RAIS RESEARCH ASSOCIATION for INTERDISCIPLINARY STUDIES
November 20-21, 2025

DOI:10.5281/zenodo.18075922

A Multi-Agent AI Framework for Musical Score Writing

Scott Josephson[1], Dr. Atif Farid Mohammad[2]

[1]*Graduate School, Capitol Technology University, Maryland, USA, sjosephson@captechu.edu (Contact)*
[2]*Graduate School, Capitol Technology University, Maryland, USA, afmohammad@captechu.edu*

Abstract: The development of artificial intelligence (AI) has significantly influenced music composition, yet current single-agent AI systems face limitations in versatility, adaptability, and collaborative dynamics, often producing repetitive and stylistically rigid compositions. To address these challenges, this research proposes a multi-agent AI framework specifically designed for musical score writing, utilizing Agentic AI principles where autonomous, specialized agents collaboratively contribute elements such as melody, harmony, rhythm, and dynamics, emulating human compositional processes. Employing a mixed-method comparative design with quantitative assessments and qualitative evaluations by expert musicians, the study hypothesizes that multi-agent interactions enhance musical complexity, creativity, and coherence compared to single-agent systems. The expected findings aim to contribute significantly to computational creativity theory and practical applications in education, entertainment, and therapy, highlighting the potential of multi-agent systems to advance human-AI collaborative creativity.

Keywords: AI Agent, Computational Creativity, Multi-Agent Systems, Multimodal Large Language Model, Musical Score

Introduction

The development of artificial intelligence (AI) has significantly impacted the field of music composition, pushing boundaries traditionally defined by human creativity. AI-driven music composition involves computational systems capable of autonomously creating musical content, redefining what is achievable beyond traditional human-led creative processes (Boden, 2009). This revolutionary approach has provided new avenues for musical exploration and experimentation, enabling composers and musicians to transcend conventional compositional limits. Various AI methods, including rule-based systems, machine learning algorithms, neural networks, and deep learning frameworks, have been employed to compose music, analyze musical styles, and even perform real-time improvisation.

Background

However, despite substantial progress, current single-agent AI systems often lack the versatility, adaptability, and collaborative nature inherent in human music composition. Single-agent AI systems typically operate under predefined parameters, limiting their ability to adapt dynamically or generate diverse musical ideas spontaneously. These constraints often result in repetitive, predictable, or stylistically rigid compositions, failing to capture the depth and dynamism of human-generated music (Eigenfeldt, Bown, & Carey, 2015; Herremans, Chuan, & Chew, 2017). A particular challenge for these systems is modeling the long-term structure of music, as they struggle to remember and coherently develop musical events over extended periods. Furthermore, most generative AI systems have limited or no memory between sessions unless explicitly programmed, hindering their ability to maintain context over prolonged interactions.

Addressing these limitations necessitates exploring innovative computational models, notably multi-agent systems, which are a key implementation of Agentic AI. Agentic AI represents an advanced paradigm of AI systems characterized by their ability to independently execute complex tasks and solve problems through chain-of-thought reasoning and continuous planning. These systems are designed for high autonomy, enabling them to independently decide, act, and learn to achieve specific goals with minimal human intervention. Multi-agent systems comprise multiple autonomous agents that interact and

cooperate to achieve collective objectives, mirroring the collaborative aspects of human creative processes (Wooldridge, 2009). Such systems possess inherent advantages, including enhanced adaptability, scalability, and distributed problem-solving capabilities. This collaborative and distributed approach provides an opportunity to develop AI music systems that more accurately emulate human compositional methods and creative interactions, thereby generating richer and more sophisticated compositions in complexity and expression (de Berardinis et al. 2025).

This research aims to address these limitations explicitly by proposing a multi-agent AI framework specifically designed for musical score writing. In this framework, specialized agents are designed to perform distinct compositional roles, such as melody creation, harmonic structuring, rhythmic development, and dynamic articulation. Each agent independently generates content while communicating and coordinating with other agents to create a cohesive musical composition (Brown, 2025). This collective effort allows for emergent creativity, where the combined outputs surpass what any individual agent could achieve alone, akin to an orchestra guided by a conductor (Brown, 2025). For instance, a harmony agent would pursue the goal of creating a coherent chord progression. In contrast, a rhythm agent would proactively anticipate rhythmic needs, continuously learning and adapting their contributions based on feedback and the overall musical context. This distributed approach is designed to overcome the challenge of modeling long-term musical structure by distributing complexity across specialized, interacting agents, enabling them to maintain context and develop musical ideas over extended compositions (Priestly, 2025).

The potential of multi-agent systems to enhance creativity and efficiency is well-documented in other fields, such as robotics, gaming, and distributed computing. While multi-agent systems have been explored in music for tasks like improvisation and interactive performance, a comprehensive, hierarchical, and truly agentic framework specifically for complex musical score writing that fully leverages the principles of Agentic AI remains comparatively underexplored. Preliminary investigations have indicated promising outcomes, suggesting that multi-agent collaborations could significantly enrich the creative possibilities in AI-generated music (Brown, 2025). These systems can explore a broader musical space through diverse and complementary interactions, leading to more innovative and varied outcomes that more closely align with human creative processes (Miranda, 2001; Deng et al., 2024). This research aims to push beyond "exploratory creativity" (exploring within an established conceptual space) towards "transformational creativity" (deliberately transforming or transcending that space) (Jordanous, 2025).

This investigation's central research question is: "How does a multi-agent AI framework enhance the complexity, creativity, and coherence of generated musical scores compared to single-agent systems?" This question seeks to empirically validate the advantages of collaborative interactions among specialized AI agents. More specifically, the research will explore whether the collaborative nature inherent in multi-agent frameworks genuinely improves musical quality, originality, and alignment with human compositional styles. Understanding these dynamics can significantly inform future development strategies for AI-driven creative tools. By addressing these critical questions, this research contributes to computational creativity theory and holds substantial implications for practical applications. Enhanced AI Music systems can transform education, entertainment, and therapy by providing dynamically adaptive compositional tools. These systems could support music instruction in educational settings by generating customized exercises or complex compositional challenges, significantly enriching students' learning experiences (Jin et al., 2025). In entertainment contexts, multi-agent AI systems could produce tailored soundtracks or interactive musical compositions responsive to real-time inputs or narrative developments (Li & Wang, 2023). Furthermore, therapeutic applications could benefit from adaptive musical experiences that promote relaxation, cognitive engagement, or emotional regulation,

thus contributing positively to mental health and well-being (66degrees, 2025). Crucially, this research emphasizes a "human-in-the-loop" approach, where the AI catalyzes human creativity, helping artists overcome creative blocks and explore new melodic territories, ensuring that the human element remains central to the artistic process (Wiafe & Fränti, 2023).

While the rapid advancement of AI in music composition has made significant strides, it still faces substantial limitations inherent in single-agent methodologies. This research proposes to overcome these limitations through a multi-agent AI framework explicitly designed to emulate the collaborative, adaptable, and versatile nature of human music composition. By rigorously investigating how these systems enhance musical creativity, complexity, and coherence, this study aims to provide critical insights that advance computational creativity, enrich practical applications, and fundamentally broaden the scope of human-machine creative collaborations.

Problem Statement

The problem to be addressed in this study is the inadequacy of existing single-agent AI music composition systems in replicating the depth, creativity, and collaborative dynamics characteristic of human-generated musical compositions. Despite significant advancements, current single-agent AI models often struggle to consistently produce outputs with the complexity, originality, and long-term structural coherence typically observed in music created by humans, frequently lacking emotional depth and consistency of the production (Hernandez-Olivan, Hernandez-Olivan, & Beltran, 2022). This significantly limits their practical utility in various domains (The IoT Academy, 2025). Industries such as entertainment, education, and therapeutic settings rely heavily on nuanced, adaptive, and sophisticated music compositions to engage audiences, facilitate learning, and support therapeutic interventions effectively (Gera, 2025).

There is a significant knowledge gap regarding the capability of multi-agent AI frameworks—systems where multiple specialized agents dynamically interact—to enhance musical complexity, creativity, and coherence. These limitations in single-agent systems stem from their lack of true agentic capabilities, such as autonomous decision-making, adaptive learning, and proactive planning (K2view, 2025). It remains unclear whether collaborative interactions among multiple AI agents, embodying these agentic principles, can overcome the identified limitations of single-agent systems. Without addressing this gap, the broader adoption of AI-driven creative technologies will remain hindered, restricting their potential transformative impact in real-world applications (Jin et al., 2025). Consequently, understanding how multi-agent Agentic AI frameworks could elevate AI-generated music compositions to match human compositional styles and preferences more closely is essential. Failure to explore and address this limitation may perpetuate the use of AI music systems that fail to meet the sophisticated requirements and expectations of practitioners and end-users across critical sectors.

Purpose of the Study

This mixed-method comparative study aims to evaluate how a multi-agent AI framework impacts musical creativity, complexity, and coherence compared to traditional single-agent AI systems. This study is designed explicitly to respond to the identified gap in computational creativity by systematically exploring whether collaborative interactions among specialized AI agents, driven by agentic principles such as autonomous decision-making, adaptive learning, and collaborative goal-seeking, can enhance music composition outputs. The mixed-method design provides a comprehensive evaluation: quantitative methods will measure objective musical properties (e.g., structural coherence, statistical diversity), while qualitative methods, such as expert evaluations from professional musicians, will capture subjective aspects like perceived creativity, emotional depth, and human-like qualities. The variables to be assessed include musical complexity,

creativity, coherence, human compositional alignment, stylistic authenticity, emotional expressiveness, and adherence to human-like musical principles. Data will be analyzed using statistical methods such as ANOVA and paired comparisons. The research will be conducted in generalized geographic locations involving music professionals and academic institutions without compromising participant confidentiality.

The findings are anticipated to significantly advance the field of computational creativity by showcasing the benefits of a multi-agent AI framework, potentially leading to the development of more sophisticated and collaborative AI systems in music composition. Insights from expert evaluations are expected to guide the design of future AI models, enhancing their capacity to create compositions that deeply resonate with human listeners.

Introduction to the Conceptual Framework

The study is grounded in the overarching paradigm of Agentic AI, which provides a robust theoretical foundation for developing intelligent systems capable of autonomous action, complex reasoning, and continuous adaptation in dynamic environments. Within this paradigm, the research integrates established theories of computational creativity (Boden, 2009), multi-agent systems (Wooldridge, 2009), and collaborative AI interactions (Eigenfeldt, Bown, & Carey, 2015). Central to these theories, and particularly relevant to Agentic AI, is the concept of distributed cognition (Hutchins, 1995) which underscores how collective interactions among autonomous agents can lead to emergent properties and enhanced creative outcomes that surpass the capabilities of individual components. This integrated theoretical framework guides all research decisions, including the definition of the problem statement, purpose, and research questions, by providing a coherent rationale for hypothesizing that multi-agent Agentic AI interactions will yield significantly improved creative outputs in music.

Theoretical Foundations and Propositions

The convergence of these theories within the Agentic AI framework yields several key propositions for this study:

- Computational Creativity and Transformational Output: Drawing from computational creativity theory, this study posits that Agentic AI, through its multi-agent architecture, can move beyond "exploratory creativity" (generating variations within established musical styles) towards "transformational creativity" (deliberately altering or transcending existing musical conceptual spaces) (Jordanous, 2025). This is achieved by enabling agents to learn and apply musical rules and strategically break or reinterpret them, fostering genuine novelty and originality in compositions. The framework aims to simulate human creative processes involving extrapolation rather than mere data interpolation (Carnovalini & Rodà, 2020).
- Multi-Agent Systems and Emergent Musical Complexity: Building on multi-agent systems theory, the research proposes that distributing complex compositional tasks among specialized, autonomous agents will lead to a higher degree of musical complexity, coherence, and depth than achievable by single-agent systems (66degrees, 2025). Each agent contributes to a collective objective with its defined musical goal (e.g., a harmony agent aiming for coherent chord progressions, a rhythm agent for intricate patterns). The emergent behavior from their dynamic interactions and communication, akin to musicians in a band or an orchestra guided by a conductor, is hypothesized to produce compositions with sophisticated inter-track dependencies and long-term structural coherence, addressing a critical limitation of current AI music generation (Hernandez-Olivan, Hernandez-Olivan, & Beltran, 2022).
- Distributed Cognition and Enhanced Collaborative Dynamics: The principle of distributed cognition suggests that intelligence and problem-solving are not confined to a single entity

but emerge from the interactions of multiple agents within an environment. In this framework, the collaborative dynamics among specialized AI agents, facilitated by shared memory and communication protocols, will enable a more nuanced and adaptive compositional process (Dadman & Bremdal, 2023). This mirrors human collaborative music-making, where individual musicians contribute their expertise while adapting to the collective musical flow (Ruiz, 2025). The system's ability to continuously learn and refine its approaches based on outcomes and feedback loops, often through reinforcement learning, further enhances its adaptive capabilities.

The Multi-Agent Score Framework: A Conceptual Model

The Figure 1 diagram below depicts the Multi-Agent Score Framework, outlining the process for generating musical scores using Agentic AI principles. This framework is designed to emulate human music composition's sophisticated, multi-faceted nature by orchestrating specialized AI agents.

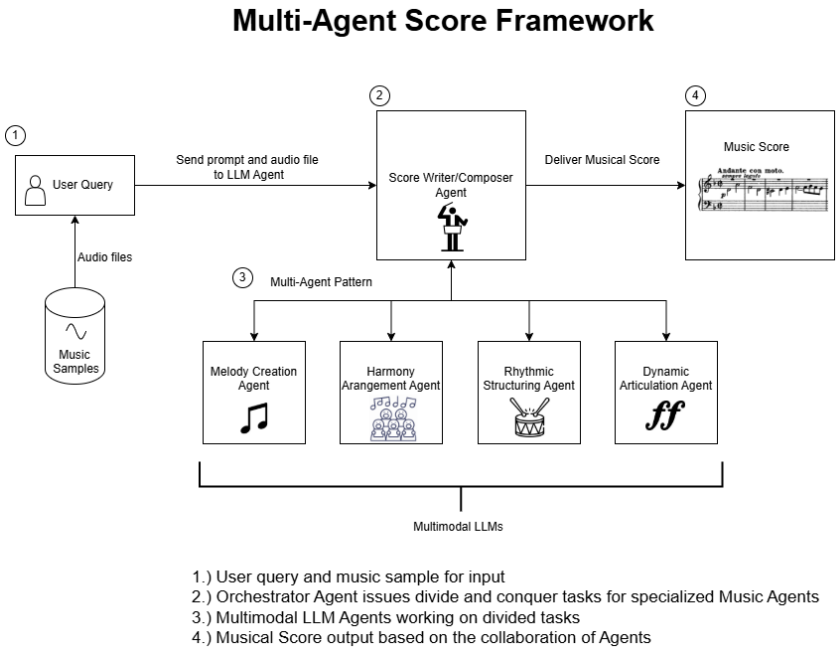

Figure 1. Muti-Agent Score Framework

The framework operates through a series of interconnected stages, each leveraging the autonomous and collaborative capabilities of Agentic AI:

User Query and Music Sample for Input: The process initiates with a user providing a natural language query, accompanied by music samples. This multimodal input is the creative brief, allowing users to articulate their desired musical style, mood, or specific compositional constraints. The ability to accept diverse inputs (e.g., audio, symbolic data, text) is a key architectural consideration for multi-agent frameworks in interactive music generation, enabling richer human-computer interaction.

Score Writer/Composer Agent (Orchestrator): The user's input is directed to the central Score Writer/Composer Agent. This agent functions as the orchestrator or "conductor" of the entire Agentic AI system, embodying high-level task management and strategic decision-making. Its primary role is to interpret complex user queries, decompose them into manageable subtasks, and dynamically assign these "divide and conquer" tasks to specialized music agents. This hierarchical control structure is crucial for handling the complexity inherent in long-form musical composition.

Multi-Agent Pattern (Specialized Music Agents): This layer comprises a team of specialized AI agents, each responsible for a distinct musical domain:

Melody Creation Agent: Focuses on generating captivating melodic lines, adhering to stylistic and emotional parameters derived from the user query.

Harmony Arrangement Agent: Develops coherent chord progressions and harmonic structures that complement the melody and overall musical intent.

Rhythmic Structuring Agent: Creates intricate rhythmic patterns and grooves, ensuring rhythmic coherence and dynamism throughout the composition.

Dynamic Articulation Agent: Manages expressive elements such as dynamics (loudness/softness), articulation (e.g., staccato, legato), and phrasing, adding emotional depth and nuance to the score.

These specialized agents operate autonomously within their defined roles, yet they continuously communicate and coordinate their efforts, often leveraging multimodal Large Language Models (LLMs) for reasoning and interaction. Their collaborative nature, where each agent contributes to a shared musical goal, is central to achieving emergent creativity and overcoming the limitations of single-agent systems in producing complex, coherent compositions. Multimodal LLMs allow these agents to process and generate musical information across various modalities, from symbolic notation to audio features, enhancing their understanding and creative output.

Musical Score Output: Finally, the Score Writer/Composer Agent integrates the cohesive musical components generated by the specialized agents into a complete musical score. This final output represents the culmination of the collaborative effort, delivering a coherent, creatively enhanced, and structurally sound composition back to the user. The emergent quality of this score is a direct result of the multi-agent system's orchestrated interactions and specialized contributions.

Methodology and Design Alignment

This conceptual framework directly informs the study's mixed-method comparative design, which is the most appropriate choice for rigorously evaluating the proposed multi-agent Agentic AI framework. The inherent complexity of musical creativity and the multi-faceted nature of the problem statement—which highlights the inadequacy of single-agent systems in replicating depth, creativity, and collaborative dynamics—necessitate a comprehensive approach that combines quantitative and qualitative methodologies.

Quantitative Methodology: The framework's emphasis on enhancing "complexity" and "coherence" in musical scores will be quantitatively assessed. This involves generating a substantial corpus of musical scores using both the proposed multi-agent Agentic AI framework and selected traditional single-agent deep learning models (e.g., Transformer-based models without multi-agent orchestration, or established rule-based systems) as baselines (Nemade, Babu, & Khan, 2025). Data collection will involve extracting objective musical features from these generated scores, such as harmonic complexity metrics, rhythmic diversity indices, and measures of long-term structural coherence (e.g., using Music Information Retrieval (MIR) techniques for feature extraction and pattern analysis). Statistical methods, including Analysis of Variance (ANOVA) and paired comparisons, will be employed to compare the outputs of the multi-agent system against the single-agent baselines across these objective metrics. This quantitative analysis will provide empirical evidence of the framework's ability to overcome the structural and complex limitations identified in the problem statement.

Qualitative Methodology: To evaluate the more subjective aspects of "creativity" and "human compositional alignment," a qualitative approach is indispensable. This will involve expert evaluations from professional musicians and music theorists. Data collection will utilize

structured surveys and human listening tests, where experts will assess the generated scores for perceived originality, emotional depth, aesthetic appeal, and overall alignment with human compositional styles. This qualitative data will capture the nuanced and often elusive dimensions of musical artistry that quantitative metrics alone cannot fully address, directly responding to the study's purpose of elevating AI-generated music to match human compositional styles and preferences. Integrating human feedback is crucial for refining Agentic AI systems, as they learn and adapt continuously from interactions.

This mixed-method comparative design is optimal because it directly aligns with the problem, purpose, and research questions. The problem statement highlights the limitations of single-agent systems in producing music with human-like depth and creativity; the purpose is to enhance these qualities. By comparing the proposed multi-agent Agentic AI framework against traditional single-agent systems using both objective and subjective measures, the study can rigorously validate whether the collaborative, autonomous, and adaptive nature of Agentic AI truly leads to superior musical outputs. The methodology's focus on empirical validation through diverse data types ensures that the central research question, "How does a multi-agent AI framework enhance the complexity, creativity, and coherence of generated musical scores compared to single-agent systems?" is directly and comprehensively answered. This approach provides a holistic understanding of the framework's impact, bridging the gap between computational capabilities and human artistic perception.

Research Questions

RQ1: To what extent does a multi-agent AI framework improve the measured complexity, creativity, and coherence metrics of generated musical scores compared to a single-agent AI system?
RQ2: How do professional musicians and composers perceive the creative processes and musical qualities of compositions generated by a multi-agent AI framework compared to those produced by a single-agent AI system?

Hypotheses

Null Hypothesis (H_0): No statistically significant difference in measured complexity, creativity, or coherence between musical scores generated by a multi-agent AI framework and those generated by a single-agent AI system.
Alternative Hypothesis (H_a): Musical scores generated by a multi-agent AI framework demonstrate significantly higher complexity, creativity, and coherence metrics than those generated by a single-agent AI system.

Significance of the Study

This research is significant to computational creativity because it addresses a critical gap in AI-driven music composition methodologies, thereby advancing the theoretical framework of multi-agent creative collaboration. By evaluating the effectiveness of a multi-agent Agentic AI framework in music generation, the study will provide empirical evidence and deepen theoretical insights into how autonomous, goal-oriented AI systems can foster emergent creativity and complex musical structures. Specifically, it will explore how Agentic AI can achieve transformational creativity in music, moving beyond mere interpolation to generate genuinely novel and original compositions that challenge existing conceptual spaces. Furthermore, it will contribute to novel models for understanding human-AI co-creation, particularly how the orchestrated interaction of specialized agents can lead to sophisticated and coherent long-form musical outputs, mirroring the complex collaborative processes of human musicians. This includes insights into how Agentic AI can discover new musical grammars, patterns, or

theoretical concepts by autonomously analyzing vast musical corpora, thereby pushing the boundaries of computational musicology.

The expected results could significantly improve how AI is applied in creative domains. The framework's inherent ability to manage long-term musical structure, adapt to nuanced emotional expressions, and continuously learn from interactions, driven by its autonomous and goal-oriented capabilities, translates into tangible benefits across various sectors:

Dynamic Adaptive Training Tools for Music Educators: The framework's capacity for personalized learning experiences, real-time feedback, and the generation of customized exercises will significantly enrich music theory instruction and instrumental practice. The Agentic AI's autonomy and proactive nature will allow it to act as a responsive tutor, adjusting content and pace without constant human prompting, thereby enhancing student engagement and understanding.

Contextually Rich Music Generation for the Entertainment Industry: By leveraging the Google Gemini Flash model for advanced audio processing, the framework will enable real-time analysis of musical input and sophisticated feature extraction, crucial for its ability to perceive and respond dynamically. This will facilitate the production of long-form, emotionally resonant soundtracks and interactive musical experiences that dynamically react to narrative developments or user input, moving beyond generic stock music.

Personalized Adaptive Music Experiences in Therapeutic Settings: The framework's capacity for emotional sensing through multimodal inputs (e.g., voice emotion recognition, facial expression analysis, biometric sensors) and its ability to generate adaptive musical interventions will support highly personalized therapeutic applications for emotional regulation, cognitive engagement, and mental well-being.

Ultimately, by addressing the research problem and answering the study’s questions through integrating Google Gemini Flash, this work is expected to positively shape both the theoretical foundations and the practical applications of AI-assisted creative processes. Furthermore, the findings will contribute to the ongoing discourse on the ethical implications of AI in music, including bias in datasets leading to cultural homogenization, intellectual property and authorship concerns, and potential job displacement. This study will help inform the development of responsible AI that augments human creativity and artistic expression rather than diminishing it.

Scope, Limitations, and Delimitations

Scope of the Study: This research’s scope is deliberately focused on the design, implementation, and evaluation of a multi-agent AI framework for musical score writing, and the comparison of its output to that of a traditional single-agent system. The study encompasses the development of a hierarchical, collaborative system of AI agents, each with a dedicated musical role (melodic composition, harmonic progression, rhythmic patterning, and dynamic articulation, among others). The framework is explicitly applied to symbolic music composition (i.e., producing written musical scores, such as MIDI-based or ABC notation-based outputs, rather than audio waveforms). By concentrating on score generation, the research stays within the domain of music theory and composition, addressing how AI agents decide and organize notes, rhythms, and dynamics. The genre and style of music are defined within the scope to ensure manageability and consistency – for example, the system might be trained and tested on Western classical or contemporary instrumental music, which provides a structured yet rich context for evaluating creativity and coherence. Within this scope, the study also defines the performance metrics and evaluation criteria: it covers objective measurements of musical complexity and coherence (such as tonal diversity, rhythmic variability, and structural repetition analysis) and subjective evaluations of creativity and musical quality by human experts. The mixed-method comparative design is a central part of the scope, meaning the research implements the multi-agent system and devises experiments to compare its outputs against those of single-agent models on the criteria.

Notably, the scope includes formulating a baseline single-agent composition model (a state-of-the-art deep learning model without agentic orchestration) to serve as a benchmark. This ensures that enhancements can be attributed to the multi-agent approach. Thus, the study spans theoretical framework building and system development to empirical evaluation. Still, it does not extend to aspects like real-time interactive music generation or direct integration with human composers in the creative loop – those aspects are acknowledged as interesting, but they lie outside the defined scope. By narrowing attention to how multiple AI agents jointly create music and how that outcome differs from a conventional AI composer, the study aims to produce clear, focused insights on the benefits and behaviors of collaborative AI in composition.

Limitations: Notwithstanding its targeted scope, this research faces several inherent limitations. First, technical and methodological constraints are associated with implementing and coordinating multiple agents. A multi-agent system of this nature introduces considerable complexity: each agent's performance is contingent on its underlying model or rules, and the coordination mechanism (the conductor/orchestrator agent) may not capture all nuances of human collaboration. One limitation is that the system's success partly depends on how effectively the orchestrator can manage the agents' contributions; if this coordination falls short, the promised gains in coherence and creativity might not fully materialize. This reflects a broader limitation in multi-agent AI design – balancing autonomy and control is difficult, and suboptimal tuning could lead to conflicts or redundancies in the generated music. In the current study, the communication and coordination protocols are necessarily simplified for feasibility, which might limit the degree of actual interaction the agents have compared to an ideal scenario. Second, the quality of creativity and musicality is challenging to measure and ensure. While the study uses expert evaluators and various metrics, these approaches have limitations. Expert judgment, for instance, is inherently subjective; what one musician deems highly creative or emotionally moving, another might find less impressive. The sample of experts and the contexts in which they evaluate the music (e.g., listening to isolated AI compositions without the rich context a human composer might provide about intent) could constrain the findings. Additionally, objective metrics for music (such as those derived from music information retrieval techniques) may not fully capture qualities like emotional expressiveness or thematic development. Therefore, there is a limitation in that some aspects of "creativity" might elude quantitative analysis and even qualitative consensus, meaning the study's conclusions can only pertain to the facets of creativity and coherence that these methods can detect. Third, like many studies in AI music, this work is limited by its data and domain. The training and testing datasets (e.g., a collection of classical piano compositions or lead sheets) impose boundaries on what the AI agents can learn and produce. If the dataset lacks diversity, the agents might not generalize well beyond certain styles, thus limiting the apparent creativity of the system's output. Moreover, the compositions generated and analyzed in this study are of a certain length and form (for example, short pieces or single-movement works); the system has not been proven on very long-form compositions such as complete symphonies or operas. Long-form composition remains challenging due to context retention issues. Although the multi-agent design aims to address this via distributed memory and roles, it may still be imperfect in maintaining global coherence over extended pieces. Current AI models often struggle with long-term structure (Hernandez-Olivan, Hernandez-Olivan, & Beltran, 2022). Despite the enhancements we introduce, the system might still exhibit some degree of limitation (e.g., a tendency toward repetitive or meandering sections in longer outputs). Another notable limitation is the computational overhead: running multiple agents (especially if any use of large neural models) can be resource-intensive, restricting the number of experiments or the complexity of models we could practically use. This means that, due to resource limitations, the agents might not have the most advanced architecture possible, potentially capping the performance level they can achieve. Finally, there are creative and cultural limits to consider. AI, whether single or multi-agent, cannot fully grasp

the cultural context or personal meaning in music that human composers contribute to. As a result, even a highly complex AI-composed score might lack the intangible "soul" or authenticity that a human composition might carry. While the study measures perceived human-likeness and expressiveness, it acknowledges that true creativity involves qualities that might remain out of reach for AI; this humbling limitation frames our expectations of what the multi-agent system can ultimately accomplish. In summary, the limitations of this study include the complexity of multi-agent coordination, the challenges of measuring a multifaceted construct like creativity, data and genre constraints, computational resource bounds, and the enduring gap between algorithmic generation and human artistry. These factors necessitate cautious interpretation of results; for instance, if improvements are observed, they will be within the confines of the test scenarios and not guarantee that the AI equals human composers in unrestricted settings.

Delimitations: Alongside inherent limitations, we have made deliberate choices to narrow the study's focus – these delimitations define what the study does and does not cover by design. One key delimitation is the choice of musical scope: the project centers on instrumental score composition, leaving any consideration of lyrics, vocal music, or the added complexities of audio production. By excluding vocals and focusing on instrumental pieces (such as solo piano works or ensemble scores), the study avoids the intricacies of natural language processing for lyrics or the timbral concerns of orchestrating specific instruments' sounds. Similarly, the compositions are constrained to styles (for example, tonal music within genres); the framework is not tested on every possible music genre. This was an intentional choice to ensure that the evaluation remains consistent and fair, comparing pieces within a coherent stylistic range, and to leverage datasets (like MAESTRO or others) that are well annotated and suitable for training the agents in those styles. Another delimitation is the structure of the multi-agent system itself. The research focuses on a specific architectural instantiation: a top-down conductor agent with predefined specialist agents (melody, harmony, rhythm, dynamics). Alternative architectures – such as a purely decentralized agent in society or a different set of musical roles (e.g., a bass line agent or form/structure agent) – are not explored in this work. The roles were selected based on traditional music theory components and prior research, and this choice confines the study to examining this division of creative labor. The implication is that if another important musical aspect (say, timbre or orchestration in terms of instrument assignment) is not explicitly handled by an agent, it falls outside this study's purview. Additionally, the study delimits itself to evaluation by comparing AI baselines and expert judgment, rather than audience reception or industry deployment. For instance, it does not include a live audience test of whether listeners prefer multi-agent compositions over human compositions, nor does it involve integrating the system into a real-world composition workflow for extensive user studies. Those aspects, while valuable, were set aside to maintain a clear focus on the core research question about multi-agent vs single-agent AI performance. In terms of methodology, the sample of human evaluators is intentionally confined to professional musicians and music theorists, instead of general listeners, to ensure informed critique. This choice limits getting high-quality feedback, though the results may reflect expert perspectives more than general audience reactions. Lastly, ethical or socio-cultural considerations (such as the impact of AI on composers' careers or listener attitudes toward AI-generated music) are not deeply probed in this technical study; they are acknowledged in passing but delineated outside the immediate research scope. By establishing these delimitations, the researcher clarified the boundaries: the study homes in on the technical and creative outcomes of a novel AI framework under controlled conditions, without straying into all possible extensions or implications.

Generalizability and Broader Implications: Given the specific scope and the above limitations and delimitations, the generalizability of this study's findings must be carefully assessed. The results and insights will most directly apply to contexts like those studied: AI-

driven symbolic music composition within the style and length ranges tested, using a structured multi-agent architecture. For example, suppose the multi-agent framework enhances musical coherence and creativity for the chosen genre (say, classical pieces). In that case, one can cautiously generalize that a comparable multi-agent approach might yield benefits in other closely related genres or composition tasks that share structural characteristics. However, generalizing beyond the conditions of this research (to very different musical traditions, to real-time interactive improvisation, or audio waveform generation) should be done with reservation. The findings are not a blanket confirmation that "multi-agent is always better" for all creative AI tasks, but rather evidence that collaborative agentic behavior offers advantages over a solitary AI composer under certain conditions. Factors like the need for a central coordinator and the balance between agent specialization and integration could play out differently in other domains. For instance, a multi-agent score-writing framework might need substantial alteration to be effective in genres where timbral innovation or production quality is paramount (like electronic pop production., Likewise, in applications like live jam sessions or adaptive game soundtracks, real-time constraints and unpredictability would test the current system's design limits.

That said, the broader implications of this work are promising for AI and music. If the multi-agent framework demonstrates tangible improvements, it suggests a new direction for AI creativity research: leveraging multiple agents as collaborators rather than seeking one all-encompassing model. This could influence how future systems are built in various creative domains – not just music, but perhaps in collaborative storytelling, art generation, or design – by highlighting the value of distributed, specialized intelligence working towards a common artistic goal. In the context of music composition specifically, the success of this framework could pave the way for more adaptive and interactive composition tools. For example, multi-agent AI systems might generate personalized music in education and entertainment settings, where different agents respond to different inputs or constraints. (Indeed, one can envision educational software in which separate AI tutors for melody and harmony help students experiment with compositions, or video game scoring systems where agents dynamically adjust the music to gameplay events – concepts aligned with the potential identified for tailored soundtracks and adaptive compositions.) Moreover, proving that AI agents can collaborate to achieve transformational creativity (beyond mere style imitation) would have implications for computational creativity theory itself, supporting the idea that constraint and interaction (as opposed to monolithic unconstrained generation) can lead to more inventive outcomes. This research also contributes to discussions on distributed cognition and collective intelligence in creative practice: it provides a concrete demonstration of how creativity can emerge from a network of AI components, which might analogize to human creative teams or symbiotic human-AI collaborations. While the current study keeps humans out of the composition process to isolate the effects of AI-AI collaboration, the findings could also inform human-in-the-loop systems. For instance, knowing which aspects of music benefit most from agent specialization could help build AI assistants that work alongside human composers, each agent handling a different suggestion (melodic idea, harmonic alternative, rhythmic variation) for the human to consider.

In terms of generalizability, it is essential to note that the evidence from this study will be context specific. The multi-agent framework is tested under conditions, and the results will indicate what is achievable within those boundaries. Extrapolation to "broader AI and music composition contexts" should account for differences in data, genre, and objectives. We anticipate that some principles – such as the utility of having dedicated sub-systems for different creative facets, or the importance of a coordinating mechanism – will hold broadly, because they stem from fundamental theories (like multi-agent coordination and cognitive distribution) rather than any one dataset. However, concrete performance gains (e.g., an

inevitable percentage increase in a creativity score) may not translate directly to another scenario without adaptation.

Thus, the research ends with demonstrating what multi-agent AI can do in a bounded scenario and discussing how these insights might inspire further innovations. If the hypothesis is validated, it could encourage more widespread adoption of multi-agent strategies in creative AI, helping overcome some current limitations of single-agent generative models that have slowed their acceptance in professional music settings. Fields like music education, therapy, and interactive entertainment, which demand high adaptability and richness in generated music, stand to benefit if the multi-agent approach produces more compelling compositions. In conclusion, while the study's findings will be most definitively applicable to its experimental context, their significance lies in opening new avenues for theoretical and applied advancements. The scope is narrow enough to allow rigorous investigation, the limitations ensure we interpret outcomes with humility, and the delimitations keep the project feasible – but collectively, these choices aim to yield insights that resonate beyond the lab, pointing toward a future where collaborative AI agents contribute meaningfully to the art of music composition. The work aspires to inform academic understanding and the next generation of AI-powered musical tools, demonstrating the potential when multiple creative minds (even if artificial) work together in harmony.

References

66degrees (2025). *Agentic AI in M&E: Challenges and opportunities*. https://66degrees.com/agentic-ai-in-media-entertainment/

Boden, M. A. (2009). Computer Models of Creativity. *AI Magazine, 30*(3), 23-34. https://web.archive.org/web/20160508042330id_/http://www.aaai.org:80/ojs/index.php/aimagazine/article/viewFile/2254/2100

Brown, E. D. (2025). *Orchestrating AI agents.* Eric D. Brown, D.Sc. https://newsletter.ericbrown.com/p/orchestrating-ai-agents

Carnovalini, F., & Rodà, A. (2020). Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Frontiers in Artificial Intelligence, 3,* 14. 2020 https://doi.org/10.3389/frai.2020.00014

Dadman, S., & Bremdal, B. A. (2023). Multi-agent reinforcement learning for structured symbolic music generation. In *International Conference on Practical Applications of Agents and Multi-Agent Systems* (pp. 52-63). Springer Nature Switzerland. https://munin.uit.no/bitstream/handle/10037/33311/article.pdf?sequence=2&isAllowed=y

de Berardinis, J., Porcaro, L., Merono-Penuela, A., Cangelosi, A., & Buckley, T. (2025). *Towards Responsible AI Music: an Investigation of Trustworthy Features for Creative Systems.* (arXiv:2503.18814). https://arxiv.org/pdf/2503.18814v1

Deng, Q., Yang, Q., Yuan, R., Huang, Y., Wang, Y., Liu, X., … Guo, Y. (2024). *Composerx: Multi-agent symbolic music composition with LLMS.* (arXiv:2404.18081). https://arxiv.org/pdf/2404.18081v1

Eigenfeldt, A., Bown, O., & Carey, B. (2015). Collaborative Composition with Creative Systems: Reflections on the Field. *ACM Computers in Entertainment, 14*(3), 1-17. https://core.ac.uk/download/pdf/196554453.pdf#page=214

Gera, S. (2025). *The Impact of Artificial Intelligence on Music Production: Creative Potential, Ethical Dilemmas, and the Future of the Industry*.

Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C. Z. A., Dieleman, S., Elsen, E., Engel, J., & Eck, D. (2019). *Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. In International Conference on Learning Representations.* https://magenta.tensorflow.org/datasets/maestro

Hernandez-Olivan, C., Hernandez-Olivan, J., & Beltran, J. R. (2022). *A survey on artificial intelligence for music generation: Agents, domains and perspectives*. (arXiv:2210.13944). https://arxiv.org/pdf/2210.13944

Herremans, D., Chuan, C.H., & Chew, E. (2017). A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR), 50*(5), 1-30. https://arxiv.org/pdf/1812.04186

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press. 1995 https://www.ida.liu.se/~729G12/mtrl/Hutchins.pdf

Jin, L., Lin, B., Hong, M., Zhang, K., & So, H. J. (2025). *Exploring the Impact of an LLM-Powered Teachable Agent on Learning Gains and Cognitive Load in Music Education.* (arXiv:2504.00636). https://arxiv.org/pdf/2504.00636

Jordanous, A. (2025). *What is computational creativity?* The Creativity Post. https://www.creativitypost.com/science/what_is_computational_creativity

K2view. (2025, January 10). What is Agentic Ai? A practical guide. https://www.k2view.com/what-is-agentic-ai/

Li, P. P., & Wang, B. (2023). Artificial Intelligence in Music Education. *International Journal of Human–Computer Interaction, 40*(16*), 4183–4192. https://doi.org/10.1080/10447318.2023.2209984*

Miranda, E. (2001). Composing Music with Computers (1st ed.). Routledge. https://doi.org/10.4324/9780080502403

Nemade, M., Babu, S., & Khan, S. (2025). From Tradition to Innovation: A Review of AI Music Generation Models, Datasets, and Evaluation Techniques. *SGS-Engineering & Sciences, 1*(1). https://spast.org/techrep/article/download/5331/486

Priestly, T. (2025). *Thinking of ai agents? You need a conductor!* Hacker Noon. https://hackernoon.com/thinking-of-ai-agents-you-need-a-conductor

Ruiz, J. V. (2025, June 3). *The ART-Official framework for AI in music education*. NAfME. https://nafme.org/blog/the-art-official-framework-for-ai-in-music-education/

The IoT Academy. (2025, February 11). *10 major challenges of Generative AI & How To Overcome Them.* The IoT Academy Blogs - Best Tech, Career Tips & Guides. https://www.theiotacademy.co/blog/challenges-of-generative-ai/

Wiafe, A., & Fränti, P. (2023). Affective algorithmic composition of music: A systematic review. *Applied Computing and Intelligence, 3*(1): 27-43. 2023 doi: 10.3934/aci.2023003 https://www.aimspress.com/article/doi/10.3934/aci.2023003?viewType=HTML

Wooldridge, M. (2009). *An Introduction to Multi-Agent Systems.* John Wiley & Sons.