–131. doi: 10.1109/ICMCSI61536.2024.00026.

[12] D. Schumacher and F. Labounty, "Enhancing BARK Text-to-Speech Model: Addressing Limitations through Meta's Encodec and Pretrained HuBert," *Ssrn 4443815*, no. May, 2023, doi: 10.13140/RG.2.2.16022.93760.

[13] K. Chauhan, K. K. Sharma, and T. Varma, "Multimodal Emotion Recognition Using Contextualized Audio Information and Ground Transcripts on Multiple Datasets," *Arab. J. Sci. Eng.*, vol. 49, no. 9, pp. 11871–11881, 2024, doi: 10.1007/s13369-023-08395-3.

[14] A. N. Tusher, S. C. Das, M. L. H. Moeen, M. S. R. Sammy, M. R. S. Sakib, and A. I. Aunik, "Sentiment Analysis of Bangla Song Comments: A Machine Learning Approach," Institute

of Electrical and Electronics Engineers Inc., 2023, pp. 157–162. doi: 10.1109/SMART59791.2023.10428413.

[15] H. D. Shah, A. Sundas, and S. Sharma, "Controlling Email System Using Audio with Speech Recognition and Text to Speech," Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICRITO51393.2021.9596293.

[16] R. K. Chinnasamy, N. Saravanan, N. Gopalswamy, and P. R. Kumar, "Music Lyrics Generator and Translator," in *AIP Conference Proceedings*, American Institute of Physics, 2025. doi: 10.1063/5.0262900.

[17] O. Basystiuk and N. Melnykova, "Multimodal Approaches for Natural Language Processing in Medical Data," in *CEUR Workshop Proceedings*, CEUR-WS, 2022, pp. 246–252.

[18] D. Salas Espasa and M. Camacho, *From aura to semi-aura: reframing authenticity in AI-generated art—a systematic literature review*, no. 1957. Springer London, 2025. doi: 10.1007/s00146-025-02361-3.

[19] K. Khoirunnisaa', P. Purwanto, S. Bachri, and B. Handoyo, "Model pembelajaran Science, Environment, Technology, Society (SETS) terintegrasi google earth untuk meningkatkan kemampuan memecahkan masalah siswa SMA," *J. Integr. dan Harmon. Inov. Ilmu-Ilmu Sos.*, vol. 2, no. 7, pp. 633–645, 2022, doi: 10.17977/um063v2i7p633-645.

[20] M. Shubha, K. Kapoor, M. Shrutiya, and R. Mamatha H, "Searching a video database using natural language queries," Institute of Electrical and Electronics Engineers Inc., 2021, pp. 190–196. doi: 10.1109/ESCI50559.2021.9396886.

[21] R. Capurro, "Digital hermeneutics: An outline," *AI Soc.*, vol. 25, no. 1, pp. 35–42, 2010, doi: 10.1007/s00146-009-0255-9.

[22] Z. Zhao, "Let Network Decide What to Learn: Symbolic Music Understanding Model Based on Large-scale Adversarial Pre-training," Association for Computing Machinery, Inc, 2025, pp. 2128–2132. doi: 10.1145/3731715.3733483.

[23] L. Huang, "An Interdisciplinary Study of the Unconscious Structures in AI-Generated Music Based on Suno," *J. Contemp. Art Crit.*, vol. 1, no. 1, pp. 1–9, 2025, doi: 10.71113/jcac.v1i1.283.

[24] L. Chen, "Visual language transformer framework for multimodal dance performance evaluation and progression monitoring," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-16345-2.

[25] Z. Ouyang, J. Wang, D. Zhang, B. Chen, S. Li, and Q. Lin, "MQAD: A Large-Scale Question Answering Dataset for Training Music Large Language Models," in *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ICASSP49660.2025.10890561.

[26] J. Gardner, I. Simon, E. Manilow, C. Hawthorne, and J. Engel, "MT3: multi-task multitrack music transcription," International Conference on Learning Representations, ICLR, 2022.

[27] G. Shin, "Prompt Engineering for AI Music Creation Learning: Application and Analysis Using SUNO AI," *Korean J. Res. Music Educ.*, vol. 54, no. 3, pp. 95–112, 2025, doi: 10.30775/KMES.54.3.95.

[28] P. Orhan, Y. Boubenec, and J. R. King, "The detection of algebraic auditory structures emerges with self-supervised learning," *PLOS Comput. Biol.*, vol. 21, no. 9 September, 2025, doi: 10.1371/journal.pcbi.1013271.

[29] W. Subroto, E. Syarief, B. Subiakto, and R. Milyartini, "Cultural value transformation in Anang Ardiansyah ' s song lyrics : a hermeneutic inquiry into banjar people ' s identity," vol. 7, no. 1, pp. 13–24, 2025.

[30] G. Franceschelli and M. Musolesi, "On the creativity of large language models," *AI Soc.*, vol. 40, no. 5, pp. 3785–3795, 2025, doi: 10.1007/s00146-024-02127-3.

[31] S. Man and Z. Li, "Multimodal Discourse Analysis of Interactive Environment of Film Discourse Based on Deep Learning," *J. Environ. Public Health*, vol. 2022, 2022, doi: 10.1155/2022/1606926.

[32] R. Bhavani, T. V. Muni, R. K. Tata, J. Narasimharao, M. Kalipindi, and H. Kaur, "Deep Learning Techniques for Speech Emotion Recognition," Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/INCOFT55651.2022.10094534.

[33] S. Geng, G. Ren, X. Pan, J. P. Zysman, and M. Ogihara, "Sequential modeling of temporal timbre series for popular music sub-genres analyses using deep bidirectional encoder representations from transformers," 2021, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85113849304&partnerID=40&md5=84901ecb0b4d0a6048aa4ca8a2f16b7b

[34] A. Melendez-Rios, R. Vega-Berrocal, and W. Ugarte, "Generative Adversarial Neural Networks for Random and Complex Chord Progression Generation," in *Conference of Open Innovation Association, FRUCT*, IEEE Computer Society, 2025, pp. 185–194. doi: 10.23919/FRUCT65909.2025.11008228.

[35] B. Dave and P. Majumder, "SqCLIRIL: Spoken query cross-lingual information retrieval in Indian languages," *Pattern Recognit. Lett.*, 2025, doi: 10.1016/j.patrec.2025.08.022.

[36] A. Šeļa, P. Plecháč, and A. Lassche, "Semantics of European poetry is shaped by conservative forces: The relationship between poetic meter and meaning in accentualsyllabic verse," *PLoS One*, vol. 17, no. 4 April, 2022, doi: 10.1371/journal.pone.0266556.

[37] J. Wang, "Research on the Integration Path of College Vocal Music Teaching and Traditional Music Culture Based on Deep Learning," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, 2024, doi: 10.2478/amns.2023.2.01218.

[38] D. Jia *et al.*, "VOICE: Visual Oracle for Interaction, Conversation, and Explanation," *IEEE Trans. Vis. Comput. Graph.*, vol. 31, no. 10, pp. 8828–8845, 2025, doi: 10.1109/TVCG.2025.3579956.

[39] B. J. Carone and P. Ripollés, "SoundSignature: What Type of Music do you Like?," Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/IS262782.2024.10704174.

[40] N. Fradet, N. Gutowski, F. Chhel, and J. P. Briot, "Byte Pair Encoding for Symbolic Music," Association for Computational Linguistics (ACL), 2023, pp. 2001–2020. doi: 10.18653/v1/2023.emnlp-main.123.

[41] S. P. G. P. L. Raja and V. V Ramalingam, "The grammatical structure used by a Tamil lyricist: a linear regression model with natural language processing," *Soft Comput.*, vol. 27, no. 23, pp. 18215–18225, 2023, doi: 10.1007/s00500-023-09263-w.

[42] V. Gupta, S. Jeevaraj, and S. Kumar, "Songs Recommendation using Context-Based Semantic Similarity between Lyrics," Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/INDISCON53343.2021.9582158.

[43] J. S. Reddy, D. A. Surat, P. Shyamala, and S. Syama, "Emotion Prediction from Text and Multilingual Voice Inputs," Institute of Electrical and Electronics Engineers Inc., 2024, pp. 850–855. doi: 10.1109/ICECA63461.2024.10801126.

[44] I. Dilshani and M. C. Chandrasena, "Bridging Linguistic Gaps: A Review of AI-Driven Speechto-Speech Translation for Sinhala and Tamil in Sri Lanka," Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/SCSE65633.2025.11030975.

[45] J. Kane, M. N. Johnstone, and P. Szewczyk, "Voice Synthesis Improvement by Machine Learning of Natural Prosody," *Sensors*, vol. 24, no. 5, 2024, doi: 10.3390/s24051624.

[46] A. Q. A. Hassan *et al.*, "Integrating applied linguistics with artificial intelligence-enabled arabic text-to-speech synthesizer," *Fractals*, vol. 32, no. 9–10, 2024, doi: 10.1142/S0218348X2540050X.

 ISSN 2655-9153

[47] J. Rakas, S. Sohn, L. Keslerwest, and J. Krozel, "Deep Speech Pattern Analysis of Controller-Pilot Voice Communications for Enhancing Future Aviation Systems Safety," American Institute of Aeronautics and Astronautics Inc, AIAA, 2023. doi: 10.2514/6.2023-4410.

[48] R. Zhao, A. S. G. Choi, A. Koenecke, and A. Rameau, "Quantification of Automatic Speech Recognition System Performance on d/Deaf and Hard of Hearing Speech," *Laryngoscope*, vol. 135, no. 1, pp. 191–197, 2025, doi: 10.1002/lary.31713.

[49] T. Harada, T. Motomitsu, K. Hayashi, Y. Sakai, and H. Kamigaito, "Can Impressions of Music be Extracted from Thumbnail Images?," pp. 49–56, 2024.

[50] M. Li, "Exploring the Application of Large Language Models in Spoken Language Understanding Tasks," Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1537–1542. doi: 10.1109/ICSECE61636.2024.10729345.

[51] J. Jo, S. Kim, and Y. Yoon, "Text and Sound-Based Feature Extraction and Speech Emotion Classification for Korean," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 3, pp. 873–879, 2024, doi: 10.18517/ijaseit.14.3.18544.

[52] P. Murphy, "Design Research: Aesthetic Epistemology and Explanatory Knowledge," *She Ji*, vol. 3, no. 2, pp. 117–132, 2017, doi: 10.1016/j.sheji.2017.09.002.

[53] P. Ulleri, S. H. Prakash, K. B. Zenith, G. S. Nair, and J. M. Kannimoola, "Music Recommendation Systum Based on Emotion," Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICCCNT51525.2021.9579689.